Technical note

A revised speech spectrum for STI calculations

Lorenzo Morales[a,*], Glenn Leembruggen[b], Stephen Dance[c], Bridget M. Shield[c]

[a] School of Computing, Science and Engineering, University of Salford, United Kingdom
[b] Acoustic Directions, Sydney, Australia
[c] School of Built Environment and Architecture, London South Bank University, United Kingdom

A B S T R A C T

The ability of the Speech Transmission Index (STI) to predict speech intelligibility under noisy conditions is highly dependent on the assumed spectrum of the speech signal. Examination of the literature showed that the long-term average speech spectrum of male talkers differs substantially from the speech spectrum recommended for STI calculations (IEC 60268-16). To explore these issues, the long-term average speech spectrum of forty male British English people was first measured, compared with the available literature and proposed for STI calculations. Then, using several voice alarm systems, the influence of the measured spectrum on STI calculations was assessed and comparisons made with the standard speech spectrum. The results showed significant STI differences under noisy conditions and considerable reductions in the required electrical power with the use of the new proposed male spectrum. This indicated that the current STI method could benefit from a revised speech spectrum.

1. Introduction

The speech transmission index (STI) [1] is one of the most widely used standardized methods for objective prediction of speech intelligibility. The STI method has been subjected to development and refinement since its introduction in the 1970s. Continuous improvements have been incorporated in successive revisions of the IEC60268-16 standard "*Objective rating of speech intelligibility by speech transmission index*" [2]. The ability of STI to predict the speech intelligibility of a transmission channel under noisy conditions is highly dependent on the assumed spectrum of the speech signal. The STI method included in IEC 60268-16 specifies a standard male speech spectrum for STI calculations.

The use of a specific speech spectrum has also been recommended in ANSI standards since 1969 for calculations of objective measures of speech intelligibility. The speech spectrum recommended in ANSI S3.5 1969 (R1986) *American National Standard methods for the Calculation of the Articulation Index* [3] was mainly based on Dunn & White [4] and French and Steinberg [5] spectra. The spectrum recommended in a later version of the standard, ANSI S3.5 1997 (R2007) *American National Standard methods for the Calculation of the Speech Intelligibility Index* [6], was derived from Pavlovic in 1987 [7]. The speech spectrum for male talkers proposed by Pavlovic was the result of a compilation of several works on speech spectra, which also included the spectra proposed by Dunn & White [4] and by French and Steinberg [5].

The IEC 60268-16 standard introduced in 1998 gender specific test signals for male and female talkers to be used for the STI calculations. The recommended spectra were taken from Steeneken works in 1992 [8]. Since its first introduction in 1998, the spectra remained unchanged in later revisions of the standard, which were published in 2003 and 2011. Fig. 1 shows the speech spectra from various sources. The spectra from ANSI S3.5 1969, ANSI S3.5 1997 (R2007) and the speech spectrum for male talkers proposed by IEC 60268-16:2011 are included in the figure. The spectra have been represented normalised to a level of 0 dB at 1 kHz.

It can be seen that the IEC curve attempts to be a possible "best fit" to the other spectra for frequencies above 400 Hz. The slope of -6 dB per octave presented by the IEC curve could have been selected due to practical reasons as it would have been easy to obtain with an analogue filter.

Most of the published data for long-term averaged speech spectra (LTASS) are for the English language as spoken in the USA [5,9–11], Australia [12], England [13], and other countries [14]. Although small differences have been found among English as spoken in different countries and among different languages, these differences are unclear as large differences among individuals have been found [15–17]. Cox and Moore [11] concluded in their studies that the variability in spectra found among several languages was mostly due to individual talker differences. Additional research presented by Pavlovic et al. [18] indicated that there was no difference in the speech spectra obtained for several languages.

---

* Corresponding author.
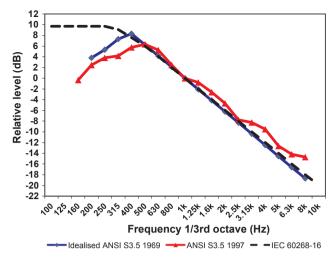 *E-mail address:* lormorales@yahoo.es (L. Morales).

Fig. 1. Male speech spectra from different sources. All the spectra levels are normalised to 0 dB at 1 kHz.

In addition to talker variation, differences among the spectra found in the literature could be due to variations in the measurement techniques. Important factors of these techniques are the angle of incidence of the microphone and its proximity to the talker's mouth utilized during the recordings. Most of the studies have used 0° incidence angle of the recording microphone to the talkers' mouth. Considerable differences may be obtained when speech is recorded with an angle of incidence substantially different from 0° [19]. Early research also highlighted the presence of artefacts in the speech recordings due to close talking conditions for recordings made at 30 cm from the subjects' lips [4].

Although the LTASS is expected to be influenced by the type of speech material that is analysed, early studies of Benson and Hirsh [9] concluded that the choice of speech material is not critical provided that is not grossly unrepresentative phonemically, such as speech passages containing repetition of **a** few phrases. It is also been established that the LTASS varies with vocal effort [14,20,21] and that it is different for men and women.

The most extensive data on LTASS was provided by Byrne and colleagues [22] in 1994. In their work, speech spectra for twelve languages were analysed. A passage from a story book was used for the recordings from which 64 s of speech were analysed. The microphone was placed in front of the talker at 20 cm in the same horizontal plane as the mouth and at an azimuth of 45° incidence, relative to the axis of the mouth. The short distance between talker and microphone was due to the lack of anechoic rooms for some of the recordings and the azimuth angle of 45° was employed to avoid artefacts related to these close talking conditions. Thirty-two talkers participated in the recordings for the British English language of which at least 10 were male [22]. No information was provided of whether the recordings for the English language took place in anechoic conditions.

Research have found a decrease in high frequency levels for a microphone placed at 45° horizontal incidence compared with levels obtained on axis to the mouth (0°) [23,24]. The sound picked-up with a microphone placed at a horizontal angle of 45° has also been defined as "slightly dull and distance", when compared with the sound recorded in front of the mouth [25].

Examination of the literature therefore showed a variety of techniques employed to measure speech spectra. It also showed that the reported long-term average speech spectrum of male talkers differs substantially from the standard STI speech spectrum recommended in the IEC 60268-16 standard. This difference in spectra could lead to potential errors in the STI calculations which are expected to be larger for noisy conditions.

Taking into account the effects on the speech spectra associated with the different techniques previously employed in the literature, an experiment was designed to obtain the male British English LTASS under anechoic conditions. The measurement results were compared with results found in the available literature and then, using several voice alarm systems, the influence of this measured spectrum was assessed on STI calculations and compared with the standard speech spectrum.

2. Experiment 1

2.1. Method

The average long-term spectrum of conversational speech was recorded under anechoic conditions with a microphone positioned in front of the subject's mouths. Several microphone-to-mouth distances and different speech material were employed in the recordings. A B&K 4188 microphone was used for the recordings and was connected to a B&K 2236 sound level meter (SLM) through an extension cable, with the SLM output connected to a Marantz PMD 661 recorder. The subjects' speech was recorded into a wav format at a sampling frequency of 44.1 kHz and 16 bits resolution. The microphone was placed at the same height as the subjects' mouths.

2.2. Subjects

Forty male native British-English participated in the recording test. The age of the talkers ranged between 18 and 62 years old as shown in Fig. 2. No subject reported any hearing or speech impairments. Ethical approval was sought and consent was given prior the experiments.

2.3. Speech material

Three Harvard phonetically-balanced sentence lists extracted from the *IEEE Recommended Practice for Speech Quality Measurement* [26] were used for the recordings. Each Harvard list contains 10 sentences incorporating specific phonemes at the same frequency as they appear in the English language.

The three selected Harvard lists provided a few seconds more than the required 64 s of speech [22]. The sentences were written on a thin sheet of cloth positioned in front of the subjects at approximately 3 m from the microphone. The impact of the cloth on the measured spectra was investigated at the microphone location and found to have negligible effect.

The sentence "Joe took father's shoe bench out, she was waiting at my lawn" was also used for the recordings. This particular sentence was chosen as it has been widely used in the previous literature and assumed to have spectrum similar to conversational speech [7,9]. This assumption was found to be reasonably correct, although at some frequencies the agreement was less satisfactory [7]. This sentence was
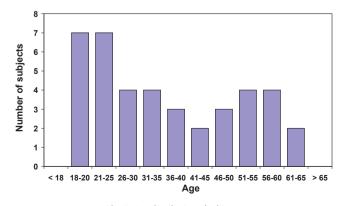

Fig. 2. Age distribution of subjects.

recorded at the same distance as employed for the Harvard sentences (0.5 m), and also at 0.2 m and 1.0 m, all taken directly in front of the subjects' lips. The recordings with this sentence were aimed to investigate the effect on the LTASS of different microphone-to-mouth distances, and they permitted a comparison with the results obtained with the Harvard sentences.

2.4. Procedure

The subjects were seated in a fixed chair placed within the anechoic chamber and given instructions to sit straight with their backs against the chair and to direct their voice towards the microphone. Prior to the recordings, the subjects held a short conversation with the researcher to allow them to adjust to the anechoic environment. This adjustment period was intended to discourage the subjects from involuntarily raising their voice to compensate for the lack of reverberation and the extremely quiet conditions.

The participants were instructed to read the sentences from the cloth sheet located in front of them at a normal speed and to try to keep a constant normal level avoiding raising their voices. The rate of the speech was of 2–3 words per second and the level of the speech was monitored on a VU meter of the sound recorder. The subjects repeated sentences in which they made a mistake or in which they felt their fluency was unsatisfactory. The microphone was placed at 0.5 m from the subjects' mouths for the Harvard sentence recordings, and at 0.5 m, 0.2 m and 1.0 m for the sentence "Joe took… waiting at my lawn".

Hanning windows 1 s long with 50% overlap were used with the waterfall function in WinMLS 2004 [27]. The frequency response of each slice was computed with a resolution of 1.0 Hz. For each slice, one third of octave band spectra were obtained from 100 Hz to 10 kHz and each segment averaged as established in the WOSA procedure [28]. Prior to the spectral analysis of the speech, each recorded sentence was normalised to the same dBA level with Adobe Audition 3.0. Additional analysis of silences recorded with the speech materials indicated that the background noise present during the recordings was at least 30 dB or lower than the speech for all the frequencies investigated (100 Hz–10 kHz). All spectra were corrected for the response of the B&K 4188 microphone used for the recordings.

2.5. Results and discussion

The L_{Aeq} level of each talker was found by first removing the gaps between sentences using Audition v3.0 and then feeding the edited speech electrically into a B&K 2260 sound level meter. The forty talker levels with the Harvard sentences were arithmetically averaged and found to be 63.8 dBA with a standard deviation of 2.7 dB, which was 0.6 dB higher than the levels obtained with the sentence "Joe took…at my lawn". These measured levels are in agreement with current data reported for vocal efforts in anechoic conditions [21]. For each subject, the measured 1/3rd octave band spectra were adjusted to normalise their overall level to 64 dBA. Fig. 3 shows the speech spectra results for the 40 subjects with the Harvard sentences measured at the distance of 0.5 m.

The LTASS results presented in Fig. 3 showed a high variability among subjects, which is in agreement with previous works found in the available literature. Fig. 4 shows the average spectrum obtained for the 40 subjects with Harvard sentences at 0.5 m. The IEC male spectrum [2], the ANSI S3.5 1997 (R2007) [6] spectrum and the male spectrum from Byrne et al. [22] were also plotted. Each of the 1/3rd octave bands was calculated on an average basis [7,22]. All spectra values were normalised to an overall broadband level of 64 dBA. Tabulated values can be seen in Table 1.

The LTASS results presented in Fig.4 for the current study showed a good agreement with the IEC male spectrum for frequencies between 400 Hz and 2 kHz, and large differences were found for frequencies outside that range.
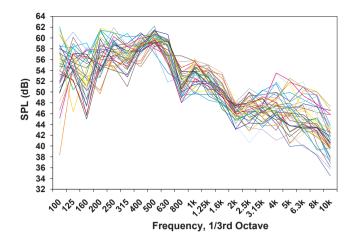

Fig. 3. LTASS for 40 subjects measured with Harvard sentences at 0.5 m. Each subject's spectrum was normalised to 64 dBA.
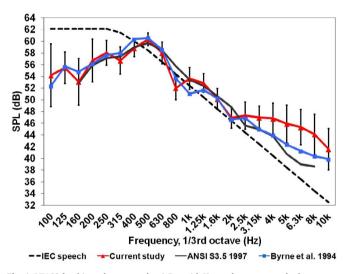

Fig. 4. LTASS for this study measured at 0.5 m with Harvard sentences and other spectra. Error bars indicate one standard deviation. All spectra were normalised to an overall level of 64 dBA.

The LTASS results presented in Fig. 4 showed a good agreement with the ANSI spectrum and with the LTASS presented by Byrne et al. [22] for frequencies below 2 kHz. For the 1/3rd octave bands of 800 Hz and 1 kHz, differences of 3–4 dB were found. The spectrum of the current study presented higher energy than the ANSI and the Byrne et al. spectra for frequencies above 2 kHz, which did not employ a microphone with a 0° angle of incidence.

The directivity of human talkers has been widely investigated [19,23–25,29]. Most of these works reported a decrease on the speech levels of approximately 1–1.5 dB for frequencies of 4–5 kHz for a microphone placed at 45° azimuth, when compared with the speech levels measured at 0°, and a decrease of 3–4 dB for frequencies of 6–8 kHz. The results of the current study presented 3 dB higher levels for frequencies of 4–5 kHz than Byrne et al. [22], which is likely due to the expected directivity of the speech. The current study results also presented approximately 4 dB higher levels for frequencies of 6–8 kHz than Byrne et al. [22], which is in line with the reported 3–4 dB directivity attenuation for a microphone placed at 45° azimuth found in the literature.

It is also noted that the average spectra of the UK male talkers presented in Fig. 1 of Byrne et al. [22] shows between 1 and 3 dB higher levels in the 2.5 kHz to 5 kHz region that the world average. This explains the 1–1.5 dB difference found between the current study and Byrne's for frequencies of 4–5 kHz. Hence, the 0° incidence of the

Table 1

SPL (dB) for the IEC spectrum and for the male spectrum obtained in the current study. All levels were normalised to an overall SPL of 64 dBA. Integrated octave band values are also given.

Frequency (Hz)	IEC 60268-16 (dB)		Current study (dB)	
100	62.1	–	54.2	–
125	62.1	66.8	55.5	59.1
160	62.1	–	53.1	–
200	62.1	–	56.7	–
250	62.1	66.6	58.0	62.0
315	61.5	–	56.7	–
400	60.0	–	58.8	–
500	58.4	63.2	60.3	63.9
630	56.5	–	58.0	–
800	54.4	–	52.0	–
1 k	52.4	57.4	53.6	57.7
1.25 k	50.4	–	52.9	–
1.6 k	48.4	–	50.2	–
2 k	46.4	51.4	46.9	53.2
2.5 k	44.4	–	47.3	–
3.15 k	42.4	–	47.0	–
4 k	40.4	45.4	46.8	51.4
5 k	38.4	–	45.9	–
6.3 k	36.4	–	45.3	–
8 k	34.4	39.4	44.1	48.7
10 k	32.5	–	41.6	–
dBA	64.0		64.0	

microphone employed in the current study can mostly explain the higher levels found here for frequencies above 2 kHz when compared with the male spectrum of Byrne et al. [22].

The ANSI speech spectrum presented similar energy distribution than Byrne et al. [22] for frequencies up to 4 kHz, and 2–3 dB lower energy for frequencies above 4 kHz. However, different recording techniques were employed to obtain both spectra and a variety of speech material was also employed to obtain the ANSI spectrum [7].

Differences were also found between the ANSI spectrum and that obtained in the current study, which could probably be attributed to the difference in the employed speech material and to the variety of recording methodologies employed in the several works compiled to obtain the ANSI spectrum [7].

The large differences found for frequencies below 400 Hz between the IEC male spectrum and the remaining spectra (ANSI, Byrne et al. and this study) suggests that the low frequency content of the standard male speech has been overestimated by IEC 60268-16 [2].

Fig. 5(a) shows the difference between the LTASS obtained with Harvard sentences and the sentence "Joe took … at my lawn", while

Fig. 5(b) shows the differences between the LTASS measured at 0.2 m, 0.5 m and 1.0 m using the sentence "Joe took … at my lawn".

The results presented in Fig. 5(a) show good agreement at frequencies below 2 kHz between the LTASS of the Harvard sentences and the sentence "Joe took…at my lawn". Worse agreement was found for frequencies above 2 kHz. Fig. 5(b) shows that the distance employed for the recordings did not influence the LTASS results substantially. Spectrum differences within ± 1.0 dB were found for most of the frequencies. The variation in the LTASS with the recording distance was found to be lower for frequencies above 1.6 kHz, and higher variation was found for frequencies between 200 Hz and 1.25 kHz. It can also be noted that variations in the LTASS with microphone distance were found to be smaller than the measured variation among talkers.

Taking into account the variety of speech material and distances talker-microphone employed in the current study, it is considered that the male spectrum obtained with Harvard sentences at the distance of 0.5 m is expected to better represent a LTASS for male British English speakers for use with voice alarm systems. This choice has been made based on the fact that the intelligibility of voice alarm systems is one of the primary uses of the STI method.

2.6. Proposed male spectrum

Table 2 shows the normalised values proposed in this study for the LTASS for British male English speakers, along with the male-talker equivalents given in IEC 60268-16 [2]. Both the LTASS and IEC spectra are normalised to an A-weighted level of 0 dB for easy scaling.

Both spectra included in Table 2 will be used in the next section for STI calculations.

3. Experiment 2

3.1. Introduction

The influence of the proposed spectrum on STI calculations was investigated with seven voice alarm systems and comparisons results were made using both the standard speech spectrum and the proposed male spectrum given in Table 2.

Impulse response measurements carried out with six voice alarm systems installed at public spaces and a virtual voice alarm system designed for the stands of a typical medium size football stadium (30,000 spectators) were employed for the STI calculations. All seven voice alarm systems utilised distributed loudspeakers and were designed to produce specific speech levels above the expected ambient
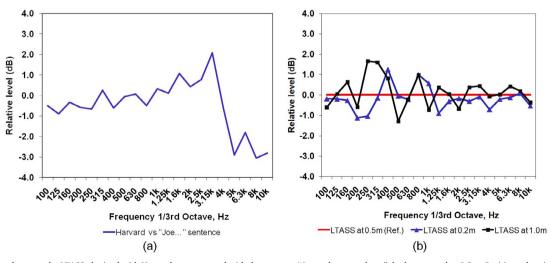

(a)

(b)

Fig. 5. (a) Difference between the LTASS obtained with Harvard sentences and with the sentence "Joe took…at my lawn", both measured at 0.5 m. Positive values indicate higher levels with the sentence "Joe…….". (b) Difference between LTASS measured at 0.2 m, 0.5 m and 1.0 m with the sentence "Joe took … at my lawn".

Table 2
Octave levels (dB) relative to the A-weighted speech level from IEC 60268-16 and for the recommended levels from this study, both adjusted to produce 0 dBA.

Octave band Hz	125	250	500	1 k	2 k	4 k	8 k	A-weighted
Male (IEC-60268-16) dB	2.9	2.9	−0.8	−6.8	−12.8	−18.8	−24.8	0.0
Male (This study) dB	−4.8	−2.0	0.0	−6.3	−10.8	−12.6	−15.2	0.0

Conference room, LSBU, London (1).

Grand Concourse at Sydney Central Station (2).

Suburban train station platform, in Sydney (3).

Overground Platform Central Station Sydney (4)

Underground station Sydney (5)

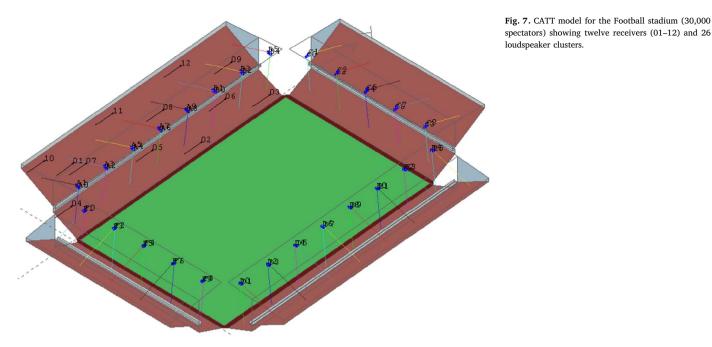Road tunnel Brisbane Australia (6).

Fig. 6. Photos of six of the seven public spaces employed for the STI calculations. The space number is included between brackets.

noise. For each of the seven systems, acoustical parameters, SPL and STI, were assessed at twelve positions. The STIs were derived from the impulse responses using the Indirect Method [2] (Schroeder equation) and post processing according to the methodology stated in IEC 60268-16:2011 [2].

3.2. Seven voice alarm systems

The seven voice alarm systems installed at seven public spaces were selected such as they represent spaces fairly used by public in real life and also provide a variety of system levels ranging from medium to very high SPL. All the seven spaces presented distributed loudspeaker systems. The distributed loudspeaker systems were composed by cabinet loudspeakers mounted on the walls for the Conference room (space 1), beam steered arrays with not time sequenced for the Train station concourse (space 2), bi-directional loudspeakers for the Suburban train station platform (space 3), distributed projector loudspeakers for the Overground station platform (space 4), recessed ceiling loudspeakers for the Underground station platform (space 5), time sequenced horns located at 40 m intervals for the Road tunnel (space 6), and 16 m spaced clusters of two horn loaded loudspeaker cabinets for the football stadium (space 7). Fig. 6 shows the six real public spaces and Fig. 7 shows the 3D computer model employed for the football stadium.

For the football stadium (medium size, 30,000 spectators), a 3D computer model was built with CATT-Acoustic software V9.0c [30]. Dimensions of the grandstands were taken from UEFA 2012 [31] and absorption coefficients taken from Egan1988 [32] and Kuttruff 2009 [33] were employed for the STI predictions. The longest stands (North and South) had a length of 120 m each and the shorter stands (East and West) had a length of 80 m. The maximum height measured from the

Fig. 7. CATT model for the Football stadium (30,000 spectators) showing twelve receivers (01–12) and 26 loudspeaker clusters.

Table 3
L_{10} ambient noise levels measured at the six public areas.

Area	Ambient noise (L_{10},dB)							
	125 Hz	250 Hz	500 Hz	1 kHz	2 kHz	4 kHz	8 kHz	A-w
Conference room	75	69	64	61	59	57	55	68
Train station concourse	69	68	67	65	62	58	55	70
Suburban train station	68	68	71	67	62	59	59	72
Overground platform	72	75	74	69	65	63	57	75
Underground platform	78	81	83	77	72	65	58	83
Road tunnel	80	85	82	80	76	70	62	85

Table 4
Octave band reverberation times (early decay times) measured with the voice alarm systems at the six public areas.

Area	Reverberation time (EDT, seconds)						
	125 Hz	250 Hz	500 Hz	1 kHz	2 kHz	4 kHz	8 kHz
Conference room	0.8	0.6	0.6	0.6	0.5	0.5	0.4
Train station concourse	2.3	3.7	4.2	4.1	3.7	2.3	1.2
Suburban train station	0.5	0.6	0.5	0.5	0.4	0.4	0.4
Overground platform	1.4	1.7	1.6	1.6	1.3	1.0	0.5
Underground platform	–	1.2	3.1	2.9	2.1	1.6	0.9
Road tunnel	–	5.7	4.9	3.8	2.8	1.9	0.9

pitch level to the front of the canopy roof was 24 m for all the stands. The depth of all the stands was 30 m from the pitch edge to the back wall (when measured parallel to the ground). All the stands included upper and lower tiers and have identical cross-sections. The dimensions of the pitch were 124 × 84 m. Fig. 7 shows the CATT model employed in the predictions.

With the exception of the Conference room, all the acoustic data for real voice alarm systems was provided by Acoustic Directions Ltd., Sydney, Australia.

3.3. Method

Acoustical parameters, output SPL and STI, were measured at twelve positions evenly distributed throughout the six real spaces. The STIs were measured using the IEC60268-16:2011 [2] Indirect Method (Schroeder equation using the impulse response). With the exception of the Conference room, the impulse responses were obtained with a B&K 2250 sound level meter equipped with a B&K 4189 microphone which was fed to a computer installed with WinMLS 2004 software.

A B&K 2236 sound level meter equipped with a B&K 4188 microphone was used for the Conference room area. The B&K 2236 sound level meter was also fed to a computer installed with WinMLS 2004 software. The B&K 4189 and 4188 are both 1/2-inch-diameter omni-directional microphones. All the STI measurement results presented in this section refer to the current revised method published in 2011, and were obtained employing the methodology and the male input spectrum recommended in IEC60268-16:2011 [2].

Octave band L_{Eq} and percentiles ambient noise levels were also measured for each area. For the Conference room, ten minutes long measurements were carried out in occupied conditions. Trains were in operation during the measurements for all the Overground, Underground, and Train station areas. The duration of the measurements matched the time of one train cycle. Three train cycles were measured and the results were averaged. An emergency evacuation scenario due to a fire alert was considered for the Road tunnel area. Hence, the measured noise produced by smoke extractor fans distributed through the tunnel was employed in the STI calculations. Ten minutes ambient noise measurements were carried out within the area. Table 3 provides the results of the ambient noise measurements (L_{A10}) for the six areas. L_{10} is the level which is exceeded the 10% of the time and it has been recommended for voice alarm system designs due to a "good practice" criterion [34]. STI results calculated with L_{10} ambient noise would be equal or higher than the reported values for the 90% of the time and lower for the remaining 10% of the time.

It should be noted that measurements for output SPL, STI and other acoustical parameters were carried out in quiet conditions. This is, all the areas were mostly free from public; trains were not running in the station areas; and the Road tunnel was free from cars and the extractor fans were not in operation. On the other hand, ambient noise measurements took place at busy times, with occupied areas and with running trains in the stations. Noise measurements for the Road tunnel

Table 5
Octave band and A-weighted output SPL employed for the STI calculations.

Area	Octave band levels with an IEC input spectrum (dB)							
	125 Hz	250 Hz	500 Hz	1 kHz	2 kHz	4 kHz	8 kHz	A-w
Conference room	82	81	76	71	66	60	58	78
Train station concourse	83	83	79	73	67	62	55	80
Suburban train station	76	82	82	77	70	62	55	82
Overground platform	81	87	84	79	73	67	60	85
Underground platform	87	93	89	84	79	72	63	90
Road tunnel	50	91	95	89	83	75	68	95

Table 6
L_{10} ambient noise and average system output levels for the Football stadium.

	Football stadium							
	125 Hz	250 Hz	500 Hz	1 kHz	2 kHz	4 kHz	8 kHz	A-w
Ambient noise (L_{10},dB)	76	81	91	91	86	77	65	94
SPL with an IEC input spectrum (dB)	107	107	103	97	91	85	79	104

Table 8
Averaged SNR calculated with the proposed male spectrum and with the IEC spectrum for the Grand concourse.

	Average signal-to-noise ratio (dB)						
	125 Hz	250 Hz	500 Hz	1 kHz	2 kHz	4 kHz	8 kHz
1. IEC male spectrum	13.7	14.8	12.3	7.9	4.7	3.2	0.0
2. Proposed male spectrum	6.0	9.9	13.1	8.3	6.7	9.4	9.6
Difference (2 − 1)	−7.7	−4.9	0.8	0.4	2.0	6.2	9.6

Table 7
STI difference between the calculation results with the IEC spectrum and the proposed male spectrum (seven voice alarm systems). CR is the Conference room (1), GC is the Grand Concourse (2), STS is the Suburban train station platform (3), OP is the Overground platform (4), UP is the Underground platform (5), Rd. T is the Road tunnel (6), and FS is the Football stadium.

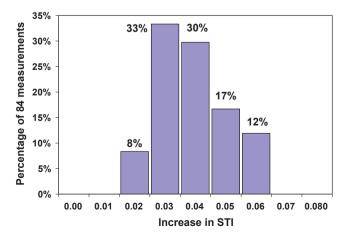|	STI difference (Proposed male spectrum minus IEC values)						
Position	CR (1)	GC (2)	STS (3)	OP (4)	UP (5)	Rd. T (6)	FS
Pos 01	0.05	0.04	0.06	0.04	0.02	0.03	0.04
Pos 02	0.06	0.03	0.06	0.04	0.03	0.03	0.04
Pos 03	0.05	0.04	0.05	0.03	0.03	0.03	0.04
Pos 04	0.05	0.05	0.05	0.04	0.02	0.03	0.04
Pos 05	0.05	0.04	0.05	0.04	0.02	0.04	0.04
Pos 06	0.06	0.04	0.05	0.04	0.03	0.03	0.03
Pos 07	0.05	0.04	0.06	0.04	0.02	0.03	0.04
Pos 08	0.05	0.04	0.06	0.04	0.02	0.03	0.04
Pos 09	0.05	0.04	0.06	0.04	0.03	0.03	0.04
Pos 10	0.05	0.05	0.06	0.04	0.03	0.03	0.04
Pos 11	0.05	0.04	0.05	0.04	0.03	0.03	0.04
Pos 12	0.06	0.04	0.06	0.04	0.03	0.02	0.04

Fig. 8. Percentage distribution of the increase in STI obtained with the proposed male spectrum for a total of 84 positions taken over the voice alarm systems.

were carried out with the area free from cars but with the smoke extractors in operation.

For the Conference room, a noisy scenario in which the area was fairly occupied was considered for the STI calculations. Due to the heavily upholstered chairs of this area, different levels of occupancy were not expected to have a large impact on the acoustic conditions. Therefore, the impulse responses obtained in empty conditions were considered to be adequate to produce STI calculations under occupied conditions. Additional acoustical parameters such as reverberation times were also computed from the impulse responses. Table 4 provides a summary of the measured averaged reverberation times for the six areas (early decay times). Early decay times have been reported here due to their good correlation with STI [35].

The output SPL of each system was measured employing a standard pink-noise with the male spectrum recommended in IEC60268-16:2011 [2]. Octave bands from 125 Hz to 8 kHz and overall A-weighted levels were obtained for each area at the twelve selected positions by means of L_{Eq} measurements 10 s long. For consistency in the STI calculations, the measured SPLs were mathematically adjusted to 10 dBA above the measured ambient noise for most of the areas. The relative levels between the octave bands were kept constant in the process. The exception was the Underground platform for which, the measured system level of 90 dBA was employed in the calculations. This level led to a signal-to-noise ratio of 7 dBA. The adjusted SPLs employed in the STI calculations can be seen in Table 5.

For the Football stadium, a scenario with 80% of the stands occupied with people was considered for the STI predictions. Scattering coefficients were selected using the guidelines included in the CATT-Acoustic software manual. Temperature of 20° and humidity of 50% were also used for the predictions. A total of 26 loudspeaker clusters comprising of two JBL PD5322-95 loudspeakers were located in the model at intervals of 16 m on the canopy roof at 3.0 m from the front edge. The directivity file of the loudspeaker employed in the predictions was supplied by its manufacturer [36]. Twelve receivers located at 1.2 m from the floor were distributed throughout the North stand area.

Impulse responses and the system output SPL were predicted at each of the twelve receivers with all the loudspeakers of the four stands in operation, as recommended in BS 7827–2011 [37]. The output SPL of the system was predicted employing the male spectrum recommended in IEC60268-16:2011. Octave bands from 125 Hz to 8 kHz and overall A-weighted levels were obtained at each of the twelve receivers.
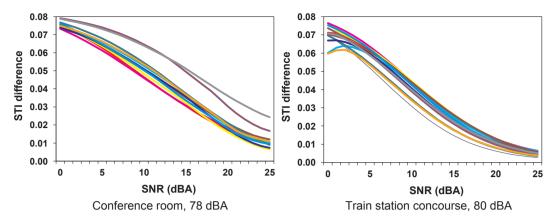
Conference room, 78 dBA

Train station concourse, 80 dBA

Fig. 9. STI difference between the use of the IEC spectrum and the proposed male spectrum, against the average SNR (dBA). STI calculations are presented for each of the 12 positions.
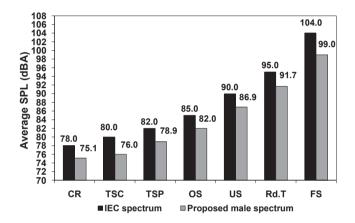

Fig. 10. Average SPLs of the voice alarm systems which are required with the proposed male spectrum to produce the same STI (5th percentile) as the IEC spectrum.

Table 9
Averaged SPL for the Conference room with two input spectra.

Conference room	IEC speech	Proposed male spectrum
Average over 12 positions	75.0 dBA	77.8 dBA
Difference with IEC speech	–	2.8 dBA

Table 10
Averaged SPL for the three additional voice alarm systems with two input spectra. Average of six positions.

Area	IEC speech	Proposed male spectrum	Difference
Cathedral	75.0 dBA	77.8 dBA	2.8 dBA
Museum foyer	75.0 dBA	77.7 dBA	2.7 dBA
Road tunnel	95.0 dBA	97.8 dBA	2.8 dBA

A level of 95 dBA (L_{A10}) was employed as it is recommended by BS 5839–8:2008 [34] for football match crowd noise in stadiums of "30,000 +" spectators. These levels were reduced to correlate to the 80% of occupancy considered here and a value of 94 dBA was obtained: 95 dBA - [10∗log (80/100)] = 94.03 ≈ 94 dBA. Available crowd noise levels measured at the Etihad Football Stadium located in Melbourne, Australia, were employed to determine the relative octave band levels. The crowd noise data was provided by Acoustic Directions Ltd., Sydney, Australia

Following recommendations included in BS 7827–2011 [37], an average of 104 dBA was the target SPL employed for the STI predictions. Table 6 provides the ambient noise levels and system output levels predicted with the CATT model (average over twelve receivers), which were both employed for the STI predictions.

3.4. STI calculations

The following calculations were performed:

1) The STIs were calculated with the standard IEC male spectrum for each position with the system output levels and the noise levels included in Section 3.3.
2) The octave band levels at each measurement position were adjusted according to the proposed male spectrum given in Table 2.
3) The STIs were then calculated with the proposed male spectrum for each position and noise levels included in Section 3.3.

Table 7 shows the differences between the STI results obtained with the IEC spectrum and with the proposed spectrum.

An increase in the STI values with the proposed male spectrum can be seen in Table 7 for all receivers. The STI increases range from a minimum of 0.02 observed for positions of the Underground platform (UP) and the Road Tunnel (Rd.T), to a maximum of 0.06 obtained for positions of the Conference room (CR) and the Suburban train station platform (STS). The percentage distribution of the STI differences presented in Table 7 has been included in Fig. 8.

3.5. Discussion

An STI increase of 0.02–0.06 was found for all the areas with the proposed male spectrum. These increases were due to the higher SNRs at mid and high frequencies with the proposed spectrum. Table 8 shows the average octave-band SNRs calculated with the IEC and the proposed male spectrum for the Grand concourse. The overall SNRs were set to identical A-weighted levels of 10 dB.

Table 8 shows that the use of the proposed male spectrum in the Grand concourse produced a decrease in SNR in the 125 Hz and 250 Hz octave bands, when compared with the IEC spectrum, which are the bands that contribute less to STI values. A small increase in SNR was found when using the proposed male spectrum at 500 Hz, 1 kHz and 2 kHz (2 dB or less), and more than 6 dB was found for the 4 kHz and 8 kHz bands. Similar differences in SNR between both spectra would be expected for all the other voice alarm systems investigated. The changes in STI shown in Fig. 8 depend on the SNRs at which the STIs were calculated. To illustrate this dependency, the STI was calculated at several SNRs and the STI change plotted against the SNR. The system SPLs were kept constant and the background noise in each situation was adjusted in 1 dB intervals whilst retaining its spectral shape. Fig. 9 shows the results for two of the seven areas investigated.

As expected, minimal STI differences were found for high (good) SNR conditions (0.01–0.02 for 25 dBA) whilst higher differences were found for low (poor) SNR conditions (0.06–0.08 for 0 dBA). Additional STI calculations with the proposed male spectrum were carried out for

each area with lower average SPLs. The average SPL with the proposed male spectrum was iteratively reduced until the 5th percentile STI for each area matched the 5th percentile obtained with the IEC spectrum. Fig. 10 shows the new average SPLs results.

Fig. 10 indicates that a lower system SPL for all the areas could be used with the proposed male spectrum, and still achieves a similar 5th percentile STI to the IEC spectrum. The decrease in required SPL varies from a minimum of 2.9 dB for the Conference room to a maximum of 5.0 dB for the Football stadium.

3.6. Electrical power calculations

The electrical power required by a voice alarm system primarily depends on the required SPL and the spectrum of the input signal; different input spectra require different electrical powers for similar SPL targets. A visual inspection of the proposed male spectrum suggests that a lower electrical power would be required to achieve similar sound pressure levels (dBA) than with the IEC male spectrum. Hence, in order to ascertain the effect of the proposed male spectrum on the electrical power requirements, electrical and on site SPL measurements were carried out using a pink noise signal with the IEC male speech spectrum and a pink noise signal with the proposed male spectrum. Both signals were digitally band-pass filtered between 80 Hz and 12 kHz with 24 dB/octave filters and normalised to the same RMS level. The signals matched the relative 1/3rd octave bands target levels (100 Hz–10 kHz) within 0.2 dB.

A-weighted voltage measurements of the normalised signals indicated 2.9 dBA higher level with the proposed male spectrum, compared with the IEC male speech spectrum. Both signals were then employed to measure the SPL at the 12 measurement positions selected for the Conference room and the averaged SPL dBA compared. The results are shown in Table 9.

The measured increase of 2.8 dBA with the proposed male spectrum reproduced by the Conference voice alarm system agreed with the expected increase of 2.9 dBA over the IEC speech level.

Additional SPL measurements were also carried out on three additional voice alarm systems employing the pink noise signals with the IEC male speech spectrum and with the proposed male spectrum described above. The three systems were installed in a Cathedral, a Museum foyer and a road tunnel (St Helena tunnel, Australia). Six measurement positions were taken per area. The results are shown in Table 10.

The SPL measurements returned an average increase of 2.8 dBA with the proposed male spectrum over the IEC speech level, which are similar to the results found for the Conference room. This increase would imply lower electrical power requirements with the proposed male spectrum to achieve similar SPL targets.

4. Conclusions

The LTASS for male British English was measured for 40 people and found to differ substantially from the IEC 60268-16 spectrum specified for STI calculations. Based on the results, a new spectrum for male speakers is proposed for use with STI calculations. Examination of the literature indicated that inter-language differences are lower than inter-talker differences. This suggests that the proposed male spectrum could be applied to other languages without introducing large errors.

The impact of the proposed male spectrum was assessed on the STI calculations and found to produce an increase of 0.02–0.06 for seven different voice alarm systems operating under conditions of 10 dBA SNR. Larger STI increases were found for lower SNR conditions whilst STI increases were minimal for quiet conditions.

The greater high-frequency content of the proposed male spectrum compared with the IEC spectrum could allow a reduction in sound system SPL capacity for the same STI values as the IEC spectrum. Reductions in SPL between 3.0 dB and 5.0 dB were found with the seven

voice alarm systems investigated. Lower electrical power requirements of 2.8 dB were also found with the proposed male spectrum to achieve similar A-weighted SPL than with the standard IEC spectrum. The possibility of achieving a specific STI target with a reduced SPL and the significant lower electrical power requirements, both found with the proposed male spectrum, could lead to a cost reduction of voice alarm systems.

To enable the proposed male spectrum to be used with the current STI method, further validation of the current STI for the English language is required with the proposed male spectrum. This would establish the relationship between STI and subjective intelligibility, and also determine whether the standard error of the obtained mathematical expression (standard deviation) is acceptable. An extensive validation of the STI and investigation into its correlation with subjective intelligibility has been undertaken and will be reported in a future paper.

References

[1] Steeneken HJM, Houtgast T. A physical method for measuring speech transmission quality. J Acoust Soc Am 1980;67:318–26.
[2] IEC 60268–16-2011. Sound system equipment. Part 16: Objective rating of speech intelligibility by speech transmission index. International Electrotechnical Commission, Geneva Switzerland.
[3] ANSI S3.5-1986. Method for calculation of the articulation index. American National Standards Institute; New York 1986.
[4] Dunn HK, White SD. Statistical measurements on conversational speech. J Acoust Soc Am 1940;11:278–88.
[5] French NR, Steinberg JC. Factors governing the intelligibility of speech of sounds. J Acoust Soc Am 1947;19:90–119.
[6] ANSI S3.5 (R2007). American National Standard methods for the Calculation of the Speech Intelligibility Index, American National Standards Institute. New York 2007.
[7] Pavlovic CV. Derivation of primary parameters and procedures for use in speech intelligibility predictions. J Acoust Soc Am 1987;82:413–22.
[8] Steeneken, H. J. M. Ph.D. thesis, University of Amsterdam, Amsterdam 1982.
[9] Benson RW, Hirsh IJ. Some variables in audio spectrometry. J Acoust Soc Am 1953;25:499–505.
[10] Stevens SS, Egan JP, Miller GA. Methods of measuring speech spectra. J Acoust Soc Am 1947;19:771–80.
[11] Cox RM, Moore JN. Composite speech spectrum for Hearing aid gain Prescriptions. J Speech Hear Res 1988;31:102–7.
[12] Byrne D, Dillon H. The National Acoustic Laboratories' (NAL) new procedure for selecting the gain and frequency response of a hearing aid. Ear Hear 1986;7:257–65.
[13] Boothroyd A. The discrimination by partially hearing children of frequency distorted speech. Int Audiol 1967;6:136–45.
[14] Tarnoczy T, Fant G. Some remarks on the average speech spectrum. Stockholm: Speech Transmission Laboratory; 1964. Q.P.S.R. Rep. No.4:13–14.
[15] Byrne D. The speech spectrum - Some aspects of its significance for hearing aid selection and evaluation. Br J Audiol 1977;11:40–6.
[16] Kiukaanniemi H, Soponen P, Mattila P. Individual differences in the long-term speech spectrum. Folia Phoniatr 1982;34:21–8.
[17] Harmegnies B, Landercy A. Intra-speaker variability of the long term speech spectrum. Speech Com 1988;1:81–6.
[18] Pavlovic CV, Rossi M, Espesser R. Statistical distribution of speech for various languages. J Acoust Soc Am 1990;88. 176-176.
[19] Studebaker GA. Directivity of the human vocal source in the horizontal plane. Ear Hear 1986;6:315–9.
[20] Pearsons KS, Bennett RL, Fidell S. Speech levels in various noise environment. Washington DC: Environmental Protection Agency; 1977. Rep. No. 600/1-77-025.
[21] Cushing IR, Li FF, Cox TJ, Worrall K, Jackson T. Vocal effort levels in anechoic conditions. Appl Acoust 2011;72:695–701.
[22] Byrne, et al. Long-term average speech spectra. J Acoust Soc Am 1994;96:2108–20.
[23] Bartlett B. Tonal Effects of Close Microphone Placement. J Audio Eng Soc 1981;29:726–38.
[24] Chu WT, Warnock ACC. Detailed directivity of sound fields around human talkers. Canada: Institute for Research in Construction; 2002. Technical Report:1–47.
[25] Brixen EB. Near field registration of the human voice: Spectral changes due to positions Preprint no. 4728 Amsterdam: 104 AES Convention; 1998.
[26] IEEE Subcommittee on Subjective Measurements. IEEE Recommended Practices for Speech Quality Measurements. IEEE Transactions on Audio and Electroacoustics

1969. Vol 17: 227-246.

[27] WinMLS Morset Sound Development, Norway. http://www.winmls.com.

[28] Welch P. The use of the fast Fourier transform for the estimation of power spectra: A method based on time averaging over short, modified periodograms. IEEE Trans Audio Electroacoust 1967;AU-15(2):70–3.

[29] Monson BB, Hunter EJ, Story BH. Horizontal directivity of low- and high-frequency energy in speech and singing. J. Acoust Soc Am 2012;132:433–41.

[30] http://www.catt.se.

[31] UEFA Guide to quality Stadiums. (UEFA 2012).

[32] Egan MD. Architectural Acoustics. New York: McGraw-Hill Inc; 1988.

[33] Kuttruff H. Room Acoustics. 5th ed. Oxon, UK: Spon Press; 2009.

[34] BS 5839–8.Fire detection and fire alarm systems for buildings – Code of practice for the design, Installation, commissioning and maintenance of voice alarm systems. London 2008, UK.

[35] Stacey A. Comparison of STIPA and sweept sine wave STI measurement methods and results. Proc IOA 2006; vol. 28 Part 8.

[36] http://www.jblpro.com.

[37] BS 7827.Code of practice for designing, specifying, maintaining and operating emergency sound systems at sports venues. London 2011, UK.