SHOULD THE MATRIX BE RELOADED?

G Leembruggen Acoustic Directions, Sydney, Australia
A Stacey AMS Acoustics, London, United Kingdom

1 INTRODUCTION

Over their many years of designing and commissioning sound systems, the authors have acquired considerable field evidence showing that an unbalanced frequency response can greatly affect speech intelligibility. Relatively small changes in an octave bandwidth, sometimes as small as 1 dB, can noticeably affect the intelligibility of conversational speech. The change in subjective intelligibility does not appear to be reflected in speech transmission index (STI) measurements. Improving intelligibility through modest equalisation would suggest that the psychoacoustic masking mechanism is not fully accounted for by the STI process. Other mechanisms may also be contributing.

Situations involving sound systems should allow listeners to understand speech without concentrating on the listening process itself. There are many situations (for example rail and bus terminals) which have sound systems meeting a specified STI performance, but in which listeners must concentrate to understand the speech, especially when that speech is delivered rapidly or with poor articulation. Other situations such as parliaments and courts are more demanding, requiring participants to listen for long periods and concentrate on the subject matter. These systems should deliver “acoustic comfort” to the listeners, ensuring intelligibility throughout the full range of voice types and speaking styles.

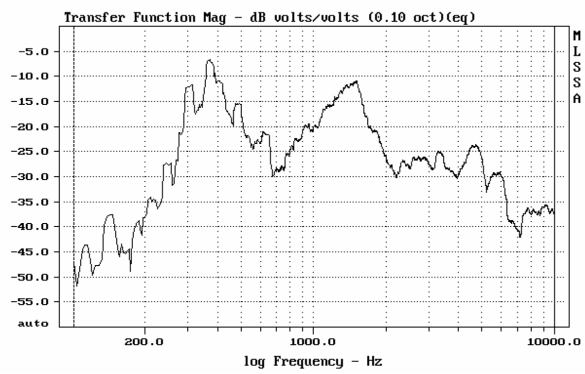
We believe that in many situations there is a real danger that using STI alone to specify a sound system can leave listeners still struggling to understand, due to a poor tonal balance that is often easily equalised.

An investigation was conducted to attempt to shed some light on the components of the STI process and how the construction of its MTF matrix might relate to the apparent lack of correlation between STI and frequency response changes. In the process of preparing this paper, the authors have examined many papers on speech structure and recognition. We have been surprised at the lack of discussion on the role of frequency response.

2 PRIOR EXPERIENCE

2.1 Example 1

One of the authors has previously tested intercom systems for use in prisons using MLSSA. Although the measured STI of one intercom system was 0.51, it was essentially useless as a speech transmission device, due to a very strong tonal colouration at 400 Hz. Fig 1 shows the frequency response and the MTF matrix of that system. The testing was carried out in a disused solitary confinement cell, which has caused the reverberant shape of the MTF matrix. Other intercoms tested had similar STIs, but gave usable speech transmission. As only the STI has been specified, argument with the distributor followed about rejection through a non-specified parameter.



MTF Matrix & Clarity C50 Ratios							
Frequency	125	250	500	1000	2000	4000	8000
C50	-9.6	-1.4	0.7	1.4	1.5	2.8	3.7
m-correction	1	1	1	0.997	0.999	0.986	0.976
0.63	0.664	0.859	0.871	0.897	0.905	0.916	0.749
0.8	0.617	0.827	0.849	0.878	0.89	0.904	0.744
1	0.53	0.769	0.807	0.844	0.86	0.882	0.736
1.25	0.404	0.683	0.744	0.793	0.815	0.848	0.72
1.6	0.261	0.588	0.668	0.729	0.752	0.799	0.693
2	0.186	0.529	0.614	0.676	0.693	0.752	0.658
2.5	0.197	0.489	0.585	0.615	0.625	0.7	0.621
3.15	0.198	0.42	0.554	0.538	0.544	0.645	0.576
4	0.126	0.295	0.495	0.467	0.471	0.589	0.519
5	0.065	0.185	0.459	0.426	0.427	0.539	0.462
6.3	0.067	0.156	0.464	0.382	0.351	0.478	0.403
8	0.054	0.108	0.431	0.337	0.226	0.411	0.356
10	0.058	0.121	0.436	0.255	0.094	0.345	0.328
12.5	0.015	0.114	0.325	0.238	0.047	0.32	0.309
octave MTI	0.285	0.451	0.565	0.559	0.528	0.613	0.539
STI value= 0.510 (0.555 modified) ALcons= 10.8% Rating= FAIR							

Fig 1. Frequency response and MTF matrix of unusable candidate prison intercom

2.2 Example 2

Recently, one of the authors was engaged to refine the equalisation of the speech sound system in the Senate at the Australian Parliament House. The author used a significant amount of vocal music for this work, with the goal of being able to understand the words of the song. As vocal music contains masking noise in the form of musical accompaniment, it is often much harder to achieve intelligibility than with spoken words. One particular artist who was difficult to understand was Tom Waits, a singer with a big, deep, poorly-articulated voice. Small changes in frequency response of no more than 1 dB in a few 1/3 octave bands, allowed us to understand the words to a much greater extent, without damaging the tonality of the music. The improvements gained by equalising with Tom Waits were also worthwhile with other artists, but were more audible with Waits.

As the equalisation also affected the music, the steady-state speech to music ratios in each octave band did not change. The additional intelligibility achieved by this type of equalisation process increases the safety margin for speech reinforcement, and allows the system to better accommodate those difficult scenarios when talkers speak quickly or do not articulate well, or when listeners are not fully concentrating.

The STI method seems to be incapable of resolving the subtlety of the role of tonal balance in producing intelligibility.

3 DISCUSSION OF STI ALGORITHM

A quick overview of how the STI matrix is generated might be useful for some readers. In producing the octave band modulation transfer index (MTI) results, the STI method uses the following process in the order given here.

i) Determine the relative modulation depths (between 0 and 1) of each of the 14 modulation frequency bands in each octave band.

ii) Adjust the modulation depths to account for the upward spread of masking in each octave band to account for the masking of signal in one octave band on the adjacent upper band.

iii) Convert these 14 modulation depths to a signal to noise ratio (SNR).

iv) Truncate the SNRs to a range within ±15 dB.

v) Compute the arithmetic average of these 14 truncated SNRs over the 14 bands.

vi) Convert the average SNR to an equivalent modulation depth to produce the octave band MTI results.

vii) Weight the MTI results according to the importance of the octave band to intelligibility for both males and females, including the redundancy factors.

3.1 Upward Spread of Masking

The original STI algorithm used a constant -35 dB/octave to model the upward spread of masking. Recently the STI algorithm has been revised [1] to include a dependence on the sound pressure level of the test signal, as shown in Table 1. No attempt is made to model the downward spread of masking.

3.1.1 Modulation Frequencies

A number of researchers[2,3,4] has concluded that speech intelligibility is crucially dependent on the preservation of the portion of the modulation spectrum between 2 and 10 Hz. Some references narrow this to 3 to 8 Hz. Drullman et al[2] concluded that modulations below 4 Hz do not appreciably aid intelligibility in everyday sentences.

Examples of modulation spectra are given in Fig 2 (from Fig 1 p126 of [5]), which shows the envelope spectra in three octave bands for clear, conversational and fast clear speech. It is noteworthy that the modulations for conversational speech have relatively more high frequency energy than clear or fast clear speech. Fig 3 from [6] gives another example of the modulation spectrum in the 1 to 2 kHz band for two minutes of speech from a single talker.

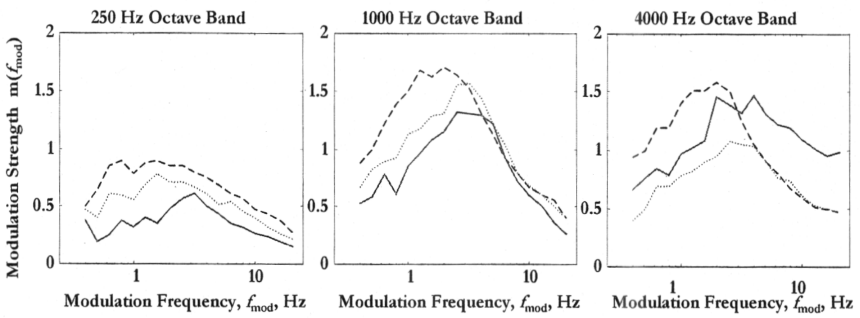


Fig 2. Comparison of envelope spectra in three octave bands for clear (dashed lines), conversational (solid lines) and fast clear speech (dotted lines).

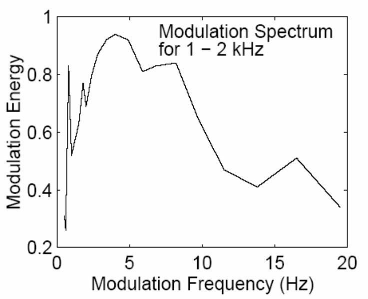


Fig 3. Modulation spectrum for two minutes of speech from a single talker.

Given the strong bias of modulation information towards the 3 to 8 Hz region, it seems surprising to the authors that the modulation transfer matrix assigns equal importance to each of the modulation frequencies when calculating the average of the SNRs. This contrasts with the STI process of weighting the octave band MTI values in accordance with their relative contributions to intelligibility.

In addition, averaging the SNRs, rather than the modulations, allows high modulation values with their associated high SNRs at low frequencies to skew the overall average. This simple averaging of the SNRs can be problematic in low-noise, reverberant environments, in which the modulation values at low frequencies are usually much higher than at frequencies above 2 Hz.

Table 2 shows the change in STI for a typical system when frequencies below 2 Hz in one octave band are removed from the averaging process. A significant decrease in the octave band MTI is evident when the averaging bandwidth is restricted to the frequencies deemed important by the aforementioned researchers.

3.2 Speech Weighting Filters

In line with the earlier standard, MLSSA uses a constant masking correction of -35 dB/octave. A MLSSA (v10w) “FULL” STI measurement of the specified speech filter yields an STI of 0.988. When the MTF matrix is adjusted using the method described by Stacey [7] to account for the new masking slopes at an SPL of 70dBA and the male/female factors, the STI becomes 0.947 for males and 0.98 7 of females. These values become the base-line for any STI measurements.

Band Lp of Test Signal	**Masking Slope dB/oct**	**Auditory Masking Factor**
30	-40	0.0001
46	-40	0.0001
56	-35	0.0003162
66	-25	0.0031623
76	-20	0.01
86	-15	0.0316228
96	-10	0.1

Table 1. Masking factors now used for STI.

Modulation Frequency	**Modulation m**	**Effective S/N**	
0.63	0.912	10.16	
0.8	0.871	8.29	Remove
1	0.812	6.35	frequencies
1.25	0.742	4.59	below 2 Hz
1.6	0.658	2.84	
2	0.571	1.24	1.24
2.5	0.477	-0.40	-0.40
3.15	0.383	-2.07	-2.07
4	0.287	-3.95	-3.95
5	0.204	-5.91	-5.91
6.3	0.151	-7.50	-7.50
8	0.132	-8.18	-8.18
10	0.144	-7.74	-7.74
12.5	0.184	-6.47	-6.47
average S/N		-0.62	-4.55
equivalent modulation		**0.479**	**0.348**

Table 2. Comparison of MTIs with frequencies below 2 Hz removed from average.

4 INVESTIGATION OF RELATIONSHIP BETWEEN SUBJECTIVE INTELLIGIBILITY AND STI

An investigation was conducted into the relationship between the subjective intelligibility of speech and the measured STI for a range of tonal balances (or frequency responses) in a reverberant environment. The method consisted of measurements of the STI for each response, subjective testing of word scores for each response, and processing of the measured STI results and word scores.

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.

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

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 change 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. This was 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. This was 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. Table 3 lists the scenarios and filters used for the tests. The frequency responses of the speaker when fed with the filter and the response of the filter itself are given in the Appendix.

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 3. Test Scenarios.

4.4 Data Processing

The word scores were then evaluated and compared against the STIs measured directly by MLSSA.

The measured STIs for each scenario were then adjusted to account for the revised variable masking thresholds and the male/female weightings. Before adjusting the STIs, the sound pressure levels for each scenario (and in each octave band) were adjusted to the level of 70 dB(A). The value of 70 dB(A) was chosen to represent a comfortable listening level for speech played over headphones.

The measured MTF matrices were then further adjusted with a weighting regimen, that changed the

a) importance of the modulation frequencies

b) weighting factors relating to upward spread of masking.

Five of the test words that were consistently misheard by the Australian listeners were analysed using MLSSA. For each of the contaminated and anechoic words, the audio spectrum was obtained. To examine the differences between the modulation spectra (in each octave band) of the anechoic and contaminated words, the waveforms of each word were filtered, squared and Fourier transformed. The resulting spectra were then normalised to the value at 0.46 Hz, being the first available data point.

5 RESULTS FROM DATA PROCESSING

5.1 STI Results

Table 4 lists the measured Full STI results for each scenario, and the Full, Male and Female STIs adjusted with the variable masking thresholds for 70 dB(A) and the male/female weightings. As a control the STI without the speech-weighting filter (and hence no masking correction) is given. In Table 4, masking factor AMF1 corresponds to a slope of -35 dB/octave, while AMF2 corresponds to the variable slopes of the revised version. 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.

	STI				
Scenario	**Full without speech filter & AMF1**	**Full with speech filter & AMF1**	**Full with speech filter & AMF2**	**Male with speech filter AMF2**	**Female with speech filter AMF2**
speech filter	NA	0.988	0.982	0.952	0.991
1	0.990	0.982	0.972	0.948	0.982
2	0.486	0.477	0.477	0.467	0.503
3	0.486	0.475	0.472	0.475	0.498
4	0.426	0.463	0.463	0.471	0.505
5	0.486	0.478	0.477	0.467	0.503
6	0.487	0.495	0.494	0.490	0.524
7	0.491	0.496	0.483	0.469	0.503
8	0.488	0.479	0.479	0.472	0.503
9	0.486	0.472	0.453	0.435	0.472

Table 4. STI values for test scenarios at right test ear.
AMF1= earlier masking slopes, AMF2 = revised masking slopes.

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. 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.

b) Minor changes to the measured Full STIs occurred when the speech filter was included, and were due mainly to the effect of the -35 dB/octave masking slope (AMF1).

c) The effect of the revised masking slopes (AMF2) on the Full STI results was comparatively small.

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.

5.2 Word Score Results

Fig 4 gives the word score results for each scenario. 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.

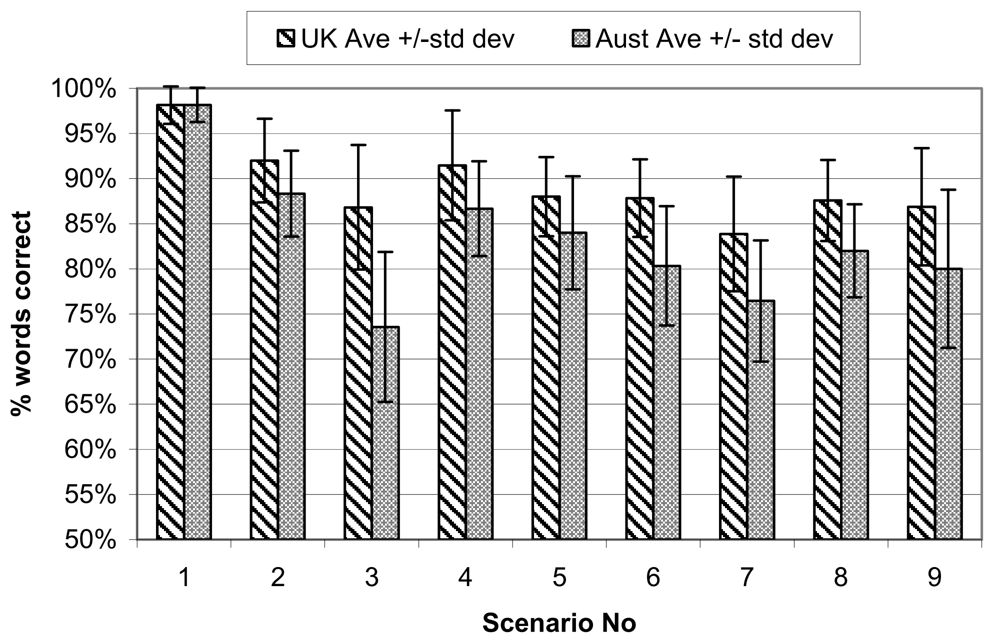


Fig 4. Graph of word score vs scenario. The error bars indicate a +/- standard deviation.

5.3 Comparison with Measured STIs

The word scores were converted to an equivalent STI value using the common intelligibility score (CIS). Fig 5 shows the equivalent word score STIs and measured STIs after modifications with the revised masking slopes and the male/female weightings.

The following comments are made.

a) The STI of 1 relating to the error bars 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, there was a reduction in the STI of between 0.01 and 0.1 for both UK and Australian listeners.

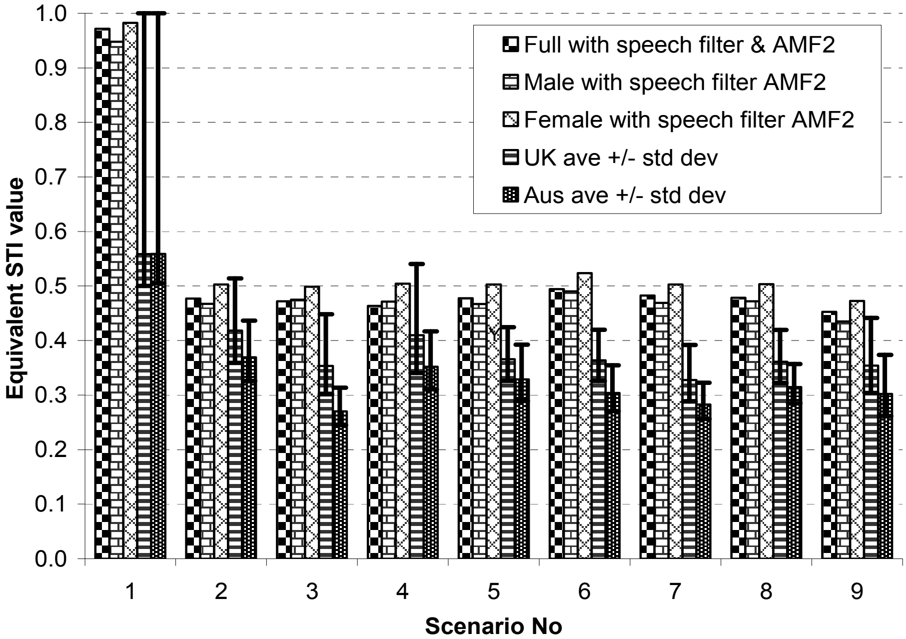


Fig 5. Comparison of equivalent STIs of PB word scores with Full, Male and Female STIs with revised weighting factors. The error bars show the range of standard deviations.

5.4 Spectra of Different Words

Five of the words that were found to be problematic were: “sheik”, “mass”, “quip”, “dung” and “salve”. These words were characterised by their relatively soft end consonants which was overpowered by the longer preceding vowels and consonants. To indicate the differences between the spectra of these words and the long term spectrum of speech, each word’s spectrum was normalised to the spectrum of the speech filter spectrum and scaled to approximately 0 dB. Fig 6 shows the difference in spectra. In particular, “sheik” and “quip” showed a relative lack of energy in the 600 to 2.3 kHz region of some 12 to 20 dB, while “salve” and “dung” showed a relative lack of some 10 dB in the region 1.6 kHz to 5.5 kHz.

The absence of energy in these spectral areas will cause the masking effects (both from a downward and upward perspective) to be significantly higher in real speech than simulated by the standard speech weighting filter. For words of these spectra, the true STIs will be lower than predicted using the standard speech spectrum.

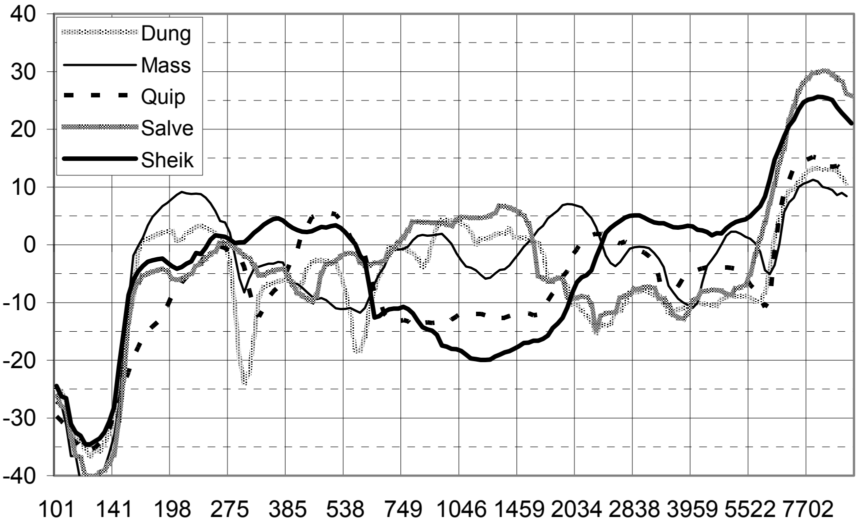


Fig 6. Difference between the spectrum of the MLSSA speech filter (and IEC specified STI speech filter) and the spectrum of each problematic word.

The octave band sound pressure levels of the speech-filtered MLS signal in Scenario 2 were replaced with the octave band levels of the word “sheik” and the overall was scaled to 70 dB(A). The STIs were then computed and were found to differ little with those of the speech filtered signal. Table 5 shows the change in STIs.

Spectrum	Full with speech filter & AMF2	Male with speech filter AMF2	Female with speech filter AMF2
speech filter	0.477	0.467	0.503
"sheik "	0.474	0.463	0.499

Table 5. Computed STIs for Scenario 2 (no tonal filter) with the spectra of speech filter and the word “sheik”.

5.5 Modulations of Specific Words

Fig 7 shows the relative modulations for the anechoic and contaminated versions of the word “quip”. For each of the words, the difference in modulations between the anechoic and contaminated versions was found in each octave band. Those differences were then averaged over the five words and are shown in Fig 8.

The average difference in each octave band shows that the modulation frequency range most affected by the reverberant contamination lies between 1.25 Hz and 6.3 Hz, with the average of maximum differences lying near 3.15 Hz.

Comparison of Fig 8 with the data of Fig 2 shows that the spectral shapes of the reduction in modulation bear a similarity to the shapes of the importance of the modulation frequencies to overall intelligibility.

This similarity suggests that weighting the SNR's in the MTF matrix according to importance might be useful.

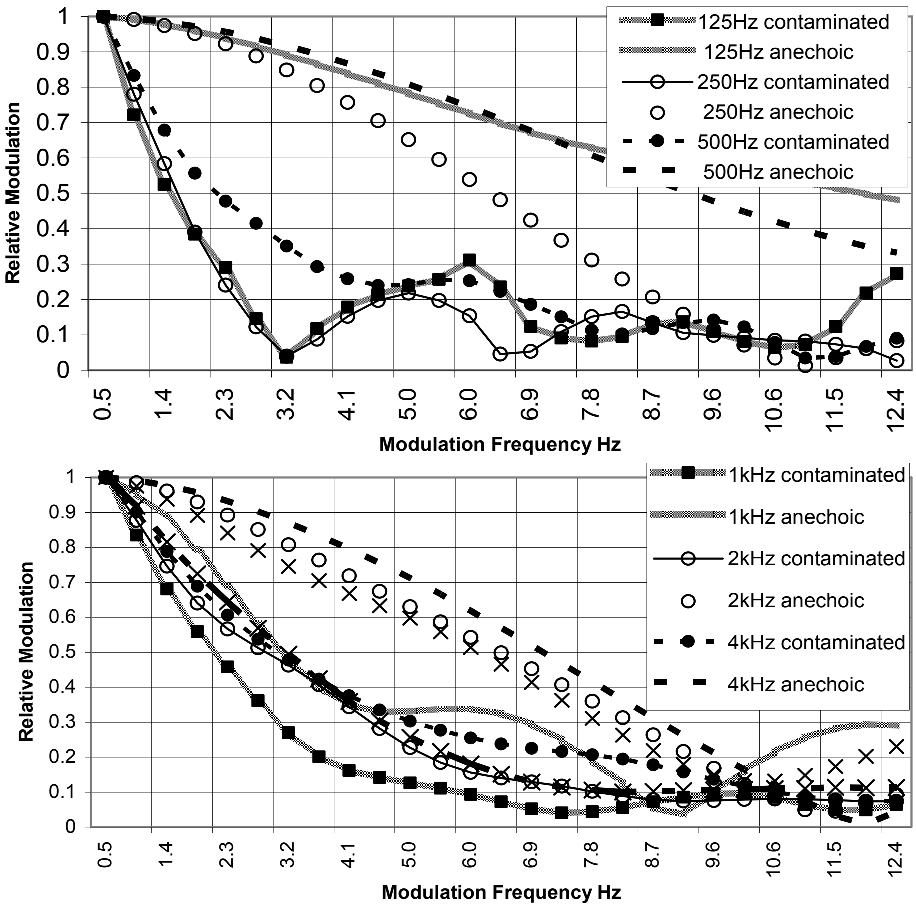


Fig 7. Comparison of normalised modulation spectra for the anechoic and contaminated forms of the word "quip".

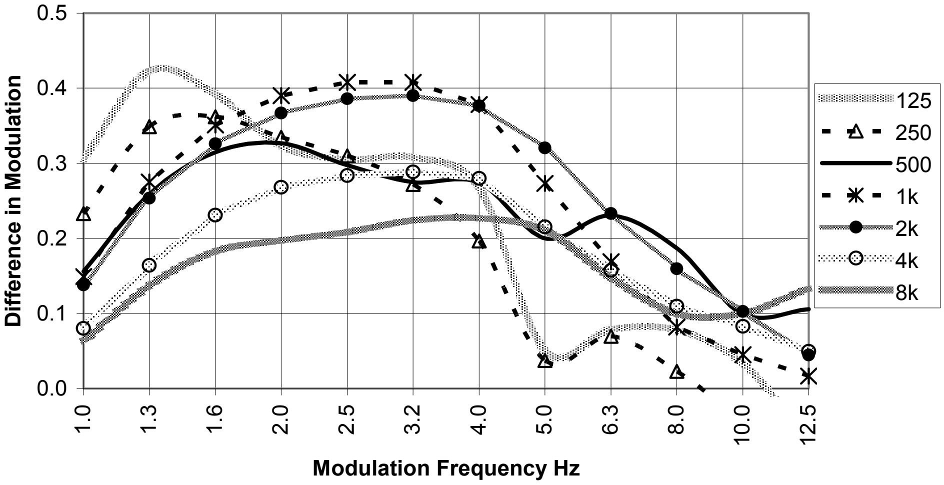


Fig 8. Average over the five words of difference between the modulations of the anechoic and contaminated versions

6 ADJUSTMENTS TO THE MTF MATRIX

6.1 Weighted Averaging of Modulation SNRs

To test the sensitivity of the STIs of each scenario to changes in the relative importance of each modulation frequency, a crude weighting system was composed for the MTF matrix. The weights are based on the octave band difference data of Fig 8 and modulation spectrum data from Fig 2 and [8], and indicate our guess of the relative importance of each frequency. Table 6 gives the specific numerical weights that were applied to the relevant modulation frequency.

frequency Hz	standard weighting	alternate weighting
0.63	1	0.08
0.8	1	0.16
1	1	0.41
1.25	1	0.77
1.6	1	1.22
2	1	1.50
2.5	1	1.62
3.15	1	1.62
4	1	1.60
5	1	1.50
6.3	1	1.30
8	1	1.01
10	1	0.73
12.5	1	0.49
sum	**14**	**14**

Table 6. Numerical weighting factors for modulation frequencies in MTF matrix.

Band Lp of Test Signal	Revised	Alternate
30	-40	-30
46	-40	-30
56	-35	-25
66	-25	-15
76	-20	-10
86	-15	-10
96	-10	-10

Table 7. Alternate masking slopes.

6.2 Masking Slopes

Masking data from [9] for a (i) 400 Hz masking tone over other tones and (ii) for a 90 Hz band of masking noise centred at 400 Hz over other tones, indicate upward masking slopes of between -10 dB/octave and -20 dB/octave when the masking signal is of the order of 65 dB. These figures contrast with the slope of -25 dB/octave listed in Table 1.

Allen[10] who also extensively refers to [9] suggests that a complete quantitative model of masking has not yet been developed.

To test the effect of different masking slopes on the STIs, the masking slopes were reduced and applied to each scenario. Table 7 lists the revised slopes (ie specified) and our crude alternative slopes.

6.3 Reloading the Matrix

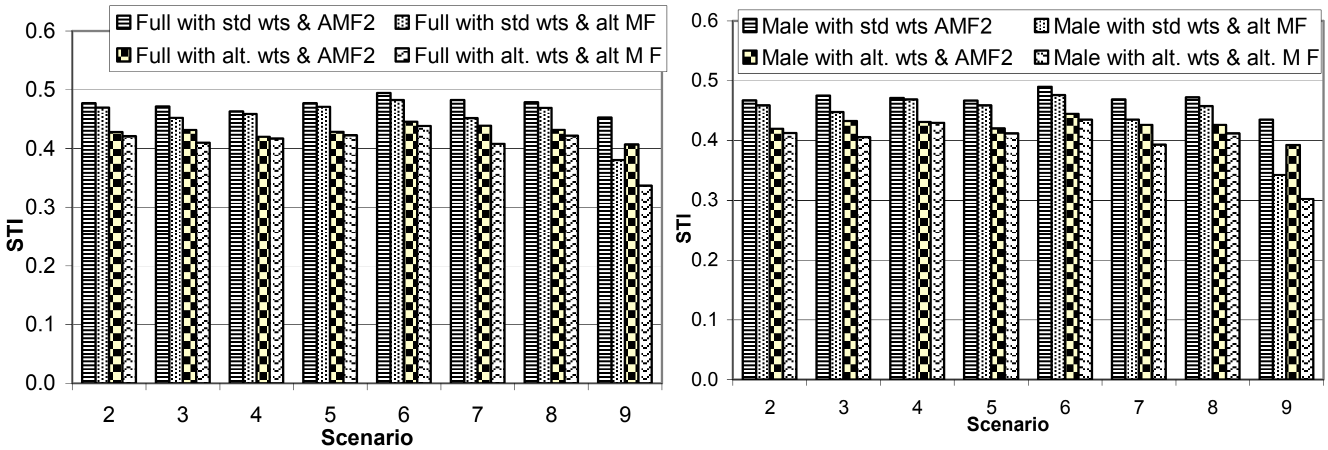
The measured MTF matrix for each scenario was reloaded with the following different combinations of factors:

i) Standard weights (ie unity) for the modulation SNRs with alternate masking slopes given Section 6.2.

ii) Alternate weights for the modulation SNRs in Section 6 .1 with the revised masking slopes.

iii) Alternate weights for the modulation SNRs with alternate masking slopes.

Fig 9 compares the Full, Male and Female STIs for each reloaded matrix with the measured STIs (after adjustment for the revised masking slopes and 70 dB(A)). While the reloaded STIs are all lower than the measured/adjusted STIs, the reduction is not sufficient to bring agreement between the reloaded STIs and the subjective equivalent STIs for each scenario.



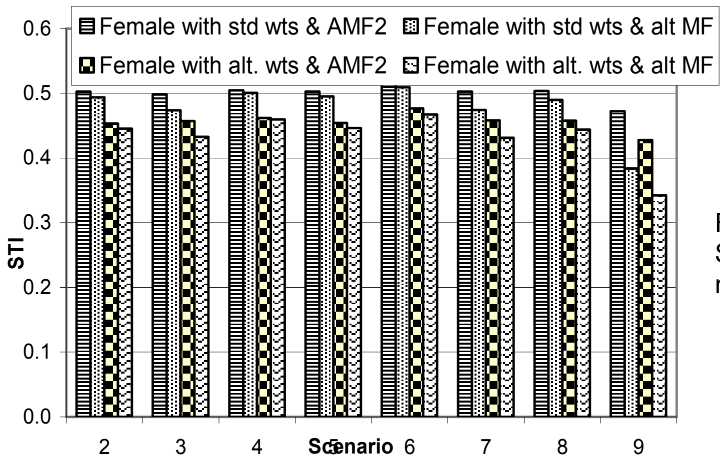


Fig 9 Comparison of Full, Male and Female STIs for each reloaded matrix with measured/adjusted STIs.

Greenberg[4,6] has suggested that intelligibility is based on both the phase and amplitude components of the modulation spectrum. Both our study here and the STI process itself used amplitude information of the modulation spectrum.

For the three combinations above, the average reduction in STI was found over scenarios 2 to 9 and is shown in Table 8. The reductions are surprisingly consistent between each STI type.

STI type	Standard modulation weights and revised masking factors	Alternative modulation weights and revised masking factors	Alternative modulation weights and alternative masking factors
full	0.025	0.044	0.068
male	0.020	0.046	0.065
female	0.024	0.045	0.068

Table 8. Reductions in STI for three combinations of reloaded MTF matrix, averaged over scenarios 2 to 9.

7 CONCLUSIONS

- An investigation was conducted to attempt to shed some light on how the construction of the MTF matrix might relate to the apparent lack of correlation between STI and frequency response changes.
- The MLSSA analyser v10w and earlier, uses the original fixed -35 dB/octave masking slopes when computing STI. These slopes have been revised to be level dependent.
- A number of researchers have concluded that modulation frequencies below 2 Hz do not contribute greatly to intelligibility.
- In reverberant environments, the process of simple averaging of SNRs that is used to form the octave band MTIs, allows the low modulation frequencies to dominate the MTI. As these frequencies are less important than mid frequencies, this can lead to STIs that are excessively high. Removal of the modulation frequencies below 2 Hz from the MTF matrix will cause a major reduction in the STI in reverberant environments.
- Objective and subjective testing of intelligibility was undertaken in a reverberant environment with seven different filters being used to shape the frequency response of a loudspeaker system.
- 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.
- The word scores for the Australian listeners were lower than those of the UK listeners.
- The word scores would have been lower had the articulation been poorer or had the listeners concentrated less.
- As expected, there were only minor differences in the corrected STIs for each frequency response scenario, and these were due to the revised masking slopes.
- The reduction in the equivalent word score STI when the filters were used was not reflected in the STI measurements.
- The spectrum of five of the words used in the test showed differences of up to 20 dB with the standard speech spectrum used for the STI measurements. While the measured STI for a system that only reproduced these words would be in error, the limited effect of the masking slopes means that these errors are minor.
- The average differences between the modulations of five test words when played anechoically and in one of the filtering scenarios, showed that modulation frequencies around 3 Hz were the most damaged by the reverberant environment and filtering.
- The similarity between the spectral shape of the modulation damage and the relative importance of each modulation frequency suggests that weighting the SNRs in the MTF matrix according to importance might be useful.
- It appears that a complete quantitative model of masking has not yet been developed.
- When the MTF matrix for each filter scenario was reloaded with crude weightings for the SNRs in each octave band and crudely decreased masking slopes, the STIs were reduced. The reductions in STIs were insufficient to match the equivalent STI of the word scores.
- More work is needed on the STI process if STI is to reliably quantify the effects of tonal balance on subjective intelligibility and listener comfort. One prominent researcher[4] has suggested that intelligibility is based on both the phase and amplitude components of the modulation spectrum, and this would imply that there is still significant work required.

REFERENCES

1 IEC 6028-16 2nd Edition 1998.

2 Drullman R, Festen JM, Plomp R. 'Effect of temporal envelope smearing on speech reception'. J. Acoust. Soc. Am., 95: pp1053-1064, 1994.

3 Arai T, Hermansky H, Pavel M, Avendado C. 'Intelligibility of speech with high-pass filtered time trajectories of spectral envelopes' Proc. I: International Conference on Spoken Language Processing, Philadelphia, 1996.

4 Greenberg S, Arai T. 'The relation between speech intelligibility and the complex modulation spectrum'. Submitted to the 7th European Conference on Speech Communication and Technology, Eurospeech - 2001.

5 'Past, present and future of the Speech Transmission Index'. ed Wijngaarden S. TNO Human Factors Soesterberg Netherlands, 2002.

6 Greenberg S, Hollenback J, Ellis D. 'Insights into spoken language gleaned from phonetic transcription of the switchboard corpus'. International Conference on Spoken Language Processing, Philadelphia, pp S32-35, 1996.

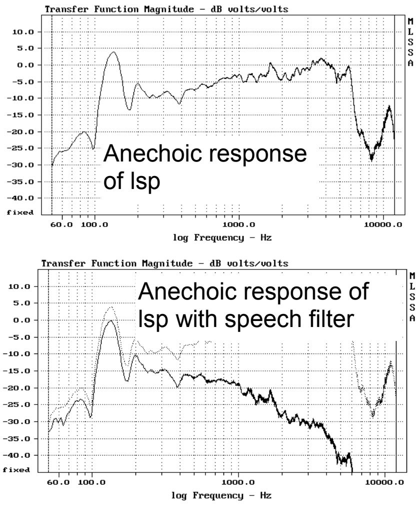
7 Stacey T. 'Checking the accuracy of MLSSA STI Measurements'. Proceedings of IOA Reproduced Sound 18, 2002.

8 Greenberg S, Arai T, Silipo R. 'Speech intelligibility derived from exceedingly sparse spectral information'. International Conference on Spoken Language Processing, pp2803-2806, 1998.

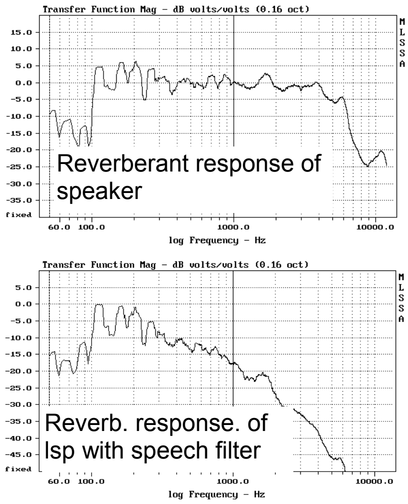
9 Egan J, Hake H. 'On the masking pattern of a simple auditory stimulus'. J. Acoust. Soc. Am., 87 No 6: pp622-630.

10 Allen J. 'Psychophysical models of masking for coding applications' (draft - source unknown) AT&T Labs, 1998.

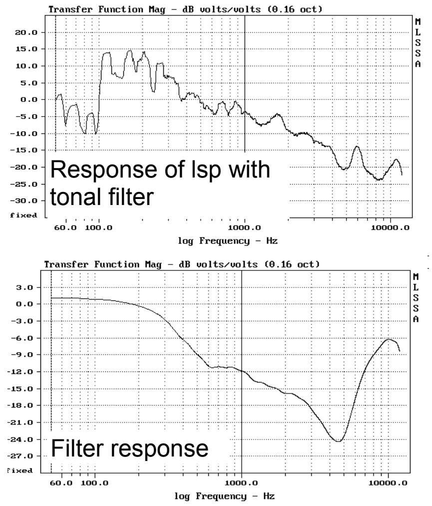
APPENDIX FREQUENCY RESPONSES OF FILTER SCENARIOS



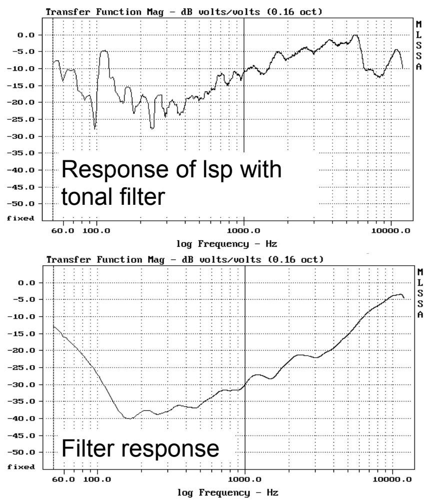
Scenario 1



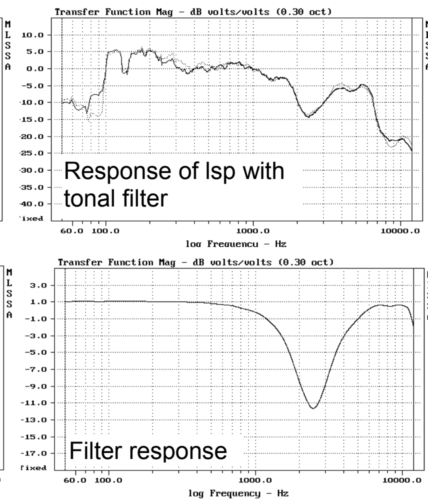
Scenario 2



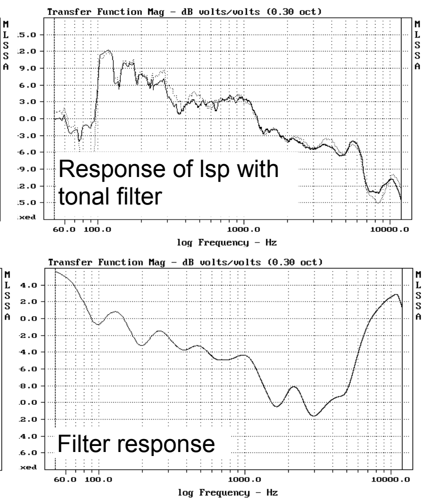
Scenario 3



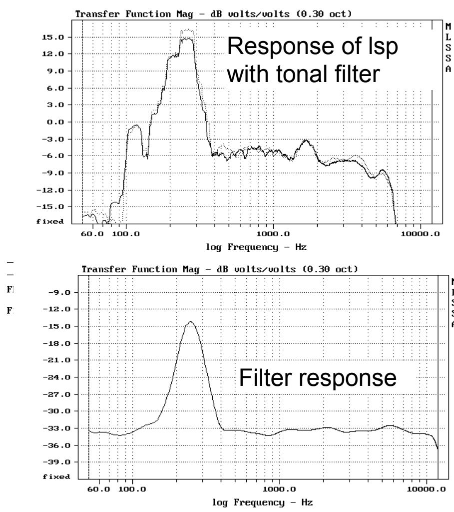
Scenario 4



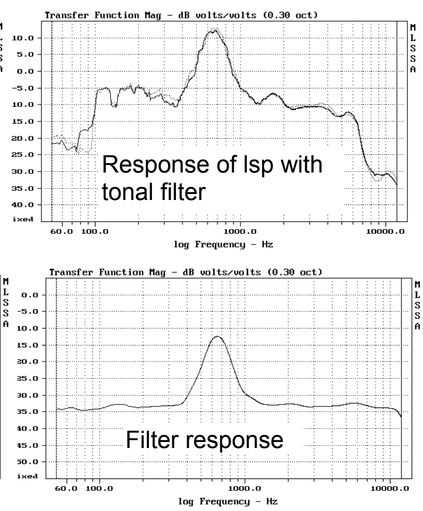
Scenario 5



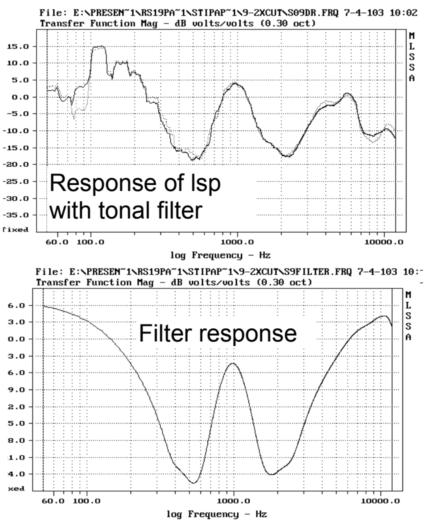
Scenario 6



Scenario 7



Scenario 8



Scenario 9