Available online www.ejaet.com

European Journal of Advances in Engineering and Technology, 2020, 7(2):7-12

Research Article ISSN: 2394 - 658X

Some Aspects of Speech Intelligibility Measuring Under Noise Dominance

Arkadiy Prodeus[1] and Maryna Didkovska[2]

[1]Department of Acoustics and Acoustoelectronics, National Technical University of Ukraine "Igor Sikorsky Kyiv Polytechnic Institute", Kyiv, Ukraine- 03056

[2]Department of Mathematical Methods of System Analysis, National Technical University of Ukraine "Igor Sikorsky Kyiv Polytechnic Institute", Kyiv, Ukraine- 03056

aprodeus@gmail.com

ABSTRACT

Specifics of acoustic signal processing prior partial signal-to-noise ratios estimating when measuring speech intelligibility under noise dominance was studied. A pre-processing algorithm was proposed and tested using real signals. The speech intelligibility was evaluated using three versions of the formant method as well as the speech transmission index. The quality criteria for this test were the closeness of speech intelligibility estimates obtained during predicting and in real measurements, as well as the reliability of detection of a test noise signal against a background of noise interference. The closeness of the measurement and prediction results of speech intelligibility testifies to the reliability of measurements performed using the proposed algorithm. A sufficiently high reliability of signal detection by the threshold method is achieved with a signal-to-noise ratio of at least 0 dB.

Key words: measurement, speech intelligibility, word intelligibility, speech transmission index, noise interference

INTRODUCTION

The problem of speech intelligibility prediction and measurement is quite old: its age at present is about 90 years [1]. Abroad, Articulation Index (AI) [2] and Speech Intelligibility Index (SII) [3] are commonly used versions of the formant method of thespeech intelligibility assessment. In former USSR, by the end of 1950s, several scientific schools were founded, headed by N. B. Pokrovsky [4], M. A. Sapozhkov [5], and Yu. S. Bykov [6], where their own versions of formant method were developed.

In 1973, a modulation method had been developed that allows to take into account both noise and reverberation effects on speech intelligibility [7]. For this reason, even statements about the “obsolescence” of formant method had been appeared in the literature [8]. However, strict comparison of capabilities of formant and modulation methods had revealed that the formant method stays preferable in certain conditions where noise dominates over the reverberation [9]. Therefore, the technology for measuring speech intelligibility proposed in this work is based on the use of the analytical and experimental base of the formant method.

PROBLEM STATEMENT

The essence of the formant method for measuring speech intelligibility is as follows [4-6, 10-12].

A loudspeaker is placed in the place where a speaker is normally located. The loudspeaker emits a segment of stationary noise as a test signal. The spectrum of this noise is similar to the speech spectrum estimated using an extended (at least 1 minute) segment of the speech signal. Acoustic power of the emitted signal is established so as to create the required sound pressure level typical for the situation being analyzed at a certain distance (for example 1 m) from the speaker. A microphone is positioned in the place of probable listener location and its output signal is processed to evaluate the speech intelligibility.

The word intelligibility W is calculated using the formant intelligibility A [4], [10]:

$$W = \begin{cases} 1.54 \cdot A^{0.25}[1-\exp(-11 \cdot A)], & A < 0.15; \\ 1-\exp\left(\frac{11 \cdot A}{1+0.7 \cdot A}\right), & A \geq 0.15. \end{cases} \quad (1)$$

$$A = \sum_{k=1}^{K} p_k \cdot P(E_k'), \tag{2}$$

where $P_k(E_k')$ is the speech perception factor; p_k is probability of formant occurrence in k-th frequency band with center frequencies f_{0k} and boundary frequencies f_{1k} and f_{2k}:

$$p_k = F(f_{2k}) - F(f_{1k}), \tag{3}$$

$$F(f) = \begin{cases} 2.57 \cdot 10^{-8} \cdot f^{2.4}, \ 100 < f \le 400 \text{ Hz}, \\ 1 - 1.074 \cdot \exp(-10^{-4} \cdot f^{1.18}), \ 400 < f \le 10000 \text{ Hz}. \end{cases} \tag{4}$$

According to N. B. Pokrovsky's technique [4],

$$P_k(E_k') = \begin{cases} \frac{0.78+5.46 \cdot \exp[-4.3 \cdot 10^{-3} \cdot (27.3-|E_k'|)^2]}{1+10^{0.1 \cdot |E_k'|}}, \ E_k' \le 0, \\ 1 - \frac{0.78+5.46 \cdot \exp\left[-4.3 \cdot 10^{-3} \cdot (27.3-|E_k'|)^2\right]}{1+10^{0.1 \cdot |E_k'|}}, \ E_k' > 0, \end{cases} \tag{5}$$

where E_k' is effective level of formant perception in k-th frequency band:

$$E_k' = E_k - \Delta B(f_{0k}), \tag{6}$$

E_k is the effective level of speech signal perception in the k-th frequency band, which is equal (at sufficiently high noise levels) to the signal-to-noise ratio q_k in this frequency band:

$$E_k = q_k = 10\lg \frac{D_{sk}}{D_{nk}}, \tag{7}$$

where D_{sk} and D_{nk} are signal and noise variances in the k-th frequency band; $\Delta B(f)$ is the difference between averaged speech and formant spectra:

$$\Delta B(f) = \begin{cases} 200/f^{0.43} - 0.37, \ f \le 1000 \text{ Hz}, \\ 1.37 + 1000/f^{0.69}, \ f > 1000 \text{ Hz}. \end{cases} \tag{8}$$

Alternatively, according to the M. A. Sapozhkov's technique, the formant spectrum is considered almost coincident with the speech spectrum, i.e. $\Delta B(f) = 0$, whence it follows that $E_k' = E_k$ [5]. It was shown in [11] that M. A. Sapozhkov's dependence $P_k(E_k')$ differs from analogous N. B. Pokrovsky's dependence.

The M. A. Sapozhkov's approach was refined in [12] by taking into consideration the dependence of perception factors on the frequency band. In this case, (1) need be replaced by

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

where perception factors $P_k(E_k)$ can be presented as polynomial functions of frequency band.

Recently, there appeared a tendency toward partial merging of the formant and modulation [2] methods of speech intelligibility assessment. In particular, according to simplified method of speech intelligibility assessment presented in GOST R ISO 24504-2015, the speech intelligibility is assessed using the STI (Speech Transmission Index) as follows:

$$STI = \sum_{k=1}^{7} \alpha_k \cdot T_k - \sum_{k=1}^{6} \beta_k \cdot \sqrt{T_k \cdot T_{k+1}}, \tag{10}$$

$$T_k = \begin{cases} 0, \ E_k < -15; \\ (E_k + 15)/30, -15 < E_k < +15; \\ 1, \ E_k > +15; \end{cases} \tag{11}$$

where α_k are weighting factors and β_k are redundancy coefficients, whose values for octave bands with center frequencies f_0 are given in GOST R ISO 24504-2015.

Obviously, abovementioned equations are used at the final stage of speech intelligibility measurement, when estimates of partial signal-to-noise ratios E_k become known. Meanwhile, problems concerned with acoustic signal pre-processing aimed to estimate parameters E_k are poorly elucidated in the literature. The purpose of this work is to eliminate this gap.

ACOUSTIC SIGNAL PREPROCESSING ALGORITHM

The final stage of speech intelligibility calculations using (1)-(11) must be preceded by the following two stages of acoustic signal preprocessing.

Stage 1. Preparation to measure:

- setting the initial data for the measurement procedure;
- recording the signal or reading the recorded signal from a memory device.

Stage 2. Signal detection and parameters evaluation:

- preliminary estimation of noise background variance σ_n^2 and calculation of threshold value tr;
- detection of the signal against the noise background using the threshold method;
- check of calculated signal start time t_{mix} and calculated signal end time $t_{mix} + T_{mix}$ validity;
- formation of a parallel combination using seven octave filters (filter bank);
- filtering of recorded acoustic signal using the filter bank;
- formation of signals and interferences for all frequency channels;
- calculation of interference variance σ_n^2 and signal variance σ_s^2;
- E_k values calculation.

Let us consider each stage in more details.
It is reasonable to start setting the initial data for the measurement procedure from selection of operation mode. Two operation modes seem to be most preferable: mode of recording and processing a real signal and mode of processing a previously recorded signal. The second mode is especially useful to work with archival data.
In addition to operation mode, the lengths of recorded signal segments should be specified in terms of variables values such as T_{n1} (duration of interference segment prior the signal), T_{mix} (duration of mixed signal/noise segment), T_{rev} (duration of reverberation interference segment), and T_{n2} (duration of interference segment after the mixed signal/noise segment). Setting of initial data finishes with specifying the version of formant method and number of octave bands (7 or 5).
To detect a signal using the threshold method, it is necessary to first estimate the variance σ_n^2 of the interference segments at the beginning and at the end of the signal record, and then calculate the threshold constant $tr = a \cdot \sigma_n$, where a is determined by the required false alarm probability. The signal detection consists in determining the signal start time t_{mix}. Since the accuracy of speech intelligibility depends on the accuracy of signal detection, it is desirable to provide visual monitoring of the detection results. An example of signal detection procedure visualization is shown in Fig. 1, where vertical red marker lines indicate the time points t_{mix}, $t_{mix} + T_{mix}$ and $t_{mix} + T_{mix} + T_{rev}$, while horizontal dashed line shows the threshold value tr. Having determined boundaries between signal and interference, it is necessary to subject the analyzed record to multi-channel filtering and then estimate partial signal-to-noise ratios E_k using the relationship

$$E_k = \frac{\sigma_{mix\ k}^2 - \sigma_{n\ k}^2}{\sigma_{n\ k}^2},$$

where $\sigma_{mix\ k}^2$ is the variance of response portion of mixed signal/noise segment from k-th filter; $\sigma_{n\ k}^2$ is the variance of response portions of noise interference segment from k-th filter.

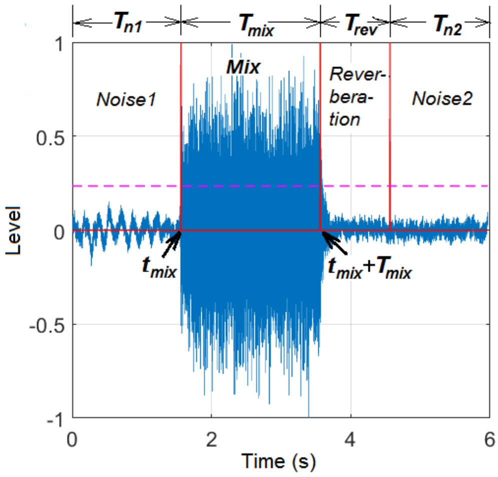


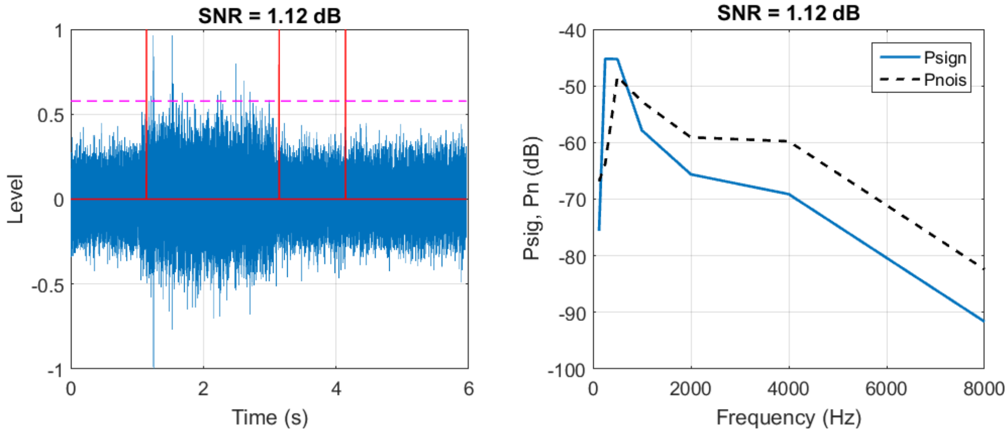
Fig. 1 Noise interference, mix of signal and noise, and reverberation segments

EXAMPLES OF NOISED SPEECH INTELLIGIBILITY MEASUREMENTS

Above presented algorithm was tested with the aid of specially designed scripts in the Matlab environment. Figs 2 to 4 present the form of analyzed signal and estimated power spectra for signal-to-noise ratios of 1.12 dB, 5.76 dB and 12.66 dB, respectively. The noise was recorded using an FM radio receiver operating in the mode of tuning off from all radio station signals.
As the presented plots show, applied threshold algorithm for detecting a noise-like signal against a noisy background reliably operates at signal-to-noise ratios of 0 dB or higher. This parameter is quite high enough for measurements of speech intelligibility in conditions typical of acoustic examination of open-plan rooms (according to ISO 3382-3-2013).
Fig. 5 shows experimentally obtained dependences of word intelligibility W and STI index on integral signal-to-noise ratio.It follows from Fig. 5 that the situation $SNR \approx 6 \dots 8$ dB can be considered as a boundary between moderate and good intelligibility, since $W \approx 0.9$ and $STI \approx 0.6$ are in this case [13], [14], [15].

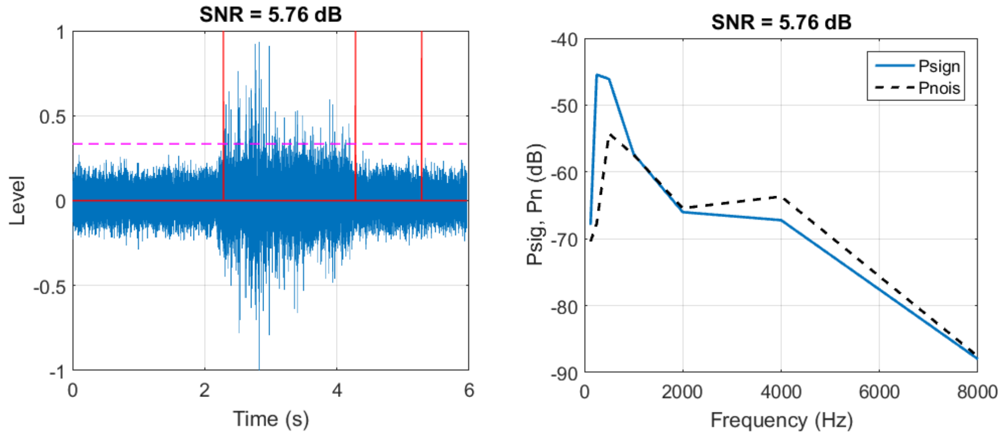
COMPARISON OF MEASUREMENT RESULTS WITH THEORETICAL ESTIMATES

Predictive assessments of speech intelligibility can be built using either computer simulation [12] or by calculations. The second approach is resource-saving as compared with the first one since it demands only analytical spectral models of speech signal and noise to be implemented.



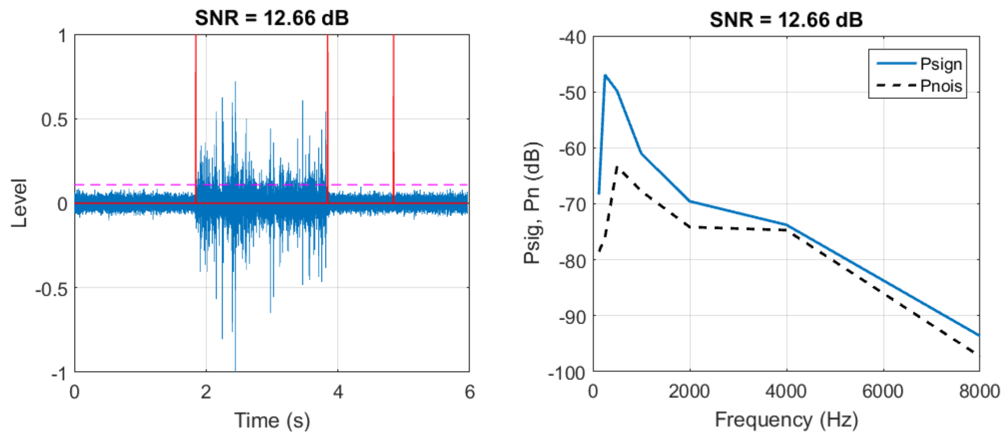
a b

Fig. 2 Waveform (a) and spectra (b) of signal and noise for SNR=1.12 dB



a b

Fig. 3 Waveform (a) and spectra (b) of signal and noise for SNR=5.76 dB



a b

Fig. 4 Waveform (a) and spectra (b) of signal and noise for SNR=12.66 dB

It can be shown that if the speech signal model has the form of variance distribution D_{sk} over frequency bands (here k is the frequency band number) and the noise model has the form of variance distribution D_{nk}, then the partial signal-to-noise ratios E_k can be calculated using the expression

$$E_k = 10\lg(D_{sk}/D_{nk}) + 10\lg(SNR_0) - 10\lg\left(\frac{\sum_{k=1}^{K} D_{sk}}{\sum_{k=1}^{K} D_{nk}}\right), \quad (12)$$

where SNR_0 is expected signal-to-noise value.

Table 1 shows an example of distribution D_{sk} [4] and distributions D_{nk} for 4 types of noise: white, pink, brown and typical for open-plan rooms. Fig. 6 presents results of predictive calculations of word intelligibility W and STI values dependences on the expected signal-to-noise ratio SNR_0 in the range of values SNR_0 = 0 ... 20 dB.

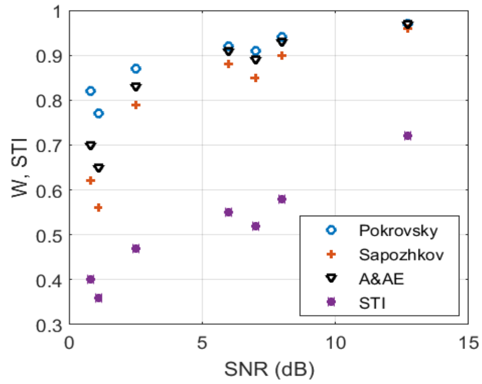


Fig. 5 Dependences of word intelligibility W and STI index on SNR

Table -1 Primary distribution of speech and noise variances

Parameters of frequency channels							
f_{0k}	125	250	500	1000	2000	4000	8000
Δf_k	90	175	355	690	1400	2800	5270
Parameters of white noise							
D_{nk}	0.130	0.254	0.514	1	2.029	4.058	7.638
Parameters of pink noise							
D_{nk}	1.043	1.014	1.029	1	1.014	1.014	0.955
Parameters of brown noise							
D_{nk}	8.348	4.058	2.058	1	0.507	0.254	0.119
Parameters of typical noise interference							
D_{nk}	5.193	2.017	1.284	1	1.023	1.273	1.5471
Parameters of long-term speech spectrum							
D_{sk}	0.146	2.846	3.642	1	0.361	0.181	0.096

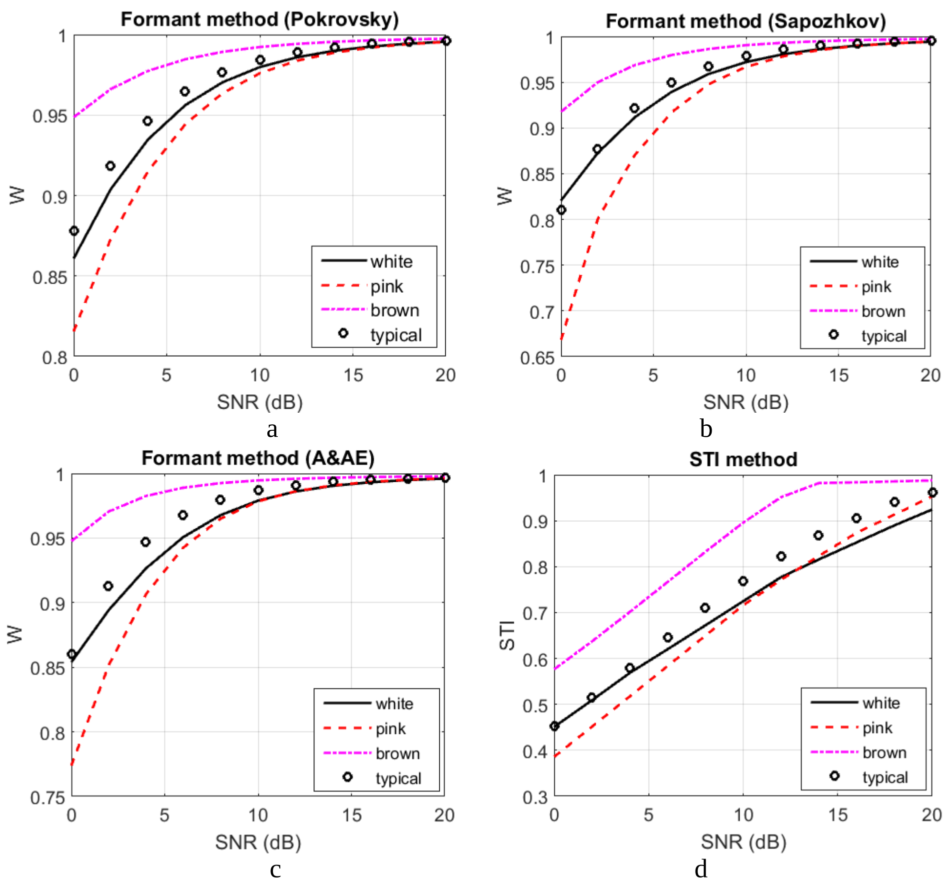


Fig. 6 Estimates of word intelligibility (a, b, c) and STI (d)

When comparing Figs 5 and 6, one can conclude that the measurement results are in good agreement with the predicted estimates. Indeed, in the case of STI estimates, situation $SNR \approx 4 \ldots 6$ dB can be considered as a boundary between moderate and good intelligibility in almost all noise models considered. In the same boundary situation, predictive estimates of word intelligibility $W \approx 0.9$ correspond to $SNR \approx 3 \ldots 5$ dB. The only exception is the brown noise model, for which boundary situation is $SNR \approx 0$ dB. Slight discrepancy between theoretical predictions and measured results can be explained by the difference between model spectra of noise interference and real noise.

CONCLUSION

Acoustic signal preprocessing algorithm for speech intelligibility measurement by formant or STI technique was proposed andits validity was tested on some examples.In addition, good agreement between the measurement and prediction results is shown. It was shown that applied threshold algorithm for detecting a noise-like signal against a noisy background reliably operates at signal-to-noise ratios of 0 dB or higher. In future, it will be reasonable try use more complicated cross-correlation processing algorithm for lower signal-to-noise ratios when detecting a noise-like signal.

REFERENCES

[1]. JA Collard, Theoretical Study of the Articulation and Intelligibility of a Telephone Circuit, *Electrical Communication*, 1929, 7, 168-186.

[2]. KD Kryter and JH Ball, SCIM – A Meter for Measuring the Performance of Speech Communication Systems, *Techn. Doc. Report No. ESD-TDR-64-674*, 1964, 1-38.

[3]. K Rhebergen 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*, 2005, 117 (4), 2181-2192.

[4]. NB Pokrovsky, *Prediction and Measurement of Speech Intelligibility*, Svyazizdat, Moscow, 1962.

[5]. MA Sapozhkov, *Voice Signal in Cybernetics and Communication*, Svyazizdat, Moscow, 1963.

[6]. YS Bykov, *Speech Intelligibility Theory and Radio Telephony Efficiency Improvement*, Gosenergoizdat, Moscow & Leningrad, 1959.

[7]. HJMSteeneken and T Houtgast, A Physical Method for Measuring Speech-Transmission Quality,*J. Acoust. Soc. Am.*, 1980, 67, 318-326.

[8]. I Aldoshina and R Pritts, *Musical Acoustics*, Kompozitor, St. Petersburg, 2006.

[9]. A Prodeus, L Dronzhevskaya, V Klimkov, D Shagitova, Formant and Formant-Modulation Methods of Speech Intelligibility Estimation. Part 2. Accuracy and Rapidity of Measurements, *Electronics and Communications*, 2011, 64(6), 16–24.

[10]. VK Zheleznyak, YK Makarov, AA Khorev Some Methodical Approaches to Evaluation of Speech Information Protection Effectiveness, *Special Equipment*, 2000, 4, 39-45.

[11]. A Prodeus, A Gavrilenko, V Didkovsky, Comparison of Versions of Formant Method for Speech Intelligibility Estimation, *Electronics and Communication. Thematic issue "Problems of electronics", Part 1*, 2008, 227-231.

[12]. V Didkovsky, M Didkovskaya, A Prodeus, *Acoustic Examination of Voice Communication Channels*, Imex-LTD, Kyiv, 2008.

[13]. T Houtgast, HJM Steeneken, SJ Van Wijngaarden, *Past, Present and Future of the Speech Transmission Index*. TNO Technische Menskunde, Soesterberg, 2002.

[14]. S Naida, O Pavlenko, Coupled Circuits Model in Objective Audiometry, *Proc. of the 2018 IEEE 38th International Conference on Electronics and Nanotechnology (ELNANO)*, Kyiv, Ukraine, 2018, 281-286.

[15]. S Naida, V Didkovskyi, O Pavlenko, N Naida, Spectral Analysis of Sounds by Acoustic Hearing Analyzer, *Proc. of the 2019 IEEE 39th International Conference on Electronics and Nanotechnology (ELNANO)*, Kyiv, Ukraine, 2019,421-424.