

Акустичні прилади та системи

UDC 004.934

DOI: [10.20535/2523-4455.2018.23.3.130367](https://doi.org/10.20535/2523-4455.2018.23.3.130367)

Automated Subjective Assessment of Speech Intelligibility in Various Listening Modes

A. M. Prodeus, Dr.Sc.(Eng.) Prof., ORCID [0000-0001-7640-0850](https://orcid.org/0000-0001-7640-0850)

e-mail aprodeus@gmail.com

K. V. Bukhta, ORCID [0000-0002-6234-7937](https://orcid.org/0000-0002-6234-7937)

e-mail k.kangrem@gmail.com

P. V. Morozko, ORCID [0000-0001-9366-812X](https://orcid.org/0000-0001-9366-812X)

e-mail p77morozko@gmail.com

O. V. Serhienko, ORCID [0000-0002-4212-1770](https://orcid.org/0000-0002-4212-1770)

e-mail sfeara86@gmail.com

I. V. Kotvytskyi, ORCID [0000-0003-2800-0080](https://orcid.org/0000-0003-2800-0080)

e-mail igorktvzk@gmail.com

O. O. Dvornyk, ORCID [0000-0003-4735-2225](https://orcid.org/0000-0003-4735-2225)

e-mail alexanderdvornyk@gmail.com

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

Abstract—In this paper, the results of automated subjective assessment of Ukrainian speech intelligibility are presented. Speech monosyllables of the consonant-vowel-consonant (CVC) type were listened in two modes: through headphones and through acoustic monitors. The assessment was carried out with the help of specially developed software that allowed automating of articulation tests. Speech listening was done for four situations: pure language; speech distorted by noise; speech distorted by reverberation; speech distorted by the combined effect of noise and reverberation. In the first case, speech monosyllables of 3 articulation tables were listened, each of which contained 50 monosyllables. In the second case, speech distorted by the additive noise with the signal-to-noise ratios (SNR) varied in the range $-15...+10$ dB was listened. In this case, models of white, pink and brown noises were used, the masking properties of which are rather well-studied. In the third case, the reverberant speech for reverberation times in the range $0.3...2.7$ s was modeled by convolution of pure speech signals with room impulse responses (RIRs) of various rooms, and in the fourth case the joint action of pink noise and reverberation was considered. It turned out that the masking ability of white noise exceeds one for brown noise for SNR less than minus 5 dB, which is not entirely consistent with preliminary predictive estimates. In addition, it turned out that listening to speech distorted by noise through acoustic monitors could lead to a significant increase in the speech intelligibility, compared to the case of listening through headphones. The analysis of possible causes of abnormal increase in speech intelligibility has been carried out. Early reflections, presence of two loudspeakers, binaural listening, psychophysical features of listeners, as well as peculiarities of software and articulatory testing organization were considered as possible reasons of the phenomenon. After correction of the software and some features of articulation tests it turned out that the results of the speech intelligibility estimation almost coincide when listening to the signals through the headphones and through acoustic monitors, if the distance between the listener and acoustic monitors does not exceed 0.6-0.8 meters. At the same time, these corrections did not differ in the behavior of the dependencies of speech intelligibility on the SNR for small (less minus 5 dB) SNR values. The general conclusion may be that listening to speech signals distorted by noise and reverberation interferences, performed with the application of the proposed automated system of articulation tests, indicates the performance and high quality of the developed system.

Ref. 13, fig. 7.

Keywords — automation of articulation tests; speech intelligibility; noise interference; reverberation interference; listening mode.



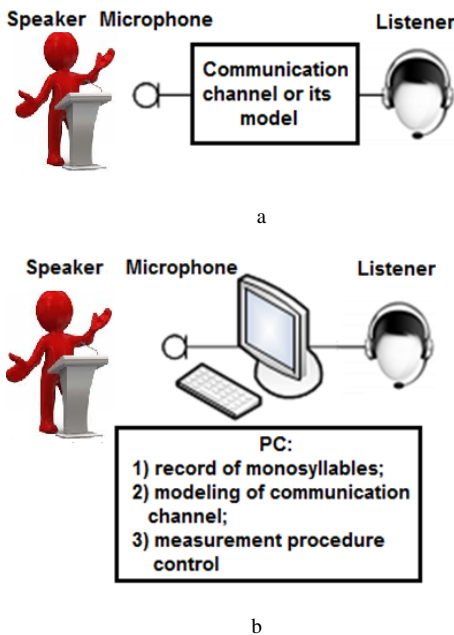


Fig. 1 Articulatory testing schemes: traditional (a) and with the use of PC (b)

I. INTRODUCTION

Speech intelligibility assessment is an important issue because high level of speech intelligibility need be ensured in communication channels, auditoriums and concert halls. At the same time, speech intelligibility need be low in rooms which are neighboring to meeting rooms [1-5].

Today, there are two approaches to measuring the speech intelligibility in communication channels: subjective and objective (instrumental) [1], [2].

Recently, due to the saving of time, financial and human resources, considerable attention is paid to the objective approach that allows automating the measurement procedure. However, despite the disadvantages of the subjective approach, it continues to be used, since the results obtained with it are necessary for the calibration of objective measurement systems.

Therefore, the task of automation of articulation tests (Fig. 1) is urgent, which will allow to speed up and facilitate the procedure of such tests as much as possible.

Russian standard GOST R 50840-95 involves the use of computer technology and regulates articulation testing according to the scheme of fig. 1b. Unfortunately, Ukrainian engineers can not use this standard in the absence of appropriate software and therefore have to use the outdated standard of the USSR GOST 16600-72.

The creation and testing of the articulation tables of Ukrainian words [3], [4] can be considered as a step towards the creation of a national standard that meets modern requirements. One of the shortcomings of [3], [4] is the lack of proposals for automating the procedure for articulation tests. Another disadvantage is the use of word tables, since it is known that the use of monosyllables tables provides higher reliability of tests [5]. It should also be noted that it was not sufficiently correct to listen

to words distorted by noise through computer speakers, since the results obtained in this way can be influenced by the features of the rooms in which the tests were conducted. Besides, there was no attempt to automate the procedure of subjective evaluation of the speech intelligibility in [3] and [4]. The disadvantages of works [3], [4] include also the uncertainty of the nature of the colored noise generated by the generator ANG-2200 [6]. Meanwhile, the question of choosing the type of masking colored noise is important, because there is no clear answer regarding the masking capabilities of white noise, compared with pink and brown noises, for low SNR values. In [7], an attempt was made to investigate this issue by assessing the quality of speech signals distorted by colored noise. However, the results obtained in this way are also not final, since the quality and intelligibility of speech signals are closely related but not identical concepts.

To eliminate these shortcomings, monosyllable articulation tables of the CVC type of the Ukrainian speech were proposed and a set of computer programs for the automation of articulation tests has been created [8]. It has been shown in [8] that listening to signals through acoustic monitors (computer speakers) can lead to much more high speech intelligibility than when listening through headphones. For example, in the case of acoustic monitors usage and noise interference action with SNR less than minus 5 dB, speech intelligibility was close to 80-90%. At the same time, speech intelligibility was close to 10-30% for headphones case. A similar situation was observed in a situation of overwhelming reverberation. Thus, in the case of reverberation action, speech intelligibility was close to 94% (instead of 65% for headphones) for 2.7 s reverberation time. These results can be explained partially by the action of early reflections in rooms [12, 13, 14]. At the same time, such a significant increase in the speech intelligibility can not be explained solely by the action of early reflections.

The object of this paper is, firstly, to highlight the peculiarities of the organization of experimental studies, which were not sufficiently reflected in [8], and secondly, to make in-depth analysis of possible causes of above-mentioned phenomenon of abnormally high speech intelligibility growth.

II. TEST ORGANIZATION

Articulatory tables are an important element of articulation tests, on which the reliability of the result depends substantially. A fragment of articulation tables of standard 50840-95 intended for an unauthorized variant of articulation tests is shown in Fig. 2a. A similar piece of tables for an automated version of articulation tests is shown in Fig. 2b. As can be seen, the possibility of uncertain perception by the listener of individual phonemes is provided in both types of tables.

Given this feature, the tables of monosyllables from the standard 50840-95 were taken as the basis and nine prototypes of articulation tables were developed within the framework of the above project [8], where the possibility of uncertain perception of individual phonemes was taken into account by presenting each table in two variants (Fig. 3). As can be seen from the comparison of Figs.



2 and 3, the proposed articulation tables were obtained by adapting into Ukrainian language the Russian-language tables, which are intended for a non-automated variant of articulation tests.

Note that the articulation tables thus obtained should be considered as prototypes, since their creation does not fully take into account the phonetic features of the Ukrainian language. Thus, in particular, the difference between frequency characteristics of phonemes of Ukrainian and Russian languages was not taken into account, as well as the peculiarities of special phonemes as дж 'dzh' and дз 'dz'.

In the experimental tests of these tables and a set of computer programs, both clear (non-distorted) speech signals and signals distorted by noise and reverberation were used.

The records of clear speech signals were made in a muffled room of the Department of Acoustics and Acoustoelectronics of the National Technical University of Ukraine "Igor Sikorsky Kyiv Polytechnic Institute". The Superlux ECM 999 microphone, the external PRESONUS AudioBox USB sound card and the Audacity version 2.1.3 audio editor and recorder were used for recording. Verbal environment was used when reading monosyllables. For example, the рок /rok/ monosyllable was read as "Запишіть рок тепер" ("Write down rok now"). The recording was performed at 16 bits bit depth and 44100 Hz sampling rate.

Listening was done for four situations:

- Clear speech;
- Speech distorted by noise;
- Speech distorted by reverberation;
- Speech distorted by the combined action of noise and reverberation.

заль	фіц	гет(д)ь	ра(б)пъ
ши(ы)н	чтаф(в)	буш(жь)	прѣт(д)

a

__як/ __ак
__юль/ __уль
__юш/ __уш
__ят/ __ат
__вум

b

Fig. 2 Fragments of the standard 50840-95 tables, intended for non-automated (a) and automated (b) variants of articulation tests

заль	фіц	геть	рап
шин	тав	буш	прьот

a

заль	фіц	хеть	раб
шин	таф	буш	прьод

b

Fig. 3 First (a) and second (b) fragments of two variants of one of the developed articulation tables

In the first case, listeners listened intermittently to sound recordings of 3 articulation tables, each of which contained 50 monosyllables.

In the second case, monosyllables distorted by the additive noise with the SNR of -10 dB, 0 dB and +10 dB were listened. Models of white, pink and brown noises were used in this case. The masking properties of these noises are considered as rather well-studied [1].

In the third case, the reverberant speech for reverberation times from 0.3 to 2.7 s was simulated by convolution of pure language signals with impulse characteristics of different rooms, and in the fourth case the case of joint action of pink noise and reverberation was considered.

The simulation of the signals distorted by the additive noise and reverberation was performed according to the general algorithm:

$$y(t) = x(t) \otimes h(t) + n(t) \quad (1)$$

where $x(t)$ is speech signal, $h(t)$ is RIR, $n(t)$ is additive noise, \otimes is convolution symbol. In the particular case of the action of an exclusively noise interference, the algorithm (1) can be simplified:

$$y(t) = x(t) + k \cdot n(t), \quad (2)$$

where $k = 10^{0.05(SNR-SNR_0)}$, $SNR = 10 \lg(D_x/D_n)$ is the "initial" signal-to-noise ratio for the clean signal $x(t)$ and the generated "initial" noise $n(t)$, D_x and D_n are the variances of the signal $x(t)$ and noise $n(t)$, respectively, SNR_0 is the desired SNR for the mixture (2).

In the case of a compatible noise and reverberation action, algorithm (1) is somewhat complicated in order to take into account the peculiarities of the RIR structure:

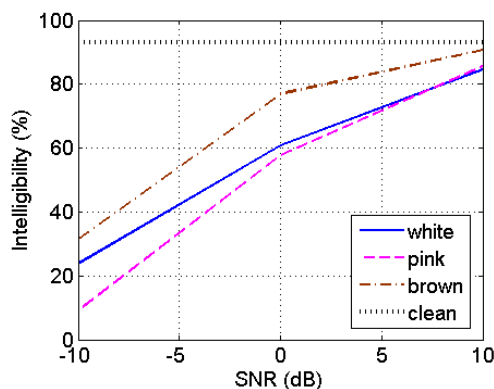
$$h(t) = h_e(t) + h_l(t) \quad (3)$$

$$h_e(t) = \begin{cases} h(t), & t \in 0 \dots 50 \text{ ms}; \\ 0, & t \notin 0 \dots 50 \text{ ms}, \end{cases}$$

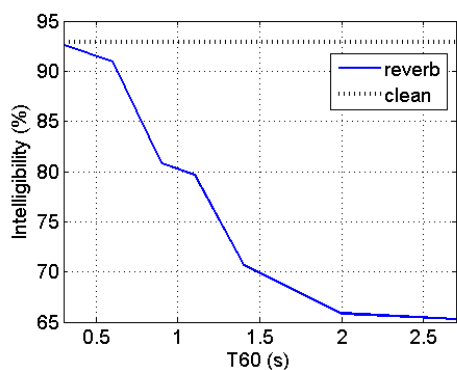
$$h_l(t) = \begin{cases} h(t), & t > 50 \text{ ms}; \\ 0, & t \leq 50 \text{ ms}, \end{cases}$$

where $h_e(t)$ is the initial part of the impulse response $h(t)$ but $h_l(t)$ is a tail of $h(t)$. Then approximately "initial" $SNR = 10 \lg D_{x_e}/D_n$, where D_{x_e} is a variance of the signal $x_e(t) = x(t) \otimes h_e(t)$. The parameters SNR_0 and k are then calculated in a manner analogous to that used in (2).

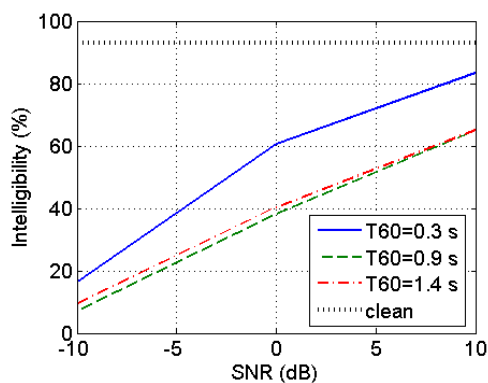
In the process of articulation testing, the students had to fix, using a computer keyboard, perceived monosyllables. This method of fixing the perceived information is fundamentally different from that offered in standard 50840-95. This is done deliberately in order to maximally approximate the procedure and the results of articulation tests to the non-automated test version, which is appropriate to consider benchmark in view of its original nature.



a



b



c

Fig. 4 Estimates of intelligibility when listening through headphones: noise (a), reverberation (b), pink noise + reverberation (c)

The processing of the obtained results consisted in evaluating the speech intelligibility by calculating the proportion of correctly recognized monosyllable. A properly recognized monosyllable was a fixed monosyllable, the symbolic composition of which fully coincided with at least one of the set of text variants of the articulation tables.

Experimental studies were carried out in two phases. At the first studies phase 26 students aged 22 were involved in listening. Listening to speech signals occurred through headphones. However, the three students did their job with deviations from the task demands

and listened signals through acoustic monitors (or through computer speakers, which in the future will also be called acoustic monitors for simplicity). Estimates of the speech intelligibility received by these three listeners were much higher than those for the remaining listeners. For example, in the case of noise interference, the speech intelligibility was 85%-93% for minus 10 dB SNR, while intelligibility was only 10%-30% when listening through the headphones [8].

One of the possible reasons of a significant increase intelligibility is the early reflection of sound in the room [9], [10], [11], [12]. However, there are other possible reasons, among which the most probable is the imperfection of the organization of experimental research and the imperfection of the developed software. Therefore, the main objective of the second phase of experimental research was to test the validity of the above assumptions. In the second phase, improved software was used, and new listener group was recruited, consisting of 20 students aged 20.

III. TESTING RESULTS AND THEIR DISCUSSION

A. First phase of experimental research

The first phase of experimental research was pilot-oriented, since its main goal was to test developed software and organization of experimental analysis.

Graphs of averaged estimates of speech intelligibility for signals distorted by noise and reverberation, for listening through headphones, are shown in Fig. 4. Similar results for listening via acoustic monitors are shown in Fig. 5.

Comparison of graphs of Figs. 4 and 5 shows that the masking properties of noise and reverberation depend not only on the signal-to-noise ratio, the color of the noise and the reverberation, but can also depend significantly on the method of listening.

Thus, in the case of noise interference, when changing the listening mode from listening through the headphones to listening through the acoustic monitors, the intelligibility increased from 10%-30% to 85%-93% in the case of SNR minus 10 dB (standard deviation of intelligibility estimates was close to 5% on average). This means that the SNR has increased by 17-20 dB when listening to signals through the acoustic systems, compared with listening through the headphones.

The resulting gain of the signal-to-noise ratio is surprising, since it is too large, although the fact of the growth of intelligibility is expected. Indeed, when listening to the noisy sound recorded in the "mono" mode through the acoustic system consisting of 2 emitters of sound, the direct sound signals from each emitter are practically coherent at the input of each ear, whereas the noise from the same emitters can be considered incoherent.

Therefore, the SNR (for the direct sound action) at the input of each ear should increase by 3 dB compared to the use of headphones.



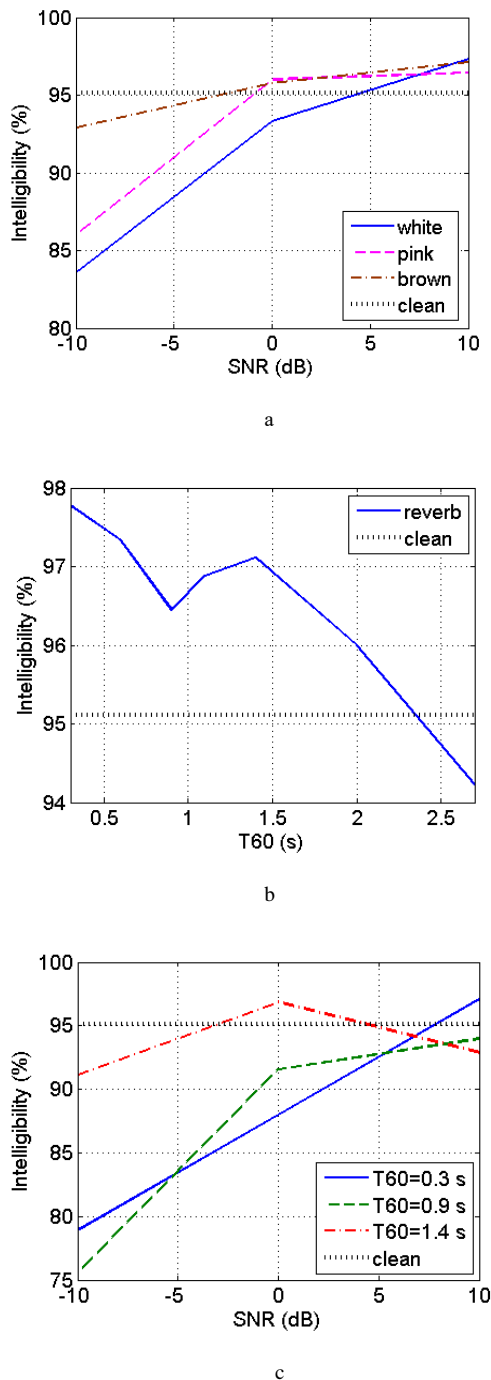


Fig. 5 Estimates of intelligibility when listening through acoustic monitors: noise (a), reverberation (b), pink noise + reverberation (c)

In addition, early reflections of the sound in the rooms play an important role. It was shown in [9], [10], [11], [12] that SNR may increase by 6-9 dB due to the effect of early reflections.

Finally, the gain of approximately 2-3 dB can be explained by binaural listening [12]. A simple summing up of these factors gives us a gain of 12-15 dB. This is already close enough to our 17-20 dB gain, although it does not get another about 3-5 dB.

One of the probable reasons for this, quite significant, difference seems to be the presence of features of the organization of experimental research, namely:

- incomplete consideration of the features of the Ukrainian language when creating articulation tables;
- limited number of speakers;
- not enough high quality audio version of articulation tables;
- insufficient number of text variants of articulation tables;
- features of the software interface.

Based on this, it is advisable in subsequent studies to eliminate these shortcomings and to check the validity of the assumption that listening of noisy signals through a pair of acoustic monitors in ordinary rooms cannot, as a rule, result in a SNR gain higher than 12-15 dB.

At the same time, it should be assumed that another important factor may be the psychophysical state of the listeners. The reason for this assumption is that, with later repetitions of the experiment of listening to signals through acoustic monitors in one of the rooms where an abnormally high degree of intelligibility was obtained, the new intelligibility estimates were elevated but did not exceed 30%-40% for SNR minus 10 dB. The discussion with the listener of the possible causes of this fact showed that the first listening, where an abnormally high degree of intelligibility was obtained, took place in a state of significant emotional upsurge. One way or another, but the hypothesis regarding this factor has the right to exist.

Finally, it is also advisable in subsequent studies to analyze in detail the acoustic characteristics of the rooms in which the aforementioned abnormal increase in the intelligibility was obtained, in order to identify possible features of the geometry of these rooms.

Returning to the situation of listening to the signals through the headphones, note that the graphs of Fig. 4a are in good agreement with the known results [1], [2], [7], [13] in the range of middle and high SNR values (0-10 dB). However, in the area of small SNR values (less than minus 5 dB), the masking property of white noise was better than that for brown noise. This contradicts the previous forecast estimates [1], [2], therefore it is also appropriate to further investigate the reasons for this discrepancy.

Regarding the effect of reverberation interference, expectations for reducing the speech intelligibility with the increase in reverberation time are confirmed. At the same time, we see that listening through monitors has also significantly enhanced the speech intelligibility. Thus, intelligibility has increased from 65% to 94% for the reverberation time of 2.7 s, and for the reverberation time of 0.3-2 s the intelligibility even exceeded one for a pure language by 1-3%. It should be noted that in this case, the standard deviation of the intelligibility estimates was quite large and close to 10%, which can be explained by the small number of listeners.

The analysis of intelligibility estimates for the case of the joint action of noise and reverberation also indicates a significant increase of the speech intelligibility when listening through acoustic monitors (the standard deviation of intelligibility values does not exceed 10%).

To summarize, we note that, based on the results of experimental studies, developed automated system of articulation tests can be considered workable. Regarding its quality, the analysis revealed a number of certain shortcomings and helped to formulate appropriate recommendations for their removal:

- when creating articulation tables, more attention should be paid to the phonetic features of the Ukrainian language;
- before recording audio versions of articulation tables, it is necessary not only to properly adjust the hardware and software, but also to instruct and train the speakers carefully, warning them from underscoring the monosyllables by pausing or increasing the volume;
- the number of text variants of articulation tables should be increased from two to three to better take into account the ambiguity of the auditory perception of some phonemes, as well as to increase the number of input variants of heard monosyllables from the keyboard;
- sounding of monosyllables should be performed in a random manner, which will reduce the risk of storing of monosyllables by listeners;
- it is necessary to make possible the correction of the results of input from the keyboard, as the listener's fatigue leads to an increase in the frequency of false keystrokes;
- it is advisable to limit the amount of visual information that is provided to the listener on the PC monitor after listening articulation tables, which

will reduce the risk of improving the results due to analysis by the listener of his own mistakes.

B. Second phase of experimental research

The purpose of the second phase of the study was to correct the organization of research, as well as the implementation of software modifications, taking into account the recommendations formed on the results of the first phase of research. This correction and modification was made as follows:

- the number of text variants of articulation tables was increased from two to three;
- sounding of monosyllables was performed in a random manner;
- it was made possible the correction by listeners of the results of input from the keyboard;
- the amount of visual information provided to the listener on the PC monitor after listening articulation tables was minimized;
- the same articulation tables were listened twice by every listener: through the headphones and through the acoustic monitors.

Results of the second experimental phase are shown in Figs. 6 and 7. For noise interference, standard deviation of intelligibility estimates was close to 6% for case of headphones and to 10% for listening through acoustic monitors. For reverberation interference, the proper standard deviation values were close to 8% and 12% respectively.

Comparing the graphs of Figs. 4 and 6 for case of listening to signals through the headphones, we see that the values of speech intelligibility practically coincide. A completely different situation is observed comparing Figs. 5 and 7 for case of acoustic monitors. As can be seen, shown in Fig. 7 intelligibility values are significantly lower than ones of Fig. 5 and practically do not differ from the results for the case of listening to the signals through the headphones (Fig. 6).

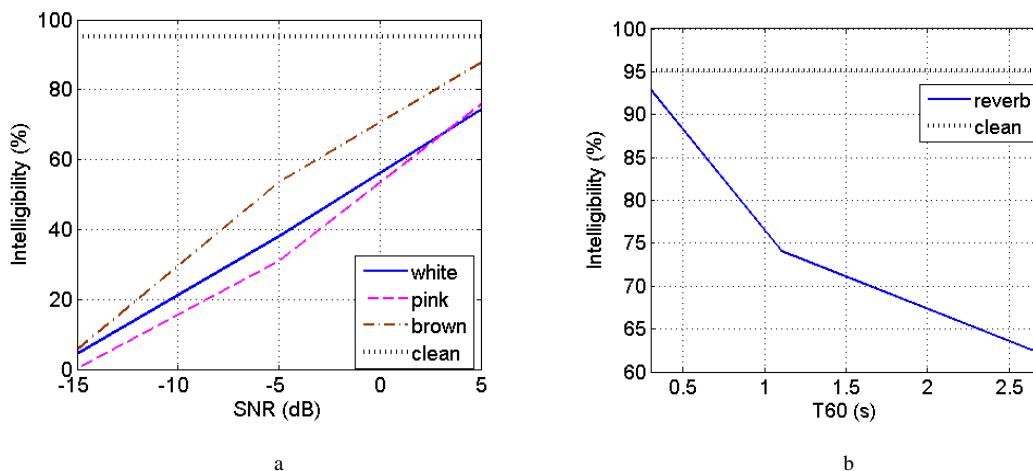


Fig. 6 Estimates of intelligibility when listening through headphones for noise (a) and reverberation (b) interferences

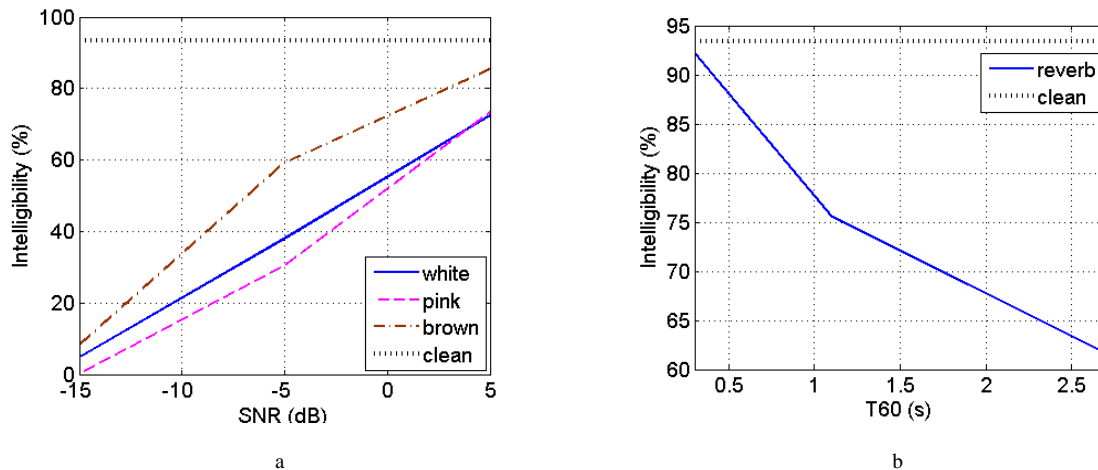


Fig. 7 Estimates of intelligibility when listening through headphones for noise (a) and reverberation (b) interferences

CONCLUSION

Thus, we can conclude that performed correction of software and research organization proved to be constructive. The results obtained indicate, firstly, that the organization of articulation tests and software features can significantly affect the test results.

Second, it is very important, in our opinion, to conclude on the practical coincidence of the results of listening through acoustic monitors and headphones. It should be noted, however, that the most likely reason for the coincidence was that the listeners are usually located 0.6-0.8 meters from the acoustic monitors, and at such distances the reverberation effect of the room on the quality of the signal received by the listener is negligible [9]. Therefore, in the future, it is advisable to try articulating tests for larger distances between listeners and acoustic monitors, where the reverberation effect is more prominent. This will allow reconciling the results with literary data on the importance of the role of early reflections in the room, which can lead to a significant increase in speech intelligibility.

Finally, in the future, it is also advisable to investigate the degree of influence of the psychophysical state of the listener on the results of articulation tests.

The general conclusion may be that listening to speech signals distorted by noise and reverberation interferences, performed with the use of the proposed automated system of articulation tests, indicates the performance and high quality of the developed system.

For small (less than minus 5 dB) signal-to-noise ratios, the incompleteness of the results obtained with the previous forecast estimates is revealed, which requires further analysis of the phenomenon.

Verification of the phenomenon of abnormally high speech intelligibility when listening to distorted signals through the acoustic systems showed that the most probable cause of such anomaly is the peculiarities of the articulation tests organization and the features of software.

A practical coincidence of the results of listening through acoustic monitors and headphones for listeners located 0.6-0.8 meters from the acoustic monitors was observed. In the future, it is advisable to try articulating tests for larger distances between listeners and acoustic monitors, where the reverberation effect is more prominent. This will allow reconciling the results with literary data on the importance of the role of early reflections in the room.

It is also advisable in the future to investigate the degree of influence of the psychophysical state of the listener on the results of articulation tests.

ACKNOWLEDGEMENTS

The authors thank the students of the Department of Acoustics and Acoustoelectronics of the National Technical University of Ukraine "Igor Sikorsky Kyiv Polytechnic Institute" for participating in the development and testing of the automated system of articulation tests.

REFERENCES

- [1] V. S. Didkovskiy, M. V. Didkovskaya, and A. N. Prodeus, Akusticheskaja ekspertiza kanalov rechevoj kommunikacii. Monografija [Acoustic examination of the channels of speech communication. Monograph], Kyiv: Imex-LTD, 2008. ISBN: 978-966-8861-85-7
- [2] A. N. Prodeus, V. S. Didkovskiy, and M. V. Didkovskaya, Akusticheskaya ekspertiza i korrekciya kommunikacionnyh kanalov. [Acoustic examination and correction of communication channels. Monograph], Saarbrücken: LAP LAMBERT Academic Publishing, OmniScriptum GmbH & Co. KG, 2017. ISBN: 978-3-330-04591-0
- [3] O. Arkhypova, V. Zhuravlov, and V. Kumeiko, "Artykuliatsiini tablytsi sliv ukrainskoi movy [Articulation tables of words of the Ukrainian language]," vol. 19, no. 2, pp. 13-17, 2009. URL: <http://ela.kpi.ua/handle/123456789/9689>
- [4] A. E. Arhipov, and E. A. Arhipova, "Analiz i obrabotka dannyh artikulyacionnyh ispytanj [Analysis and processing of articulatory test data]," *Scientific and Practical Journal "Information Protection"*, no. 4, pp. 34-42, 2012. URL: http://aprodeus.narod.ru/Teaching/ZAI/Art_2012_Arhipova_Analis_artikul_tabl.pdf



- [5] V. G. Mihajlov, "Diagnosticcheskie artikulyacionnye tablicy [Diagnostic articulation tables]," *Acoustic Journal*, vol. 48, no. 5, pp. 705-712, 2002.
URL: http://www.akzh.ru/pdf/2002_5_705-712.pdf
- [6] "ANG-2200. Acoustic Noise Generator. Owner's Guide," Research Electronics, Intl. 455 Security Place Algood, TN 38506, USA. [Online]. Available: https://reiusa.net/wp-content/uploads/2017/11/ANG_Manual_revG.pdf
- [7] A. M. Prodeus, A. V. Vityk, and D. Yu. Didenko, "Sub'iektyvne otsiniuvannia yakosti ta rozbirlyvosti movlennievkykh syhnaliv, spotvorenykh syntezovanyymi shumamy [Subjective evaluation of quality and intelligibility of speech distorted by synthesized noise]," *Microsystems, Electronics and Acoustics*, vol. 22, no. 6, pp. 45-63, 2017.
DOI: [10.20535/2523-4455.2017.22.6.101929](https://doi.org/10.20535/2523-4455.2017.22.6.101929).
- [8] A. Prodeus, K. Bukhta, P. Morozko, O. Serhienko, I. Kotvytskyi, and I. Shherbenko, "Automated System for Subjective Evaluation of the Ukrainian Speech Intelligibility," in *Proceedings of IEEE 38th International Conference on Electronics and Nanotechnology (ELNANO)*, Kyiv, Ukraine, April 24-26, 2018. DOI: [10.1109/TCSET.2018.8336263](https://doi.org/10.1109/TCSET.2018.8336263).
- [9] J. Bradley, H. Sato, M. Picard, "On the importance of early reflections for speech in rooms," *Journal of Acoustical Society of America*, vol. 113, no. 6, pp. 3233-3244, June 2003.
DOI: [10.1121/1.1570439](https://doi.org/10.1121/1.1570439)
- [10] J. S. Bradley, "Speech intelligibility studies in classrooms," *Journal of the Acoustical Society of America*, vol. 80, no. 3, pp. 846-854, 1986. DOI: [10.1121/1.393908](https://doi.org/10.1121/1.393908)
- [11] W. Yang and J. S. Bradley, "Effects of room acoustics on the intelligibility of speech in classrooms for young children," *Journal of the Acoustical Society of America*, vol. 125, no. 2, pp. 922-933, 2009. DOI: [10.1121/1.3058900](https://doi.org/10.1121/1.3058900)
- [12] I. Arweller, J. Buchholz, and T. Dau, "Speech intelligibility enhancement by early reflections," in *Proceedings of 2nd International Symposium on Auditory and Audiological Research (ISAAR 2009)*, Elsinore, Denmark, August 2009. URL: [http://orbit.dtu.dk/en/publications/speech-intelligibility-enhancement-by-early-reflections\(368caba2-dca2-461f-bc45-d8615f1dd675\).html](http://orbit.dtu.dk/en/publications/speech-intelligibility-enhancement-by-early-reflections(368caba2-dca2-461f-bc45-d8615f1dd675).html)
- [13] Yu. K. Makarov, and A. A. Horev, "K ocnke effektivnosti zashity akusticheskoy (rechevoj) informacii [To assess the effectiveness of protection of acoustic (speech) information]," *Information Security*, no. 4, 2005. URL: [http://www.vrsystems.ru/stati/k_ocenke_effektivnosti_zashiti_akusticheskoi_\(rechevoj\)_informacii.htm](http://www.vrsystems.ru/stati/k_ocenke_effektivnosti_zashiti_akusticheskoi_(rechevoj)_informacii.htm)

Надійшла до редакції 06 травня 2018 р.

УДК 004.934

Автоматизоване суб'єктивне оцінювання розбірливості мови при різних способах прослуховування

Продеус А. М., д.т.н. проф., ORCID [0000-0001-7640-0850](https://orcid.org/0000-0001-7640-0850)

e-mail aprodeus@gmail.com

Бухта К. С., ORCID [0000-0002-6234-7937](https://orcid.org/0000-0002-6234-7937)

e-mail k.kangrem@gmail.com

Морозко П. В., ORCID [0000-0001-9366-812X](https://orcid.org/0000-0001-9366-812X)

e-mail p77morozko@gmail.com

Сергиєнко О. В., ORCID [0000-0002-4212-1770](https://orcid.org/0000-0002-4212-1770)

e-mail sfeara86@gmail.com

Котвицький І. В., ORCID [0000-0003-2800-0080](https://orcid.org/0000-0003-2800-0080)

e-mail igorktvzk@gmail.com

Дворник О. О., ORCID [0000-0003-4735-2225](https://orcid.org/0000-0003-4735-2225)

e-mail alexanderdvornyk@gmail.com

Національний технічний університет України

«Київський політехнічний інститут імені Ігоря Сікорського» kpi.ua

Київ, Україна

Реферат—У даній роботі представлені результати автоматизованої суб'єктивної оцінки розбірливості української мови. Односкладові звукосполучення типу «приголосний-голосний-приголосний» прослуховувалися двома способами: через навушники та через акустичні монітори. Оцінка розбірливості мови виконувалася із застосуванням спеціально розробленого програмного забезпечення, що дозволяє автоматизувати процедуру артикуляційних випробувань. Прослуховування мови виконувалося для чотирьох ситуацій: чиста мова; мова, спотворена шумом; мова, спотворена реверберацією; мова, спотворена спільною дією шуму та реверберації. Виявилось, що маскувальна здатність білого шуму перевищує таку для коричневого шуму при відношеннях сигнал-шум, менших за мінус 5 дБ, що не зовсім узгоджується з попередніми прогнозними оцінками. Крім того, виявилось, що прослуховування мови, спотвореної шумом, через акустичні монітори може привести до значного збільшення розбірливості мови



в порівнянні з прослуховуванням через навушники. В якості можливих причин цього явища розглянуто ранні відбитки, наявність двох гучномовців, бінауральне прослуховування, психофізичні особливості слухачів, а також особливості програмного забезпечення та організації артикуляційних випробувань. Після корекції програмного забезпечення та процедури артикуляційних випробувань виявилось, що результати оцінки розбірливості мови практично не відрізняються для обох способів прослуховування за умови, що відстань між слухачем і акустичними моніторами не перевищує 0,6-0,8 метра. У той же час виконана корекція не відбилася на поведінці залежностей розбірливості мови від відношення сигнал-шум при малих (менших за мінус 5 дБ) значеннях відношення сигнал-шум.

Бібл. 13, рис. 7.

Ключові слова – автоматизація артикуляційних випробувань; розбірливість мови; шумова завада; ревербераційна завада; спосіб прослуховування.

УДК 004.934

Автоматизированное субъективное оценивание разборчивости речи при различных способах прослушивания

Продеус А. Н., д.т.н. проф., ORCID [0000-0001-7640-0850](https://orcid.org/0000-0001-7640-0850)

e-mail aprodeus@gmail.com

Бухта К. С., ORCID [0000-0002-6234-7937](https://orcid.org/0000-0002-6234-7937)

e-mail k.kangrem@gmail.com

Морозко П. В., ORCID [0000-0001-9366-812X](https://orcid.org/0000-0001-9366-812X)

e-mail p77morozko@gmail.com

Сергиенко А. В., ORCID [0000-0002-4212-1770](https://orcid.org/0000-0002-4212-1770)

e-mail sfeara86@gmail.com

Котвицкий И. В., ORCID [0000-0003-2800-0080](https://orcid.org/0000-0003-2800-0080)

e-mail igorktvzk@gmail.com

Дворник А. А., ORCID [0000-0003-4735-2225](https://orcid.org/0000-0003-4735-2225)

e-mail alexanderdvornyk@gmail.com

Национальный технический университет Украины

«Киевский политехнический институт имени Игоря Сикорского» kpi.ua

Киев, Украина

Реферат—В данной работе представлены результаты автоматизированной субъективной оценки разборчивости украинской речи. Односложные звуко сочетания типа «согласный-гласный-согласный» прослушивались двумя способами: через наушники и через акустические мониторы. Оценка разборчивости речи производилась с помощью специально разработанного программного обеспечения, позволяющего автоматизировать артикуляционные испытания. Прослушивание речи выполнялось для четырех ситуаций: чистая речь; речь, искаженная шумом; речь, искаженная реверберацией; речь, искаженная совместным действием шума и реверберации. Оказалось, что маскирующая способность белого шума превышает таковую для коричневого шума при отношениях сигнал-шум менее минус 5 дБ, что не совсем согласуется с предварительными прогнозными оценками. Кроме того, оказалось, что прослушивание речи, искаженной шумом, через акустические мониторы может привести к значительному увеличению разборчивости речи по сравнению с прослушиванием через наушники. В качестве возможных причин данного явления рассмотрены ранние отражения, наличие двух громкоговорителей, бинауральное прослушивание, психофизические особенности слушателей, а также особенности программного обеспечения и организация артикуляторного тестирования. После коррекции программного обеспечения и процедуры артикуляционных испытаний оказалось, что результаты оценки разборчивости речи практически не отличаются для обоих способов прослушивания при условии, что расстояние между слушателем и акустическими мониторами не превышает 0,6-0,8 метра. В то же время выполненная коррекция не отразилась на поведении зависимости разборчивости речи от отношения сигнал-шум при малых (менее минус 5 дБ) значениях отношения сигнал-шум.

Библ. 13, рис. 7.

Ключевые слова — автоматизация артикуляционных испытаний; разборчивость речи; шумовая помеха; реверберационная помеха; способ прослушивания.

