# Increasing Robustness in the Calculation of the Speech Transmission Index from Impulse Responses

*by*

**Densil Cabrera, Doheon Lee, Glenn Leembruggen and Daniel Jimenez**

*Reprinted from*

JOURNAL OF
BUILDING ACOUSTICS
Volume 21 · Number 3 · 2014

MULTI-SCIENCE PUBLISHING CO. LTD.
5 Wates Way, Brentwood, Essex CM15 9TB, United Kingdom

# Increasing Robustness in the Calculation of the Speech Transmission Index from Impulse Responses

**Densil Cabrera[1], Doheon Lee[1], Glenn Leembruggen[1, 2] and Daniel Jimenez[1]**

*[1]Faculty of Architecture, Design and Planning, The University of Sydney, NSW 2006, Australia. densil.cabrera@sydney.edu.au*

*[2]Acoustic Directions, PO Box 205 Summer Hill NSW 2130, Australia*

## ABSTRACT

There are many factors that can affect the measured values of the speech transmission index (STI), and this paper examines how and why identical inputs into STI calculation software are yielding varying results. The study involved a survey of a number of software implementations of the Indirect Method for computing the STI from an impulse response, one of which was written by the authors. Results are presented for artificial and measured impulse responses, and for signal and noise spectra that were designed to test particular aspects of the STI calculation. While most deviations between implementations were within 0.01 STI, some were not, revealing a need for greater robustness in the design of software and greater clarity in the STI standard (IEC60268-16), including more support for validation. This paper provides some data for such validation.

## 1. INTRODUCTION

As personal computers have become widely used and more powerful, measurement and analysis techniques in acoustics have been developed and refined to take advantage of the computing capability that is now available to practitioners and researchers. Standards in acoustics have followed this trend, with some standards involving quite extensive algorithms to derive values from acoustic measurements. Often commercial, closed-source software is available to implement such algorithms, which provides a convenient alternative to writing one's own implementation in computer code, and presumably provides a reliable implementation of the output metrics. However, the performance of software implementations varies, and Katz has shown that even a comparatively simple and well-understood metric such as reverberation time can yield results that vary markedly between implementations [1].

The speech transmission index (STI) is widely used to objectively predict the speech intelligibility performance of a transmission channel. Its areas of application include rooms for speech [2], speech distraction in open-plan offices [3], and intelligibility in audio systems [2,4], including emergency warning and information systems [4]. As

such it is a valuable tool for practitioners of audio and acoustics concerned with speech in the built environment, and in this context understanding the STI is an important part of underlying education. While the algorithm to compute the STI from an impulse response [2] is significantly more complex than the standard algorithm to calculate reverberation time from an impulse response, it is not a large task to implement it using an audio signal processing environment such as MATLAB or similar. STI calculations are also integrated into commercially available acoustic measurement and prediction software, which provide relatively simple user interfaces and workflow to derive values. This paper aims to identify issues in the calculation of STI values that become evident by testing and comparing various STI implementations.

The basis of the STI method is the computation of the modulation transfer function (MTF), which describes the loss of modulation in a signal over the octave frequency bands contained in a speech signal (125 Hz to 8 kHz).

The evolution of the STI method is described by van Wijngaarden et al. [5]. Editions 1 and 2 of IEC60268-16 required the MTF to be the measured directly from the test signal. Edition 3 of IEC60268-16 [6], released in 2003, provided a method of computing the MTF and hence STI from an impulse response, with the mathematical introduction of a signal-to-noise ratio term (SNR). Edition 4 (dated 2011) [2] calls this the Indirect Method and provides substantially more detail than [6]. Measurement of the MTF was first proposed by Schroeder [7] (without the SNR adjustment in its original form), and is shown in equation 1.

$$m_k\left(f_{\mathrm{m}}\right)=\frac{\left|\int_0^{\infty} h_k\left(t\right)^2 e^{-j2\pi f_{\mathrm{m}} t}\,dt\right|}{\int_0^{\infty} h_k\left(t\right)^2\,dt}\times\left(1+10^{-SNR_k/10}\right)^{-1} \tag{1}$$

Here $h_k(t)$ is an impulse response filtered to octave band $k$, $f_{\mathrm{m}}$ is a modulation frequency, $t$ is time, and $SNR_k$ is the SNR in octave band $k$ expressed in decibels.

The result, $m_k$, is the modulation transfer ratio, which as a function of modulation frequency $f_{\mathrm{m}}$ over the relevant octave bands forms the modulation transfer function (MTF). The MTF that is used in the calculation of the STI is constructed from 14 modulation frequencies (0.63 Hz to 12.5 Hz, 1/3-octave spaced) and seven octave bands (125 Hz to 8 kHz). Note that this equation in IEC60268-16 (Edition 4) has a typographical error, in that it omits the squaring of the impulse response in the numerator of the first term (p. 26). Notwithstanding this, the first term is quite straightforward to implement. The second term is the ratio of signal to signal-plus-noise, which is derived from supplementary data rather than from the impulse response itself. The noise term represents not only physical noise, but also threshold and masking effects in the auditory system. The equation can be more explicitly expressed as per equation 2.

$$m_k\left(f_{\mathrm{m}}\right)=\frac{\left|\int_0^{\infty}h_k\left(t\right)^2 e^{-j2\pi f_{\mathrm{m}}t}dt\right|}{\int_0^{\infty}h_k\left(t\right)^2 dt}\times\frac{I_{\mathrm{s},k}}{I_{\mathrm{s},k}+I_{\mathrm{n},k}+I_{\mathrm{rt},k}+I_{\mathrm{am},k}} \tag{2}$$

Here, $I$ is an intensity value ($I_{\mathrm{s}}$ for speech, $I_{\mathrm{n}}$ for background noise, $I_{\mathrm{rt}}$ for auditory reception threshold and $I_{\mathrm{am}}$ for auditory masking due to the intensity in the octave band below). $I$ is related to the respective sound pressure level, $L$, by equation 3.

$$I=10^{L/10} \tag{3}$$

Edition 4 of IEC60268-16 included some important changes, including refinement of the auditory-masking algorithm (used to derive $I_{\mathrm{am}}$). Although the more explicit second term in equation 2 also appears to be straightforward, it is easy to lose sight of this simple operation when implementing the intricacies of the STI calculation (e.g., calculating $I_{\mathrm{am}}$), and the present study finds that some implementations have errors in this area. There is certainly scope for greater clarity in the standard, including in section A.5.3, which presents this term relating signal to signal-plus-noise. However, Annex M of the standard presents a worked example that helps to resolve uncertainty in interpretation.

The Indirect Method of deriving the STI from impulse responses (rather than directly from a modulated noise signal) has advantages of significant convenience and versatility. Impulse responses are commonly acquired for a variety of reasons in measuring audio systems and room acoustics – and it is convenient to be able to derive an STI value from the same measurement that is used to derive a host of other parameters (notwithstanding the need for particular transducer types in some circumstances). The method is versatile in that the octave band signal and noise levels must be accounted for subsequent to the impulse response measurement, which allows for various signal and noise scenarios to be evaluated from a single impulse response. The main potential disadvantage of the Indirect Method is that it may not be immediately obvious how the sound pressure level of speech is derived (although Annex J of the standard provides guidance on this), whereas the speech level is automatically included in correctly calibrated Direct Method measurements. This is one of the aspects of the STI method where education can provide considerable benefit.

## 2. TEST MATERIAL

### 2.1 Implementations

The impetus for this study originates from the authors' efforts in writing an implementation of the STI (Indirect Method) in MATLAB. As a check of the code's correctness, the code was tested against commercially and freely distributed implementations of STI computations. The unexpected differences in results between the various implementations pointed to several issues, which are discussed in this paper.

In response to these differences, the authors' implementation was improved and validated to the point that they are confident that its results are correct. This checking included:

(i) confirming that the function matches the worked example of Annex M in the standard;
(ii) confirming that Annex M (specifically, the second half of it) is a correct representation of the algorithm expressed in the standard;
(iii) confirming that the initial signal processing to derive the MTF was correct by reference to the standard, comparison with other implementations and theory; and
(iv) comparing results with those of other implementations.

The other implementations surveyed in this paper are referred to using letters 'B'-'I', with 'A' signifying the authors' implementation. They include six commercial and two free implementations. Some of these implementations are popularly used in professional practice and education. The tested versions were current at the end of the year 2013.

**2.2 Impulse responses**

Although many impulse responses were tested in the work leading to this paper, the paper reports on results of four impulse responses, three of which test particular features in the software implementations. This set of test impulse responses consists of a delta function, two impulse responses generated from exponentially decaying octave-spaced tones, and an impulse response measured in a lecture theatre (i.e. a realistic example). These are described in more detail below.

The delta function impulse response is a single digital impulse (Kronecker delta function) after 10 ms of silence, and followed by silence (comprising a 2 s duration waveform sampled at 48 kHz, 32-bit precision). The theoretical MTF and STI values of this are all 1, but measured values are likely to be slightly lower (due to the octave bandpass filters' own MTF, and perhaps for other reasons). The reason for the 10 ms initial silence is that one of the tested implementations appeared unable to calculate STI and related values if the impulse peak was on the first sample of the waveform.

The two exponentially decaying octave-spaced tone impulse responses were designed to test the implementations' response to particular combinations of reverberation times. The purpose of using exponentially decaying octave-spaced tones (rather than exponentially decaying noise, which would sound more like a real room impulse response) is that the theoretical modulation transfer function of this signal is exactly known, and therefore deviations from it can be used as a diagnostic tool.

As the MTF is derived from envelopes of the octave-band filtered impulses, the actual frequency content within each octave band (whether narrow band, or spanning the full octave band) does not influence the MTF (apart from its incidental effect on the envelope). Such impulse responses have been used in other studies to test the performance of various reverberation time analysis algorithms for the same underlying reason, either explicitly [8] or implicitly [9].

The theoretical modulation transfer ratio ($m$) of an exponentially decaying impulse response with a given reverberation time ($T$) is a function of modulation rate ($f_m$), as described by equation 4. The value 13.8 in the equation comes from $\ln(10^6)$, relating to the 60 dB range used to define reverberation time. This theoretical equation does not account for irregularities that often occur at the start of a room impulse response (the strong direct sound and distinct early reflections). The use of artificial impulse responses based on decaying tones, avoids such irregularities. When equation 4 is evaluated for the 14 standard values of $f_m$ using the seven octave band reverberation times, the theoretical MTF matrix is derived, from which the theoretical STI value can be calculated.

$$m\left(f_m\right)=\frac{1}{\sqrt{1+\left(\frac{2\pi f_m T}{13.8}\right)^2}} \tag{4}$$

The exponentially decaying tone impulse responses, $y(t)$, were generated using equation 5 (where $t$ is time, starting at 0 s). They consist of seven simultaneous exponentially decaying pure tones tuned to octave band centre frequencies.

One of the impulse responses, *IR2*, was generated with a reverberation time, $T$, of exactly 1 s in all seven octave bands ($k$ = 1 to 7, $f_0$ = 125 Hz). The other decaying tone impulse response, *IR3*, was generated with reverberation varying markedly between adjacent bands: chosen values are $T_k$ = 2.00, 0.31, 2.00, 0.31, 2.00, 0.31, and 2.00 s for $k$ = 1 to 7 (i.e., frequencies of 125 Hz to 8000 Hz respectively). This test waveform is used to examine the effect of octave band filter selectivity on MTF and STI values, and the particular values were chosen to yield an STI value of 0.60 (without noise, threshold and masking effects).

These test impulse responses were generated and preceded by a 10 ms delay within 2 s wave files (sampling rate of 48 kHz, with 32-bit precision).

$$y\left(t\right)=\sum_{k=1}^{7}\sin\left(2^k\pi f_0 t\right)e^{-3\ln(10)t/T_k} \tag{5}$$

The fourth test impulse response was measured in a lecture theatre using a Brüel & Kjær head and torso simulator (HATS, type 4128C) as the source at the lecturing position, and a free field microphone as receiver (Brüel & Kjær type 4190) located towards the rear of the audience seating (source-receiver distance of 8 m). Other equipment used for the measurement included a Brüel & Kjær Nexus microphone preamplifier and power supply, an RME Babyface audio interface, Brüel & Kjær power amplifier (type 47216C) and a MATLAB-based software environment for acoustic measurement written by the authors. The lecture theatre seats 110, and has a volume of

610 $m^3$. The unoccupied octave band reverberation times (T30), derived only from the particular impulse response, are 1.28, 0.92, 0.74, 0.84, 0.88, 0.80 and 0.58 s in the 125 Hz – 8 kHz octave bands respectively. Further details of this lecture theatre are given by Cabrera et al. [10]. The impulse response was measured using an exponentially swept sinusoid (45 s duration, spanning 50 Hz to 15 kHz). Background noise was measured, and octave band speech level was predicted from room acoustic measurements and theory, taking into account the 1 m free-field speech levels stated in IEC60268-16 Edition 4, the source-receiver distance, the room's volume and reverberation time, and the measured directivity pattern of the head and torso simulator [11]. This test impulse response was truncated to 2 s, with a sampling rate of 44.1 kHz and 16-bit precision. In a sense, the particular details of this measurement are unimportant – the important point is that this impulse response represented a realistic result from measurement, to complement the highly unrealistic synthetic impulse responses described above.

## 3. RESULTS

### 3.1 Effect of filterbank on the MTF and the STI

The calculation of the STI from an impulse response involves several steps, some of which can be omitted for testing purposes. The derivation of the original MTF (without adjustments for SNR) involves octave band filtering, the application of the first half of equation 1 (or 2) to derive the 98-valued MTF, and the derivation of the STI value from the MTF. Most implementations make the MTF available for viewing or further analysis by the user.

Although the very simple test stimuli are highly unrealistic, they are helpful identifying possible limitations in filtering and calculating the MTF, because the theoretical result is exactly known. The key limitation in reaching the theoretical values should be the design of the octave bandpass filters. Before considering the results of various implementations, we first examine the effect of bandpass filter design. IEC60268-16 specifies that the filters should meet IEC 61260 class 0 or class 1 criteria [12], and have as linear phase response as possible. For computer-based implementations, linear (including zero) phase filters that meet class 0 requirements can be implemented, and this was the authors' approach. However, class 0 filters can have a wide range of frequency selectivity values, from shallow to very steep filter skirts, and the choice of filter selectivity can affect the resulting MTF for two reasons: firstly, greater selectivity slightly reduces the MTF values due to the MTF of the filter itself; and secondly, greater selectivity reduces the influence of out-of-band modulation. The first of these effects is examined by analysing the impulse waveform, and the second by analysing IR3 (exponentially decaying sinusoids with contrasting reverberation times between the octave bands). Results are given in Table 1. The tested filters were IEC 61260 class 0 compliant, zero phase, with filter skirts designed at 30, 72 and 144 dB/octave. Their measured responses are shown in Figure 1.

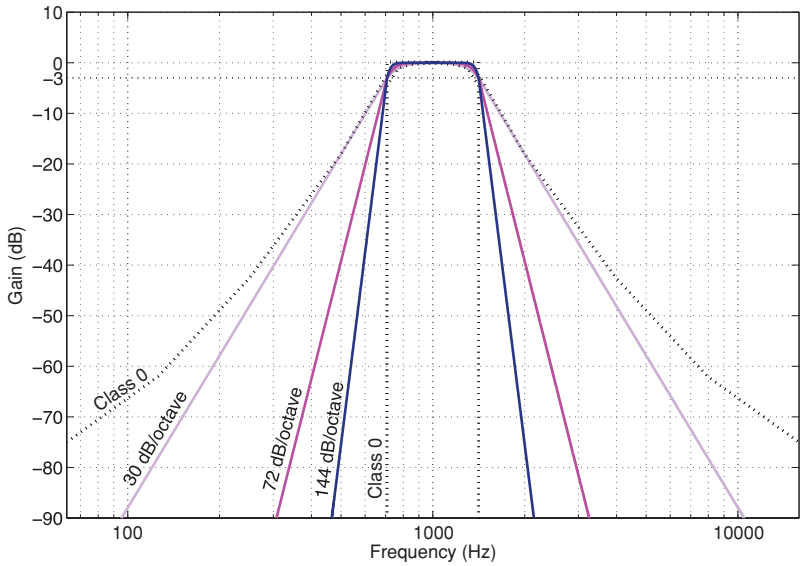


**Figure 1.** Measured magnitude response of the three zero phase filters (showing the 1 kHz octave band filter as an example), with 30 dB/octave (mauve), 72 dB/octave (magenta) and 144 dB/octave (dark blue) skirts. Limit curves for IEC 61260 class 0 octave band filters are also shown (dotted), and all three of the filters' measured responses are fully within these limits for all seven of the octave band filters (using a 2 s duration test signal with a sampling rate of 48 kHz).

A zero phase filter has an 'acausal' pre-ring (a build-up before time zero) in its impulse response, and the authors' STI implementation provides 100 ms of zero padding to the waveform prior to filtering to avoid significantly truncating the filterbank's response – effectively introducing a linear phase delay. However, the synthetic impulse responses tested for this paper also have 10 ms of silence at their start, which allows the pre-ring to be largely accounted for (if present) when analysed by other STI implementations.

As seen in Table 1, the STI value derived from the delta function is negligibly affected by the filterbank's frequency selectivity, although small effects can be seen in both the modulation transfer index (MTI) at the 125 Hz octave band, and the modulation transfer ratio ($m$) for $f_m$ of 12.5 Hz at 125 Hz (which is the most sensitive of the values in the MTF to the filterbank design).

Table 1 also shows that the STI value derived from the exponentially-decaying sinusoids with varying octave-band reverberation time is significantly affected by the filters' frequency selectivity. The 30 dB/octave filterbank (similar to a 5$^{th}$ order filter outside the bandpass limits, but with its in-band response modelled on a 7$^{th}$ order Butterworth filter) is not sufficiently selective for this waveform, while the 72 dB/octave and 144 dB/octave filterbanks are (these are similar to 12$^{th}$ order and 24$^{th}$

order filters respectively). Of the three filterbanks, the 72 dB/octave version is the best choice considering its results from both of these impulse responses, and this version is used in the other tests reported in this paper.

**Table 1: Values calculated for three class 0 compliant zero phase filters with filter skirts of 30, 72 and 144 dB/octave. The upper part of the table shows values calculated from the delta function waveform, while the lower part shows values calculated from the exponentially decaying sinusoids with reverberation time varying between octave bands, as described in the text. STI values in the table are calculated without adjusting for noise, masking or threshold.**

| **Delta function waveform** | **30 dB/oct** | **72 dB/oct** | **144 dB/oct** |
|---|---|---|---|
| STI (male), theoretical value is 1 | 1.00 | 1.00 | 1.00 (0.999) |
| MTI for 125 Hz octave band | 1.00 | 0.99 | 0.98 |
| $m$, 125 Hz band, $f_m$ = 12.5 Hz | 0.975 | 0.951 | 0.922 |
| Rms deviation of MTF from theory | 0.003 | 0.007 | 0.011 |

| **Decaying sinusoids IR3** | **30 dB/oct** | **72 dB/oct** | **144 dB/oct** |
|---|---|---|---|
| STI (male), theoretical value is 0.60 | 0.55 | 0.60 | 0.60 |
| Rms deviation of MTF from theory | 0.060 | 0.001 | <0.001 |

Figure 2 gives a more finely grained representation of the STI error resulting from the calculation of STI using class 0 compliant octave band filters with skirt slopes from 30 dB/octave to 72 dB/octave, when IR3 is analysed. IEC60268-16 allows 0.01 error in STI due to filter design, which for this impulse response requires filters with at least 45 dB/octave skirt slopes. The 45 dB/octave filterbank tested yields an rms error of 0.004 in the MTF of the delta function (compared to 0.007 rms error for the 72 dB/octave filterbank), but neither error affects the STI derived from the delta function. Of course, it must be remembered that IR3 is an extreme case, and measured impulse responses would not normally have such great contrasts between adjacent octave bands, meaning that less selective filters would normally not result in excessive errors.

Venturi *et al*. [8] conducted a similar investigation into the effect of filter selectivity on reverberation time analysis, finding that a more selective filter-bank (144 dB/octave) improves the accuracy of results for T30 (reverberation time evaluated from -5 dB to -35 dB of the decay, i.e. with a 30 dB evaluation range) for octave-band reverberation times ranging between 0.5 s and 2 s. However they also conclude that less selective filters would be appropriate for short reverberation times and for parameters like early decay time (EDT) or T10 that have small evaluation ranges (which would be more vulnerable to being influenced by the filters' time response).

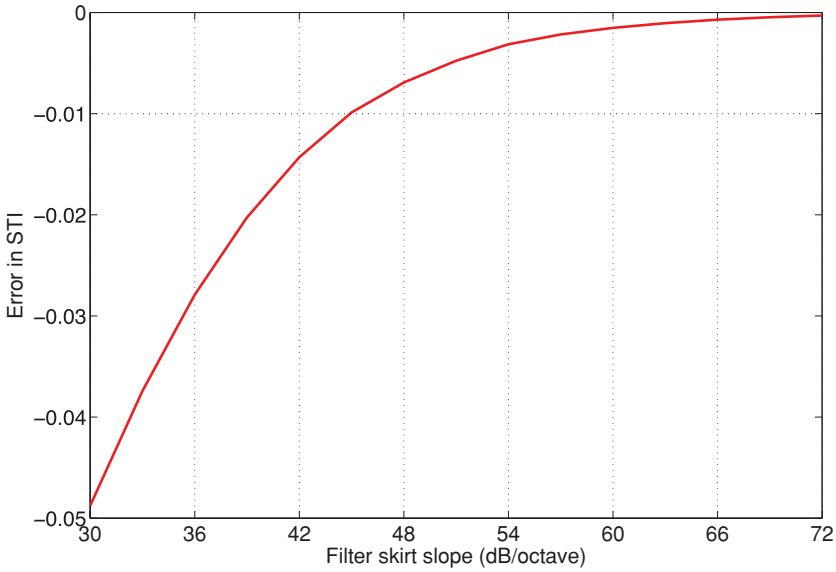


**Figure 2.** Error in the calculated STI as a function of octave band filter selectivity (which is characterised by the slope of the filter skirts) for IR3, where the reverberation time varies markedly between adjacent octave bands. Error values are negative because the STI is underestimated for IR3 when there is leakage between bands. The filters chosen for the authors' implementation have 72 dB/octave skirts.

Another issue that may yield differences between implementations is whether base 2, base 10 or nominal centre frequencies are used for octave band filters and for the modulation frequencies (which are 1/3-octave-spaced) [12]. For the seven octave band filters used for STI, the base 2 and nominal centre frequencies are identical, whereas the base 10 centre frequencies are $f_c=10^{3n/10}$, where $n$ is an integer from 7 to 13. Tests indicate that the choice has a negligible effect on results of indirect measurements (which is the focus of this paper). However, this issue becomes important with respect to the modulation frequencies when *direct* measurements, such as STIPA [2], are made. With STIPA, the upper modulation frequency in each octave band must be exactly 5 times the lower frequency, which precludes the use of base 2 or base 10 frequencies. The definition of the octave in IEC60268-16 suggests that base 2 is preferred for octave band frequencies, although for the impulse responses tested, the root mean square deviation between MTFs derived from base 2 versus base 10 filters (72 dB/octave) is less than 0.001, and similarly the effect on STI is less than 0.001. With regard to the choice of modulation frequency, theoretical testing (using equation 4) shows that results from nominal, base 10 and base 2 modulation frequencies are negligibly different (e.g., for $T$ = 0.5, 1 and 2 s, the effect on STI is less than 0.001).

### 3.2 Testing of STI Implementations

The various implementations of STI were tested using the synthetic impulse responses, deriving the MTF without noise, masking and threshold adjustments) and STI (male) values. Table 2 presents the results. STI values are given with a precision of two decimal places, as per IEC60268-16, which permits a 0.01 STI deviation due to octave band filter design (for STI values in the 0.10 to 0.90 range).

For the delta function, STI was 1.00 in all cases, but the other two impulse responses yielded some values that deviated from theory. In the case of IR2 ($T$=1 s in all octave bands), only implementation I has an unexpected STI value; but in the case of IR3, C, F, G, H and I yielded reduced STI values. Note that there is some uncertainty in the result for H because it was not possible to disable masking in its STI calculation. To check the various implementations' derivation of STI from the MTF, the STI was also calculated outside of each implementation from their MTF (apart from H and I), and in all cases the calculated STI was the same as that output by the program. Hence the deviations in STI values in Table 2 result from the derivation of the MTF, rather than from subsequent stages in STI calculation.

**Table 2: Comparison of STI calculated by implementations A-I from three synthetic impulse responses with known theoretical MTF and STI values. The upper table shows the STI values (without noise, masking or threshold adjustments), where the theoretical values are 1.00 (IR1), 0.59 (IR2) and 0.60 (IR3). The lower table shows the root mean square error of the modulation transfer functions. Values that deviate significantly from theoretical values are shown in bold. Implementations H and I do not provide MTF values, and so their MTF error is not available. Implementation H's values (italicized) could be masking-affected (because masking could not be disabled in that implementation). Implementation A was written by the authors.**

| STI | A | B | C | D | E | F | G | H | I |
|---|---|---|---|---|---|---|---|---|---|
| delta function (STI = 1) | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | *1.00* | 1.00 |
| IR2 (STI = 0.59) | 0.59 | 0.59 | 0.59 | 0.59 | 0.59 | 0.59 | 0.59 | *0.58* | **0.55** |
| IR3 (STI = 0.60) | 0.60 | 0.60 | **0.57** | 0.59 | 0.60 | **0.57** | **0.56** | *0.52* | **0.56** |

| MTF rms error | A | B | C | D | E | F | G | H | I |
|---|---|---|---|---|---|---|---|---|---|
| delta function | 0.01 | 0.01 | <0.01 | <0.01 | <0.01 | <0.01 | <0.01 | - | - |
| IR2 | <0.01 | 0.01 | **0.02** | <0.01 | **0.02** | <0.01 | 0.01 | - | - |
| IR3 | <0.01 | 0.01 | **0.04** | **0.02** | **0.02** | **0.04** | **0.04** | - | - |

Examination of the MTF values provides further insight into the differences between implementations, and the second part of Table 2 summarises these by their root mean square error (from the theoretical MTF). Implementations A and B returned error values of 0.01 or less in all cases. Larger errors were found from implementations C and E for IR2, and for C, D, E, F and G for IR3. Since IR3 was designed to test the effect of filter selectivity on MTF, the filters in C, D, E, F and G might not be optimally selective for

impulse responses that have reverberation times or overall energy varying substantially between adjacent bands. The STI results of IR3 for C and F are consistent with 6th order (36 dB/octave) filters (drawing on the analysis in Figure 2), and G, H and I might also have relatively weak filter selectivity (probably combined with other issues) contributing to their lower STI values for IR3.

However, there is a different explanation (or partial explanation) for the performance of C, E and G, revealed by their results with IR2, which comes from the particular modulation frequencies used in deriving the MTF. Previously it was noted that the differences between using nominal, base 2 or base 10 modulation frequencies was negligible – but these implementations do not use any of these. Instead they round the modulation frequencies to a linearly distributed frequency scale (rather than a logarithmically distributed scale), presumably because a fast Fourier transform or similar technique is used to derive the MTF instead of directly implementing the first half of Equation 1. This would not produce a significant deviation if the impulse response were long, as frequency precision is the reciprocal of the analysis period, and therefore a 10 s analysis period would yield 0.1 Hz precision, which should be adequate in almost all circumstances. However, the impulse responses tested here have a duration of 2 s, yielding a resolution of 0.5 Hz on a linear frequency scale, which is too coarse to distinguish some of the low modulation frequencies that are used to calculate MTF.

Examination of the values in the MTFs and rearranging Equation 4 to find the corresponding modulation frequency reveals that implementation E indeed has 0.5 Hz precision, whereas implementation C has 0.37 Hz precision and G has 0.18 Hz precision. It is likely that this finer precision comes from zero-padding the 2 s impulse response, with the values being consistent with $2^{17}$ and $2^{18}$ samples in the analysis window. Examples of distorted MTFs are shown in Figure 3, although with its finer frequency precision, the distortion in G is quite small.

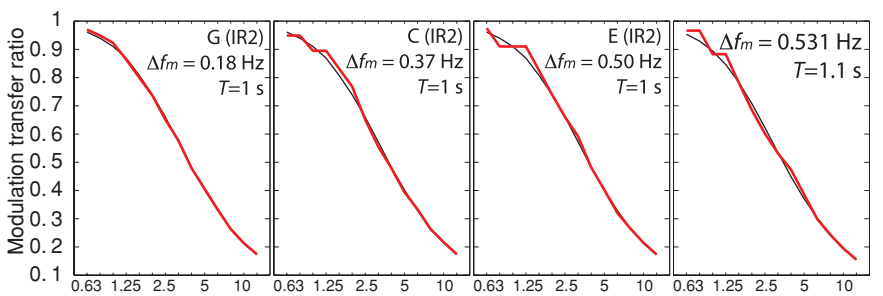


**Figure 3.** Modulation transfer functions for a reverberation time of 1 s (first three charts from the left) and 1.1 s (right chart). The theoretical result (from Equation 4) is shown as a black line on each of the charts. The first three charts (from the left) show the measured results (in the thicker red line) from IR2 using implementations G, C and E in the 1 kHz octave band. The right chart shows the result that would be obtained with 0.531 Hz precision (the thicker red line), which is the case where the error in STI is greatest (refer to Figure 4).

IEC60268-16 allows impulse responses to be no shorter that 1.6 s (which is the period corresponding to the lowest modulation frequency of 0.63 Hz) and the mapping of linear to logarithmic frequencies is likely to result in greater errors for shorter duration impulse responses if zero-padding is not done. While rounding of modulation frequencies to a linear frequency scale distorts the MTF, it may not result in a significant error in the final STI value because some modulation frequencies will be rounded up, and others are rounded down – which tends to balance out the error in the STI calculation. Figure 4 confirms this: over a wide range of reverberation times and modulation spectrum precision values (using Equation 4), the STI error is mostly less than 0.01 STI (the greatest error is +0.0128 STI). Errors are largest at 0.531 Hz, which could occur with an impulse response duration of 1.88 s without zero-padding. If the minimum impulse response duration (after zero-padding, if applicable) is 2.5 s, then the worst case error is 0.005 STI; if the minimum duration is 6 s, then the worst case error is 0.003 STI.

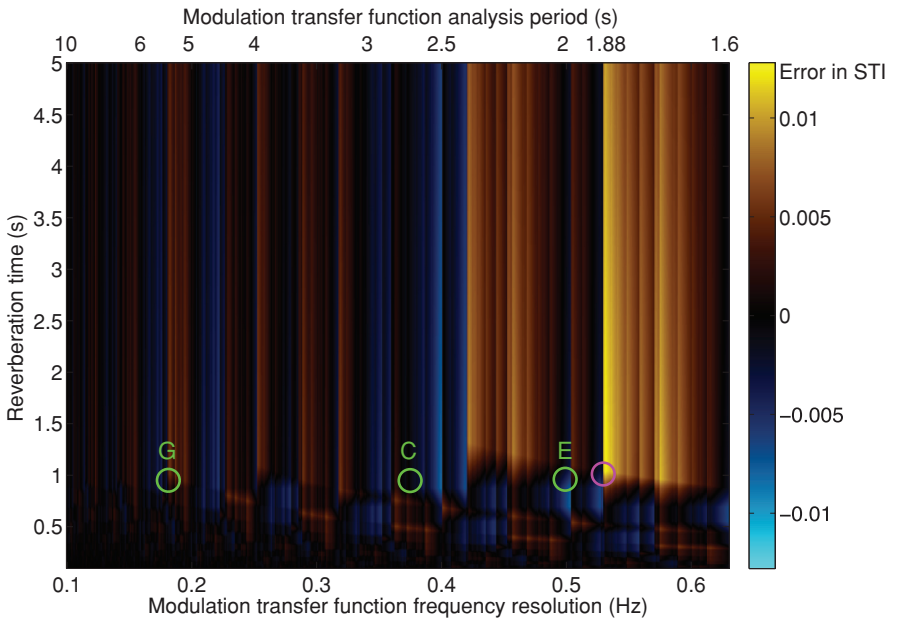


**Figure 4.** Calculated error in STI value resulting from using linearly distributed modulation frequency values, as a function of spectrum precision ($\Delta fm$ from 0.1 to 0.63 Hz) and reverberation time. Orange indicates that the calculated STI value is higher than had it been calculated with exact modulation frequencies (a positive error), and blue indicates a negative error. The largest errors are positive, and error is greatest for a reverberation time of 1.1 s combined with a frequency resolution of 0.531 Hz (identified by the magenta circle). The small STI errors predicted for analysing IR2 by implementations C, E and G are identified by green circles.

### 3.2 Adjusting for signal-to-noise ratio, threshold and masking

In most circumstances in architectural acoustics, STI measurements need to be adjusted to account for signal and noise sound pressure levels, including the effects of auditory reception threshold and masking. Although these adjustments are achieved by simple mathematical operations, deviations could result from misinterpretations of the standard. A further problem can be that an implementation does not provide a clear method of entering the octave band speech and noise sound pressure level data that is required for these adjustments. If these adjustments cannot be made, then the resulting STI values will be artificially high, potentially significantly so. Implementation I is not included in this part of the paper because there was no clear way of providing the data to make these adjustments.

The implementations were tested extensively, but for succinctness the results of only four pairs of signal (S) and noise spectra (N) are presented as examples in Table 3. The first pair (SN1) has measured spectra from the lecture theatre as previously described. The second (SN2) has the same speech spectrum, but a much reduced signal-to-noise ratio. The third (SN3) has a jagged speech spectrum, which will have much stronger masking effects. The last (SN4) is the same as the first but is attenuated by 30 dB, with the signal near the auditory reception threshold. Hence, the four signal and noise spectrum pairs are designed to test general performance (SN1), the effect of noise (SN2), the effect of masking (SN3) and the effect of threshold (SN4) on the resulting STI value.

**Table 3: Octave band sound pressure levels (dB re 20 µPa) of four signal (S) and noise (N) spectrum pairs used to test implementations of the STI. The auditory reception threshold used for the STI calculations is also shown.**

| | 125 Hz | 250 Hz | 500 Hz | 1 kHz | 2 kHz | 4 kHz | 8 kHz |
|---|---|---|---|---|---|---|---|
| **S1** | 60.1 | 57.4 | 53.8 | 48.4 | 40.7 | 33.5 | 25.9 |
| **N1** | 41.3 | 33.2 | 32.2 | 22.3 | 23.1 | 20.9 | 14.8 |
| **S2** | 60.1 | 57.4 | 53.8 | 48.4 | 40.7 | 33.5 | 25.9 |
| **N2** | 55.0 | 55.0 | 50.0 | 45.0 | 40.0 | 30.0 | 20.0 |
| **S3** | 80.0 | 65.0 | 60.0 | 70.0 | 80.0 | 60.0 | 30.0 |
| **N3** | 10.0 | 10.0 | 10.0 | 10.0 | 10.0 | 10.0 | 10.0 |
| **S4** | 30.1 | 27.4 | 23.8 | 18.4 | 10.7 | 3.5 | -4.1 |
| **N4** | -10.0 | -10.0 | -10.0 | -10.0 | -10.0 | -10.0 | -10.0 |
| **Thresh.** | 46 | 27 | 12 | 6.5 | 7.5 | 8 | 12 |

Two impulse responses were tested with these combinations of signal and noise spectra: the synthetic impulse response with a reverberation time of exactly 1 s (IR2), and the measured impulse response from the lecture theatre.

One of the potential issues here is that some implementations are based on Edition 3, while others are based on Edition 4, of IEC60268-16 – and these two editions employ slightly different masking algorithms. However, for the particular inputs chosen, the

absolute difference in STI value between the masking algorithms of Edition 3 and 4 is 0.0014 or less. In cases where an implementation allowed either edition to be used, both were tested, but only Edition 4 results are given here because the values did not significantly differ (by more than 0.01 STI).

The results given in Table 4 show a similar pattern for the synthetic and measured impulse responses. For SN1 and SN2, there are no significant differences between the results of the eight tested implementations (A-H). For those signal and noise conditions, most results are the same as A, with a maximum deviation of 0.01 STI, which is not significant. However four of the implementations underestimate the reduction in STI due to the effect of masking for SN3 (B, C, D and F), one slightly overestimates the effect of masking (H), and four of the implementations do not correctly account for auditory threshold for SN4 (C, F, G and H).

**Table 4. STI values calculated for the signal and noise spectra specified in Table 3, for the artificial impulse response with a reverberation time of exactly 1 s (IR2), and the measured impulse response from the lecture theatre. The letters A-H refer to the implementations tested. Values that have significant errors are in bold.**

| STI from IR2 | A | B | C | D | E | F | G | H |
|---|---|---|---|---|---|---|---|---|
| SN1 (measured) | 0.57 | 0.57 | 0.57 | 0.57 | 0.57 | 0.56 | 0.57 | 0.56 |
| SN2 (high noise) | 0.42 | 0.42 | 0.43 | 0.42 | 0.42 | 0.42 | 0.42 | 0.42 |
| SN3 (masking) | 0.53 | **0.58** | **0.59** | **0.58** | 0.53 | **0.58** | 0.52 | **0.51** |
| SN4 (threshold) | 0.39 | 0.39 | **0.57** | 0.39 | 0.40 | **0.27** | **0.36** | **0.34** |

| STI from measured IR | A | B | C | D | E | F | G | H |
|---|---|---|---|---|---|---|---|---|
| SN1 (measured) | 0.64 | 0.64 | 0.64 | 0.64 | 0.64 | 0.63 | 0.64 | 0.64 |
| SN2 (high noise) | 0.47 | 0.47 | 0.47 | 0.47 | 0.47 | 0.46 | 0.47 | 0.48 |
| SN3 (masking) | 0.59 | **0.65** | **0.67** | **0.66** | 0.59 | **0.66** | 0.59 | **0.55** |
| SN4 (threshold) | 0.43 | 0.43 | **0.63** | 0.43 | 0.43 | **0.29** | **0.40** | **0.39** |

When interpreting these results, it must be remembered that all of these implementations correctly calculate the STI for IR2 when masking and threshold do not influence the result. Apart from the edition (3 or 4) of the standard in each implementation (which has a negligible effect on the result in these tests), there is no flexibility within the masking and threshold algorithms that would allow results to differ at all. Therefore, the deviations seen in Table 4 result from errors in the implementation of the algorithm.

The error in masking (SN3) is particularly interesting because four of the eight implementations exhibit approximately the same error. Some of these implementations are of Edition 3, and some of Edition 4 (and in one case both were tested, yielding approximately the same error). Although the source of this error is not known (and it is not known if the error has the same source in all four implementations), the erroneous result can be replicated by incorrect indexing of the octave band number. The intensity

of masking in octave band $k$ should be calculated from the product of the intensity in the octave band below ($k$-1) and the auditory masking factor, *amf* (which is also calculated from the intensity in the band below). This is expressed in Equation 6, but if the index on the right-hand side of Equation 6 is changed to $k$ instead of $k$-1, then the erroneous result is replicated (yielding STI of 0.59 and 0.66 for IR2 and the measured impulse response respectively). This equation is common to the two editions.

$$I_{am,k} = I_{k-1} \times amf \tag{6}$$

## 4. DISCUSSION

The majority of tests reported in this study return STI values that match expectations, but there are some circumstances where the STI values deviate appreciably. With the exception of implementation I, the various implementations provided correct results when the input impulse response (together with signal and noise spectra if relevant) did not have extreme characteristics. However, by testing extreme inputs, this study has identified vulnerabilities in various implementations, and has shown that a more robust implementation is feasible.

One area where additional robustness is possible is the design of the octave band filterbank. Steep filter skirts can be achieved (e.g. 72 dB/octave) without artificially reducing the calculated STI (even for a delta function impulse response). The filter can be linear (or zero) phase, thereby avoiding the introduction of any phase distortion within the octave band signals that are used to derive the envelopes (the effect of which might be to artificially reduce STI). Although it appears to be unnecessary, it would be possible to compensate for a filterbank's own MTF, but it is certainly not practical to compensate for leakage between octave band filters. For a personal computer software implementation, there seems to be little reason not to use a robust filterbank, unless perhaps less selective filters are actually required for other types of impulse response analysis done concurrently by the software.

Another issue that arose was the distortion of MTFs that is introduced by remapping a linear distribution of modulation transfer function frequencies, which was present in three implementations. Although this distortion was visually and numerically obvious, the error introduced into the STI values was relatively small because positive and negative errors tended to balance each other. Generally there seems to be little reason to use a fast Fourier transform to derive the MTF (which would result in linear frequency distribution), but if one is used, then the error can be reduced by increasing the analysis period (e.g. by zero-padding the impulse response). Modelling indicates that an analysis period of several seconds makes the error negligible, and this seems to be the approach taken by implementation G. Perhaps a greater problem with a coarse linear frequency distribution is that the MTF itself is harder to interpret, but this could be ameliorated by plotting against the actual frequencies evaluated rather than against the nominal frequencies. Furthermore, a positive by-product of the linear distribution is likely to be that more detail is available at higher modulation frequencies – which could be returned to the user (even if it is not used for STI calculations).

Remarkably, a majority of the implementations tested had an error in the masking calculation. Only three (including A, by the authors) correctly accounted for auditory masking. The effect of the error in four of the implementations is that STI values can be higher than they should be, especially when the signal and/or noise spectra vary markedly between adjacent octaves. Although the tested spectrum of SN3 is artificial (and such contrasts in octave band speech level are unlikely to be tolerated in a real situation), it is not uncommon for masking to contribute significantly to the STI calculation in everyday situations. The masking in SN3 was only from the speech signal with the 250, 4000 and 8000 Hz bands being affected by the speech levels in the band below, but in more realistic situations masking could also come from background noise. While Annex M of the Edition 4 standard provides data that can be used for validation, it is likely that most of the code in the implementations with the masking error was written prior to Edition 4, although all have apparently been updated since the publication of Edition 4. Conceivably the masking error could have propagated if the software developers used each other's implementations for validation.

Although errors in threshold calculation are evident in some of the implementations, typical situations in which speech intelligibility is important are unlikely to have speech spectra near the auditory threshold (except perhaps in the lowest and highest bands). While errors in threshold adjustment are unlikely to affect most practical measurements, the adjustment is a simple operation to implement, so there seems to be little reason not to include it as described by the standard.

This study highlights the need for software to be validated, and it identifies some ways by which Indirect Method STI software can be validated. Although impulse responses consisting of exponentially decaying octave-spaced sinusoids are highly unrealistic, their MTF values can be exactly predicted from theory, allowing the first part of any implementation to be validated. The authors also tested the implementations with a pseudo-random repeatable 'white noise' waveform using an algorithm in [13] with an exactly known reverberation time (i.e. an exponentially decaying envelope), although this was not included in the data presented in this paper. As a test stimulus, this waveform had the advantage of providing energy over the full range of each octave band (and beyond), but the disadvantage of a fluctuating short-term envelope that made the MTF values deviate from their theoretically calculated values. This advantage did not provide any additional insights into the performance of the implementations, and the disadvantage made the results harder to interpret. There are many ways by which the derivation of MTFs could be further tested through synthetic impulse responses – for example, their robustness to the presence of substantial low frequency energy (below the 125 Hz octave band, although arguably noise in the 63 Hz octave band should also be used to contribute to the masking calculation), their response to sparse impulse responses (e.g., with strong echo-like features), or their vulnerability to leakage from energy near the upper and lower limits of the octave bands.

Validation of the second part of the STI calculation (deriving the STI from the MTF, taking signal and noise levels into account) is currently possible using Annex M of the Edition 4 standard. However, that is not the main purpose of Annex M – instead it provides a worked example of the way that the effects of known signal and noise spectra can be first removed from an MTF, and then new signal and noise data applied.

Hence the second half of Annex M provides the worked example needed to validate software implementations, but this might not be clear to a software writer. Considering the errors that were found in a majority of the software implementations, it would be helpful to have a worked example that simply showed how to derive the STI from the MTF and signal and noise spectra.

This paper is concerned with implementations that were current in the year 2013, and it is hoped that many of the anomalies found will be fixed in current or future versions of implementations.

## 5. CONCLUSION

This paper has found some unnecessary vulnerabilities in a number of commercial and freeware implementations of the Indirect Method of measuring the STI. Deviations between implementations are due to: (i) the design of the octave band filter-bank; (ii) the way in which the MTF is calculated; (iii) perhaps the edition of the standard implemented (although it should not have caused deviations in the current study); (iv) errors in implementation of masking and/or threshold; (v) the capacity of the user interface to allow the input of required data for calculation; and (vi) further unknown issues. Almost all implementations (from 2013) performed well when non-extreme inputs were tested, but it is possible to make an implementation robust to extreme inputs (which the authors have done in their implementation), thereby minimising the risk of significant error in STI calculation.

## 6. NOTE ON THE AUTHORS' IMPLEMENTATION

The authors' implementation of the Indirect Method for STI calculation is part of a suite of MATLAB functions for audio system and room acoustics measurement, known as AARAE [14]. The results reported here are from version 1.09 of the function STI_IR, within the AARAE project. AARAE is freely available from http://aarae.org.

## REFERENCES

[1] Katz, B.F.G., International round robin on room acoustical impulse response analysis software 2004, *Acoustics Research Letters Online (ARLO)*, 2004, 5(4), 158-164.

[2] IEC 50268-16 Ed. 4, Sound System Equipment – Part 16: Objective Rating of Speech Intelligibility by Speech Transmission Index, International Electrotechnical Commission 2011.

[3] ISO 3382-3: 2013: Acoustics – Measurement of room acoustic parameters – Part 3: Open plan offices, International Organization for Standardization, 2013.

[4] AS 1670.4-2004: Fire detection, warning, control and intercom systems – System design, installation and commissioning – Sound systems and intercom systems for emergency purposes, Standards Australia, 2004.

[5] van Wijngaarden, S., Verhave, J. and Steeneken, H., The speech transmission index after four decades of development, *Acoustics Australia*, 2012, 40(4), 134-138.

[6] IEC 50268-16 Ed. 3, Sound System Equipment – Part 16: Objective Rating of Speech Intelligibility by Speech Transmission Index, International Electrotechnical Commission 2003.

[7] Schroeder, M., Modulation transfer functions: Definition and measurement, *Acustica*, 1981, 49(3), 179-182.

[8] Venturi, A., Farina, A. and Tronchin, L., On the effects of pre-processing of impulse responses in the evaluation of acoustic parameters on room acoustics, *Journal of the Acoustical Society of America*, 2013, 133(5), 3224 (POMA 19, 015006).

[9] Guski, M. and Vorländer, M., Comparison of noise compensation methods for room acoustic impulse response evaluations, *Acta Acustica united with Acustica*, 2014, 100, 320-327.

[10] Cabrera, D., Lee, D., Collins, R., Hartmann, B., Martens, W.L. and Sato, H., Variation in oral-binaural room impulse responses for horizontal rotations of a head and torso simulator, *Building Acoustics*, 2011, 18(1-2), 227-252.

[11] Chu, W.T. and Warnock, A.C.C., Detailed directivity of sound fields around human talkers. Institute for Research in Construction, National Research Council of Canada, Ottawa, ON, Canada, Technical Report IRC-RR-104, 2002.

[12] IEC 61260 Ed. 1, Electroacoustics - Octave-band and fractional-octave-band filters, International Electrotechnical Commission 1995.

[13] Park S. K. and Miller K. W., Random number generators: Good ones are hard to find, *Communications of the ACM*, 1988, 31, 1192-1201.

[14] Cabrera, D., Jimenez, D. and Martens, W.L., Audio and Acoustical Response Analysis Environment (AARAE): a tool to support education and research in acoustics, *Proceedings of Internoise*, Melbourne, Australia, 2014.