

Measuring Intelligibility using STOI Algorithms



Zoran Milivojevic, Dijana Kostic, and Zoran Velikovic

College of Applied Technical Sciences in Niš

Aleksandra Medvedeva

Niš 18000, Serbia

{zoran.milivojevic@vtsnis.edu.rs} {koricnac@yahoo.com} {zoran.velickovic@vtsnis.edu.rs}

ABSTRACT: *In this paper, we have prepared an objective model for the measure of intelligibility and proposed STOI algorithm which produces parameters. Then we proposed a testing process for analysing intelligibility of the sentences formed from the Matrix Sentence Test. This test is based on a subjective MOS (Mean Opinion Score) test and b). Objective test STOI algorithm. We also compared the results to find an analytical model for connecting pairs. We conducted the MOS test and calculated mean absolute error of the STOI algorithm.*

Keywords: Intelligibility, Objective Measure, STOI, MOS

Received: 3 April 2021, Revised 16 July 2021, Accepted 25 July 2021

DOI: 10.6025/jitr/2021/12/4/111-118

Copyright: Technical University of Sofia

1. Introduction

When designing audio systems for speech transfer, one of the most important features should be predicted - intelligibility of speech, taking care about one-third octave band. In speech technology, term “intelligibility”, is ranging from 0 to 100% intelligibility. Intelligibility of phoneme identification by listeners is a direct measure of intelligibility. The intelligibility of speech is the property of a speech signal that can be measured using: subjective and objective methods. Subjective methods of evaluation intelligibility are realized with the MOS test. These tests are often long-lasting and cost very well. They are realized using different sets of word (logatoms [1], Phonetic Balanced [2]...) and sentences (everyday [3] and matrix [4]).

Objective methods for evaluation intelligibility implies speech tests are carried out using a computer. Parameters such as: AI [5], SII [6], STI [7], CSTI [8] are used for testing. They are suitable for evaluation of intelligibility of speech for several types of degradation: reverberation, additive noise, filtering and clipping.

AI is an articulation index developed in the AT & Bell Labs [5]. Based on AI the speech intelligibility index - SII was developed [6]. Following the index SII, Houtgast and Steeneken developed a speech transfer index - STI [7], which can predict the intelligibility of reverberated speech and nonlinearity of distortion.

However, if we want to make an objective evaluation of intelligibility using a method where the disorder signal is processed by some kind of time-frequency variable (TF) gain function, these are not appropriate methods.

In the paper [9], STOI algorithm for objective testing of sentences was demonstrated. The work is based on the analysis of the speech signal in the time frequency domain (TF). The efficiency of the STOI algorithm was tested under the conditions of a superposed noise: car, caffè and bottle.

In this paper, the effectiveness of the STOI algorithm was analyzed in evaluation the intelligibility of sentences spoken in the Serbian language in the presence of the variable level of the White Gaussian Noise (WGN). The authors created an algorithm for determining the optimal parameters of the STOI algorithm (β_{opt} , β_{opt}). The experiment was realized at the Collage of Applied Science in Nis in which it was performed: a). subjective MOS test and b). an objective test using the STOI algorithm. The experiment was realized for $SNR = \{-7, -5, -2, 0, 2\}$ dB. The tests were implemented on students of the Collage of Applied Tehnical Science of Nis, gender structure: 10 male and 10 female, age from 19 to 26 years. The subjective MOS test determined the intelligibility of sentences spoken in the Serbian language and was used as a reference for the purpose of comparing the results obtained with the STOI algorithm. Comparative analysis of the results determined the optimal parameters of the STOI algorithm β_{opt} , β_{opt} , as well as their analytical dependence. In the end, a mean absolute error of estimation of the STOI algorithm was determined, for measuring the results obtained by the MOS test.

The organization of work is the follow. Section 2 describes the STOI algorithm. Section 3 describes the algorithm for evaluation of optimal parameters. Section 4 describes an experiment of evaluation intelligibility and performs comparative analysis of the results. Section 5 is a conclusion.

2. STOI Algorithm

The STOI algorithm is described in [9] and consists of the following steps:

Input: x - clean speech signal, y - speech signal with superimposed noise,

Output: d - intelligibility.

Step 1: TF decomposition of the x and y signals.

Step 2: Forming frames of length 256 and modification using the Hann - windowed.

Step 3: Removing the silence regions.

Step 4: Find the frame with the maximum energy of signal x .

Step 5: Finding the k^{th} DFT-bin of the m^{th} frame. The standard of One-third octave is defined as:

$$X_j(m) = \sqrt{\sum_{k=k_1(j)}^{k_2(j)-1} |x(k, m)|^2}, \quad (1)$$

where k_1 and k_2 represented the boundaries of one-third octave.

Step 6: Forming a short-time envelope vector of clean signal:

$$x_{j,m} = [x_j(m-N+1), X_j(m-N+2), \dots, X_j(m)]^T, \quad (2)$$

where is $N = 30$, which corresponding to a frame length of 384 ms, for $f_s = 10$ kHz.

Step 7: The normalization and clipping y signal

$$\bar{y}_{j,m}(n) = \min \left(\frac{\|x_{j,m}\|}{\|y_{j,m}\|} y_{j,m}(n), \left(1 + 10^{-\beta/20}\right) x_{j,m}(n) \right), \quad (3)$$

where is $\beta = -15$ dB.

Step 8: Defining a measure of intelligibility $d_{j,m}$:

$$d_{j,m} = \frac{(x_{j,m} - \mu_{x_{j,m}})^T (y_{j,m} - \mu_{y_{j,m}})}{\|x_{j,m} - \mu_{x_{j,m}}\| \|y_{j,m} - \mu_{y_{j,m}}\|}, \quad (4)$$

Step 9: The average measure of intelligibility for all bands and frames is d :

$$d = \frac{1}{JM} \sum_{j,m} d_{j,m}, \quad (5)$$

where M represent total number of frames, and J number of one-third octaves.

3. Algorithm for Evaluation of Optimal Parameters

The algorithm for determining the optimal parameters of the STOI algorithm is realized in the following steps:

Input: SNR_{min} , SNR_{max} , ΔSNR , β_{min} , β_{max} , $\Delta\beta$, N_{min} , N_{max} , ΔN , N_r .

Output: Trajectory of optimal values y_{fit}

Step 1: Created the test of sentence, with fixed syntax word structure from the SMST base (signal x_r)

FOR $SNR = SNR_{min} : \Delta SNR : SNR_{max}$

FOR $r = 1 : N_r$

Step 2: Superposed White Gaussian noise, n ($\sigma_2 = 1$, $\mu = 0$)

$$x_{1r} = x_r + k_r \cdot n, \quad (6)$$

where k_r is the coefficient determining the SNR of the speech signal x_{1r} .

Step 3: Realization of a subjective MOS test and generating a evaluation of intelligibility MOS_{xr}.

FOR $\beta = \beta_{min} : \Delta\beta : \beta_{max}$

FOR $N = N_{min} : \Delta N : N_{max}$

Step 4: Application of STOI algorithm $d_r(N, \beta) = \text{STOI}(N, \beta)$

END N ;

END β ;

END r ;

Step 5: Forming of equivalent matrix of intelligibility coefficient:

$$d(N, \beta) = \overline{d_r(N, \beta)}, \quad (7)$$

Step 6: The mean value of coefficient for STOI intelligibility is: $d_{STOI} = \overline{d(N, \beta)}$, (8)

and statistical parameters are mean value $\mu = d_{STOI}$ and variance σ^2 .

Step 7: Forming matrix error of evaluation: $e = |d - MOS_{xr}|$, (9)

Step 8: Generating a trajectory of minimum error: $y_k = f(N_{opt}, \beta_{opt})$, (10)

Step 9: Generating an equivalent trajectory of optimal values by fitting method: $y_{fit} = \beta_{fit} = a_3 N^3 + a_2 N^2 + a_1 N + a_0$, (11)

where is $a_3 = -4.927 \cdot 10^{-5}$, $a_2 = 0.01177$, $a_1 = -0.9856$, $a_0 = 2.942$.

4. Experimental Results and Analysis

In this section, an experiment was implemented in order to find the optimal parameters of the STOI algorithm.

4.1. Experiment

In order to test the effectiveness of evaluation using the STOI algorithm, experiment was implemented at the Collage of Applied Science Niš, where is applied a algorithm described in Section 3 for the evaluation of intelligibility. An experiment was implemented in order to find the optimal parameters of the STOI algorithm. The input parameters are: a). $SNR_{min} = -7$ dB, $SNR_{max} = 2$ dB, $\Delta SNR = 2$ dB, b). $N_r = 20$, number of test sentences for each value SNR , c). $N_{min} = 10$, $N_{max} = 100$, $N = 10$, $\beta_{min} = -40$, $\beta_{max} = -5$, $\Delta b = 5$. The output parameters of the algorithm are: a). analytic function of optimal values $\beta_{opt} = y_{fit}$, b) the results of subjective test MOS_{xr} , c). the results of objective test d_{STOI} i d). statistical parameter: mean value m and variance σ^2 for d_{STOI} . Results are shown in tabular and graphical form.

4.2. The Base

A SMST base has been created which contains words (name, verb, number, adjective and object) spoken in Serbian language [10]. Using the algorithm described in section 3 (step 1), sentences with a fixed syntax structure were formed. The words are formed by a random selection of the word from the SMST base to the syntax structure: name-verb-numberadjective-object. It is possible to create 100000 different sentences, which made test unpredictable for the tested group.

4.3. Test Group

A test group was formed from students of the Collage of Applied Science Nis, age from 19 to 26 ($\mu = 20.85$), gender structure: 10 male and 10 female respondents.

4.4. The Results

Table 1 shows the results of MOS (step 3) and dSTOI (step 4) and statistical of mean values μ and variance σ^2 for various SNR values.

The values of the MOS and STOI coefficients for each sentence are shown graphically for: a) $SNR = 0$ dB (Figs. 1. a) and b) $NR = -7$ dB (Figs. 2.a). The value of equivalent matrix coefficients $d(N, b)$, (step 5) are shown: a) $SNR = 0$ dB (Figs.1.b) and b) $SNR = -7$ dB (Figs. 2.b). A matrix 5error of intelligibility evaluation e (step 7) are shown: a) $SNR = 0$ dB (Figs. 1.c) and b) SNR

= -7 dB (Figures. 2.c). Trajectories of minimal error of intelligibility evaluation $y_k(N, \beta)$, (step 8) are shown on: a) $SNR = 0$ dB (Figures. 1.d) and b) $SNR = -7$ dB (Figures. 2.d). In Figure 3 are shown: a) trajectory y_k minimal error for $SNR = \{-7, -5, -2, 0, 2\}$ dB and b) equivalent trajectory y_{fit} (step 9). Intelligibility determined with: a) subjective MOS test, MOS_{xr} , and objective test using STOI algorithm, d_{STOI} are shown in Figure 4.

SNR (dB)	MOS_{xr} (%)	σ^2	μ (%)	d_{stoi}
2	71	4.4834	70.1975	70.1975
0	61	4.8871	66.5205	66.5205
-2	57	5.0465	61.5471	61.5471
-5	47	538.5059	45.0886	45.0886
-7	34	364.4125	44.0291	44.0291

Table 1

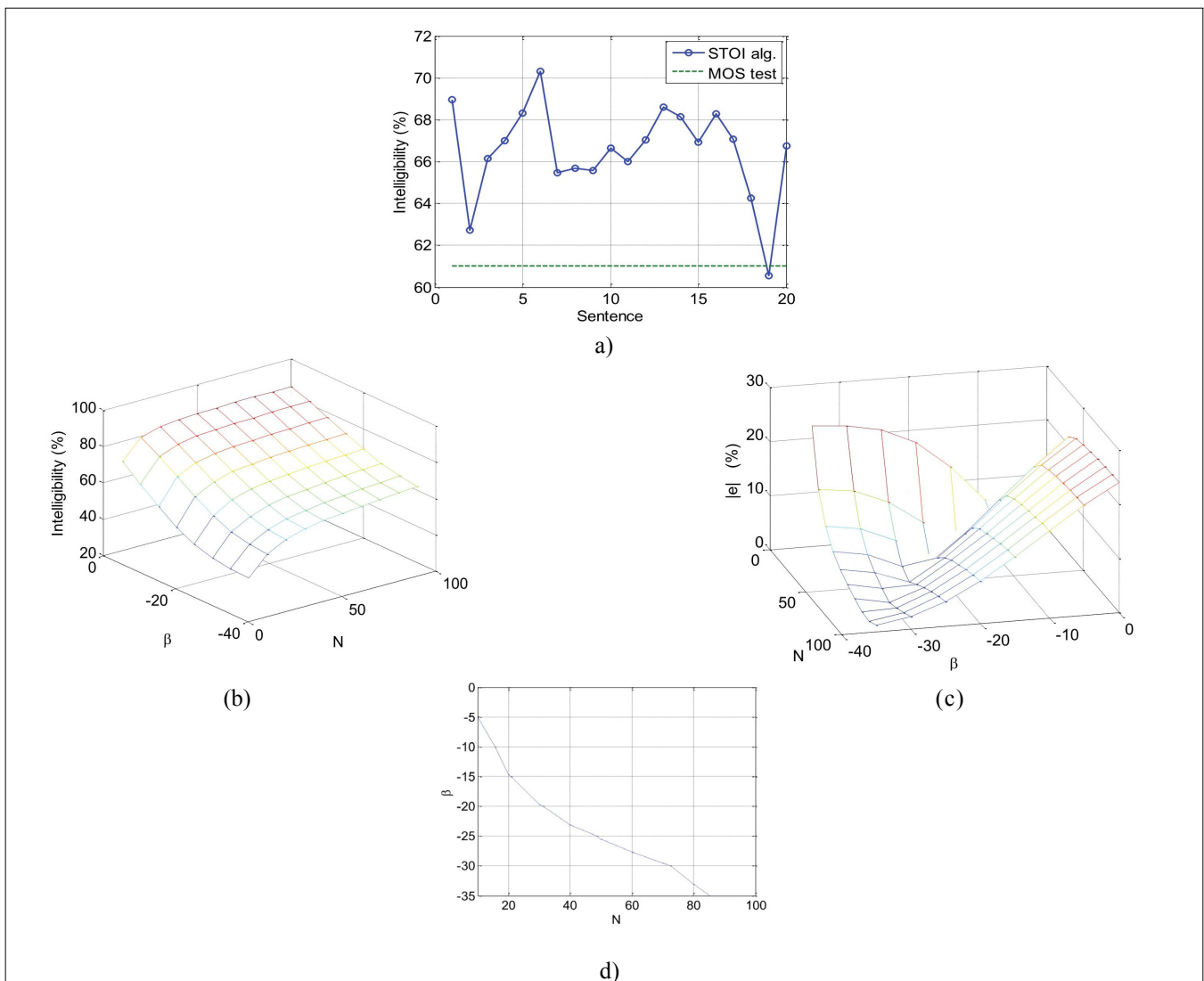


Figure 1. $SNR = 0$ dB: a) intelligibility of sentence x_r ($r = 1, \dots, 20$), coefficient of STOI algorithm and MOS evaluation, b) matrix of mean value d, c) matrix of error e and d) trajectory of minimal error of evaluation intelligibility y_k

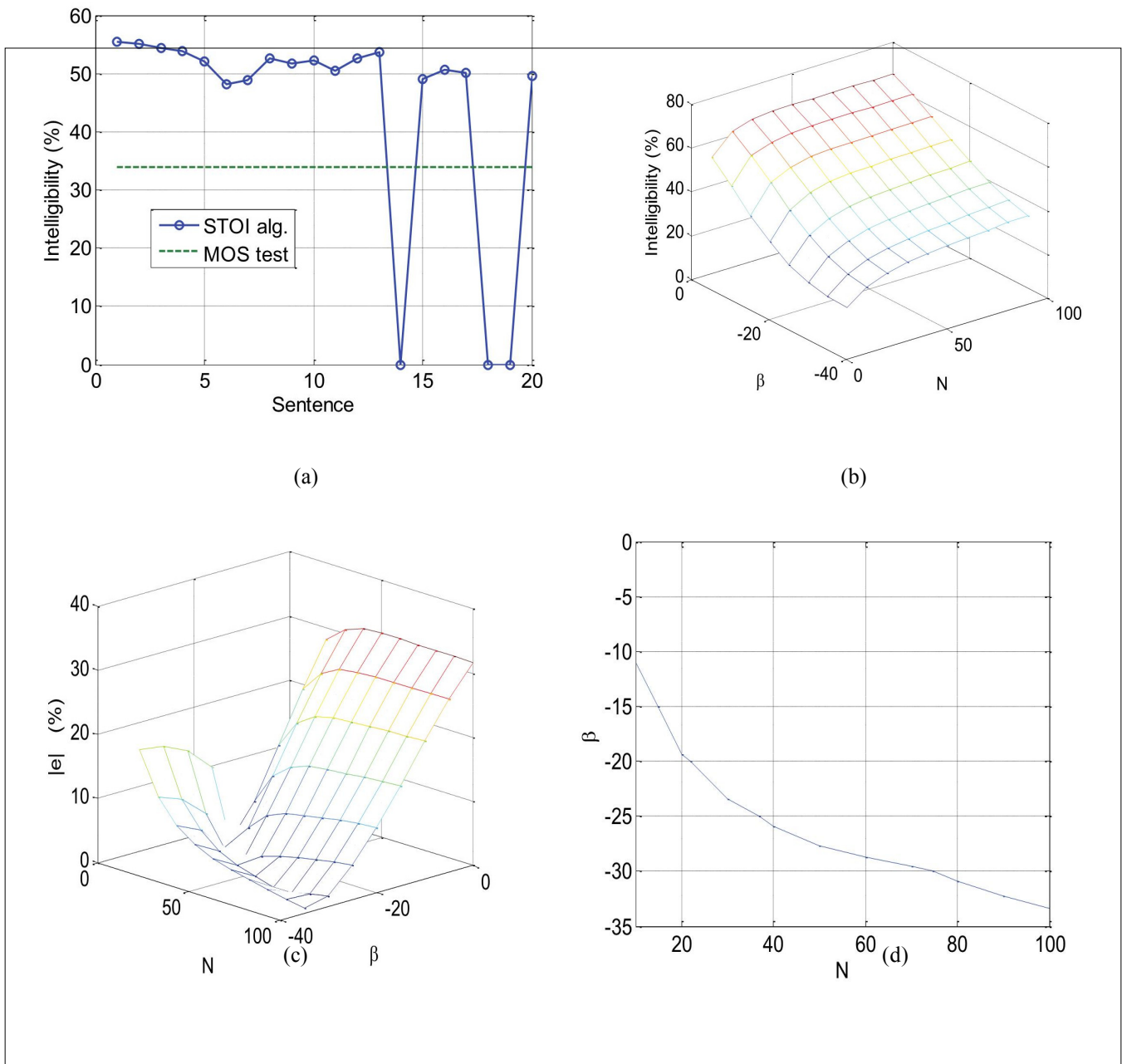


Figure 2. SNR=-7dB: a) intelligibility of sentence x_r ($r = 1, \dots, 20$), coefficient of STOI algorithm and MOS evaluation, b) matrix of mean value d , c) matrix of error e and d) trajectory of minimal error of evaluation intelligibility y_k

4.5. Analysis of Results

By analysis of results shown in Table 1 and Figs. 1 ÷ 4, it can be concluded that:

a). There are many numbers of pairs (N_{opt}, β_{opt}) , which can analytical described with Eq. (11) and graphically in Fig. 3 with trajectory;

b). Taking into considering the results of the subjective MOS test as a reference, the mean absolute error of the intelligibility evaluation determined by the STOI algorithm (Table 1, Figure 4) is:

$$e_{SNR} = \frac{1}{5} \sum_{SNR=-7dB}^{2dB} |MOS_{XR}(SNR) - d_{STOI}(SNR)| = 4.5622\%$$

which is considered a good evaluation.

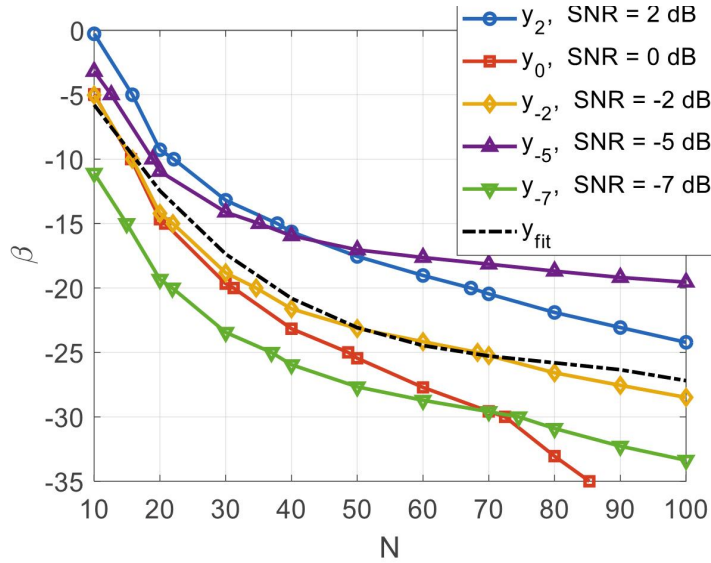


Figure 3. Trajectory of minimal error of intelligibility evaluation for various levels of SNR

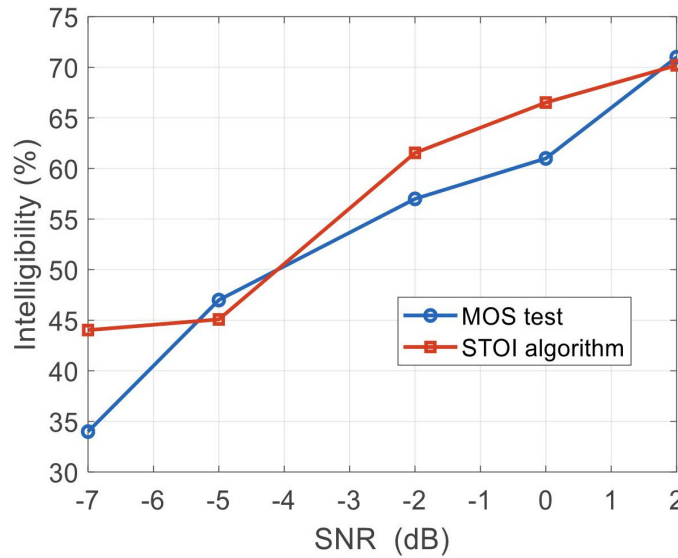


Figure 4. Intelligibility determined by: a) MOS test and b) STOI algorithm

5. Conclusion

The paper presents an algorithm for determining the optimal parameters of the STOI algorithm in the presence of WGN. A detailed analysis of the results of the conducted experiment showed that there are an infinitely large number of values of optimal parameters and their analytical formula is determined. Taking into account the results of the subjective MOS test as a reference, it has been shown that the mean absolute error of evaluation intelligibility of the STOI algorithm is $e_{SNR} = 4.5622\%$.

This result points to the quality of the STOI algorithm and makes recommendations for its application when evaluating speech intelligibility.

References

- [1] Kosti, D., Milivojevi, Z., Stojanovi, V. (2016). The Evaluation of Speech Intelligibility in the Orthodox Church on the Basis of MOS Test Intelligibility Logatom Type CCV, Icest 2016, Ohrid, Macedonia, 153-156.
- [2] Kruger, K., Gough, K., Hill, P. (1991). A Comparison of Subjective Speech Intelligibility Test in Reverberant Environments, *Canadian Acoustic*, 19 (4) 23 - 4.
- [3] Plomp, R., Mimpen, A. M. (1979). Improving the Reliability of Testing the Speech Reception Threshold for Sentences, *Audiology*, 18, 43 - 52.
- [4] Hagerman, B., Sentences for Testing Speech Intelligibility in Noise, *Scand Audio*, 11, 79-87.
- [5] French, N. R., Steinberg, J. C. (1947). Factors Governing the Intelligibility of Speech Sounds, *J. Acoust. Soc. Amer.*, 19 (1), 90–119.
- [6] Methods for Calculation of the Speech Intelligibility Index, S3.5-1997, ANSI, New York.
- [7].Steeneken, H. J. M., Houtgast, T. (1980). A Physical Method for Measuring Speech-Transmission Quality, *J. Acoust. Soc. Amer.*, 67 (1) 318–326.
- [8] Goldsworthy, R. L., Greenberg, J. E. (2004). Analysis of Speech-Based Speech Transmission Index Methods with Implications for Nonlinear Operations, *J. Acoust. Soc. Amer.*, 116 (6) 3679–3689.
- [9] Tall, C. H., Hendriks, R. C., Heusdens, R., Jensen, J. (2011). An Algorithm for Intelligibility Prediction of Time-Frequency Weighted Noise Speech, *EEE Transactions on audio, speech, and language processing*, 19 (7), September.
- [10] Milivojevi, Z., Kosti, D., Veli kovi, Z., Brodi, D. (2016). Serbian Sentence Matrix Test for Speech Intelligibility Measurement in Different Reverberation Conditions, UNITEH Gabrovo II, 173-178.