Audio Engineering Society

Convention Paper

Presented at the 143rd Convention
2017 October 18–21, New York, NY, USA

A novel procedure for direct-method measurement of the full-matrix Speech Transmission Index

Jan A. Verhave[1] and Sander J. van Wijngaarden[1]

[1] *Embedded acoustics BV, Ampèreweg 18, 2627 BG Delft, The Netherlands*

Correspondence should be addressed to Jan Verhave (jan@embeddedacoustics.com)

ABSTRACT

When measuring the Speech Transmission Index (STI), until now one had to choose between two alternatives: impulse-response based full STI measurements (indirect method), or measurements based on modulated STIPA signals (direct method). Limitations apply when using either method. A novel procedure is proposed to measure the full STI through the direct method. The procedure combines advantages of indirect full STI measurements and direct STIPA measurements, completing a full STI measurement in 63 seconds. Similar to STIPA, the test signal is simultaneously modulated with 2 modulation frequencies per octave band. However, a rotation scheme is applied that uses a different set of modulation frequencies during different stages of the measurement, ending up with a full matrix (7 octaves x 14 modulation frequencies).

1 Introduction

Since its introduction by Steeneken and Houtgast in the early 1970s [1], the Speech Transmission Index (STI) has gradually evolved into a widely applicable and commonly accepted method for measuring and predicting speech intelligibility [2]. Successive extensions and adaptations to the model have widened the scope of its applicability, and enabled it to predict the effects of almost every intelligibility-degrading factor that occurs in real-life speech communication, with a few notable exceptions (such as low-bandwidth digital voice compression) [3]. The concern that different implementations or interpretations of the STI model might result in different outcomes was removed by standardization efforts [4]. At the same time, the publication of successive versions of the STI standard added to the popularity of the STI and widened its user population.

The original hardware and software for measuring the STI took up to 15 minutes per measuring condition. A need was soon felt to shorten the measuring time. This led to the development of new versions of the STI, including STITEL and RASTI [5] (which are now obsolete) and STIPA [6]. STIPA, which is currently (de facto) the standard way of measuring the STI, takes only 18-25 seconds per condition. The reduction in measuring time is mostly achieved by sampling only a part of the Modulation Transfer Function (MTF) matrix on which the STI is based. In many cases, the effect on the accuracy of the STI result is negligible. As a result, STIPA quickly gained popularity while Full STI faded into the background.

The original Full STI implementation is no longer even commercially available, while STIPA is implemented by about a dozen vendors.

However, there are cases where STIPA is not considered an adequate replacement for Full STI. In particular, STIPA may be inaccurate when discrete echoes occur, or when highly irregular frequency transfer functions occur in combination with strong reverberation. The only alternative currently available is so-called "indirect" estimate of the Full STI. Indirect STI estimates are obtained by measuring the impulse response of a room (or transmission channel), from which the MTF is computed. Room impulse responses can be measured quickly, and the Full STI is then easily calculated. Unfortunately, there are concerns with indirect STI estimates as well.

2 Concerns with STIPA

In order to understand the limitations of the STIPA and the indirect method, one must consider the role of the Modulation Transfer Function (MTF) as the central concept within the STI model.

The Modulation Transfer Function quantifies how the envelope spectrum of a signal is affected during transmission, by reduction of the modulation depth. It is expressed as a 7x14 matrix, comprising 7 octave bands (125 Hz – 8 kHz) times 14 modulation frequencies (0.63 Hz - 12.5 Hz). Each element of the matrix (sometimes called an m-value) expresses, on a 0-1 scale, the reduction in modulation depth that the intensity function of the signal is suffering. The MTF is measured by means of a test signal (the STI signal), that is 100% modulated. This signal is transmitted through the channel under test, after which the modulation depth is measured for each octave band and each modulation frequency. The reduction in modulation depth is the value of the MTF.

STIPA does not sample the full MTF matrix, but only 2 modulations per octave band (14 modulations in total, all applied simultaneously). A different pair of modulation frequencies is used in each octave band, which means that all 14 modulation frequencies are still considered, just not in every band. The accuracy is not adversely affected, unless (1) the modulation frequency dependent behaviour differs across bands, while not all bands contribute equally, or (2) some modulation frequencies are strongly affected in certain bands, whereas others are not. The first condition may occur with irregular frequency transfer functions in combination with strong reverberation, the second in case of echoes. In practice, the primary concern is with echoes. These occur in especially large venues (such as cathedrals), but also in PA systems with multiple loudspeakers (unless properly synchronized to compensate for propagation times by means of delay lines).

Figure 1. Computation of the Modulation Transfer Function matrix (after Steeneken and Houtgast [3]).

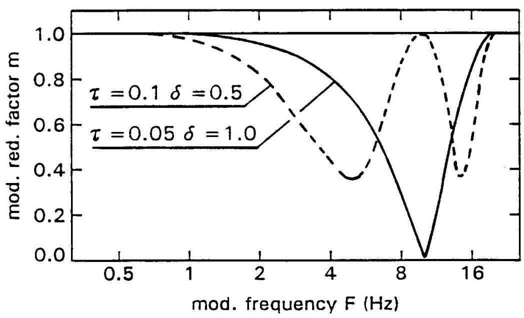

Figure 2. Impact of an echo on the MTF (after Steeneken and Houtgast [3]).

The effect of an echo on the MTF is to selectively cancel out specific modulation frequencies. Undersampling the modulation frequency axis (as STIPA does) means that the impact of the echoes on the MTF may be missed.

3 Concerns with the indirect method for estimating the STI

When implementing the indirect method for measuring the STI, it makes no sense to work with sparse versions of the MTF matrix such as STIPA. When the MTF is derived computationally from the impulse response, it inherently covers the full modulation frequency range. Indirect STI measurements are therefore always full STI measurements; concerns do exist, but these are separate from the concerns when using STIPA.

The STI originally uses modulated test signals for good reasons. These test signals resemble speech (in a physical sense) with regard to all parameters relevant to the MTF. If the channel under test comprises components that exhibit non-linear behaviour, these components will respond to the modulated test signal in a similar fashion as they would respond to speech (which contains the same range of modulations).

If the MTF is derived from an impulse response, the way in which the impulse response was measured determines how non-linear behaviour of the channel under test is dealt with. Most commonly, impulse responses are measured with swept sines or maximum-length sequences. Both types of test signals are usually short in duration; neither resemble speech in terms of intensity modulation content. Regardless of the accuracy with which the impulse response is measured, and the MTF is subsequently calculated, this procedure implicitly assumes that the channel under test is a linear time-invariant (LTI) system.

Public Address systems and Voice Evacuation systems, to which the STI is often apply, do not satisfy the requirements for an LTI system. Compression, automatic gain control and other common components of these systems will have an impact on speech intelligibility that is registered with original (direct) full STI method, but not properly dealt with when using the indirect method.

Another concern is the influence of background noise, which is automatically dealt with correctly when using the direct method. With the indirect method, the influence of background noise is suppressed by the measurement procedure. The noise spectrum needs to be measured separately and explicitly added in by means of post-hoc calculations. Software for indirect STI measurement is sometimes unclear and inconsistent in the way background noise is incorporated.

As a final concern, many software packages do not compute the MTF directly from the impulse response through the established mathematical relationship proven by Schroeder [7], but rather by an approximation based on the reverberation times computed from the impulse response (following a procedure introduced by Houtgast [8]). In fact, this is a “double indirect” method, which is only valid when adopting additional assumptions (such as purely linear energy decay).

Based on the above considerations, it is fair to say that use of the indirect method requires in-depth knowledge of the underlying principles of the STI model, and is not recommended for non-specialists. Even when executed properly, the indirect method cannot be used for a large category of real-life (non-LTI) transmission channels.

4 General approach towards developing a novel Full STI

Observing the shortcomings of indirect STI and STIPA, a need was felt by some, including the authors of this paper, to re-introduce Full STI. Although the original Full STI implementation has not been commercially available for many years, one option would be to simply re-implement it as it was before, but this time on the current generation of acoustical analyzers. This is not practical for two reasons: (1) the long duration of the original Full STI was the whole reason why it became disused in the first place, so any practically meaningful implementation will need to be considerably faster; (2) the original Full STI requires synchronization between the test signal source and the analyser, which is not needed for STIPA, and probably unacceptable to current STI users.

To elaborate on the latter point: the original Full STI method went through 98 consecutive stages of the measurement; each element of the MTF-matrix was evaluated separately. This means that the STI analyser must go through these stages in sync with the STI test signal source. However, the current practice is that a STIPA signal is simply generated continuously (or looped indefinitely). Measurements can be started and stopped at will, at any moment, without synchronization concerns. The current generation of STI test equipment does not even comprise a trigger line or data link between signal source and analyser that could be used for synchronization.

In short, in order to be successful a new Full STI method needs to be considerably faster and must not require explicit synchronization between test signal source and analyser.

The general approach towards the novel Full STI is to use a modulation scheme similar to STIPA, with two simultaneous modulation frequencies per octave band, but to rotate through a series of 7 stages in order to fill the entire MTF matrix. Synchronization does not take place between analyser and *generator*, as was the case with the original Full STI. Instead, the analyser synchronizes to the beginning of each of the 7 stages through a cross-correlation technique.

It was found that increasing the number of simultaneous modulations per band to more than 2 is inefficient; the length of the signal would have to be increased disproportionally to compensate for the additional statistical fluctuations.

5 The modulation scheme for the novel Full STI

When adhering to 2 simultaneous modulation frequencies per octave, the complete signal consists of 7 stages. Each individual stage is similar to STIPA. In fact, we choose to have one of the stages to be identical to STIPA, while the other stages are derived from the original STIPA through a rotation scheme. The original STIPA signal uses 7 base modulation frequencies: [0.63 0.80 1.00 1.25 1.60 2.00 2.50] Hz. The other modulation frequencies are multiples of 5 of the base frequencies. Each band simultaneously contains a base modulation and its multiple (1:5), the modulation depth of each modulation being 0.55 and the two modulations being 180 degrees out of phase. In mathematical form, the function for the modulator *m(φ),* with φ as the base modulation frequency, is as follows:

$$m(\varphi) = \sqrt{\frac{1}{2}(1 + 0.55[\sin(\varphi) - \sin(5\varphi)])} \quad (1)$$

The assignment of the modulation frequencies to octave bands within STIPA is given in Fig. 3.

	125 Hz	250 Hz	500 Hz	1 kHz	2 kHz	4 kHz	8 kHz
0.63			x				
0.8						x	
1.0		x					
1.25					x		
1.6	x						
2.0				x			
2.5							x
3.15			x				
4.0						x	
5.0		x					
6.25					x		
8.0	x						
10.0				x			
12.5							x

Figure 3. Assignment of modulation frequencies to octave bands in standard STIPA. All combinations of modulation frequencies and octave bands are applied simultaneously.

The original STIPA signal is the first of 7 stages of the new full STI signal, from which the other 6 stages are derived. The rotation scheme ensures that all modulation frequencies have been applied in all octave bands after 7 stages. Multiple alternative rotation schemes are possible, and have been evaluated. The rotation scheme may have an effect on the outcome of a full STI test if it is ended prematurely (before completion of all 7 stages). The optimal rotation scheme makes sure that the modulation frequencies are balanced across bands as much as possible at the end of each stage. It was determined that the optimum rotation scheme shifts up 4 places in the base modulation vector (modulo 7) after each stage. For example: in the 500 Hz band, the first stage starts with the (standard STIPA) base modulation frequency of 0.63 Hz. For the second stage, it shifts the base frequency up by 4 steps, making the base modulation frequency 1.6 Hz. For the third stage, it goes up another 4 steps (modulo 7), arriving at 0.8 Hz. This operation is applied for all bands, for all 7 stages. The rotation scheme is summarized in Fig. 4.

	125	250	500	1k	2k	4k	8k
0.63 Hz	7	4	1	5	2	6	3
0.80 Hz	2	6	3	7	4	1	5
1.00 Hz	4	1	5	2	6	3	7
1.25	6	3	7	4	1	5	2
1.60	1	5	2	6	3	7	4
2.00	3	7	4	1	5	2	6
2.50	5	2	6	3	7	4	1

Figure 4. Rotation scheme showing the order in which the base modulation frequencies are applied to each band. The original STIPA signal (stage 1) is marked in red. The second modulation frequency per octave band is always 5x the base frequency.

The 7 stages of the signal are, in principle, 7 independent STIPA-like signals, that fill the same space as full STI when combined. However, the design of the new Full STI signal covers more than just the rotation scheme. For instance, the transitions between the stages also need consideration; the stages need to be spliced together so that the envelope function is seamless in each of the octave bands. This requires careful design of the modulation function, making sure that no audible (and measurable) discontinuities occur when transition between one modulation frequency and the next. This transition phase needs to be as short as possible, since it adds to the overall measurement time.

6 Measuring time

The total measuring time will be 7x the time allocated for each individual stage. If each stage were to be made as long as an regular STIPA measurement (for which the standard requires a minimum of 15s), this would mean that a full STI measurement would take 1 minute and 45 seconds. This is already much faster than the original STI. In practice, we can reduce the measuring time further; the individual elements of the MTF matrix will be determined with less accuracy than with STIPA, but this is offset by the advantage of filling the complete matrix (hence having many more independent m-value estimates).

It would be unwise to choose the length of the measuring stage arbitrarily, because of constraints associated with the structure of the test signal. Ideally, each stage would contain an exact multiple of all signal modulation periods, which would allow for a perfect splice between stages. In practice, this would make the signal about as long as the original Full STI signal. We are therefore allowing some slack around the transitions between stages, during which the test signal is generated, but results not included in the analysis (to prevent splicing artifacts from influencing the result). The slack needed to keep the influence of splicing artifacts down to an acceptable mimimum depends on the measuring time. A duration of 9s per stage was found to be the optimal compromise. This means that the total Full STI measuring time is 63 seconds.

7 Synchronization

The STI analyser needs to synchronize to the signal, making sure that it is aligned properly to the current measuring stage. A measurement can be arbitrarily started at any of the 7 stages, at any moment within that stage. In practice, data needs to be discarded for the first (partial) stage that is analysed; the analysis truly starts at the beginning of the first complete stage.

The synchronization process is based on a correlator function, which is similar to the modulator (Eq.1). Hence, the synchronization process takes place in the same domain as the computation of the MTF.
Depending on the current stage of the signal, the correlator will peak for a different set of modulation frequencies. Even if the test signal structure is degraded by processing through very poor channels (0.15<STI<0.30), the correlator still shows sufficiently sharp transitions in its response when going from one stage to the next, that the beginning of the stage is known with sufficient accuracy.

8 Differences between the original full STI and the novel implementation

The original Full STI and our novel implementation are identical in terms of the evaluated frequency bands and modulation frequencies. The way in which the MTF is calculated is also identical, as is the computation of the STI from the MTF.

Apart from the aspects mentioned before (faster measuring time and the need for synchronization to the signal), there is one other difference. The original STI explicitly introduced random modulations in all octave bands, except for the octave band currently under test. In the novel Full STI all bands are tested at the same time, in the same way as STIPA. The reason for introducing random modulations in the original STI was that (in case of nonlinear distortion) distortion products will spill over from adjacent bands into the band under tests, leading to a representative reduction in modulation depth. In STIPA, it is the signal itself that is spilling over; but since the modulation frequencies are different in each band, contributions from adjacent bands are still recognized as "noise" rather than signal, and result in a penalty rather than a bonus. The novel Full STI works in the same way as STIPA. The results from the original STIPA validation study were found to also apply to the new method.

9 Implementation results

A full validation study, a full description of which is beyond the scope of this paper, was carried out in order to ensure that the accuracy, precision and stability of the novel Full STI are (at least) equal to STIPA and the original full STI throughout a series of reference conditions.

The novel Full STI was first implemented in Matlab. This Matlab implementation of Full STI was tested with the standard series of sine-based validation test vector, and passed. Next, a real-time implementation of the Full STI was developed for the Bedrock SMxx series of acoustic analyzers.

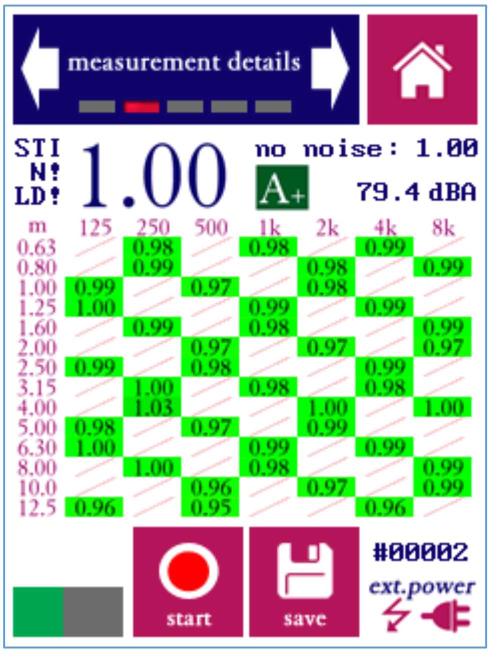

Figure 5. Full STI implementation on the Bedrock SMxx, while measuring. The analyser is synchronized and in the second stage of the signal.

The measurement takes up to 9 seconds to synchronize to the test signal, after which the MTF matrix is gradually filled as each stage of the signal is completed. The current STI value is displayed and updated as the measurement progresses.

In many cases, echoes are not a concern; in those cases the full STI result is identical to the STIPA result, and the STI does not differ from one stage to the next. In those cases where STIPA is not suitable for use, fluctuations are observed between stages.

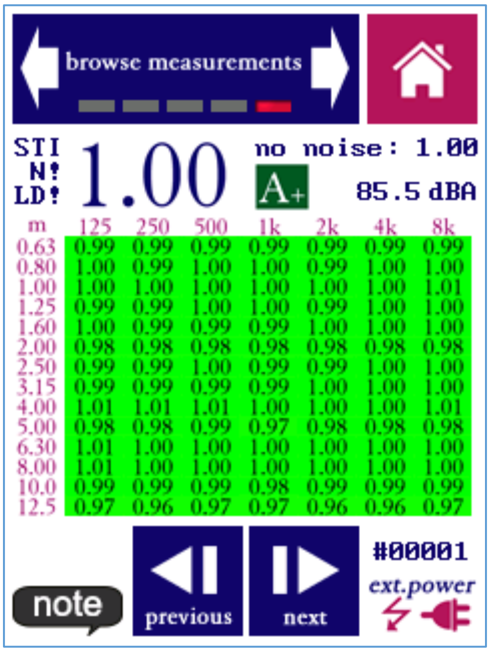

Figure 6. The full STI measurement has completed (in this case, for a perfect transmission channel). The MTF matrix is completely filled.

The full STI app was beta-tested between April and June 2017, and released as part of a firmware update to all users of the Bedrock SM50 and SM90 in June 2017. The feedback from the user community has been very positive. The fact that a Full MTF matrix is available not only means that the result can be trusted in cases where STIPA cannot; it also means that more diagnostic data is available when troubleshooting a transmission channel.

10 Conclusions

A novel version of the Full STI was implemented, that has the same scope of applicability and reliability as the original STI, while allowing for much shorter measuring times. The novel Full STI is similar in its use as STIPA, but is also fully applicable in the presence of echoes and harsh reverberation. It is a direct method, which means that it can be used on channels that feature non-linear or time-variant characteristics. The novel method was implemented on a commercial measuring device.

References

[1] Houtgast, T. and Steeneken, H.J.M. (1971). "Evaluation of Speech Transmission Channels by Using Artificial Signals", *Acustica* 25, 355-367.

[2] Steeneken, H. J., van Wijngaarden, S. J., & Verhave, J. A. (2011, May). "The Evolution of the Speech Transmission Index." In: *Audio Engineering Society Convention 130*. Audio Engineering Society.

[3] S.J. van Wijngaarden (Ed., 2002). "Past, present and future of the Speech Transmission Index." *Soesterberg: TNO*.

[4] International Electrotechnical Commission (2011). IEC-60268-16 rev. 4 "Sound system equipment-Part 16: Objective rating of speech intelligibility by speech transmission index". *International Electrotechnical Commission*.

[5] Houtgast, T., & Steeneken, H. J. M. (1984). "A multi-language evaluation of the RASTI-method for estimating speech intelligibility in auditoria.`' *Acta Acustica united with Acustica*, 54(4), 185-199.

[6] van Wijngaarden, S., Verhave, J., & Steeneken, H. (2012). "The Speech Transmission Index after four decades of development". A*coustics Australia*, 40(2).

[7] Schroeder, M. R. (1981). `'Modulation transfer functions: Definition and measurement." *Acta Acustica united with Acustica*, *49*(3), 179-182.

[8] Hou Houtgast, T., Steeneken, H. J., & Plomp, R. (1980). "Predicting speech intelligibility in rooms from the modulation transfer function. I. General room acoustics." *Acta Acustica united with Acustica*, *46*(1), 60-72.