# Influence of time-varying talker directivity on the calculation of speech transmission index from speech in a room acoustical context

Adam OPSATA[1]; Densil CABRERA[2]; Manuj YADAV[3]

[1,2,3] Faculty of Architecture, Design and Planning, The University of Sydney, Australia

## ABSTRACT

Standard calculations of the Speech Transmission Index (STI) are most frequently determined using specialised test signals, but methods have also been developed that estimate an STI value by using real speech as the stimulus signal. However, these methods have previously been studied in highly controlled scenarios, and have rarely been tested using real human talkers in real acoustical environments. We conducted a test to study how the natural movements of real talkers within an actual room environment affects the calculation of speech-based STI, by comparing results to those produced by a stationary head and torso simulator. In addition to the challenge of analysing a non-ideally modulated test signal, using a human talker introduces time variance in the transfer function from reference microphone to receiving microphone (due to incidental body movements and time-varying directivity). This paper examines the extent of this effect using a 130 m$^3$ room with long and short reverberation times.

Keywords: Speech Transmission Index, Room Acoustics, Signal Processing, I-INCE Classification of Subjects Number(s): 63.3

## 1. INTRODUCTION

The speech transmission index (STI) is an objective metric designed to predict the intelligibility of speech transmitted through a system such as an acoustic environment or sound reinforcement system (1-3). It measures the extent that signal envelope intensity modulations are preserved in the presence of spectral and/or temporal degradations, such as noise and/or reverberation (4). It is based on the idea that the reduction of these temporal envelope modulations indicate a reduction in speech intelligibility. The speech transmission index has been shown to be highly correlated with speech intelligibility in a variety of acoustic environments (5). The STI metric can be applied to accurately predict the recognition of speech by both normal-hearing and hearing-impaired listeners (6-8), and has been validated in several languages (9, 10).

There are, however, cases where the calculation of STI using other signals, speech in particular, is desired, which has been addressed in some studies that derive STI from speech based signals with some success (see (11) for review). But testing these speech-based STI methods has been largely limited to simulated and highly controlled environments, which do not take into account the complexities of speaking in a real environment with factors such as variable background noise, reverberance, etc. Moreover, the effect of time-varying directivity of human speech on STI values has also not been explored. This study aims to examine the effect of talker directivity in real room acoustical environments, on the calculation of STI using speech-based methods. These methods were tested with unrehearsed speech from a human talker, which exhibited natural time-varying directivity, and the same speech played back through a source of fixed directivity (a head and torso simulator (HATS)). Furthermore, the talker and the HATS were recorded in two room acoustic environments, characterising furnished and non-furnished rooms.

The next subsection briefly reviews the traditional STI methods and the speech-based STI methods. This is followed by a description of the implementation of two speech-based STI methods that were used to determine the time-varying STI from speech. This is followed by the results and conclusions.

[1] aops9981@uni.sydney.edu.au
[2] densil.cabrera@sydney.edu.au
[3] manuj.yadav@sydney.edu.au

### 1.1 *Traditional* STI methods

The original STI method proposed by (12) uses artificial test signals as a probe stimulus. Octave bands of noise spanning the range of the speech spectrum (125Hz to 8kHz) are intensity modulated, one at a time, at frequencies typically ranging from 0.6 to 12.5 Hz in one-third octave band increments. The stimulus signals have a known modulation depth and can therefore be used to determine any reduction in modulation depth caused by the transmission system. These results are used to form a modulation transfer function (MTF) in which the fractional reduction in modulation depth is represented at each modulation frequency in each frequency band. A transmission index (TI per octave band) is formulated by averaging MTF values over each octave band. TI values per octave band are then weighted according to the importance of each band's contribution to intelligibility, and summed to form a single number STI result.

More recent versions of the STI method attempt to simplify the measurement process. For example, the STIPA method produces an STI value derived using a single test signal, which possesses simultaneous modulation of all frequency bands. STIPA uses fewer modulation frequencies in each band in order to create a sparse MTF (3). In the indirect method, the STI is calculated from the application of acoustic theory by using speech-to-noise ratios, and either reverberation times or the Fourier transform of the squared system impulse response (3, 13).

### 1.2 Speech-based STI methods

Determining STI from actual recorded speech was considered by Steenikan and Houtgast (4), however, further development was curtailed due to a lack of computing power at that time (11). Several methods have since been proposed to produce STI values by using actual speech as the stimulus signal (14-25).

Using speech as a probe stimulus has many potential benefits. Chiefly, it can be observed that the STI methods that are most often used are the ones with the easiest measurement procedure, and so an effective STI method that only uses two microphones and a talker would likely be popular. It would also allow for measurement in a wider range of situations: the actual intelligibility of speech in a populated auditorium could be obtained without subjecting an audience to unsettling test signals. Payton and Braida (19) assert that it is necessary to use speech as a stimulus probe if the effect of speaking style is to be included in the assessment of intelligibility. This would allow the metric to reflect the benefit of clear articulation in noisy or reverberant spaces found by (8, 26).

Payton *et al.* (20) point out that clearly articulated speech typically has envelope spectra with significantly larger contributions at low modulation frequencies (less than 5 Hz) than conversationally articulated speech, which contradicts the observation by Steeneken and Houtgast (4) that the envelope spectra for the two speaking styles are very similar. While this can be caused by the fact that clear speech is typically delivered at a slower rate, this trend can be observed even when conversational and clear speech are delivered at identical rates.

Short-time intelligibility metrics offer the potential to compute STI in time-varying background noise, which is a noted limitation in traditional STI methods. A speech-based short-time STI metric can also provide greater insight into how intelligibility fluctuates over the course of a sentence. Through their implementation of the short-time envelope regression method, Payton and Shrestha (27) suggest that people tend to speak in a manner such that intelligibility decreases over the course of a sentence when speaking conversationally, but holds constant when speaking clearly. There is significant interest in short-time intelligibility metrics that can track speech intelligibility in fluctuating environments, such as classrooms, offices, etc.

Aside from being hindered by computing power, the other limitation in Houtgast and Steeneken (4, 28) attempts to arrive at a speech-based STI was that when analysing the modulation envelope spectra of real speech signals, it is difficult to differentiate between the useful, natural modulations present in speech and detrimental components with similar spectra (11). The measured envelope spectra of speech degraded solely by additive noise and reverberation are often dominated by artifacts that contradict predictions made by acoustic theory (11, 15, 19, 28, 29, 30). These artifacts increase in intensity in instances when acoustic theory predicts a decrease, which impedes the creation of accurate MTFs.

Payton and Braida (19) introduced a method that attempted to overcome the influence of these artifacts by generating a MTF from speech signals and then truncating it at frequencies where the coherence between the clean and degraded signal envelopes in that octave band falls below 0.8. This method achieved limited success, and in many instances the MTF was severely truncated which made it less robust in a variety of conditions.

Speech has also been considered as a means to measure STI in certain types of nonlinear systems (14, 15, 17, 21, 30, 31, 32). Many speech transmission systems such as telephony and hearing aids use processing such as amplitude compression and noise suppression. This processing can introduce non-linear distortion, which can spuriously increase or decrease the modulation depth of standard test signals (21), and render traditional STI probes ineffective. Methods by (4,14) attempt to use speech or speech like noise as a test signal, however (21, 33) found that these methods are not able to predict the intelligibility of non-linearly processed speech. These speech-based STI methods were improved by (21), however, Ma (25) points out that their results were not validated by the perception of human listeners.

Overall, the success of speech-based STI methods proposed to date is contentious. According to Payton and Shrestha (27), speech-based techniques have been shown to provide nearly the same result as the traditional STI, in both noise and noise plus reverberation conditions. However, according to a review by Wijngaarden et al. (11), speech-based methods show promise but have only had limited success and need to be tested in a wider variety of environments. Speech-based methods have rarely been studied outside of highly controlled computer simulations, and so the influence of factors present in real acoustic contexts with time-varying directivity of talkers is not well known. Furthermore, previous studies have either focused on conditions of noise or noise plus reverberation. In the current study, speech-based STI methods that are presented in the following section were tested in reverberant conditions with very low background noise.

## 2. REVIEW AND IMPLEMENTATION OF SELECT SPEECH-BASED METHODS

The following methods involve placing a microphone near the mouth of the talker (reference signal) and microphones in several receiver positions (received signals). Each received signal is naturally degraded by the transmission system or acoustic environment. The reference signal aims to represent the anechoic stimulus speech of the talker, although in real world applications it is likely to be contaminated by the talker's environment.

These speech-based methods are similar in that they all extract envelopes from band-filtered speech, and have final steps that weight frequency band metrics. They mainly differ in how the envelope signals are computed and how they compute a modulation metric in each frequency band.

### 2.1 The Envelope Regression Method

The Envelope Regression (ER) metric was proposed by Ludvigsen et al. (14) and reformulated by Goldsworthy and Greenberg (21). The ER method can be applied when the analysis window contains the entire length of the speech signal, however it is particularly of interest when applied to small portions of speech to calculate time-varing STI, as implemented by Payton and Shrestha (27).

The ER method differs from traditional STI calculations in that it does not attempt to create a MTF. Instead, a "modulation metric" $M$ is calculated from the speech intensity envelope in each octave band: each segment is filtered into octave bands by using 6th order Butterworth filters, then squared, and low pass filtered by convolution with a 10 ms Hamming window. The envelopes are resampled to a rate of 408 Hz. The modulation metric $M$ for each band is calculated as:

$$M_i = \frac{\mu_{xi}}{\mu_{yi}} \frac{\frac{1}{N}\sum_{k=1}^{N}[x_i(k)y_i(k)] - \mu_{xi}\mu_{yi}}{\frac{1}{N}\sum_{k=1}^{N}[x_i(k)]^2 - (\mu_{xi})^2} \tag{1}$$

where $x$ is the clean intensity envelope, $y$ is the degraded intensity envelope, $i$ is the octave band, $\mu_{xi}$ and $\mu_{iy}$ are the means of the of the clean and degraded intensity envelopes respectively, and $N$ is the number of samples ($k$) in the analysis window.

The modulation metric is used to find the apparent signal to noise ratio, aSNR, in each band in accordance with the IEC standard (3) as

$$a\mathrm{SNR}_i = 10\log_{10}(M_i/1 - M_i) \tag{2}$$

The remainder of the calculation is performed in accordance with formulas and weightings prescribed in (3).

### 2.2 Normalised Covariance Method

Ma *et al.* (25) studied the performance of the STI-based Normalised Covariance Method (NCM), using short frames of speech degraded by environmental noise. The NCM was initially proposed by Hollube and Kollmeier (34) and later revised by Goldsworthy and Greenberg (21). The NCM was implemented as described Loizou (35). Like the ER method, NCM also calculates the STI value as a weighted sum of transmission index (TI) values, and avoids the use of a full MTF. The TI calculation is derived from the covariance between the octave-band envelopes of clean and degraded speech.

More specifically, speech segments are filtered into 20 frequency bands from 0 to 16 kHz. Speech Envelopes were extracted by using the Hilbert transform and then downsampling to 25 Hz. Next, the normalised covariance in each frequency band ($r_i$ )is calculated as

$$r_i = \sum_t (x_i(t)-\mu_i)(y_i(t)-v_i) \Big/ \left( \sqrt{\sum_t (x_i(t)-\mu_i)^2} \sqrt{\sum_t (y_i(t)-v_i)^2} \right) \tag{3}$$

where for each octave band ($i$), $x$ and $y$ are the envelopes of clean and degraded speech respectively, $\mu$ is the mean of $x$ and $v$ is the mean of $y$. Resulting values of $r_i$ are limited to a range between 0 and 1 and converted to a SNR in each band using

$$\mathrm{SNR}_i = 10\log_{10}(r_i^2 / 1 - r_i^2) \tag{4}$$

$\mathrm{SNR}_i$ is limited to ± 15 dB and used to obtain a frequency band TI by

$$\mathrm{TI}_i = (\mathrm{SNR}_i + 15)/30 \tag{5}$$

The NCM index is computed by averaging the transmission indices using

$$\mathrm{NCM} = \left( \sum_{i=1}^{K} W_i \times \mathrm{TI}_i \right) \Big/ \sum_{i=1}^{K} W_i \tag{6}$$

where $K$ is the total number of frequency bands, $W$ is frequency band weighting. Instead of the default weights provided in (36), the current study used the weights defined as

$$W_i^{(1)} = \left( \sum_t x_i^2(t) \right)^p \tag{7}$$

as these were found to be more consistent with subjective intelligibility Ma *et al.* (2009). Here $p$ is the exponent set to 3 by default in the code. NCM does not include weightings to model the upward spread of masking prescribed in (3).

This method was found by Ma *et al.* (25) to perform well in measuring nonlinear systems in the presence of fluctuating background noise and was used in the current study to analyse the measurements.

## 3. METHOD

The effect of time-varying talker directivity on speech-based STI metrics was studied by comparing results calculated from unrehearsed speech of a human talker whose directivity naturally varies over time, with results from same speech reproduced through a stationary head and torso simulator.

Two sets of measurements were conducted in the reverberant chamber at the University of Sydney. In the first measurement session the room was acoustically treated to have the characteristics of a furnished room. In the second set of measurements the room was untreated and empty other than the measurement equipment and two people (Figure 1).

In both measurement sessions, the talker was fitted with two DPA 4066 microphones: one mounted on the nose and one positioned 7 cm from centre of the mouth. For the purpose of analysis, the signal from the nose-mounted microphone is considered as the reference signal. Four Earthworks M31 omnidirectional microphones were placed around the room at varying heights and at least 1.2 meters from any reflective surface, in accordance with (3). The signals recorded by these microphones are considered to be the receiver signals. In the first measurement, a talker spoke for seven minutes and was instructed to use natural head movements as though speaking to a room full of people. All signals were recorded at a sampling rate of 48 kHz. In the second measurement, the speech recorded by the nose-mounted microphone was then reproduced through a Brüel & Kjær 41286c HATS. This speech was recorded by both the reference and receiver microphones.

The STI at each receiver location was also determined by using the indirect method. An impulse response was created for each receiver location, using a 60 s exponential sinusoidal sweep produced though the HATS. The level of the speech used for the indirect calculation was determined from the calibrated speech recordings according to Annex J.2 of (3). In order to compensate for differences in the level and spectrum of the human and HATS stimulus speech, separate indirect STI results were calculated using the J.2 speech levels from both the human and hats sources. The octave band background noise level at each receiver position was calculated from calibrated recordings from each receiver microphone.

The speech-based STI was calculated using the short form envelope regression method as described in Payton and Shrestha (27) and the Normalised covariance method as described by Ma *et al.,* (25), which are discussed in section 2 above. In both cases the measurements were analysed with 5 s analysis windows overlapping by 75 percent. Analysis was also conducted using different window sizes in order to study the effect window size has on the results.

The level of the HATS signal was reproduced within 1 dB of the original speech, however the HATS was not equalised, so the spectrum of the human and hats recordings is not identical. It was also found that noise introduced by the recording system was compounded in the HATS recording which can partially explain the lower speech-based STI results. This can be further refined in future studies.

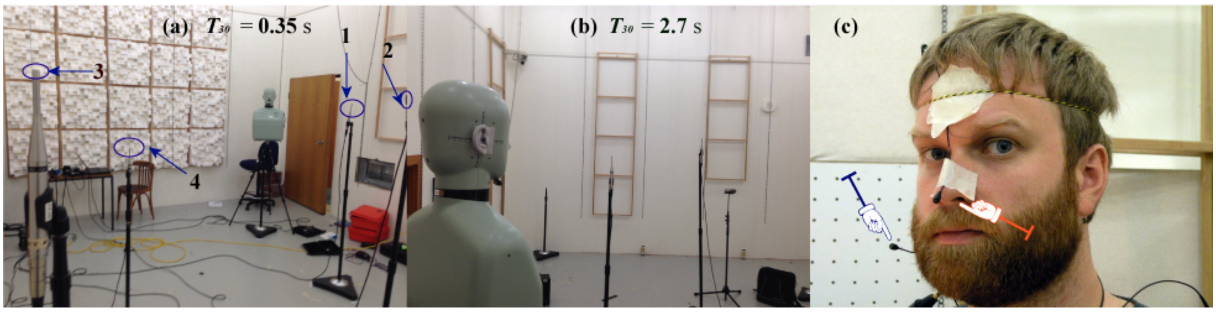


Figure 1 – (a) and (b) show the two measurement conditions with and without acoustical treatment, respectively. The reverberation time ($T_{30}$; mean of the 0.5, 1 and 2 kHz octave bands) for these two conditions were 0.35 s and 2.7 s, respectively. (a) also shows the 4 receiver microphone positions, marked 1-4, and the HATS in the talker position. (c) shows the nose-mounted (red indicator) microphone and the microphone placed 7 cm from the centre of the mouth (black indicator)

## 4. RESULTS AND DISCUSSION

The ER results were compared to theoretical method results in conditions of short and long reverberation times (M1 and M2 in Figure 2 (a)). The ER results are characterised by both their mean and 90th percentile values. The 90th percentile (S90) of the ER STI is shown to be a better summary statistic overall, although it slightly over predicts intelligibility when using long analysis windows (results from long analysis window are not presented in this paper). For very large window sizes the mean ER result approaches the theoretical method. In M2, the results from human and HATS are very similar in terms of both S90 STI and RMS error. The S90 STI is also well correlated with the indirect STI in M2. In M1, the speech-based short-term STI values for human and HATS diverge more than M2 (as quantified in higher RMS errors) in all the receiver positions, and produce results that are less reliable than when comparing them to the indirect STI. Assuming other factors being more or less consistent between M1 and M2, time-varying talker directivity is seen to have a greater impact on speech-based STI results in the lesser reverberant environment of M1. In conditions where traditional methods predict a high STI value, the ER results appear to be a less accurate measure of intelligibility, and more sensitive to small changes in SNR due to noise introduced by the measurement system. However considering it is not useful to differentiate between very high STI values (as 0.75 and 0.85 are both rated as excellent), this variance is not likely to impact this method's usefulness.

The NCM results are compared to theoretical results and shown in Figure 2(b). While the NCM results are generally in the same range as the ER, the S90 of the NCM results typically under-predict intelligibility in M1 and over predict in M2.

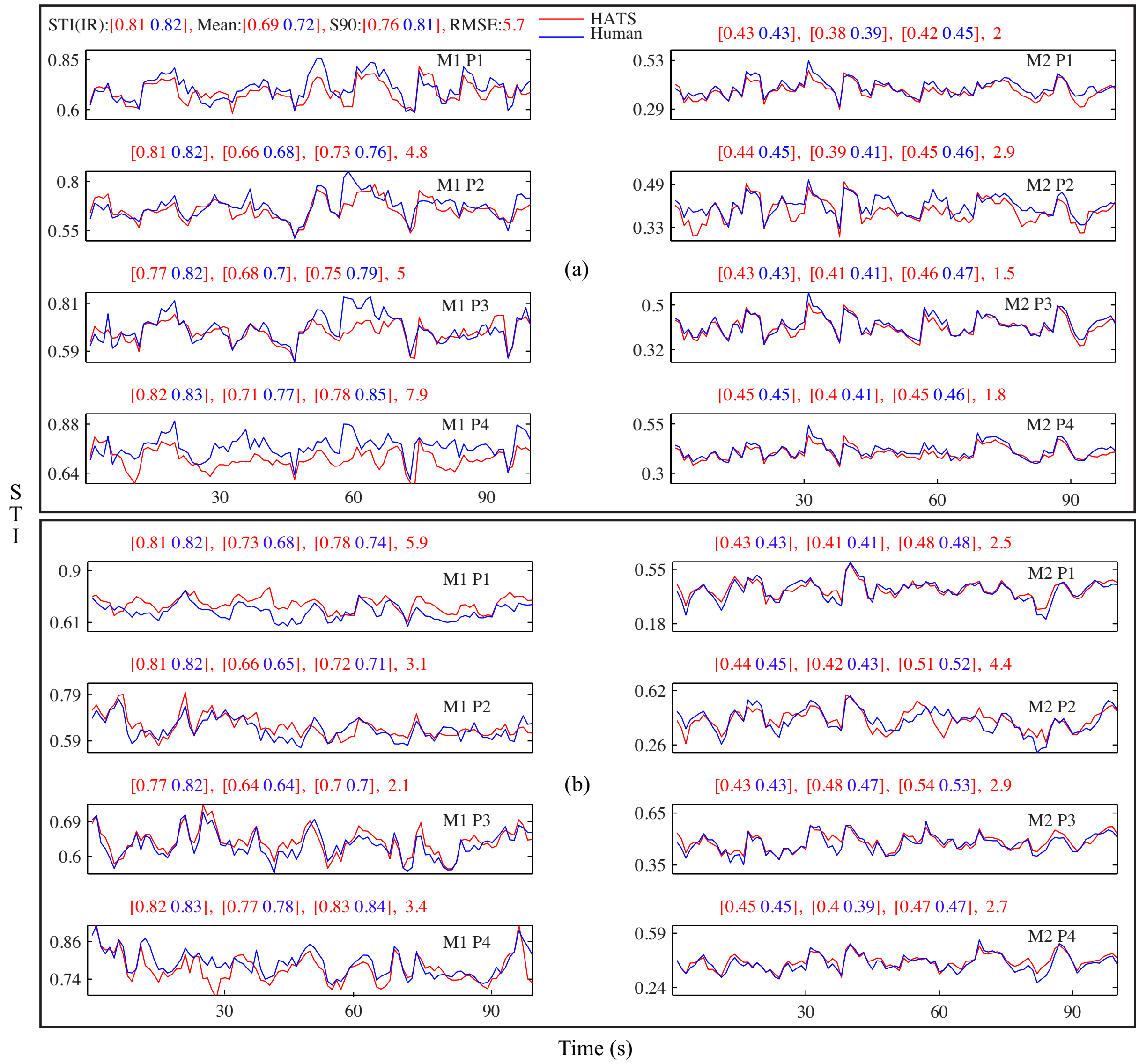


Figure 2 - The time-varying speech-based STI derived from the (a) ER method and (b) NCM for two room environments (M1 and M2) with four measurement positions (P1-4) per environment, as seen in Figure 1. Each subfigure (with ordinate scaled between the maximum and minimum STI values) plots the time-varying STI values for the HATS (red) and Human (blue) subject over a 5 s moving window. The title of each subfigure has four group of values: the overall STI (IR) values calculated from the *traditional* method using room impulse responses in the first bracket, followed by the mean of the time-varying STI, the 90$^{th}$ percentile (S90) of STI values and the root-mean-squared error in percentages. It can be observed that when speech is more severely degraded by reverberation (M2), the results from both speech based methods exhibit less variation due to talker directivity than in conditions where traditional methods predict a high STI value (M1).

## 5. CONCLUSIONS

The first aim of this study was to test the efficacy of speech-based STI measures in two real room acoustic environments characterised by relatively low and high reverberation for a fixed volume, using natural speech by a human talker. The second aim was to explore the effect of time-varying directivity of human speech, by comparing results from the speech-based STI measures derived using recorded speech of a human and a fixed directivity source (HATS).

The results indicate that both the speech-based STI measures (ER and NCM) provide useful indicators of intelligibility in quiet reverberant conditions regardless of time-varying talker directivity. For the relatively less reverberant condition, speech-based results were generally found to be less precise, and time-varying talker directivity showed a greater influence, and vice-versa for the more reverberant condition. It can, however, be argued that the lower precision in the results may not significantly impede the methods' usefulness, as the spread of error is acceptable within the range of STI values.

When short analysis windows are used, the overall intelligibility is better characterised by results in the 90th percentile. For longer analysis window (greater than 100 s) the overall intelligibility is better characterised by the mean of the results.

Overall, under both measurement conditions considered, the methods provided useful results, regardless of time-varying talker directivity. This is encouraging as far as further exploration of these speech-based STI methods in more room environments is considered, which are less intrusive to measure in occupied rooms when compared with the traditional STI methods.

## 6. REFERENCES


1. IEC Part 16: Objective rating of speech intelligibility by speech transmission index (2nd edition), in IEC 60268, 1998, Sound System Equipment (Int. Electrotech. Commiss., Geneva, Switzerland).
2. IEC Part 16: Objective rating of speech intelligibility by speech transmission index (3rd edition), in IEC 60268, 2003, Sound System Equipment (Int. Electrotech. Commiss., Geneva, Switzerland).
3. IEC Part 16: Objective rating of speech intelligibility by speech transmission index (4th edition), in IEC 60268, 2011, Sound System Equipment (Int. Electrotech. Commiss., Geneva, Switzerland).
4. Steeneken, H.J.M. and Houtgast, T., A physical method for measuring speech-transmission quality, *J. Acoust. Soc. Am.*, Vol. 67, 1980, pp. 318-326.
5. Houtgast , T. and Steenekcn, H.J.M., A review of the MTF concept in room acoustics and its use for estimating speech intelligibility in auditoria, *J. Acoust. Soc. Am.,* Vol. 77, 1985, pp. 1069-1077.
6. Humes, L. E., Dirks, D. D., Bell, T. S., Ahlstrom, C., and Kincaid, G. E. (1986). Application of the Articulation Index and the Speech Transmission Index to the recognition of speech by normal-hearing and hearing- impaired listeners, *J. Speech Hear. Res.* 29, 447–462.
7. Ludvigsen, C. ''Prediction of speech intelligibility for normal-hearing and cochlear hearing-impaired listeners,'' *J. Acoust. Soc. Am.* 82, 1987, 1162 – 1171.
8. Payton, K. L., Uchanski, R. M., and Braida, L. D.. Intelligibility of conversational and clear speech in noise and reverberation for listeners with normal and impaired hearing, *J. Acoust. Soc. Am. 95,* 1994, 1581–1592.
9. Anderson, W. B., and Kalb, J. T. English verification of STI method for estimating speech intelligibility of a communications channel, *J. Acoust. Soc. Am.* 81, 1987, 1982–1985.
10. Brachmanski, S. Estimation of logatom intelligibility with STI method for polish speech transmitted via communication channels, *Arch. Acoust*. 29, 2004, 555–562.
11. van Wijngaarden, S., Verhave, J., Steeneken, H. The speech transmission index after four decades of development, *Acoustics Australia,* Vol. 40 Issue 2, 2012, p134.
12. Houtgast, T., and Steeneken, H. J. M. Evaluation of speech transmission channels by using artificial signals,'' *Acustica* 25, 1971, 355 – 367.
13. Schroeder, M. R. Modulation Transfer Functions: Definition and Measurement *Acustica,* Vol. 49, 1981, pp. 179-182.
14. Ludvigsen, C., Elberling, C., Keidser, G., and Poulsen, T. Prediction of intelligibility of non-linearly processed speech, *Acta Otolaryngol*. Suppl. 469, 1990, 190–195.
15. Ludvigsen, C. The use of objective methods to predict the intelligibility of hearing aid processing speech, in Recent Developments in Hearing Instrument Technology, edited by J. Beilin and G. R. Jensen (Danavox/Stougaard Jensen, Copenhagen, Denmark), 1993, pp. 81–94.
16. Ludvigsen, C., Elberling, C., and Keidser, G. Evaluation of a noise reduction method-comparison between observed scores and scores predicted from STI, *Scand. Audiol.* 22, 1993, 50–55.
17. Drullman, R., Festen, J. M., and Plomp, R. Effect of reducing slow temporal modulations on speech reception, *J. Acoust. Soc. Am.* 95, 1994, 2670–2680.
18. Drullman, R. Temporal envelope and fine structure cues for speech intelligibility, *J. Acoust. Soc. Am.* 97, 1995, 585–592.
19. Payton, K. L., and Braida, L. D. A method to determine the speech transmission index from speech waveforms, J. Acoust. Soc. Am. 106, 1999, 3637–3648.

20. Payton, K. L., Braida, L. D., Chen, S., Rosengard, P., and Goldsworthy, R. Computing the STI using speech as a probe stimulus, in Past, Present and Future of the Speech Transmission Index, edited by S. J. v. Wijngaarden (TNO Human Factors, Soesterburg, The Netherlands), Chap. 11, 2002, pp. 125–138.
21. Goldsworthy, R. L., and Greenberg, J. E. Analysis of speech-based speech transmission index methods with implications for nonlinear operations, *J. Acoust. Soc. Am.* 116, 2004, 3679–3689.
22. van Wijngaarden, S., and Houtgast, T. Effect of talker and speaking style on the speech transmission index, *J. Acoust. Soc. Am.* 115, 2004, 38L– 41L.
23. Payton, K. L., and Shrestha, M. Analysis of short-time speech transmission index algorithms, *Acoustics* '08 Paris, 2008, 633–638.
24. Payton, K. L., and Shrestha, M. Evaluation of short-time speech-based intelligibility metrics, in 9th Internat. Congress on Noise as a Public Health Problem, edited by B. Griefahn (IfADo, Dortmund, Germany), 2008, pp. 243–251.
25. Ma, J., Hu, Y., and Loizou, P. Objective measures for predicting speech intelligibility in noisy conditions based on new band-importance functions, *J. Acoust. Soc. Am.* 125, 2009, 3387–3405.
26. Krause, J. C., and Braida, L. D. Investigating alternative forms of clear speech: The effects of speaking rate and speaking mode on intelligibility, *J. Acoust. Soc. Am.* 112, 2002, 2165–2172.
27. Payton, K. L., & Shrestha, M. Comparison of a short-time speech-based intelligibility metric to the speech transmission index and intelligibility data. *The Journal of the Acoustical Society of America*, 134(5), 2013, 3818-3827.
28. Houtgast, T., and Steeneken, H. Envelope spectrum and intelligibility of speech in enclosures, in *IEEE-AFCRL Conference on Speech Communications and Processing,* IEEE, Newton, MA, 1972, pp. 392–395.
29. Payton, K. L., Uchanski, R. M., and Braida, L. D. Computation of modulation spectra for the Speech Transmission Index using real speech, in *Proceedings of the 1993 IEEE Workshop on Applications of Signal Processing to Audio and Acoustics* (IEEE Signal Processing Society), 1993, Paper 5.4.
30. Hohmann, V., and Kollmeier, B. The effect of multichannel dynamic compression on speech intelligibility, *J. Acoust. Soc. Am.* 97, 1995, 1191–1195.
31. Plomp, R. The negative effect of amplitude compression in multichannel hearing aids in light of the modulation transfer function,'' *J. Acoust. Soc. Am.* 83, 1988, 2322–2327.
32. Villchur, E. Comments on 'The negative effect of amplitude compression in multichannel hearing aids in the light of the modulation transfer function,''' *J. Acoust. Soc. Am.* 86, 1989, 425–427.
33. van Buuren, R., Festen, J., and Houtgast, T. Compression and expansion of the temporal envelope: Evaluation of speech intelligibility and sound quality, *J. Acoust. Soc. Am.* 105, 1999, 2903–2913.
34. Hollube, I., and Kollmeier, K. Speech intelligibility prediction in hearing-impaired listeners based on a psychoacoustically motivated perception model, *J. Acoust. Soc. Am.* 100, 1996, 1703–1715.
35. Loizou, P. C. *Speech enhancement: theory and practice.* 2013, CRC press.
36. ANSI Methods for calculation of the speech intelligibility index, S3.5, 1997, American National Standards Institute, New York.