Accuracy of speech transmission index predictions based on the reverberation time and signal-to-noise ratio

Citation for published version:
Galbrun, L & Kitapci, K 2014, 'Accuracy of speech transmission index predictions based on the reverberation time and signal-to-noise ratio', *Applied Acoustics*, vol. 81, pp. 1-14.
https://doi.org/10.1016/j.apacoust.2014.02.001

Digital Object Identifier (DOI):
10.1016/j.apacoust.2014.02.001

Published In:
Applied Acoustics

Accuracy of speech transmission index predictions based on the reverberation time and signal-to-noise ratio

Laurent Galbrun* and Kivanc Kitapci

School of the Built Environment, Heriot-Watt University, Edinburgh, EH14 4AS, UK

Abstract

This paper examines the accuracy of the speech transmission index (STI) calculated from the reverberation time (T) and signal-to-noise ratio (L_{SN}) of enclosed spaces. Differences between measured and predicted STIs have been analysed in two rooms (reverberant vs. absorbent), for a wide range of absorption conditions and signal-to-noise ratios (sixteen tests). The STI was measured using maximum length sequence analysis and predictions were calculated using either measured or predicted values of T and L_{SN}, the latter assuming diffuse sound field conditions. The results obtained for all the conditions tested showed that STI predictions based on T and L_{SN} tend to underestimate the STI, with differences between measured and predicted STIs always lower than 0.1 (on a 0.0 – 1.0 scale), and on average lower than 0.06. According to previous research, these differences are noticeable and therefore non-negligible, as 0.03 is the just noticeable difference in STI. The use of either measured or predicted values of T and L_{SN} provided similar STI predictions (i.e. non-noticeable changes), with differences between predictions that are on average lower than 0.03 for the absorbent room, and lower than 0.01 for the reverberant room.

1. Introduction

Speech intelligibility is a sound quality descriptor that can be used to analyse the suitability of spaces where speech is crucial (e.g. teaching rooms, meeting and conference rooms). The speech intelligibility properties of a space can be quantified either through listening tests or through physical measurements. The latter method is normally used in room acoustics, as it is an objective measurement and is also much faster than listening tests which are subject based. In particular, the speech transmission index (STI) is commonly used to measure the speech intelligibility of enclosures. The STI is an electronic method which was developed by Houtgast and Steeneken [1] and which normally requires specialist equipment or software to calculate the modulation transfer function (MTF). The MTF forms the basis of the STI method and is typically determined from impulse responses [2], but can also be estimated from the reverberation time and signal-to-noise ratio present in the space [3]. However, the accuracy of this simple acoustic method is not documented in the literature, and is therefore examined in this paper.

The importance of being able to quantify the STI from simple room acoustic parameters lies in the fact that this allows determining a fundamental design parameter without the need of specialist equipment or software (e.g. maximum length sequence software or ray tracing software), as a simple spreadsheet can be used. This method can therefore be used by non-specialists for design purposes or acoustic assessments. Its accuracy needs however to be known, in order to define its applicability and limitations. This is achieved in the current study by comparing STI values obtained from the impulse response method based on maximum length sequence analysis [2] and for which accuracy is known, with STI values calculated from the reverberation time and signal-to-noise ratio [3]. Two rooms have been tested under sixteen different acoustic conditions (different reverberation times and signal-to-noise ratios), allowing to examine a wide range of STI values (0.1 – 0.8) and carrying out a detailed analysis.

The paper begins with the background theory to the STI and a description of the methodology used. Results are then presented and analysed, followed by conclusions where the main findings are summarised.

2. Background theory

In the 1970s, Houtgast and Steeneken [1] developed an electronic method which has since then been used to measure speech intelligibility. In this method, speech is modelled as modulated bands of noise, and the signal's distortion between the source and receiver can be quantified using the modulation transfer function (MTF). The MTF effectively mimics the behaviour of speech and quantifies the reduction in modulation between the source and receiver using the modulation reduction factor $m(f_m)$, where f_m is the modulation frequency. This method is well established and its details can be found in the literature [1,4-5].

In practice, the MTF can be calculated by producing an electronic signal and measuring the reduction in modulation of the signal at a receiving position; this is normally achieved through the use of the impulse response of the signal. A common method used to calculate the MTF from impulse responses, and which is used in this study, is to

---

* Corresponding author. Tel.: +44 131 4513145
E-mail address: L.G.U.Galbrun@hw.ac.uk

apply maximum length sequence analysis [2]. Maximum length sequences are binary periodic sequences which correlate with impulses; the accuracy of maximum length sequence methods has been demonstrated [2,6], and Rife [2] provided a calculation example in which the STI value was found to be within 0.6% of the exact STI value. This method is applied here through the use of the MLSSA software (Maximum Length Sequence System Analyzer, DRA Laboratories (Sarasota, USA)).

Alongside signal processing analysis, Houtgast and Steeneken [3] have shown that the MTF can be calculated from the room acoustic properties of a space, and more specifically from the reverberation time and signal-to-noise ratio present in the space, using the equation

$$m(f_m) = \frac{1}{\sqrt{1+\left(2\pi f_m \frac{T}{13.8}\right)^2}} \times \frac{1}{1+10^{-0.1L_{SN}}} \tag{1}$$

where $m(f_m)$ is the modulation reduction factor, L_{SN} is the signal-to-noise level (dB), f_m is the modulation frequency (Hz) and T is the room's reverberation time (s). Equation (1) shows that simple room acoustic parameters are sufficient for calculating the MTF, thus removing the need for using impulse responses. The comparison of this method with impulse response methods, as well as its accuracy, is however not documented in the literature.

Regardless of the MTF method used, speech intelligibility results are expressed using the speech transmission index (STI) [4]. The STI is calculated from the modulation reduction factor $m(f_m)$, with f_m ranging from 0.63 to12.5 Hz in 1/3 octave intervals, and each $m(f_m)$ is calculated for octave bands from 125 Hz to 8 kHz. To obtain the STI, the apparent signal-to-noise ratio, L_{SNapp} (dB), should first be calculated from

$$L_{SNapp} = 10log\frac{m(f_m)}{1-m(f_m)} \tag{2}$$

where L_{SNapp} is the signal-to-noise ratio that would have produced the modulation reduction factor $m(f_m)$, had all the distortion been caused by interfering noise [5]. L_{SNapp} is then averaged over all modulation frequencies for each octave band frequency (125 Hz – 8 kHz) to give seven average L_{SNapp} values. These average L_{SNapp} values are then summed to give a single weighted average apparent signal-to-noise ratio, $\overline{L_{SNapp}}$, (dB) according to

$$\overline{L_{SNapp}} = \sum_{i=1}^{7} w_i\left(L_{SNapp}\right) \tag{3}$$

where w_i is the weighting used for octave bands from 125 Hz to 8 kHz (= 0.13, 0.14, 0.11, 0.12, 0.19, 0.17, 0.14) [5]. Lastly, the STI can be calculated from this single $\overline{L_{SNapp}}$ using the formula [5]

$$STI = \left(\overline{L_{SNapp}} + 15\right)/30 \tag{4}$$

noting that STI = 1 when $\overline{L_{SNapp}} \geq 15$ dB and STI = 0 when $\overline{L_{SNapp}} \leq -15$ dB.

The results presented in this paper are based on this STI calculation procedure, and further details about the methodology applied are described in the following section.

3. Methodology

Two medium sized rooms were selected within the School of the Built Environment of Heriot-Watt University (material properties listed in Table 1). One was an empty laboratory chamber (Fig. 1) with reflective surfaces and no windows (named 'reverberant room' from here forward), and the other one was a teaching room (Fig. 2), with twenty seats, furniture and absorbent surfaces (named 'absorbent room'). These two spaces allowed examining different absorption conditions, from highly reverberant to highly absorbent, for dimensions representative of spaces where the intelligibility of unamplified speech can be essential (e.g. meeting rooms and teaching rooms). Furthermore, the use of twelve foam absorption panels (Auralex 2 inch wedges (Indianapolis, USA) of dimensions 1.2 m × 0.6 m × 0.025

Table 1
Materials and furniture present in the reverberant and absorbent rooms.

	Reverberant room	Absorbent room
Walls	Brickwork, plasterboard	Concrete blocks, single glazed windows, whiteboards, screen, pin board
Floor	Concrete	Thin carpet
Ceiling	Concrete	Suspended ceiling with mineral fibre tiles
Furniture	None	Desks (veneered chipboard), upholstered chairs

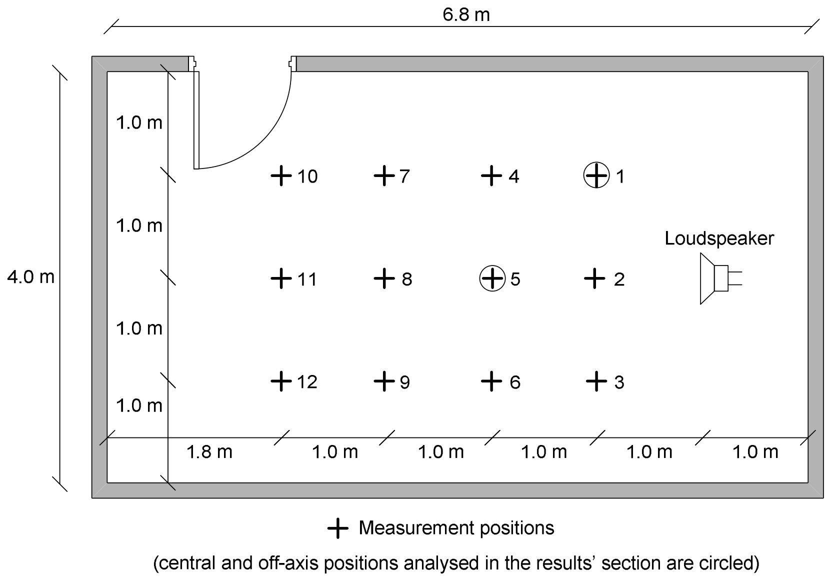


Fig. 1. Floor plan of the reverberant room showing the sound source and 12 receiver positions tested.

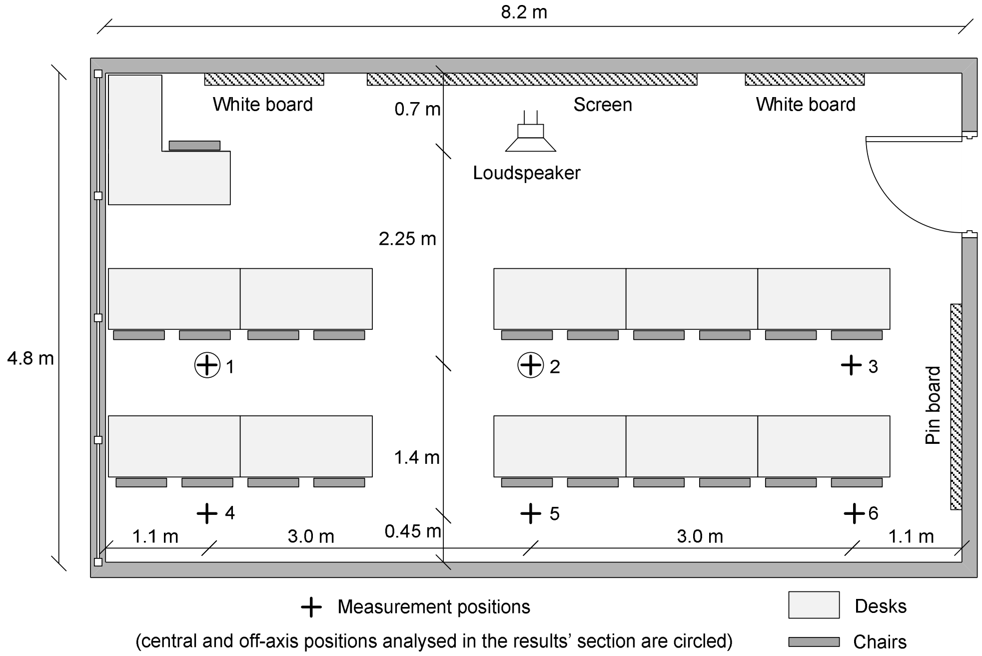


Fig. 2. Floor plan of the absorbent room showing the sound source and 6 receiver positions tested.

m) allowed testing an additional reverberation time per room. When used, the panels were distributed uniformly across the rooms' walls, with their largest dimension placed horizontally, and with a centre panel height of approximately 1.2 m.

The combination of different reverberation time conditions and signal-to-noise ratios (sixteen different acoustic conditions) allowed testing a wide range of STI values (0.1 – 0.8). To illustrate this, the impact that the reverberation time, T, and signal-to-noise ratio, L_{SN}, have on the STI is shown in Fig. 3 (data based on constant values of T and L_{SN}, i.e. the data is not frequency dependent). The curves show that the STI is constant (equal to 0) for L_{SN} values lower than -15 dB, and also tends to be independent from the signal-to-noise level for L_{SN} values greater than 20 dB (with the exception of T values lower than 0.4 s). However, large variations in STI occur for -10 dB ≤ L_{SN} ≤ +10 dB. The curves also indicate that the STI is more sensitive to changes in T occurring at low reverberation times. These observations justify the use of the reverberation times and signal-to-noise levels selected for the study, which are outlined below.

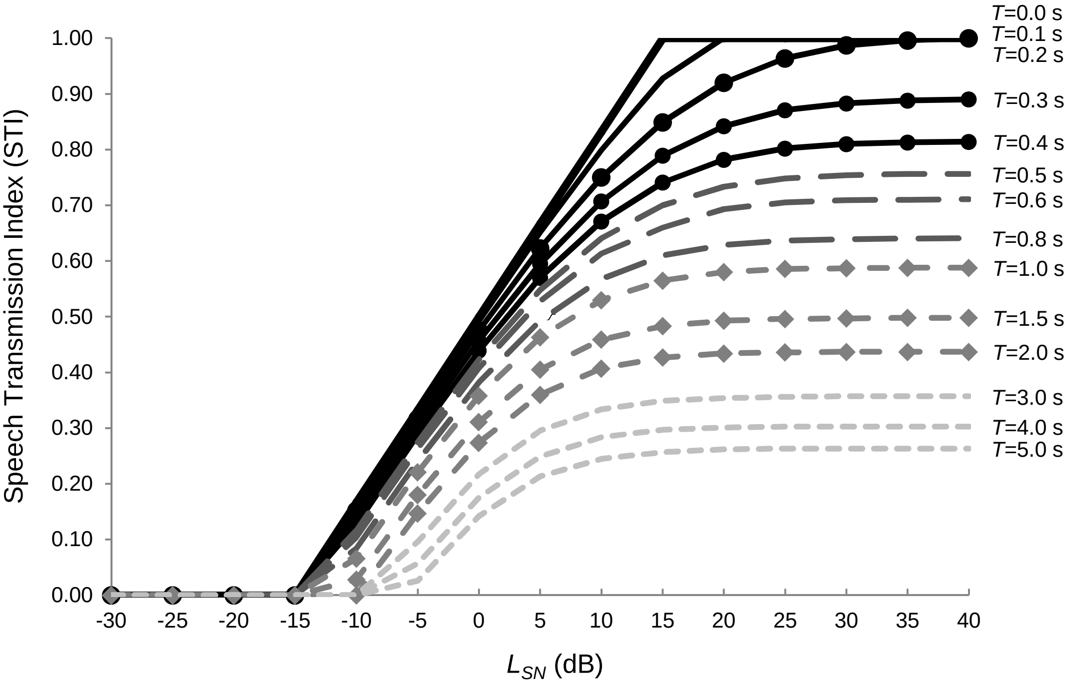


Fig. 3. Speech transmission index variation with signal-to-noise ratio L_{SN} and reverberation time T.

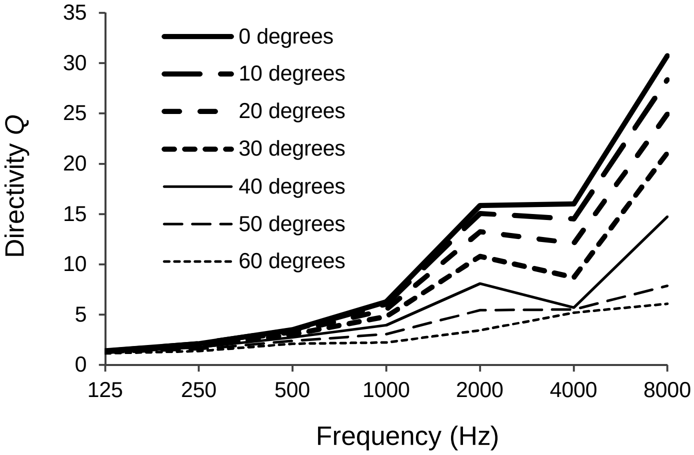


Fig. 4. Horizontal directivity Q of the loudspeaker used during the tests (KEF Coda III), for the range 0-60 degrees, where the reference of 0 degrees corresponds to positions on-axis with the loudspeaker.

The reverberant room had a mid-frequency reverberation time (average of 500 Hz, 1 kHz and 2 kHz octave bands) of 3.1 s when empty, and 1.2 s with the absorption panels; the absorbent room had a mid-frequency reverberation time of 0.5 s and 0.4 s without and with the absorption panels respectively.

The use of electronically generated white noise made it possible to vary the background noise, i.e. the signal-to-noise ratio. The white noise was output from the high quality loudspeaker KEF Coda III (Maidstone, UK) used to generate the speech signal, thus ensuring a constant signal-to-noise ratio across the room. Four signal-to-noise ratios were tested: high (no white noise used), + 10 dB, 0 dB and -10 dB. In all measurements, the signal level was calibrated as 60 dBA at 1 m on-axis from the centre of the loudspeaker [7], which was positioned at a propagating height of 1.2 m (mid-way between the woofer and tweeter). Similarly, the receiver (microphone) was always placed at a height of 1.2 m, which is representative of the floor to ears' height for an adult seated. Measurements were undertaken at 12 receiving positions for the reverberant room (Fig. 1), and 6 receiving positions for the absorbent room (Fig. 2). The reason for having less positions examined in the absorbent room followed a simple observation: after measurements had been carried out in the reverberant room, it was found that differences between measured and predicted STIs showed little variations at different positions; the use of a large number of receiving positions was therefore found to be unnecessary, which is why only 6 positions (instead of 12) were examined in the absorbent room which was tested afterwards.

In the results' section, data is presented in detail only for one central and one off-axis position (circled in Fig. 1 and Fig. 2), and as an average taken across all measurement positions (refer to the results' section for the justification of this approach).

The software MLSSA (Maximum Length Sequence System Analyzer, DRA Laboratories (Sarasota, USA)) was used to measure the STI from impulse responses (refer to section 2 for further details about this method). The computer used to run MLSSA was connected via its sound card to the loudspeaker KEF Coda III and to a half inch microphone Brüel & Kjaer Type 4190 (Naerum, Denmark), which was in turn connected to a microphone power supply Brüel & Kjaer Type 2804. The sound pressure level of the speech signal output from MLSSA and the background noise level were measured using an integrating-averaging sound level meter Brüel & Kjaer Type 2250, with a data averaging period of 20 s. The reverberation time was measured two times and then averaged at each receiving position using a Norsonic sound measuring system Type 823 (Lierskogen, Norway), with the loudspeaker KEF Coda III used as the sound source. All measurements were undertaken in octave bands for the frequency range 125 Hz – 8000 Hz.

Three predictions (1, 2 and 3) were calculated from T and L_{SN} using equations (1)-(4). Prediction 1 was obtained from the measured reverberation time, T, and the measured signal-to-noise ratio, L_{SN}. On the other hand, Predictions 2 and 3 were based on predicted values of T and L_{SN}. For Prediction 2, T was calculated from Sabine's formula

$$T = \frac{0.161V}{A} = \frac{0.161V}{\sum S\alpha + A_{air}} \tag{5}$$

where V is the volume of the room (m^3), A is the absorption (m^2), S is the area of each surface (m^2), α is the absorption coefficient of each surface and A_{air} is the air absorption (m^2). The absorption coefficients were obtained from published data of common materials [8-11] and from the Auralex panels' data [12]. The air absorption, A_{air} (m^2), was included in the calculation and was obtained from [13]

$$A_{air} = \frac{4\Delta L_{air} V}{4.34} \tag{6}$$

where ΔL_{air} is the air attenuation in dB/m, which was taken from tables of the standard ISO 9613-1 [14], for a temperature of 20°C and a relative humidity of 50%. Prediction 3 was based on a different calculation of T: for conditions without the absorption panels, the measured reverberation time was used, whilst for conditions with the added panels, the reverberation time was based on the absorption measured for the room without panels, A, and the added absorption provided by the panels', $\delta A_p = S_p \times \alpha_p$ (where S_p is the total surface covered by the panels and α_p is the absorption coefficient of the panels), the reverberation time being then equal to

$$T = \frac{0.161V}{A + \delta A_p} \tag{7}$$

Prediction 3 effectively represents a case where measurements have been undertaken in a room, and predictions are needed to check the amount by which added absorption can increase speech intelligibility within that space. This is a real case scenario which can often be encountered in practice, which is why Prediction 3 was included in the analysis.

The signal-to-noise ratio, L_{SN}, also needed to be predicted for Predictions 2 and 3. This was done by calculating the sound pressure level of the source, L_S, and subtracting from it the background noise level measured when the source was turned off, L_N , to give L_{SN}. In order to calculate L_S, the sound power level output from the speaker, L_W, had to be known. This was derived from the sound pressure level measured 1 m in front of the speaker, using the classical room acoustic equation [15]

$$L_W = L_S(1\text{m}) - 10log\left(\frac{Q}{4\pi} + \frac{4(1-\bar{\alpha})}{A}\right) \tag{8}$$

where L_S(1m) is the sound pressure level measured on-axis at 1 m from the source (dB), Q is the directivity of the source, $\overline{\alpha}$ is the average absorption coefficient of the room and A is the room absorption (m^2). The absorption of the room was calculated from the measured reverberation time (equation (5)) and $\overline{\alpha}$ was derived from the equation $\overline{\alpha} = A/S_{total}$, where S_{total} is the total surface area of the room. Using the value found for L_W, L_S could be calculated at any receiving position from [15]

$$L_S(r) = L_W + 10log\left(\frac{Q}{4\pi r^2} + \frac{4(1-\bar{\alpha})}{A}\right) \tag{9}$$

where r is the distance between the source and receiver (m), and all the other parameters are identical to those previously defined for equation (8). It is important to note that equations (8) and (9) assume a diffuse reverberant sound field, a condition which is not applicable to rooms with low reverberation times. These equations were nevertheless used for their simplicity, as other methods such as ray tracing and empirical models could not be considered as simple methods for predicting the STI, and would have therefore defeated the purpose of the present study which is aimed at quantifying the accuracy of a simple STI prediction method.

It is also important to point out that Predictions 2 and 3 assume a known power output. Whilst this was easily calculated in this study by measuring the MLSSA speech spectrum output from the loudspeaker (L_S(1m)), in practice this represents a complication which can render the calculation of Predictions 2 and 3 unpractical. However, the power output from a talker can be derived from a known speech spectrum measured at 1 m from a talker, which can for example be taken from the standard ANSI S3.5 [16], and be used as L_S(1m) in equation (8) (where the term including absorption should be ignored, if L_S(1m) was measured in a free-field).

The source directivity, Q, of equations (8) and (9) represents another input which might not be readily available. For the loudspeaker used in this study, Q was measured in the anechoic chamber of Heriot-Watt University, and all the results presented in this paper use this measured data (Fig. 4). However, a value of Q = 1 (spherical radiation) might be used for simplicity, although this affects the accuracy of the signal level predicted (and hence the STI), as pointed out in the discussion of results (end of section 5.2).

All of the above indicates that, as long as T and L_{SN} can either be measured in situ or predicted, the STI prediction method based on room acoustic properties can be applied. The accuracy of results obtained from this method is presented in the following section.

4. Results

Three sets of STI results are presented in this section for the reverberant and absorbent rooms: 1) Results obtained at a central position within the rooms (Fig. 5 and Tables 2 and 3); 2) Results obtained at an off-axis position within the rooms (Fig. 6 and Tables 4 and 5); 3) Results obtained by averaging the STI across all measurement positions within the room (Fig. 7 and Tables 6 and 7). Data is not presented for every position, considering the relatively low variability in STI differences between measurements and predictions observed across the rooms (Table 8). Only positions 1 (off-axis) and 5 (central) of Fig. 1 are discussed in detail for the reverberant room, and positions 1 (off-axis) and 2 (central) of Fig. 2 for the absorbent room. The off-axis positions discussed correspond to a horizontal angle of 45 degrees for the reverberant room and 53 degree for the absorbent room, and are therefore comparable (the vertical angle being zero for all positions).

The three sets of results allow examining on-axis intelligibility, off-axis intelligibility, as well as intelligibility representative of the overall rooms' properties and including both on-axis and off-axis propagation of speech.

It is also important to note that MLSSA measurements showed a maximum change in the STI of around ±0.001 when measurements were repeated several times, demonstrating the reliability of the STI measurements and justifying their use as reference data in the comparisons with prediction data.

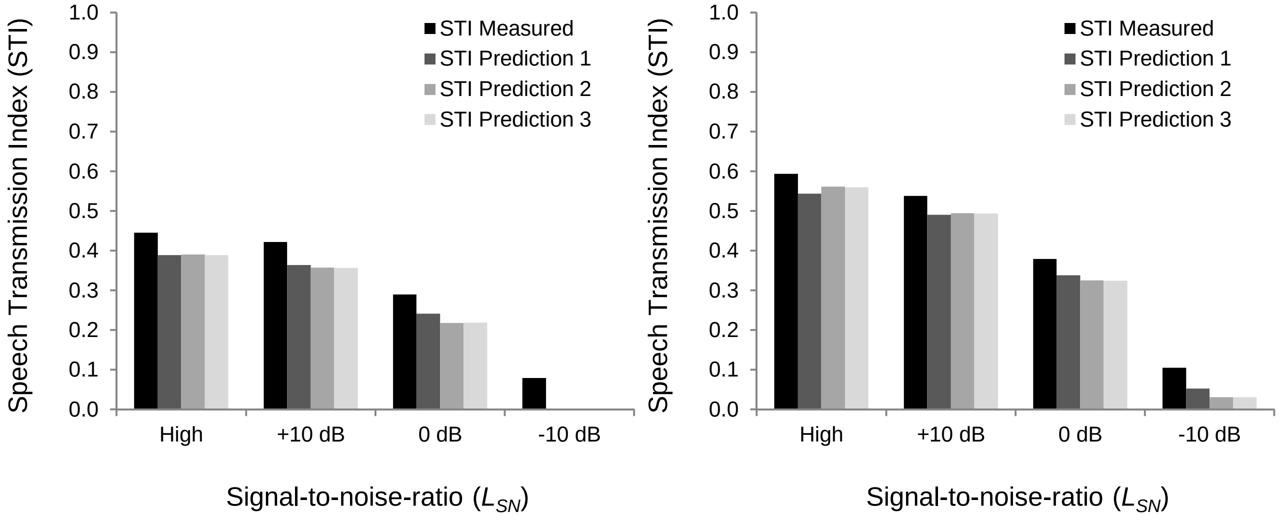
The errors presented in this section are shown as STI differences (0.0-1.0 scale) and no percentages are given, as these were found to be irrelevant for this scale (i.e. larger percentages obtained at lower STI values are not representative of lower accuracy).

4.1. Reverberant room

At the central position (Fig. 5(a)-(b)), differences between the STI measured and STI predictions are on average equal to 0.058 (Table 2(e)), with all differences within the range +0.032 – +0.079 (Table 2(a)-(d)). Predictions are always lower than measured results, and differences between the predictions are on average equal to 0.008 (Table 3(e)), and all within the range -0.023 – +0.018 (Table 3(a)-(d)).

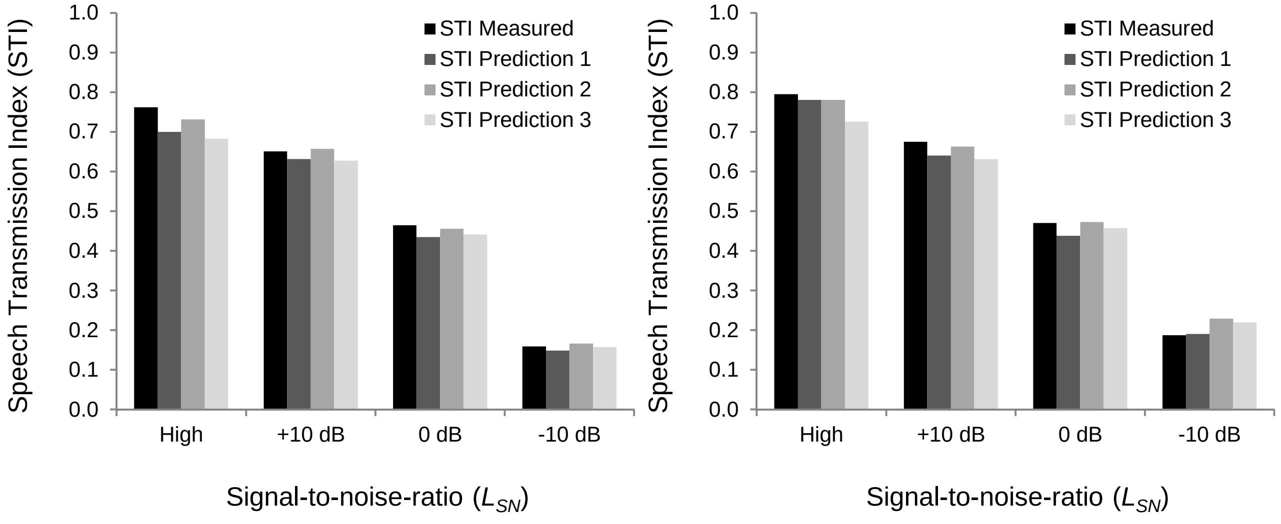
At the off-axis position (Fig. 6(a)-(b)), differences between the STI measured and STI predictions are on average equal to 0.042 (Table 4(e)), with all differences within the range +0.016 – +0.073 (Table 4(a)-(d)). Predictions are always lower than measured results, and differences between the predictions are on average equal to 0.007 (Table 5(e)), and all within the range -0.003 – +0.021 (Table 5(a)-(d)).

For the average taken across all positions (Fig. 7(a)-(b)), differences between the STI measured and STI predictions are on average equal to 0.050 (Table 6(e)), with all differences within the range +0.026 – +0.078 (Table 6(a)-(d)). Predictions are always lower than the measured results, and differences between the predictions are on average equal to 0.005 (Table 7(e)), with all differences within the range -0.015 – +0.017 (Table 7(a)-(d)).



(a) Reverberant room without panels

(b) Reverberant room with panels



(c) Absorbent room without panels

(d) Absorbent room with panels

Fig. 5. Speech transmission index (STI) measured and predicted at a central position in the reverberant and absorbent rooms.

4.2. Absorbent room

At the central position (Fig. 5(c)-(d)), differences between the STI measured and STI predictions are on average equal to 0.026 (Table 2(e)), with all differences within the range -0.042 – +0.079 (Table 2(a)-(d)). Predictions tend to be lower than the measured results, the most notable exception being the predictions for tests with panels and a signal-to-noise ratio of -10 dB (Fig. 5(d) and Table 2(d)). Differences between the predictions are on average equal to 0.023 (Table 3(e)), and all within the range -0.054 – +0.055 (Table 3(a)-(d)).

At the off-axis position (Fig. 6(c)-(d)), differences between the STI measured and STI predictions are on average equal to 0.041 (Table 4(e)), with all differences within the range -0.053 – +0.093 (Table 4(a)-(d)). Predictions tend to be lower than measured results, with some exceptions observed at a high signal-to-noise ratio and L_{SN} = +10 dB. Differences between the predictions are on average equal to 0.030 (Table 5(e)), and all within the range -0.063 – +0.045 (Table 5(a)-(d)).

For the average taken across all positions (Fig. 7(c)-(d)), differences between the STI measured and STI predictions are on average equal to 0.030 (Table 6(e)), with all differences within the range -0.028 – +0.061 (Table 6(a)-(d)). Predictions are always lower than the measured results, with some exceptions observed for tests with a high signal-to-noise ratio (Fig. 7(c)-(d) and Table 6(a)). Differences between the predictions are on average equal to 0.021 (Table 7(e)), and all within the range -0.045 – +0.047 (Table 7(a)-(d)).

Table 2

Differences between STI values obtained from measurements and predictions obtained from room acoustic data. Reverberant and absorbent rooms' results obtained at a central position, for different signal to noise levels.

⬜	0.000 – 0.020
░	0.021 – 0.040
▒	0.041 – 0.060
▓	0.061 – 0.080
█	0.081 – 0.100

Room condition	STI difference between measurements and predictions					
	Reverberant room			Absorbent room		
	Pred. 1	Pred. 2	Pred. 3	Pred. 1	Pred. 2	Pred. 3
No panels	0.056	0.055	0.056	0.062	0.030	0.079
With panels	0.050	0.032	0.034	0.014	0.014	0.069

(a) High L_{SN} (no artificial noise)

Room condition	STI difference between measurements and predictions					
	Reverberant room			Absorbent room		
	Pred. 1	Pred. 2	Pred. 3	Pred. 1	Pred. 2	Pred. 3
No panels	0.058	0.064	0.065	0.019	-0.006	0.023
With panels	0.048	0.043	0.044	0.035	0.012	0.043

(b) L_{SN} = +10 dB

Room condition	STI difference between measurements and predictions					
	Reverberant room			Absorbent room		
	Pred. 1	Pred. 2	Pred. 3	Pred. 1	Pred. 2	Pred. 3
No panels	0.049	0.072	0.072	0.030	0.009	0.024
With panels	0.041	0.054	0.055	0.032	-0.003	0.012

(c) L_{SN} = 0 dB

Room condition	STI difference between measurements and predictions					
	Reverberant room			Absorbent room		
	Pred. 1	Pred. 2	Pred. 3	Pred. 1	Pred. 2	Pred. 3
No panels	0.079	0.079	0.079	0.011	-0.008	0.002
With panels	0.052	0.074	0.074	-0.004	-0.042	-0.032

(d) L_{SN} = -10 dB

Room condition	STI difference between measurements and predictions					
	Reverberant room			Absorbent room		
	Pred. 1	Pred. 2	Pred. 3	Pred. 1	Pred. 2	Pred. 3
No panels	0.060	0.067	0.068	0.030	0.013 *0.006*	0.032
With panels	0.048	0.051	0.052	0.021 *0.019*	0.018 *-0.005*	0.039 *0.023*
Overall	0.058			0.026 *0.018*		

(e) Average of L_{SN} conditions (a)-(d). Average calculated from absolute values, with true average shown in italic when different.

Table 3

Differences between STI values obtained from room acoustic predictions 1, 2 and 3. Reverberant and absorbent rooms' results obtained at a central position, for different signal to noise levels.

□	0.000 – 0.020
░	0.021 – 0.040
▒	0.041 – 0.060
▓	0.061 – 0.080
█	0.081 – 0.100

Room condition	STI difference between predictions					
	Reverberant room			Absorbent room		
	Pred. 2-1	Pred. 3-1	Pred. 2-3	Pred. 2-1	Pred. 3-1	Pred. 2-3
No panels	0.002	0.000	0.002	0.031	-0.017	0.049
With panels	0.018	0.016	0.002	0.000	-0.054	0.055

(a) High L_{SN} (no artificial noise)

Room condition	STI difference between predictions					
	Reverberant room			Absorbent room		
	Pred. 2-1	Pred. 3-1	Pred. 2-3	Pred. 2-1	Pred. 3-1	Pred. 2-3
No panels	-0.006	-0.007	0.001	0.025	-0.004	0.030
With panels	0.004	0.003	0.001	0.022	-0.009	0.031

(b) L_{SN} = +10 dB

Room condition	STI difference between predictions					
	Reverberant room			Absorbent room		
	Pred. 2-1	Pred. 3-1	Pred. 2-3	Pred. 2-1	Pred. 3-1	Pred. 2-3
No panels	-0.023	-0.023	0.000	0.021	0.006	0.015
With panels	-0.013	-0.014	0.001	0.035	0.019	0.016

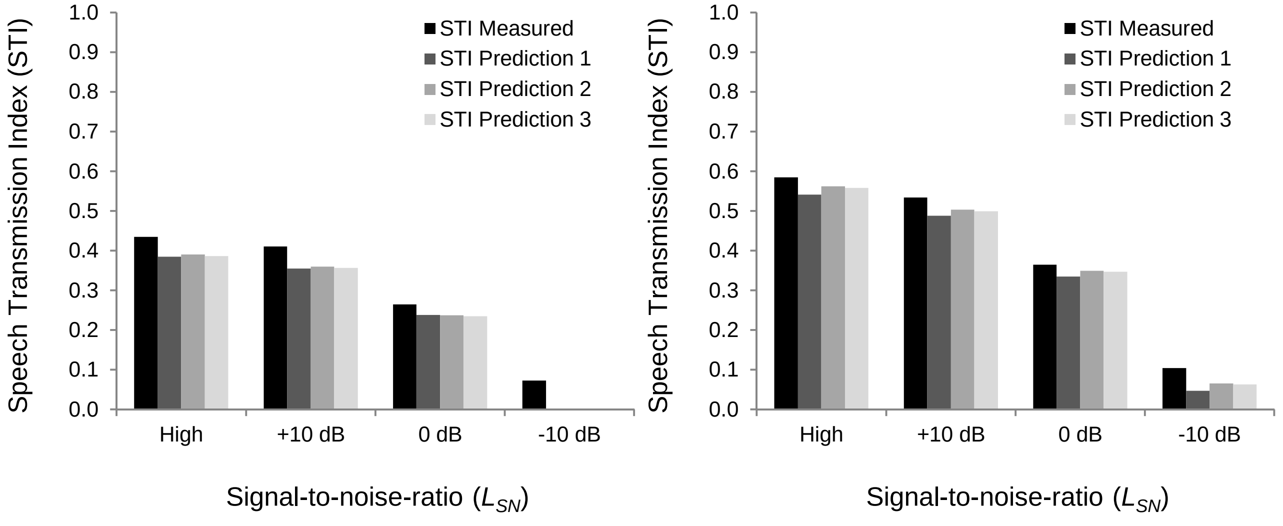
(c) L_{SN} = 0 dB

Room condition	STI difference between predictions					
	Reverberant room			Absorbent room		
	Pred. 2-1	Pred. 3-1	Pred. 2-3	Pred. 2-1	Pred. 3-1	Pred. 2-3
No panels	0.000	0.000	0.000	0.018	0.009	0.010
With panels	-0.022	-0.022	0.000	0.039	0.029	0.010

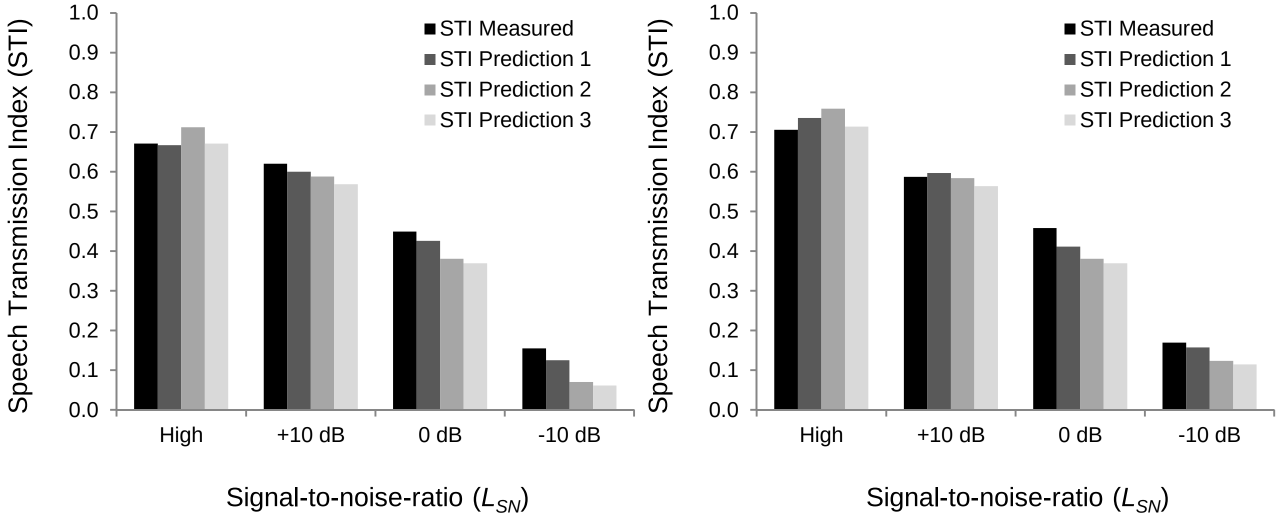
(d) L_{SN} = -10 dB

Room condition	STI difference between predictions					
	Reverberant room			Absorbent room		
	Pred. 2-1	Pred. 3-1	Pred. 2-3	Pred. 2-1	Pred. 3-1	Pred. 2-3
No panels	0.008 *-0.007*	0.007 *-0.007*	0.001 *0.000*	0.024 *-0.002*	0.009	0.026
With panels	0.014 *-0.003*	0.014 *-0.004*	0.001	0.024 *-0.004*	0.028	0.028
Overall	0.008 *-0.003*			0.023 *0.016*		

(e) Average of L_{SN} conditions (a)-(d). Average calculated from absolute values, with true average shown in italic when different.



(a) Reverberant room without panels (b) Reverberant room with panels



(c) Absorbent room without panels (d) Absorbent room with panels

Fig. 6. Speech transmission index (STI) measured and predicted at an off-axis position in the reverberant and absorbent rooms.

5. Discussion

5.1. Main findings

Differences in STI between measured results and simple predictions based on T and L_{SN} are always lower than 0.1 (on a 0.0-1.0 scale), and on average always lower than 0.06. These differences are noticeable and therefore non-negligible, as a change in STI of 0.03 has been demonstrated as a just noticeable difference by Bradley et al. [17].

The prediction method based on simple acoustic parameters tends to underestimate the STI. This is observed both for central positions (with only one notable exception out of sixteen tests in Fig. 5 and Table 2), as well as off-axis positions (with only two notable exceptions out of sixteen tests in Fig. 6 and Table 4). This underestimation is not due to the exclusion of the direct field in equation (1), as its inclusion provides negligible increases in STI, and only for positions very close to the source (+0.003 for position 2 in the reverberant room, with predictions obtained by using the $m(f_m)$ formula given by Long [9]). Inaccuracies in the reverberation time are also not expected to be the cause for this, as the reliability of the reverberation time data used was confirmed by repeating measurements using different techniques and equipment. There is then no obvious reason why this underestimation occurs.

Table 4

Differences between STI values obtained from measurements and predictions obtained from room acoustic data. Reverberant and absorbent rooms’ results obtained at an off-axis position, for different signal to noise levels.

□	0.000 – 0.020
░	0.021 – 0.040
▒	0.041 – 0.060
▓	0.061 – 0.080
█	0.081 – 0.100

Room condition	STI difference between measurements and predictions					
	Reverberant room			Absorbent room		
	Pred. 1	Pred. 2	Pred. 3	Pred. 1	Pred. 2	Pred. 3
No panels	0.050	0.044	0.049	0.004	-0.042	0.000
With panels	0.043	0.022	0.027	-0.030	-0.053	-0.008

(a) High L_{SN} (no artificial noise)

Room condition	STI difference between measurements and predictions					
	Reverberant room			Absorbent room		
	Pred. 1	Pred. 2	Pred. 3	Pred. 1	Pred. 2	Pred. 3
No panels	0.056	0.051	0.054	0.020	0.032	0.051
With panels	0.046	0.031	0.034	-0.010	0.003	0.023

(b) $L_{SN} = +10$ dB

Room condition	STI difference between measurements and predictions					
	Reverberant room			Absorbent room		
	Pred. 1	Pred. 2	Pred. 3	Pred. 1	Pred. 2	Pred. 3
No panels	0.027	0.028	0.030	0.023	0.068	0.079
With panels	0.030	0.016	0.018	0.047	0.077	0.089

(c) $L_{SN} = 0$ dB

Room condition	STI difference between measurements and predictions					
	Reverberant room			Absorbent room		
	Pred. 1	Pred. 2	Pred. 3	Pred. 1	Pred. 2	Pred. 3
No panels	0.073	0.073	0.073	0.030	0.084	0.093
With panels	0.057	0.039	0.041	0.013	0.046	0.055

(d) $L_{SN} = -10$ dB

Room condition	STI difference between measurements and predictions					
	Reverberant room			Absorbent room		
	Pred. 1	Pred. 2	Pred. 3	Pred. 1	Pred. 2	Pred. 3
No panels	0.051	0.049	0.052	0.019	0.056 *0.036*	0.056
With panels	0.044	0.027	0.030	0.025 *0.005*	0.045 *0.018*	0.044 *0.040*
Overall	0.042			0.041 *0.029*		

(e) Average of L_{SN} conditions (a)-(d). Average calculated from absolute values, with true average shown in italic when different.

Table 5

Differences between STI values obtained from room acoustic predictions 1, 2 and 3. Reverberant and absorbent rooms' results obtained at an off-axis position, for different signal to noise levels.

□	0.000 – 0.020
░	0.021 – 0.040
▒	0.041 – 0.060
▓	0.061 – 0.080
█	0.081 – 0.100

Room condition	STI difference between predictions					
	Reverberant room			Absorbent room		
	Pred. 2-1	Pred. 3-1	Pred. 2-3	Pred. 2-1	Pred. 3-1	Pred. 2-3
No panels	0.006	0.001	0.005	0.045	0.004	0.041
With panels	0.021	0.017	0.005	0.023	-0.022	0.045

(a) High L_{SN} (no artificial noise)

Room condition	STI difference between predictions					
	Reverberant room			Absorbent room		
	Pred. 2-1	Pred. 3-1	Pred. 2-3	Pred. 2-1	Pred. 3-1	Pred. 2-3
No panels	0.005	0.002	0.003	-0.012	-0.031	0.020
With panels	0.015	0.012	0.003	-0.013	-0.033	0.020

(b) L_{SN} = +10 dB

Room condition	STI difference between predictions					
	Reverberant room			Absorbent room		
	Pred. 2-1	Pred. 3-1	Pred. 2-3	Pred. 2-1	Pred. 3-1	Pred. 2-3
No panels	-0.001	-0.003	0.002	-0.046	-0.056	0.011
With panels	0.015	0.012	0.003	-0.031	-0.042	0.011

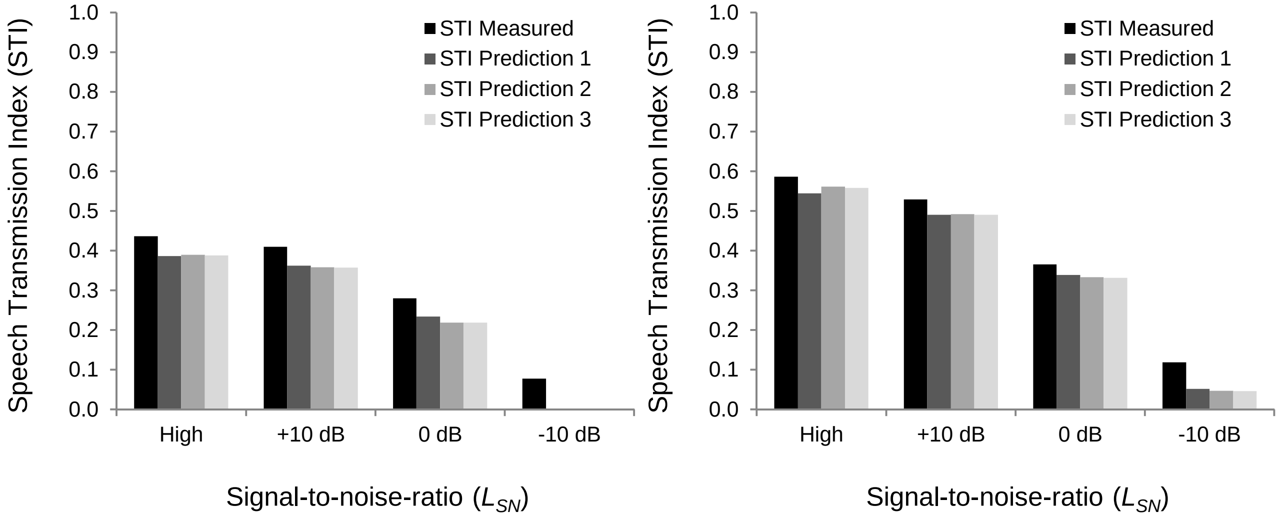
(c) L_{SN} = 0 dB

Room condition	STI difference between predictions					
	Reverberant room			Absorbent room		
	Pred. 2-1	Pred. 3-1	Pred. 2-3	Pred. 2-1	Pred. 3-1	Pred. 2-3
No panels	0.000	0.000	0.000	-0.054	-0.063	0.009
With panels	0.018	0.016	0.002	-0.033	-0.043	0.009

(d) L_{SN} = -10 dB

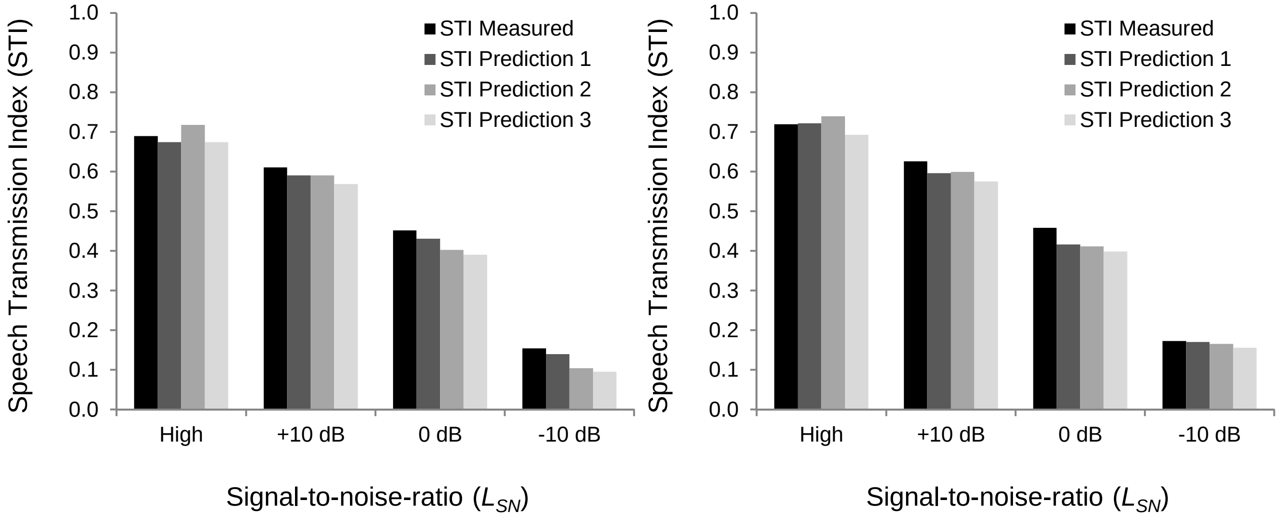
Room condition	STI difference between predictions					
	Reverberant room			Absorbent room		
	Pred. 2-1	Pred. 3-1	Pred. 2-3	Pred. 2-1	Pred. 3-1	Pred. 2-3
No panels	0.003 *0.002*	0.001 *0.000*	0.003	0.039 *-0.017*	0.039 *-0.037*	0.020
With panels	0.017	0.014	0.003	0.025 *-0.014*	0.035	0.021
Overall	0.007			0.030 *-0.010*		

(e) Average of L_{SN} conditions (a)-(d). Average calculated from absolute values, with true average shown in italic when different.



(a) Reverberant room without panels

(b) Reverberant room with panels



(c) Absorbent room without panels

(d) Absorbent room with panels

Fig. 7. Average speech transmission index (STI) measured and predicted in the reverberant and absorbent rooms.

5.2. Differences between predictions

Looking back at Tables 2, 4 and 6, it can be seen that the use of either measured (Prediction 1) or predicted (Prediction 2 and 3) input data has little effect on the accuracy of predictions. Differences between the three predictions examined (Tables 3, 5 and 7) are on average lower than 0.03 for the absorbent room (range: 0.021-0.030), and lower than 0.01 for the reverberant room (range: 0.005 – 0.008); apart from one exception, all these differences are therefore not noticeable, as they are lower than 0.03. The larger variations observed in the absorbent room are related to its poor sound diffuseness, which affects the predicted signal-to-noise ratio and reverberation time. Firstly, the direct sound field is more pronounced in absorbent spaces, so that the uncertainty in its prediction adds to the uncertainty in the prediction of the reverberant field which is otherwise dominant. And secondly, small differences between low reverberation times are more significant in percentage terms and have a larger impact on STI predictions. For example, the predicted reverberation time of the absorbent room without panels is only slightly underestimated by 0.07 s on average, but this corresponds to a non-negligible percentage difference of 13%, which can increase the STI of up to +0.08 according to Fig. 3. For these reasons, comparable inaccuracies in input data result in larger STI differences in absorbent spaces compared to reverberant spaces. This proves that the accuracy of the predicted input data is particularly important in absorbent spaces.

Table 6

Differences between STI values obtained from measurements and predictions obtained from room acoustic data. Reverberant and absorbent rooms' results averaged over all measurement positions, for different signal to noise levels.

□ 0.000 – 0.020
░ 0.021 – 0.040
▒ 0.041 – 0.060
▓ 0.061 – 0.080
█ 0.081 – 0.100

Room condition	STI difference between measurements and predictions					
	Reverberant room			Absorbent room		
	Pred. 1	Pred. 2	Pred. 3	Pred. 1	Pred. 2	Pred. 3
No panels	0.050	0.046	0.048	0.016	-0.028	0.015
With panels	0.042	0.026	0.028	-0.002	-0.020	0.027

(a) High L_{SN} (no artificial noise)

Room condition	STI difference between measurements and predictions					
	Reverberant room			Absorbent room		
	Pred. 1	Pred. 2	Pred. 3	Pred. 1	Pred. 2	Pred. 3
No panels	0.047	0.052	0.053	0.020	0.020	0.042
With panels	0.039	0.037	0.039	0.030	0.027	0.050

(b) L_{SN} = +10 dB

Room condition	STI difference between measurements and predictions					
	Reverberant room			Absorbent room		
	Pred. 1	Pred. 2	Pred. 3	Pred. 1	Pred. 2	Pred. 3
No panels	0.045	0.061	0.061	0.020	0.049	0.061
With panels	0.026	0.032	0.034	0.043	0.047	0.060

(c) L_{SN} = 0 dB

Room condition	STI difference between measurements and predictions					
	Reverberant room			Absorbent room		
	Pred. 1	Pred. 2	Pred. 3	Pred. 1	Pred. 2	Pred. 3
No panels	0.076	0.078	0.078	0.014	0.050	0.059
With panels	0.067	0.072	0.073	0.002	0.007	0.017

(d) L_{SN} = -10 dB

Room condition	STI difference between measurements and predictions					
	Reverberant room			Absorbent room		
	Pred. 1	Pred. 2	Pred. 3	Pred. 1	Pred. 2	Pred. 3
No panels	0.055	0.059	0.060	0.017	0.037 *0.022*	0.044
With panels	0.044	0.042	0.043	0.019 *0.018*	0.025 *0.015*	0.039
Overall	0.050			0.030 *0.026*		

(e) Average of L_{SN} conditions (a)-(d). Average calculated from absolute values, with true average shown in italic when different.

Table 7

Differences between STI values obtained from predictions 1, 2 and 3. Reverberant and absorbent rooms' results averaged over all measurement positions, for different signal to noise levels.

□	0.000 – 0.020
░	0.021 – 0.040
▒	0.041 – 0.060
▓	0.061 – 0.080
█	0.081 – 0.100

Room condition	STI difference between predictions					
	Reverberant room			Absorbent room		
	Pred. 2-1	Pred. 3-1	Pred. 2-3	Pred. 2-1	Pred. 3-1	Pred. 2-3
No panels	0.003	0.001	0.002	0.044	0.001	0.043
With panels	0.017	0.014	0.003	0.018	-0.029	0.047

(a) High L_{SN} (no artificial noise)

Room condition	STI difference between predictions					
	Reverberant room			Absorbent room		
	Pred. 2-1	Pred. 3-1	Pred. 2-3	Pred. 2-1	Pred. 3-1	Pred. 2-3
No panels	-0.004	-0.005	0.001	0.000	-0.022	0.022
With panels	0.002	0.001	0.002	0.003	-0.021	0.024

(b) L_{SN} = +10 dB

Room condition	STI difference between predictions					
	Reverberant room			Absorbent room		
	Pred. 2-1	Pred. 3-1	Pred. 2-3	Pred. 2-1	Pred. 3-1	Pred. 2-3
No panels	-0.015	-0.015	0.000	-0.028	-0.040	0.012
With panels	-0.006	-0.007	0.001	-0.005	-0.017	0.013

(c) L_{SN} = 0 dB

Room condition	STI difference between predictions					
	Reverberant room			Absorbent room		
	Pred. 2-1	Pred. 3-1	Pred. 2-3	Pred. 2-1	Pred. 3-1	Pred. 2-3
No panels	-0.001	-0.001	0.000	-0.036	-0.045	0.009
With panels	-0.005	-0.006	0.001	-0.005	-0.015	0.009

(d) L_{SN} = -10 dB

Room condition	STI difference between predictions					
	Reverberant room			Absorbent room		
	Pred. 2-1	Pred. 3-1	Pred. 2-3	Pred. 2-1	Pred. 3-1	Pred. 2-3
No panels	0.006 *-0.004*	0.006 *-0.005*	0.001	0.027 *-0.005*	0.027 *-0.027*	0.022
With panels	0.007 *0.002*	0.007 *0.000*	0.002	0.008 *0.003*	0.021 *-0.021*	0.023
Overall	0.005 *-0.001*			0.021 *-0.001*		

(e) Average of L_{SN} conditions (a)-(d). Average calculated from absolute values, with true average shown in italic when different.

Results also show that Prediction 2 is always greater than Prediction 3, although differences are not large and on average always lower than 0.028 (Tables 3, 5 and 7). The higher STIs of Prediction 2 are due to the fact that the reverberation time of Prediction 3 is normally higher than the one of Prediction 2 for which T is underestimated (Table 9). As predictions tend to underestimate the STI, these results suggest that the use of measured data for predicting the reverberation time when panels are added (i.e. the method used in Prediction 3), is not a guarantee of greater accuracy for STI predictions, as the underestimations can be larger (i.e. in Prediction 2, an underestimated T can improve predictions by accident, in comparison to Prediction 3). It should however be noted that Predictions 2 and 3 will always be subject to the accuracy of the absorption data used to calculate the reverberation time T; this accuracy was relatively good for the tests carried out in this study (±0.1 s on average for T; Table 9), but larger differences between STI predictions should be expected from poorly predicted values of T. Reliable absorption data is therefore of paramount importance for obtaining reliable STI predictions based on predicted T, and for this reason, predictions based on measured input data (Prediction 1) are more reliable.

Looking back at Fig. 5 and Fig. 6, it can also be noted that Predictions 2 and 3 show different trends in STI for the central and off-axis positions. These differences are particularly marked in the absorbent room, where Predictions 2 and 3 tend to be larger than Prediction 1 at central positions (Fig. 5), whilst the opposite occurs at off-axis positions (Fig. 6). This is due to the large differences observed in the absorbent room for ΔL_{SN} (measured minus predicted signal-to-noise ratio) between central and off-axis positions (Table 10). These ΔL_{SN} differences occur because of the inaccurate assumption of a diffuse sound field in equations (8) and (9): the sound field is very absorbent and the presence of furniture both blocks and reflects sound in the proximity of the receiver positions, two reasons why the sound field cannot be truly diffuse in the absorbent room tested. However, the use of alternative methods for predicting the signal level (e.g. ray tracing or empirical models [15]) would increase the complexity of the method to a point that would defeat the purpose of the current study. And although the values shown in Table 10 indicate large differences in L_{SN}, these inaccuracies appear to be important only for the STI predictions of spaces where the non-diffuseness is quite marked.

Finally, it should be noted that Predictions 2 and 3 were calculated using exact values of the directivity Q. However, if directivity data is not available for the sound source, a value of $Q = 1$ might be used for simplicity. For the reverberant room tested, the use of $Q = 1$ in equations (8) and (9) leads to relatively small differences in STI (in comparison to predictions obtained from the exact Q) for the high signal-to-noise ratio and L_{SN} = +10 dB (up to +0.03 maximum), but large differences of up to +0.1 are found for L_{SN} = 0 dB and L_{SN} = -10 dB; these differences are similar for central and off-axis positions. In the absorbent room, the differences are small at central positions (±0.01 for all signal-to-noise ratios), but much larger at off-axis positions: up to +0.12 for the high signal-to-noise ratio and L_{SN} = +10 dB, and up to +0.19 for L_{SN} = 0 dB and L_{SN} = -10 dB. These results suggest that the use of a directivity $Q = 1$ is then acceptable only for high signal-to-noise ratios and central positions, but otherwise the exact directivity should always be used for the method to be reliable. Incidentally, it can be noted that the use of $Q = 1$ can result in better STI predictions, as the increases in STI can compensate the underestimations normally obtained from STI predictions based on L_{SN} and T. However, such improvements occur by accident rather than because of better accuracy of the input data used.

6. Conclusions

This paper examined differences between measured and predicted speech intelligibility in two rooms, for a wide range of absorption conditions and signal-to-noise ratios (sixteen tests). Speech intelligibility was quantified using the speech transmission index (STI), which was measured using maximum length sequence analysis and compared with predictions obtained from the room acoustic parameters T (reverberation time) and L_{SN} (signal-to-noise ratio). These parameters were either measured or predicted from fundamental room acoustic theories assuming diffuse sound field conditions. Three predictions were examined, based on T and L_{SN} values which were either measured (Prediction 1),

Table 8
Average differences between STI measurements and predictions for all the positions examined in the rooms (refer to Fig. 1 and 2 for the positions' numbering). Averages calculated from all signal-to-noise ratios and absorption conditions.

Position	1	2	3	4	5	6	7	8	9	10	11	12
STI difference	0.042	0.070	0.037	0.046	0.058	0.041	0.055	0.062	0.046	0.037	0.049	0.047

(a) Reverberant room

Position	1	2	3	4	5	6
STI difference	0.041	0.026	0.040	0.057	0.026	0.033

(b) Absorbent room

Table 9
Average differences between predicted and measured reverberation time averaged across all positions (single value obtained by averaging results over the octave bands 125 Hz – 8 kHz). The sign (+) indicates that the predicted reverberation time is greater than the measured reverberation time at most frequencies, whilst the sign (-) indicates that it is lower at most frequencies.

	$T_{\text{pred.2}} - T_{\text{meas.}}$		$T_{\text{pred.3}} - T_{\text{meas.}}$	
	(s)	(%)	(s)	(%)
Without panels	+0.03 (+)	+1%	0.0	0%
With panels	-0.11 (-)	-8%	-0.09 (-)	-6%

(a) Reverberant room

	$T_{\text{pred.2}} - T_{\text{meas.}}$		$T_{\text{pred.3}} - T_{\text{meas.}}$	
	(s)	(%)	(s)	(%)
Without panels	-0.07 (-)	-13%	0.0	0%
With panels	-0.03 (-)	-8%	+0.04 (+)	+11%

(b) Absorbent room

Table 10
Average differences between predicted and measured signal to noise ratio, averaged across all positions and averaged over the octave bands 125 Hz – 8 kHz (identical differences obtained for all the L_{SN} conditions tested). The sign (+) indicates that the predicted signal-to-noise ratio is greater than the measured signal-to-noise ratio at most frequencies, whilst the sign (-) indicates that it is lower at most frequencies.

	ΔL_{SN} (predicted - measured)	
	Reverberant room (dB)	Absorbent room (dB)
Without panels	-0.9 (-)	-0.4 (-)
With panels	-0.9 (-)	+0.7 (+)

(a) Central position

	ΔL_{SN} (predicted - measured)	
	Reverberant room (dB)	Absorbent room (dB)
Without panels	+0.2 (-)	-2.3 (-)
With panels	+0.4 (+)	-1.4 (-)

(b) Off-axis position

	ΔL_{SN} (predicted - measured)	
	Reverberant room (dB)	Absorbent room (dB)
Without panels	-0.3 (-)	-2.4 (-)
With panels	-0.1 (-)	-1.3 (-)

(c) Average across all positions

predicted (Prediction 2) or a combination of measured and predicted values (Prediction 3).

Results showed that the differences between measured and predicted STIs were always lower than 0.1, and on average lower than 0.06. According to previous research [17], these differences are noticeable, and therefore non-negligible, as they are greater than the just noticeable difference in STI of 0.03. The results also indicated that STI predictions based on T and L_{SN} tend to underestimate the STI, but no reason could be found to justify this.

The use of either measured or predicted reverberation time and signal-to-noise ratio showed little variations between predictions, with average differences in STI always below 0.03 for the absorbent room, and below 0.01 for the reverberant room. It was shown that the higher differences found in the absorbent room are due to inaccuracies in the prediction of its more pronounced direct sound field and its lower reverberation time. The non-diffuseness of the absorbent room and presence of furniture also resulted in more marked variations between STI predictions made on-axis and off-axis, due to larger differences between the measured and predicted signal-to-noise ratios.

Finally, it was pointed out that the relatively small variations observed between STI predictions using either measured or predicted input data cannot be generalised, as larger inaccuracies due to poorly predicted input data can be expected, especially if inappropriate absorption coefficients were to be used for estimating the reverberation time.

References

[1] Houtgast T and Steeneken H. The modulation transfer function in room acoustics as a predictor of speech intelligibility. Acust 1973;28:66-73.

[2] Rife DD. Modulation transfer function measurement with maximum-length sequences. J. Audio Eng. Soc. 1992;40(10): 779-790.

[3] Houtgast T, Steeneken H, Plomp R. Predicting speech intelligibility in rooms from the modulation transfer function. I. General room acoustics. Acust 1980;11:195-200.

[4] Steeneken H, Houtgast T. A physical method for measuring speech-transmission quality. J Acoust Soc Am 1980;67(1):318-326.

[5] Houtgast T, Steeneken H. A review of the MTF concept in room acoustics and its use for estimating speech intelligibility in auditoria. J Acoust Soc Am 1985;77(3):1069-1077.

[6] Rife DD, Vanderkoy J. Transfer function measurement with maximum length sequences. J Audio Eng Soc 1989;37:419-444.

[7] BS EN 60268-16. Sound system equipment – Part 16: Objective rating of speech intelligibility by speech transmission index; 1998.

[8] Fry A. Noise control in building services. Oxford, UK: Pergamon Press, Sound Research Laboratories; 1988.

[9] Long M. Architectural Acoustics. London, UK: Academic Press; 2006.

[10] Parkin P, Humphreys H, Cowell J. Acoustics, noise and buildings. London, UK: Faber and Faber; 1979.

[11] Building Research Establishment. BRE – Absorption coefficients. Watford, UK. Available online at www.bre.co.uk (last accessed February 2008).

[12] Auralex 2" Studiofoam Wedges. Sound Absorption Test RAL-A93-58. Illinois, USA: Riverbank Acoustical Laboratories of IIT Research Institute; 1993.

[13] Astolfi A, Corrado V, Griginis A. Comparison between measured and calculated parameters for the acoustical characterization of small classrooms. Appl Acoust 2008;69:966-976.

[14] ISO 9613-1. Acoustics – Attenuation of sound during propagation outdoors – Part 1: Method of calculation of the attenuation of sound by atmospheric absorption; 1993.

[15] Vér IL, Beranek LL. Noise and vibration control engineering. New Jersey, USA: John Wiley & Sons, second edition; 2006.

[16] ANSI/ASA S3.5-1997 (R2012). American national standard methods for calculation of the speech intelligibility index; 1997.

[17] Bradley JS, Reich R, Norcross SG. A just noticeable difference in C_{50} for speech. Appl Acoust 1999;58:99-108.