Audio Engineering Society

Convention Paper

Presented at the AES 156th Convention
2024 June 15–17, Madrid, Spain

MATLAB Implementation of STIPA (Speech Transmission Index for Public Address Systems)

Pavel Záviška[1], Pavel Rajmic[1], and Jiří Schimmel[1]

[1]*Brno University of Technology, Faculty of Electrical Engineering and Communication, Czech Republic*

Correspondence should be addressed to Pavel Rajmic (pavel.rajmic@vut.cz)

ABSTRACT

STIPA is a popular method for the prediction of speech intelligibility when speech is passed through a transmission channel. Though, implementations of STIPA are not publicly available and are often limited to the use with a proprietary measurement hardware. We present a Matlab implementation of STIPA direct method according to the IEC 60268-16:2020 standard. The proposed implementation meets prescribed requirements, which is demonstrated on reference test signals. We also carried out a successful verification measurement with respect to a commercial measurement device. Our software has open source code.

1 Introduction

The increasing importance of public address (PA) systems used for emergency announcement purposes has led to greater emphasis being placed on the speech intelligibility provided by the sound system. A number of standards specify speech intelligibility requirements as a system parameter to be met and then verified, once the installation is complete, see [1, 2] and more. This applies not only to purposes of emergency states but also to environments such as lecture halls, theaters, etc., where speech intelligibility is important.

Speech intelligibility can be understood as the amount of information available in the transmitted speech signal and thus it is a fundamental aspect of quality of the speech signal transmitted through the transmission channel. Standardized methods exist to objectively assess speech codecs, such as ITU-T P.862 (PESQ) [3], ITU-T P.863 (POLQA) [4] or ITU-T P.563 [5] and other methods such as PEMO-Q [6] or ViSQOL [7]. These methods estimate the quality focusing on signal processing in voice coders and common problems in voice transmission through data networks. However, these issues are not directly related to the intelligibility of PA systems. The room acoustics must be taken into account as well. More recent approaches for assessing intelligibility exist beyond STI(PA), such as STOI (short-time objective intelligibility) [8, 9], but they are not yet subject to standardization.

The Speech transmission index (STI) is a well-established objective predictor of how much speech intelligibility is degraded after passing a transmission channel. The STI of a particular transmission channel is obtained based on a comparison between measured signal at the output of the channel and the test input signal. The quantification of STI is a standardized procedure [10].

The history of STI dates back to 1970s, when a metric of speech intelligibility was first proposed and then adjusted [11]. Starting from the first edition in 1988, the methodology of the standard has undergone several

modifications, corrections and extensions. The third edition of the standard (2003) came up with a significantly accelerated way to determine the STI, under the abbreviation STIPA. The currently valid standard [10] from 2020 is actually its fifth edition.

Despite the fact that the usage of the STIPA method is widespread, it seems that reliable, high-quality implementation of STIPA remains in the sector of audio measurement equipment producers. STIPA modules are present (or can be purchased separately) in the devices of NTi Audio, Audio Precision, Brüel&Kjær, Embedded Acoustics, Bedrock Audio and others.

This fact was the first motivation point to implement the STIPA method and make it publicly available. A few repositories can be found online today, but we did not come across a code that would closely follow the standard [10]; as an example, the GitHub repository of Jon Polom[1] implements an STI estimation based on acquisition of real speech, which actually does not rely on the standard.

The second motivation was much practical: STIPA as an integral part of an audio measurement device limits the scope of possible scenarios of its usage. For example, one could need to compute STI offline, even multiple times, with different speech enhancement filters plugged in the channel. Actually, in our research described in [12], we needed an offline STI estimation where the digital audio signal was obtained via demodulation of interferences from an optical cable exposed to acoustic vibrations.

2 STIPA Theory & Implementation

A speech signal passes through a transmission channel, which can be simply a room, a telephone line, or an electro-acoustic channel consisting of a microphone, amplifier and speaker. The transmission can involve certain types of signal processing, either in the analog or digital form. The typical usage, limitations of STI model and of the STIPA method are explained in [10].

The STI takes into account physical properties of the transmission channel, and summarizes the ability of the channel to preserve speech intelligibility numerically, based on the difference between the input and output signals. The STI is a single real number ranging between 0 and 1; STI closer to 1 means a better speech intelligibility and vice versa, see Fig. 1. For example, an STI of 0.58 corresponds to a situation of 'high quality PA systems', present in concert halls and modern churches; with such an STI, a native listener should be able to understand complex messages in a familiar context.

Fig. 1: The indicative STI scale as present in annexes H and I of the standard [10].

The standard specifies two options how to derive the Speech transmission index. The *direct method* utilizes a speech-like measurement signal, while the *indirect method* is based on the measurement of the impulse response and is thus only applicable to linear, time-invariant systems. Thus, the direct method takes into account the effect of non-linear distortion to the speech intelligibility [10]. For our implementation, we chose the direct method since it covers a greater scope of measurement applications, including nonlinear distortions and strong additive noise component that we faced in our research [12].

In the following, we describe the fundamental steps of the STIPA direct method, as defined in the standard [10], and we make comments on our actual implementation. The Matlab source code is available at GitHub.[2]

Input signal The direct STIPA measurement requires an input signal which is actually a broadband pink noise modulated by two amplitude modulations independently in each of seven frequency bands. The resulting signal resembles the behavior of speech signals, but its advantage is the simplicity, reproducibility and independence on the language. To produce such a signal (of recommended length 15–25 seconds), a noise generator and a set of seven filters are the main components needed. We used a half-octave filterbank of order 20. A spectrogram of an excerpt from a STIPA signal is depicted in Fig. 2.

Formally, the STIPA signal is a mixture

$$\sum_{k=1}^{7} G_k N_k(t) A_k(t),$$

[1] https://github.com/jmpolom/sti-wav

[2] https://github.com/zawi01/stipa

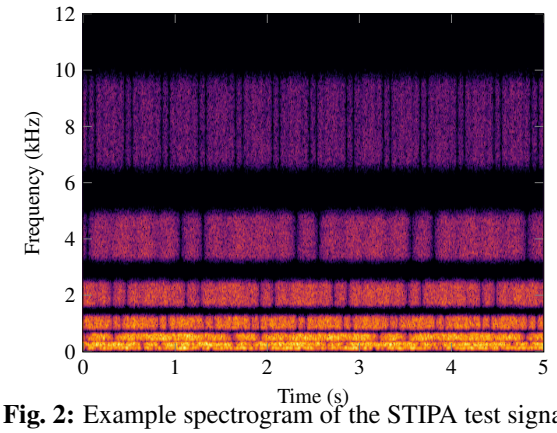

Fig. 2: Example spectrogram of the STIPA test signal.

where

$$A_k(t) = \sqrt{0.5\,(1+0.55(\sin(2\pi f_{1k}t) - \sin(2\pi f_{2k}t)))}.$$

The factor of 0.55 is actually the modulation depth, common for all modulation frequencies and frequency bands.

Notation:

k	octave band number
G_k	octave band weighting factor, $G_k = 10^{L_k/20}$
L_k	level in dB in octave band k (given by the standard)
$N_k(t)$	band-limited noise-carrier signal
$A_k(t)$	amplitude modulator, see above
f_{1k}, f_{2k}	two modulation frequencies per band.

Measurement chain The input signal, shortly the STIPA signal, is fed into the transmission channel and the output is acquired in the digital form.

STI computation To obtain the STI, a number of steps must be followed on the recorded signal that are enumerated below. In a nutshell, the transmission channel decreases modulation depth of the STIPA signal components; such reductions have to be quantified and mixed into a single final number.

1. **Band-filtering** – the signal is fed into a filterbank to split it to seven frequency bands. The standard defines the filters rather generally but they must achieve a minimum of 42 dB attenuation at the center frequency of each adjacent band [13]. We use filters of order 18, designed by the Matlab tool octaveFilter. The filter design technique uses mapping the desired filter to a Butterworth analog prototype, which is then mapped back to the digital domain [14]. Based on our practical observations, we also cut off the first 200 ms of all resulting signals to safely avoid transient effects of the IIR octave filters.

2. **Envelope detection** – the intensity envelope has to be determined by squaring the outputs of the bandpass filters, followed by a low-pass filter with a cut-off frequency of approximately 100 Hz. We used the lowpass function of Matlab with precisely 100 Hz as the passband frequency.

3. **Calculation of modulation depths** – the modulation depths for each octave band and modulation frequency have to be estimated. Such procedure must be always carried out over a whole number of periods for each modulation frequency, otherwise the estimation of the depths would be biased. We achieve this by simply cutting off the suitable number of signal samples from the signal end.
The modulation transfer ratio m_{k,f_m} in band k at the frequency modulation f_m is calculated as the ratio of the output and input depths,

$$m_{k,f_m} = m_\mathrm{o}(k,f_m)/m_\mathrm{i}(k,f_m),$$

where

$$m_\mathrm{o}(k,f_m) = \frac{2\sqrt{[\sum_k I_k(t)\cdot\sin(2\pi f_m t)]^2 + [\sum_k I_k(t)\cdot\cos(2\pi f_m t)]^2}}{\sum I_k(t)}$$

where $I_k(t)$ is the envelope in octave band k as the function of time (see previous point).
Our implementation allows to set all of $m_\mathrm{i}(k,f_m)$ to 0.55, which is the nominal modulation depth of the input STIPA signal. This regime is default when no reference signal is passed to the `stipa` function. Otherwise, these depths are calculated analogously to the above expression.

4. **Limiting** m_k – to avoid complex values in the SNR value computed further, the modulation transfer values are limited to 1 if they exceed it:

$$m_{k,f_m} = \min(m_{k,f_m}, 1).$$

5. **Taking ambient noise into account** – When STIPA measurements are carried out in noiseless conditions, this step can predict the intelligibility index for the case of presence of ambient noise. This assumes that the intensity of the ambient noise is measured separately. STIPA adjusts the modulation indexes following

$$m_{k,f_m} = m_{k,f_m} \cdot \frac{I_{s,k}}{I_{s,k} + I_{n,k}}.$$

Here, $I_{s,k}$ represents the acoustic intensity of the test signal in band k, and $I_{n,k}$ is the corresponding intensity of ambient noise.
In our implementation, this step is performed when both vectors of test signal levels and ambient noise levels in octave bands are provided. Otherwise, no adjustment is done and coefficients m_{k,f_m} are just retrieved from step 4.

6. **Taking auditory effects into account** – Since frequency-dependent auditory effects take place in real situations, the STIPA methodology incorporates these effects in the form of a reduction of the modulation transfer function. The modulation indexes are adjusted such that

$$m_{k,f_m} = m_{k,f_m} \cdot \frac{I_k}{I_k + I_{\mathrm{am},k} + I_{\mathrm{rt},k}}.$$

Here, $I_k = I_{s,k} + I_{n,k}$ denotes the total acoustic intensity discussed in step 5. The term $I_{\mathrm{am},k}$ stands for the auditory masking in octave band k and is computed via

$$I_{\mathrm{am},k} = 10^{L_{k-1}/10} \cdot 10^{L_{\mathrm{a},k}/10}$$

where L_{k-1} is the intensity in band $k-1$, and $L_{\mathrm{a},k}$ are given by Table A.2 in the standard [10]. Thus, masking effects are not modeled in band $k = 1$.

Further on, $I_{\mathrm{rt},k}$ takes into account the absolute reception threshold, which involves the absolute threshold of hearing and the minimal required dynamic range for a correct recognition of speech. This quantity is again dependent on frequency, and

$$I_{\mathrm{rt},k} = 10^{A_k/10}$$

is determined by coefficients A_k from Table A.3 of the standard. Auditory effects are taken into account only when the signal is obtained acoustically and when the total intensity in octave bands are known. In our implementation, step 6 is performed when the vector of the octave-band signal levels is provided.

7. **SNR computation** – the value of the effective SNR is computed from the limited modulation transfer values,

$$SNR^{\mathrm{eff}}_{k,f_m} = 10\log_{10}\frac{m_{k,f_m}}{1 - m_{k,f_m}},$$

and the result is limited such as not to exceed the range of $[-15, 15]$ dB.

8. **Transmission index** – the index is determined for the SNR value in each band:

$$TI_{k,f_m} = \frac{SNR^{\mathrm{eff}}_{k,f_m} + 15}{30}.$$

Clearly, such an index resides in the interval $[0, 1]$.

9. **Modulation Transfer index (MTI)** – the MTI of each band is computed via taking the average value over the frequencies:

$$M_k = MTI_k = \frac{1}{n}\sum_{m=1}^{n} TI_{k,f_m}.$$

In the case of STIPA, actually, $n = 2$.

10. **STI computation** – Calculate the final value of the Speech transmission index as

$$STI = \sum_{k=1}^{7} \alpha_k M_k - \sum_{k=1}^{6} \beta_k \sqrt{M_k \cdot M_{k+1}},$$

where the first part STI takes into account the intra-band modulations and the second part depends on MTIs of adjacent bands. The factors α_k, β_k in the expression are gender-specific factors for octave band k, given in Annex A of the standard [10]. In the event that STI is greater than one, the result is clipped to one.

3 Validation

IEC 60286-16:2020 standard requires to verify any STIPA implementation using the test signals described in its Annex C. Additionally, Annex A contains useful suggestions that could be supported by tests. To test our implementation of STIPA, we utilize test signals developed by Embedded Acoustics, which are available

along with the description and Matlab source codes.[3] The implementation presented in this paper satisfies all the below-described verification tests. Once the test signals are downloaded, the tests can be run using `verificationTests.m` script from the repository.

Annex A.2.2 – weight factor test

- Five test signals, each with sine carriers with only two neighboring bands (125/250, 250/500, 500/1000, 1000/2000, 2000/4000, and 4000/8000 Hz) are modulated to check the weight (α) and redundancy (β) factors.
- When level-dependent features are disabled, the six target STI values are specified.
- In this test, we compute STI on each of the six signals, and when the difference between the target and the actual STI value is smaller than 0.001, the test is considered successful.

Annex A.3.1.2 – filter bank phase test

This annex describes several requirements and recommendations on the filters used in the octave filter bank:

- Shape of filters should comply with IEC 61260-1 [13], class 1.
- Input signal shall be band-split without loss of power.
- Filters should provide 42 dB minimum attenuation at the center frequency of each band.
- Filters can be either FIR or IIR.
- Phase should be as linear as possible to avoid distortions of the phase relationship of the amplitude modulations by the settling behavior of the filters. Phase characteristics of the filters shall not give rise to a systematic error higher than 0.01 STI for the range between 0.1 and 0.9 STI.

The properties of the filterbank were already discussed in the very first part of the STI computation procedure. Our filters accomplish the first four requirements; as for the last point, the phase test itself uses test signals generated based on two sine carriers per octave band located at the edge of the central 1/2-octave. Again, we compare computed STI obtained from test signals with the reference STI values. According to the standard, a systematic error higher than 0.01 should not be obtained. Our maximum error is 0.0019, which occurs for STI 0.9. The Mean Absolute Difference (MAD) is $4.5 \cdot 10^{-4}$.

Annex C.3.2 – direct method modulation depth test

Since the direct method uses a noise band carrier signal as the excitation signal, it is fairly easy to replace the noise with sine wave carriers and subsequently control the modulation depths to test the capabilities of the measuring algorithm. This test contains 11 test signals with different modulation depths from 0.0 to 1.0 in 0.1 steps. The computed modulation indexes, also called m-values, are then compared with the target modulation depths for each octave band and modulation frequency. The absolute value of the error between the computed and theoretical m-value shall not exceed 0.05 and the overall m-value errors shall not yield a systematic absolute error (offset) in the STI results greater than 0.01. The target STI values are provided with the testing signals. Our implementation provides a maximum absolute m-value error of 0.003 and has zero systematic error on the STI values.

Annex C.4.2 – direct method filter bank slope test

Filter bank slopes are checked using 100% modulated sine carrier in the observe band and non-modulated sine carriers in the adjacent octave bands. If the steepness is exactly 41 dB/octave, an m-value of 0.5 should be obtained, corresponding to the SNR of 0 dB. The m-values should be 0.5 ± 0.05 or higher. Our minimum m-value in this test is 0.53.

4 Verification Measurement

The STIPA test signal sampled at 48 kHz with 16 bit depth was generated using the `generateStipaSignal` function, and loaded into the NTi Audio MR-Pro device, which was responsible for the signal playback during the entire measurement process. An active loudspeaker (with no equalization active) was positioned at the typical location of the speaker, and its volume was adjusted as to achieve a "normal" sound level of 60 dBA at a distance of

[3] http://www.stipa.info/index.php/download-test-signals

Fig. 3: Measurement setup used in verification

1 meter from the loudspeaker. According to [15, 10], this level should be used when corrected speech level is unknown.

For capturing the broadcasted STIPA signal, a calibrated NTi Audio M4260 microphone was employed. The microphone was placed at various positions corresponding to the several listener's locations within the auditorium. The captured signal was routed to NTi Audio XL2 audio analyzer, which measured both the sound pressure level in dBA and the Speech transmission index (STI) using NTi's internal implementation. Simultaneously, the signal was recorded using laptop with Steinberg UR44 audio interface and Steinberg Wavelab v9.5 software. A schematic of the verification setup is in Fig. 3; an example setup of the loudspeaker and the microphone placement can be seen in Fig. 4. Characteristics of UR44 were verified using Audio Precision APx525 analyzer; the dynamic range according to AES17 is 93.3 dB and the frequency characteristic ripple is 0.1 dB in the range of 20 Hz to 20 kHz.

Subsequently, the recorded audio files were imported into Matlab. The initial and ending silence sections were cropped, and the STI values were computed using the `stipa` function. STI results from the NTi XL2 analyzer and the computed STI values are listed in Table 1, along with the measured loudness level.

To test the implementation in various conditions, measurements number 11 and 12 were intentionally designed to produce low STI values, even though the measured loudness level was relatively high. Specifically, measurement no. 11 was performed with a microphone placement on the floor facing away from the testing loudspeaker, and measurement no. 12 was performed in the presence of an additive white noise from another sound source.

The Mean Absolute Difference (MAD) of STI values provided by NTi XL2 and our implementation is 0.0033 and the Pearson's correlation coefficient is 0.9983. The results indicate a strong similarity between the STI values obtained from NTi XL2 and the proposed implementation, suggesting high agreement and reliability

Fig. 4: An example of loudspeaker and microphone placement in the auditorium.

of our method when compared to the commercially licensed device.

5 Conclusion

An open-source Matlab implementation of STIPA direct method was presented.[4] It closely follows the current standard [10] and it has been numerically verified. Our implementation, independent of a hardware measurement device, can widen the range of possible applications. In future, more extensions like those presented in Annexes of the standard may be implemented to build a more complete open-source tool. In this sense, authors are invited to contribute to the code.

Acknowledgments

Research described in this paper was supported by the Ministry of the Interior of the Czech Republic, program IMPAKT1, under Grant VJ01010035, project "Security risks of photonic communication networks".
The authors thank anonymous reviewers for their insightful comments and suggestions.

[4] https://github.com/zawi01/stipa

Table 1: Results from the verification measurements.

Meas. number	Level (dBA)	STI	
		NTi XL2	computed
1	61.2	0.73	0.73
2	60.7	0.65	0.65
3	59.4	0.51	0.50
4	60.6	0.59	0.59
5	60.3	0.53	0.53
6	62.4	0.72	0.72
7	62.4	0.69	0.70
8	59.2	0.49	0.50
9	59.2	0.50	0.51
10	64.5	0.76	0.75
11	65.6	0.52	0.52
12	63.0	0.43	0.43

References

[1] International Organization for Standardization, "Fire detection and alarm systems – Part 19: Design, installation, commissioning and service of sound systems for emergency purposes," revision 1, 2007.

[2] National Fire Protection Association, "NFPA 72 – National Fire Alarm and Signaling Code," 2022.

[3] Rix, A. W., Beerends, J. G., Hollier, M. P., and Hekstra, A. P., "Perceptual evaluation of speech quality (PESQ)-a new method for speech quality assessment of telephone networks and codecs," in *Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing*, volume 2, pp. 749–752, 2001, doi:10.1109/ICASSP.2001.941023.

[4] The International Telecommunication Union Telecommunication Standardization Sector (ITU-T), "P.863: Perceptual objective listening quality prediction," 2018.

[5] The International Telecommunication Union Telecommunication Standardization Sector (ITU-T), "P.563: Single-ended method for objective speech quality assessment in narrow-band telephony applications," 2004.

[6] Huber, R. and Kollmeier, B., "PEMO-Q—A New Method for Objective Audio Quality Assessment Using A Model of Auditory Perception," *IEEE Trans. Audio Speech Language Proc.*, 14(6), 2006, doi:10.1109/TASL.2006.883259.

[7] Hines, A., Skoglund, J., Kokaram, A., and Harte, N., "ViSQOL: an objective speech quality model," *EURASIP Journal on Audio, Speech, and Music Processing*, 2015 (13), pp. 1–18, 2015.

[8] Taal, C. H., Hendriks, R. C., Heusdens, R., and Jensen, J., "An Algorithm for Intelligibility Prediction of Time-Frequency Weighted Noisy Speech," *IEEE Transactions on Audio, Speech, and Language Processing*, 19(7), pp. 2125–2136, 2011, doi:10.1109/TASL.2011.2114881.

[9] Jensen, J. and Taal, C. H., "An Algorithm for Predicting the Intelligibility of Speech Masked by Modulated Noise Maskers," *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, 24(11), pp. 2009–2022, 2016, doi:10.1109/TASLP.2016.2585878.

[10] International Electrotechnical Commission, "Sound system equipment – Part 16: Objective rating of speech intelligibility by speech transmission index," 2020. Number IEC 60268-16:2020, edition 5.0.

[11] Steeneken, H. J. M. and Houtgast, T., "A physical method for measuring speech-transmission quality." *The Journal of the Acoustical Society of America*, 67 1, pp. 318–26, 1980.

[12] Dejdar, P., Mokrý, O., Čížek, M., Rajmic, P., Münster, P., Schimmel, J., Pravdová, L., Horváth, T., and Číp, O., "Characterization of sensitivity of optical fiber cables to acoustic vibrations," *Scientific Reports*, 13(1), 2023, doi:10.1038/s41598-023-34097-9.

[13] International Electrotechnical Commission, "Electroacoustics – Octave-band and fractional-octave-band filters – Part 1: Specifications," 2014. Number IEC 61260-1:2014, edition 1.0.

[14] Orfanidis, S. J., *Introduction to Signal Processing*, Prentice Hall, 1995, ISBN 0132091720.

[15] International Organization for Standardization, "Ergonomics – Assessment of speech communication," 2003. Number 9921:2003.