PROCEEDINGS of the
23rd International Congress on Acoustics

9 to 13 September 2019 in Aachen, Germany

Calculating the Speech Transmission Index in fluctuating noise: a data-driven approach in the short-term implementation

Nicola PRODI[1] and Chiara VISENTIN[2]

Department of Engineering, University of Ferrara, Italy

ABSTRACT

Everyday communication takes place in the concurrent presence of reverberation and background noise; the latter may have a fluctuating character and a speech-like spectrum, being for instance the result of multiple speakers talking together in the background (i.e., babble noise). The objective characterization of these listening conditions can be achieved by using a time-frame implementation of the Speech Transmission Index (STI) in the indirect scheme, named eSTI. One prerequisite of using the method is that the optimal time frame has to be determined. In this study, an experimental approach was used to determine the optimal time frame, defined as the one that provides coincident psychometric curves under stationary and fluctuating background noises. Matrixed-word listening tests were presented to 79 young adults with normal hearing. The speech reception task was presented under 26 listening conditions, created by varying signal-to-noise ratio, reverberation and noise type. By comparing the psychometric curves for the two noises, an interval of suitable frame durations was identified, ranging between 200 and 345 ms. Using a time frame within this interval thus ensures that the same eSTI value corresponds to the same predicted intelligibility, irrespective of the noise type.

Keywords: speech reception, fluctuating noise, speech transmission index

1 INTRODUCTION

Everyday listening takes place inside rooms with a longer or shorter reverberation, and often the noise background may have a fluctuating character, as happens for instance inside a public space with one or more unattended talkers. Evidence has shown that acoustical conditions that include a fluctuating noise masker yield improved speech intelligibility for normal hearing subjects due to the so-called "fluctuating masker release". Listeners are able to glimpse part of the signal when the masker is low, so that their performance improves compared to a stationary noise when both interferers are played at the same long–term root mean square (rms) level ((1),(2),(3),(4)). In order to objectively qualify speech intelligibility (SI) in the presence of a fluctuating masker, it is possible to extend the metrics based on the rms analysis of signal and masker. The basic idea is to follow the temporal fluctuations of both signal and noise by calculating their levels in consecutive time frames. Rheberghen and Versfeld (5) introduced this type of short-term analysis of sound levels to adapt the Speech Intelligibility Index SII (6) to modulated noises, and termed the resulting metric the Extended Speech Intelligibility Index (ESII). The maskers used for the validation in (5) were anechoic and included speech-like fluctuating noise, sinusoidally intensity-modulated speech and multi-talker noise. Later on, in order to achieve a better compliance with the SII, Rhebergen et al. (7) used shorter frames and added a forward masking calculation scheme. Finally, a validation of the ESII was presented that employed anechoic target signals masked by several types of maskers including anechoic speech and speech-like fluctuating noise, as well as real-life background sounds having unspecified environmental reverberation (8).

The effect of reverberation is neglected in the ESII/SII models and, to overcome this limitation, George et al. (4) complemented the ESII metric with the usage of the Speech Transmission Index STI (9). This twofold procedure proved sufficiently accurate in many tested cases.

Besides revising the SII concept, elaborations of the original STI indicator have also been explored to cope with noise fluctuations. A simple attempt in this direction was accomplished at first without the effect of reverberation (10) and was developed using speech–shaped stationary noise as the probe

[1] nicola.prodi@unife.it
[2] chiara.visentin@unife.it

signal in compliance with normative indications. Both signal and noise were cut into time frames whose duration was frequency-dependent and hence the SNR was calculated for each frame. This short-term "anechoic" STI returned results consistent with ESII under the same fluctuating background noises, thus including fluctuating speech-like noise maskers. A mandatory choice to employ the short-term STI was to resort to the indirect calculation scheme (9), where the estimate of the modulation losses due to reverberation is achieved by the modulation transfer function (MTF) of the impulse response, and the effect of SNR is added independently.

Payton and Shrestha (11) firstly implemented a frame application of the STI (called eSTI henceforth) but the eSTI has received little attention in the past. Only recently it has been further investigated by van Schoonhoven et al. (12). The authors discussed the conditions under which an impulse response measured in a noisy setting can still be suitable for the eSTI approach.

The present work investigates experimentally the application of eSTI to speech-spectrum maskers having both stationary and fluctuating envelopes in the presence of reverberation. To implement eSTI, the signal and the noise levels are calculated separately on a time-frame basis, while the effect of reverberation on the MTF is calculated from simulated noise-free impulse responses. As outlined in (11), a central issue to be clarified in the calculation scheme of eSTI is the appropriate duration of the time frame. This point is addressed in the present study and the existence of a suitable range of frame values is investigated for a group of sound fields.

2 MATERIALS AND METHODS

2.1 Acoustical conditions

A rectangular room with dimensions (length: 12 m; width: 8 m; height: 4 m) was simulated within the CAD acoustic software Odeon (Version 14.0, 2017). The modelled room boundaries were flat and the sound absorption coefficient was changed uniformly at all frequency bands in order to achieve four reverberant conditions. The reverberation time (T30) values, averaged across the 500-2000 Hz octave bands (called T_{mid} henceforth), were respectively equal to 0.30 s, 0.65 s, 1.00 s and 1.54 s. The scattering coefficient (δ) of all surfaces was set to the value of $\delta = 0.1$ to ensure a sufficiently even distribution of reflections. In Figure 1 the octave band values of T30 are shown.

A frontal speaker with the directivity of a human talker was placed slightly off the symmetry axis, at 2.5 m from the listener; both speaker and receiver were placed at 1.5 m height from the floor.

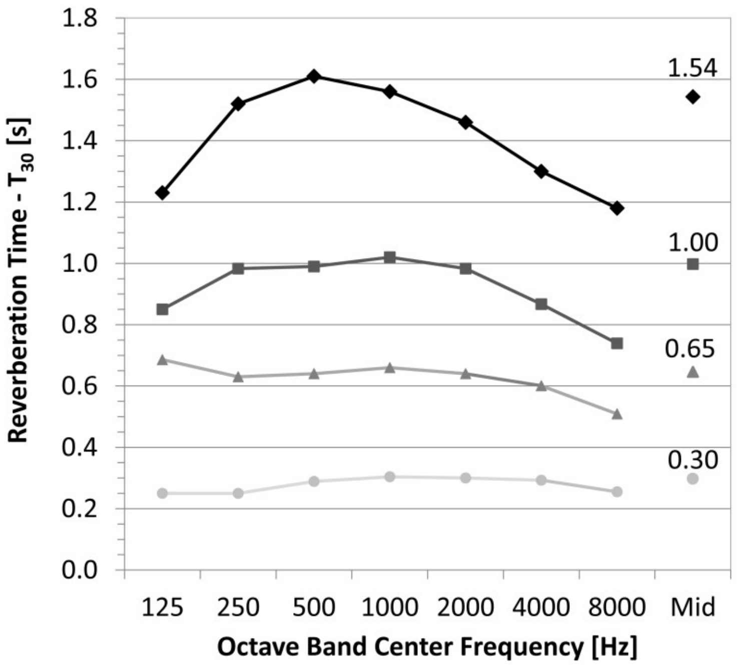

Figure 1 – Four simulated reverberation times T30, in the octave bands from 125 Hz to 8 kHz. The single numbers in the legend are the T_{mid}, which is the average T30 of the 0.5–2 kHz octave bands.

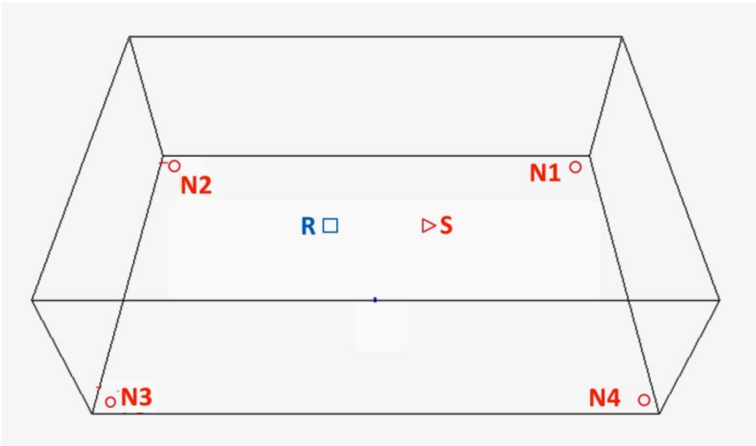

Figure 2 – A view of the room model with the target directional source (S) and the receiver in front of it (R). The four omnidirectional noise sources at the room lower corners are indicated as N1..N4.

Four omnidirectional sound sources were located at the four lower corners of the room in order to simulate a spatially distributed noise background. A view of the room model with the sources and the receiver is shown in Figure 2.

Binaural room impulse responses (BRIRs) were calculated separately for the speech signal source and for the four noise sources. Two continuous (no silent gaps) noise signals were used as maskers and they both had the octave-band spectral characteristics of female speech (9). The first background noise was stationary, and was derived from a steady–state pink noise signal which was spectrally shaped in octave bands to meet the required spectrum, and will be referred to as SSN. The second noise had speech-like fluctuations; it was obtained by processing Italian phrases spoken by a native female speaker, according to the established ICRA procedure (13). In order to reduce the possibility of associating a direction of arrival to the noise the coherence of the masker at the ears of the listener in each sound field was minimized.

As regards the target signal, two versions of the speaker's spectrum were developed, one for the shorter reverberations (T_{mid} = 0.30; 0.65 s) and one for the longer ones (T_{mid} = 1.00; 1.54 s). In particular the target signal kept the natural spectral character of the speaker in shorter reverberations whereas for the latter group the spectrum of the target signal was altered.

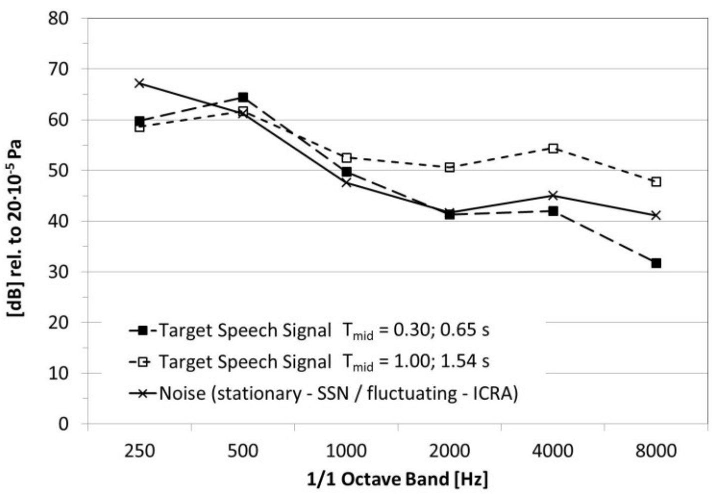

Figure 3 – Spectral content of the reverberated target signals: recorded signal, used in the conditions with T_{mid} = 0.30; 0.65 s (filled squares and long–dashed line), and signal manipulated, to mimic Lombard speech, used in the conditions with T_{mid} = 1.00; 1.54 s (white squares and short–dashed line). The spectral content of the reverberated background noises is also depicted (crosses and solid line).

This was done to account in a simple manner for the changes that a longer reverberation and hence a higher background noise has on speech production. The phenomenon is known as the Lombard reflex (14) and the so-called “Lombard speech” is characterized by several spectro-temporal alterations compared to the speech produced in quiet and anechoic conditions (15). Data from (16) were taken as a model in developing the tilted version of the speaker’s spectrum. This alteration of the speaker’s spectrum for the higher reverberations added generality and realism to the experimental conditions. The octave–band spectra of target signals and of noises are showed in Figure 3.

The level of the target speech was fixed at a long term rms value of 63 dB(A). The noise level was varied to achieve long-term rms SNRs spanning from +4.0 dB to -13.4 dB. In total, 26 sound fields were created. A group of 24 conditions were obtained by combining two types of noise (stationary and fluctuating) at three reverberation conditions (0.30, 0.65 and 1.00 s) each played back at four long-term SNRs (3 reverbs x 4 SNRs x 2 noises). Two further conditions were added, having reverberation time T_{mid}=1.54 s and one SNR only (1 reverb x 1 SNR x 2 noises).
The obtained STI values ranged between 0.10 and 0.47 and correspond to ratings of the SI either “poor” or “fair”, consistently with a challenging communication (17).

2.2 Implementation of the short–term analysis and of eSTI

The values of eSTI were calculated by using a speech-shaped stationary noise as a probe signal and thus fluctuations were left only to the fluctuating masker, not to the target signal. A Matlab script was used for the eSTI calculations which implemented the frame subdivision and calculated the SNR, the MTF from the impulse responses, the STI in each time frame according to the indirect method and finally output the eSTI values as averages of the frame ensemble values. Instead of averaging the two slightly different left and right eSTI values, it was decided to choose the highest in analogy with the better-ear approach described in (18).

By construction, if eSTI is evaluated over a sufficiently long time frame the values obtained for the stationary and the fluctuating noises coincide when the two noises share the same long-term rms spectrum. This communal eSTI for long frames corresponds numerically to the STI value for the stationary masker. When the same signals are processed with increasingly shorter frames the stationary noise shows negligible eSTI changes because its SNR does not change. On the contrary the eSTI values for the fluctuating noise undergo a remarkable increase because the SNRs increase with the shortening of the frames. This was already pointed out in (10) and the finding was motivated by the occurrence of larger eSTI values at the instants when the noise amplitude modulation is low and hence the SNR increases. Thus the resulting average eSTI value is larger for the fluctuating noise than for the stationary noise. Consistent with this finding, the eSTI metric can be employed to quantify the increase in SI which is expected due to the fluctuating masker release (4). In this study, listening tests were implemented to investigate which frame or frame intervals could be appropriate for calculating the eSTI value.

2.3 Listening tests

There were 79 participants, all of them native Italian speakers. They were recruited among the students and the academic staff of the local University, and paid a small allowance for their participation. Prior to the experiment, all listeners performed a self–administered hearing screening using the IOS–device based uHear application (19). All listeners obtained test results in the category of “normal hearing” (up to 25 dB HL) for the frequencies under testing (500–8000 Hz). They provided verbal consent prior to the experiment. Due to the extended design of the experiment and the large number of listening conditions investigated, the participants were randomly assigned to three groups, which had only a slight discrepancy in the gender distribution. No significant difference was found between the age distributions of the three groups when a Kruskal–Wallis test was performed (H=0.65, p=0.72).

The speech material used for the experiment was the recently developed Word Sequence Test (WST) in the Italian language, whose details are described in (20). In brief, the test is based on sequences of four disyllabic meaningful words (CVCV structure), that were selected among the corpus of the already available Diagnostic Rhyme Test in the Italian language (21), thus respecting the language-specific consonant phoneme distribution. Twenty-eight words were organized in a (7x4) base word matrix and the test sequences were created by sequentially selecting the words from the base matrix. The test was administered in a closed set format.

2.4 Procedures

The experiment took place in a sound treated room. As in previous studies (22), a three-dimensional audio rendering system based on seven pairs of loudspeakers surrounding a single listener seated in the center of the room was employed in the playback. In the system each pair of loudspeakers is processed independently from the others with cross-talk filters for trans-aural rendering.

Prior to the experiment, one test list composed of 12 sequences was presented to the listeners, at a fixed SNR of +10 dB, in stationary noise and anechoic conditions. The aim was to get the listeners familiarized with the test procedure and the stimulus material. After this phase the experiment started, and a test list of 12 trials was presented for each listening condition. In order to minimize the influence of sequential and learning effects, acoustic conditions, background noises and test lists were randomized among each group of participants. Furthermore, to avoid listeners' fatigue, a small break was proposed after the conclusion of the first half of the experiment. For each participant, the score (correct/incorrect) for each word composing a sequence was acquired and used to evaluate the SI, defined as the percentage of words correctly recognized within a sequence.

2.5 Statistical analysis

The dependent variable of interest in the study was SI; the independent variables were noise type (two levels: stationary and fluctuating) and acoustic condition, varying at 13 different levels for each noise type. A Generalized Linear Mixed Model (GLMM) was used to analyze the data. Post hoc analyses were based on pairwise comparisons of the means predicted by the GLMM model above; in order to account for planned multiple comparison, a Bonferroni correction was applied. All statistical analyses were conducted using the software R (23) and the significance threshold was set at 0.05.

3 RESULTS

3.1 SI data as a function of eSTI (time window: 12 ms)

Figure 4 (left panel) shows the SI results averaged across the participants for each listening condition as a function of eSTI calculated here using a time frame of 12 ms. A psychometric function was employed for describing the listener performance in the speech reception task as a function of the objective metric eSTI and the following logistic function was fitted to the data points:

$$\mathrm{SI\ (eSTI)} = \frac{100}{1+\exp\left(4s_{50}(\mathrm{eSTI}_{50}-\mathrm{eSTI})\right)}, \qquad (1)$$

where eSTI50 and s50 are the constants fully defining the logistic function. Specifically, the eSTI50 is defined as the eSTI required for a 50% intelligibility score, whereas s50 describes the slope of the function at its midpoint expressed in %/JND. In this definition it is assumed for the eSTI the same JND value as for the STI. The logistic curves in Figure 4 are the best-fitting regressions, found using a non-linear least squares method in Matlab (MathWorks, 2015). Despite the fact that the data points were derived from listening conditions with different SNRs and reverberation times, and there was also a mismatch between the speakers' and maskers spectra, the overall fits could still be considered as "average" according to the rating of psychometric curves from audiological testing.

The thresholds eSTI50 were equal to 0.18 and 0.23 for the stationary and the fluctuating background noise respectively. In contrast, very similar slopes were found for the two noises, equal to 7.0 %/JND for the stationary noise and 6.9 %/JND for the fluctuating one. As it can be seen in Fig. 4 (left), using eSTI with a frequency-independent 12 ms frame window for mapping SI, two almost parallel psychometric curves were obtained, whose distance can be approximated by the difference of the eSTI50 values. This difference is equal to 0.05, and is thus larger than the reference JND for STI.

3.2 SI data: identification of an appropriate interval of time frames for the eSTI metric

eSTI values were obtained for different time windows with a 5 ms step, and step by step the statistical significance of the effect of noise type on the SI results was checked. This systematic statistical analysis showed that it was possible to identify a range of time frames for which the two psychometric curves cannot be statistically discriminated since the effect of noise on SI becomes statistically not significant. In particular, this behavior was verified within the interval (200, 345) ms. Further increasing (or decreasing) the time window beyond these limits yielded a significant main effect of noise type (time window of 350 ms: p=0.040; time window of 195 ms: p=0.038) but no significant interaction between noise and eSTI.

Unfortunately, the procedure did not output an unambiguous single time frame but, for practical purposes, the time window having the mid interval duration, that is 272 ms, was deemed appropriate and was taken as a reference for later elaborations. Specifically, the statistical analysis indicated that, when this reference time window was used, the main effect of eSTI was still significant ($\chi 2(1)=2122.1$, $p<0.001$) but neither the noise type nor the interaction of the two factors had significant influence on the SI results (main effect: p=0.52; interaction: p=0.69). Figure 4 (right panel) shows the SI results as a function of eSTI with the reference time frame. The RMSE values of the regression curves from the data points were 4.0% (stationary noise) and 4.2% (fluctuating noise). The thresholds and the slopes coincided, being eSTI50 = 0.17 and s50 = 6.3 %/JND.

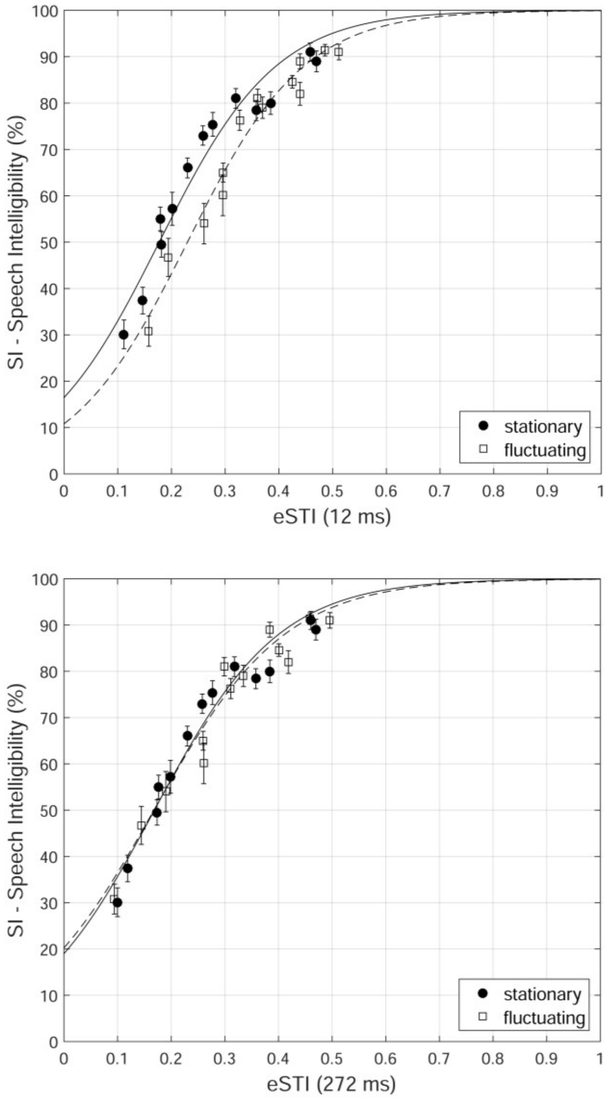

Figure 4 – Speech intelligibility (SI) results averaged across the listeners, with 95% confidence intervals, as a function of the eSTI. Results are divided according to the background noise type, and the best–fitting regression curves with a logistic shape are also included: stationary noise (black circles, solid line), fluctuating noise (white squares, dashed line). Left: eSTI calculated using a time window of 12 ms; eSTI calculated using a time window of 272 ms.

4 DISCUSSION

The present results show that eSTI is a suitable metric to build psychometric curves describing the behavior of SI for room-acoustical conditions with added stationary or fluctuating speech-like noise. This is proved by the goodness-of-fit of the obtained regression curves in terms of RMSE. The values of RMSE are in fact in line with literature studies despite the fact that present data spring from an unusually large range of experimental conditions, including source/noise spectral mismatch, variable reverberation and a wide range of SNRs. Limiting such variation would probably have further improved the RMSE.

The present range of reverberation times was selected to investigate the impact of noise fluctuations in practical room-acoustics applications. In particular, higher reverberation values will most probably smear out the masker fluctuations, thus entirely cancelling the fluctuating masker release and making eSTI unnecessary. On the other hand, as regards short reverberations, values such as $T_{mid} < 0.3$ s are of limited practical importance being the least recommended in technical norms of room acoustics design, for instance in (24).

It is to be remarked that eSTI does not simply integrate the effect of SNR and reverberation but, as for STI, it includes also the position-specific modifications coded in the impulse response. It is known that the pattern of early reflections is crucial in the transmission of speech, and its effect is considered in the MTF by the Schroeder integral (25). The ability of eSTI to follow the position-specific characteristics was not entirely exploited in the present study, which employed a single position inside a room. This capacity will be verified in future studies, for instance adding the listening distance in the experiments and considering it as a dependent variable in the statistical models.

The potential fields of usage of eSTI in practical applications are multiple. First, in room acoustics, a description of the speech reception in more realistic noise backgrounds would help in tailoring the design, according to the expected impact of fluctuating noise on reception accuracy. So, one natural application would be testing several types of speech-related noises in the process of design for speech transmission, and the evaluation of how the room acoustics can be shaped to optimize their control by taking both reverberation time and early reflections into consideration. Second, the STI approach was tested in the past with listeners having hearing impairment (26, 27) and its potentials in investigating the role of reverberation on this group of listeners was highlighted. The single number STI showed some advantage over SRT, and also an equivalent SRT for reverberant conditions was proposed, named SRRT (27). Anyway, much work needs to be done to confirm the present findings for hearing-impaired clinical population.

5 CONCLUSIONS

The present work investigated an adaptation of the STI indirect method to deal with a speech-like fluctuating masker. A range of fitting time windows 200 - 345 ms was found for the eSTI calculation, which all guarantee that the psychometric curves for the stationary and the fluctuating noise cannot be separated statistically, and can be thus considered coincident. Thus, given an eSTI value calculated within the interval (for instance with a conventional reference duration equal to 272 ms), it is ensured that the same SI results are obtained, irrespective of the character of the noise type. The present two maskers match those used in some previous anechoic models which were validated by using very short frame values. For this reason it is believed that the elongation of the time frame duration is mainly due to reverberation and not to the specific noise types. Anyway the borders of the interval may depend on the nature of fluctuating masker employed in the present study and further work is needed to investigate on the refinement of a suitable interval in order to fit a larger set of relevant fluctuating noises. An extended account of the present study can be found in (28).

REFERENCES

1. Festen, J. M., & Plomp, R. (1990). “Effects of fluctuating noise and interfering speech on the speech reception threshold for impaired and normal hearing,” J. Acoust. Soc. Am., 88(4), 1725-1736.
2. Gustafsson, H. A., and Arlinger, S. D. (1994). “Masking of speech by amplitude modulated speech,” J. Acoust. Soc. Am. 95, 518 - 529.
3. Versfeld, N. J., and Dreschler, W. A. (2002). “The relationship between the intelligibility of time-compressed speech and speech in noise in young and elderly listeners,” J. Acoust. Soc. Am. 111, 401–408.
4. George, E. l. J., Festen, J. M., and Houtgast, T. (2008). “The combined effects of reverberation and nonstationary noise on sentence intelligibility,” J. Acoust. Soc. Am. 124(2), 1269–1277.
5. Rhebergen, K. S., and Versfeld, N. J. (2005). “A Speech Intelligibility Index-based approach to predict reception threshold for sentences in fluctuating noise for normal-hearing listeners,” J. Acoust. Soc. Am. 117(4), 2181–2192.

6. American National Standard Institute (1997). ANSI–S3.5. Methods for Calculation of the Speech Intelligibility Index (American National Standard Institute, New York).
7. Rhebergen, K. S., Versfeld, N. J., and Dreschler, W. A. (2006). "Extended speech intelligibility index for the prediction of the speech reception threshold in fluctuating noise," J. Acoust. Soc. Am. 120(6), 3988–3997.
8. Rhebergen, K. S., Versfeld, N. J., and Dreschler, W. A. (2008). "Prediction of the intelligibility for speech in real-life background noises for subjects with normal hearing," Ear Hear. 29(1), 169–175.
9. International Electrotechnical Commission (2011). IEC60268–16 (edition 4.0). Sound system equipment, Part 16: Objective rating of speech intelligibility by speech transmission index (International Electrotechnical Commission, Geneva, Switzerland).
10. Ferreira, M., and Payton, K. (2014). "Using the short-time speech transmission index to predict speech reception thresholds in fluctuating noise," J. Acoust. Soc. Am. 135(4), 2224–2225.
11. Payton, K. L., and Shrestha, M. (2013). "Comparison of a short-time speech-based intelligibility metric to the speech transmission index and intelligibility data," J. Acoust. Soc. Am. 134(5), 3818–3827.
12. van Schoonhoven, J., Rhebergen, K. S., and Dreschler, W. A. (2017). "Towards measuring the Speech Transmission Index in fluctuating noise: Accuracy and limitations," J. Acoust. Soc. Am. 141(2), 818–827.
13. Dreschler, W. A., Verschure, H., Ludvigsen, C., and Westermann, S. (2001). "ICRA noises: artificial noise signals with speech-like spectral and temporal properties for hearing instrument assessment," Audiology 40(3), 148–157.
14. Lombard, E. (1911). "Le Signe de l'Elevation de la Voix (The sign of the rise in the voice)," Ann. Maladiers Oreille, Larynx, Nez, Pharynx (Annals of diseases of the ear, larynx, nose and pharynx) 37, 101–119.
15. Lu, Y., and Cooke, M. (2008). "Speech production modifications produced by competing talkers, babble, and stationary noise," J. Acoust. Soc. Am. 124(5), 3261–3275.
16. Valentini-Botinhao, C., (2014). "Intelligibility enhancement of speech in noise," Proceedings of Conference "Reproduced Sound 2014", Institute of Acoustics, St. Albans (UK) 36(2).
17. International Organization for Standardization (2003) ISO 9921, Ergonomics – Assessment of speech communication, Geneva, Switzerland.
18. van Wijngarden, S., and Drullman, R. (2008). "Binaural intelligibility prediction based on the speech transmission index," J. Acoust. Soc. Am. 123(6), 4514–4523.
19. Unitron. (2009). uHear (Mobile application software). Retrieved from http://itunes.apple.com.
20. Visentin, C., and Prodi, N. (2018). "A Matrixed Speech-in-Noise Test to Discriminate Favorable Listening Conditions by Means of Intelligibility and Response Time Results," J. Speech Lang. Hear. 61(6), 1497–1516.
21. Bonaventura, P., Paoloni, F., Canavesio, F., and Usai, P. (1986). "Realizzazione di un test diagnostico di intelligibilità per la lingua italiana (Development of a diagnostic intelligibility test in the Italian language)," (International Technical Report No. 3C1286), Fondazione Ugo Bordoni, Rome, Italy.
22. Prodi, N., Visentin, C., and Farnetani, A. (2010). "Intelligibility, listening difficulty and listening efficiency in auralized classrooms," J. Acoust. Soc. Am. 128(1), 172–181.
23. R Core Team (2017). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL https://www.R-project.org/.
24. Deutsche Institut für Normung (2016). DIN18041. Hörsamkeit in Räumen – Anforderungen, Empfehlungen und Hinweise für die Planung (Acoustic quality in rooms – Specifications and instructions for the room acoustic design (Deutsche Institut für Normung, Berlin, Germany).
25. Bradley, J. S., Reich, R. D., and Norcross, S. G. (1999). "On the combined effects of signal-to-noise ratio and room acoustics on speech intelligibility," J. Acoust. Soc. Am. 106(4), 1820–1828.
26. Duquesnoy, A. J., and Plomp, R. (1980). "Effect of reverberation and noise on the intelligibility of sentences in cases of presbyacusis," J. Acoust. Soc. Am. 68(2), 537–544.
27. George, E. L., Goverts, S. T., Festen, J. M., and Houtgast, T. (2010). "Measuring the effects of reverberation and noise on sentence intelligibility for hearing-impaired listeners," J. Speech Lang. Hear. Res. 53(6), 1429–1439.
28. Prodi, N., Visentin, C. (2019). "An experimental study of a time-frame implementation of the Speech Transmission Index in fluctuating speech-like noise conditions", Appl. Acoust. 152, 63–72.