CHECKING THE ACCURACY OF MLSSA STI MEASUREMENTS

Tony Stacey AMS Acoustics Ltd

1. INTRODUCTION

The STI (**S**peech **T**ransmission **I**ndex) and its little brother RASTI (**R**oom **A**coustics or **RA**pid **S**peech **T**ransmission **I**ndex) are functions of the following parameters:

(i.) Reverberation time (T)
(ii.) Signal-to-noise ratio (SNR)
(iii.) Direct-to-reverberant ratio (DRR)
(iv.) Receiver Q (Qr)

Equation 1 shows the formula for predicting the *m(F)* as a function of the above parameters. This equation can be derived from the first principles given in [2].

$$m(F) = \frac{\sqrt{\left(Q_r 10^{DRR/10} + \frac{1}{1+(\omega\tau)^2}\right)^2 + \left(\frac{\omega\tau}{1+(\omega\tau)^2}\right)^2}}{Q_r 10^{DRR/10} + 1} \cdot \frac{1}{1+10^{-SNR/10}} \qquad (1)$$

where:

DRR is the direct-to-reverberant ratio at the receive position in dB
Q_r is the receiver Q and acts as a *DRR* modifier
ω is the radial frequency $2\pi F$ in Hz
F is the modulation frequency in Hz
τ is the exponential time constant $T/6ln(10)$ in seconds
T is the reverberation time at the receive position in seconds
SNR is the signal-to-noise ratio at the receive position in dB

As the DRR becomes increasingly negative, equation 1 reduces to the more commonly seen equation 2:

$$m(F) = \frac{1}{\sqrt{1+(\omega\tau)^2}} \cdot \frac{1}{1+10^{-SNR/10}} \qquad (2)$$

MLSSA applies two corrections to the MTF matrix when computing calibrated STI values, namely:

(i.) Auditory Correction Factors (*ACF*), called m-correction in MLSSA, and (ii.) Speech shape Filter Corrections (*SFC*) using the MLSSA speech.ini file.

The *ACF*'s and *SFC*'s are worthy of some explanation.

1.1. ACF's

The *ACF*'s take account of the masking effect on an octave band of speech from the adjacent lower octave band [1]. The slope of the masking is given as -35dB/octave and acts to decrease the *m(F)* in an octave band *k* as follows:

$$m'(F)_k = m(F)_k \frac{I_k}{I_k + I_{k-1} 10^{-35/10}} \qquad (3)$$

where I_k is the mean intensity in octave band *k* and I_{k-1} the mean intensity in the adjacent lower octave band.

It is important to note that I_k and I_{k-1} are the total intensities and so include direct, reverberant and noise components.

The consequence of the *ACF*'s is a reduced STI for systems with an over emphasised low frequency response or a response with significant amplitude deviations between adjacent octave bands. This can be caused by background noise or reverberant energy, and is not necessarily due to the signal source itself.

1.2. SFC's

The *SFC*'s in MLSSA are used to correct for differences between the response of a speech filter inserted into the signal chain and that of standard speech spectra tabulated in the MLSSA speech.ini file.

Each column in the file is first A-weighted and then normalised to a common dBA level. The resulting differences in each octave band are then used to correct the *m(F)*'s by adjusting the octave band apparent SNR's appropriately (assuming a correctly shaped spectrum had been used). In this way, the corrections can either increase or decrease the octave TI values.

Table 1 below shows the resulting *SFC*'s using the speech filter and the standard reference spectra (used for standard STI and RASTI calculations). These values are added to the apparent SNR's and the MTF matrix, and hence STI, adjusted accordingly.

Table 1 – Calculation of *SFC*'s

	125	250	**500**	1k	**2k**	4k	8k	dBA
Speech Filter	-0.8	0.0	-1.2	-4.3	-9.0	-15.7	-24.3	0.4
Adjusted Filter	-1.2	-0.4	-1.6	-4.7	-9.4	-16.1	-24.7	0.0
Reference	-2.5	0.0	-1.0	-5.0	-10.0	-17.0	-23.0	0.0
SFC's	-1.3	0.4	0.6	-0.3	-0.6	-0.9	1.7	-

It should be noted that the *SFC*'s are only applied when MLSSA calculates the STI and <u>not</u> at the time when the stimulus is captured. It is therefore possible to artificially increase or decrease the resulting STI value by altering the speech.ini file. It should also be noted that the corrections are applied to 'calibrated' STI's and RASTI's whether or not a speech filter has actually been used.

For so called ‘calibrated’ measurements, MLSSA v10 requires that the following criteria be satisfied:

1) 65535-point impulse response,
2) >= 1 second impulse response duration (sti.set = 1.8s),
3) >= 10kHz bandwidth (sti.set = 12kHz),
4) Speech filter (or adequate SNR in each octave band),
5) Only 1 pre-average cycle.

If the above criteria are not satisfied (excluding item 4), an ‘uncalibrated’ measurement will be made and the speech.ini file will not be used to adjust the RASTI or STI values.

Not until all of the above parameters and corrections are accounted for should we begin to compare MLSSA STI measurements with our mathematical predictions.

2. MEASUREMENT METHOD

All measurements were made using the MLSSA v10W sti.set set-up file and the speech filter recommended in the user manual [3]. MLSSA was running under Windows 98.

Three scenarios were investigated as follows:

(i.) MLS stimulus corrupted with noise only (electronic domain),
(ii.) MLS stimulus corrupted with reverberation only (electro-acoustic domain) and
(iii.) MLS stimulus corrupted with reverberation and noise (electro-acoustic domain).

2.1. (i.) MLS Stimulus Corrupted with Noise Only

For this scenario, the apparatus was set-up as shown in figure 1 below.

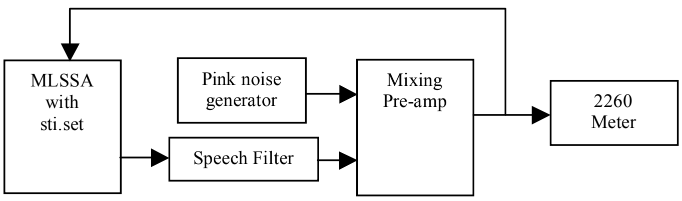


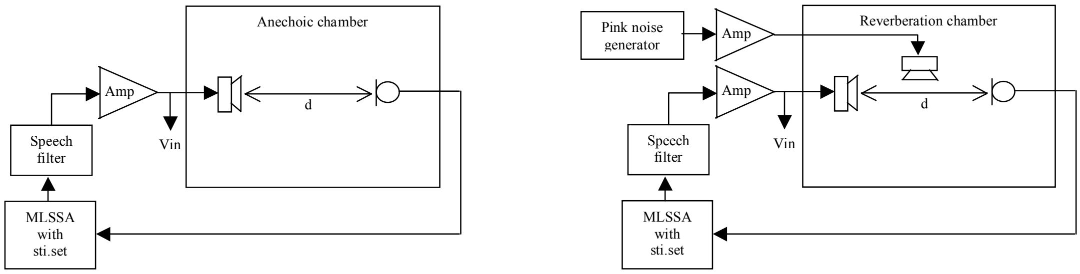
Figure 1 – Set-up for scenario (i.)

The pink noise generator was adjusted in steps of 5dB such that the measured SNR at the output of the pre-amplifier ranged from –15dB to +15dB in the 1kHz octave band. For each SNR the octave band levels of the speech shaped MLS stimulus and pink noise were measured at the pre-amplifier output using the 2260-meter.

The octave SNR’s were then used to mathematically predict the STI and RASTI for comparison with the measured MLSSA values.

2.2. (ii & iii.) MLS Stimulus Corrupted with Reverberation Only and Reverberation & Noise

The apparatus was set up as shown in figure 2 below.



(a) - Anechoic (b) - Reverberant

Figure 2 – Set-up for scenarios (ii & iii.)

The source loudspeaker was first placed in an anechoic chamber (figure 2a) at a known distance (d) from the measuring microphone and the speech shaped MLS stimulus applied across the loudspeaker terminals at a known voltage (Vin). The octave band sound pressure levels (i.e. direct levels) were measured at the microphone position.

The set-up was then transferred to a reverberation chamber (figure 2b) keeping d and Vin constant. In this way, the octave band direct sound pressure level at the microphone was always known and hence the DRR could be determined. The volume of the reverberation chamber was approximately $40m^3$.

A pink noise generator and omni-directional loudspeaker was used as the noise source for scenario (iii.).

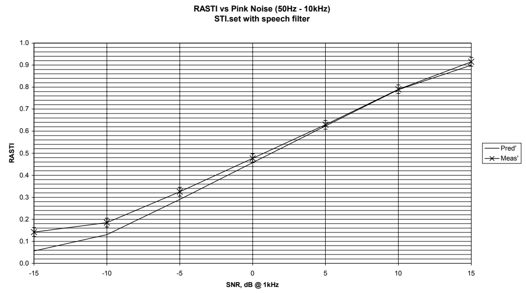
The reverberation time of the chamber was adjusted using foam wedge absorbers. For each reverberation time, the total octave band noise and signal levels (direct + reverberant) were measured and recorded for use in the mathematical predictions.

All reverberation times and sound pressure levels were measured at the microphone position using the source loudspeaker in its respective position.

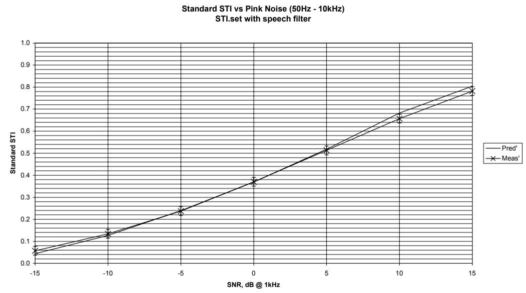
Reverberation times used in the mathematical predictions were EDT's taken from the MLSSA acoustics tables. Paper trace decays were also measured for comparison but the results are not included in this paper.

3. RESULTS

The results are presented in the graphs of figures R1 to R6.

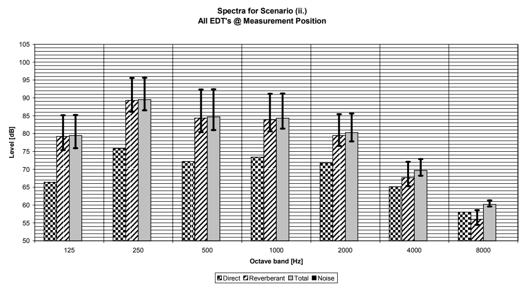


(a)

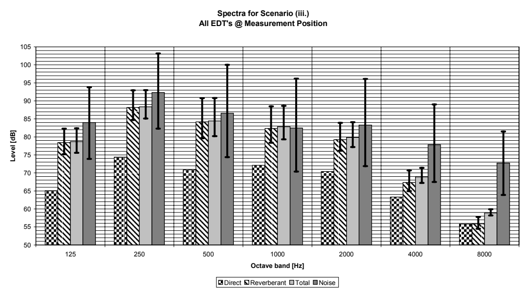


(b)

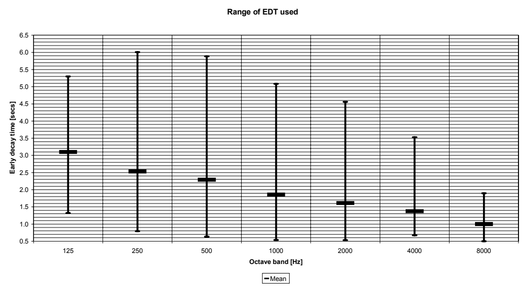
Figure R1 – Scenario (i.), Noise contamination



(a)



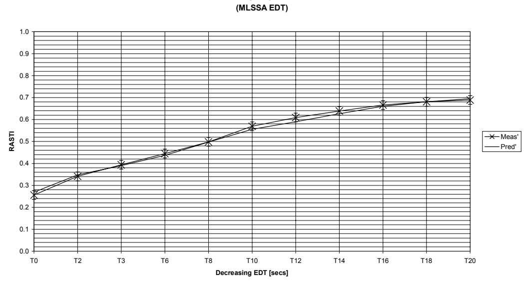
(b)



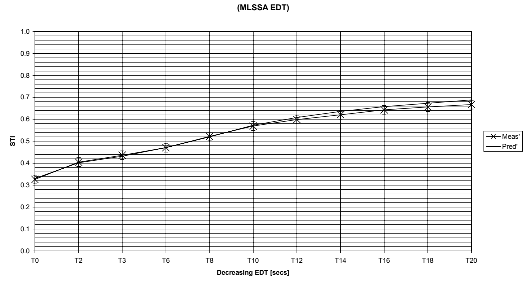
(c)

Note: The error bars in these graphs show the maximum and minimum values used in the investigation.

Figure R2 – Spectra & EDT's for Scenarios (ii.) & (iii.)

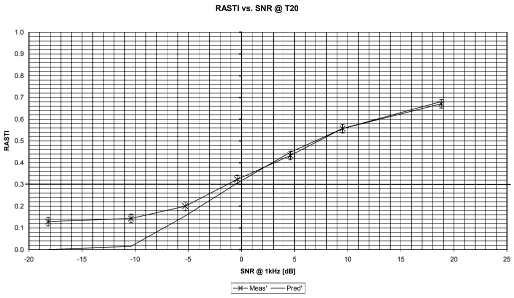


(a)



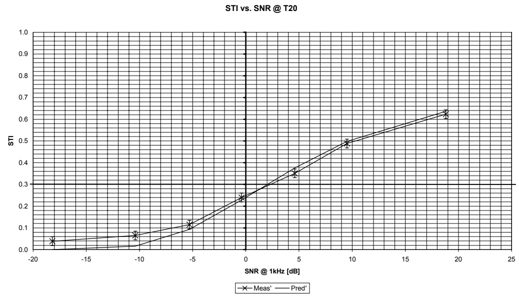
(b)

Figure R3 – Scenario (ii.), Reverberation contamination

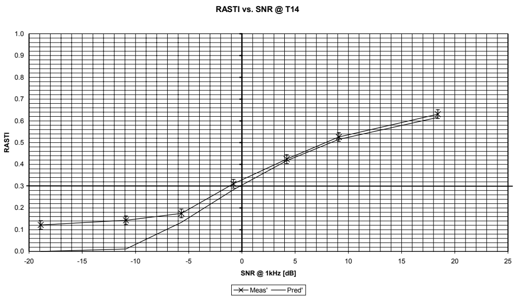


(a)

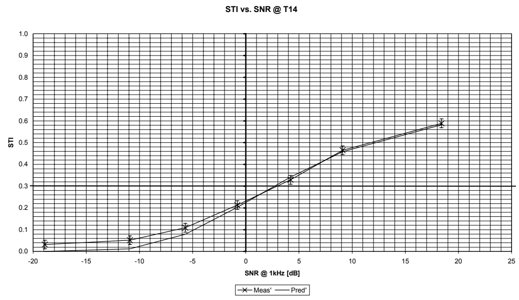
< Low EDT >



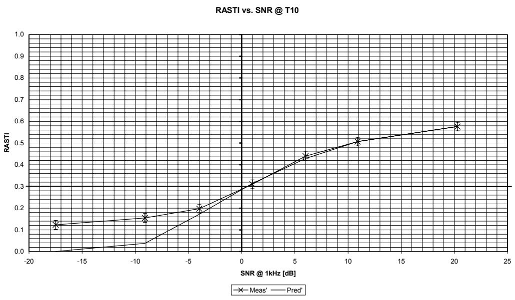
(b)



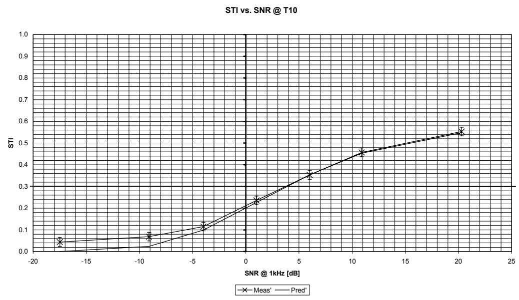
(c)



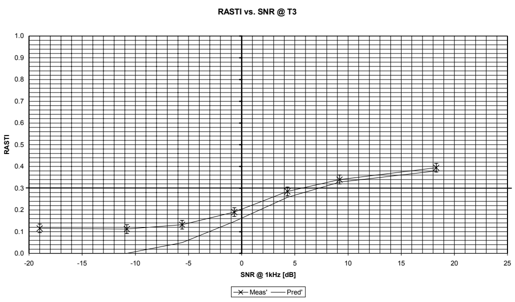
(d)



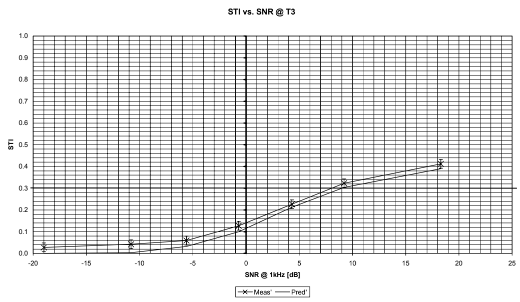
(e)



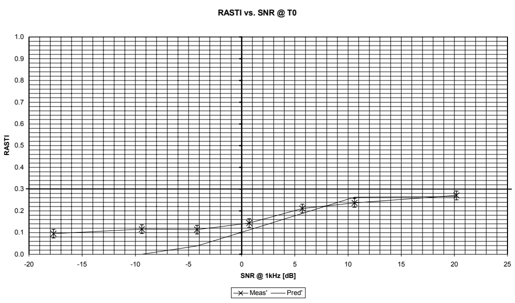
(f)



(g)

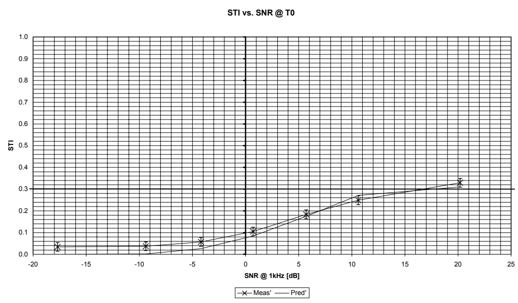


(h)



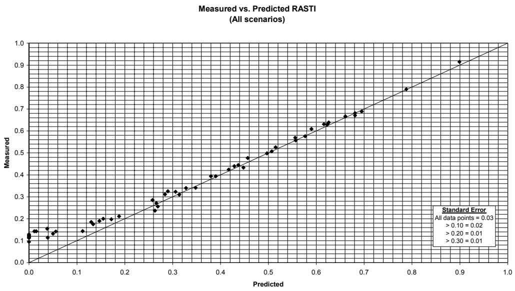
(i)

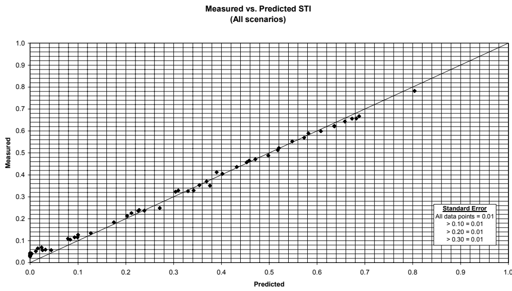
< High EDT >



(j)

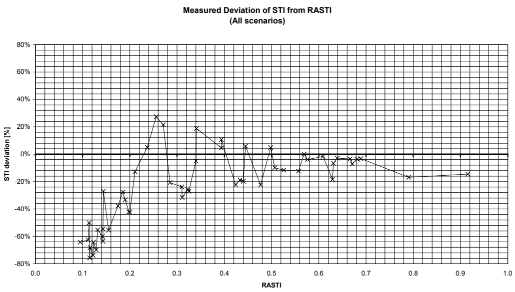
Figure R4 – Scenario (iii.), Reverberation & Noise contamination

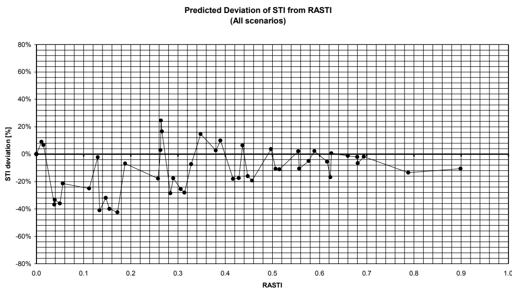




(a) (b)

Figure R5 – Measured vs. Predicted Results





(a) (b)

Figure R6 – STI vs. RASTI

4. OBSERVATIONS

Figures R1 show the results for scenario (i.) for both RASTI and standard STI. As can be seen from figure R1a, there is good agreement between the measured and predicted RASTI values when the SNR is positive. However, this agreement deteriorates as the SNR becomes increasingly negative and the measured RASTI ‘bottoms out’. Figure R1b shows that a better agreement is obtained when we measure the standard (full) STI. In practice, this discrepancy would only be an issue if we were, for example, attempting to quantify a speech privacy system. It would be unwise to operate a voice alarm system if the SNR in the 1kHz octave band was negative.

The MLSSA STI output matrices showed that the m-corrections (*ACF*’s) were in agreement with the predicted values for scenario (i.). Further, as the noise increased, the *ACF*’s had less effect on the MTF’s and hence the STI. This is due to a ‘flattening out’ of the total signal as the pink noise begins to dominate, i.e. at high SNR’s, the auditory masking is a function of the spectral shape of the MLS stimulus (speech shape).

Figures R2 show the measured sound pressure levels and Early Decay Times used for scenarios (ii.) and (iii.). The vertical bars depict the maximum and minimum values in each octave band for all samples in the scenario. As can be seen, with the exception of the 8kHz octave band, the measured DRR's were negative for all samples. This is due to the relatively large EDT/V ratio (i.e. small V) even for low EDT's. It is worth noting that with the exception of cinemas and other similar spaces, public areas generally have hard reflecting surfaces and this, coupled with a distributed loudspeaker system, results in a negative DRR in most instances.

Figures R3 show there is very good agreement between predicted and measured results for both RASTI and standard STI when the stimulus is corrupted with reverberation only. The *ACF*'s are also well predicted with a maximum error of 0.001 for some reverberation times and octave bands.

Figures R4 show the results for scenario (iii.). For all EDT's the agreement between the measured and predicted results deteriorates as the SNR's become more negative in the same way as in scenario (i.). The standard STI again performs better in this respect. The *ACF*'s in this scenario showed a maximum error of 0.002. This error occurred in the 8kHz octave band.

The measured vs. predicted results for all scenarios are shown in Figures R5. As can be seen, the measured RASTI 'bottoms out' in the region 0.15 while the measured STI agrees with the predictions down to about 0.05. This is in general agreement with the graph of figure 2 given in IEC 60268-16: "Objective rating of speech intelligibility by speech transmission index".

Figure R6a shows the deviation of the measured standard STI as a percentage of the measured RASTI. It can be seen that above approximately 0.20 RASTI, the STI deviation tends to decrease with increasing RASTI. Further, as the RASTI increases, the STI is more likely to be lower than the RASTI. The decreasing points below 0.20 RASTI show the effect of the later 'bottoming out' of the standard STI seen in figure R5b.

5. CONCLUSIONS

Using the results of figures R5, we can conclude that predicted RASTI values falling below approximately 0.20 and STI values below 0.10 are likely to be underestimates of MLSSA's measurements. This would appear to be a limitation in the measurement process.

To put figures R6 into perspective, let us suppose that we audited a VA system using the RASTI method and measured a minimum value of 0.50 (the pass criteria in IEC 60849). From figure R6a, we see that at 0.50 RASTI, the STI is approximately 12% down, i.e. 0.44. Had we audited using the standard (full) STI, then not only would the system fail, but also the expected subjective rating would go from FAIR down to POOR.

Figure R6b shows the predicted deviations and, with the exception of low RASTI, is in good general agreement with the measured values of figure R6a. These graphs show that RASTI is <u>not</u> a **RA**pid version of the standard (full) STI but is in fact a different measurement entirely. Which measure is the better predictor of speech intelligibility is beyond the scope of this investigation.

6. REFERENCES

[1]. H. J. M. Steeneken and T. Houtgast. “A physical method for measuring speech-transmission quality. J. ASA 67(1) Jan 1980

[2] T. Houtgast, H. J. M. Steeneken and R. Plomp. “Predicting Speech Intelligibility in Rooms from the Modulation Transfer Function”, Part I. General Room Acoustics. Acoustica Vol. 46:1980

[3] MLSSA v.10W users manual