SPEECH TRANSMISSION INDEX: TOO WEAK IN TIME AND FREQUENCY ?

T Steinbrecher Bose Professional Systems, Friedrichsdorf, Germany

1 INTRODUCTION

A method widely used in past years for the evaluation of the speech intelligibility of sound systems, is the Speech Transmission Index STI. The STI and its derivatives (today, the STIPA-method is often applied) are specified in EN IEC 60268-16 [1], with the most recent changes being made in 2003 and with the standard currently being under regular maintenance. Both the simulation and the measurement of speech intelligibility have gained increased importance within the context of standardizing the performance of voice alarm systems. Thus, consultants, sound system designers and measurement engineers that apply the STI, are naturally interested in the ability of the method to properly account for the various types of distortion that are typically associated with such systems.

The STI is based on the research of Steeneken and Houtgast (see e.g. [2]) and allows the objective evaluation of a transmission channel that is compromised by various distortions in the time and frequency domain. By implementation of a simplified model, it also accounts for level-dependant upward-masking of frequencies, an important component of the human auditory system. The STI has been validated with subjective intelligibility tests for a large number of test conditions [3]. The derivation of the STI is based on determining and post-processing the Modulation Transfer Function (MTF) of the transmission channel. The measurement of the MTF can either be performed in an analogue fashion (by applying modulated noise signals) or by Fourier transform of the system's squared impulse response ('Schroeder-method', [4]). For the simulation of the STI, it is recommended to apply the latter method, since the representation of non-statistical properties of the sound-field (e.g. an echo) is only possible in a fully simulated impulse response.

In the past years, two main problems have often been associated with the STI: the relatively low and irregular sensitivity against certain distortions in the time domain (especially late-arriving echoes) as well as against aberrations in the frequency response of the transmission channel. A detailed analysis of both effects is presented in this paper. First, section 2 discusses the consequences of the time-symmetry of the STI-MTF concept with a focus on echo distortions and the ratio of early (typically useful) to late (typically detrimental) energy. In section 3, the results of a systematic experiment are presented, where the frequency response of a communication channel including background noise was altered by filtering and band-pass limiting and the influence on the STI was evaluated. Workarounds are proposed to overcome some aspects of the described limitations.

2 TIME DOMAIN DISTORTIONS

2.1 General

In general, the STI-MTF method is sensitive to distortions in the time domain with the most well-known behavior being the response to a transmission channel that employs exponential reverberation. Here, the MTF shows a distinctive low-pass character with higher modulation frequencies (faster modulations) being more strongly reduced in modulation than the lower modulation frequencies. It has been known for quite some time and a variety of examples have been documented where the STI method fails to correctly represent the influence of echoes (another time-domain distortion) on speech intelligibility (see e.g [5]).

Engineers and researchers who apply the STI method under typical room acoustical conditions, no matter whether the source is natural or electronically amplified, have thus been concerned about the general applicability of the method to properly distinguish between early and late energy. Therefore, this section attempts to analyze the matter by taking a detailed look at some properties of the MTF when distorted by single echoes and through evaluation of more complex impulse responses that were simulated in a computer model. The latter analysis was performed in version 6.5 of the sound system simulation software Bose Modeler [6, 7].

2.2 Echoes

If the Schroeder-method is applied to a system that is characterized by an impulse and a delayed and attenuated repetition of that impulse after a time interval ΔT, the following relation between the modulus of the modulation m and the modulation frequency F, the delay time ΔT and the ratio δ of the intensities of direct sound (first arrival) and echo can be derived (Formula (1)):

$$m(F) = \frac{\sqrt{1 + 2\delta\cos(2\pi F\Delta T) + \delta^2}}{1+\delta} \qquad (1)$$

It can be seen from equation (1) that *m(F)* – due to the cos-term incorporated – is a function with periodic characteristics. The typical shape of an MTF which is distorted by an echo (and no other distortions present) is characterized by one or multiple notches, with the depth of the notch being determined by the level or intensity ratio of the first and the delayed arrival (see. Fig. 1). For an intensity ratio of 1 (same level of first and delayed arrival), the modulation will decrease to 0 for certain modulation frequencies. If the levels of the two arrivals are different, the notch in the MTF will not reach down to m=0. Under these circumstances, the MTF response takes the shape of one or more classic comb-filters.

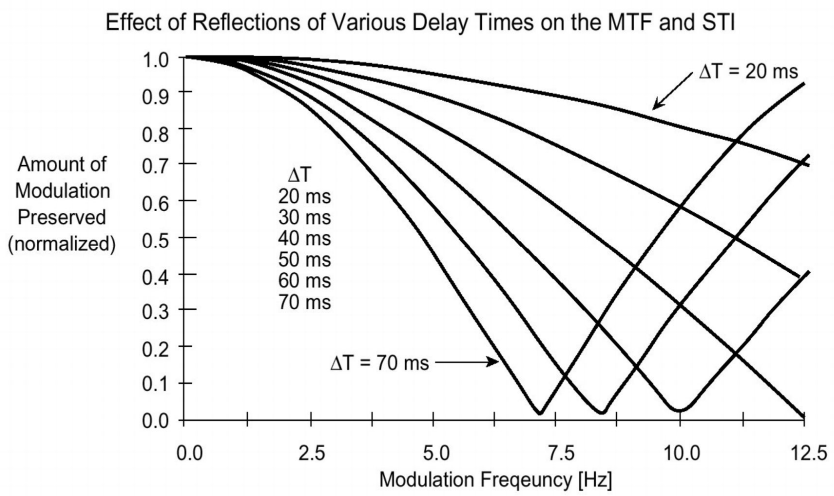

Figure 1: MTF distorted by a single echo for various delay-times. Here, the levels for first and delayed arrivals are identical.

It is important to note that for such simple transmission systems that basically only consist of two arrivals, each secondary arrival – no matter how short the delay time is – will always cause a reduction in modulation. For very short delay times up to a couple of milliseconds, this reduction appears very high in modulation frequency and more or less outside of the frequency range of interest for the STI standard which extends from 0.63 Hz to 12.5 Hz. For increasing delay times, the notch moves towards lower modulation frequencies and since the notch is repeating, eventually multiple notches will appear in the MTF as can be seen in Figure 2 for echo arrivals with delays of 350 and 400 ms:

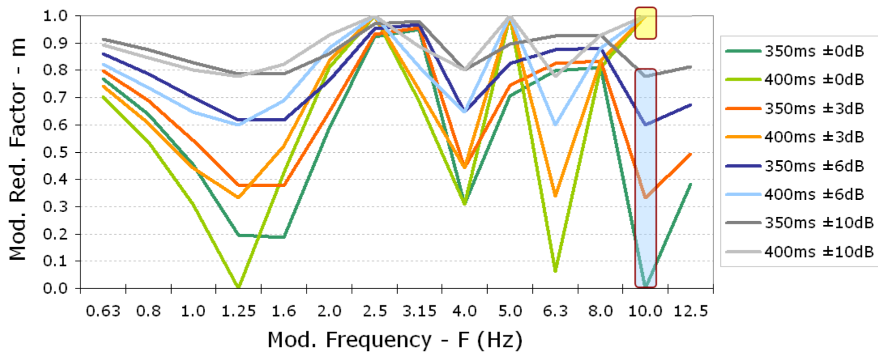

Figure 2: MTF distorted by a single echo at 350 and 400 ms of delay for various echo-levels.

The location of the notch within the modulation frequency spectrum is repeating periodically at uneven multiples of the inverse of twice the delay time. Here, the cos-term in (1) gets minimal (equal to -1). As an example, an echo with 100 ms arrival delay will cause a first notch in the MTF at 1/(200 ms) = 5Hz. Additional notches will then appear at 15Hz, 25 Hz, etc. It is important to note that for even multiples of 5Hz (10 Hz, 20 Hz, etc.), the modulation will become maximum (m=1). As a consequence, this periodicity has an equivalent effect on the STI, as shown in Fig. 3.

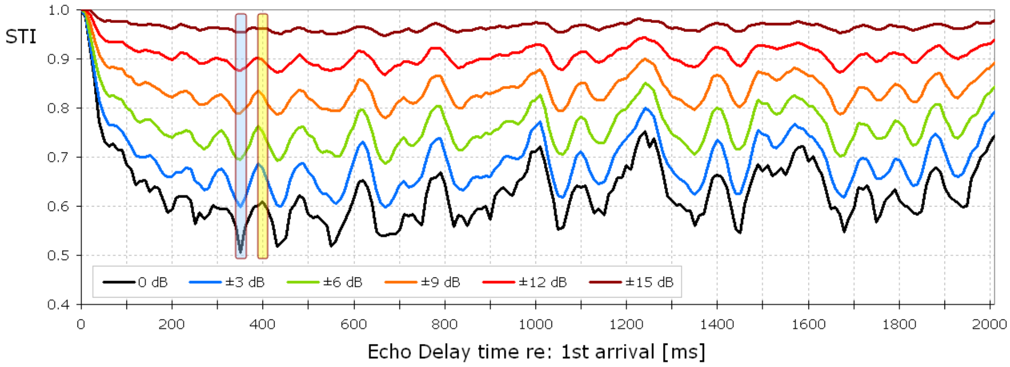

Figure 3: STI versus delay time of a single echo for various ratios of the levels of first arrival and echo. No other arrivals are present. The highlighted regions refer to Fig. 2.

For echoes with higher levels, the initial drop of STI with delays increasing from 0 up to 50 to 100 milliseconds is pronounced. Then, the STI starts to bottom out because the notch moves so far downward in modulation frequency that modulation will increase again for higher modulation frequencies. The response is then further characterized by repeating local minima and maxima, from which two neighboring regions are marked at 350 and 400 ms. Every 50 ms, the modulation index at F = 10 Hz alternates between minimum (blue) and maximum (yellow, see Fig. 2 for comparison). For all other modulation frequencies, this alternation repeats at different time intervals. The longer the delay time, the more modulation frequencies will reach a maximum in sync with others. Eventually, with a long enough delay time for the second arrival, the STI will become 1.0 again. It is obvious at first glance that this doesn't make sense from a psycho-acoustical perspective.

In general, it has to be regarded that the periodicity of the STI and the associated ripple in Fig. 3 is pronounced by the discrete sampling of the MTF in third-octave band resolution. The periodicity is further emphasized by the fact that some of the standardized modulation frequencies (see Fig. 2) are multiples of each other. In order to magnify this latter effect, Fig. 3 has been calculated by applying non-preferred frequencies F (0.625 Hz, 0.8 Hz, 1.0 Hz and their exact multiples). If a Discrete Fourier Transform (DFT) is applied to calculate the MTF as required in the STI-standard and as applied in this exemplary case, the sampled modulation spectrum repeats after an additional delay time that is equivalent to the least common multiple of the period lengths of all modulation frequencies. For the frequencies used in Fig. 3, this additional time delay equals 40 seconds, so for two delay times of 0.1 and 40.1 seconds, the MTF is exactly identical. For the frequencies depicted in Fig. 2 (0.63 Hz, 0.8 Hz, etc.) this time delay equals 100 seconds. While an echo with such a long delay time has no practical meaning in acoustics, this extreme case may illustrate that at certain points in time, many modulations are phase shifted by exact multiples of their respective period lengths and thus cause the peaks in the rippled response of Fig. 3.

Another complication arises from the symmetry of the cos-function (cos(x) = cos(-x)), which thus yields identical results for negative time delays. Accordingly, this is also applicable to the intensity ratio δ: If the levels of direct sound and echo are reversed, the MTF and thus the STI don't change. It is only the delta in time and the delta in level that determines the MTF-shape, not the specific order of the arrivals.

Apparently, there is no equivalent functionality within the human hearing system that would represent the benefit of a periodically increasing modulation for large delay times. A modulation shifted by several periods in time does certainly not improve speech intelligibility. Also, the time- or level-reversal of echo and direct sound typically results in strong subjective differences due to the ability of the hearing system to lock onto strong early arrivals (and employing temporal masking).

In this context, it may be worth noting that the subjective validation of the STI as performed by Steeneken and Houtgast, actually yielded a slight underestimation of intelligibility via the STI-method for the echo test conditions that employed high signal-to-noise ratios and large arrival delays. The echoes analyzed had arrival delays of 50, 100 and 200 ms and a relative level of -3dB. This somehow surprising result may have been a consequence of the fact that the parameters of the STI have been optimized on the basis of subjective validation tests where the intelligibility was determined on the basis of single CVC-nonsense-words embedded in carrier sentences rather than sentence intelligibility tests. Under the influence of echo-distortions, it may be easier to pick up a single CVC word out of an identical and well-known carrier phrase.

Given the blindness of the STI-MTF concept to time- and level reversals in simple echo-only transmission channels, it is of special importance to know about the general ability of the STI-method to correctly represent the early-to-late energy ratio with regards to intelligibility. In the next section, simulated impulse responses will be analyzed for the beneficial and detrimental influences of discrete secondary arrivals in an otherwise constant and semi-complex sound field including reverberation and noise.

2.3 Single arrivals in a semi- complex sound field

To evaluate the influence of single arrival in an otherwise constant sound field, a series of controlled experiments was conducted in the Bose Modeler Design Program. The program predicts the STI based on an MTF that is derived from a simulated impulse response by the 'Schroeder-method'. The MTF is post-processed by applying the 2003 version of the standard. Thus, absolute signal and noise levels in each frequency band as well as the reception threshold are correctly taken into account.

The test model had a basic shoebox shape with a room volume of 2100 cubic meters. All surfaces carried the same absorption properties and were marked as totally diffuse. A source with the radiation characteristics of a human talker as specified in the STI standard (here: the NTI Talkbox) was placed in the model together with a monaural receiver on axis of the speaker and at a distance of 4m. The source level was adjusted to "Lombard" level, which is 10 dB higher than the "normal" speech level as specified in the STI standard. The raised speech level equals 70 dB(A) broadband at a distance from the talker of 1m. The direct field component of the talker response was equalized to flat at the receiver, resulting in a broadband speech level of 58 dB (A) at a distance from the talker of 4m. By changing the surface absorption, the reverberation time was adjusted to 1, 2 and 3 seconds, which very roughly results in a critical distance that is – given the standardized directivity of the talker-source – in the range of one to two meters around the selected receiver location. The model allows the addition of arbitrary as well as standardized noise spectra. For this analysis, the STI's were predicted in a noiseless model as well as with an ANSI NCB 35 noise rating.

With no other discrete arrivals than the direct component, this described reference test setup results in a simulated (squared) impulse response that appears as follows at the receiver location:

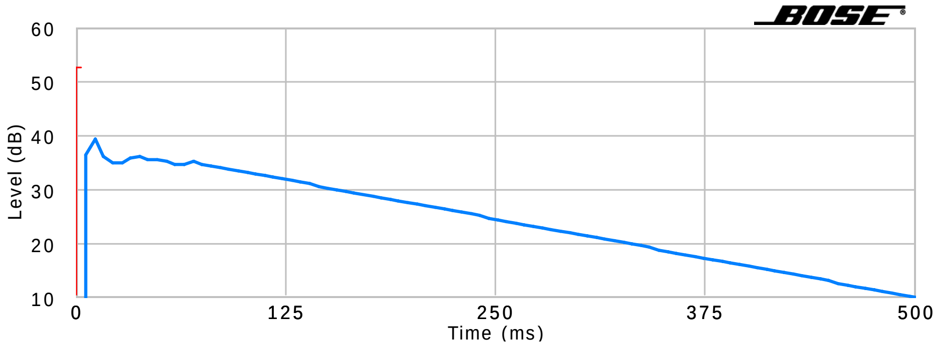

Figure 4: Simulated squared impulse response at the receiver location for the 1 kHz one-third-octave band with no secondary arrival in the 1 second room. Red pin: Direct arrival. Blue curve: Envelope function of buildup and decay of reverberant energy.

Note the quick buildup of diffuse energy through early floor reflections, the plateau caused by subsequently contributing diffuse reflections from other surfaces as well as the final exponential decay.

This reference condition yields STI values with fair to excellent ratings for the three test reverberation times in the case where no noise is present in the model except the noise intensity used to model the auditory reception threshold. For the three reference test conditions that include an NCB 35 background noise spectrum, the STI values had ratings of fair to good. The exact STI results are as follows:

Reverberation Time [s]	1	2	3
No noise	0.78	0.64	0.55
NCB 35	0.69	0.58	0.51

Table I: STI results for the reference conditions.

To evaluate the influence of a single early reflection varying in time and level, a second sound source was placed in very close proximity to the receiver. Due to the close proximity, the second source is driven at less than -25dB level compared to the primary talker while achieving the same direct SPL at the receiver. Thus, the influence of the second source on the level of the diffuse field can be neglected. This is also visible in Figure 5, where a second arrival with a relative delay of 50 ms and an equal level as the direct component is added: the envelope function for the diffuse sound is unaffected.

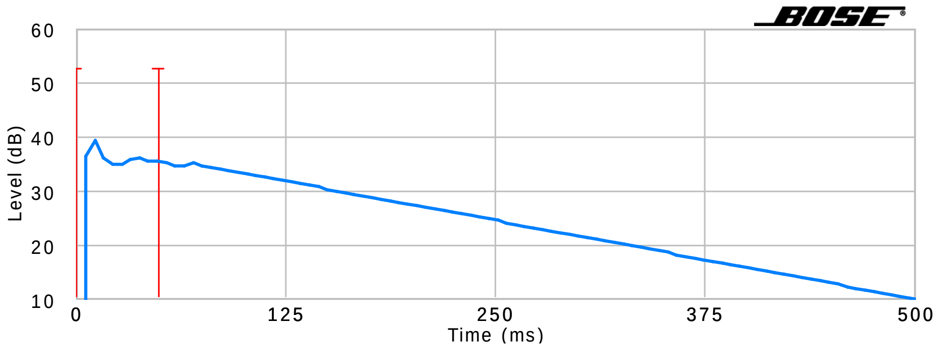

Figure 5: Simulated squared impulse response at the receiver location for the 1 kHz third-octave band with one secondary arrival of equal level at 50 ms delay. Note the unaltered reverberant envelope (blue) when compared to the case shown in Fig. 4.

For the subsequent analysis depicted in Figures 6 and 7, the relative delay of the secondary arrival was changed in 10 ms steps from 0 ms to 150 ms while the level was changed in 3 dB steps from -9 to +9 dB relative to the level of the direct arrival.

Figure 6 shows the STI values achieved for the various conditions. The left column (from top to bottom) represents the 1, 2 and 3 second rooms without additional background noise. The right column represents the same rooms including an NCB 35 background noise spectrum. NCB 35 is a common noise criterion for a variety of applications employing HVAC noise such as courtrooms, conference rooms, churches, classrooms and many other similar facilities [8,9].

The octave band noise values for NCB 35 in the speech frequency range are as follows, the summed level over all speech bands is 52 dB (-) and 43 dB (A):

Octave band [Hz]	125	250	500	1k	2k	4k	8k
NCB 35 [dB-SPL]	51	44	40	37	33	30	27

Table II: Octave band noise levels for NCB 35 noise criterion.

Figure 7 is a different representation of the same data: Here, the differences in STI to the reference conditions are displayed.

In each representation, the curves for the various levels of the second arrival are encoded in colors according to the legend. The curve for the reference condition is depicted in black and includes markers. The curves for the two highest arrival levels (+6 and +9 dB re: the direct arrival) are displayed with a dashed line, since these conditions will less likely appear in real world systems. Nonetheless, it should be noted in this context that while the analysis was performed for a single secondary arrival, multiple reflections with lower levels that arrive at roughly the same time will add up in energy and will cause a similar effect than a single arrival with equal energy.

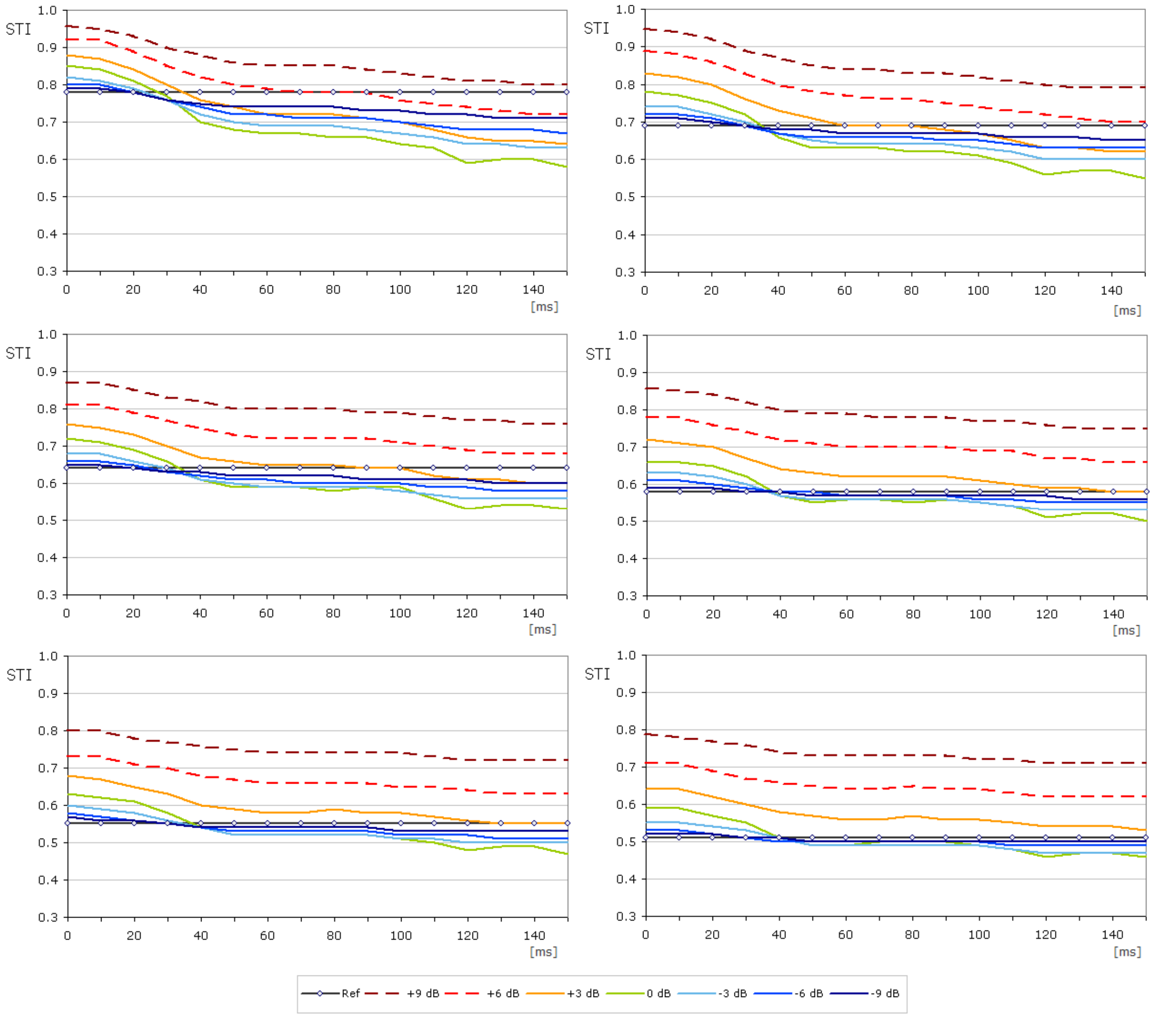

Figure 6: STI versus delay time and level of a single reflection in an otherwise constant sound field for reverberation times of 1, 2 and 3 seconds. Left column: without additional background noise. Right column: including NCB 35 noise. The dashed lines represent secondary arrival levels that may only seldom appear in practice.

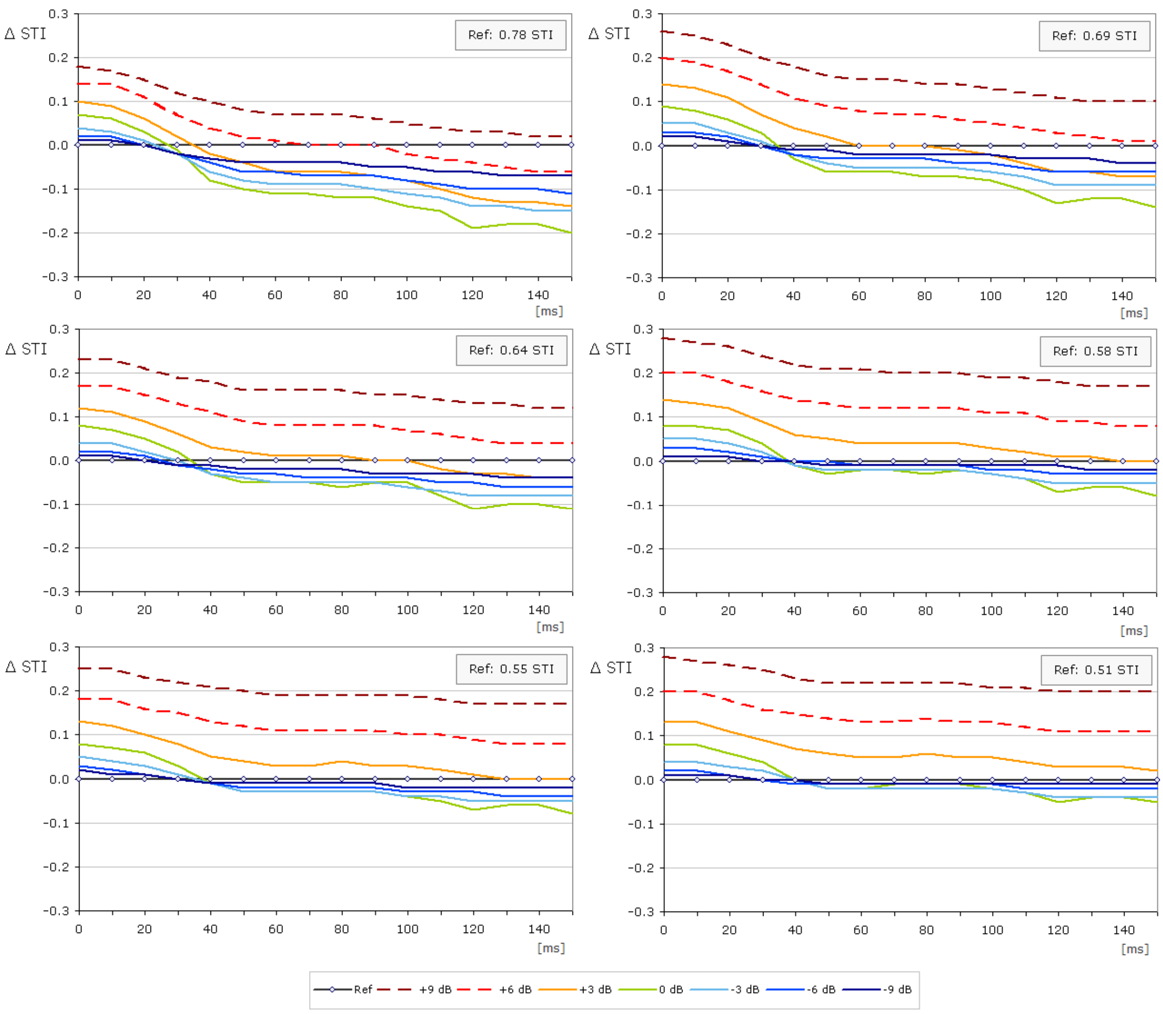

Figure 7: STI difference to reference condition (ΔSTI = 0.0) versus delay time and level of a single reflection in an otherwise constant sound field for reverberation times of 1, 2 and 3 seconds. Left column: without additional background noise. Right column: including NCB 35 noise.

Among others, several conclusions can be drawn from the results depicted in Figures 6 and 7:

- Early arrivals up to delays of roughly 30 ms will always improve the STI. This effect is due to the improved “early” level and is even noticeable with arrivals that are -6 or -9 dB down compared to the direct arrival. As an example, a secondary arrival with 20 ms delay at -3dB level will yield a typical improvement in STI of 0.03 to 0.05, (at a typical JND of 0.03).
- Typically, for later arrival times, the STI starts to deteriorate due to the increased echo-effect (notches appear in the respective MTF’s). Very strong secondary arrivals result in a high direct field level since the source for the secondary arrival is very close. In this case, the significantly improved direct to reverberant ratio partly suppresses the deterioration by the echo-effect. (Also, the time-reversal property describes in section 2.2 becomes effective).
- In practical sound fields as the one analyzed, the absolute level of the second arrival does matter. No symmetry as with the simple echo-only conditions can be observed.
- The beneficial effect of early secondary arrivals increases significantly when background noise is added since the additional arrival will also improve the signal-to-noise ratio.
- The beneficial effect of early secondary arrivals increases slightly with higher reverberation times.

From Figure 6, it appears that the detrimental effect of the background noise is less apparent for higher reverberation times, i.e. the differences in reference STI for the 1 second room are much larger than for the 3 second room. It should be noted in this context that, due to the higher reverberant field sound pressure level, the total SPL at the receiver location gets higher for the more reverberant rooms, e.g. the effective speech level in the 2 second room is roughly 2 dB higher then in the 1 second room, thus further improving the signal-to-noise ratio.

3 FREQUENCY DOMAIN DISTORTIONS

3.1 General

According to the STI-standard, the MTF – as the foundation of the STI – has to be determined for the octave bands from 125 Hz to 8 kHz. In this context, it should be noted that the calculation of the modulation loss in each octave band is performed independent of the other bands. The only exception is the modeling of the auditory upward-masking, which is level-dependent.

Since for all linear systems, the influence of other typical distortions like reverberation or echoes is independent of the frequency response[1], the only signal-level-dependent variable remaining in each band is the signal-to-noise ratio (SNR). Thus, in case measurements or simulations are performed under conditions with very low background noise, even systems with a very irregular or limited frequency response will still achieve very high STI ratings.

Under such conditions, the only remaining noise source in the STI model is the threshold of hearing (in STI terminology: Absolute Reception Threshold), which is modeled as a frequency-dependant noise intensity. With moderate presentation levels of speech, e.g. in an environment with an un-amplified talker, these noise intensities may already affect the SNR, especially at the high frequencies, where typical speech is significantly attenuated in level. It is therefore critically important that a realistic noise spectrum is always added to any STI prediction.

In practice, two typical situations occur with (electro-) acoustic systems: Often, the STI is calculated or post-processed in a predictive program (e.g. a spreadsheet) that excludes the absolute reception threshold and the user may also omit the specification of noise in the actual transmission channel resulting in potentially infinite signal-to-noise ratio for each frequency band.

As a second example, typically applicable when modern impulse response based measurement technology is used, the user may lose information about the actual (absolute) speech and noise levels in each octave band. Typically, the excitation spectrum of sweep- or MLS-based measurements is not speech-shaped and averaging methods may be used to intentionally suppress the noise floor. Typically, this leads to over-estimated signal levels and under-estimated masking effects, especially in the four highest frequency bands, where the speech spectrum is already attenuated by 6 to 25 dB compared to a pink spectrum (see also table III). Unfortunately, these are the speech bands with the highest contribution to the overall STI rating. While the application of the mentioned methods may be advantageous under certain conditions, such measurements may significantly over-estimate the STI unless the user applies appropriate corrections to the results. Fortunately, when using modulated noise signals and calibrated meters (STIPA), the absolute signal and apparent background noise levels as well as the resulting masking intensities are always properly included in the result.

[1] It should be noted that throughout this paper, the term 'frequency response' is used, although this term actually describes the complex transfer function of a transmission system and therefore also includes the phase response. The analysis in this paper is restricted to changes of the modulus of the frequency response, which is also termed amplitude response.

3.2 Analysis of frequency responses

In order to evaluate the sensitivity of the STI with regards to distortions of the frequency response, the STI was calculated as a function of a nominal signal-to-noise ratio for six different attenuation spectra. No other distortions (e.g. reverberation) were present. The surveyed conditions include limitations of the transmission bandwidth that are partly representative of practical conditions in voice alarm systems, such as 300 Hz to 3.5 kHz analogue telephone bandwidth or a high-shelving filter that reduces energy only in the two upper frequency bands. Additionally, some more radical frequency responses were analyzed, e.g. a 15 dB cut in the important 2 kHz octave band or a 'small-bandwidth' transmission with signal only available in the 1 kHz octave band. Together with the standardized spectrum for male speech, all conditions are described as relative octave band levels in Table III and depicted in Figure 8.

	Octave Band level [dB rel]						
EQ	125	250	500	1k	2k	4k	8k
Flat	0	0	0	0	0	0	0
Knee 4k	0	0	0	0	0	-6	-12
TNO BP2	-40	-8	0	0	0	-4	-40
Dip 2k	0	0	0	0	-15	0	0
LP 500	0	0	0	-18	-36	-40	-40
BP 1k	-40	-40	-24	0	-24	-40	-40
Speech Male	+2.9	+2.9	-0.8	-6.8	-12.8	-18.8	-24.8

Table III: Octave band attenuation levels of spectra analyzed.

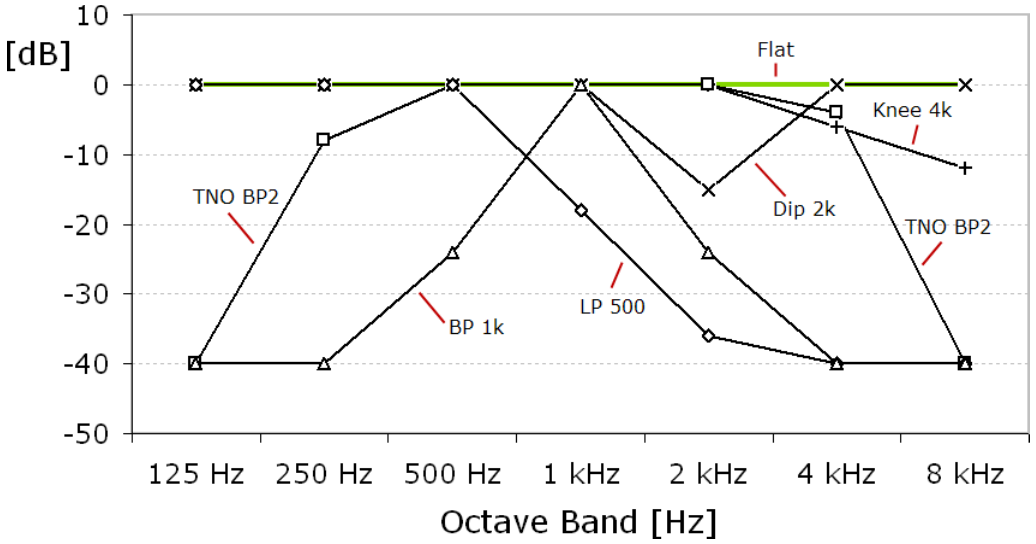

Figure 8: Frequency responses analyzed.

The analysis was performed as follows:

- A male speech spectrum was generated with the standardized relative octave-band levels and a broadband un-weighted speech level of 75 dB SPL. This speech level is within the SPL range that is not yet affected strongly by the upward-masking function of the STI.
- Noise was generated with the same spectrum as the speech signal. For identical octave-band levels, this yields the 0 dB SNR condition which is equivalent to an STI of 0.5.
- The overall level of the noise is adjusted to yield various SNR's in the range of -50 to +50 dB. E.g. with the noise at 45 dB SPL, the nominal SNR is 30 dB.
- The reference condition is a flat system response, where the speech signal passes without any attenuation, yielding an SNR independent of frequency.
- For the other conditions, the signal level was reduced in the applicable frequency bands while the noise was kept constant from the reference condition. Note that, depending on the frequency response applied, the total speech level will be reduced from the initial 75 dB SPL. This overall level reduction is strongly depending on the spectrum applied, e.g. a 500 Hz high-pass would remove much more energy from a speech signal than a low-pass at 4 or 8 kHz where the relative signal level is very low.
- There was no make-up gain applied to the speech signal in order to maintain the speech level at 75 dB SPL once a response with much less total SPL was applied. This condition is equivalent to a system where the overall gain cannot be further increased, e.g. because the system would otherwise run into acoustic feedback.

Figure 9 shows the results for nominal signal-to-noise ratios of -50 to +50 dB. An undistorted frequency response (the reference condition) yields the green curve, which is clipped at ± 15 dB SNR according to the STI standard.

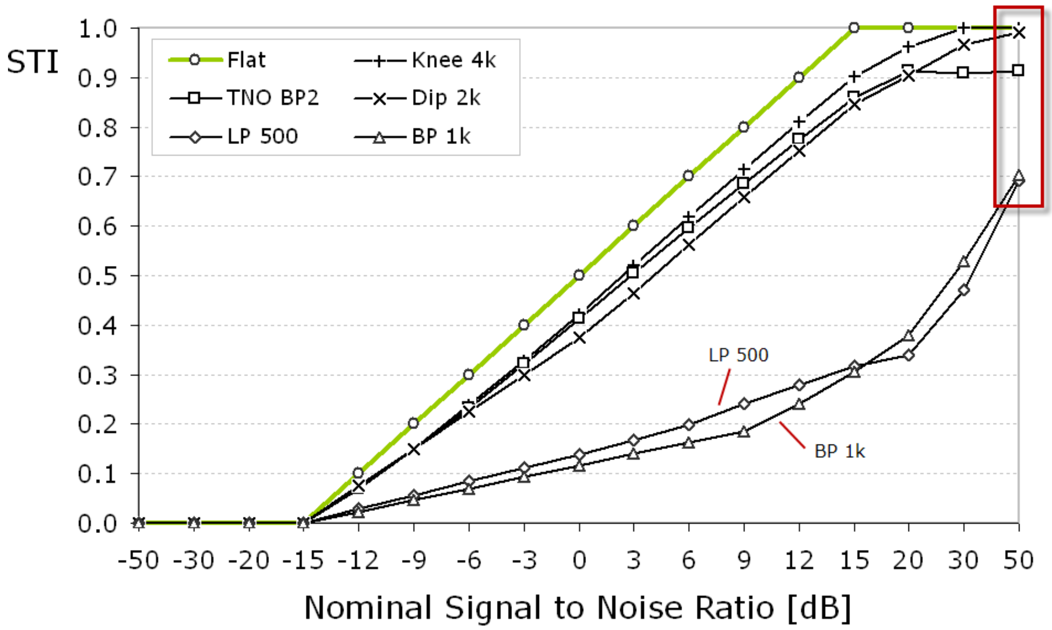

Figure 9: STI versus nominal signal-to-noise ratio SNR (see text) for different distortions of the frequency response according to Table III. The nominal speech level is 75 dB. Note non-linear divisions in the SNR-scale.

At very high signal-to-noise ratios (e.g. 50 dB, see marker frame), the limited frequency responses show only small degradations of the STI. Even the severely deteriorated responses still achieve STI values in the region of STI = 0.7 (classification: "good"). If the calculations would have omitted the absolute reception threshold, the values would be even higher since infinitely small noise levels will result in infinitely high SNR values, even if the frequency response is attenuated by as much as 40 dB. (At 30 and 50 dB nominal SNR, the noise intensity of the absolute reception threshold is as high or higher than the simulated noise.)

For all signal-to-noise ratios applicable to many sound and voice alarm systems (roughly the range of 0 to 20 dB), even moderate restrictions of the frequency response, like a 4 kHz low pass behaviour, result in a reduction of the STI of up to 0.1 with the given noise spectrum. In general, the relationship between the STI and subjective intelligibility scores is nonlinear and also depending on the testing method applied. It is therefore difficult to relate a 0.1 improvement in STI to a specific improvement in intelligibility. As one example, with an STI of 0.5, a 60% score is achieved for phonetically balanced CVC-nonsense-words. With a 0.1 reduction in STI, a score of only 45% would be achieved. For the conditions with severe restrictions (e.g. the 500 Hz low-pass filter or the 1 kHz band-pass system), the STI is reduced extremely, which proves the inefficiency of such systems under real-world conditions. It should be noted that, with an electro-acoustic system, the frequency response as described here is composed of the sum of all system-components – from the microphone to the receiver location in the sound field.

4 CONCLUSIONS

Two commonly reported weaknesses of the STI-method and the underlying Modulation Transfer Function MTF have been analyzed. A systematic analysis showed that, in general, the STI-method is indeed sensitive to the controlled addition of early arrivals, thus reflecting the importance of strong early arrivals for speech intelligibility. The relatively weak sensitivity of the STI with regards to strong and late-arriving echo-distortions in the time-domain is inherent to the method and can not be improved suitably by post-processing. Subsequently, for the analysis of late-arriving, strong echoes, additional criteria should be applied or an auralisation should be evaluated.

The sensitivity of the STI with regards to irregularities in the frequency response of the transmission channel can be significantly improved by ensuring that appropriate speech and background noise levels as well as the speech reception threshold are always included in the calculation. This is a necessary pre-requisite for an accurate STI-assessment, since otherwise, the signal-to-noise ratios in some or all speech bands and subsequently the STI will be over-predicted. Nonetheless, even if optimum care is taken to correctly apply the method, conditions will still exist where the STI will fail to correctly account for the described effects of limitations in the frequency response: systems that always maintain more than 15 dB SNR in all frequency bands, will be almost insensitive to changes of the frequency response. Currently, the most suitable workaround to overcome this limitation is to perform a separate measurement of the frequency response of the transmission channel and to evaluate this measurement with regards to deviations from a flat response.

5 ACKNOWLEDGMENT

This paper would have been impossible without the help of and many fruitful discussions with two highly respected colleagues: Chris Ickler, senior acoustics and DSP engineer at Bose Research and co-inventor of the Bose simulation tools Modeler and Auditioner, and Jan Verhave, senior research engineer and STI-expert at TNO in the Netherlands.

6 REFERENCES

1. International Standard EN (IEC) 60268-16 (2003), “Sound System Equipment – Part 16: Objective Rating of Speech Intelligibility by Speech Transmission Index”.
2. H.J.M. Steeneken and T. Houtgast, “A Physical Method for measuring Speech Transmission Quality”, *J. Acoust. Soc. Am., Vol.* ***67****, 318-326 (1980).*
3. H.J.M. Steeneken, Ph.D. thesis, University of Amsterdam, Amsterdam, 1992.
4. M.R. Schroeder, “Modulation Transfer Functions: Definition and Measurement”, *Acustica, Vol.* ***49,*** *179-182 (1981)*.
5. Peter Mapp, “Some effects of equalization on sound system intelligibility & STI measurement error”, *Proc. IoA, Vol.* ***25*** *Pt. 8 (2003)*.
6. Kenneth D. Jacob, Thomas K. Birkle and Christopher B. Ickler, “Accurate Prediction of Speech Intelligibility without the Use of In-Room-Measurements”. *J. Audio Eng. Soc., Vol.* ***39,*** *No. 4 (April 1991)*.
7. Bose Modeler Design Program. http://www.bose.com/modeler
8. Leo L. Beranek, “Applications of NCB and RC noise criterion curves for specification and evaluation of noise in buildings”, *Noise Control Eng. J., Vol.* ***45,*** *209-216 (1997)*.
9. ANSI S12.2-1995, “Criteria for Evaluating Room Noise”.