COMPUTER SCIENCES AND INFORMATION TECHNOLOGIES

UDC 621.391.83(045)
DOI:10.18372/1990-5548.78.18249

[1]A. M. Prodeus,
[2]O. O. Dvornyk,
[3]A. S. Naida,
[4]O. P. Grebin

COMPARISON OF TWO METHODS FOR MEASURING SPEECH INTELLIGIBILITY

National Technical University of Ukraine "Ihor Sikorsky Kyiv Polytechnic Institute," Kyiv, Ukraine
E-mails: [1]aprodeus@gmail.com ORCID ID: 0000-0001-7640-0850, [2]alexanderdvornyk@gmail.com,
[3]naida.a.s.2001@gmail.com, [4]alexgstudio.2016@gmail.com

Abstract*—Comparing of the full speech transmission index method and full formant-modulation method of speech intelligibility measurement has been made. The methods were compared according to the accuracy of the measurements under the conditions of noise interference. The dependences of the STI estimates errors on the signal-to-noise ratio and on the duration of the test signals were obtained by means of computer simulation. It is shown that the accuracy of speech transmission index estimates is practically the same for both methods. In particular, it is shown that the use of test signals with a duration of 4 s is minimally acceptable and provides an estimation error of 0.03. Using 8 s and 16 s test signals reduces the speech transmission index estimation error to 0.02 and 0.01, respectively, for a wide signal-to-noise ratio range from minus 28 dB to plus 28 dB. The obtained results are close to those for the case of the joint action of noise and reverberation.*

Index Terms—Speech transmission index; measurement method; formant-modulation method; bias; standard deviation.

I. INTRODUCTION

Since noise interference is always present in flight control and navigation centres, one of the important hardware and software components of flight control and navigation systems should be a system for evaluating the intelligibility of speech masked by noise interference. The presence of such a system will allow not only to assess the quality of the acoustic characteristics of control centres, but also to provide recommendations for improving these characteristics.

In addition, measurement and prediction of speech intelligibility in speech information transmission channels is necessary when testing communication lines and during acoustic examination of premises [1], [2], [3], [4]. The modulation method for evaluating speech intelligibility [3] is the most common today due to its versatility. Unlike the formant method [1], the modulation method allows taking into account not only the effect of noise, but also the effect of reverberation on speech intelligibility. The formant-modulation (FM) method proposed in [5] is also universal, as it is a type of modulation method. The advantage of the FM method is the possibility of calculating both the speech transmission index (STI) index and the articulation index (AI). Note that the FM method got its name precisely because of the ability to calculate both STI and the AI.

The disadvantage of the FULL STI [6] method is the long duration of the measurement procedure. Therefore, simplified "fast" methods STIPA and STITEL are used in practice instead of the FULL STI method [6]. The FM method also exists in full (FULL FM) and fast versions. However, the cost of using fast methods is a decrease in the accuracy of STI estimation [7]. Taking into account this fact, the possibility of performing STI measurement using the FULL STI method has been realized in some modern measurement systems [8].

When measuring STI by the FULL STI method, the duration of the test signal is chosen to be close to 10–15 seconds [6]. However, it is very difficult to find a justification for such a choice in the literary sources in the form of the dependence of the STI measurement error on the duration T of the test signals. For the FULL FM method, such dependencies are also unknown. In addition, the ratio of STI estimation errors by FULL STI and FULL FM methods is unknown. The objective of this paper is to eliminate these shortcomings.

II. PROBLEM STATEMENT

A. FULL STI Method

Before STI measurements by the FULL STI method, 14 test signals are generated

$$x_i(t) = \xi(t)\sqrt{f_i(t)},\ f_i(t) = 1 + \sin 2\pi F_i t,\ i = \overline{1,14}, \quad (1)$$

F_i = 0.63, 0.8, 1, 1.25, 1.6, 2, 2.5, 3.15,…
…4, 5, 6.3, 8, 10, 12.5 Hz,

$\xi(t)$ is a stationary noise with speech spectrum, $f_i(t)$ is a modulation function, F_i is a modulation frequency. In speech transmission index measurements, the signals $x_i(t)$ are alternately emitted by a sound source located at the point where the speaker is usually located.

At the point where the listener is located, a signal $y_i(t) = x_i(t) + n(t)$, $n(t)$ is the noise interference, is formed for everyone $x_i(t)$ with a certain signal-to-noise ratio (SNR). The signals $y_i(t)$ are then filtered by a seven parallel-connected octave filters, resulting in a set of 98 signals $y_{k,i}(t),\ k = \overline{1,7},\ i = \overline{1,14}$.

The speech transmission index is calculated as

$$\mathrm{STI} = \sum_{k=1}^{7} \alpha_k \cdot \mathrm{MTI}_k - \sum_{k=1}^{6} \beta_k \cdot \sqrt{\mathrm{MTI}_k \cdot \mathrm{MTI}_{k+1}}, \quad (2)$$

MTI_k is a modulation transfer index in kth frequency band, α_k and β_k are the weight coefficients [6].

Values MTI_k are calculated as

$$\mathrm{MTI}_k = \frac{1}{14}\sum_{i=1}^{14} \mathrm{TI}_{k,i}, \quad (3)$$

$$\mathrm{TI}_{k,i} = \begin{cases} \dfrac{\mathrm{SNR}_{\mathrm{eff}\,k,i} + 15}{30}, & -15 < \mathrm{SNR}_{\mathrm{eff}\,k,i} < 15, \\ 0, & \mathrm{SNR}_{\mathrm{eff}\,k,i} \le -15, \\ 1, & \mathrm{SNR}_{\mathrm{eff}\,k,i} \ge 15, \end{cases} \quad (4)$$

$$\mathrm{SNR}_{\mathrm{eff}\,k,i} = 10\lg\frac{\tilde{m}_{k,i}}{1 - \tilde{m}_{k,i}}, \quad (5)$$

$$\tilde{m}_{k,i} = \frac{2\,|A_{k,i}(F_i)|}{|A_{k,i}(0)|}, \quad A_{k,i}(F_i) = \frac{1}{T}\int_0^T y_{k,i}^2(t) e^{-j2\pi F_i t} dt, \quad (6)$$

$\mathrm{TI}_{k,i}$ is a transfer index, $\mathrm{SNR}_{\mathrm{eff}\,k,i}$ is an effective signal-to-noise ratio, $|\cdot|$ is a module symbol, ~ is an estimate symbol.

B. FULL FM Method

The algorithm of the FULL FM method differs from the above algorithm of the FULL STI method only in calculation of modulation transmission index as

$$\mathrm{MTI}_k = \begin{cases} \dfrac{E_k + 15}{30}, & -15 < E_k < 15, \\ 0, & E_k \le -15, \\ 1, & E_k \ge 15, \end{cases} \quad (7)$$

$$E_k = \frac{1}{14}\sum_{i=1}^{14} \mathrm{SNR}_{\mathrm{eff}\,k,i} \quad (8)$$

C. Comparison of FULL STI and FULL FM methods

It can be seen that the difference between the FULL FM method and the FULL STI method is the interchange of linear and non-linear operations. Indeed, in the FULL STI method, the $\mathrm{SNR}_{\mathrm{eff}\,k,i}$ values are first subjected to a nonlinear transformation according to (4), and then averaged according to (3). In the FULL FM method, the opposite is done: the $\mathrm{SNR}_{\mathrm{eff}\,k,i}$ values are first averaged according to (8), and then subjected to a nonlinear transformation according to (7).

The usefulness of the change in the operations order lies in the ability to calculate not only the STI, but also the AI [1]

$$A = \sum_{k=1}^{7} p_k \cdot P_k(E_k), \quad (9)$$

p_k is the probability of formants presence in the kth frequency band, $P_k(E_k)$ is the perception coefficient, $E_k = 10\lg D_{sk}/D_{nk}$ is the signal-to-noise ratio in the kth frequency band, D_{sk} and D_{nk} are signal and noise variances, respectively, in the kth frequency band.

Given the piecewise linear nature of (4) and (7) dependencies, it can be expected that in a certain neighborhood of $\mathrm{SNR}_{\mathrm{eff}\,k,i}$ = 0 dB, the average values and variances of STI estimates obtained by full modulation and full FM methods will be close. However, outside this range, the differences can be significant. Since this issue has not been investigated to date, the objective of this paper is to fill this gap.

III. Set up of the Study

The research was carried out by means of computer simulation. Signals $y_i(t) = x_i(t) + n(t)$, $n(t)$ is stationary pink noise, were generated with a sampling frequency of 22050 Hz. During each calculation session, the duration T of the signals was varied to 4, 8, 16, 32, and 64 seconds, and the SNR was varied from minus 28 dB to plus 28 dB in 4 dB increments. For each combination of SNR and T parameters, 30 STI estimates were calculated, which

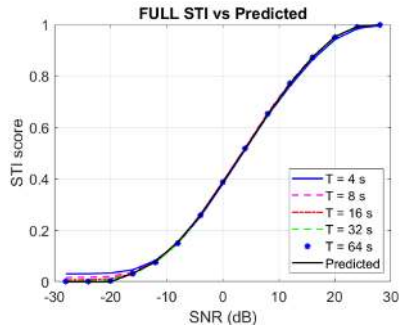
made it possible to estimate the expectation and standard deviation of STI estimates with sufficient for practical applications accuracy. The signals $y_i(t)$ were filtered by a seven parallel-connected bandpss octave filters with center frequencies of 125, 250, 500, 1000, 2000, 4000, 8000 Hz and a transmission coefficient outside the passband of minus 60 dB.

The software proposed in [5] was used for STI calculations by the FULL FM method. This software has been modified to calculate STI by the FULL STI method. Calculations were performed in the Matlab R2022a.

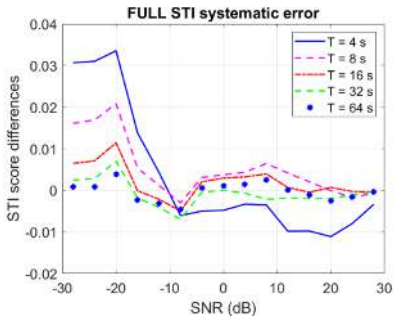
IV. Results of the Study

A. FULL STI Method

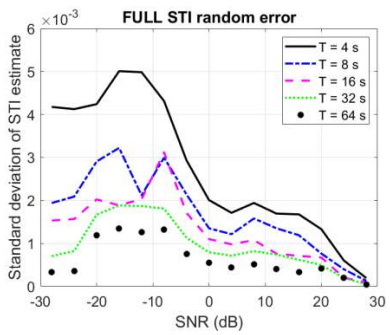
The results of estimating the expectation, bias and standard deviation of the STI estimates for the FULL STI method are shown in Fig. 1.



a)



b)



c)

Fig. 1. FULL STI method: estimates of expectation (a), bias (b) and standard deviation (c)

Expectation estimates (Fig. 1a) indicate the presence of a bias of STI estimates, which decreases with increasing duration T. To obtain quantitative values of the bias magnitude, the predictive STI estimate for the case of noise interference obtained using (2), (7) and (8) was used as a benchmark.

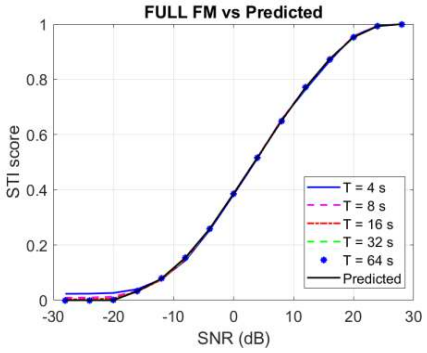
As can be seen in Fig. 1b, the STI estimate obtained by the FULL STI method is some shifted towards higher values at SNR < –15 dB. At SNR > – 15 dB, the STI estimate is also somewhat biased, but the magnitude of the bias decreases with increasing SNR. For T = 16 s, the value of the bias does not exceed 0.01 in the range of SNR values from minus 28 dB to plus 28 dB.

The standard deviation of the STI estimate (Fig. 1c) has a maximum in the range of SNR values from minus 20 dB to minus 10 dB and decreases to very small values as the SNR approaches 28 dB.

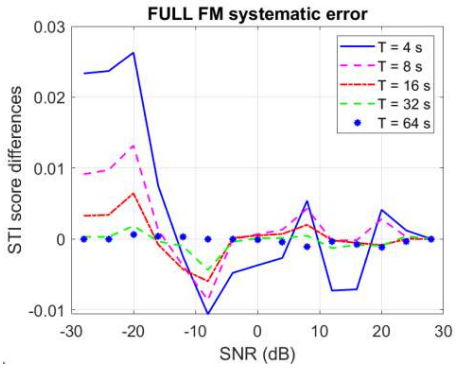
A slight decrease in the standard deviation of the STI score also occurs as the SNR approaches minus 28 dB. For T = 16 s, the value of the standard deviation does not exceed 0.003 in the range of SNR values from minus 28 dB to plus 28 dB.

B. FULL FM Method

The results of estimating the expectation, bias and standard deviation of the STI estimates for the FULL FM method are shown in Fig. 2.

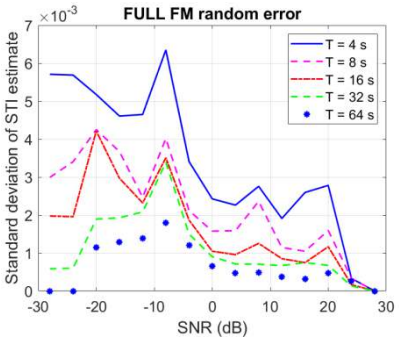


a)



b)

Fig. 2. FULL FM method: estimates of expectation (a), bias (b) and standard deviation (c)



c)

Fig. 2. Ending. (See also p. 11)

As can be seen in Figs. 2a and 2b, the STI estimate obtained by the FULL FM method is noticeably biased towards higher values at SNR< –15 dB. At SNR > –15 dB, the STI estimate is also slightly biased, but the magnitude of the bias decreases with increasing T and SNR. For T = 16 s, the value of the bias does not exceed 0.006 in the range of SNR values from minus 28 dB to plus 28 dB.

The standard deviation of the STI estimate (Fig. 2c) has a maximum in the range of SNR values from minus 20 dB to minus 8 dB and decreases to very small values as the SNR approaches 28 dB. A slight decrease in the standard deviation of the STI score also occurs as the SNR approaches minus 28 dB. For T = 16 s, the value of the standard deviation does not exceed 0.004 in a wide range of SNR values from minus 28 dB to plus 28 dB.

C. *Comparison of the Methods*

The above results indicate a significant similarity in the accuracy of STI measurements between FULL STI and FULL FM methods. In order to compare more clearly the STI estimates for these methods, the difference in the expectation estimates was calculated

$$\Delta_{\mathrm{FM},mdl} = \overline{\mathrm{STI}}_{\mathrm{FM}} - \overline{\mathrm{STI}}_{mdl}, \qquad (10)$$

$\overline{\mathrm{STI}}_{\mathrm{FM}}$ and $\overline{\mathrm{STI}}_{mdl}$ are average values of STI estimates obtained by FULL FM and FULL STI methods, respectively (Fig. 3a).

In addition, the ratio of the estimated standard deviations was calculated

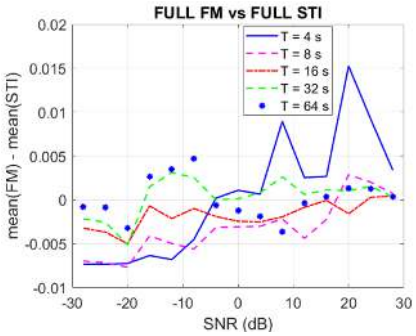
$$\Lambda_{\mathrm{FM},mdl} = \overline{\sigma\mathrm{STI}}_{\mathrm{FM}} / \overline{\sigma\mathrm{STI}}_{mdl}, \qquad (11)$$

$\overline{\sigma\mathrm{STI}}_{\mathrm{FM}}$ and $\overline{\sigma\mathrm{STI}}_{mdl}$ are estimates of the corresponding standard deviations (Fig. 3b).

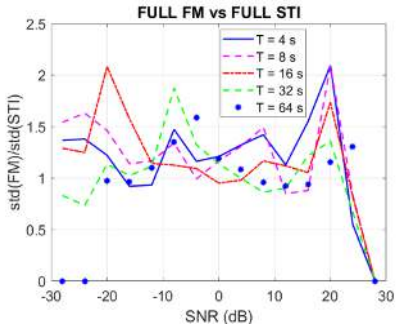
Shown in Fig. 3a results mean that at T = 16 s the difference in the average values of STI estimates does not exceed 0.005 in a wide range of *SNR* values from minus 28 dB to plus 28 dB. Shown in Fig. 3b results of calculations according to (11) mean that the standard deviations of the estimates also differ little because their ratio is close to 1. A graph in Fig. 3c shows the dependence of the averaged, over the *SNR* interval from minus 20 dB to plus 20 dB, values of the ratio (11) on the duration T of the test signals (1). It can be seen that the standard deviations of the STI estimates for the FULL FM method are only 10–30% higher than those for the FULL STI method.

V. Discussion

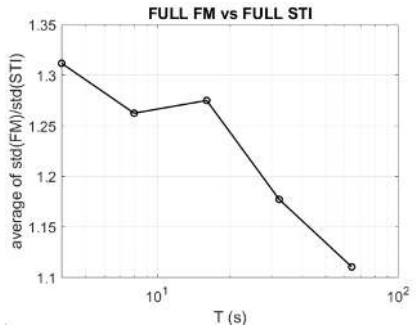
Some obtained results for T = 4 s, T = 8 s and T = 16 s are summarized in the Table I, where Δ is the maximum bias within the interval – 28 dB < SNR < 28 dB, Σ is the maximum standard deviation, $\Omega = \sqrt{\Delta^2 + \Sigma^2}$ is the maximum total measurement error.



a)



b)



c)

Fig. 3. Comparison of the methods: difference in expectations (a), ratio of standard deviations (b), average of standard deviations ratio (c)

TABLE I. ERRORS FOR NOISE DISTURBANCE

Method	***T* (s)**	Δ	Σ	Ω
FULL STI	**4**	0.033	0.005	0.034
	8	0.020	0.003	0.020
	16	0.010	0.003	0.011
FULL FM	**4**	0.027	0.006	0.028
	8	0.012	0.004	0.018
	16	0.006	0.004	0.007

Since the value of just noticeable difference (JND) for STI is 0.03 [8], it can be seen from the Table 1 that this requirement is practically satisfied even at $T = 4$ s, where the maximum total error of STI estimation is close to 0.03 for both methods. At $T = 8$ s and $T = 16$ s, the maximum total STI estimation error for both methods is close to 0.02 and 0.01, respectively.

Similar results were obtained in [10] for the case of the combined action of noise and reverberation. In this case, the model of the signal to be analyzed has the form $y_i(t) = x_i(t) \otimes h(t) + n(t)$, $h(t)$ is the room impulse response (RIR), $\otimes$ is the convolution symbol. The record of the RIR of a real university auditorium with a volume of 370 m^3 and a reverberation time of 0.8 s was borrowed from [11].

Some results from [10] for $T = 4$ s, $T = 8$ s and $T = 16$ s are shown in the Table II. However, it is noted in [10] that the case of the combined action of noise and reverberation is verified for the situation when the reverberation time in the room does not exceed 1 s. The case of a longer reverberation time requires additional verification.

TABLE II. ERRORS FOR NOISE AND REVERBERATION [10]

Method	***T* (s)**	Δ	Σ	Ω
FULL STI	**4**	0.032	0.004	0.032
	8	0.016	0.004	0.016
	16	0.007	0.003	0.008
FULL FM	**4**	0.022	0.007	0.023
	8	0.011	0.006	0.013
	16	0.004	0.004	0.006

Note that the results from Tables I and II are in good agreement with [6] where it is specified that the FULL STI estimation error be close to 0.02 for a test signal duration of 10 s.

VI. CONCLUSION

The FULL STI and FULL FM methods provide almost the same STI measurement accuracy in the range of SNR from minus 28 dB to plus 28 dB and in the range of test signals duration from 4 s to 64 s.

The obtained results are close to those for the case of the joint action of noise and reverberation.

Test signals duration $T = 4$ s is minimally acceptable and provides an estimation error of 0.03. The use of test signals durations of 8 s and 16 s allows reducing the STI estimation error to 0.02 and 0.01, respectively.

In the future, it is appropriate to investigate the influence of the shape of the long-term speech spectrum on the results of STI assessment [12], [13].

REFERENCES

[1] J. Collard, “Theoretical Study of the Articulation and Intelligibility of a Telephone Circuit,” *Electrical Communication*, vol. 7, 1929, p. 168. Available at: https://www.worldradiohistory.com/Archive-ITT/20s/ITT-Vol-07-1929-03.pdf

[2] K. Kryter, *The Effects of Noise on Man*, Academic Press, New York and London, 1970, 612 p. Available at: https://www.perlego.com/book/1897278/the-effects-of-noise-on-man-pdf

[3] H. Steeneken, and T. Houtgast, "A physical method for measuring speech-transmission quality,” *J. Acoust. Soc. Am.,* 67, 1980, pp. 318–326, https://doi.org/10.1121/1.384464

[4] K. Rhebergena, “A Speech Intelligibility Index-based approach to predict the speech reception threshold for sentences in fluctuating noise for normal-hearing listeners,” *J. Acoust. Soc. Am.*, 117 (4), Pt. 1, April 2005, pp. 2181-2192, https://doi.org/10.1121/1.1861713

[5] A. Prodeus, “Formant-Modulation Method of Speech Intelligibility Evaluation: Measuring and Exactness,” *Proc. VII Int. Conf. MEMSTECH 2011*, Lviv, Polyana, Ukraine, 2011, pp. 54–60. Available at: https://ieeexplore.ieee.org/document/5960267

[6] British Standard BS EN 60268-16. Sound system equipment. Part 16. Objective rating of speech intelligibility by speech transmission index. 2011.

[7] A. Prodeus, "Rapid version of a formant-modulation method of speech intelligibility estimation," *Proc. VII Int. Conf. MEMSTECH 2011*, Lviv, Polyana, Ukraine, 2011, pp. 61–63. Available at: https://ieeexplore.ieee.org/document/5960269

[8] NTi Audio, Application note. Speech Intelligibility. Measurement with the XL2 analyzer. Dec. 2020, 28 p. Available at: https://www.nti-audio.com/en/

[9] J. Bradley, R. Reich, and R. Norcross, “A just noticeable difference in C50 for speech,” *Applied Acoustics*, (58), 1999, pp. 99–108, https://doi.org/10.1016/S0003-682X(98)00075-9

[10] A. Prodeus, O. Dvornyk, A. Naida and M. Didkovska, "The Accuracy of Speech Transmission Index Estimation under Conditions of Joint Action of

Noise and Reverberation," *2023 IEEE 13th International Conference on Electronics and Information Technologies (ELIT)*, Lviv, Ukraine, 2023, pp. 257–260, https://doi.org/10.1109/ELIT61488.2023.10310682

[11] M. Jeub, M. Schafer, and P. Vary, "A binaural room impulse response database for the evaluation of dereverberation algorithms," *Proc. Int. Conference on Digital Signal Processing (DSP)*, Santorini, Greece, 2009, https://doi.org/10.1109/ICDSP.2009.5201259

[12] D. Byrne, H. Dillon, and K. Tran, "An international comparison of long-term average speech spectra," *J Acoust Soc Am.* 1994, 96 (4): 2108–2120, http://dx.doi.org/10.1121/1.410152

[13] L. Morales, G. Leembruggenb, S. Dancec, and B. Shield, "A Revised Speech Spectrum for STI Calculations," *Applied Acoustics*, 2018, 132: 33–42, https://doi.org/10.1016/j.apacoust.2017.11.008

Received August 29, 2023

Prodeus Arkadiy. orcid.org/0000-0001-7640-0850. Doctor of Engineering Science. Professor.
Acoustic and Multimedia Electronic Systems Department, National Technical University of Ukraine "Igor Sikorsky Kyiv Polytechnic Institute," Kyiv, Ukraine.
Education: Kyiv Polytechnic Institute, Kyiv, Ukraine, (1972).
Research interests: digital signal processing.
Publications: 181.
E-mail: aprodeus@gmail.com

Dvornyk Oleksandr. Post-graduate Student.
Acoustic and Multimedia Electronic Systems Department, National Technical University of Ukraine "Igor Sikorsky Kyiv Polytechnic Institute," Kyiv, Ukraine.
Education: Kyiv Polytechnic Institute, Kyiv, Ukraine, (2019).
Research interests: digital signal processing.
Publications: 4.
E-mail: alexanderdvornyk@gmail.com

Naida Anton. Post-graduate Student.
Acoustic and Multimedia Electronic Systems Department, National Technical University of Ukraine "Igor Sikorsky Kyiv Polytechnic Institute," Kyiv, Ukraine.
Education: Kyiv Polytechnic Institute, Kyiv, Ukraine, (2022).
Research interests: digital signal processing.
Publications: 5.
E-mail: naida.a.s.2001@gmail.com

Grebin Oleksandr. Candidate of Science (Engineering). Associate Professor.
Acoustic and Multimedia Electronic Systems Department, National Technical University of Ukraine "Igor Sikorsky Kyiv Polytechnic Institute," Kyiv, Ukraine.
Education: Kyiv Polytechnic Institute, Kyiv, Ukraine, (1985).
Research interests: digital signal processing.
Publications: 44.
E-mail: alexgstudio.2016@gmail.com

А. М. Продеус, О. О. Дворник, А. С. Найда, О. П. Гребінь. Порівняння двох методів вимірювання розбірливості мовлення

Виконано зіставлення повного методу вимірювання індексу передачі мовлення та повного формантно-модуляційного методу вимірювання розбірливості мовлення. Методи порівнювалися за точністю вимірювань в умовах дії шумової завади. Залежності похибок оцінок індексу передачі мовлення від відношення сигнал/шум та тривалості тестових сигналів отримано за допомогою комп'ютерного моделювання. Показано, що точність оцінок індексу передачі мовлення є практично однаковою для обох методів. Зокрема, показано, що використання тестових сигналів тривалістю 4 с є мінімально припустимим й забезпечує похибку оцінки 0,03. Використання тестових сигналів тривалістю 8 і 16 с дозволяє зменшити похибку оцінки індексу передачі мовлення до 0,02 та 0,01, відповідно, для широкого діапазону відношення сигнал/шум від мінус 28 дБ до плюс 28 дБ. Отримані результати є близькими до таких для випадку спільної дії шуму і реверберації.

Ключові слова: індекс передачі мовлення; метод вимірювання; формантно-модуляційний метод; зміщення; стандартне відхилення.

Продеус Аркадій Миколайович. ORCID 0000-0001-7640-0850 Доктор технічних наук. Професор.
Кафедра акустичних та мультимедійних електронних систем, Національний технічний університет України «Київський політехнічний інститут ім. І. Сікорського», Київ, Україна.
Освіта: Київський політехнічний інститут, Київ, Україна, (1972).
Напрямок наукової діяльності: цифрова обробка сигналів.
Кількість публікацій: 181.
E-mail: aprodeus@gmail.com

Дворник Олександр Олександрович. Аспірант.
Кафедра акустичних та мультимедійних електронних систем, Національний технічний університет України «Київський політехнічний інститут ім. І. Сікорського», Київ, Україна.
Освіта: Київський політехнічний інститут, Київ, Україна, (2019).
Напрямок наукової діяльності: цифрова обробка сигналів.
Publications: 4.
E-mail: alexanderdvornyk@gmail.com

Найда Антон Сергійович. Аспірант.
Кафедра акустичних та мультимедійних електронних систем, Національний технічний університет України «Київський політехнічний інститут ім. І. Сікорського», Київ, Україна.
Освіта: Київський політехнічний інститут, Київ, Україна, (2022).
Напрямок наукової діяльності: цифрова обробка сигналів.
Кількість публікацій: 5.
E-mail: naida.a.s.2001@gmail.com

Гребінь Олександр Павлович. Кандидат технічних наук. Доцент.
Кафедра акустичних та мультимедійних електронних систем, Національний технічний університет України «Київський політехнічний інститут ім. І.Сікорського», Київ, Україна.
Освіта: Київський політехнічний інститут, Київ, Україна, (1985).
Напрямок наукової діяльності: цифрова обробка сигналів.
Кількість публікацій: 44.
E-mail: alexgstudio.2016@gmail.com