IS SII BETTER THAN STI AT RECOGNISING THE EFFECTS OF POOR TONAL BALANCE ON INTELLIGIBILITY?

G. Leembruggen   Acoustic Directions, Sydney Australia

1 INTRODUCTION

1.1 Motivation for the Work

With over fifteen years of experience in the commissioning of sound systems for parliamentary, court and stadium situations, the author and close colleagues are convinced that small changes in the tonal balance of speech yield important changes in subjective intelligibility. This relationship between tonal balance and subjective intelligibility is not reflected by the Speech Transmission Index (STI) metric.

The author has found that subtle changes in tonal balance allow listeners to better understand the nuances of words and articulation styles of everyday talkers. These changes in tonal balance often mean that listeners do not need to strain to understand and are much more comfortable.

It is important that perception of intelligibility is a comfortable experience for listeners in situations such as parliaments, courts and stadia. Courts and parliaments are daily, eight-hour working environments in which speech is the most important aspect of their operation.

Tonal adjustments make an important difference to intelligibility with talkers who do not articulate clearly and who are being reproduced in the presence of reverberation. Examples of the types of equalisation that are effective are:

- a 1 dB boost over an octave in the 1 to 2.5 kHz region
- removal of colouration at low-mid frequencies due to a resonance – an example might be a notch of 2 dB at 415 Hz with a bandwidth 0.2 octaves

The author uses a significant amount of vocal music to commission systems, with the goal of being able to understand the lyrics. As vocal music contains masking noise in the form of musical accompaniment, it is often much harder to achieve intelligibility than with spoken words. At the Australian Parliament, we found small frequency-response changes of no more than 1 dB in a few 1/3 octave bands allowed us to significantly improve the intelligibility of a difficult Tom Waits vocal track whilst adding a refinement to the sound quality for other artists.

It should be noted that in making these adjustments, we are adjusting for word recognition, not just for sound quality.

When the author spoke with UK colleagues about the inability of STI to include the effects of frequency response on intelligibility, it was suggested that this was included in the recent Speech Intelligibility (SII) standard.

As this SII standard includes the effective signal to noise ratio resulting from a reduction in speech modulation and a more sophisticated algorithm to predict the effects of the upward spread of masking, it is pertinent to investigate if SII is able provide a more comprehensive assessment of subjective intelligibility than STI.

1.2 Foundation for this Work

In a paper[1] presented at Reproduced Sound 19, the author and Tony Stacey of AMS Acoustics explored the relationship between the subjective intelligibility of speech and the measured STI for a range of tonal balances produced by different filters in a noise-free, reverberant environment.

The method consisted of making an STI measurement of a loudspeaker system with each filter response, subjective testing of word scores with each filter response, and processing the measured STI results and word scores.

Using the objective and subjective data obtained in [1], the SII values were calculated for the different tonal balances and compared with the subjective word scores.

2 OVERVIEW OF SII

A detailed description of the computation of SII is given in the American National Standard S3.5-1997 'Methods for Calculation of the Speech Intelligibility Index'. This standard claims that SII is "highly correlated with the intelligibility of speech under a variety of adverse listening conditions such as noise masking, filtering, and reverberation".

SII values range between 0 and 1, and SII is intended to reflect the proportion of total speech cues available to the listener. An SII of 0.5 indicates that half of the speech cues are delivered to the listener.

As per STI, SII is completely based on the signal to noise ratios in specific frequency bands, with every parameter contributing to intelligibility loss being converted to an equivalent noise level.

The SII algorithm is more complex than STI with respect to its mechanisms to account for the upward spread of masking and hearing acuity, but includes the signal to noise ratios derived from STI's modulation transfer indices (MTIs).

Four computational procedures are provided for SII, corresponding to four types of bandwidths. In order of accuracy, they are:

i) Critical band -21 bands

ii) 1/3 octave bands - 18 bands

iii) Equally contributing critical bands - 17 bands

iv) Octave bands - 6 bands

We have used the 1/3 octave band procedure as it corresponds with normal electro-acoustic analysis practice.

The SII calculation assumes i) the listener faces the speech source, and ii) the speech source is assumed to be omni-directional. Monaural or binaural reception is accounted for with different gain factors.

3 SII CALCULATION METHOD

3.1 Calculation of SII

The methodology presented below for calculation of SII is merely a summary of the important aspects of SII as they pertain to this investigation, and is not intended to be rigorous.

The following parameters are to be calculated in order for each 1/3 octave band between 160 Hz and 8 kHz inclusive:

1. **Speech spectrum level E'i** for the direct field of speech at the centre of listener's head.
   E'_i is the L_{eq} of speech in the band of interest over a 1 Hz bandwidth.

   $E'_i = E_i + G_i$

   where E_i is the speech spectrum level and G_i is the insertion gain in each band.

   We have used a nominal level LA_{eq} of 59.2 dB and L_{eq} 62.4 dB, corresponding to the standard speech level for 'normal' vocal effort given in the Standard.

2. **Equivalent noise spectrum level N'_i** at centre of listener's head from uncorrelated (ambient) noise and correlated (temporal response) noise.
   The noise levels N'_i are derived from the MTI data as described in Section 3.2.

3. **Equivalent hearing threshold T_i** from tables in the standard.
   For people in the 18-30 age group with no hearing loss, T'_i is 0 dB.

4. **Equivalent masking spectrum level Z'_i**
   The equivalent masking spectrum level comprises masking from within-band, upward spread, and self-speech masking. It includes ambient noise, loss of modulation and is calculated from:

   i) Self-speech masking spectrum level V_i

   $V_i = E_i - 24$ (1)

   ii) Determine B_i as the larger of V_i and N'_i

   iii) **Slope C_i per octave of the spread of masking** from all lower frequency bands up to the (i -1)th band.

   $C_i = -80 + 0.6[B_i + 10\log F_i - 6.35]$ (2)

   where F_i is the centre frequency of the 1/3 oct band.

   iv) **Equivalent masking spectrum level Z'_i**

   $$Z_i = 10\log\left\{10\text{^}(N'/10) + \sum_{k}^{i-1} 10\text{^}(B_k + 3.32\, C_k \log(0.89F_i/F_k)/10\right\} \qquad (3)$$

5. **Equivalent internal noise spectrum level X'_i**
   X'_i is the reference internal noise spectrum level given in tables in the Standard
   $X'_i = X_i + T'_i$ (4)

6. **Equivalent disturbance level D_i** - the larger of Z_i and X'_i and represents the equivalent level of interfering noise.

7. **Speech level distortion factor L_i**
   Accounts for the decrease in intelligibility of speech at high presentation levels. Its value decreases to zero at high levels.

   $$L_i = 1 - (E'_i - U_i - 10)/160 \qquad (5)$$

   where U_i is standard speech spectrum level at normal vocal effort. (L_i is truncated at 1)

8. **Band audibility function A_i**
   A_i is the effective portion of the speech dynamic range within the band that contributes to speech intelligibility under conditions less than optimal.

   $$K_i = (E'_i - D_i + 15)/30 \qquad (6)$$

   $$A_i = L_i K_i \qquad (7)$$

   K_i is the total effective signal to noise ratio and is limited to the range of 0 to 1.

9. **Speech intelligibility index SII**
   where I is the band-importance function.

   $$SII = \sum_{i=1}^{n} I_i A_i \qquad (8)$$

   The band-importance function depends on the nature of the speech type and the proficiency of the talkers and listeners. A range of band-importance functions is given in Table B.2 of the Standard.

3.2 Calculation of Equivalent Noise Spectrum N_i

Using the modulation transfer indices in the STI matrix, the average apparent speech to noise ratio (over all modulation frequencies) can be found from Eqn 9. From Eqn 10, the equivalent noise spectrum for correlated and decorrelated noise can then be found.

$$R_i = 10\log[M_i/(1 - M_i] \quad \text{where } R_i \text{ is truncated to } +/-15\text{dB} \qquad (9)$$

$$N'_i = E'_i - R \qquad (10)$$

4 METHODOLOGY USED FOR THIS INVESTIGATION

Sections 4.1 and 4.2 below are reproduced directly from [1].

4.1 Measurement Procedure

A loudspeaker and dummy head with binaural microphones were set up in an anechoic chamber at AMS Acoustics. The loudspeaker was fed with a MLS signal via a speech weighting filter and power amplifier. The response of the speaker was then measured at each ear by the binaural microphones at a distance of 1.5 m from the speaker on axis and processed by MLSSA v10w to yield the speaker's anechoic frequency response of the speaker and the system STI. This was Scenario 1.

The speaker and associated speech filter and amplifier were relocated to a reverberation chamber at AMS Acoustics. Again the system STI was measured at a distance of 1.5 m from the speaker on axis using the binaural microphones. This was Scenario 2.

Using acoustic absorption material, the reverberation time of the chamber was adjusted so that the measured STI was approximately 0.5.

Seven different frequency response shaping filters (Scenarios 3 to 9) were then sequentially inserted into the drive chain to alter the speaker's frequency response. For each filter, the impulse response was captured and the frequency response and STI of the system measured with and without the speech-weighting filter connected in series with the response-shaping filter.

4.2 Subjective Procedure

A CD of anechoically recorded speech was prepared and consisted of 1000 carrier sentences and words arranged into 20 groups of 50 words. The words were spoken by a female, and were single syllable, phonetically balanced (PB) types situated at the end of each sentence.

Three groups (of 50 words) were then played through the speaker in the anechoic chamber. The sound was picked up by binaural microphones on the dummy head at a distance of 1.5 m from the speaker and recorded onto digital media. (Scenario 1)

The speaker and associated speech filter and amplifier were relocated to a reverberation chamber. Another thee groups of words played through the speaker and received by the binaural microphones at a distance of 1.5 m from the speaker on axis and recorded. (Scenario 2)

For each of the seven response-shaping filters, three lists of 50 words were replayed and recorded for Scenarios 3 to 9. When the groups were exhausted, a reshuffled version of the lists was used.

The recordings of the nine scenarios were then distributed to listeners in the UK and Australia. In the UK, seven listeners evaluated all or part of the three lists for each of the nine scenarios, to give a total of 135 listening sessions. In Australia, three listeners evaluated all of the three lists for each of the nine scenarios, to give a total of 81 listening sessions. The sentences were presented to listeners through headphones, and the listener wrote down the word at the end of the sentence.

4.3 Test Scenarios and Frequency Response Filters

The responses of the tonal filters were chosen from our experience to exaggerate the subjective difficulties with intelligibility.

Table 1 lists the scenarios and filters used for the tests, while Figure 1 shows the frequency responses of the filters used to modify the tonal balances.

Scenario	Description	Tonal Filter	Comment
1	anechoic	none	
2	reverberant	none	Absorption adjusted to produce STI ≈ 0.5
3	reverberant	5dB/octave cut	
4	reverberant	5dB/octave boost	
5	reverberant	2.5kHz 12dB notch @ Q=0.7	
6	reverberant	Plateau @-3dB 400Hz to 1kHz Plateau @ -10 dB 1.2kHz to 6kHz	Typical poor sound system
7	reverberant	250Hz 18dB boost @ Q = 1.5	
8	reverberant	630Hz 18dB boost @ Q = 1.5	
9	reverberant	Notches @ 500Hz & 2kHz at -18 dB	

Table 1 Frequency response scenarios

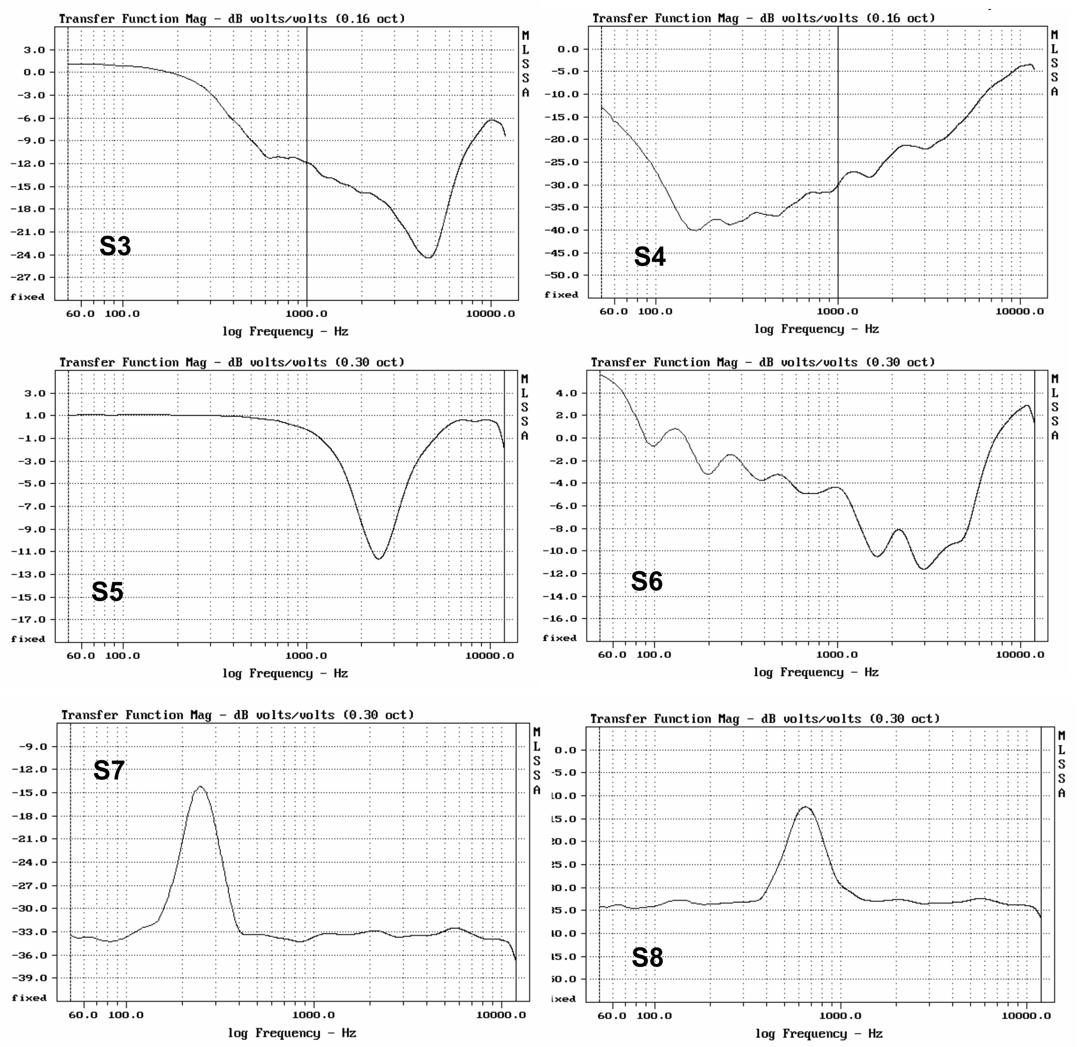
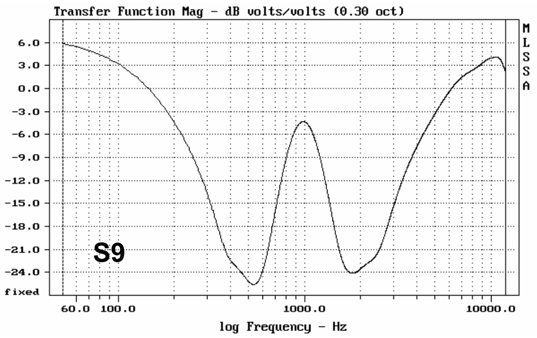


Figure 1 Frequency responses of tonal filters for Scenarios 3 to 9



4.4 Data Processing

The following steps were used:

1. The speech spectrum level E_i in each one third octave band with a 'normal' voice from Table 3 in the SII Standard was used and modified by the frequency response G_i of the tonal filter for each Scenario.

2. For each Scenario, the total (integrated) level of the filtered speech was adjusted to produce a sound pressure level (SPL) of 59.2 dB(A). The associated linear weighted SPLs varied between Scenarios. Steps 1 and 2 yielded the speech spectrum levels E'_i.

3. The MTF matrices for Scenarios 3 to 9 were exported from MLSSA. Using the method of [2], the masking corrections applied by MLSSA to compute the MTIs were found, and the unmasked MTIs computed for each Scenario. As the MTIs were only available in octave bands, the MTIs in the one-third octave bands were made equal to the relevant octave band MTI.

4. Using the equations (9) and (10), the equivalent noise spectrum levels N'_i were found.

5. The steps listed in Section 3.1 were then undertaken to yield the SII for each Scenario. SII results were computed with band-importance functions for “standard’ and “short passages of easy reading material”.

6. The word scores were evaluated and were converted using the Common Intelligibility Scale to equivalent STIs.

5 RESULTS FROM DATA PROCESSING

5.1 Measured STI Results

Table 2 lists the measured STI results for each scenario, and the Full, Male and Female STIs using with the revised masking thresholds for 70 dB(A). Although the table lists only the right-ear measurements, the left-ear results were between 0.01 and 0.02 above the right-ear results.

The following comments are made.

a) As expected, the Full STIs without the speech filter did not change significantly when the tonal filters were inserted.

b) The low value of 0.426 for Scenario 4 was due to poor signal to noise ratios in the bands up to 1 kHz, caused by a combination of the rising high frequency response, limited amplifier power and the residual noise in the test chamber.

c) The low value of 0.426 for Scenario 4 was due to poor signal to noise ratios in the bands up to 1 kHz, caused by a combination of the rising high frequency response, limited amplifier power and the residual noise in the test chamber.

d) The effect of the Male weighting on the Full STI was also relatively minor.

e) The Female weighting increased the STIs, but also reduced the spread of the STIs over the scenarios.

	Full with speech filter	**Male with speech filter**	**Female with speech filter**
speech filter	0.982	0.952	0.991
1	0.972	0.948	0.982
2	0.477	0.467	0.503
3	0.472	0.475	0.498
4	0.463	0.471	0.505
5	0.477	0.467	0.503
6	0.494	0.490	0.524
7	0.483	0.469	0.503
8	0.479	0.472	0.503
9	0.453	0.435	0.472

Table 2 Measured STIs for each Scenario

Word Score Results

Figure 2 gives the word score results for each scenario.

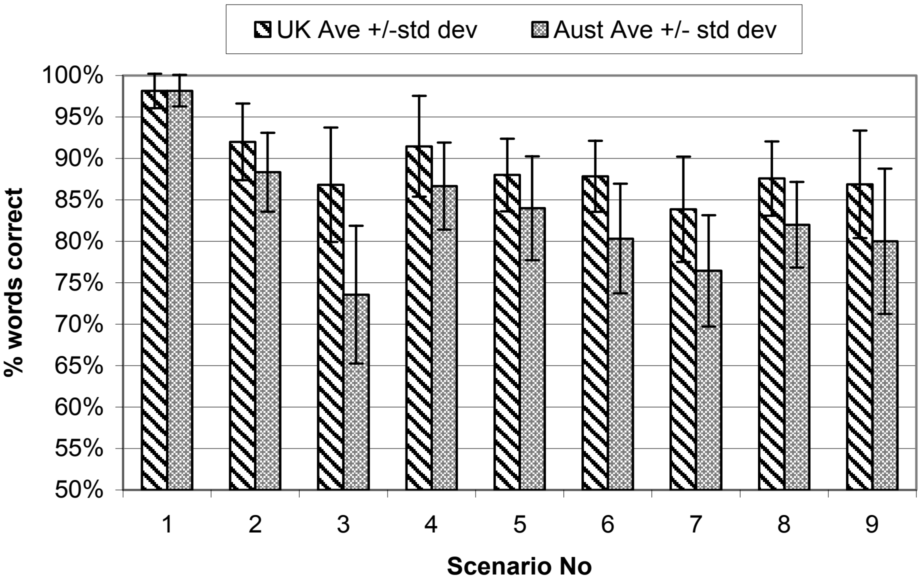


Figure 2 Word score for each Scenario. Error bars indicate a +/- standard deviation.

The following comments are made.

a) Although the word score testing was not carried out rigorously in accordance with the ISO TR 4870 standard, and there was a wide range in the results, the trends were clear.

b) The average Australian scores for each scenario were generally lower than the corresponding UK scores. This was highly likely to result from accent differences.

c) The UK and Australian average scores showed a similar trend over the range of scenarios.

d) There was a noticeable reduction in the word score when the tonal filters were inserted.

e) Even though the test words were well-articulated, each of the Australian listeners found it necessary to concentrate while listening, in order to discern the test words. More concentration was required for the filtered words. If this concentration had not been applied, the scores would have been lower.

f) The Australian listeners found the process to be quite tiring, and yet the measured STI was of the order of 0.5, which is a value that is typically specified for sound systems.

Figure 3 compares the equivalent PB word-score STIs and the measured STIs with the revised masking slopes and the full, male and female weightings.

The following comments are made.

a) The STI of 1 for the upper error bar for Scenario 1 resulted from the CIS conversion which greatly amplified the STI with word scores greater than 97%. At this value, a 3% change in PB score results in a STI change of 0. 45; ie from 0.55 to 1.0.

b) When the tonal filters were introduced, the reduction in equivalent STI was up to 0.16 for UK listeners and up to 0.2 for Australian listeners.

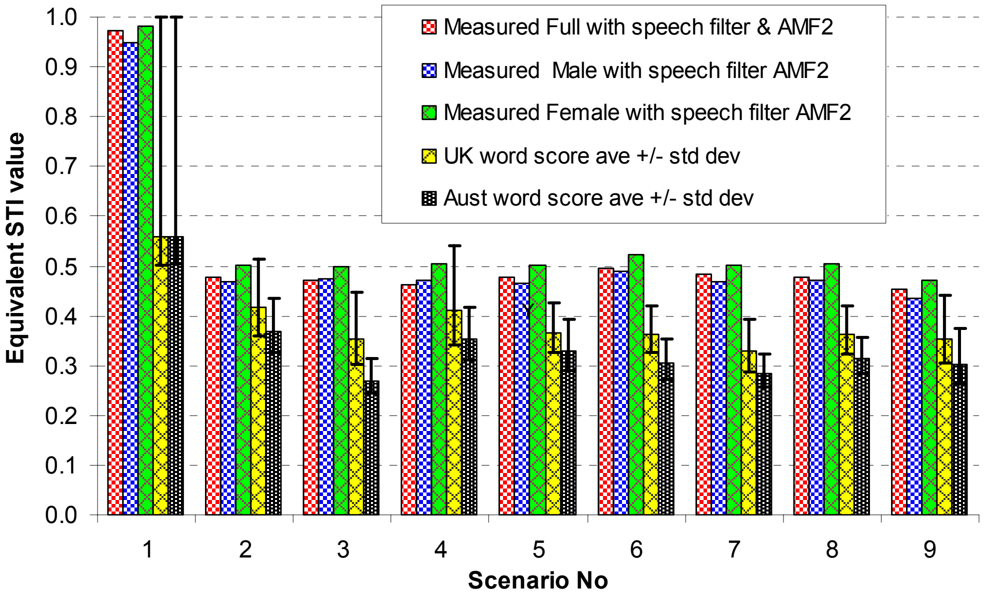


Figure 3 Comparison of equivalent STIs of PB word scores with measured Full, Male and Female STIs with revised weighting factors. The error bars show the range of standard deviations.

5.2 SII Results

Table 3 compares the predicted SII results using the band-importance functions for 'standard' and 'short passages' with the measured STI values. The linear SPLs resulting from each Scenario being normalised to 59.2 dB(A) are also given.

The following comments are made:

a) In general, the SII and STI values are similar, although significant differences exist for some filters.

b) The SII results with 'short passages' weightings are approximately 0.01 lower than with 'standard weightings.

c) There is a marked increase (approx 0.08) in the SIIs with Scenario 7 for which the associated frequency response is a large boost at 250 Hz. This result is unexpected, and is entirely contrary to our experience in the field. We have found that excess energy in this low-mid region is very detrimental to intelligibility.

	SII		STI		
	Standard	**Short Passages**			**Lp(lin)**
Scenario 1	0.989	0.992	0.972		62.4
Scenario 2	0.477	0.466	0.477		62.4
Scenario 3	0.495	0.482	0.472		65.8
Scenario 4	0.476	0.459	0.463		59.6
Scenario 5	0.480	0.470	0.477		62.9
Scenario 6	0.499	0.484	0.494		63.3
Scenario 7	0.577	0.570	0.483		67.0
Scenario 8	0.501	0.494	0.479		61.2
Scenario 9	0.529	0.523	0.453		63.5

Table 3 Comparison of STI and SII results with reverberation

If the correlated noise due to reverberation is removed from the calculations, the SIIs listed in Table 4 result.

	SII	
	Standard	**Short Passages**
Scenario 1	0.996	0.997
Scenario 2	0.996	0.997
Scenario 3	0.982	0.985
Scenario 4	0.997	0.998
Scenario 5	0.996	0.997
Scenario 6	0.999	0.999
Scenario 7	0.988	0.991
Scenario 8	0.965	0.973
Scenario 9	0.999	1.000

Table 4 Comparison of STI and SII results without reverberation

The following comments are made:

a) SII results are close to unity for all Scenarios. This is contrary to our field experience in dry acoustic environments, which indicates that filters of this severity are quite detrimental to subjective intelligibility.

b) With Scenario 3 (5dB/octave cut), the SIIs are slightly lower than Scenario 1. We would expect these SIIs to be significantly lower.

c) The SII for Scenario 8 (18 dB boost at 630 Hz) has the lowest SII. This is an interesting result, as our experience suggests it would have been no worse than with the other filters.

A comparison of the STI, SIIs and the equivalent STI of the word-scores is given in Figure 4. It is evident that the STI and SII scores are similar, and considerably higher than the word score results.

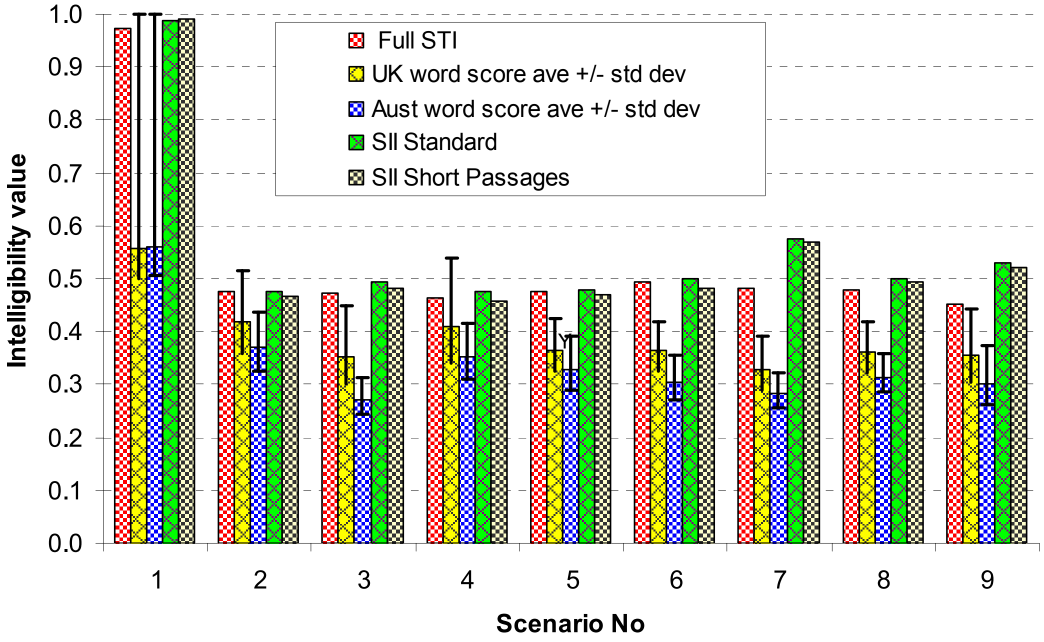


Figure 4 Comparison of STI, SII and equivalent STI values of word scores

6 CONCLUSIONS

1. An investigation was conducted to assess if the Speech Intelligibility Index SII metric was better than STI at recognising the effects of tonal balance (frequency response) on subjective intelligibility.

2. Objective and subjective testing of intelligibility was undertaken in a noiseless reverberant environment with seven different filters being used to shape the frequency response of a loudspeaker system. This work was undertaken for an earlier paper[1].

3. From the measured modulation transfer indices, the correlated noise levels were calculated, representing the reduction in modulation. Using these noise levels and the system frequency responses, the SII was calculated for each scenario.

4. Word score testing showed a clear trend that intelligibility in a noiseless, reverberant environment can be worse when the natural spectrum of speech is damaged by different types of frequency response-shaping filters.

5. The word scores for the Australian listeners were lower than those of the UK listeners.

6. The word scores would have been lower had the articulation been poorer or had the listeners concentrated less.

7. In general, the SII and STI values were relatively similar, although significant differences existed for some filters.

8. The SII values with the seven different filters were within a range of 0.1, with six of the seven lying within a range of 0.06.

9. The differences between the equivalent STI value of the word scores were up to 0.2 below those of the SII.

10. The reductions in the equivalent STI for the word scores (converted to STI using the Common Intelligibility Scale) with the filters inserted were not reflected in the STI measurements nor the SII calculations.

11. More work is needed on the SII and STI processes if these metrics are to reliably quantify the effects of tonal balance on subjective intelligibility and listener comfort.

7 REFERENCES

1. G. Leembruggen and A. Stacey, Should the matrix be reloaded? Proc. IOA, Vol. 25, Part 8 2003

2. A Stacey, Checking the accuracy of MLSSA STI measurements. Proceedings of IOA, Reproduced Sound 18, 2002