**Speech Based Transmission Index for All:
an intelligibility metric for variable hearing ability**

Nader Mechergui,^{a)} Sonia Djaziri-Larbi, and Mériem Jaïdane

*Université Tunis El Manar, Ecole Nationale d'Ingénieurs de Tunis,
Signals and Systems Lab, Tunisia.*

(Dated: 27 January 2017)

A method to measure the speech intelligibility in public address systems for normal hearing and hearing impaired persons is presented. The proposed metric is an extension of the speech based Speech Transmission Index to account for accurate perceptual masking and variable hearing ability: The sound excitation pattern generated at the ear is accurately computed using an auditory filters model, which shapes depend on frequency, sound level and hearing impairment. This extension yields to a better prediction of the intensity of auditory masking which we use to rectify the Modulation Transfer Function and thus to objectively assess the speech intelligibility experienced by hearing impaired as well as by normal hearing persons in public spaces. The proposed metric was developed within the framework of the European Active and Assisted Living research program, and was labeled SB-STI *for All*. Extensive subjective In-Lab and In-Vivo tests have been conducted and the proposed metric proved to have a good correlation with subjective intelligibility scores.

PACS numbers: 43.71.Gv

Keywords: intelligibility; speech; auditory model; presbycusis; hearing impaired.

^{a)}Electronic mail: nader.mechergui@gmail.com; Also at CEA Linklab-Telnet Innovation Labs, Tunisia

I. INTRODUCTION

According to Eurostat¹, 25% of the European population will have an age higher than 65 in 2050. This forecast highlights the need of the adjustment of public services for aged people and especially public transportation services. Aged people go through different sensory losses that affect their transportation services experience. For example, age related hearing loss (presbycusis) affects the comprehension of vocal announcements in railway stations. This sensory loss gradually affects everyone from the age of 18 years as measured in ISO7029⁷ (statistical distribution of hearing thresholds as a function of age) and can reduce speech intelligibility for aged persons. However, public address systems are designed for normal hearing persons and do not take into account the diversity of hearing ability of the population. Through the I'CityForAll² project under the Active and Assisted Living (AAL) program, we aim to enhance and assess intelligibility for both normal hearing and hearing impaired persons in public and urban spaces. The present paper proposes a method to assess speech intelligibility for a variable hearing ability in confined public spaces.

Nowadays, the most used standard to assess objectively the intelligibility of public address (PA) systems is the Speech Transmission Index (STI)⁸. The STI assesses noise and reverberation distortions to measure the effect of room acoustics on intelligibility. However, it does not take into consideration the difference in hearing ability of the listeners. The STI is based on the speech envelope distortion which is measured by the Modulation Transfer Function (MTF)^{9,10}. The MTF quantifies the loss of depth in the amplitude modulation of the speech envelope that affects speech intelligibility. To measure the MTF, Steeneken *et al.*¹¹ propose a modulated speech shaped noise (SSN) as a probe signal which mimics the speech envelope in order to take into account non-linear and time domain related distortions. On the other hand, Schroeder¹² deduces the MTF from the Room Impulse Response (RIR). Both methods are not adequate for real time intelligibility monitoring in actual conditions; for example, to measure the intelligibility of vocal announcements in a railway station at rush hours, STI would require the broadcasting of SSN in the railway station and this would annoy the passengers. RIR based STI requires also the measure of actual impulse response of the considered space. In both cases, carrying out the measures when the spaces are empty can be limiting knowing that noise level and reverberation, and thus intelligibility, depend on the occupancy of the station^{13–15}. Real speech (vocal announcement) can be used as

an alternative to synthesized probe to assess the intelligibility in public spaces at any time while taking advantage of the actual test conditions (ambient noise and reverberation at rush hours for example).

Several methods were developed to predict MTF from real speech to measure a speech based STI¹⁶. The envelope regression method proposed by Ludvigsen *et al.*¹⁷ has proven to be effective in the prediction of MTF using speech^{16,18}. This method was used by Payton *et al.*¹⁹ in order to predict speech intelligibility for different analysis windows lengths and they obtained a good correlation between theoretical STI and SB-STI using Ludvigsen’s method for windows length of 0.3 second. This work motivated the use of the envelope regression method to compute the MTF of speech signals. In this study, we extend the SB-STI using the envelope regression method to assess the intelligibility of PA systems for normal hearing and hearing impaired persons. Previous works have shown how the STI scale can be adjusted to take into consideration listeners with presbycusis²⁰. This scale modification was introduced in the ISO standard based on empirical assessment of intelligibility experienced by hearing impaired persons⁸. We modify in this study the auditory masking effect introduced in the ISO60268 standard revision of 2003 in order to measure the STI for both normal hearing (NH) and hearing impaired (HI) persons while keeping the same STI scores scale. We rectify the MTF values according to the audiogram of persons using Glasberg and Moore’s model of loudness perception applied to cochlear hearing loss²¹. This allows to assess the intelligibility experience of HI but also to accurately compute the masking effect for both, HI and NH. Finally, we evaluate the proposed metric SB-STI *for All* via subjective intelligibility tests carried out in in-lab and in in-vivo conditions.

The paper is organized as follows. In section II, we introduce the STI metric standardized by the ISO and we emphasize the importance of the MTF calculation. In section III, we introduce Glasberg and Moore’s auditory model to rectify the MTF in order to account for hearing impairment. Section IV is dedicated to subjective in-lab and in-vivo intelligibility test procedures for NH and HI persons, designed to assess and validate the SB-STI *for All*. Finally, we compare in section V between subjective measures in in-lab and in in-vivo contexts with the proposed objective intelligibility metric SB-STI *for All*.

FIG. 1. Block diagram of STI: filtering clean and distorted signals in 7 octave bands, envelope extraction (wave rectifier method), MTF computation in each octave band, MTFs rectification using auditory model and STI score computation using octave band importance and information redundancy weights.

II. MAIN STEPS OF THE STI METRIC

In this section, we remind the main steps of the ISO60268 STI metric, as they are identical to those of the SB-STI¹⁸ and of the proposed SB-STI *for All*. The main difference between the 3 STI algorithms lies in the calculation of the MTF from the probe signal (the STI metric is an intrusive measure, i.e. it needs both clean and distorted signals). The standard STI uses modulated SSN to compute the MTF, while both SB-STI algorithms use real speech as probe signal. In our work, the choice of real speech as probe signal has two motivations. The first one is to have a real-time monitoring metric for PA systems which captures intelligibility experience without causing any annoyance to the people present in the public space. The second motivation is the possibility to use a precise prediction of the auditory masking effects which undoubtedly affect the intelligibility experience. This latter aspect is detailed in section III.

A. Computing the MTF

As depicted in Fig. 1, both clean and distorted signals are filtered in seven octave sub-bands. The envelope of each sub-band signal is extracted and used to compute the MTF related to that sub-band. The MTF reflects the modification caused by the acoustic channel

FIG. 2. Effect of noise and reverberation on the MTF of octave band 1 kHz (Schroeder's method) with variable RT60 and SNR.

on the envelope intensity as a function of frequency. The longer the reverberation and higher the noise, the smaller the MTF is (in $[0, 1]$). We show on Fig. 2 how different acoustic distortions affect the MTF. In case of a stationary noise, the MTF value is independent of modulation frequencies, but the lower the SNR, the lower the MTF value is. In presence of reverberation, we note (Fig. 2) that the MTF values at higher modulation frequencies are more affected than those at lower frequencies.

Two different MTF calculation methods are proposed in the ISO60268 standard⁸: the direct method which uses modulated SSN as probe signal and the indirect method which uses the RIR and is also named Schroeder's method¹². The direct method consists in broadcasting modulated SSN in a room with different modulation frequencies f_m ranging from 0.63 to 12.5 Hz. The received signal is correlated with sine and cosine signals in order to extract the MTF for each sub-band k (with center frequencies from 125 Hz to 8 kHz) and at each modulation frequency f_m as follows:

$$\text{MTF}_k(f_m) = 2 \frac{\sqrt{\left(\sum_{n=1}^N I_k(n) \cdot y_1(n)\right)^2 + \left(\sum_{n=1}^N I_k(n) \cdot y_2(n)\right)^2}}{\sum_{n=1}^N I_k(n)}, \quad (1)$$

where $y_1(n) = \sin(2\pi f_m n)$, $y_2(n) = \cos(2\pi f_m n)$, and $I_k(n)$ is the received envelope of octave band k as a function of time index n . The indirect method uses the room impulse response

and is defined as follows:

$$\text{MTF}_k(f_m) = \frac{\| \int_0^\infty h_k^2(t) e^{-i2\pi f_m t} dt \|}{\int_0^\infty h_k^2(t) dt} \cdot \frac{1}{1 + 10^{-\text{SNR}/10}}, \quad (2)$$

where h_k is the filtered RIR by octave band filter k and SNR is the signal to noise ratio. Neither direct nor indirect method take into account the nature of the emitted speech and that of the background noise; the noise is considered as a stationary signal in both methods in order to simplify the prediction of the MTF. However, we know that noise affects differently the intelligibility depending on its nature: modulated, stationary or speech competing noise²². Besides, these two methods can not be used (as standardized) to monitor the intelligibility in real time in a space full of people as they would necessitate the interruption of the normal function of the audio system. This is mainly why we opted for another method to compute the MTF which uses real speech as probe.

During the development of STI, Houtgast *et al.* have used speech as input to compute the MTF. They observed artifacts when measuring the envelope spectra of degraded speech¹⁰. These artifacts consist in increases of the envelope intensity spectra when theory predicts decreases. This over estimation of the intensity introduces bias in the calculation of the MTF.

To avoid artifacts, other methods can be used to calculate the MTF using normalized covariance, cross-power spectrum or envelope regression methods. A good review of those methods is given by Goldsworthy *et al.*¹⁶. In this study, we are not comparing between these methods but we are interested in rectifying the speech based MTF for different hearing abilities. We computed the MTF by the Envelope Regression method (ER) proposed by Ludvigsen *et al.*¹⁷ which was used by Payton *et al.*¹⁸ in real time conditions with noise and reverberation distortion. This method is useful for real time intelligibility monitoring in public spaces, knowing that it was proven to have a good correlation with standard STI scores, when computed on short speech frames with 0.3 second duration¹⁸ (due to real-time constraint). In the experiment reported in¹⁸, real speech was broadcasted to assess the intelligibility. The ER method consists in comparing the probe envelope signal and the distorted envelope signal in each band k using linear regression analysis (cf. Appendix) and the MTF

120 is then expressed as follows:

$$\text{MTF}_k = \frac{\mu_x \sum_{n=1}^N (X_k[n] - \mu_x)(Y_k[n] - \mu_y)}{\mu_y \sum_{n=1}^N (X_k[n] - \mu_x)^2}, \quad (3)$$

121 where $X_k[n]$ and $Y_k[n]$ are respectively the clean and distorted temporal envelopes in band
 122 k , $\mu_x = \frac{1}{N} \sum_{n=1}^N (.)$ is the sample mean, N is the number of samples and n is the time index.
 123 The ER method is used in the proposed metric SB-STI *for All*.

124 B. Rectifying the MTF using an auditory model

125 To account for the auditory perception, The MTF is rectified to take into consideration
 126 two psychoacoustic phenomena: the auditory masking and the absolute reception threshold.
 127 The intensity of these two effects is calculated by an auditory model and used to rectify the
 128 MTF in each octave band.

129 We remind here that the auditory masking intensity is the minimum required intensity of
 130 a signal at a given frequency to be heard by a listener in the presence of another masking
 131 signal, and the absolute reception threshold is the minimum required intensity at a given
 132 frequency band for the speech to be understood in an anechoic environment⁸.

133 The auditory model used in ISO60268 to compute the auditory masking intensity I_{am_k}
 134 was designed mainly to take into account high level speech²³ knowing that the intelligibility
 135 decreases for speech levels higher than 85 dB. The auditory masking intensity in band $k + 1$
 136 is deduced from the speech level of band k using piecewise linear regression equations for
 137 different speech level intervals. This model was deduced from empirical measures of upward
 138 masking slopes for a low frequency noise in the band 50-800 Hz with variable level within
 139 46-96 dB²³. We know, however, that the auditory filter shape is broadened not only as a
 140 function of the sound pressure level but also as a function of frequency and hearing ability.
 141 As a consequence, the intensity of masking in band $k + 1$ should be higher for hearing
 142 impaired persons and should be taken into account in high frequency octave bands as well
 143 as in the low frequency bands.

144 Furthermore, in ISO60268, only the masking effect of band k on the upper band $k + 1$
 145 is considered. The masking effect on higher and lower bands is ignored as depicted in

FIG. 3. Auditory masking intensity of octave band k on band $k + 1$ following the empirically deduced masking slope.

Fig. 3. Besides the intensity of auditory masking, the STI standard introduces also the absolute reception threshold (ART), not to be confused with the absolute hearing threshold. The ART was measured empirically and used in the standard ISO60268⁸, where these two phenomena are assumed to have the same effect as an additive stationary noise that affects the intelligibility and were integrated into the MTF as follows:

$$\text{MTFr}_k = \text{MTF}_k \cdot \frac{I_k}{I_k + I_{am_k} + I_{art_k}}, \quad (4)$$

where MTFr_k is the rectified MTF_k in band k , I_k is the intensity of the distorted speech in band k , I_{am_k} is the intensity of auditory masking and I_{art_k} is the intensity of absolute reception threshold.

In section III, we will explain how we modify the auditory model to account for hearing impairment and to give a better prediction of auditory perception effects while preserving the same assumption (stationary noise as in Eq. (4)).

C. Computing the STI score

To compute the STI score, we applied the same procedure as the ISO standard but using the speech based MTF computed with the envelope regression method (cf. Eq. (3)). We note here that the ER method based MTF is a function of octave band index k and it does not depend on modulation frequencies, while the MTF computed with the direct method depends on octave bands and modulation frequencies.

164 Firstly, the effective SNR is computed as follows:

$$\text{eSNR}_k = 10 \log_{10} \left[\frac{\text{MTFr}_k}{1 - \text{MTFr}_k} \right], \quad (5)$$

165 where eSNR_k takes into consideration background noise and reverberation as "noise" and
 166 MTFr_k is the rectified MTF of octave band k . Then the modulation transfer index for each
 167 octave band k is obtained from eSNR_k :

$$\text{MTI}_k = \frac{\text{eSNR}_k + 15}{30}. \quad (6)$$

168 The former equation is used to normalize the values of eSNR_k between 0 and 1 while keeping
 169 some dynamics.

170 Finally, knowing that octave bands participate differently to the intelligibility, each MTI_k
 171 value is weighted according to band importance and redundancy weights to obtain the SB-
 172 STI score as follows:

$$\text{SB-STI} = \sum_{k=1}^7 \alpha_k \text{MTI}_k - \sum_{k=1}^6 \beta_k \sqrt{\text{MTI}_k \cdot \text{MTI}_{k+1}}, \quad (7)$$

173 where α_k is the weight factor of octave band k , β_k is the redundancy factor between octave
 174 band k and octave band $k + 1$ and SB-STI is the speech based STI score obtained with ER
 175 method. Note that these weights are different if male or female voice is used.

176 The standard STI score is known to be highly correlated with intelligibility for NH per-
 177 sons: for scores lower than 0.3 the intelligibility is considered as bad and for scores beyond
 178 0.6 it is considered as good⁸. **However, for other languages²⁴⁻³⁰ and for non-native**
 179 **listeners³¹, there exist different matching scales fitted for each case. In partic-**
 180 **ular, for subjects with presbycusis²⁰, the scale proposed in ISO60268⁸ lacks on**
 181 **precision and varies according to hearing loss category.** We propose in this study
 182 to modify the auditory model used to rectify the MTF in order to take into consideration
 183 NH and HI listeners (sensorineural hearing loss) while keeping the same matching scale.

III. AUDITORY MODEL FOR VARIABLE HEARING ABILITY

In this section we explain how we modify the auditory model to design the proposed SB-STI *for All*, an intelligibility metric that uses a speech based MTF, which we extend to account for variable hearing ability. Indeed, the MTF is the core of different intelligibility metrics, it describes the acoustic transmission channel from the source to the receiver. In this study, we chose to use a speech based MTF rather than the direct and indirect methods of the ISO60268 standard, because our aim is to develop a pseudo real-time intelligibility metric dedicated for PA systems, which captures the variations of ambient noise, reverberation and vocal announcements, in addition to the variable hearing ability of listeners. Particularly, the main core of this study is the extension of the speech based MTF to account for the internal sound representation at the ear level, **contrarily to other studies^{26,32,33} which propose objective intelligibility metrics, different from STI, that are also based on auditory models. The goal of our approach is to extend the standard STI** in order to i) take into consideration variable hearing ability (including age related hearing loss), and ii) to compute a more accurate auditory masking. To do this, we introduce a different auditory model to predict I_{am_k} and I_{art_k} in order to cope with the limits of the model used in the ISO60268.

As depicted in Fig. 4, the inputs of the auditory model are the distorted signal (the received speech degraded by noise and reverberation) and the audiogram of the listener which represents the total hearing loss as a function of frequency compared to the audiogram of an 18 years aged person. According to the World Health Organization (WHO), a person is considered as hearing impaired if his/her Pure Tone Average (PTA) exceeds 25 dB³. In the following, we predict I_{am_k} using the auditory filters modeled as roex filters. On the other hand, I_{art_k} is modified according to the provided audiogram of the listener.

A. Auditory filters for a variable hearing ability

To measure an auditory masking intensity that takes into consideration the variation of sound pressure level, frequency and hearing ability, we base our approach on Moore and Glasberg's model of loudness perception. The updated model described in²¹ shows how cochlear hearing loss affects the auditory filter shape for a given center frequency and sound

FIG. 4. Auditory model diagram: the power in each ERB band is calculated from the amplitude spectrum of the distorted speech, then the auditory filters are computed which also vary with hearing loss of outer hair cells (HL_{ohc}). The excitation pattern in each octave band is computed using the generated auditory filters and speech spectrum in each band. From that the intensity of auditory masking (I_{am}) is deduced. The intensity of absolute reception threshold (I_{art}) is deduced from the audiogram.

pressure level. This model predicts the auditory excitation pattern that is generated by the audio signal in the basilar membrane of the cochlea. The shape of the roex auditory filter is given by:

$$W_{erb}(f) = \begin{cases} (1 + Pl_{erb} \cdot g) e^{-Pl_{erb} \cdot g} & \text{if } g \leq 0 \\ (1 + Pu_{erb} \cdot g) e^{-Pu_{erb} \cdot g} & \text{if } g > 0, \end{cases} \quad (8)$$

$$\text{where } g = |f - f_c|/f_c, \quad (9)$$

and W_{erb} is the auditory filter shape for each Equivalent Rectangular Bandwidth (ERB), erb is the ERB band index, g is the absolute value of the frequency deviation, f_c is the center frequency of the auditory filter that varies for each erb , and Pl_{erb} and Pu_{erb} are respectively the slopes parameters of lower and upper filter sides. The loudness model of Moore and Glasberg²¹ describes how Pl_{erb} and Pu_{erb} depend on sound pressure level and hearing loss. The slopes parameters for NH were deduced from psychoacoustic measures carried out with a signal at 51 dB/ erb effective input level (after outer and middle ear filtering). However, the slope Pl_{erb} of the lower filter side is decreased (filter is broadened) with increasing sound

level. To account for this variability regarding speech level, Pl_{erb} for NH is computed as:

$$Pl_{erb}(L) = Pl_{erb}(51) - \gamma(L - 51) \frac{Pl_{erb}(51)}{Pl_{1k}(51)}, \quad (10)$$

where $Pl_{erb}(L)$ is the slope parameter of the lower filter side as a function of the effective level L in dB and γ is the equation's slope parameter equal to 0.35. L is deduced for each auditory filter from the amplitude spectrum of the outer and middle ear filtered signal.

Moore and Glasberg propose another filter broadening factor to account for hearing loss: if the impairment is mainly due to a cochlea malfunction, the total hearing loss measured by the audiogram has two main origins; the outer and inner hair cells. In that case, the total hearing loss is expressed as:

$$HL = HL_{ohc} + HL_{ihc}, \quad (11)$$

where HL is the audiogram (as a function of frequency) measured by audiometry, HL_{ohc} and HL_{ihc} are respectively the hearing loss induced by outer and inner hair cells. HL_{ohc} represents between 80% and 90% of the total hearing loss with a maximum of "57.6+G(dB)", where G is the low level gain of the cochlear amplifier for NH persons²¹. Furthermore, the broadening of the auditory filters and the variation of Pl_{erb} and Pu_{erb} are caused exclusively by HL_{ohc} . Hence, the proposed broadening factor due to hearing impairment is expressed as a function of HL_{ohc} by:

$$B(f_c) = \begin{cases} 10^{0.01HL_{ohc}(f_c)}, & f_c \geq 500\text{Hz} \\ 10^{0.01HL_{ohc}(f_c)[1-\frac{(f_c-0.5)^2}{c}]}, & f_c < 500\text{Hz} \end{cases} \quad (12)$$

where $HL_{ohc}(f_c)$ is the hearing loss due to outer hair cells as a function of center frequency and $c = 1.23$. Note that in Eq. (8) and Eq. (10) frequency is expressed in the ERB scale, while in Eq. (12) f_c is expressed in Hz. ERB index erb and center frequency f_c are two related frequency scales that refer to the same filter index and we can switch between the two scales following the expression:

$$erb = 21.366 \log_{10} \left[\frac{4.368}{1000} f_c + 1 \right]. \quad (13)$$

FIG. 5. Auditory filters centered at $f_c = 1$ kHz and 4 kHz for different HL_{ohc} values and sound pressure levels (L_{rms}). The filters centered at 1 kHz: $L_{rms}=75$ dB and $HL_{ohc}=0, 40$ and 90 dB. The filters centered at 4 kHz: $L_{rms}=30, 75$ and 80 dB and $HL_{ohc}=0$ dB.

Finally, the filter slopes Pl_{erb} and Pu_{erb} are divided by $B(f_c)$ in order to take into consideration the broadening of the auditory filter caused by hearing loss. Besides, it appears that the effect of sound pressure level on the auditory filter broadening, described in Eq. (10) is less significant for HI than for NH. To account for this, Moore and Glasberg made the slope γ dependent on $B(f_c)$:

$$\gamma = 0.35 - 0.35 \frac{B(f_c) - 1}{3}. \quad (14)$$

Fig. 5 summarizes the variation of the auditory filter shape as a function of hearing loss, sound pressure level and frequency: the first group of auditory filters centered at 1 kHz were generated with a fixed sound pressure level of 75 dB and a varying HL_{ohc} from 0 to 80 dB. We see clearly that the filter is broadened when the hearing loss increases. The second group of auditory filters centered at 4 kHz corresponds to normal hearing auditory filters with a varying sound intensity from 30 to 90 dB. We note here that the increase of intensity affects only the lower frequency side of the filter as expressed in (10). When comparing the two filter groups, we can also note that at the same sound pressure level (75 dB) and hearing ability ($HL_{ohc}=0$ dB) the auditory filter is wider at higher center frequencies.

FIG. 6. Excitation pattern in octave band centered at 2 kHz for a NH (dashed) and a HI person (dotted) with a flat hearing loss of 60 dB ($HL_{ohc} = 0.9HL$). The speech spectrum was computed using a Hamming window equal to the length of the sentence (2 seconds) and 44.1 kHz sampling frequency.

B. Intensity of auditory masking from excitation pattern

The auditory excitation pattern is a perceptual representation of the audio signal in the frequency domain. It is generated by filtering the received audio signals with the auditory filters. Our goal is not to compute an excitation pattern for loudness prediction but to quantify the auditory masking caused by the signal in one octave band on its neighboring octave bands. This step yields 7 excitation patterns relative to the octave bands with center frequencies from 125 Hz to 8 kHz. The excitation pattern in octave band k as a function of erb is given by:

$$EP_k(erb) = \frac{\sum_{f=1}^{N_{f_k}} W_{erb}(f) \cdot FY_k(f)}{ERB_{erb}}, \quad (15)$$

where

$$FY_k(f) = \begin{cases} FY(f) & \text{if } f \in [f_{klow}, f_{khigh}], \\ 0 & \text{otherwise.} \end{cases} \quad (16)$$

$W_{erb}(f)$ is the magnitude of auditory filter indexed by erb , $FY_k(f)$ is the amplitude spectrum of the distorted speech in band k (cf. Fig. 4), ERB_{erb} is the equivalent rectangular bandwidth of the auditory filter W_{erb} , f_{klow} and f_{khigh} are respectively the lower and upper cutoff frequencies of octave band k , f is the frequency and N_{f_k} is the number of frequency bins in

$[f_{klow}, f_{khigh}]$. Note that, in order to compute the excitation pattern in each octave band, we computed directly the amplitude spectrum $FY(f)$ of the distorted speech and $FY_k(f)$ was obtained by putting to zero all the values of $FY(f)$ lying outside the octave band limits, as explained by Eq (16). As shown in Fig. 6, the speech spectrum in band k , $FY_k(f)$, generates an excitation pattern in the neighboring bands. We note also that the upward masking is higher than the downward masking.

Particularly, Fig. 6 shows that the spread of masking of the excitation pattern for HI persons is more important than that of NH persons and can affect octave bands beyond $k + 1$. So rather than considering only the effect of masking of band k on band $k + 1$ we take into account the intensity of masking generated by the excitation pattern on all the neighboring bands. The I_{am_k} in each octave band k is obtained by summing the auditory masks generated by the other 6 excitation patterns in that octave band k . The auditory masking intensity that depends on hearing loss is then expressed for band k as follows:

$$I_{am_k}^{HL} = \frac{1}{N_{fc_k}} \sum_{\substack{i=1..7 \\ i \neq k}} \sum_{f_c=f_{klow}}^{f_{khigh}} EP_i(f_c), \quad (17)$$

where i and k are octave band indices, f_c the center frequencies of auditory filters (equivalent to *erb* according to Eq. (13)), f_{khigh} and f_{klow} are respectively the upper and lower limits of octave band k and N_{fc_k} is the number of center frequencies within band k .

C. Absolute reception threshold for variable hearing ability

The absolute reception threshold intensity I_{art_k} was also modified to account for hearing impairment by adding the intensity of hearing loss as follows:

$$I_{art_k}^{HL} = I_{art_k}^{NH} + HL, \quad (18)$$

where $I_{art_k}^{NH}$ is the intensity of absolute reception threshold for NH, HL the intensity of hearing loss measured by audiometry (audiogram). Knowing that the absolute reception threshold and the auditory masking are the effects of the same biological process at the cochlea³⁴, we consider that their intensities should not be summed and we modify the $MTFr_k$ of Eq. (4)

FIG. 7. Mapping between standard STI (Schroeder’s method) and SB-STI *for All* for NH persons. 56 test conditions were used with RT60 in $[0.5, 5]$ seconds and SNR in $[-12, 10]$ dB.

according to this as follows:

$$\text{MTF}r_k = \text{MTF}_k \cdot \frac{I_k}{I_k + \max(I_{am_k}^{HL}, I_{art_k}^{HL})}. \quad (19)$$

To obtain the SB-STI *for All* score, we continue the same procedure as for the classic SB-STI using Eq. (5), (6) and (7) as described in subsection II C.

D. SB-STI *for All* vs. standard STI

The difference between the SB-STI *for All* and the standard STI lies in the computation and rectification of the MTF. In this study, we compute the speech based version of the MTF using the envelope regression method of Eq. (3) and we add the contribution of auditory masking intensity and absolute reception threshold using Moore and Glasberg’s auditory model. Knowing that the standard STI is well correlated with subjective intelligibility measures for NH persons¹¹, we parameterized the SB-STI *for All* algorithm with a NH audiogram in order to compare both metrics. The in-lab French speech database used for the comparison test was developed by the “Collège National d’Audioprothèse”³⁵ and was originally used in hearing in noise tests (HINT) by audiologists. The HINT database is composed of 5 phonetically balanced lists with 20 sentences each, each sentence has a duration ranging from 1.5 to 2.5 seconds. All sentences are in French, pronounced by the

same male speaker and sampled at 44.1kHz. The data base was distorted using 7 RIRs with reverberation time RT60 ranging from 0.5 to 5 seconds. The RIRs were computed with the image source model³⁶. The reverberated sentences were then distorted with speech shaped noise (like in¹¹) with SNR values in $[-12, 10]$ dB with a step of 2dB. This procedure yielded 56 different test conditions.

The mapping between the SB-STI *for All* scores for NH persons (hearing loss $HL=0$ dB) and standard STI scores computed with indirect method is displayed in Fig. 7. The results show that there is a linear dependence between the two scores with a correlation coefficient $\rho = 0.94$. This shows that the proposed SB-STI *for All*, parameterized for NH persons, provides a good prediction of the standard STI. Besides, we note that the STI scores are higher than the SB-STI *for All* scores. This can be explained by the fact that the standard STI scores here are computed using Schroeder's MTF method (cf. Eq. (2)), which cannot be rectified regarding masking effects and absolute reception threshold.

To measure the performance of the SB-STI *for All* algorithm for variable hearing ability we designed subjective intelligibility tests using both, NH and HI panels, presented in the following section.

IV. SUBJECTIVE INTELLIGIBILITY ASSESSMENT

Two subjective intelligibility tests, in-lab and in-vivo, were designed to assess the performance of the SB-STI *for All* metric. The in-lab test was realized in a listening room using headsets and is described in details in³⁷. The in-vivo test was conducted in a shed which simulates the acoustic properties of a railway station. Both test procedures are described in the following.

A. In-lab intelligibility test with NH and HI persons

1. In-lab speech material and listening panel

For this in-lab test we used the HINT database (cf. section IIID) distorted with reverberation using RIRs generated by the image-source model³⁶ and additive babble noise. Table I summarizes the test conditions of the intelligibility assessment. The HINT database is composed of 5 phonetically balanced sentences. The three first lists (L_1 , L_2 , L_3) were

FIG. 8. Age-related hearing loss distribution of HI test participants. 10 hearing impaired persons (5 female, 5 male)

distorted with the same test conditions (SNR, RT60). This subjective intelligibility test was also used to evaluate intelligibility enhancement algorithm as described in previous work³⁷. The speech database was normalized to have the same RMS level at -26 dBFs.

The listening panel was composed of 20 subjects with:

- 10 NH persons with an average age of 26,
- 10 HI persons (age related hearing loss) with an average age of 70,
- 10 female and 10 male.

All were naive subjects, i.e. they had no prior experience with the intelligibility test procedure nor with the speech material. The test followed the guidelines for ethical considerations provided by the AAL program⁴ which recommends a transparent procedure for the test subjects by signing an informed-consent³⁸. The HI participants were examined and supervised by an Otolaryngologist (ENT), who provided their audiograms: they had mild to moderate symmetrical sloping hearing loss for both ears with sensorineural origins. The hearing loss distribution of HI listeners is depicted in Fig. 8.

2. *In-lab test procedure*

The experiment took place in a listening room with noise level in quiet 30 dBA and reverberation time $RT60 = 0.2s$. The stimuli were presented to the participants through

headphones (Beyerdynamic DT770) at 75 dB SPL, measured with a Head And Torso Simulator (HATS, B&K type 4128). The participants were asked to repeat the heard sentences and the technician underlined the correctly recognized words in each sentence. The subjective intelligibility score is defined as the percentage of correctly recognized words per sentence.

B. In-vivo intelligibility test with NH persons

1. *In-vivo speech material and listening panel*

The used database for the in-vivo test was composed of real French vocal announcements provided by the French national railway company SNCF. The database is composed of 8 vocal announcements pronounced by a female voice (Simone) and another 8 pronounced by a male voice (David). To avoid bias due to memorization of the vocal announcements, we asked the participants to fill in the blanks only 4 key words for each vocal announcement. The key words can be the destination, provenance, number of the train, platform and departure time.

The listening panel was composed of 17 NH persons (10 women and 7 men) and were tested by groups of variable size. All the test procedure and selection of participants were supervised by professionals from the National Expertise Center of Information Technologies and Communication for Autonomy (CENTICH⁵).

2. *In-vivo test procedure*

To simulate the acoustic environment of a railway station, the test was conducted in a shed with dimensions [L60×W20×H5] meters and with a reverberation time RT60=2.4 seconds. Loudspeakers were used for broadcasting vocal announcements and railway station ambient noise. Six passive loudspeakers were used to ensure homogeneous sound pressure level of the noise in the shed and were positioned in such a way that the listeners are not exposed to direct sound field. Two types of PA systems were used for the playback of the vocal announcements; one composed of 8 standard wide band "sound projectors", positioned around the users and the second composed of 2 column loudspeakers (1m) provided by Active Audio⁶. All the loudspeakers were calibrated to provide a specific sound pressure level controlled by an Automatic Gain Control. This setup allows to have different configurations of SNR, type

FIG. 9. Scatter plot of subjective vs. objective scores of intelligibility of in-lab test. The SB-STI *for All* scores of NH correspond to a flat audiogram with 10 dB HL. The subjective and objective intelligibility scores of each test condition are the average of the 10 scores of each group (NH and HI).

of loudspeaker and speaker voice. Table II summarizes the different test conditions of the in-vivo test.

V. RESULTS AND DISCUSSION

In order to assess the performance of the proposed metric, we computed the SB-STI *for All* with the same test conditions as the subjective in-lab and in-vivo tests, taking into account the audiograms of the NH and HI participants in the algorithm and the broadcasting level.

A. Results of in-lab tests

In case of the in-lab test, the SB-STI score for a given test condition with HI is the average of the scores of the 10 HI participants (using their respective audiograms).

In case of the in-lab test with the NH participants, and according to the definition of a normal hearing adult given by the WHO organization, an audiogram with hearing loss below 25dB for all frequencies should be used. For the computation of the SB-STI *for All* score for NH, a flat audiogram with 10dB hearing loss for all NH participants was used, which yielded

FIG. 10. Mapping of subjective vs. objective intelligibility scores for all the panel of in-lab test (average of 20 persons). The fitting curve is obtained using Eq. (20) where $(a, b, c, d) = (39.52, 33.83, 20.03, 3.612e6)$.

the best correlation with the subjective intelligibility scores. As in the HI case, the SB-STI score for a given test condition is the average of the 10 NH scores.

Fig. 9 shows the scatter plot of subjective vs. objective scores for NH and HI persons. We obtain for both scatter plots a correlation coefficient between objective and subjective scores of about $\rho = 0.84$. If we consider this participants panel as a representative sample of a population, the mapping between subjective and objective scores will be as depicted in Fig. 10. The fitting model of Fig. 10 was computed using the following Eq. (20) to best fit the well known "S" shape between subjective and objective intelligibility measures:

$$f(x) = \frac{a \cdot e^x - c \cdot e^{-x}}{b \cdot e^x + d \cdot e^{-ax}}, \quad (20)$$

where a , c , b and d are the parameters to optimize. On Fig. 10, the correlation coefficient between subjective and predicted intelligibility scores ($f(\text{SB-STI})$) reached $\rho = 0.895$ and the standard deviation was $\sigma = 0.196$.

B. Results of in-vivo tests

The distorted vocal announcements were live recorded during the in-vivo test (with NH participants) using omni-directional microphones. Each recorded vocal announcement was

FIG. 11. Scatter plot and mapping of subjective vs. objective intelligibility in-vivo scores for NH persons. Each point is an average of the scores of 17 NH for a vocal announcement measured objectively and subjectively with the same test condition. The fitting curve is obtained using Eq. (20) where $(a, b, c, d) = (39.48, 33.25, 14.12, 4.013e6)$.

then time-aligned (using a cross-correlation based algorithm) with its reference before the computation of the SB-STI *for All* score. The test results are reported in the scatter plot of Fig. 11. The correlation coefficient between predicted (using the best fitting model of Eq. (20) and subjective intelligibility scores was $\rho = 0.95$ with a standard deviation $\sigma = 0.04$. These results prove that the proposed SB-STI *for All* yields a good average intelligibility prediction for a NH population in real environment.

It is also worth to note that when gathering all the in-lab and in-vivo test results (NH and HI) in the same scatter plot (cf. Fig. 12), we may conclude that the in-lab and in-vivo intelligibility scores overlay and can thus be easily associated to the same intelligibility scale.

C. Discussion

The present study puts together two known models, the first is the speech based STI for intelligibility prediction and the second is Moore and Glasberg' auditory model for perceived loudness assessment using roex auditory filters. The goal of this study is to develop a useful model that can predict the intelligibility for different hearing abilities. This becomes a necessity knowing that the population is getting older and the services should be adapted to the hearing diversity that constitutes a target group like passengers of railway station,

airports, etc.

The effect of age-related hearing loss and hearing impairment on speech intelligibility was investigated in²⁰ and the ISO60268⁸ proposes a method to re-scale the standard STI metric to fit the intelligibility qualification (from bad to excellent) for different hearing abilities. For example, to obtain a *good* intelligibility we need to achieve at least an STI score of 0.6 for NH and of 0.72 for HI, measured with a PTA loss of 15dB. However, the proposed SB-STI *for All* avoids any scale modification and offers a direct assessment of the intelligibility qualification through the use of the specific audiogram profile. Furthermore, the scale adjustment proposed in ISO60268 is based only on the PTA hearing level at 0.5kHz, 2kHz and 4kHz, relative to 18 years old NH listeners. This adjustment lacks on precision because different hearing profiles (i.e. audiograms, and hence different intelligibility experience) may correspond to the same PTA hearing level. Besides, for a PTA equal to 30 dB we can reach excellent (100%) intelligibility (SB-STI score of 0.75) in contrary to what is shown in ISO60268, where for the same PTA fair intelligibility could not be achieved.

This can be explained by the definition of a NH person given in ISO60268, which is different from that of WHO organization: the PTA for NH should be between 0 and 5dB according to ISO60268 and between 0 and 25dB according to WHO organization. This mismatch can be explained by a difference in the definition of the absolute hearing threshold of the 18 years NH listener.

The above mentioned drawbacks are avoided using the SB-STI *for All*, as it is based on a precise auditory model and audiograms to provide a robust intelligibility prediction on a unique scale for HI and NH, as confirmed by the result of Fig. 12, where HI and NH intelligibility ratings correspond to the predicted SB-STI for All scores with a correlation of $\rho = 0.95$ and on a unique scale.

VI. CONCLUSION

The STI metric is a criterium widely used by acousticians to assess the acoustical characteristics of a place. The aim of this study is to extend this metric to account for the diversity of hearing ability and to improve its usability in a real time monitoring context.

We based our approach on the speech based STI which broadcasts real speech in the space to compute the MTF; this technique enables in-vivo intelligibility assessment, taking

FIG. 12. Scatter plot of all (in-lab and in-vivo) intelligibility scores for NH and HI. 37 participants (27 NH and 10 HI) for 47 different test conditions with additive noise and reverberation. The fitting curve is obtained using Eq. (20) where $(a, b, c, d) = (39.48, 33.25, 14.12, 4.013e6)$.

into account the actual conditions of the space (reverberation and noise vs. occupancy). We then rectified the SB-MTF by including Glasberg and Moore’s auditory model to accurately account for perceptual masking intensity and absolute reception threshold; the proposed SB-STI is able to predict the intelligibility experience of HI as well as NH persons, based on the audiogram of a specific auditory profile and the received distorted speech. From that we obtained an audiogram dependent SB-MTF to compute a new STI score, the SB-STI *for All*, which is highly correlated with the standard STI (Schroeder’s method) for the NH profile.

The results of in-lab and in-vivo subjective intelligibility tests show that the SB-STI *for All* yields a correlation of 0.84 and 0.94 with subjective ratings of HI and NH respectively. Furthermore, both the in-vivo and in-lab tests yielded coherent results, i.e. using the same fitting model yielded a good correlation between objective and subjective measures for both NH and HI cases.

However, the proposed metric is not thought to assess the intelligibility experience of a specific person in a particular place, it is rather dedicated to predict the intelligibility experience of a target group with a particular hearing profile.

To perform a comprehensive measure for a given public place, we suggest two strategies. The first is to provide a distribution of SB-STI scores for a given population using the distribution of hearing thresholds as a function of age provided by the standard ISO7029⁷. The

second strategy is to provide four SB-STI scores, for normal hearing and for slight, moderate and severe hearing impairment according to the classification of the WHO organization. Indeed, the metric's objective is to provide a *for All* score that assesses all the system for a given target population with its hearing diversity.

ACKNOWLEDGMENTS

The authors would like to thank all the partners of the I'City for All project under the European Active and Assisted Living Program.

Appendix A: Envelope regression method

In the ER method, it is assumed that for each octave band k ¹⁷:

$$Y_k[n] = A_k X_k[n] + B_k, \quad (\text{A1})$$

where A_k and B_k are the parameters that produce the best linear fit, $X_k[n]$ and $Y_k[n]$ are respectively the clean and distorted envelope samples and n is the time index. The MTF is deduced from the effective Signal to Noise Ratio (eSNR) for each octave band k as follows¹¹:

$$\text{eSNR}_k = 10 \log_{10} \left[\frac{A_k \mu_x}{B_k} \right] = 10 \log_{10} \left[\frac{\text{MTF}_k}{1 - \text{MTF}_k} \right]; \quad (\text{A2})$$

where the MTF of the k^{th} octave band is:

$$\text{MTF}_k = \frac{A_k \mu_x}{A_k \mu_x + B_k}, \quad (\text{A3})$$

and $\mu_x = E\{X_k[n]\}$ ($E\{\cdot\}$ denotes the expected value). Goldsworthy¹⁶ proposes to use the minimum mean square error (MMSE) criterion for the optimal fit and thus the MTF for each band k is given by:

$$\text{MTF}_k = \frac{\mu_x \lambda_{xy}}{\mu_y \lambda_x} \quad (\text{A4})$$

where $\lambda_{xy} = E\{(X_k[n] - \mu_x)(Y_k[n] - \mu_y)\}$ and $\lambda_x = E\{(X_k[n] - \mu_x)^2\}$. We obtain from Eq. (A1) and using unbiased covariance estimation, the MTF expression given by Eq. (3).

REFERENCES

- ¹Website: <http://www.ec.europa.eu/eurostat> (last viewed 01/10/17).
- ²Website: <http://www.icityforall.eu/> (last viewed 01/10/17).
- ³Website: <http://www.who.int/> (last viewed 01/10/17).
- ⁴Website: <http://www.aal-europe.eu/> (last viewed 01/10/17).
- ⁵Website: <http://www.centich.fr/> (last viewed 01/10/17).
- ⁶Website: <http://www.activeaudio.fr/> (last viewed 01/10/17).
- ⁷ISO 7029:2000, “Acoustics – Statistical distribution of hearing thresholds as a function of age,” Standard (International Organization for Standardization, 2000).
- ⁸IEC 60268-16:2011, “Sound system equipment - Part 16: Objective rating of speech intelligibility by speech transmission index,” Standard (International Organization for Standardization, 2011).
- ⁹T. Houtgast and H. Steeneken, “A review of the MTF concept in room acoustics and its use for estimating speech intelligibility in auditoria,” *The Journal of the Acoustical Society of America* **77**, 1069–1077 (1985).
- ¹⁰T. Houtgast and S. van Wijngaarden, *Past, present and future of the Speech Transmission Index*, 159 (TNO Human Factors, 2002).
- ¹¹H. Steeneken and T. Houtgast, “A physical method for measuring speech transmission quality,” *The Journal of the Acoustical Society of America* **67**, 318–326 (1980).
- ¹²M. R. Schroeder, “Modulation transfer functions: Definition and measurement,” *Acta Acustica united with Acustica* **49**, 179–182 (1981).
- ¹³Y. Mzah, S. Maarfi, R. Ghozi, and M. Jaïdane, “Non-monotonic impact of occupancy level on reverberation indicators: Case of a public confined eating establishment,” in *Audio Engineering Society Conference on Dereverberation and Reverberation of Audio, Music, and Speech*, 2–7 (2016).
- ¹⁴V. Desarnaulds, A. Carvalho, and G. Monay, “Church acoustics and the influence of occupancy,” *Building Acoustics* **9**, 29–49 (2002).
- ¹⁵M. Hodgson and E. Nosal, “Effect of noise and occupancy on optimal reverberation times for speech intelligibility in classrooms,” *The Journal of the Acoustical Society of America* **111**, 931–939 (2002).

- ¹⁶R. Goldsworthy and J. Greenberg, “Analysis of speech-based speech transmission index methods with implications for nonlinear operations,” *The Journal of the Acoustical Society of America* **116**, 3679–3689 (2004).
- ¹⁷C. Ludvigsen, C. Keidser, and T. Poulsen, “Prediction of intelligibility of non-linearly processed speech,” *Acta Oto-laryngologica Suppl* **469**, 190–195 (1990).
- ¹⁸K. Payton and M. Shresha, “Evaluation of short-time speech-based intelligibility metrics,” in *9th International Congress on Noise as a Public Health Problem*, 243–251 (2008).
- ¹⁹K. L. Payton and M. Shrestha, “Comparison of a short-time speech-based intelligibility metric to the speech transmission index and intelligibility data,” *The Journal of the Acoustical Society of America* **134**, 3818–3827 (2013).
- ²⁰A. Duquesnoy and R. Plomp, “Effect of reverberation and noise on the intelligibility of sentences in case of presbycusis,” *The Journal of the Acoustical Society of America* **68**, 537–544 (1980).
- ²¹B. Moore and R. Glasberg, “A revised model for loudness perception applied to cochlear hearing loss,” *Hearing Research* **188**, 70–88 (2004).
- ²²J. Festen and R. Plomp, “Effects of fluctuating noise and interfering speech on the speech-reception threshold for impaired and normal hearing,” *The Journal of the Acoustical Society of America* **88**, 1725–1736 (1990).
- ²³J. Sander and H. Steeneken, “Objective prediction of speech intelligibility at high ambient noise levels using the speech transmission index,” *EUROSPEECH* (1999).
- ²⁴J. Bradley, “Predictors of speech intelligibility in rooms,” *The Journal of the Acoustical Society of America* **80**, 837–845 (1986).
- ²⁵B. W. Anderson and J. T. Kalb, “English verification of the STI method for estimating speech intelligibility of a communications channel,” *The Journal of the Acoustical Society of America* **81**, 1982–1985 (1987).
- ²⁶C. Christiansen, M. S. Pedersen, and T. Dau, “Prediction of speech intelligibility based on an auditory preprocessing model,” *Speech Communication* **52**, 678–692 (2010).
- ²⁷A. Tisseya, A. Moulinier, and Y. Rouard, “Intelligibility in various rooms: Comparing its assessment by (RA)STI measurement with a direct measurement procedure,” *Applied Acoustics* **53**, 1–3 (1998).
- ²⁸J. Rennies, H. Schepker, I. Holube, and B. Kollmeier, “Listening effort and speech intelligibility in listening situations affected by noise and reverberation,” *The Journal of the*

- Acoustical Society of America **136**, 2642–2653 (2014).
- ²⁹S. Brachmanski, “Estimation of logatom intelligibility with the STI method for polish speech transmitted via communication channels,” *Archives of Acoustics* **29**, 555–562 (2004).
- ³⁰P. Jianxin, “Relationship between chinese speech intelligibility and speech transmission index using diotic listening,” *Speech Communication* **49**, 933–936 (2007).
- ³¹S. J. van Wijngaarden, H. Steeneken, and T. Houtgast, “Quantifying the intelligibility of speech in noise for non-native listeners,” *The Journal of the Acoustical Society of America* **111**, 1906–1916 (2002).
- ³²M. Elhilali, T. Chi, and S. A. Shamma, “A spectro-temporal modulation index (STMI) for assessment of speech intelligibility,” *Speech Communication* **41**, 331–348 (2003).
- ³³I. Holube and B. Kollmeier, “Speech intelligibility prediction in hearing impaired listeners based on a psychoacoustically motivated perception model,” *The Journal of the Acoustical Society of America* **100**, 1703–1716 (1996).
- ³⁴B. Moore, “Perceptual consequences of cochlear hearing loss and their implications for the design of hearing aids,” in *Ear and Hearing* **17**, 133–161 (1996).
- ³⁵College National d’Audioprothese, “HINT, audiometrie vocale [CD],” Audition France Innovation (2015).
- ³⁶E. Lehman and A. Johansson, “Prediction of energy decay in room impulse responses simulated with an image-source model,” *The Journal of the Acoustical Society of America* **124**, 269–277 (2008).
- ³⁷A. Ben Jemaa, N. Mechergui, G. Courtois, A. Mudry, S. Djaziri-Larbi, M. Turki, H. Lissek, and M. Jaïdane, “Intelligibility enhancement of vocal announcements for public address systems: A design for all through a presbycusis pre-compensation filter,” in *Interspeech*, 70–74 (2015).
- ³⁸Ambient Assisted Living Joint Programme Call 6, “Guide for applicants”, 9 (2013).

TABLE I. In-lab intelligibility assessment: test conditions for each sentence list of the HINT database.

SNR (dB)	$(\mathbf{L}_1, \mathbf{L}_2, \mathbf{L}_3)$	$\mathbf{L}_4$	$\mathbf{L}_5$
-8	RT60 = 2s RT60 = 4s	RT60 = 0.5s RT60 = 2s RT60 = 4s	-
-6	-	RT60 = 0.5s RT60 = 2s RT60 = 4s	-
-4	RT60 = 0.5s RT60 = 2s RT60 = 4s	RT60 = 0.5s RT60 = 2s RT60 = 4s	-
-2	RT60=0.5s	RT60=0.5s	RT60 = 2s RT60 = 4s
0	RT60 = 2s RT60 = 4s	-	RT60 = 0.5s RT60 = 2s RT60 = 4s
2	-	-	RT60 = 0.5s RT60 = 2s RT60 = 4s
4	RT60 = 2s RT60 = 4s	-	RT60 = 2s RT60 = 4s

TABLE II. Test conditions of the in-vivo test (RT60=2.6 seconds). VA: vocal announcement.

VA	Speaker	PA system	SNR (dB)
#1 to 4	F	Sound Projector	5
#5 to 8	M	Column	8.5
#9 to 12	F	Column	9
#13 to 16	M	Sound Projector	5

LIST OF FIGURES

1	Block diagram of STI: filtering clean and distorted signals in 7 octave bands, envelope extraction (wave rectifier method), MTF computation in each octave band, MTFs rectification using auditory model and STI score computation using octave band importance and information redundancy weights.	4
2	Effect of noise and reverberation on the MTF of octave band 1 kHz (Schroeder's method) with variable RT60 and SNR.	5
3	Auditory masking intensity of octave band k on band $k + 1$ following the empirically deduced masking slope.	8
4	Auditory model diagram: the power in each ERB band is calculated from the amplitude spectrum of the distorted speech, then the auditory filters are computed which also vary with hearing loss of outer hair cells (HL_{ohc}). The excitation pattern in each octave band is computed using the generated auditory filters and speech spectrum in each band. From that the intensity of auditory masking (I_{am}) is deduced. The intensity of absolute reception threshold (I_{art}) is deduced from the audiogram.	11
5	Auditory filters centered at $f_c = 1$ kHz and 4 kHz for different HL_{ohc} values and sound pressure levels (L_{rms}). The filters centered at 1 kHz: $L_{rms}=75$ dB and $HL_{ohc}=0, 40$ and 90 dB. The filters centered at 4 kHz: $L_{rms}=30, 75$ and 80 dB and $HL_{ohc}=0$ dB.	13
6	Excitation pattern in octave band centered at 2 kHz for a NH (dashed) and a HI person (dotted) with a flat hearing loss of 60 dB ($HL_{ohc} = 0.9HL$). The speech spectrum was computed using a Hamming window equal to the length of the sentence (2 seconds) and 44.1 kHz sampling frequency.	14
7	Mapping between standard STI (Schroeder's method) and SB-STI <i>for All</i> for NH persons. 56 test conditions were used with RT60 in $[0.5, 5]$ seconds and SNR in $[-12, 10]$ dB.	16
8	Age-related hearing loss distribution of HI test participants. 10 hearing impaired persons (5 female, 5 male)	18

623	9	Scatter plot of subjective vs. objective scores of intelligibility of in-lab test.	
624		The SB-STI <i>for All</i> scores of NH correspond to a flat audiogram with 10 dB	
625		HL. The subjective and objective intelligibility scores of each test condition	
626		are the average of the 10 scores of each group (NH and HI).	20
627	10	Mapping of subjective vs. objective intelligibility scores for all the panel of in-	
628		lab test (average of 20 persons). The fitting curve is obtained using Eq. (20)	
629		where $(a, b, c, d) = (39.52, 33.83, 20.03, 3.612e6)$	21
630	11	Scatter plot and mapping of subjective vs. objective intelligibility in-vivo	
631		scores for NH persons. Each point is an average of the scores of 17 NH for	
632		a vocal announcement measured objectively and subjectively with the same	
633		test condition. The fitting curve is obtained using Eq. (20) where $(a, b, c, d) =$	
634		$(39.48, 33.25, 14.12, 4.013e6)$	22
635	12	Scatter plot of all (in-lab and in-vivo) intelligibility scores for NH and HI. 37	
636		participants (27 NH and 10 HI) for 47 different test conditions with additive	
637		noise and reverberation. The fitting curve is obtained using Eq. (20) where	
638		$(a, b, c, d) = (39.48, 33.25, 14.12, 4.013e6)$	24