



Published by Avanti Publishers
**International Journal of Architectural
Engineering Technology**

ISSN (online): 2409-9821



Comparison of Speech Quality and Intelligibility Assessments in University Classrooms

Arkadiy Prodeus^{1,*}, Kateryna Kukharicheva¹ and Maryna Didkovska²

¹Department of Acoustic and Multimedia Electronic Systems, ²Department of Mathematical Methods of System Analysis, National Technical University of Ukraine "Igor Sikorsky Kyiv Polytechnic Institute", Kyiv, Ukraine

ARTICLE INFO

Article Type: Review Article

Keywords:

Speech intelligibility

Early sound reflection

Binaural room impulse response

Timeline:

Received: November 14, 2021

Accepted: December 20, 2021

Published: December 28, 2021

Citation: Prodeus A, Kukharicheva K, Didkovska M. Comparison of Speech Quality and Intelligibility Assessments in University Classrooms. Int J Archit Eng Technol. 2021; 8: 52-60.

DOI: <https://doi.org/10.15377/2409-9821.2021.08.5>

ABSTRACT

Estimates of speech quality and intelligibility for three university classrooms of small, medium and large sizes are presented. The quality and intelligibility of speech were assessed by objective methods using binaural room impulse responses, measured at 5-6 points of the premises. The measures of speech quality were log-spectral distortion (LSD), bark spectral distortion (BSD) and perceptual evaluation of speech quality (PESQ), and the objective measure of speech intelligibility was the speech transmission index (STI).

Among the quality measures considered, only BSD is shown to be highly correlated with STI measures for all three classrooms. In this case, correlation coefficient R varies from minus 0.6 for a small room to minus 0.98 for a large room. The close relationship between PESQ and STI is observed only in the case of a large classroom ($R = 0.96-0.99$), and the LSD measure was found to be uncorrelated with STI for premises of all sizes. The obtained results can serve as a justification for the use of BSD instead of STI, and vice versa, in the acoustic examination of classrooms of different sizes.

*Corresponding Author

Email: aprodeus@gmail.com

Tel: +380-67-232-28-96

1. Introduction

One of the conditions for obtaining a quality education is the good quality and intelligibility of the teacher's speech. However, factors such as noise level and its spectrum features, as well as the features of reverberation in the room, significantly affect the quality and intelligibility of the speaker's speech.

The problem of ensuring good speech intelligibility is especially relevant for students with hearing impairments [1] and primary school students [2]. As for the researches in the field of room acoustics, it was concluded in [3, 4] that there are a number of issues that need further development.

Unlike the acoustics of school classrooms and concert facilities, the acoustics of university classrooms are insufficiently studied. It was suggested in [5] that a possible reason for this situation may be the "intermediate" position of university premises, the sizes of which is larger than ones for school classrooms, but smaller than the sizes of concert halls. However, the size of university classrooms can also vary significantly. Classification by the size of university classrooms was proposed in [6]. According to this classification, there are small, medium and large size premises, with a volume of less than 350 m³, 350-650 m³ and more than 650 m³, respectively. In [7], small, medium and large classrooms include premises with a volume of less than 230 m³, 230-350 m³ and more than 350 m³, respectively.

Estimates of the C50 clarity index and the STI for medium and large classrooms were provided in [8, 9], but small premises were not considered. In [10], speech intelligibility in small and medium-sized university classrooms was assessed by both articulation and instrumental methods. However, large size classrooms were not considered. In addition, the parameters EDT, T30 and C50 were evaluated instead of the STI index.

A possible STI increase to 7-14% near the back wall of the room compared to STI values in the center of the room was indicated in [11, 12]. However, no link was found between speech intelligibility assessments and classroom size, and the relationship between STI and objective measures of speech quality was not analyzed. In [5] these shortcomings have been eliminated partly, but the question of the relationship between STI and objective quality measures remains unclear, which hinders the understanding of the relationship between quality and intelligibility of speech in classrooms of different sizes.

The objective of this paper is to establish a statistical relationship between objective measures of speech quality LSD, BSD and PESQ, on the one hand, and an objective measure of speech intelligibility STI, on the other hand.

2. Research Organization

The research was performed in three stages:

- 1) LSD, BSD and PESQ evaluation;
- 2) STI assessment;
- 3) calculation of correlation coefficients between LSD, BSD and PESQ, on the one hand, and STI, on the other hand.

The binaural room impulse responses (BRIR) of small (177 m³), medium (270 m³) and large (370 m³) classrooms were used for the research. Photos of these rooms are shown in Fig. (1), and their features are given in [5]. The plans of these rooms and locations of sound sources and artificial heads with a couple of microphones are shown in Fig. (2).

Classrooms №1 and №2 are located in the educational building №12 of the National Technical University of Ukraine "Igor Sikorsky Kyiv Polytechnic Institute" (Ukraine). The technology of BRIRs measuring in these classrooms is described in [5]. A self-manufactured artificial head and audio equipment of non-professional level quality were used to measure the BRIRs of these rooms. These are the household active speaker Genius SP-HF 2.0

500 (Taiwan), omnidirectional condenser measuring microphones Superlux ECM-999 (Taiwan), and outside sound card Superlux ECM-999 Steinberg UR242 (Germany-China). According to ISO 3382-1 [16], the sound source was placed at the height of 1.5 m, and microphones were placed at the height of 1.2 m during measuring in all rooms.



Figure 1: Photos of small (a), medium (b) and large (c) rooms.

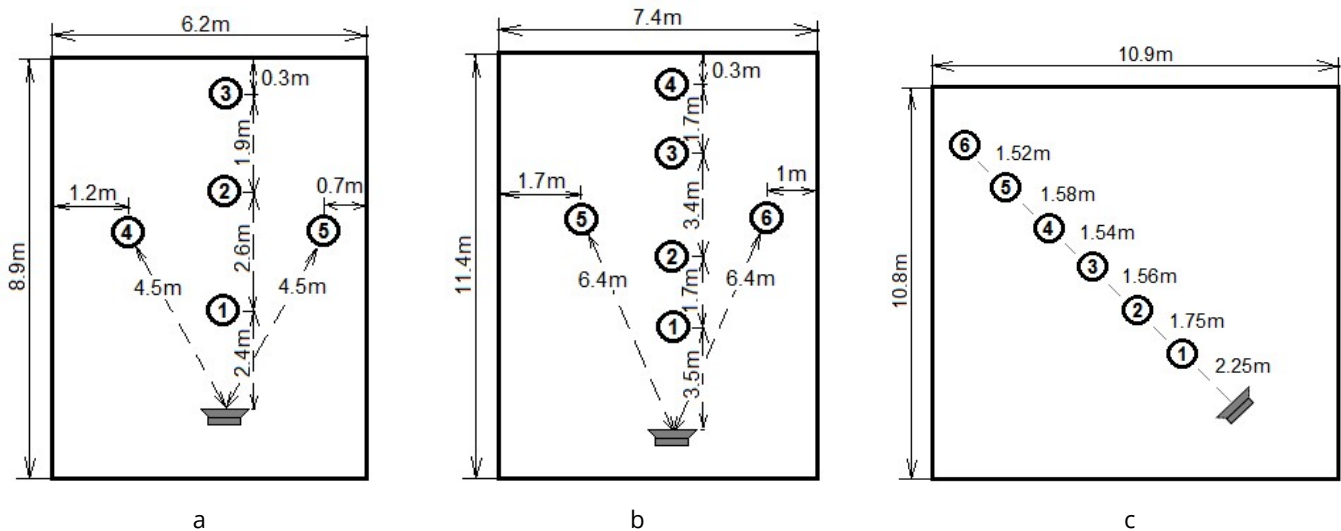


Figure 2: Plans of small (a), medium (b) and large (c) rooms.

BRIRs records for classroom №3 were taken from the Aachen Impulse Response Database [13]. The technology of estimation of these BRIRs is described in [14].

An mls-sequence having the length of 2^{16} samples, which corresponds to the duration of the signal of 1.49 s with the sampling rate of 44.1 kHz was used for small and middle-size classrooms. A similar mls-sequence was used for large classrooms, but its duration was 1.36 s because of the sampling rate of 48 kHz. This mls-sequence was repeated 17 times during radiation. When calculating the BRIRs of premises, the last 16 bursts of the BRIR score were averaged to increase the signal-noise ratio by 12 dB.

The technology of correction of the frequency response of the measuring path of the artificial head system developed, proposed in [16], was used. This correction is necessary because the amplitude-frequency response of the loudspeaker and the microphone are not perfectly uniform in the frequency band of analysis. It was shown in [16] that such correction can be made by controlled dividing the frequency response of the loudspeaker-room-microphone system by the previously obtained estimate of the amplitude-frequency response of the loudspeaker-microphone subsystem. The problem with such calculations is the division operation, because the amplitude-frequency response of the loudspeaker-microphone subsystem may contain small numerical values. However, it is clear that if proper control over the amplitude-frequency characteristics of the loudspeaker-microphone subsystem is ensured, such a division can be practically implemented. Since the amplitude-frequency response of

the loudspeaker-microphone subsystem takes the smallest values at the edges of the frequency range, and the variance of the evaluation of the mutual spectrum of the system "loudspeaker-room-microphone" is the largest on the right edge of the frequency range, it is advisable to apply the method of regularization to achieve necessary calculation accuracy. The role of the regularizing factor can be played by a spectral weighting window whose values are close to one at low and medium frequencies, and the values of this window are close to zero with the approaching to the right edge of the frequency range. The regularization parameter is thus the width of such a weighting window. The Hann (Hanning) window have been used as a regularizing factor and the width of the Hann window was close to 80% of the total frequency range analyzed. As a result of these actions, the sidelobe level of the impulse response of the loudspeaker-microphone subsystem was reduced to minus 44-45 dB, which is only 3-4 dB higher than the theoretically achievable value.

In assessing the LSD, BSD and PESQ parameters, the records of 4 female and 4 male speakers were used, who read the same text of legal content. The length of the analyzed records was close to 20 s, the sampling frequency of the records was 22050 Hz, bit depth 16 bits.

Processing of these records was performed as follows:

- 1) after resampling the pure signal $x(t)$ from 22050 Hz to 44100 Hz or 48000 Hz, it was convoluted with the BRIR $h(t)$, resulting in the formation of a signal $y(t) = x(t) \otimes h(t)$, where \otimes is the convolution symbol;
- 2) the values of quality measures LSD, BSD and PESQ were calculated.

Estimates of the LSD measure were calculated using signals $x(t)$ and $y(t)$ amplitude spectra

$$LSD = \frac{2}{RM} \sum_{m=1}^M \sum_{j=1}^J |G\{X(j, m)\} - G\{Y(j, m)\}| \quad (1)$$

$$G\{X(j, m)\} = \max\{20 \lg(|X(j, m)|), \delta\}, \quad \delta = \max_{l, k} \{20 \lg(|X(j, m)|)\} - 50,$$

$X(j, m)$ and $Y(j, m)$ are discrete Fourier transforms of the signals $x(n)$ and $y(n)$ frames, respectively, j is frequency sample number, J is a number of frequency samples.

Estimates of the BSD measure

$$BSD = \frac{\sum_{m=1}^M \sum_{k=1}^K [B_x(k, m) - B_y(k, m)]^2}{\sum_{m=1}^M \sum_{k=1}^K [B_x(k, m)]^2} \quad (2)$$

were $B_x(k, m)$ and $B_y(k, m)$ are bark spectra of the m th frame of the signals $x(n)$ and $y(n)$, respectively, k is critical band number.

In this paper, a broadband version of the PESQ measure (ITU-T standard P.862.2) [15] was used. The description of the algorithm for calculating the PESQ measure is very cumbersome.

Estimation of STI was performed according to (3), after which Pearson correlation coefficients between LSD, BSD and PESQ, on the one hand, and STI, on the other hand, were calculated.

$$STI = \sum_{k=1}^7 \alpha_k \cdot MT_k - \sum_{k=1}^6 \beta_k \cdot \sqrt{MT_k \cdot MT_{k+1}}, \quad (3)$$

$$MT_k = \frac{1}{14} \sum_{i=1}^{14} T_{ki}, \quad T_{ki} = \begin{cases} 0, & E_{ki} < -15; \\ (E_{ki} + 15)/30, & -15 \leq E_{ki} \leq +15; \\ 1, & E_{ki} > +15, \end{cases}$$

$$E_{ki} = 10 \lg \frac{m_{ki}}{1 - m_{ki}}, \quad m_{ki} = \left| \int_0^{\infty} h_{rk}^2(t) \exp(-j2\pi F_i t) dt \right| / \int_0^{\infty} h_{rk}^2(t) dt,$$

where α_k are weight coefficients, β_k are redundancy coefficients, $h_{rk}(t)$ are the result of filtration of $h_r(t)$ with k th band octave filter, $F_i = 0,63-12,5$ Hz. Filtration of $h_r(t)$ was performed using 7-octave filters with center frequencies from 125 Hz to 8 kHz [27].

This signal processing was performed for the BRIRs corresponding to the locations of the artificial head at different points in each of the rooms.

3. Research Results

The results of LSD, BSD and PESQ estimating are shown in Figs. (3, 4 and 5), respectively.

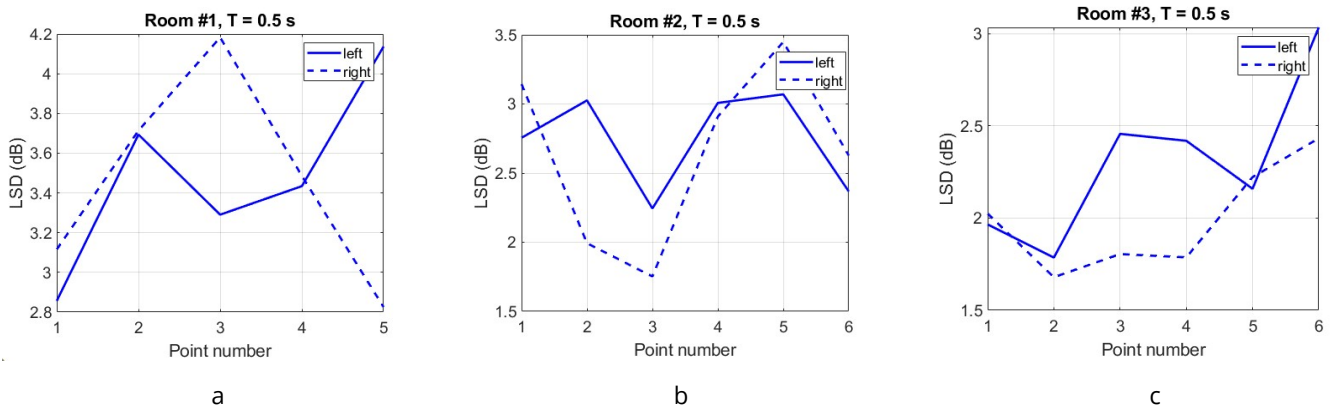


Figure 3: LSD estimates for small (a), medium (b) and large (c) rooms.

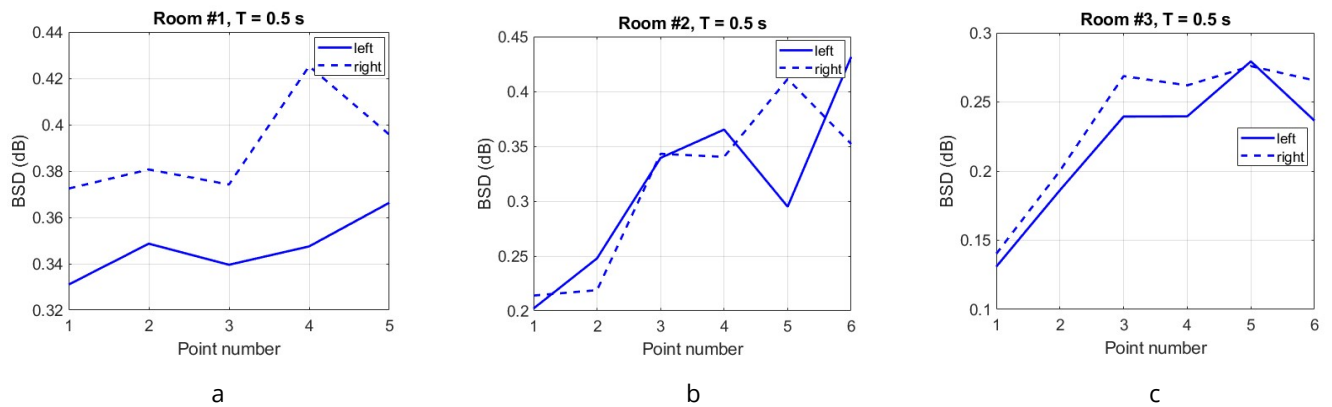


Figure 4: BSD estimates for small (a), medium (b) and large (c) rooms.

The results of the STI assessment are shown in Fig. (6).

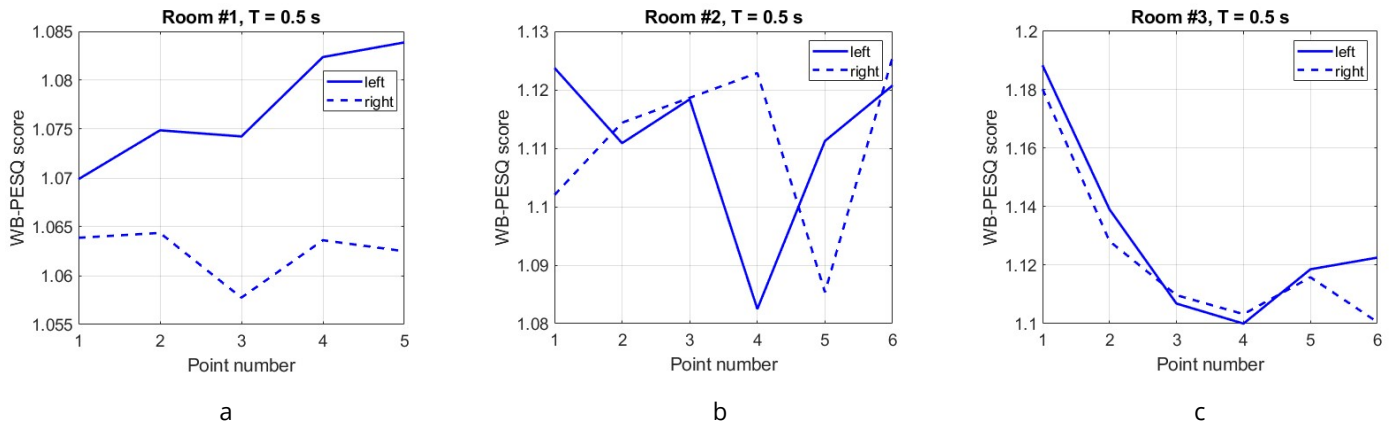


Figure 5: PESQ estimates for small (a), medium (b) and large (c) rooms.

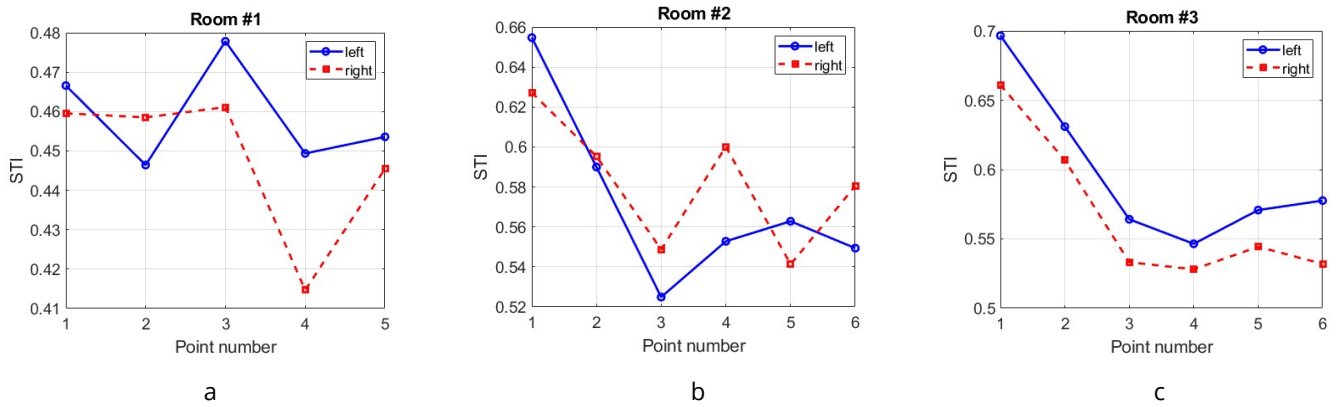


Figure 6: STI estimates for small (a), medium (b) and large (c) rooms

A common feature presented in Fig. (6) results of STI assessment is a decrease in speech intelligibility in the middle of the audience, as well as a slight increase in speech intelligibility near (approximately 30 cm) the back wall of the room. The decrease in speech intelligibility in the middle of the room can be explained by the decrease in the signal-to-noise ratio with the increase in the distance between the source and the receiver of the sound. The increase in speech intelligibility at the back wall of the room can be explained by the addition of the energies of the direct sound and the sound reflected from the back wall. Given the relationship between speech quality and intelligibility, it would naturally be expected that the behavior of objective measures of quality would be similar to that of an STI. Since it is difficult to visually discern such a similarity, it is advisable to use Pearson's correlation coefficient as a measure of such similarity.

The results of estimating the correlation coefficients between LSD, BSD and PESQ, on the one hand, and STI, on the other hand, are given in Table 1 and in Fig. (7).

Table 1: Values of correlation coefficients R .

	LSD-STI		BSD-STI		PESQ-STI	
	L	R	L	R	L	R
room 1	-0,56387	0,216488	-0,56866	-0,98966	-0,54194	-0,30362
room 2	0,366821	0,145747	-0,80145	-0,79546	0,301594	0,249534
room 3	-0,58519	-0,1982	-0,92239	-0,98263	0,985637	0,957655

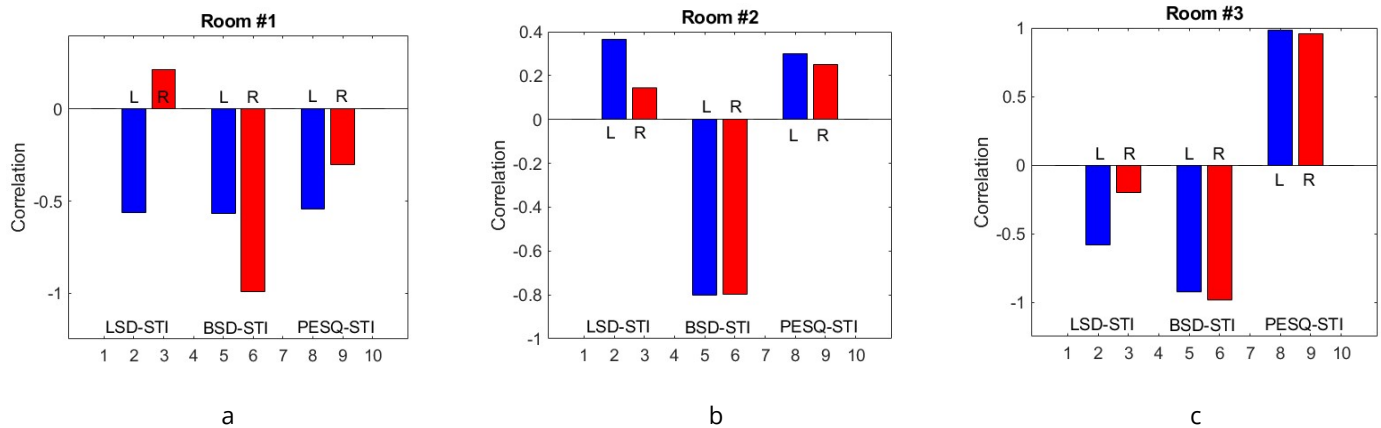


Figure 7: Correlation coefficients for small (a), medium (b) and large (c) rooms.

As follows from the presented results, the values of the correlation coefficient are different for different quality measures and for the channels of the measuring system. The difference in the values of the correlation coefficient in the left and right channels of the measuring system can be explained by the shielding effect of the artificial head. The behavior of the values of the correlation coefficient for various measures of quality in rooms of different sizes is of much greater interest.

4. Discussion

Presented in Table 1 and in Fig. (7) results of the evaluation of correlation coefficients R indicate that there is a close correlation between BSD and STI estimates for classrooms of all sizes. Interestingly, this correlation increases with the size of the room. Indeed, if for a small room it takes R values from minus 0.57 (left channel) to minus 0.99 (right channel), then for a medium-sized room the R values are close to minus 0.8, and for a large classroom, they reach minus 0.92 -0.98. As for the low R values for a small room, they can be explained by the low absorption capacity of the rear wall of the room, the upper half of which is glass. As a result, strong early reflections are formed, which lead to low intelligibility (Fig. 6a) and low quality of speech signals (Fig. 3a, 4a, and 5a).

The close relationship between PESQ and STI is observed only in the case of a large classroom (Fig. 7c), and the LSD measure was found to be uncorrelated with STI for premises of all sizes (Fig. 7).

The obtained results testify to the fundamental possibility of using BSD as a measure of speech intelligibility in classrooms of different sizes, where the effect of reverberation outweighs the effect of noise. The use of PESQ as a measure of speech intelligibility is also possible, but, as can be seen, such use is limited to the case of large premises. Finally, the LSD proved unsuitable for the role of a speech intelligibility measure.

Given the limited amount of statistics, the results obtained in this paper should be considered as indicative of certain trends. Obviously, the amount of statistics should be increased to obtain reliable statements about the existence of clear patterns, as well as the reasons for their existence.

An adaptive measure termed speech to reverberation modulation energy (SRMR) was proposed in [17] for non-intrusive quality diagnosis of reverberant and dereverberated speech. Given the importance of non-intrusive measures of quality, it is worthwhile for future research to investigate their relationship to speech intelligibility as well.

It was shown also in [18] that when listening in noise, modification performance on improving intelligibility is more important than its potential negative impact on speech quality. However, when listening in quiet or at SNRs in which intelligibility is no longer an issue to listeners, the impact on speech quality due to modification becomes a concern. In this regard, it seems advisable to repeat such experiments with the only difference that reverberation is used instead of noise. And in the future, it is logical to consider the combination of noise and

reverberation as interfering factors. It is possible that the results obtained in [19, 20] can serve as a starting point for such studies.

5. Conclusion

The Pearson correlation coefficients between estimates of objective measures LSD, BSD and PESQ of the speech signals quality, on the one hand, and estimates of objective speech intelligibility measure STI, on the other hand, were calculated for small, medium and large classrooms. It is shown that for the BSD-STI pair the correlation is high in premises of all sizes, and for the PESQ-STI pair, the correlation is high only for large rooms. Regarding the LSD measure, its estimates are weakly correlated with those for STI for premises of all sizes.

The obtained results can be used in the acoustic examination of university classrooms. However, due to the lack of statistics, the results obtained should be considered preliminary.

References

- [1] Hu Y, Kokkinakis K. Effects of early and late reflections on intelligibility of reverberated speech by cochlear implant listeners. *J Acoust Soc Am.* 2013; 135(1): <https://doi.org/10.1121/1.4834455>
- [2] Yang W, Bradley J. Effects of room acoustics on the intelligibility of speech in classrooms. *J Acoust Soc Am.* 2009; 125(2): 1-12. <https://doi.org/10.1121/1.3058900>
- [3] Bradley J. Review of objective room acoustics measures and future needs. *Appl Acoust.* 2011; 72(10): 713-720. <https://doi.org/10.1016/j.apacoust.2011.04.004>
- [4] Arweiler I, Buchholz J, Dau T. Speech intelligibility enhancement by early reflections. *ISAAR 2009: Binaural Processing and Spatial Hearing, 2nd International Symposium on Auditory and Audiological Research, Elsinore, Denmark 2009*; Available at: <https://proceedings.isaar.eu/index.php/isaarproc/article/view/2009-29>
- [5] Prodeus A, Didkovska M. Assessment of speech intelligibility in university lecture rooms of different sizes using objective and subjective methods. *E-Europ J Enterprise Technol.* 2021; 35(111): 47-56. <https://doi.org/10.15587/1729-4061.2021.228405>
- [6] Leccese F, Rocca M, Salvadori G. Fast estimation of Speech Transmission Index using the Reverberation Time: Comparison between predictive equations for educational rooms of different sizes. *Appl Acoust.* 2018; 140: 143-149. <https://doi.org/10.1016/j.apacoust.2018.05.019>
- [7] Nestoras C, Dance S. The Interrelationship Between Room Acoustics Parameters as Measured in University Classrooms Using Four Source Configurations. *Build Acoust.* 2013; 20(1): 43-54. <https://doi.org/10.1260/1351-010X.20.1.43>
- [8] Eldakdoky S. Optimizing acoustic conditions for two lecture rooms in Faculty of Agriculture, Cairo University. *Ain Shams Eng J.* 2017; 8: 481-490. <http://dx.doi.org/10.1016/j.asej.2016.08.013>
- [9] Duran S, Ausiello L, Battaner-Moro J. Acoustic Design Criteria for Higher-Education Learning Environments. *Proc Inst Acoust.* 2019; 41(3): 1-12. Available at: <https://pure.solent.ac.uk/en/publications/acoustic-design-criteria-for-higher-education-learning-environmen>
- [10] Choi Y-J. The intelligibility of speech in university classrooms during lectures. *Appl Acoust.* 2020; 162: 107211. <https://doi.org/10.1016/j.apacoust.2020.107211>
- [11] Prodeus A, Didkovska M, Motorniuk D, Dvornyk O. The Effects of Noise, Early and Late Reflections on Speech Intelligibility. *Proc. IEEE 40th Int. Conf. on Electronics and Nanotechnology (ELNANO`2020)*, Kyiv, Ukraine 2020; 488-492. <https://doi.org/10.1109/ELNANO50318.2020.9088854>
- [12] Prodeus A, Didkovska M. Objective assessment of speech intelligibility in small and medium-sized classrooms. *Proc. IEEE Int. Scientific-Practical Conf. on Problems of Infocommunications, Science and Technology (PIC S&T`2020)*, Kharkiv, Ukraine 2020; Available at: https://www.researchgate.net/publication/347490796_Objective_Assessment_of_Speech_Intelligibility_in_Small_and_Medium-Sized_Classrooms
- [13] Aachen Impulse Response Database. Available at: <https://www.iks.rwth-aachen.de/en/research/tools-downloads/databases/aachen-impulse-response-database/>
- [14] Jeub M, Schäfer M, Vary P. A binaural room impulse response database for the evaluation of dereverberation algorithms. *Int. Conf. Proc. on Digital Signal Processing (DSP)*, Santorini, Greece 2009. <https://doi.org/10.1109/ICDSP.2009.5201259>
- [15] Perceptual Evaluation of Speech Quality (PESQ) ITU-T Recommendations P.862, P.862.1, P.862.2. Version 2.0 - October 2005.
- [16] Dvornyk O, Motorniuk D, Didkovska M, Prodeus A. Artificial Software Complex "Artificial Head". Part 1. Adjusting the Frequency Response of the Path," *Microsystems, Electronics and Acoustics*, 2020' 22(1): <https://doi.org/10.20535/2523-4455.me.198431>
- [17] Falk T, Zheng C, Chan W-Y. A Non-Intrusive Quality and Intelligibility Measure of Reverberant and Dereverberated Speech. *IEEE Transactions on Audio, Speech, and Language Processing*, 2010; 18(7): <https://doi.org/10.1109/TASL.2010.2052247>
- [18] Tang Y, Arnold C, Cox TJ. A Study on the Relationship between the Intelligibility and Quality of Algorithmically-Modified Speech for Normal Hearing Listeners. *J Otorhinolaryngol Hear Balance Med.* 2018; 1(1): 5. <https://doi.org/10.3390/ohbm1010005>

- [19] Dong H, Lee C. Speech intelligibility improvement in noisy reverberant environments based on speech enhancement and inverse filtering. *J Audio Speech Music Proc.* 2018; 3. <https://doi.org/10.1186/s13636-018-0126-8>
- [20] Prodeus A, Didkovskiy V. Objective estimation of the quality of radical noise suppression algorithms. *Radioelectronics and Communications Systems*, 2016; 59(11): 502-509. <https://doi.org/10.3103/S0735272716110042>.