Dear author,

Please note that changes made in the online proofing system will be added to the article before publication but are not reflected in this PDF.

We also ask that this file not be used for submitting corrections.

Applied Acoustics xxx (xxxx) xxx

Contents lists available at ScienceDirect

Applied Acoustics

journal homepage: www.elsevier.com/locate/apacoust

An experimental study of a time-frame implementation of the Speech Transmission Index in fluctuating speech-like noise conditions

Nicola Prodi, Chiara Visentin

Department of Engineering, University of Ferrara, Ferrara, Italy

ARTICLE INFO

Article history:
Received 18 January 2019
Received in revised form 22 March 2019
Accepted 25 March 2019
Available online xxxx

Keywords:
Speech reception
Fluctuating noise
Speech Transmission Index

ABSTRACT

Everyday communication takes place in the combined presence of reverberation and background noise, the latter having in some cases fluctuating characteristics and speech-like spectrum. To predict the speech intelligibility for fluctuating maskers, a time-frame implementation of the *Speech Transmission Index* (STI) in the indirect measurement scheme, named *Extended STI* (eSTI), has been recently proposed. Stationary speech spectrum noise is used as the probe signal and the a priori knowledge of the impulse response characterizing the transmission chain is required. A key issue of eSTI is the selection of the appropriate time frame for the calculations, whose duration is here assessed by a systematic experimental approach. Listening tests are developed for the scope, using speech-like stationary and fluctuating noises as maskers and rendering sound fields with various reverberation times and over a wide range of signal-to-noise ratios. An interval of frame durations providing equivalent values of eSTI under the two noises is identified. Within the target interval, statistically coincident psychometric curves are obtained for the two noises, thus ensuring that the same eSTI corresponds to the same speech intelligibility.

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 ([1–4]). 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 duration of the time frames ranged from 35 ms in the lowest critical band considered (150 Hz) to 9.4 ms in the highest (8 kHz); the values were derived by multiplying by a factor 2.5 the time intervals reported in studies on temporal gap detection [7]. 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. [8] 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 [9].

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 [10]. The values of the latter indicator were derived for several reverberation times and signal-to-noise ratios (SNRs) from a chart under the hypothesis of diffuse sound field [11]. This procedure gave an approximate evaluation of the effect of reverberation. The two effects (background noise and reverberation) were separated by means of the "STI at the 50% threshold" concept, meaning that a STI equal to 0.33 was matched with the measure of speech reception threshold (SRT the SNR granting 50% accuracy) for the

E-mail addresses: nicola.prodi@unife.it (N. Prodi), chiara.visentin@unife.it (C. Visentin)

https://doi.org/10.1016/j.apacoust.2019.03.026
0003-682X/

specific test material. By doing so, for a given reverberation time, it was possible to extract from the STI chart an equivalent SRT_{stat} for stationary noise which was then translated into an SRT_{fluc} for fluctuating noise by the usage of ESII charts for different types of modulated maskers. This twofold procedure proved sufficiently accurate in many tested cases. However, this approach has limitations. First, it is chart-based and thus off-line. Second, the position-specific details of the impulse responses (i.e. timing and energy of peculiar early reflections) are not considered in the estimate, since a purely exponential decay is hypothesized by the underlying diffuse field assumption. Third, the noise fluctuations, the SNR and the reverberation are assumed to influence SI independently. This is only an approximation since both signal and fluctuating noise are reverberated in the same physical space, and it is known that the combined effect of noise and reverberation on the speech reception is greater than the sum of the individual effects [12].

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 [13] and was developed using speech-shaped stationary noise as the probe signal in compliance with normative indications. Both signal and noise were cut into time frames whose duration was frequency-dependent [7] 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 [10], 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.

The main theoretical requirement to be fulfilled is the absence of non-linear processing or other critical elaborations in the transmission chain [14]. In this respect a linear system is expected most often in room acoustics, when a natural (that is, un-aided by electro-acoustics and/or by hearing aids) transmission between a talker and a listener is considered. Within this framework, after the impulse response is processed, the effect of the reverberated fluctuating noise is accounted for through a short-term analysis, which follows the course of reverberated signal and masker and hence their running SNR. By doing so, for each frame a single STI value is calculated and finally their average value describes the whole ensemble. Payton and Shrestha [15] implemented this frame application of the STI (called eSTI henceforth) as the theoretical reference model to validate another variation of the STI called sSTI; the latter quantity uses speech as a probe signal [16] and was conceived for the run-time monitoring of SI. Despite its simplicity and its straightforward derivation from the indirect STI method, the eSTI has received little attention in the past and only recently it has been further investigated by van Schoonhoven et al. [17]. The authors discussed the conditions under which an impulse response measured in a noisy setting can still be suitable for the eSTI approach.

Since it is derived from the well-established STI metric, the eSTI is deemed a candidate for the analysis of both stationary and fluctuating noise in the presence of reverberation and may have relevant potential in applications where the original STI is unreliable [10]. In particular, the most important characteristic required for such a metric is that it should provide output values directly comparable with the long-term STI ones. The property maintains backwards compatibility.

The present work investigates experimentally the application of eSTI to speech-spectrum maskers having both stationary and fluctuating envelopes in the presence of reverberation. The task is not that of developing a novel SI model but rather the work aims at a systematic experimental validation of one of the most well-known and used models, the STI, for its usage in a short-term design. 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 [15], a central issue to be clarified in the calculation scheme of eSTI is the appropriate duration of the time frame and its eventual dependence on frequency. Therefore, in the present study a preliminary step to the application of eSTI is to compare frequency-dependent and frequency-independent frames. Dependence of eSTI on frame duration is addressed and finally the existence of a suitable range of frame values is investigated for a group of sound fields. The tested conditions include reverberation times and a large span of SNR values which are representative of common and more demanding listening situations. Room-acoustical simulations of a mid-sized room, renderings and speech-in-noise tests are used to provide a set of realistic sound fields. Two types of maskers are employed in separate acoustical conditions. Both noises are continuous and have speech-like spectrum, but while the former is stationary the latter is fluctuating. Each sound field comprises a frontal speech signal and one of the two masking noises, which is output simultaneously from four omnidirectional loudspeakers located at the room corners. Since STI and eSTI use monaural information, they are not able to take into account specific directions of arrival of either noise or target signal. For this reason the sound fields to be used in the eSTI validation are processed in order to remove the direction of arrival of the noise. By doing so a spatially diffuse interferer is provided. In particular this is obtained by the superposition of noise sources that result incoherent at the ears of the listener (interaural coherence), and under this circumstances it is expected that the binaural SI is minimized due to the least binaural unmasking [18]. The usage of a diffuse masker rather than one fixed in space or even co-located at the target source position can also better approximate a spatially distributed disturbance which can be experienced for instance inside real-life public spaces, with spread and unattended speech sources. More specifically this study focused on three main research questions:

1. What is the course of SI with respect to eSTI in the presence of reverberation for speech-like stationary or fluctuating maskers in an incoherent noise?
2. Do the short-frame lengths of refs. [5] and [15] that validated ESII and STI for anechoic target and anechoic fluctuating and stationary speech-like noises also apply in the presence of reverberation with the same maskers?
3. Alternatively, is it possible to find an interval of frame lengths other than those used previously that ensure the overlapping of the psychometric curves for both stationary and fluctuating speech-like maskers? If so, the output eSTI values would be directly comparable with the long-term STI.

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 (T_{30}) 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 simulated T_{30} octave band values are reported in Fig. 1. The scattering coefficient (δ) of all surfaces was set to the value of $\delta = 0.1$ to ensure a suffi-

Please cite this article as: N. Prodi and C. Visentin, An experimental study of a time-frame implementation of the Speech Transmission Index in fluctuating speech-like noise conditions, Applied Acoustics, https://doi.org/10.1016/j.apacoust.2019.03.026

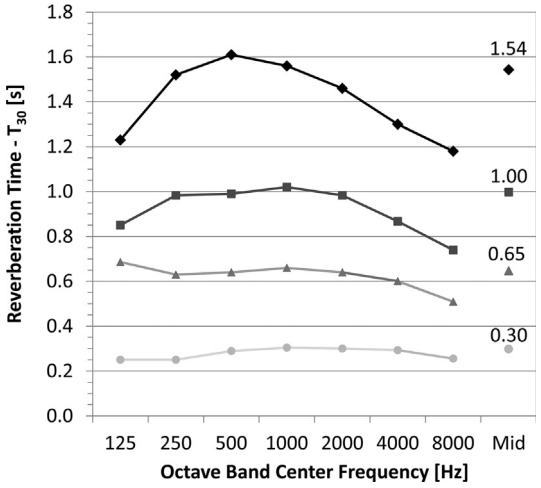

Fig. 1. Four simulated reverberation times T_{30}, in the octave bands from 125 Hz to 8 kHz. The single numbers in the legend are the T_{mid}, which is the average T_{30} of the 0.5–2 kHz octave bands.

ciently even distribution of reflections thus avoiding spurious acoustics effects from repeated lateral reflections (*flutter echoes*).

A frontal speaker with the directivity of a female 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. A set of Head Related Transfer Functions (HRTFs) of a B&K 4100 head and torso simulator, which had been measured previously by the authors, were inserted in the simulation software to obtain spatialized renderings. 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 Fig. 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 [10]. 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 [19]. The resulting ICRA signal is characterized by the same amplitude modulations at all frequency bands (co-modulation) that follow the envelope of the Italian speech. The obtained Italian ICRA noise signal is completely unintelligible. Moreover, 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. In particular the BRIRs of the four noise sources were convolved with four different short samples (5 ms) of white noise prior to mixing, thus ensuring a broadband randomization of the phases. By doing so, the noise shared the same room reverberation as the target signal and the fluctuations were kept too; however, the masker sounded entirely diffuse and the locations of the noise sources could not be identified.

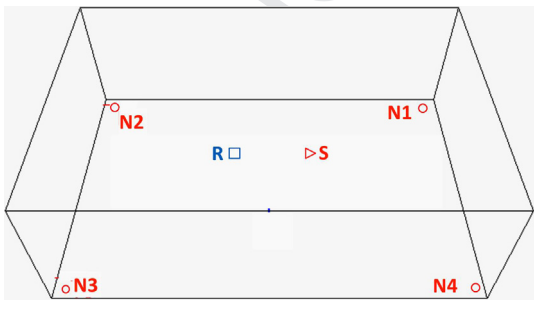

Fig. 2. A view of the room model with the target directional source (S) and the receiver in front of it (R). The four omnidirectional noise sources at the room lower corners are indicated as N1..N4.

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. This was done to account in a simple manner for the changes that a longer reverberation and hence an higher background noise has on speech production. The phenomenon is known as the Lombard reflex [20] and the so-called "Lombard speech" is characterized by several spectro-temporal alterations compared to the speech produced in quiet and anechoic conditions. In fact "Lombard speech" is more intelligible than speech in quiet when both are mixed with stationary noise at the same SNR, and the more the energetic masking, the more the speech production is altered to compensate for it [21]. Furthermore, the same authors [22] indicated that the most effective change in "Lombard speech" to improve intelligibility was the spectral tilt, that is a shift of the speech energy towards higher frequency bands (>1kHz). This known adaptation of speech was implemented in the present work for the target signal in the most reverberant conditions (T_{mid} = 1.00 s and 1.54 s), which were considered more prone to Lombard effects in practical applications. Data from [23] were taken 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 Fig. 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 both cases, the level was measured as the energetic average of the signal at the two ears by means of a B&K 4100 head and torso simulator placed at the listener position in the sound treated room where the listening tests took place. In total, 26 sound fields were created. A group of 24 conditions were obtained by combining two types of noise (stationary and fluctuating) at

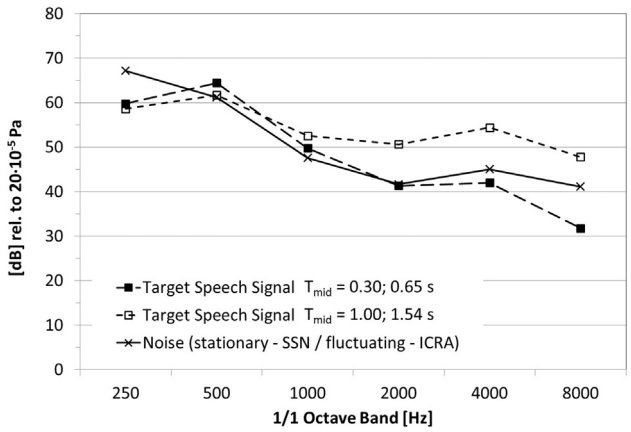

Fig. 3. Spectral content of the reverberated target signals: recorded signal, used in the conditions with T_{mid} = 0.30; 0.65 s (filled squares and long-dashed line), and signal manipulated, to mimic Lombard speech, used in the conditions with T_{mid} = 1.00; 1.54 s (white squares and short-dashed line).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

Please cite this article as: N. Prodi and C. Visentin, An experimental study of a time-frame implementation of the Speech Transmission Index in fluctuating speech-like noise conditions, Applied Acoustics, https://doi.org/10.1016/j.apacoust.2019.03.026

three reverberation conditions (0.30, 0.65 and 1.00 s) each played back at four long-term SNRs (3 reverbs × 4 SNRs × 2 noises). Two further conditions were added, having reverberation time T_{mid} = 1.54 s and one SNR only (1 reverb × 1 SNR × 2 noises). With longer reverberation stationary and fluctuating noises tend to behave similarly due to the smearing of the dips in the fluctuating noise. For this reason one SNR at T_{mid} = 1.54 s was deemed sufficient to extend the range of listening conditions without increasing the number of sound fields too much.

The complete set of the SNRs values, together with the STI values calculated for the stationary noise are reported in Table 1. STI values range between 0.10 and 0.47 and correspond to ratings of the SI either "poor" or "fair", consistently with a challenging communication [24]. In Table 1 one can evaluate the effect of the Lombard speech on the STI. For instance this is evident when comparing the conditions D and N. In fact they share almost the same SNR but, the latter being much more reverberant, one would expect a STI lower than in D. On the contrary, in N the spectral tilt of the higher octave bands which was implemented to account for the "Lombard speech" affects the STI calculation and causes a substantial increase of the parameter compared to condition D.

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. In the first place, the anechoic probe signal and the noise files (length 40 s) were convolved with the simulated BRIRs. Afterwards, the convolved files were played back through the audio rendering system in the sound treated room used for the experiments (see Section 2.3 for a description) and recordings were made with a B&K 4100 head and torso simulator placed at the listening position. The long-term levels of the probe signal were adjusted to coincide with the levels of the target signal which were also measured by the same means. Then, after level adjustment, the recorded probe and noise signals and the target impulse responses were input in a Matlab (MathWorks, 2015) script. The script implemented the frame subdivision and calculated the SNR, the MTF from the impulse responses, the STI in each time frame according to the [10] indirect method and finally output the eSTI values as averages of the frame ensemble values. The only improvement with respect to the normative algorithm for the MTF calculation consisted in extending the set of 14 modulation frequencies to 18, in order to improve effectiveness with respect to the evaluation of reverberation [25]. The resulting modulation frequencies span from 0.63 Hz to 31.5 Hz and are spaced 1/3 octave apart. The differences between the measured eSTI data at the two ears of the listener were always smaller than the just noticeable difference (JND) of the quantity which can be set equal to 0.03 in congruence with that of the STI [26]. 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 [27].

To start with, the calculation of the eSTI for the whole set of sound fields was accomplished with the frequency-dependent time frames reported in Table 2, which stem from the studies of [13] and [7]. Later on, the eSTI values were calculated according to the suggestion provided by [5], indicating a likely frequency-independent time frame 12 ms long. Fig. 4 reports the correlation between the two obtained eSTI data sets respectively with frequency-dependent and independent (12 ms) values. The two choices are equivalent given the strong positive linear correlation (R^2 = 0.99, $p < 0.001$) found for both noises and for the whole set of sound fields. Thus, the 12 ms frame replaced the frequency-dependent group of values in the subsequent elaborations. It has to be remarked that the MTF function is not touched by the choice of the frame duration because the octave band MTF values enter the eSTI calculation scheme as multiplicative factors independent of the frame.

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 [13] 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 bigger for the fluctuating noise than for the stationary noise. The behavior is exemplified in Fig. 5 where the trends of eSTI difference between fluctuating and stationary noise values (indicated as $eSTI_{ICRA} - eSTI_{SSN}$) are reported across the set of thirteen conditions. The plots are grouped from the top to the bottom panel according to reverberation time. In Fig. 4 for ease of reading the longest frame is set to 12288 ms ($12*2^{10}$ ms). The difference between the eSTI values of the two noises has always a decreasing trend with the lengthening of the time frame, and as expected it tends to vanish when the duration approaches the long term rms average. For most frames across the different reverberation times the values are larger than the expected JND which was set at 0.03. Also in the Figures one can observe a tendency to have a smaller dependence of $eSTI_{ICRA} - eSTI_{SSN}$ from SNR for lower reverberation times as well a slight trend of increase of the gap with SNR at fixed reverberation times. 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]. Although this qualitative analysis confirmed the potentials of the indicator, it did not specify which frame or frame interval could be appropriate for the scope. This latter task was accomplished experimentally in the next part of the work with the help of listening tests.

Table 1
Outline of the thirteen acoustical conditions listed from A to O. Each condition was used for both background noises giving raise to twenty-six sound fields. Reverberation times (T_{mid}), signal-to-noise ratios (SNR) and Speech Transmission Index (STI) values are reported. The T_{mid} values refer to the arithmetical average of the reverberation times in the octave bands centered at 500 Hz, 1 kHz and 2 kHz. The speech and noise levels were measured at the listener's positions with a B&K 4100 head and torso simulator during the playback. The STI values were calculated only for the listening conditions with the stationary noise as a masker.

Acoustical condition	T_{mid} [s]	SNR [dB]	STI
A	0.30	−9.6	0.12
B		−6.6	0.20
C		−5.6	0.23
D		−2.6	0.32
E	0.65	−9.0	0.10
F		−5.8	0.18
G		−2.3	0.26
H		+4.0	0.46
I	1.00	−13.4	0.17
L		−7.1	0.36
M		−5.9	0.38
N		−2.5	0.47
O	1.54	−4.8	0.28

2.3. Listening tests

2.3.1. Participants

There were 79 participants, all of them native Italian speakers. They were recruited among the students and the academic staff

Please cite this article as: N. Prodi and C. Visentin, An experimental study of a time-frame implementation of the Speech Transmission Index in fluctuating speech-like noise conditions, Applied Acoustics, https://doi.org/10.1016/j.apacoust.2019.03.026

Table 2
Time frames as a function of the octave bands used for the calculation of the eSTI values (Ferreira and Payton, 2014; Moore, 1997).

Octave band [Hz]	250	500	1000	2000	4000	8000
Duration [ms]	19.7	14.0	1.0	7	5	4

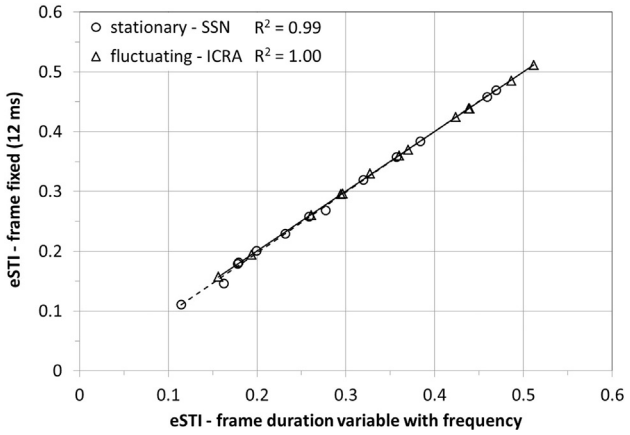

Fig. 4. Relationship between the eSTI calculated with a frame duration variable with frequency (Ferreira and Payton, 2014; Rhebergen and Versfeld, 2005) and with the fixed time-frame of 12 ms suggested in (Rhebergen and Versfeld, 2005).

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 [28], to rule out the presence of impairments ([29,30]). All listeners obtained test results in the category of "normal hearing" (up to 25 dB HL) for the frequencies under testing (500–8000 Hz). All the listeners 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 main task of the experiment was testing the effect of noise type, which was designed as a "within" participants variable. Indeed, each group was presented with a different subset of the 13 acoustic conditions but, in each group, the participants were presented with both noise types for a given condition. The assessment of the groups was deemed appropriate for the experimental design and for the later usage of the statistical model (see below par. "Statistical analysis").

2.3.2. Speech material

The speech material used for the experiment was the recently developed Word Sequence Test (WST) in the Italian language, whose details are described in [31]. 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 [32], thus respecting the language-specific consonant phoneme distribution. Twenty-eight words were organized in a (7×4) base word matrix and the test sequences were created by sequentially selecting the words from the base matrix. Each sequence includes a carrier phrase ("Ora diremo le parole…" which means: "Now we will say the words…") and four target words. By construction, the WST is low-context [33] and, unlike the matrix sentence test in the Italian language [34], the WST lacks a syntactic structure linking the target words. The WST was devised to be presented in a closed-set format.

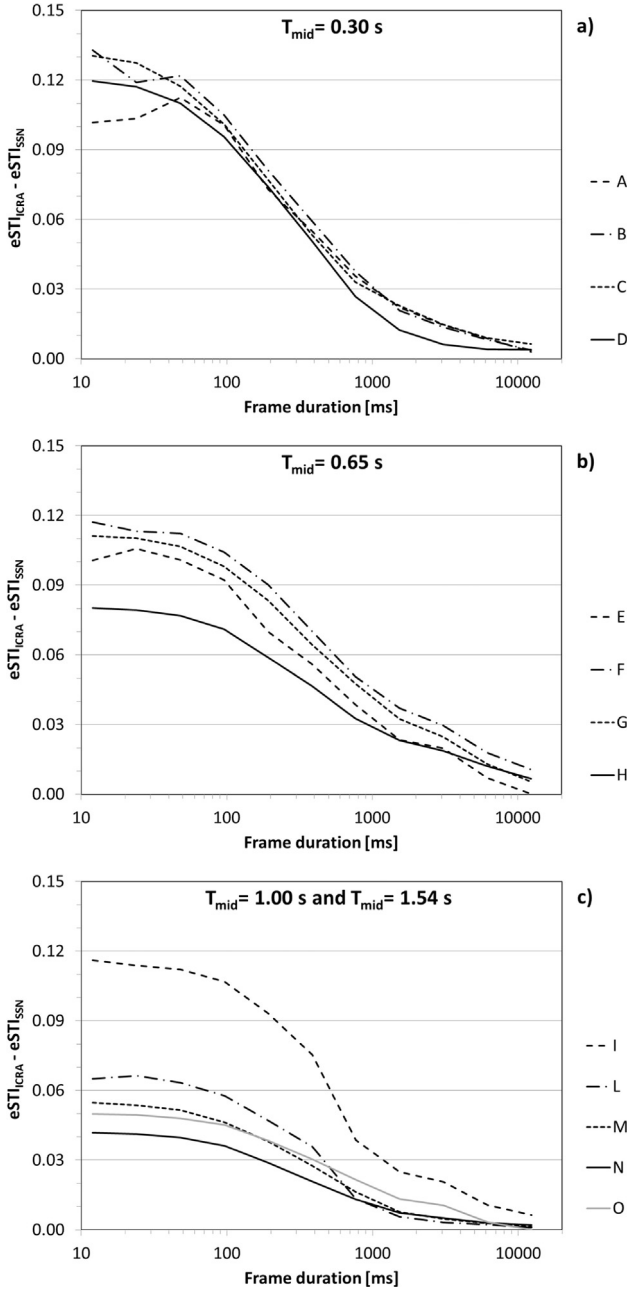

Fig. 5. Plots of the differences $eSTI_{ICRA} - eSTI_{SSN}$ calculated employing increasingly longer time frames ranging from 12 ms to an upper limit of 12288 ms ($12*2^{10}$ ms). The panels show data for the thirteen conditions used in the experiments (marked A to O as in Tab. 1). Data are grouped from top to bottom according to the different reverberation times. Upper panel a) $T_{mid} = 0.30$ s; mid panel b) $T_{mid} = 0.65$ s; lower panel c) $T_{mid} = 1.00$ s and $T_{mid} = 1.54$ s.

For the experiment, the test sequences were recorded in a sound-attenuated booth by an adult, native Italian female speaker, with a trained voice and expertise in stage reading. She was instructed to pronounce the test sequences at a conversational rate and maintaining a constant vocal effort.

Please cite this article as: N. Prodi and C. Visentin, An experimental study of a time-frame implementation of the Speech Transmission Index in fluctuating speech-like noise conditions, Applied Acoustics, https://doi.org/10.1016/j.apacoust.2019.03.026

2.3.3. Procedures

The experiment took place in a sound treated room. As in previous studies [35], 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. During the listening tests, the listener input the selected words via a touch-screen located close in front of her/him. The alteration to the sound propagation due to the touchscreen was negligible. The test presentation and the data collection were managed by means of an in-house LabVIEW® (Version 13.0; National Instruments) script. The application controlled through MIDI commands an audio rendering engine consisting of an AudioMulch® software [36] with the X–Volver VST plug-in [37] for real-time auralization hosted on a control PC placed outside the sound-attenuated room. The 14 signals were delivered through a Solid State Logic® Alpha-Link MX sound card to a set of Tannoy® Precision 8D loudspeakers.

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 and background noise to be evaluated. The participants listened to a sequence at a time; the background noise started almost one second before the carrier phrase and ended simultaneously with the final item of the sequence. Immediately at the end of the audio playback, the word matrix was displayed on the touchscreen. The participants had to mark the words they identified in serial order from the leftmost to the rightmost column; it was not possible to change a response once it had been selected. As the last word was selected, the next sequence was automatically played back. 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.

2.3.4. 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, on account of two main issues. First, the repeated-measures experimental design, in which multiple measures were collected for each participant. The model accounted for the random effects introduced by individual variability. Second, the non-normal distribution of the data. Indeed, the normal distribution of SI data cannot be granted for all acoustic conditions, and especially so for the most favorable ones where the metric undergoes a ceiling effect. In the statistical model, noise type, acoustic condition (tracked by the eSTI, considered as a continuous variable) and their interaction were included as fixed factors. As explained, the model also included random effects, modeling the participant as a random factor nested within the acoustic condition. A binomial distribution was used in the GLMM. The selection of the most effective model was based on a forward procedure using a likelihood ratio test. The consistency of the finally selected GLMM model was investigated by checking its assumptions; in particular, this implied a control of the normality of the random effect terms and the residuals, as suggested by [38]. 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* [39], and the *lme4* [40] and the *lsmeans* [41] packages. The statistical significance threshold was set at 0.05.

3. Results

In Section 3.1., the SI results are presented as a function of the eSTI in the form of suitable psychometric curves, calculated using a time frame of 12 ms in order to check if this duration fits conditions including reverberation over the fluctuating speech-like noise too. It is demonstrated that for this window duration the psychometric curves obtained for the stationary and the fluctuating noise are not coincident. Then, in Section 3.2, an appropriate interval of time windows for the eSTI metric is identified which yields coincident psychometric curves for the two maskers.

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

Fig. 6 shows the SI results averaged across the participants for each listening condition as a function of eSTI, which was calculated here using a time frame of 12 ms. Since reverberation is accounted for by the objective metric, all sound fields are grouped into one plot, and only one regression curve for each background noise was fitted to the data. 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:

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

where eSTI_{50} and s_{50} are the constants fully defining the logistic function. Specifically, the eSTI_{50} is defined as the eSTI required for a 50% intelligibility score, whereas s_{50} 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 Fig. 6 are the best-fitting regressions, found using a non-

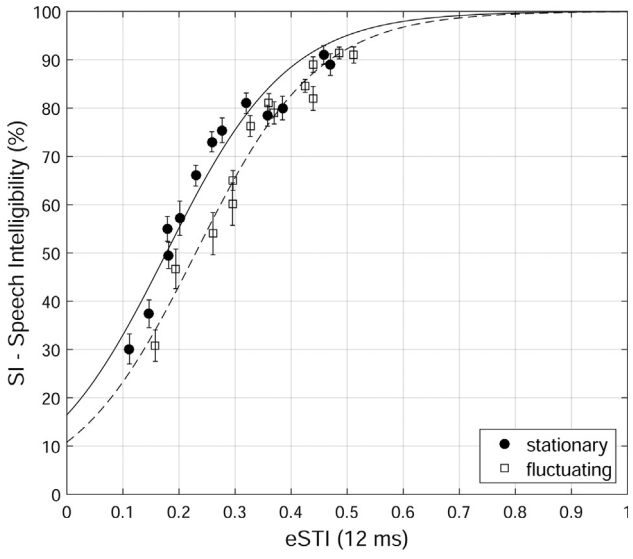

Fig. 6. Speech intelligibility (SI) results averaged across the listeners, with 95% confidence intervals, as a function of the eSTI, calculated using a time window of 12 ms. Results are divided according to the background noise type, and the best-fitting regression curves with a logistic shape are also included: stationary noise (black circles, solid line), fluctuating noise (white squares, dashed line).

Please cite this article as: N. Prodi and C. Visentin, An experimental study of a time-frame implementation of the Speech Transmission Index in fluctuating speech-like noise conditions, Applied Acoustics, https://doi.org/10.1016/j.apacoust.2019.03.026

linear least squares method in Matlab (MathWorks, 2015). The root mean squared error (RMSE) value of the curve from the data points was calculated and taken as an indicator of how well the logistic function fitted the points; the fits were regarded as good because the RMSEs were small. In particular, the RMSE values of the psychometric curves reported in Fig. 6 were 4.7 and 3.5% for the stationary and the fluctuating noise respectively. 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 developed in [42].

The thresholds $eSTI_{50}$ 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. 6, using eSTI with a 12 ms frame window for mapping SI, two almost parallel psychometric curves were obtained, whose distance can be approximated by the difference of the $eSTI_{50}$ values. This difference is equal to 0.05, and is thus larger than the reference JND. This indicates that in order to obtain a 50% SI score, a greater eSTI is needed for the fluctuating masker. Alternatively, if the same eSTI is taken for the two noises in the region of 50% accuracy, a gap in the predicted scores close to 12% would be obtained for a wide range of eSTI values.

To determine if the observed difference between the regression curves was statistically significant, a GLMM analysis was performed as described in Section 2.3. The statistical model indicated that the main effects of eSTI ($\chi^2(1) = 2149.5$, $p < 0.001$) and noise ($\chi^2(1) = 134.4$, $p < 0.001$) were significant, as well as their interaction ($\chi^2(1) = 5.2$, $p = 0.022$). In order to investigate the effect of the noise type on SI at different eSTI values, the continuous variable was fixed at the values of 0.10, 0.25 and 0.50 where the post hoc analyses were performed. The values were chosen as the minimum, the mean and the maximum of the eSTI used for the listening tests. The effect of noise type was found to be significant in all cases ($p < 0.001$ for the three comparisons); the estimated difference between the SI means was 10.7, 10.3 and 2.6% for eSTI values respectively equal to 0.10, 0.25 and 0.50.

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

One of the goals of the study was to explore the time windows for eSTI calculation and possibly determine the length or range of lengths providing coincident psychometric curves for the two maskers. This goal was pursued by calculating eSTI values for different time windows with a 5 ms step and by checking step by step the statistical significance of the effect of noise type on the SI results. 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. For both borderline time windows, the post hoc analyses indicated that the difference between the estimated SI values in the two maskers (averaged over the eSTI interval) was equal to 1.3%.

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: $p = 0.69$). Fig. 7 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 $eSTI_{50} = 0.17$ and $s_{50} = 6.3\%$/JND.

4. Discussion

4.1. Including room acoustics in the short-term framework through eSTI and applications

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 [43].

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

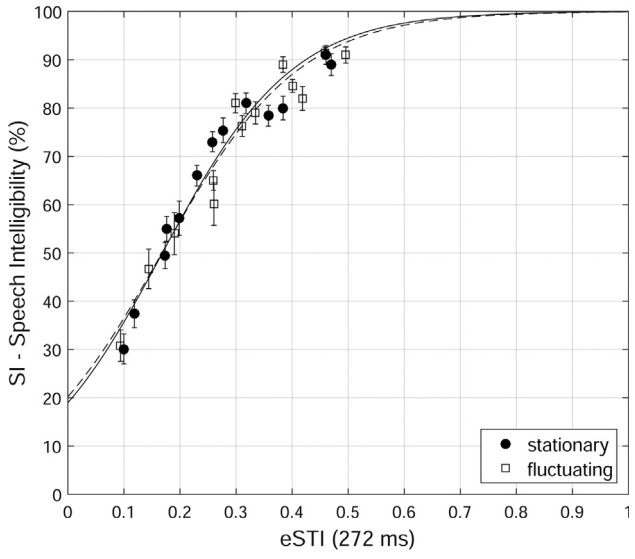

Fig. 7. Speech intelligibility (SI) results averaged across the listeners, with 95% confidence intervals, as a function of the eSTI, calculated using a reference time window of 272 ms. Results are divided according to the background noise type, and the best-fitting regression curves with a logistic shape are also included: stationary noise (black circles, solid line), fluctuating noise (white squares, dashed line).

Please cite this article as: N. Prodi and C. Visentin, An experimental study of a time-frame implementation of the Speech Transmission Index in fluctuating speech-like noise conditions, Applied Acoustics, https://doi.org/10.1016/j.apacoust.2019.03.026

the Schroeder integral [26]. 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.

As expected, it was found that eSTI at threshold depended on the test material. An $eSTI_{50}$ equal to 0.17 was found in the present study, which employs the WST speech material; the value has the same meaning of the "STI at the SRT" equal to 0.33 found in [44] and [4]. It is to be noted that, since for a stationary noise the eSTI coincides with the STI, the values from the two studies can be directly compared. The difference stems from the different speech corpus employed, and in particular on the diverse predictability and on the scoring method (in the studies of George *et al.* a rating based in sentence scoring was used). As all these factors affect the shape and the slope of the psychometric curve [42] and different eSTI/STI at threshold are to be expected. Here, it is to be remarked that the psychometric curve obtained for the WST material is compatible with literature data since it is intermediate between the curves for the phonetically balanced (PB) words and the modified rhyme test (MRT) reported in [45].

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 ([46,44]) 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 [44]. In this perspective, the present work may provide a further key to achieve information on listeners' performance, adding fluctuating noise over reverberation. It is believed that such enhanced picture could help in better setting the performance of people with hearing impairment under reverberation and fluctuating noise with respect to listeners with normal hearing. Anyway, much work needs to be done to confirm the present findings for hearing-impaired clinical population. In particular, to start with, investigations on a suitable time frame interval to equate stationary and fluctuating noise should be carried out, to check if it differs from the present findings for normal-hearing population.

4.2. Study limitations

The single parameter which was optimized in the study was the frame duration, whereas the rest of the STI model was kept unchanged and directly refers to the indirect method for the MTF evaluation [17] with only slight modifications [25]. So, as for the STI, one of the limitations of the present approach is that it is confined to a linear system whose impulse response is analyzed [10]. Moreover, due to its construction, this eSTI approach is configured as an off-line evaluation method which is not suitable for run-time monitoring since noise-free impulse responses will not be easily available [17].

Moreover, the process of selection of the time frame was not driven by any specific hypothesis on the mechanism of speech reception, but solely by the property of mismatch of the SNRs between the noises when one passes from the long- to the short-term analysis. With the shortening of the frame duration, the SNR of fluctuating noise increased, whilst the SNR of stationary noise was constant, and this process directly affected the eSTI values. By this procedure the findings show that the 12 ms frame is not appropriate to achieve a unified description of both noises under reverberant room acoustics conditions. The short time frames typical of the ESII (12 ms) was validated for anechoic target and with speech-like fluctuating noise too ([5,9]). In addition, Ferreira and Payton [13] elaborated an anechoic version of STI substantially equivalent to it. Since the type of noise employed in the present study is quite comparable with the speech-like fluctuating noises of previous studies one can argue that the stretching of time frames found here is primarily due to reverberation and not the specific type of noise. However, the present [200; 350] ms interval might not be appropriate for other types of fluctuating noises. For instance, different ranges might be output by this matching procedure in case of other types of fluctuating noises and they could be either overlapping or not overlapping with the present interval. It has to be recalled that the present noise was selected because it represents a common type of disturbance in public spaces, and has a very high practical relevance in many room acoustics applications involving speech in rooms. Nonetheless further work is needed to extend and integrate the present findings to a larger set of fluctuating interferers, for instance using reverberated sound fields with multiple talkers or intermittent noise. In general, for a given interferer, its inherent masking potential and its resilience in the presence of reverberation is highly dependent on its spectro-temporal characteristics [47]. For instance, high frequency noise bursts with a fast duty-cycle are more harmful to SI, but they are very sensitive to reverberation since rapid fluctuations are easily smeared [8]. The present eSTI approach allowed the combination in a viable manner of the increased masking due to the smearing effect of reverberation on noise with to the self-masking of the signal also caused by reverberation.

In addition, when the time frame interval is translated into a frequency scale, it covers the range 2.89–5 Hz. This fact points to a match of the identified interval with the expected syllables frequency of the present speech corpus, which is composed of four disyllabic words pronounced in approximately two seconds [31] thus covering a prominent frequency modulation of the speech (4–5 Hz) [48]. This indication from the experiments highlights that, in view of qualification, the SNR at the syllable level might be important to provide a good description of the energetic masking of both fluctuating and stationary speech-like noise, and could properly complement the modulation loss of the target signal due to room acoustics. This aspect was not fully clarified within this work and should be further investigated for other masker types as well.

Finally, it is to be noted that one of the reverberation times investigated here (T_{mid} = 0.3 s) falls in the interval of suitable frame durations. Concerns have been raised upon this specific point in [15] since the MTF correction due to the impulse response is by construction independent of the time-frame, while the SNR is

Table 3
Sample size of the listening tests. The listeners were divided in three groups (A–C) and each group was presented with a different subset of the overall 13 listening conditions. The mean and the standard deviation of the participants' ages are also indicated.

Group	# of conditions presented	Sample size (female)	Age [years]
A	4	21 (9)	M = 27.3 (SD = 6.1)
B	4	29 (10)	M = 26.7 (SD = 7.3)
C	5	29 (10)	M = 26.7 (SD = 5.8)

Please cite this article as: N. Prodi and C. Visentin, An experimental study of a time-frame implementation of the Speech Transmission Index in fluctuating speech-like noise conditions, Applied Acoustics, https://doi.org/10.1016/j.apacoust.2019.03.026

not. The present experimental data indicate that the method is robust with respect to this limit, since for both short (12 ms) and longer frames the effect of room acoustics is correctly accounted for. The evidence presented here is that reliable logistic curves have been obtained by listening tests in all cases and thus the indicator is consistent in tracing the performance irrespective of the frame duration.

5. Conclusions

The present work investigated an adaptation of the STI indirect method to deal with a speech-like fluctuating masker. The previous literature indications that employed a frequency-dependent set of frame values in anechoic speech-like fluctuating masker were revised, and an equivalent frequency-independent frame duration was employed. While keeping speech-like stationary and fluctuating maskers this work introduced reverberation on both signal and interferers. This addition stretched the time frame duration that is appropriate to equate the values of the objective indicator eSTI under stationary and fluctuating noises so that they are matched to the same accuracy performance (Table 3).

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.

References

[1] Festen JM, Plomp R. Effects of fluctuating noise and interfering speech on the speech-reception threshold for impaired and normal hearing. J Acoust Soc Am 1990;88(4):1725–36.
[2] Gustafsson HA, Arlinger SD. Masking of speech by amplitude modulated speech. J Acoust Soc Am 1994;95:518–29.
[3] Versfeld NJ, Dreschler WA. The relationship between the intelligibility of time-compressed speech and speech in noise in young and elderly listeners. J Acoust Soc Am 2002;111:401–8.
[4] George ELJ, Festen JM, Houtgast T. The combined effects of reverberation and nonstationary noise on sentence intelligibility. J Acoust Soc Am 2008;124 (2):1269–77.
[5] Rhebergen KS, Versfeld NJ. A Speech Intelligibility Index-based approach to predict reception threshold for sentences in fluctuating noise for normal-hearing listeners. J Acoust Soc Am 2005;117(4):2181–92.
[6] American National Standard Institute (1997). ANSI–S3.5. Methods for Calculation of the Speech Intelligibility Index (American National Standard Institute, New York).
[7] Moore B. (1997). "Temporal Processing in the Auditory System," in: An Introduction to the Psychology of Hearing (4th ed.) (Academic Press, San Diego, California), Chap. 4, pp. 148–176.
[8] Rhebergen KS, Versfeld NJ, Dreschler WA. Extended speech intelligibility index for the prediction of the speech reception threshold in fluctuating noise. J Acoust Soc Am 2006;120(6):3988–97.
[9] Rhebergen KS, Versfeld NJ, Dreschler WA. Prediction of the intelligibility for speech in real-life background noises for subjects with normal hearing. Ear Hear 2008;29(1):169–75.
[10] International Electrotechnical Commission. IEC60268–16 (edition 4.0). Sound system equipment, Part 16: Objective rating of speech intelligibility by speech transmission index. Geneva, Switzerland: International Electrotechnical Commission; 2011.
[11] Houtgast T, Steeneken HJM, Plomp R. Predicting speech intelligibility in rooms from the modulation transfer function. I. General room acoustics. Acustica 1980;46:59–72.
[12] Harris RW, Swenson DW. Effects of reverberation and noise on speech recognition by adults with various amounts of sensorineural hearing impairment. Audiology 1990;29(6):314–21.
[13] Ferreira M, Payton K. Using the short-time speech transmission index to predict speech reception thresholds in fluctuating noise. J Acoust Soc Am 2014;135(4):2224–5.
[14] Ludvigsen C, Elberling C, Keidser G. Evaluation of a noise reduction method: Comparison between observed scores and scores predicted from STI. Scand Audiol 1993;22(Suppl. 38):50–5.
[15] Payton KL, Shrestha M. Comparison of a short-time speech-based intelligibility metric to the speech transmission index and intelligibility data. J Acoust Soc Am 2013;134(5):3818–27.
[16] Goldsworthy RL, Greenberg JE. Analysis of speech-based transmission index methods with implications for non-linear operations. J Acoust Soc Am 2004;116(6):3679–89.
[17] van Schoonhoven J, Rhebergen KS, Dreschler WA. Towards measuring the Speech Transmission Index in fluctuating noise: accuracy and limitations. J Acoust Soc Am 2017;141(2):818–27.
[18] Lavandier M, Culling JF. Speech segregation in rooms: effects of reverberation on both target and interferer. J Acoust Soc Am 2007;122(3):1713–23.
[19] Dreschler WA, Verschure H, Ludvigsen C, Westermann S. ICRA noises: artificial noise signals with speech-like spectral and temporal properties for hearing instrument assessment. Audiology 2001;40(3):148–57.
[20] Lombard E. "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) 1911;37:101–19.
[21] Lu Y, Cooke M. Speech production modifications produced by competing talkers, babble, and stationary noise. J Acoust Soc Am 2008;124(5):3261–75.
[22] Lu Y, Cooke M. The contribution of changes in F0 and spectral tilt to increased intelligibility of speech produced in noise. Speech Comm. 2009;51:1253–62.
[23] Valentini-Botinhao C. Intelligibility enhancement of speech in noise. Proceedings of Conference "Reproduced Sound 2014. St. Albans (UK): Institute of Acoustics; 2014.
[24] International Organization for Standardization. ISO 9921, Ergonomics – Assessment of speech communication. Geneva: Switzerland; 2003.
[25] van Wijngaarden SJ, Houtgast T. Effect of talker and speaking style on the Speech Transmission Index (L). J Acoust Soc Am 2004;115(1):38–41.
[26] Bradley JS, Reich RD, Norcross SG. On the combined effects of signal-to-noise ratio and room acoustics on speech intelligibility. J Acoust Soc Am 1999;106 (4):1820–8.
[27] van Wijngarden S, Drullman R. Binaural intelligibility prediction based on the speech transmission index. J Acoust Soc Am 2008;123(6):4514–23.
[28] Unitron. uHear [Mobile application software] Retrieved from http://itunes.apple.com, ; 2009.
[29] Szudek J, Ostevik A, Dziegielewski P, Robinson-Anagor J, Gomaa N, Hodgetts B, et al. Can Uhear me now? Validation of an iPod-based hearing loss screening test. J. Otolaryngology-Head Neck Surgery 2012;41(1):S78–84.
[30] Wang JC, Zupancic S, Ray C, Cordero J, Demke JC. Hearing test app useful for initial screening, original research shows. The Hearing J 2014;67(10):32–4.
[31] Visentin C, Prodi N. A matrixed speech-in-noise test to discriminate favorable listening conditions by means of intelligibility and response time results. J Speech Lang Hear. 2018;61(6):1497–516.
[32] Bonaventura P, Paoloni F, Canavesio F, 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.
[33] Kalikow DN, Stevens KN, Elliott LL. Development of a test of speech intelligibility in noise using sentence materials with controlled word predictability. J Acoust Soc Am 1977;61(5):1337–51.
[34] Puglisi GE, Warzybok A, Hochmuth S, Visentin C, Astolfi A, Prodi N, et al. An Italian matrix sentence test for the evaluation of speech intelligibility in noise. Int. J. of Audiology 2015;54(sup2):44–50.
[35] Prodi N, Visentin C, Farnetani A. Intelligibility, listening difficulty and listening efficiency in auralized classrooms. J Acoust Soc Am 2010;128(1):172–81.
[36] Bencina R. (2010). AudioMulch, Version 2.0.4 [Computer software]. Retrieved from http://www.audiomulch.com/.
[37] Farina A. X-volver VST plug-in [Computer software] Retrieved from http://pcfarina.eng.unipr.it/X-volver.htm, ; 2017.
[38] Everitt BS, Hothorn T. A handbook of statistical analysis using R. Chap: Second edition (Chapman and Hall/CRC, New York); 2010. p. 12.
[39] R Core Team. R: A language and environment for statistical computing. Vienna, Austria: R Foundation for Statistical Computing; 2017. , https://www.R-project.org/.
[40] Bates D, Maechler M, Bolker B, Walker S. Fitting linear mixed effects models using lme4. J Stat Softw 2015;67:1–48.
[41] Lenth RV. Least-squares means: the R package. J Stat Softw 2016;69(1):1–33.
[42] MacPherson A, Akeroyd MA. Variations in the slope of the psychometric functions for speech intelligibility: a systematic survey. Trends Hear. 2014;18:1–26.

Please cite this article as: N. Prodi and C. Visentin, An experimental study of a time-frame implementation of the Speech Transmission Index in fluctuating speech-like noise conditions, Applied Acoustics, https://doi.org/10.1016/j.apacoust.2019.03.026

[43] Deutsche Institut für Normung. DIN18041. Hörsamkeit in Räumen – Anforderungen, Empfehlungen und Hinweise für die Planung (Acoustic quality in rooms – Specifications and instructions for the room acoustic design. Berlin, Germany): (Deutsche Institut für Normung; 2016.
[44] George EL, Goverts ST, Festen JM, Houtgast T. Measuring the effects of reverberation and noise on sentence intelligibility for hearing-impaired listeners. J Speech Lang Hear Res 2010;53(6):1429–39.
[45] Anderson BW, Kalb JT. English verification of the STI method for estimating speech intelligibility of a communications channel. J Acoust Soc Am 1987;81 (6):1982–5.
[46] Duquesnoy AJ, Plomp R. Effect of reverberation and noise on the intelligibility of sentences in cases of presbyacusis. J Acoust Soc Am 1980;68(2):537–44.
[47] Schubotz W, Brand T, Kollmeier B, Ewert SD. Monaural speech intelligibility and detection in maskers with varying amounts of spectro-temporal speech features. J Acoust Soc Am 2016;140(1):524–40.
[48] Dubbelboer F, Houtgast T. The concept of signal-to-noise ratio in the modulation domain and speech intelligibility. J Acoust Soc Am 2008;124 (6):3937–46.

Please cite this article as: N. Prodi and C. Visentin, An experimental study of a time-frame implementation of the Speech Transmission Index in fluctuating speech-like noise conditions, Applied Acoustics, https://doi.org/10.1016/j.apacoust.2019.03.026