Audio Engineering Society

Convention Paper

Presented at the 128th Convention
2010 May 22–25 London, UK

Further Investigations into Improving STI's Recognition of the Effects of Poor Frequency Response on Subjective Intelligibility.

Glenn Leembruggen[1], Marco Hippler[2] and Peter Mapp[3]

[1] Acoustic Directions, ICE Design Sydney Australia and University of Sydney
Glenn@acousticdirections.com

[2] University of Applied Sciences Cologne Germany

[3] Peter Mapp and Associates Colchester UK

ABSTRACT

Previous work has highlighted deficiencies in the ability of the STI metric to satisfactorily recognise the subjective loss of intelligibility that occurs with sound systems having poor frequency responses, particularly in the presence of reverberation. In a recent paper, we explored the changes to STI values resulting from a range of dynamic speech spectra taken over differing time lengths with different filter responses. That work included determining the effects on STI values of three alternative spreading functions simulating the ear's upward masking mechanism. This paper extends that work and explores the effects on STI values of two masking methods used in MPEG-1 audio coding.

1. INTRODUCTION

The speech transmission index (STI) (1), (2) has gained international acceptance as a useful measure of the ability of a transmission path to faithfully transmit intelligible speech.

However, a number of acoustical engineers working in the design and commissioning of sound systems for major public spaces have noted that significant degradation occurs in subjective speech intelligibility when the frequency response of the sound system at the listener is poor (3), (4). Relatively small changes to the frequency response, sometimes as small as 1 dB, can noticeably affect the intelligibility of conversational speech and the degree of listening concentration that is required. These improvements are not associated with increases in the measured STI values and can be found in relatively low-noise situations.

This issue was first explored by Leembruggen and Stacey at a 2003 IOA conference (4), in which tests using speech reproduced with gross frequency response variations showed much lower subjective intelligibility scores than the STIs which were measured for the system with those frequency responses.

Possible reasons for this mismatch between subjective and objective intelligibility are:

- STI's use (2) of a long term speech spectrum rather than short term spectra. (Short-term spectra show significant differences to the long term spectrum)
- The masking function used in STI may not properly model the ear's psycho-acoustic mechanism of self-speech masking.

The ability of the Speech Intelligibility Index (SII) (5) to satisfactorily account for self-speech masking with the frequency response filters of (4) was investigated by Leembruggen in (6), with little change to the SII being observed with the filtered speech.

Subsequent work (7) presented at a 2009 IOA conference explored the effects on STI values of short-term speech spectra and three alternative masking functions, when those spectra were shaped with the above-mentioned frequency responses with gross variations.

Those three alternative masking algorithms were:

1. The spreading function used by the SII was built (8), (9) on masking curves developed by Schroeder in 1963 (10), (11). This method uses the slope of the spreading function to compute the upward-masking for the speech in each one-third octave band. The equivalent masking noise in each one-third octave band was then reduced to octave-wide bands.
2. The excitation pattern (EP) model of hearing developed by Moore and Glasberg (12), (13), (14), (15), (16), (17). For each filtered speech spectrum, the EP was computed, from which the self-speech masking levels were found and converted to signal to noise ratios in each octave band.
3. Using slopes derived from the excitation pattern model of Moore and Glasberg, the masking was computed in each third-octave band. This method is similar to the SII method.

The work presented in this paper follows directly from work described in (4) and(7).

2. PRIOR WORK IN 2003

To give this paper a suitable context, details of the initial work undertaken in 2003 are reproduced from (4).

2.1. Measurement Procedure

A loudspeaker and dummy head with binaural microphones were set up in an anechoic chamber. The response of the speaker was then measured at each ear with binaural microphones at a distance of 1.5 m from the speaker on axis and processed by MLSSA v10w to yield the loudspeaker's anechoic frequency response of the speaker and the system STI. The system was then relocated to a reverberation chamber. Again the system STI was measured at a distance of 1.5 m from the speaker and 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 (Filter shapes 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 a speech-weighting filter connected in series with the response-shaping filter.

2.2. Subjective Procedure

A CD of anechoically recorded female speech was prepared and consisted of 1000 carrier sentences with single-syllable, phonetically-balanced (PB) words situated at the end of each sentence. Three groups of 50 words were then played through the speaker in the anechoic chamber (Filter shape 1) and recorded on the dummy head at a distance of 1.5 m from the loudspeaker. The system was relocated to a reverberation chamber and another thee groups of words played through the loudspeaker and recorded binaurally at a distance of 1.5 m (Filter shape 2). For each of the seven response-shaping filters, three lists of 50 words were replayed and recorded for filter shapes 3 to 9. When the groups were exhausted, a reshuffled version of the lists was used.

The recordings of the nine shapes 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 shapes. In Australia, three listeners evaluated all of the three lists for each of the nine shapes. The sentences were presented to listeners through headphones, and the listener wrote down the word at the end of the sentence. The playback level of the recordings was approximately 70 dBA at the listener's ear for all filter shapes.

2.3. Filter Shapes

The frequency responses of the tonal filters were chosen to exaggerate subjective listening difficulties. Figure 1 shows the relative frequency responses of those filters, and to allow easier comparison, each response is normalised to its value at 1 kHz.

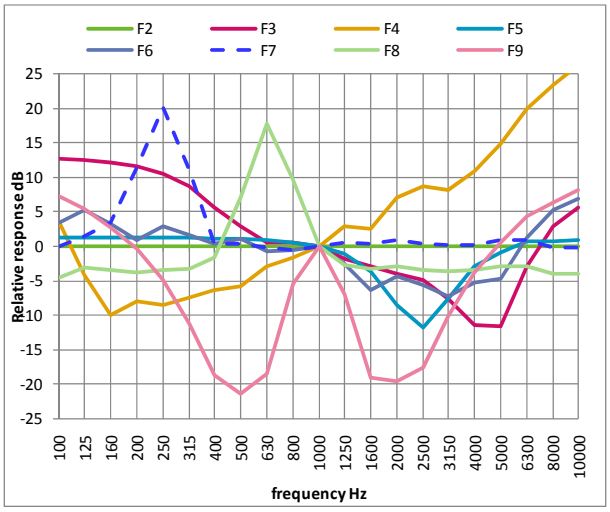


Figure 1. Relative frequency responses of filters used to modify the speech spectrum. Each response is normalised to its value at 1 kHz

2.4. Word Score Results

Figure 2 gives the word score results for each filter shape.

The following comments are made.

1. 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.
2. The average Australian scores for each filter shape were generally lower than the corresponding UK scores. This was likely to result from accent differences.

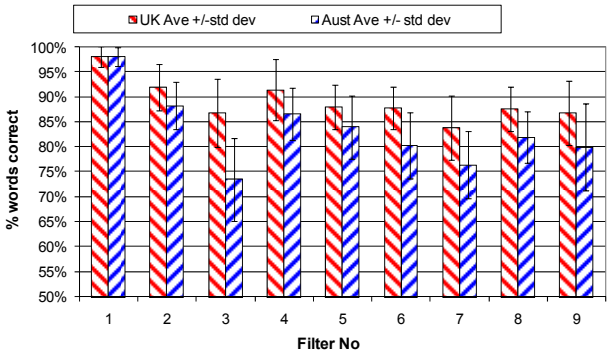


Figure 2. Word scores for the different filter shapes. Note that Filters 1 and 2 have flat responses and Filter 1 is anechoic, while filter shapes 2 to 9 are reverberant. The error bars show the standard deviations

3. The UK and Australian average scores showed a similar trend over the range of filter shapes.
4. There was a noticeable reduction in the word score with the filters inserted.
5. 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.
6. The Australian listeners found the process to be tiring, and yet the measured STI was of the order of 0.5, which is a value that is typically specified for sound systems.

2.5. Comparison with Measured STIs

The word scores were converted to an equivalent STI value using the common intelligibility score (CIS). Figure 3 shows those word-equivalent STI values and measured values of male STIr, calculated according to the 2003 STI IEC standard. Although the talker was a female and the STI measurements are male, that difference does not change the results appreciably.

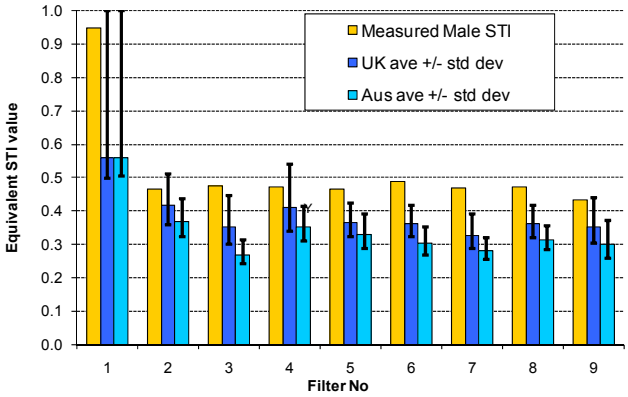


Figure 3. Comparison of equivalent STIs of PB word scores with measured Male STIrs. The error bars show the range of standard deviations

The following comments are made.

- In filter shape 1 (anechoic and flat response), the error bars extend up to an STI of 1. This is caused by the CIS conversion which amplifies the STI when word scores exceed 97%. At this value, a 3% change in PB score results in a STI change from 0.55 to 1.0.
- The word scores are always lower than the measured STIs.
- The effect of talker accent on intelligibility can be seen. Wijngaarden et al (18) noted that non-native talkers and listeners require a higher STI score for similar intelligibility than with native listeners.

3. METHODOLOGY

As the speech spectra used in (7) were re-used for this work, information from (7) regarding their preparation is reproduced below.

3.1. Speech Spectra

Six talkers (5 male, 1 female) were recorded anechoically and a 10 second segment of each talker extracted. The anechoic data was then reverberated using FIRverb software with a reverberation time of approximately 2 s in each octave bandwidth. The spectra of each talker in specific time slices was then found using scan analysis provided by the waterfall function in the software WinMLS2004.

The following spectra were found using a Hanning window with 50% overlap for each talker for both the anechoic and reverberant environments.

- 10 slices of 1 s length
- 40 slices of 250 ms length
- 200 slices of 50 ms length

3.2. Preparation of Spectra for Analysis

The speech spectra were then prepared for analysis as follows:

1. All spectra were bundled into one-third octave bands.
2. The total rms level of the ten one-second slices was computed for each talker to form the long-term L_{eq} in each one-third octave band.
3. The long-term L_{eq} levels were A weighted and summed to give the long term LA_{eq} of each talker and normalised to the long-term operational speech level of 75 dBA. The resulting normalization factor D was stored for subsequent use.
4. Each of the 1 second, 250 ms and 50 ms time-slice spectra was then adjusted by the normalization factor D.
5. To ensure that the statistics were not skewed by spectra representing soft syllables or gaps between words, any spectrum whose total level was less than 50 dBA was removed from the analysis

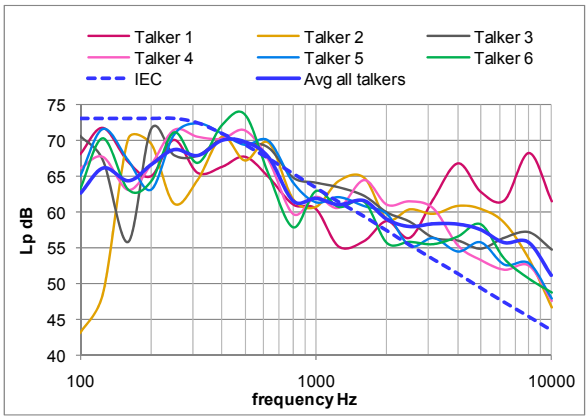


Figure 4. Comparison of long term L_{eq} reverberated spectra of 6 talkers and their average with IEC spectrum.

Figure 4 compares the long term L_{eq} in one-third octave bands with the IEC spectrum interpolated from (2).

An example of the range of short-term spectra is given in Figure 5, which compares the IEC spectrum with 1/3rd octave band data for Talker 1 in the reverberant environment.

The following data is shown for the 50 ms time slices:

- normalised IEC spectrum
- mean level of each 1/3rd octave band
- 10th percentile of each 1/3rd octave band
- 90th percentile of each 1/3rd octave band
- spectrum of an individual time-slice showing strong differences with the IEC spectrum

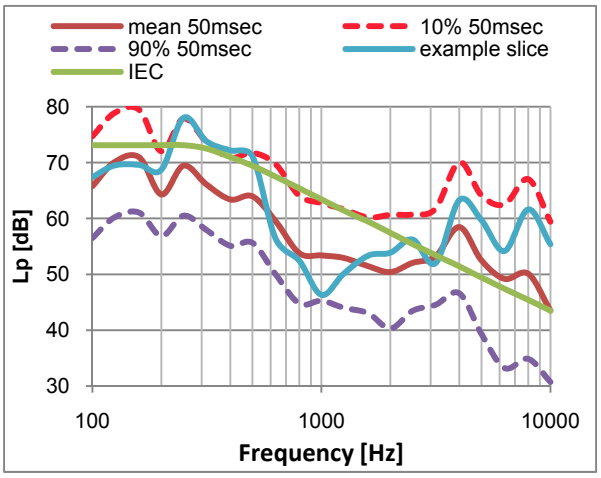


Figure 5. Statistics of speech spectra in 1/3rd octave bands in reverberant environment for 50 ms time slices.

4. MASKING ALGORITHMS

4.1. Loss of SNR due to Masking

Two new masking models have been used to compare their relative impacts on STI values with the range of filtered speech spectra described above:

1. Psycho-acoustic Model 1 used in MPEG-1 encoding.
2. Psycho-acoustic Model 2 used in MPEG-1 encoding

Each of the models produces an equivalent self-masking noise term in each octave band. These terms are then used to adjust the measured modulation indices.

The STIs of these two models are compared with the STIs using the following four masking models used in (7).

- model used in STI.
- model used in the Speech Intelligibility Index SII.
- model using excitation patterns (of Moore and Glasberg) with 0.1ERB resolution and 1/3rd octave spectral lines with and without ear-filtering.
- model with and without ear-filtering derived from slopes of the excitation patterns computed by the excitation pattern model of Moore and Glasberg.

4.1.1. Calculating the masking in STI

To determine the auditory masking level in say octave band *k*, the sound pressure level of the speech and ambient noise in the preceding octave band *k*-1 must first be found. Using the relationships between the acoustic level and the associated masking level given in Table 1, the equivalent masking noise *am*dB is found for band *k*.

As the auditory masking factor amf is an intensity parameter, Eqn 1 is used to convert the *amdB* into that form. Eqn 2 is then used to calculate the intensity of the audio masking signal in each octave band.

$$amf = 10^{(\frac{amdB}{10})} \quad \text{Eqn 1}$$

$$I_{am,k} = I_{k-1} * amf \quad \text{Eqn 2}$$

where:

$I_{am,k}$ is the audio masking intensity in octave band k

I_{k-1} is the intensity of the signal in octave band $k-1$

Item	Range 1	Range 2	Range 3	Range 4
Octave band level *Lk-1* dB SPL	< 63	≥ 63 and < 67	≥ 67 and < 100	≥ 100
Auditory masking *amdB*	0,5 × L k-1 - 65	1,8 × L k-1 – 146.9	0,5 × L k-1 – 59.8	-10

Table 1 Auditory masking levels as a function of the acoustic octave band level.

The masking Intensity $I_{am,k}$ is then used to adjust each modulation index m_{kf} as per Eqn 3.

$$m'_{kf} = m_{kf} \frac{I_k}{I_k + I_{am,k} + I_{rs,k}} \qquad \text{Eqn 3}$$

where

I_k is the intensity of the signal in octave band k

$I_{rs,k}$ is the absolute reception threshold which is not discussed further.

4.2. MPEG-1 Masking Models

The spreading function that describes the upward and downward spreading of masking relates to the difference in frequency between the masker and the maskee. That difference is expressed in Barks, which is a measure of the ear's critical bandwidth (19) p 182.

Eqn 4 gives the relationship between Barks and frequency in Hertz.

$$Z = 13\arctan\left(\frac{0.76f}{1000}\right) + 3.5\arctan\,(f/7500)^2 \qquad \text{Eqn 4}$$

where Z is the frequency band in Barks, f is the frequency in Hz.

Integer Bark numbers represent the lower frequency of a critical band.

4.2.1. Psycho-acoustic Model 1

The two-piece model of psycho-acoustic spreading is given in Equation 1 from (19). Different slopes are used depending on the power spectrum level of the masking frequency and whether the frequency of the maskee is within plus or minus half a critical band from the masker.

The model is meant to mimic the masking data for tones masking tones. (19) p 188.

The total spreading function resulting from a filtered spectrum is the sum of the intensities of the maskees in each bark interval.

$$\left\{\begin{array}{ll} L_{sf}(f) = (0.4\mathrm{L_m} + 6)\Delta z & \text{for } -1 < \Delta z < 0 \\ L_{sf}(f) = -17 + \Delta z - 0.4\mathrm{L_m} + 11 & \text{for } \Delta z < -1 \\ L_{sf}(f) = -17\Delta z & \text{f or } 0 < \Delta z < 1 \\ L_{sf}(f) = (0.15\mathrm{L_m} - 17)\Delta z - 0.15\mathrm{L_m} & \text{for } \Delta z > 1 \end{array}\right\} \qquad \text{Eqn 5}$$

where:

$L_{sf}(f)$ is the level of the maskee frequency

Lm is the level of the masking frequency

Δs is the difference in the Bark of the masking and maskee frequencies

The psycho-acoustic spreading function described by this model is illustrated in Figure 6 for a number of masker levels at 9 Bark.

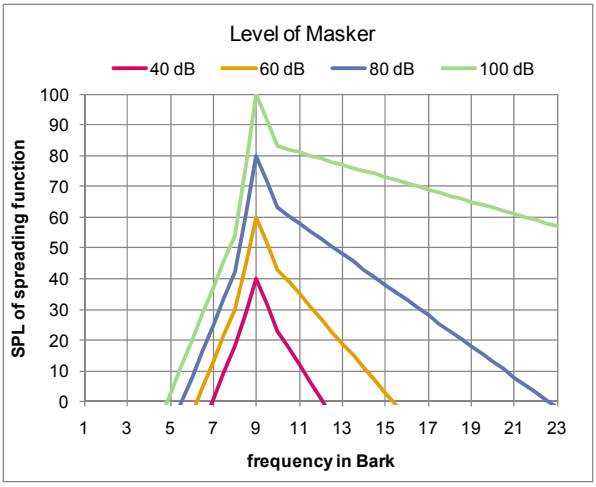


Figure 6. Plot of spreading function of Model 1 with a range of masking levels centred at 9 Bark.

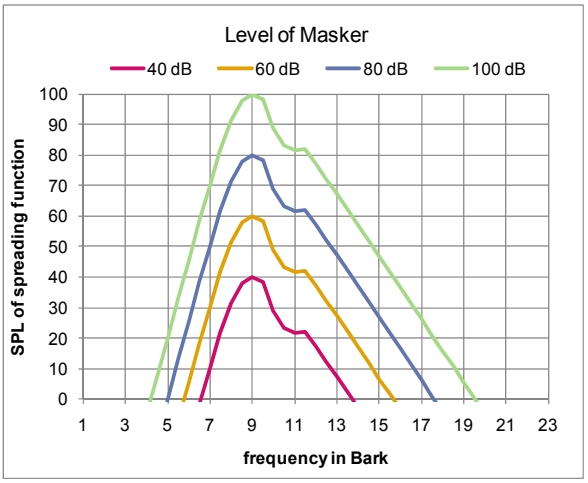


Figure 7. Plot of spreading function of Model 2 with a range of masking levels centred at 9 Bark

4.2.2. Psycho-acoustic Model 2

The model of psycho-acoustic spreading is given in Eqn 6 from (19) and is derived from the Schroder spreading function (19) p185. In this reference, the authors note that this spreading function is not level dependent, in contrast with the clear level-dependence seen in experimental data.

$$L_{sf}(f) = L_m + 15.8111389 + 7.5(1.05\Delta z + 0.474) - 17.5\sqrt{(1 + (1.05\Delta z + .474)]^2} + 8 * \text{MIN}(0, (1 + 1.05\Delta z - 0.52 - 2(1.05\Delta z - 0.5)$$

Eqn 6

where

$L_{sf}(f)$ is the level of the maskee frequency

Lm is the level of the masking frequency

Δs is the difference in Bark of the masking and maskee frequencies

The total masking level resulting from a spectrum spread over N Bark is the sum of the maskee intensities in each Bark interval resulting from N spreading functions.

The psycho-acoustic spreading function described by this model is illustrated in Figure 7 for a number of masker levels at 9 Bark.

5. COMPUTATION OF STI WITH THREE MASKING MODELS

The STI values were computed with the two MPEG-1 models for the range of filters and speech spectra, and compared with those obtained in (7).

5.1. Adjustments to the Measured MTF Matrix

As; i) the MLSSA analyser used to measure the MTF matrices in (4) had applied masking to those matrices, and ii) some SNRs were less than 30 dB, the MTF matrices were de-noised and then de-masked, by applying the inverse of the specified masking abd noise adjustments.

5.2. Preparation of Spectra for STI Calculations

The 1/3rd octave speech spectra were prepared for insertion into the STI calculations as follows:

1. The process described above in Section 3.2 was used.
2. All adjusted spectra were logarithmically summed into octave bands for inputting into the STI algorithm as the Speech Signal.
3. Each 1/3rd octave spectrum was converted into a 1 Bark wide spectrum using Eqn 4.

The process of (7) used both anechoic and reverberated spectra and found that the differences in STI values were relatively small between the two types of spectra. As these differences yielded little additional information, it was decided that only the anechoic spectra would be used for this paper.

5.3. Inclusion of Background Noise

Noting Steinbrecher's concerns in (20), a realistic amount of background noise was introduced into the calculations of STI. A noise spectrum corresponding to NR20 (approximately 33 dBA) was used to ensure that under operational situations where background noise is almost universally present, the reduction in signal to background noise ratio due to a depressed frequency response was accounted for.

5.4. Computing STI using STI masking

The STI was calculated using the STI masking model for each processed time-slice spectra and talker.

5.5. Computing STI using MPEG-1 masking

For each processed time-slice spectra in a given Bark interval, the masking intensity levels in all other Bark intervals were computed using Eqn 5 and Eqn 6. The total intensity in each Bark was then calculated.

The total masking levels in each Bark were allocated/divided into the appropriate octave bands to yield the masking noise in octave bands. Those noise levels were converted back to intensity $I_{am,k}$ and using Eqn 3 to adjust each modulation index m_{kf}, the STI was calculated for each processed time-slice spectrum and talker.

6. RESULTS

6.1. Comparison of Masking Levels

Figure 8 compares the octave-band sound pressure levels of a selected speech spectrum with Filter 9 with the masking noise levels produced by the STI, SII, EP slope (filtered and unfiltered) MPEG-1 Model 1 and MPEG1-Model 2 algorithms.

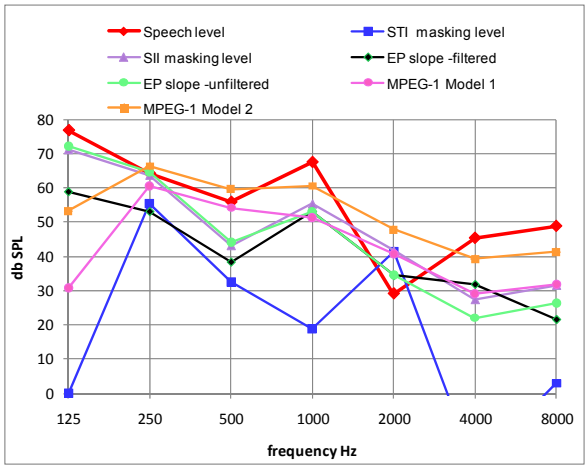


Figure 8. Comparison of masking-noise levels produced by six different masking algorithms

6.2. STIs with IEC Speech Spectrum

Figure 9 compares the STIs of the three methods for the IEC speech spectrum given in (2). The following trends are observed:

- The differences in the STI values approximately range from 0.07 to 0.14.
- MPEG-1 Model 2 masking method yields the lowest STI values.
- Within a masking scheme, the changes in STI values with filter shape are relatively minor. Only Model 2 with Filter 9 shows a significant change.

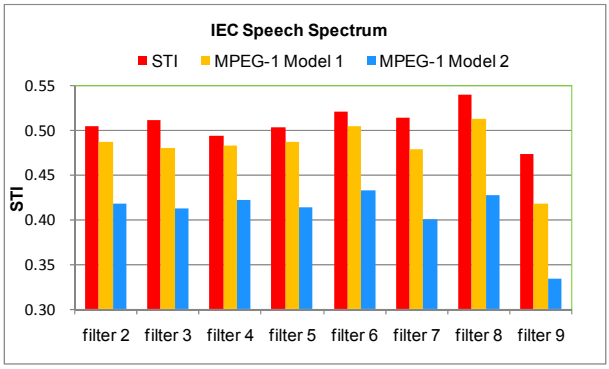


Figure 9. Comparison of STIs predicted by the six masking models with the IEC speech spectrum.

The octave band MTI values were examined for the eight filter shapes to help understand the contribution of each octave band to the overall STI value. Comparisons of the MTIs for Filters 3 and 9 are given in Figure 10 and Figure 11which shows some of the extremes of the overall behavior.

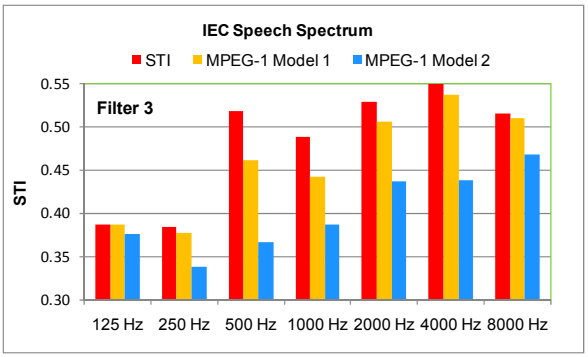


Figure 10. Octave band MTI values of Filter 3 for six masking methods with IEC spectrum

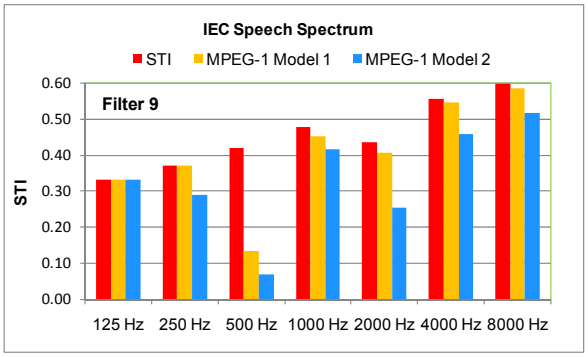


Figure 11. Octave band MTI values of Filter 6 for six masking methods with IEC spectrum

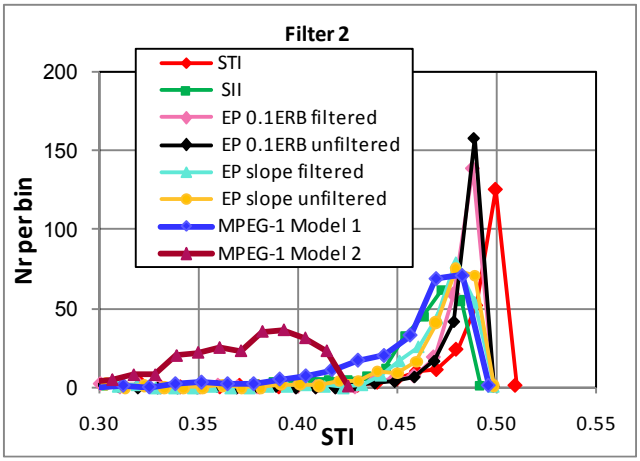
6.3. STIs with spectra of talkers

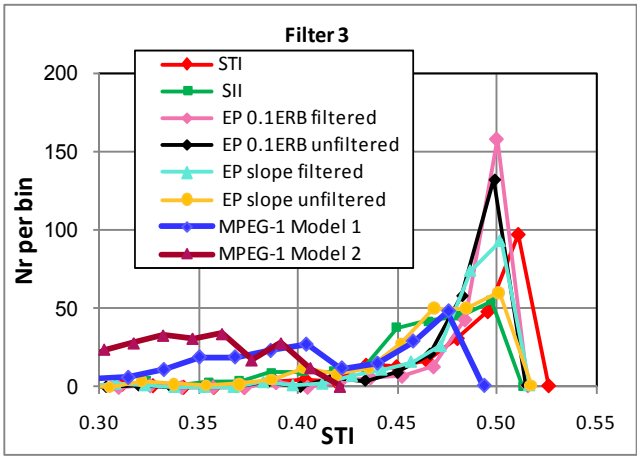
6.3.1. Histograms of STI values for Talker 1

Using the range of anechoic spectra obtained for Talker 1, the STI values for each filter and masking model were examined for their distribution of STI values. Twenty bin-ranges were formed between the maximum and minimum values of STI for each masking model. Figure 12 shows histograms of the STI values in bin sizes equal to (max-min)/20.

The following trends are observed for the MPEG-1 Models 1 and 2:

- Model 2 shows significantly lower STIs than the other masking models.
- The results of Model 2 generally show greater a greater range of STI values than the other masking models.
- Model 1 shows slightly lower STIs than the non MPEG-1 masking models, and its values also show slightly greater distribution than the non MPEG-1 models examined.





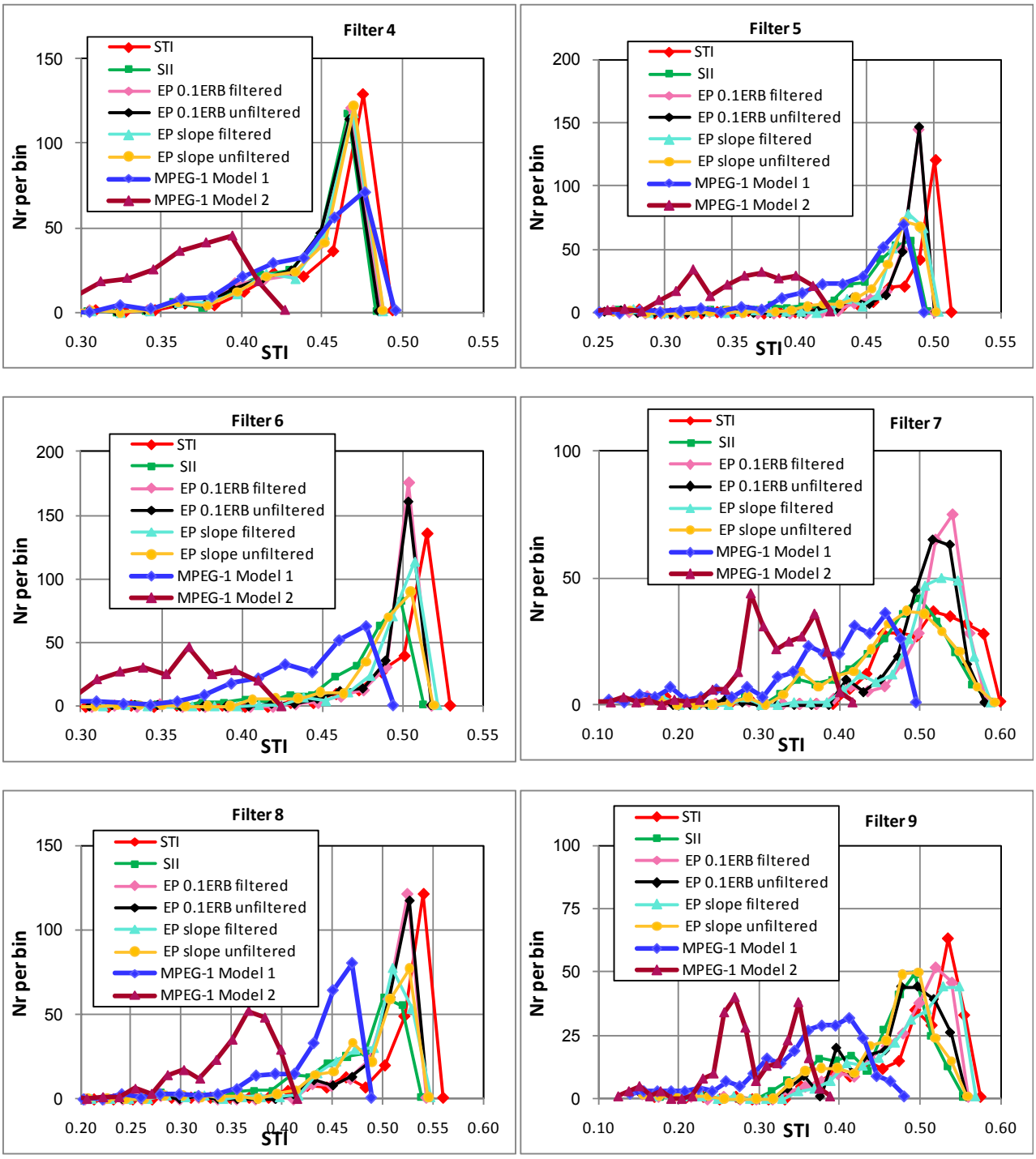


Figure 12. Histogram of STI values (using 20 bins) for six masking models for the nine filters with anechoic Talker 1 and all time slices. Note the differences in scales between graphs.

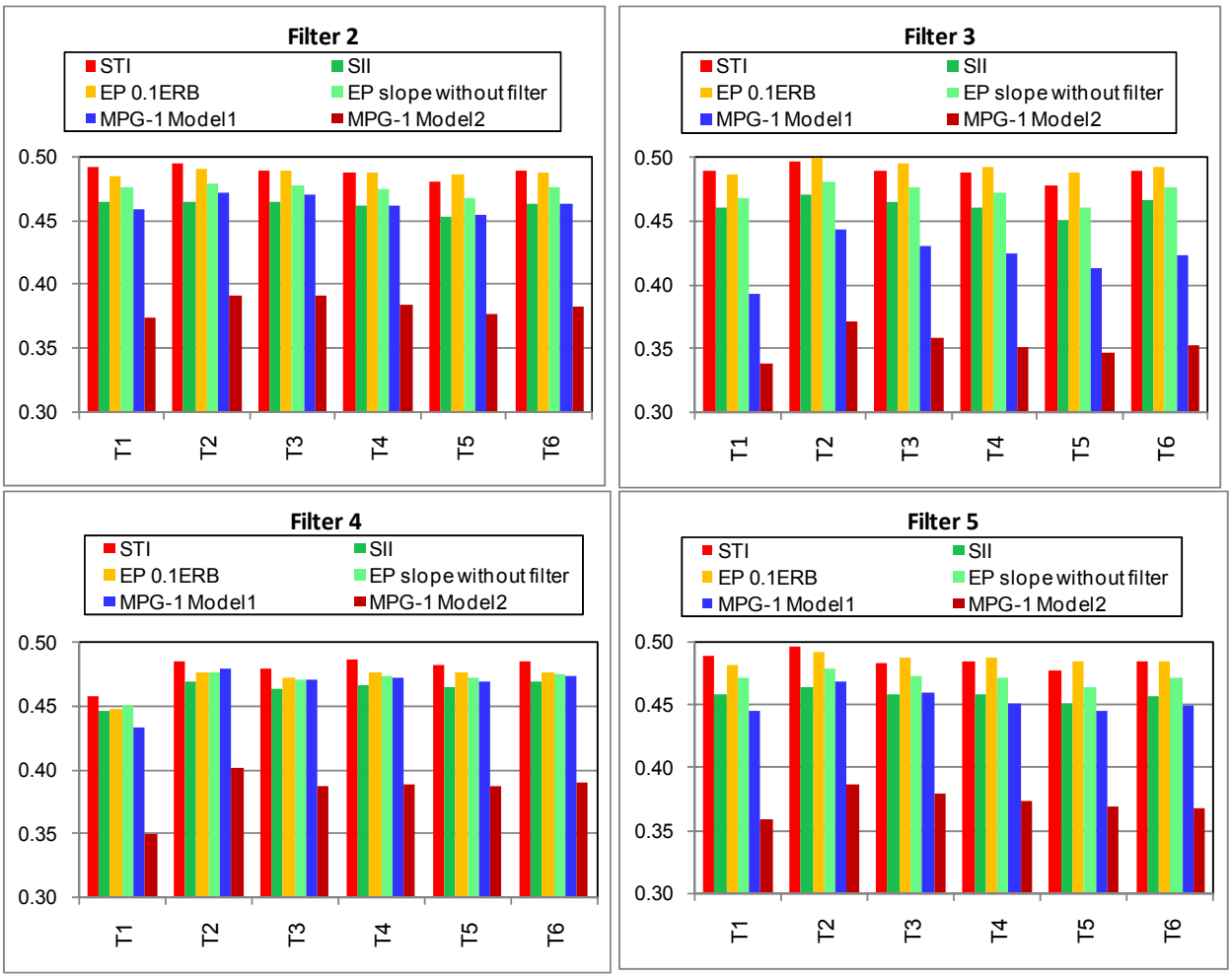
6.3.2. Mean STIs for all Talkers

The mean STI value for each talker and filter was computed for six masking methods using all short-term anechoic spectra. Figure 13 compares the mean STI values for all talkers and filters.

Comparison of the STIs in (with the IEC spectrum) with those in Figure 13 (with short-term spectra) indicates that the mean STIs of the MPEG models with short-term spectra are noticeably lower than with the IEC spectrum.

The following trends are observed for the MPEG-1 Models 1 and 2:

- The STI values with the MPEG-1 masking Model-2 are universally the lowest and are typically 0.15 below those with STI masking. This is a significant difference, given the JND of STI being 0.03The STI values with Model 2 do not show significant variation with Filter number.
- The STI values with the MPEG-1 Model-1 are almost consistently the second lowest and range between 0.01 and 0.14 below the values with STI masking.
- For Filters 2, 4 and 5, the Model 1 STI values are similar to the other non MPEG models, while with the other filters, the STI values are considerably (approximately 0.1) lower than the other non MPEG models.



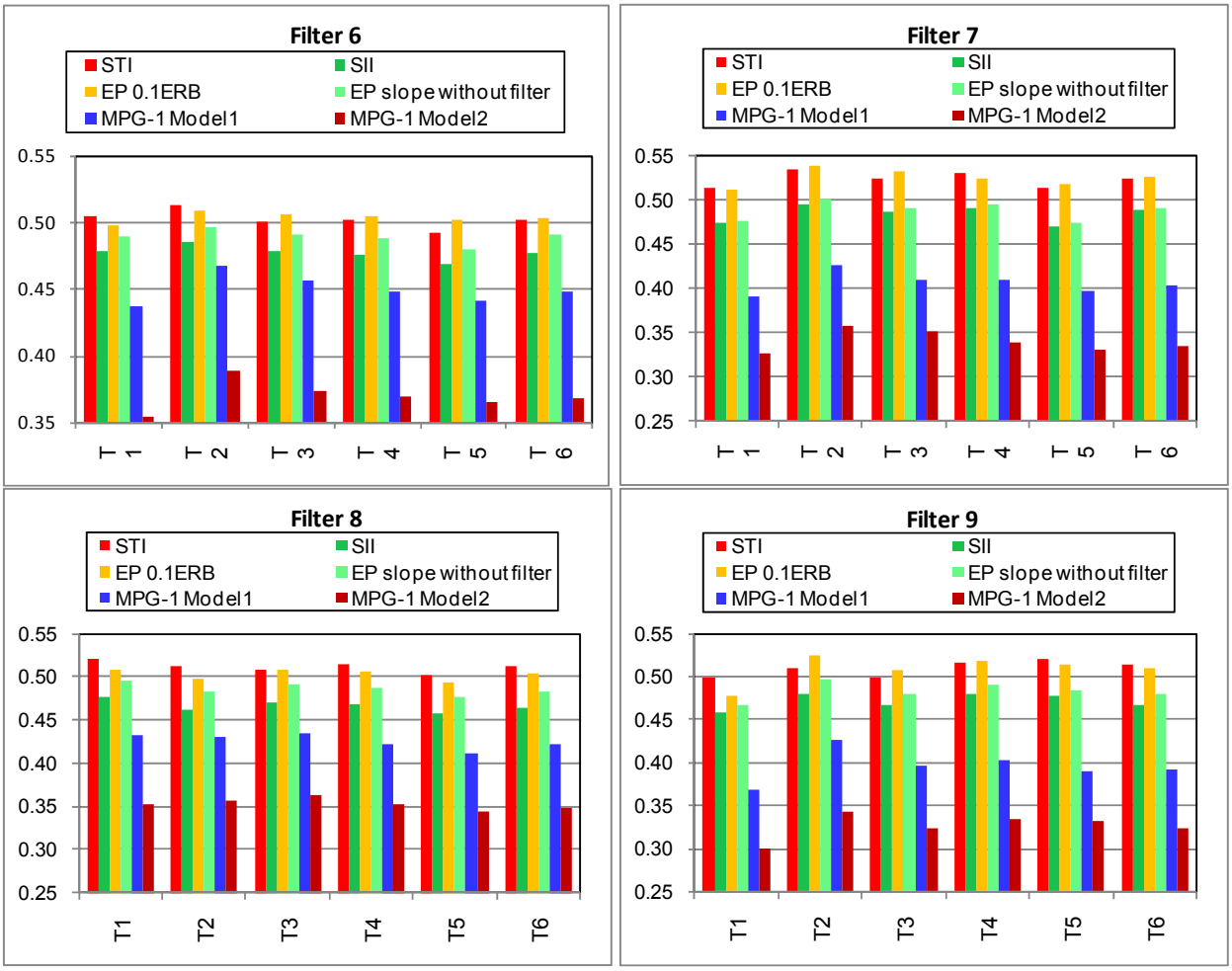


Figure 13. Mean values of STI with the STI, SII and EP slope masking methods.
Data is for anechoic speech and Tn indicates Talker n . Note the different ordinate scales.

7. SUMMARY AND CONCLUSIONS

This paper has investigated the extent of changes to the values of the STI resulting from modifications to two key parameters of that metric. These two aspects are i) the spectrum of speech and ii) the model of the ear's masking mechanism. These two parameters are used to predict the level of equivalent noise resulting from the self-masking of speech.

7.1. Speech Spectra

The standard STI methodology uses a specific long-term spectrum of speech. To investigate changes to the STI values resulting from short-term speech spectra, short-term spectra of six talkers were found using time intervals of 1 s, 250 ms and 50 ms with both anechoic and reverberated speech.

Analysis of these spectra showed variations of up to +12 and -40 dB relative to the IEC spectrum.

7.2. STI Values with Six Masking Methods

The effects of two alternative methods of psychoacoustic masking on STI values were calculated for a large range of speech spectra and compared to the STI values obtained with the specified STI masking method.

As this paper directly extends the 2009 work by the authors, it is pertinent to include the 2009 results in this summary.

Table 2 summarises the eight masking models for which STI were computed for this work and in 2009.

The basis for the calculation was a measured MTF matrix for which the raw STI value was 0.5. The new STI values were computed for six talkers, each with eight filter shapes. The filter shapes had severe frequency response aberrations, and were intended to reflect a sound system with an extremely poor frequency response.

The long-term LA_{eq} level of each talker with the applied filter shape was normalised to 75 dB, and all the resulting short-term spectra computed with this normalisation. A background noise level of NR20 (approximately 33 dBA) was also applied to the STI calculations.

All these masking models are based on masking with stationary signals, and do not consider temporal masking mechanisms. Only Methods 2, 3, 4, 7 and 8 take the ear's downward masking into account.

Method	Type and source	Comment	Ear filtering	Assumes speech "spectral lines" at	Calculation interval
1	STI Specified in IEC standard 60268-16	Uses defined equation to predict masking.	no	octave intervals	octave
2	SII Specified in ANSI S3.5-1997	Uses defined equations to predict masking. Note: the specified attenuation of 24 dB for the speech level was not used.	no	1/3rd octave intervals, which are ultimately integrated into octave bands.	1/3rd octave
3	Difference of two excitation patterns in the inner ear. With and without filtering of the ear	Computes difference between EP with only one band and the EP with all bands other than that band.	yes	1/3rd octave intervals, ultimately integrated into octave bands.	0.1xERB
4			no		
5	Slopes derived from excitation pattern responses.	Uses equations that we developed to predict masking.	yes	1/3rd octave intervals, which are ultimately integrated into octave bands.	1/3rd octave
6			no		
7	MPEG-1 Model 1	Uses defined equation to predict masking.	no	Bark intervals which are ultimately integrated into octave bands	Bark
8	MPEG-1 Model 2		no		

Table 2 Parameters of the eight masking models

7.3. Primary Findings

Our principal findings are:

1. When the mean STI values are calculated with the two MPEG-1 masking models using the range of short-term spectra and filter shapes, the resulting values noticeably differ from the value obtained with the STI masking method and the long-term IEC speech spectrum.
2. The mean STI values with the MPEG-1 masking Model-2 are universally the lowest and are typically 0.15 below and up to 0.2 below those with STI and the other non MPEG masking methods. These are significant differences, given the Just Noticeable Difference of STI being 0.03.
3. The mean STI values with masking Model-1 are almost consistently the second lowest and range between 0.01 and 0.14 below the values with STI masking. and up to 0.07 below those with other non MPEG masking methods.
4. Comparison of the STIs based on the IEC spectrum with those based on short-term spectra indicates that the mean STIs of the MPEG models with short-term spectra are noticeably lower than with the IEC spectrum.
5. The mean STI values with Model 2 do not show significant variation with Filter number.
6. For Filters 2, 4 and 5, the mean STI values for Model 1 are similar to the other non MPEG models, while with the other filters, the STI values are considerably (approximately 0.1) lower than the other non MPEG models.
7. With short-term speech spectra, the spread of STI values with Model 2 is generally much greater than with the other non MPEG masking models.
8. With short-term speech spectra, the spread of STI values with Model 1 is generally slightly greater than with the other non MPEG masking models.

7.4. Conclusions

Our principal conclusions follow from the findings:

1. As masking can occur in bandwidths that are much narrower than an octave, we conclude that the concept of octave bands used in STI may be contributing to the mismatch between measured STIs and subjective intelligibility.
2. Over the range of spectra and masking models, the STI values with MPEG-1 Model 2 are much closer to reflecting the equivalent STI values associated with the subjective word-scores for each filter shape shown in Figure 3.

   This result is somewhat unfortunate, as this masking model does not include any level-dependence, and is therefore the least sophisticated of all the models examined in this and our 2009 paper.
3. A different masking model that also includes the temporal effects of pre and post masking is probably required if STI is to satisfactorily reflect the subjective experience of listeners under conditions of poor spectral balance.

   Processes such as those discussed by Goldsworthy and Greenberg (21) incorporating temporal effects might be useful in narrowing the gap between subjective experience and the objective measure of STI.

8. REFERENCES

1. H. J. M. Steeneken, T. Houtgast. A physical method for measuring speech-transmission quality. *Journal of the Acoustical Society of America.* 1980, Vol. 67, 1, pp. 318-326.

2. IEC. Sound System Equipment Part 16: Objective rating of speech intelligibility by Speech Transmission Index. 2nd Edition 2003. International Standard No. 60268-16.

3. Mapp, P. Some Effects of Equalisation on Sound System Intelligibility and Measurement. *Preprint AES 115th Convention* . 2003.

4. Leembruggen, G.A, Stacy A. Should the Matrix be Reloaded? *Proc IOA.* 2003.

5. American National Standards Institute. Methods for calculation of the Speech Intelligibility Index. New York : s.n., 1997. ANSI S3.5-1997.

6. Leembruggen, G. Is SII better than STI at recognising the effects of poor tonal balance on intelligibility? *Proc IOA.* 2006, Vol. 28, Pt 6.

7. Leembruggen, G., Hippler, M., Mapp, P. Exploring ways to improve sti's recognition of the effects of poor spectral balance on subjective intelligibility. *Proc. IOA.* 2009, Vol. 31, Pt 4.

8. Ludvigsen, C. Relations among some psychoacoustic parameters in normal and cochlearly impaired listeners. *Journal of the Acoustical Society of America.* 1985, Vol. 78, 4, pp. 1271-1280.

9. Palvovic, C.V. Derivation of primary parameters and procedures for use in speech intelligibility predictions. *Journal of the Acoustical Society of America.* 1987, Vol. 82, 2, pp. 413-422.

10. Zwicker, Eberhard. Ueber die Lautheit von ungedrosselten und gedrosselten Schallen. *Acustica.* 1963, Vol. 13, pp. 194-211.

11. Zwicker, E, Fastl, H. *Psychoacoustics Facts and Models.* 3rd Edition. Berlin : Springer, 2007.

12. Glasberg, B.R. and Moore, B.C.J. Derivation of auditory filter shapes from notched-noise data. *Hearing Research.* 1990, Vol. 47, pp. 103-138.

13. Moore, B.C.J. *An Introduction to the Psychology of Hearing.* 5th Edition. Bingley : Emerald Group, 2008.

14. Moore, Brian C. J. and Glasberg, Brian R. Formulae describing frequency selectivity as a function of frequency and level, and their use in calculating excitation patterns. *Hearing Research.* 1987, Vol. 28, pp. 209-225.

15. Moore, B.C.J., Glasberg, B.R and Baer, T. A Model for the Prediction of Thresholds, Loudness, and Partial Loudness. *Journal of the Audio Engineering Society.* 1997, Vol. 45, 4, pp. 224-239.

16. Moore, Brian C.J. and Glasberg, Brian R. Suggested formulae for calculating auditory-filter bandwidths and excitation patterns. *Journal of the Acoustical Society of America.* 1983, Vol. 74, 3, pp. 750-753.

17. Glasberg, B.R., Moore, B.C.J. Prediction of absolute thresholds and equal loudness contours using a modifed loudness model (L). *Journal of the Acoustical Society of America.* 2006, Vol. 120, August 2006.

18. Wijngaarden, S.J., Steeneken, H.J.M. and Houtgast, T. Quantifying the intelligibility of speech in noise for non-native listeners. *J. Acoust. Soc Am.* 2002, Vols. 112 p 3004-3013.

19. Bosi, M., Goldberg, R.,. *Introduction to Digital Audio Coding and Standards.* s.l. : Kluwer Acadamic Publishers, 2003.

20. Steinbrecher, T. Speech Transmission Index: Too weak in time and frequency? *Proc IOA.* 2008, Vol. 30, Part 6.

21. Goldsworthy, R; Greenberg, J. Analysis of speech-based transmission index methods with implications for non-linear operations. *Journal of Acoustical Society of America.* 2004, Vol. 116 pp 3679 to 3689, Dec 2004.