

ANOTHER LOOK AT THE RELATIONSHIPS BETWEEN FREQUENCY RESPONSE AND THE SPEECH TRANSMISSION INDEX, WITH RESPECT TO WORD SCORES AND ROAD TUNNELS

G. Leembruggen

Acoustic Directions and ICE Design, Sydney Australia

1 INTRODUCTION

This paper takes another look at the relationships between frequency response and the speech transmission index (STI). Experience gained by the author and colleagues over many years of designing and commissioning sound systems has shown that a degraded frequency response can greatly reduce speech intelligibility. In reverberant, low noise environments, major degradation in responses can make speech essentially not understandable, and this loss of intelligibility is often barely reflected in the STI performance. Additionally, relatively small changes over an octave bandwidth, sometimes as small as 1 dB, can noticeably improve the perceived intelligibility of conversational speech and this change in perceived intelligibility is virtually not reflected in the STI.

Both of these situations suggest that the psychoacoustic masking mechanism that occurs with relatively low levels is not fully accounted for by the STI process.

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 not only deliver satisfactory speech intelligibility, but should provide “acoustic comfort” for listeners, so that they can actively engage with the speech without straining to understand the content. To achieve this robust intelligibility, the system must accurately reproduce the full range of voice types and speaking styles.

1.1 Part 1: Revisit of 2003 STI Work

The first section of this paper revisits work presented by the author and a colleague¹ in 2003 to the Reproduced Sound conference. That paper explored the apparent inability of the STI metric to properly account for the perceived loss of speech intelligibility that occurs with poor frequency response in public address situations, particularly in reverberant environments.

Measurements of STI were made in a reverberation chamber using a loudspeaker with seven frequency response scenarios. The scenarios were derived from seven sets of equalisation filters that were grossly different. Those STI results were compared with the equivalent STIs derived from word scores obtained from listening tests with recordings made in the same reverberation chamber with the same seven frequency responses. As the ambient noise levels were low in both the speech-recording and listening situations and there were negligible echoes in the reverberation chamber (compared to the reverberation), intelligibility degradation was only due to reverberation and the frequency-response filters. The shape of the filters was selected to represent a highly exaggerated situation, that would rarely be encountered in real-life public address situations.

The paper concluded that in reverberant, noise free situations, STI measurements of situations with highly degraded frequency responses were essentially unable to match the equivalent STI from the word scores.

This paper uses the measured data from 2003 to revisit the relationship between measured and word-score STIs by addressing the following issues:

- a) Critique from colleagues concerning the true ambient noise and listening levels during the word score tests.

A range of ambient noise levels and spectra that could be more representative of the listening environments was therefore used in this analysis.

- b) Application of the speech spectrum during the STI calculation, rather than applying a speech shaped filter to the actual test signal, which was the case in 2003. This provides a higher signal-to-noise ratio for the speech signal.
- c) Application of the continuous auditory masking mechanism specified in STI standard IEC 60268-16 -2011² (Rev. 4).
- d) The new relationship between phonetically-balanced word scores and STI for real-world listening between recently developed by Morales et al³ et al.

1.2 Part 2: STI and Frequency Response in Two Road Tunnels

Part 2 of this paper continues the theme established in Part 1 concerning the comparative independence of frequency response and STI.

With the experience of designing and/or commissioning emergency sound systems for ten road tunnels in Australia and New Zealand, the author is aware that the specifications for emergency sound system performance in these types projects generally focus on STI, which is typically to be assessed in the presence of the smoke-exhaust jet-fans. Little regard seems to be given to the importance of frequency response and perceived intelligibility. A consequence of this focus on STI is that many tunnel systems provide very poor subjective intelligibility although they may just meet specification.

The author is also becoming increasingly concerned at the willingness of many audio and electro-acoustic practitioners to focus on the operational performance of systems, without any critical listening work to understand and optimise the subjective performance of the system.

In the context of these two concerns, the author considers it possible that when the acoustic performance of a tunnel system is being benchmarked after commissioning, if the system complies with or is close to the STI specification, the system will be regarded as “good enough”, without regard for the perceived intelligibility and listening comfort.

It was therefore instructive to explore the effect of gross frequency response changes on the STI performance in the presence of jet-fan noise with two tunnels in New Zealand. These two tunnels were recently designed and commissioned by the author and the filters selected for this investigation would produce a very large degradation in the sound quality during emergencies compared to the sound without these filters.

2 PART 1

2.1 Overview of 2003 Work

2.1.1 Objective Measurements

An investigation was conducted into the relationship between the subjective intelligibility of speech and the measured STI for a range of 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.

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

analyser to yield the anechoic frequency response of the speaker and the STI. This anechoic measurement is Scenario 1.

The loudspeaker was moved into a reverberation chamber at AMS Acoustics, and the STI was again measured at a distance of 1.5 m from the speaker on axis using binaural microphones. Using acoustic absorption material, the reverberation time of the chamber was adjusted so that the measured STI was approximately 0.5; this is Scenario 2. The measured reverberation times in the chamber are shown in Table 1.

| | | | | | | | |
|------------------|-----|-----|-----|------|------|------|------|
| Frequency | 125 | 250 | 500 | 1000 | 2000 | 4000 | 8000 |
| RT (secs) | 3.5 | 3.3 | 2.2 | 1.7 | 1.7 | 1.4 | 1.1 |

Table 1 Measured reverberation times in the reverberation chamber with applied sound absorption material to produce an STI of approximately 0.5.

Seven different filters were then used sequentially to shape the frequency response of the loudspeaker (Scenarios 3 to 9) with the same arrangement of sound absorption material. For each filter, the impulse response was captured and the frequency response and STI of the system computed.

2.1.2 Subjective Measurements

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 voice artist, and were single syllable, phonetically balanced (PB) types from Harvard situated at the end of each sentence.

Following a similar process to the objective measurements in the anechoic and reverberation chambers, three groups of fifty words were played through the loudspeaker and recorded onto digital media using binaural microphones on the dummy head at a distance of 1.5 m from the speaker. Recordings made in the reverberation chamber were made with the same arrangement of sound absorption material as the objective measurements, with a total of seven scenarios being recorded corresponding to the seven response shaping filters. When the groups of words 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 three lists for each of the nine scenarios, to give a total of 81 listening sessions. The sentences were presented to listeners through headphones or loudspeakers, and the listener wrote down the word at the end of the sentence. The ambient noise level in the listening areas was relatively low, and typical of a living room environment.

2.2 Test Scenarios and Frequency Response Filters

The responses of the tonal filters were chosen from our experience to grossly exaggerate the difficulties for perceived intelligibility. Table 2 lists the environment and filters pertinent to each scenario. The frequency responses of the loudspeaker when fed with the filter and the responses of the filter itself are given in the Appendix.

| Scenario | Description | Tonal Filter |
|-----------------|--------------------|--------------------------|
| 1 | anechoic | None |
| 2 | reverberant | None |
| 3 | reverberant | 5 dB/octave cut |
| 4 | reverberant | 5 dB/octave boost |
| 5 | reverberant | 2.5 kHz 12dB notch Q=0.7 |

| Scenario | Description | Tonal Filter |
|----------|-------------|--|
| 6 | reverberant | Plateau @-3dB 400Hz to 1 kHz, plateau @ -10 dB 1.2kHz to 6 kHz |
| 7 | reverberant | 250 Hz 18 dB boost Q = 1.5 |
| 8 | reverberant | 630 Hz 18 dB boost Q = 1.5 |
| 9 | reverberant | Notches: 500 Hz & 2 kHz -18 dB |

Table 2 Details of temporal and filter parameters for each scenario.

2.3 Computation of STIs with Refined Parameters

To compute the STIs for the conditions under which the word scores were obtained, a two-part process was used in accordance with the method laid out in Annex M of [2].

- In Part 1 of the process, the MTF data obtained using the MLSSA analyser was corrected to remove the effect on the STIs of ambient noise and masking computed by MLSSA. This process yielded MTF matrices that included degradation only by temporal effects, and hearing threshold.
- In Part 2, adjustments to the corrected MTF matrix were made for new ambient noise and listening levels and the masking algorithm specified in Rev 4 of the STI standard.

2.3.1 Listening Conditions

Five possible ambient noise levels for the listening situations were used for the STI calculations. An air-conditioned environment is assumed with noise levels ranging from a typical low-noise situation to an unlikely high-noise situation. The octave band L_{eq} levels corresponding to the five noise situations are shown in Table 3 and Figure 1.

| Noise Situation | Frequency Hz | | | | | | | |
|-----------------|--------------|-----|-----|-----|------|------|------|------|
| | dBA | 125 | 250 | 500 | 1000 | 2000 | 4000 | 8000 |
| AC Low | 32.3 | 36 | 34 | 29 | 27 | 22 | 18 | 15 |
| NR 25 | 33.3 | 44 | 35 | 29 | 25 | 22 | 20 | 18 |
| AC High | 36.3 | 46 | 38 | 32 | 30 | 26 | 20 | 20 |
| NR 30 | 38.0 | 48 | 40 | 34 | 30 | 27 | 25 | 23 |
| NR 35 | 42.8 | 52 | 45 | 39 | 35 | 32 | 30 | 28 |

Table 3 Possible noise levels for the listening situation applied to the STI calculations.

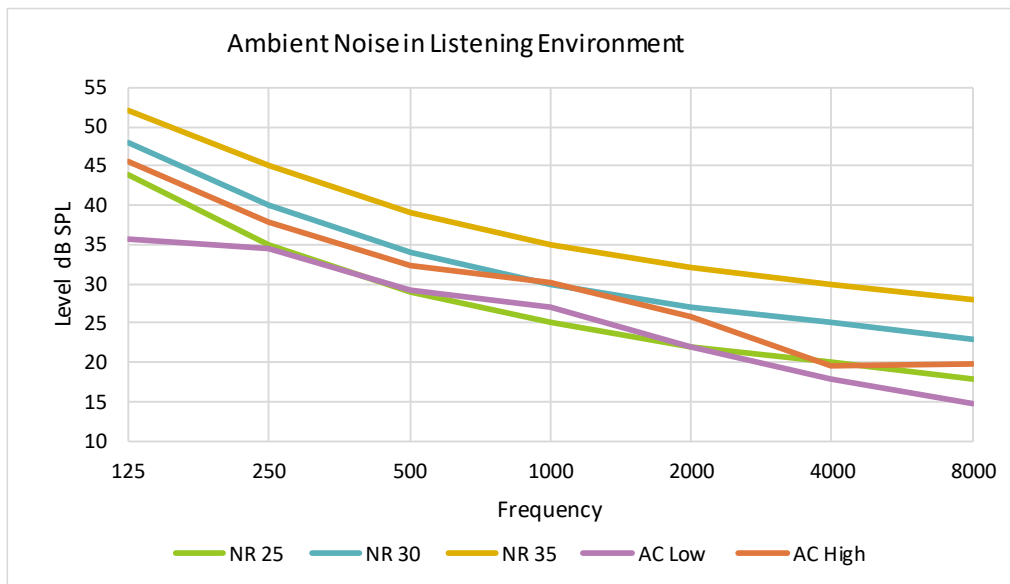


Figure 1 Spectra of the range of ambient noise levels applied to the listening environment.

The speech level for the STI calculations was set to 66 dBA, which is a comfortable level for speech in a small environment that would be regarded as quiet and suitable for critical listening. The male speech spectrum presented by Cushing et al⁴ was used for the calculations as it has greater similarity to spectra measured by the author, Byrne et al⁵, and the ANSI standard S3.5⁶, compared to the spectrum specified by the STI standard.

The overall level of each filtered speech signal was adjusted (normalised) to 66 dBA. Figure 2 and Figure 3 show the octave band levels after normalisation with two alternative views to assist visualisation of the differences in the filtered spectrum.

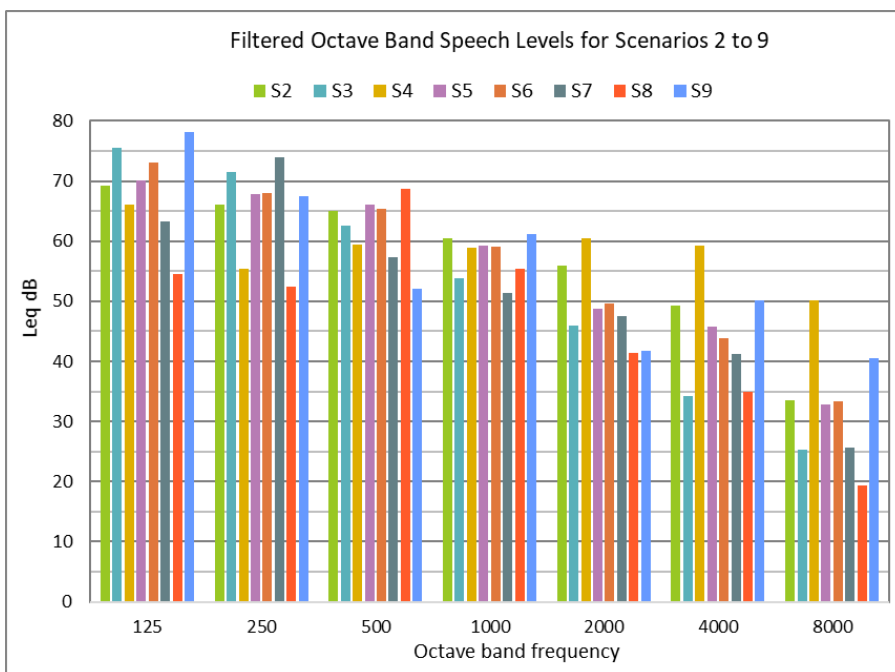


Figure 2 Octave band levels of speech when filtered with Scenarios 2 to 9 (frequency is abscissa).

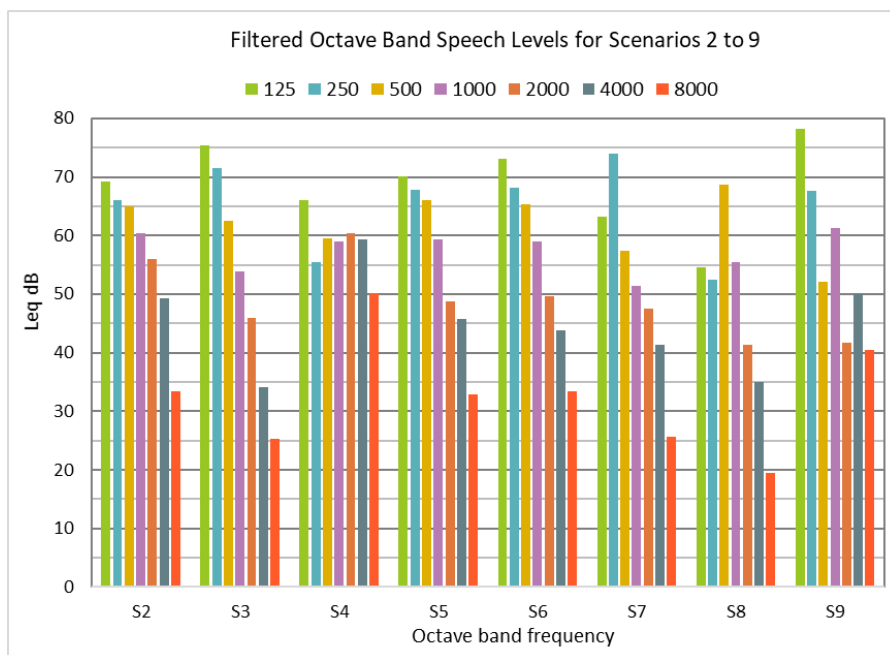


Figure 3 Octave band levels of speech when filtered with Scenarios 2 to 9 (Scenario is abscissa).

2.3.2 STI Results

The MTF matrix for the left ear was used for the STI calculations, as this ear gave the highest measured MTF values in the reverberation chamber. Table 4 shows the computed STIs for each filter scenario.

| Filter scenario | AC Low 32.3 dBA | NR 25 33.3 dBA | AC High 36.3 dBA | NR 30 38.0 dBA | NR 35 42.8 dBA |
|-----------------|--------------------|-------------------|---------------------|-------------------|-------------------|
| 2 | 0.52 | 0.52 | 0.52 | 0.52 | 0.51 |
| 3 | 0.51 | 0.51 | 0.50 | 0.49 | 0.46 |
| 4 | 0.51 | 0.51 | 0.51 | 0.51 | 0.50 |
| 5 | 0.52 | 0.52 | 0.52 | 0.51 | 0.50 |
| 6 | 0.53 | 0.53 | 0.52 | 0.52 | 0.51 |
| 7 | 0.50 | 0.49 | 0.49 | 0.48 | 0.46 |
| 8 | 0.49 | 0.48 | 0.47 | 0.46 | 0.42 |
| 9 | 0.52 | 0.52 | 0.52 | 0.52 | 0.50 |

Table 4 STI result for each filter scenario and ambient noise level for a listening level of 66 dBA.

2.4 Estimated STI from Word Scores

Figure 4 shows the word scores for both the UK and Australian listeners.

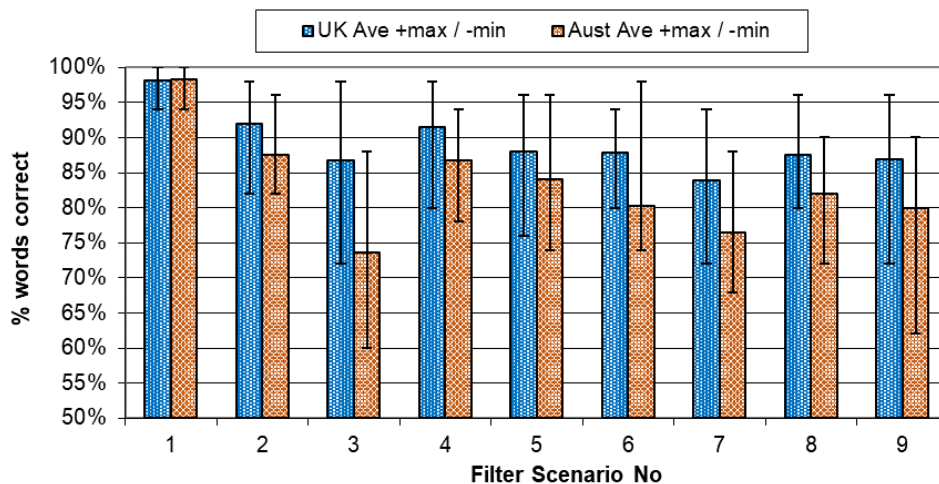


Figure 4 Plot of word score vs filter scenario. The error bars indicate the maxima and minima.

The following points are pertinent to the results.

- 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. Judging by the difficulty the author has with the intelligibility with some BBC television productions, it is concluded that this difference was likely to result from accent differences.
- c) There was a noticeable reduction in the word score when the tonal filters were inserted.
- d) Even though the test words were well-articulated, each Australian listener 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 and the words had not been so well articulated in the original recording, the scores would have been lower.
- e) Overall, the Australian listeners found the process to be quite tiring.

2.5 Associated STI Results

In 2003, the equivalent STIs were computed from the word scores using the relationship between phonetically-balanced words and STI published in 1985 by Anderson and Kalb⁷ (AK). That relationship is presented in Figure E1 of Annex E in both the 2003 (Rev 3) and 2011 versions of standard 60268-16. Using a different experimental method to AK, Morales et al revised the relationships between the intelligibility of Harvard PB words and STI for real-world listening in both reverberant and noisy environments.

The relationships between the STIs and real-life listening presented by Morales et al were published between 2012⁸ and 2018⁹. Refs [3] and [5] present relationships for reverberant situations only, whereas Ref [6] presents relationships for degradation due to combinations of reverberation, echoes, additive noise and bandpass limiting. As this assessment relates to reverberation only, Morales' relationship between the monaural measured STI and the average scores of the real-life PB word intelligibility tests shown in Figure 4 of [3] is used. Figure 5 below shows this relationship along with averaged PB score results for reverberation distortions from AK shown in Figure E1 of 60268-16 2011.

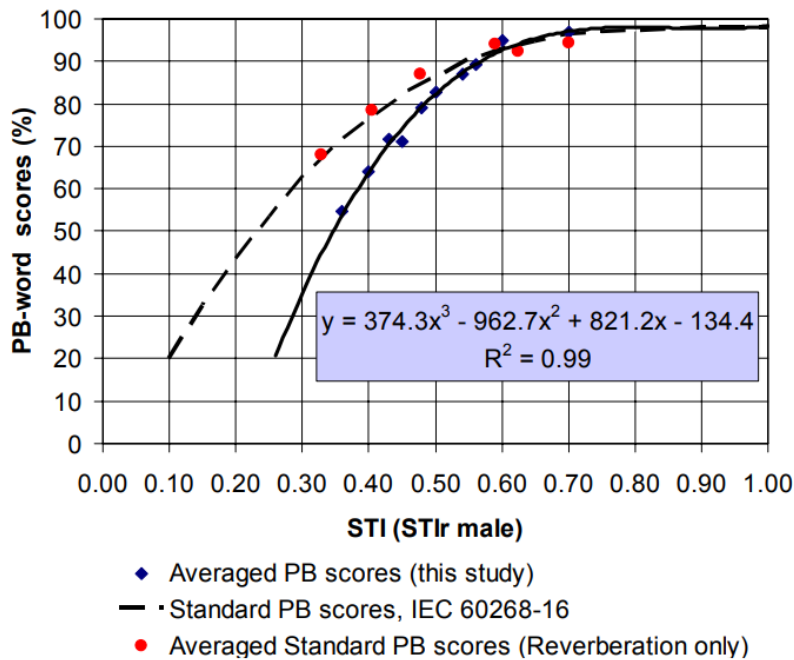


Figure 5 Relationship between the monaural measured STI and the average PB scores.

Table 5 presents the STI results for each scenario in the estimated listening situation.

| Scenario | Equivalent STI from Word Score | | Description | Tonal Filter |
|----------|--------------------------------|------|-------------|---|
| | UK | Aust | | |
| 1 | 0.82 | 0.83 | anechoic | None |
| 2 | 0.60 | 0.55 | reverberant | None |
| 3 | 0.54 | 0.45 | reverberant | 5 dB/octave cut |
| 4 | 0.59 | 0.54 | reverberant | 5 dB/octave boost |
| 5 | 0.55 | 0.52 | reverberant | 2.5 kHz 12dB notch Q=0.7 |
| 6 | 0.55 | 0.49 | reverberant | Plateau -3dB 400 Hz to 1 kHz Plateau -10 dB 1.2 kHz to 6 kHz |
| 7 | 0.52 | 0.47 | reverberant | 250 Hz 18 dB boost Q = 1.5 |
| 8 | 0.55 | 0.50 | reverberant | 630 Hz 18 dB boost Q = 1.5 |
| 9 | 0.54 | 0.49 | reverberant | Notches 500 Hz & 2 kHz -18 dB |

Table 5 Computed STI results for each scenario in the estimated listening situation.

Figure 6 compares the equivalent STIs for the UK and Australian word scores and the measured STIs for each filter scenario and noise situation. The maximum and minimum of the word-score STIs are also indicated. Although the STI for NR35 is included for completeness, the author considers it is unlikely that this noise spectrum would have been present when the word scores were obtained.

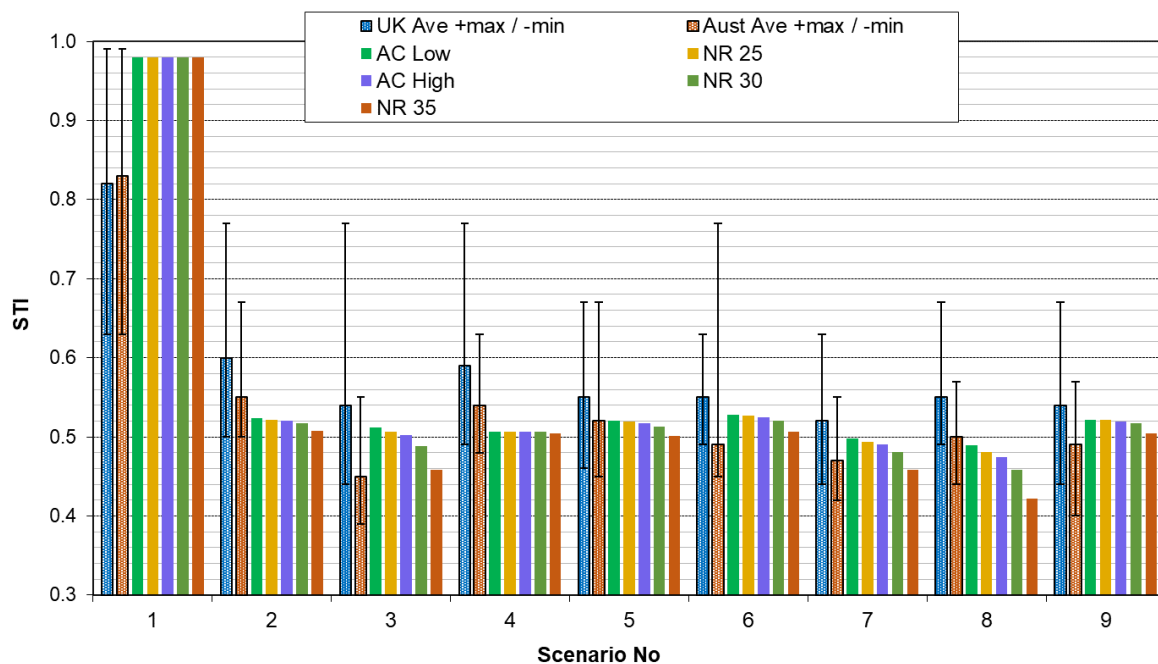


Figure 6 Comparison of equivalent STIs from word scores and measured STI vs filter scenario for the five noise situations. The error bars indicate the maxima and minima.

2.6 Discussion

The following points are noted, and for the reason noted above, the NR35 ambient noise situation is not included as it represents an unlikely listening scenario.

- If the words had not been spoken by a voice artist with good articulation, the word scores would have likely been lower.
- The STIs for Scenario 2 are relatively close (within 0.02) to the minimum equivalent STI for the Australian listeners.
- With filter Scenarios 3, 6, 7 and 9, the measured STIs are lower than the Australian average, whereas for Scenarios 2, 4, and 8, the measured STIs are higher than the Australian average. The measured STIs in Scenario 5 are close to the Australian average. However, Scenario 8 is an unusual case as it is a narrow peaking boost of 20 dB at 630 Hz.
- Other than the NR30 situation and Scenario 8, none of the STIs for the filter scenarios and noise situations is able to satisfactorily mirror the minimum STI scores of the Australian listeners, particularly at the lower ambient noise levels. This result highlights the need to recognise that more words could be lost than suggested by the average STI relationship developed by Morales.
- Figure 7 shows an expanded view of Morales' relationship between word score and equivalent STI in reverberation-only situations, with the axes reversed. When viewed in this form, it is more readily recognised that a loss in PB words of 5% to 40% covers only a range of 0.25 in STI.

In terms of the general awareness of STI users, the range of 0.64 to 0.39 is within a familiar range. What may be more unfamiliar is that this range encompasses such a large extent of lost words. Given that complex instructions or unfamiliar words could be regarded as having some similarity to PB words in their degree of difficulty for recognition, this potential loss in word recognition is much greater than the STI value suggests.

- f) Morales' relationship assigns a given loss of intelligibility to a higher STI than the AK relationship. For example, Morales equates a 40% loss of PB words to an STI of 0.39 compared to AK's STI of 0.29. Morales' relationship therefore compresses the range of word losses into a smaller STI range with higher absolute values than before. As such, an organisation wishing to approve a given STI performance based on a limited understanding of STI would approve the result without an understanding of the potential loss of intelligibility, particularly with respect to unfamiliar words.
- g) A primary difficulty is that no STI verification has been done for speech in which the tonal structure within the 250 Hz to 4 kHz has been highly distorted. As such, Morales' relationship may not properly represent the loss of intelligibility that occurs with self-speech masking when the frequency responses are poor.
- h) The measured STI values are based on a long-term average speech spectrum, whereas the spectra of individual words can show significant differences with that average spectrum. This is illustrated in Figure 8, which shows the difference between the standardised IEC speech spectrum and the spectra of five PB words used in the 2003 work. Differences of up to 20 dB are evident. The possibility of self-speech masking becomes much more likely during words with soft consonants at the end such as "salve", "sheik", "quip" and "dung" with this type of spectra. The self-masking that can occur with these types of words is essentially not recognised by STI's masking algorithm.

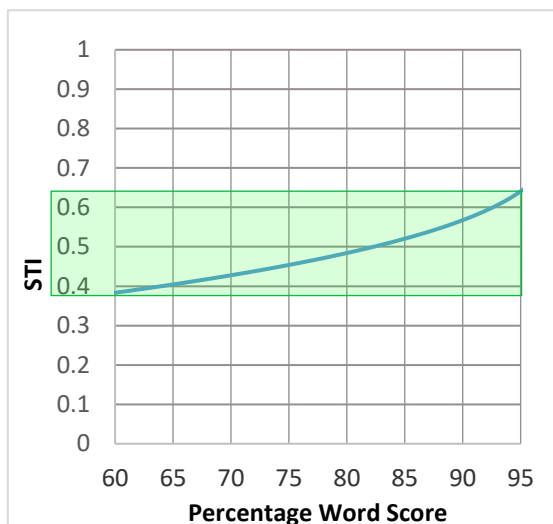


Figure 7. Expanded view of reversed relationship between STI and word score.

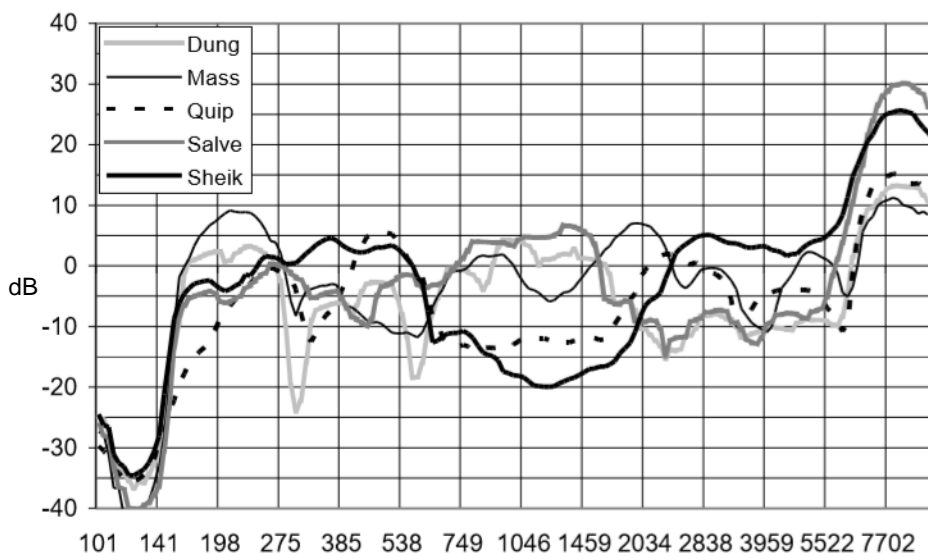


Figure 8 Differences in the spectra of five PB words and the standard IEC speech spectrum.

- i) As a "broad brush" approach, the measured STI results appear to reasonably represent the word scores. However, when the receipt by listeners of critical information is at stake and the talkers are not trained actors, the broad brush of STI is not able to satisfactorily quantify the actual loss of intelligibility.

3 PART 2

3.1 Site Work

The electro-acoustic optimisation process of each road tunnel sound system consisted of equalisation, setting signal delays and optimising the gain structure of the system.

- a) A suitable equalisation was developed using a combination of acoustic measurements and listening to optimise subjective speech intelligibility, naturalness of tone, and listening comfort at high levels.

When listening to speech at levels exceeding 90 dBA, strong high-frequency content such as occurs in sibilant words, can rapidly become very uncomfortable. As good high-frequency response is also essential for speech intelligibility, the approach we used was a combination of a slight reduction in the equalised response above 5 kHz, and careful de-essing of specific words, which was generally those words ending in the letter 's'.

- b) A goal of the work to develop the gain structure of the system so that the long-term L_{Aeq} level was as high as possible in order to maximise the STI, whilst ensuring that the system did not produce strongly audible distortion. The gain structure was developed using a combination of broadband level measurements, capturing the amplifier output using a digital oscilloscope and critical listening.

To achieve this outcome, a crest factor of 21 dB needed to be provided. At face value this crest factor seems excessive but is required to accommodate the necessary high-frequency equalisation boost of 22 dB at 11 kHz. Although the long-term RMS level of speech was low, short-term peaks in this frequency range can be up to 21 dB above the long-term overall RMS level.

- c) As amplifiers in road tunnel systems can be of variable quality and often do not overload (clip) cleanly, distortion can be clearly audible even with modest amounts of clipping. In this context, care is required to ensure distortion is not particularly audible.

After the sound system in each tunnel was optimised for perceived intelligibility and tonal comfort, the performance of the system was benchmarked using the direct STI method STIPA. In Tunnel 1 (VPT), measurements were made at 58 locations along its 410 m length, while in Tunnel 2 (JHT), measurements were made at 47 locations along its 320 m length.

The level of the STIPA signal was carefully set so that its spatially-averaged level was equivalent to the spatially-averaged level of speech. To achieve this, the STIPA level was set 2 dB higher than the speech level to account for the reduction in long-term speech level that gaps between words produce. That reduction was computed from the speech audio file as the difference in RMS levels between the announcement with gaps and without gaps.

Measurements were also made in octave bands of the noise produced by the smoke-exhaust jet-fans in the tunnels. Figure 9 shows the total L_{Aeq} sound pressure level of the fan noise expressed as $L_{Aeq-1second}$ along VPT (northerly direction) and JHT (southerly direction). Each one-second interval produced a spatial average of the sound levels over a traversed distance of approximately 1 m.

The STIPA measurements at each position provided i) the level in octave bands of the test signal shaped to the standard IEC speech spectrum and ii) the MTF matrix. Using a bulk calculation process, the jet-fan noise was introduced into each STIPA measurement. That process also incorporates the masking and threshold calculations and follows the method of Annex M in [2]. Table 6 shows the averages and standard deviations of the STI performance with and without jet-fan noise and the equivalent L_{Aeq} level of speech in the tunnel.

Neither of the two systems can be regarded as optimally designed, as the project brief was to make improvements to the existing systems by changing only the loudspeakers and cabling. No additional amplifiers were to be installed. This constraint resulted in distances between loudspeakers that would normally be regarded as large in the case of VPT and very large in JHT.

| VPT | | | JHT | | | | |
|----------------|----------------|-------------------|------------------|----------------|----------------|-------------------|------------------|
| | STI with noise | STI without noise | Speech L_{Aeq} | | STI with noise | STI without noise | Speech L_{Aeq} |
| Average | 0.48 | 0.51 | 93.5 | Average | 0.38 | 0.45 | 92.0 |
| Std Dev | 0.07 | 0.05 | 1.19 | Std Dev | 0.05 | 0.05 | 1.5 |

Table 6 STI results in the two tunnels measured during the benchmarking process.



Figure 9 Broadband noise levels along the two tunnels.

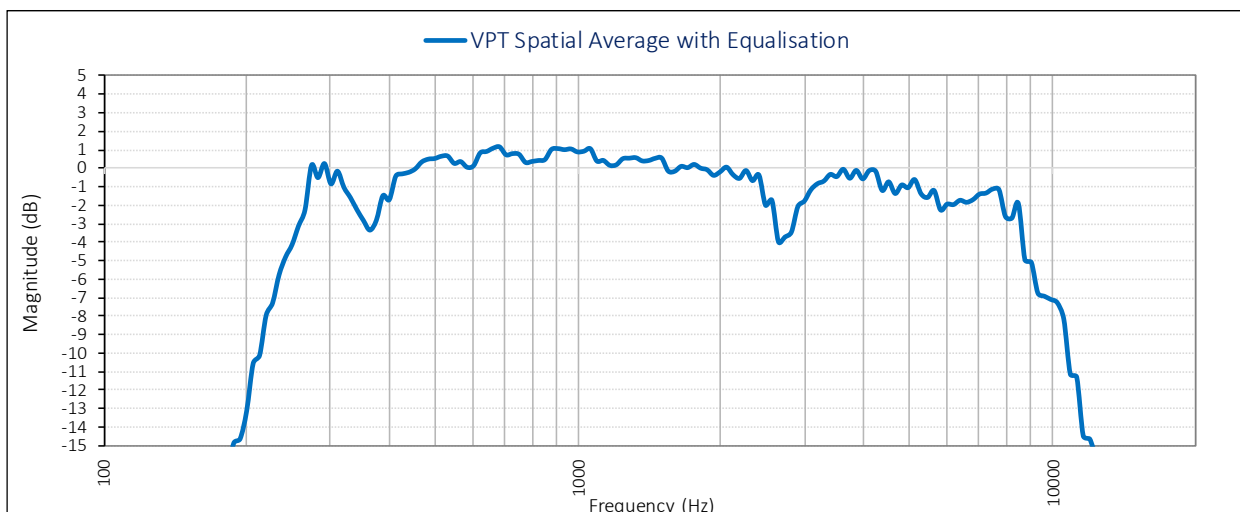


Figure 10 Example of spatially-averaged frequency response after equalisation and optimisation of the perceived intelligibility and listening comfort.

3.2 Effect of Changes to Frequency Response on the STI

3.2.1 Frequency Responses Used

To assess the change in STI performance resulting from a change in the frequency response of the system for this paper, six different filters were applied in bulk to the STI calculations for each tunnel. These filters were selected to demonstrate the insensitivity of the STI results to changes in frequency response that would sound extremely poor and would result in highly degraded

intelligibility. The filters took the form of boosts or cuts to a number of octave bands. Table 7 shows the levels in each octave band for the applied filter.

As the boost or cuts affect the overall level of the speech signal in the calculations, an overall level adjustment was applied so that the L_{Aeq} of the speech signal remained constant. To illustrate the octave band levels in Table 7 that would result with filtered pink noise, Figure 11 provides a graphical examples of frequency response that produce the octave band levels in each filter response. Subjective descriptions are also given.

| Filter No | Octave Band Frequency Hz | | | | | | | Overall level adj. |
|-----------|--------------------------|-----|-----|------|------|------|------|--------------------|
| | 125 | 250 | 500 | 1000 | 2000 | 4000 | 8000 | |
| 1* | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| 2 | 0 | 0 | 0 | -3 | -3 | -3 | -5 | 1.2 |
| 3 | 0 | 4 | 2 | -4 | -4 | 0 | 0 | -0.6 |
| 4 | 0 | 0 | 0 | 2 | -5 | -2 | 0 | -0.2 |
| 5 | 0 | 0 | -4 | 2 | -4 | -4 | 0 | 1.0 |
| 6 | 0 | 0 | 0 | -9 | -9 | 0 | 0 | 1.9 |
| 7 | 0 | 0 | 0 | -6 | 0 | 6 | 0 | 0.2 |

Table 7 Octave band levels applied to shape the frequency responses in the tunnels. All values are in dB. *Existing response without filter.

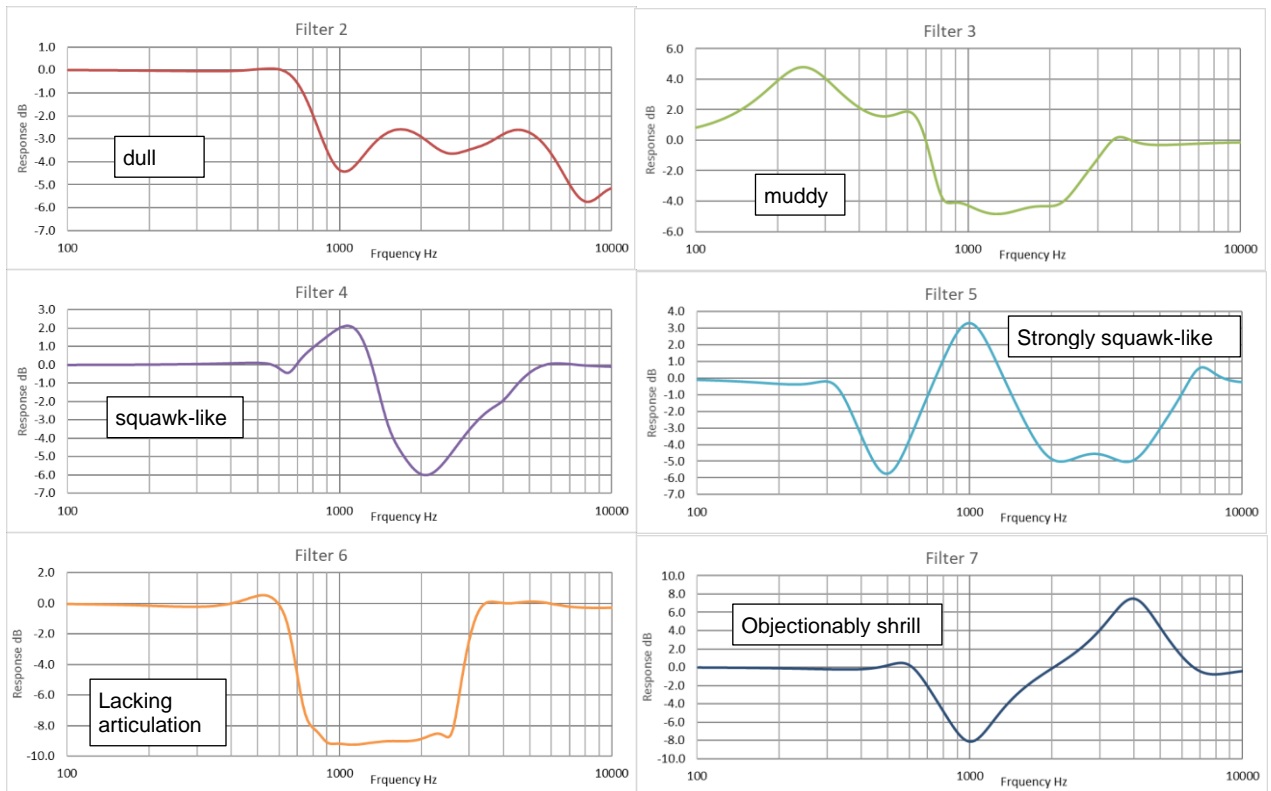


Figure 11 Example frequency responses that produce the octave band levels in Table 7 with pink noise. Subjective descriptions are also given.

3.2.2 STI results

| Filter No | VPT | | | JHT | | |
|-----------|-------------|---------|---------------|-------------|---------|---------------|
| | Average STI | Std Dev | Change in STI | Average STI | Std Dev | Change in STI |
| 1* | 0.48 | 0.07 | | 0.38 | 0.05 | |
| 2 | 0.46 | 0.07 | 0.02 | 0.36 | 0.05 | 0.02 |
| 3 | 0.45 | 0.07 | 0.03 | 0.35 | 0.05 | 0.03 |
| 4 | 0.44 | 0.07 | 0.04 | 0.36 | 0.05 | 0.02 |
| 5 | 0.45 | 0.07 | 0.03 | 0.36 | 0.05 | 0.02 |
| 6 | 0.43 | 0.07 | 0.05 | 0.34 | 0.05 | 0.04 |
| 7 | 0.47 | 0.06 | 0.01 | 0.37 | 0.04 | 0.01 |

Table 8 Average STIs and standard deviations for the two tunnels with the actual response (No 1*) and six shaping filters.

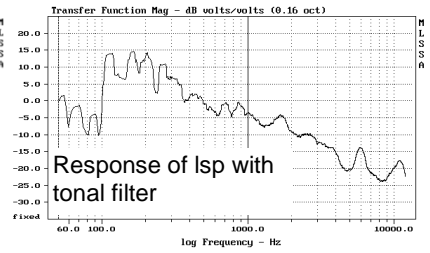
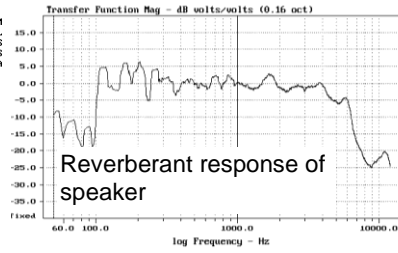
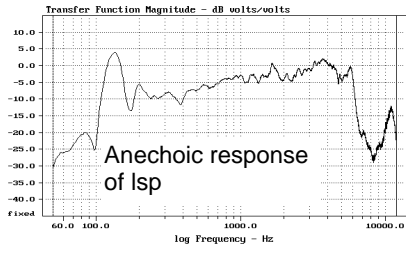
Table 8 shows the STI results with the filters inserted into the calculations. Inspection of the results reveals that significant reductions in the sound energy within octave bands that are deemed more critical for intelligibility (1 kHz to 8 kHz) have only minor effects on the STI, even during situations with high noise levels.

The author concludes that when specifying or commissioning an emergency sound system in a road tunnel, blind adherence to simply achieving a STI performance without paying significant attention to the frequency response of the system over the entire listening area, can yield a system which produces poor perceived intelligibility and yet its performance meets the project specification.

4 REFERENCES

- 1 Leembruggen, G.A, Stacy A. Should the Matrix be Reloaded? Proc IOA. 2003
- 2 IEC. Sound System Equipment Part 16: Objective rating of speech intelligibility by Speech Transmission Index. 3rd Edition, 2011. International Standard No. 60268-16
- 3 Morales, L; Dance, S; Shield, B; Leembruggen, G. Speech Transmission Index for the English Language Verified Under Reverberant Conditions with Two Binaural Listening Methods: Real-Life and Headphones. J. Aud. Eng. Soc. Vol 62 Issue 7/8 pp. 493-504; July 2014
- 4 Cushing, Li, Cox, Worrall, Jackson. Vocal effort levels in anechoic conditions. Applied Acoustics. 2011, Vol. 72, pp. 695-701
- 5 Byrne, et al. An international comparison of long-term average speech spectra. JASA 1994, Vol. 96
- 6 ANSI. Methods for Calculation of the Speech Intelligibility Index. American National Standard. S3.5-1997
- 7 Anderson, B.W., and Kalb, J.T. "English verification of the STI method for estimating speech intelligibility of a communications channel," J. Acoust. Soc. Am. Vol 81, 1987, pp 1982-1985f
- 8 Morales, L; Dance, B; Leembruggen, G, Preliminary validation of the revised STI male for the English Language. Proc IOA Vol 34 Pt 4 2012
- 9 Morales, L; Li, F A new verification of the speech transmission index for the English language. Speech Communication 105 (2018) 1–11

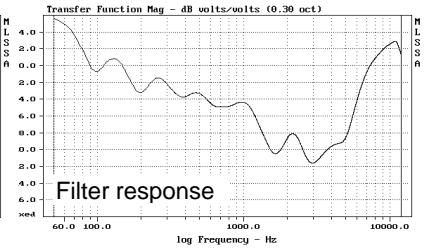
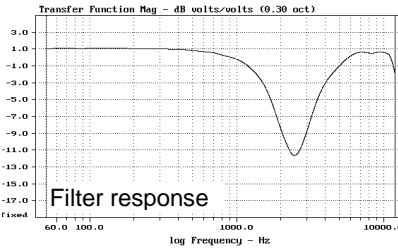
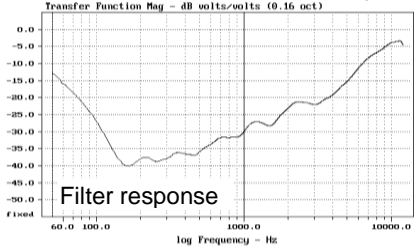
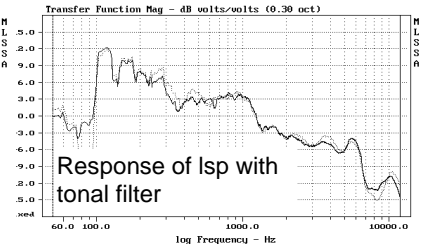
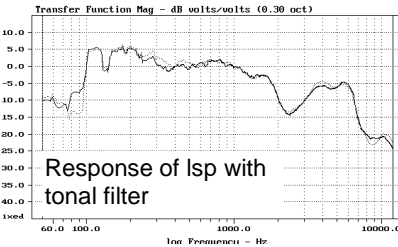
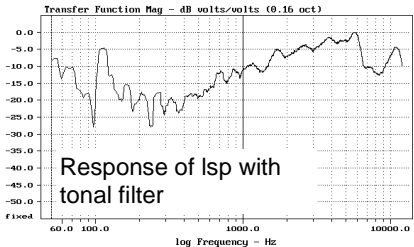
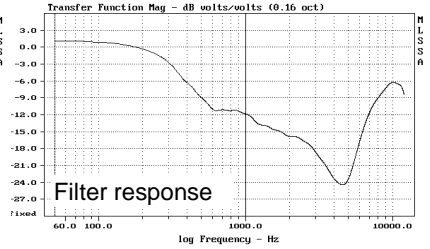
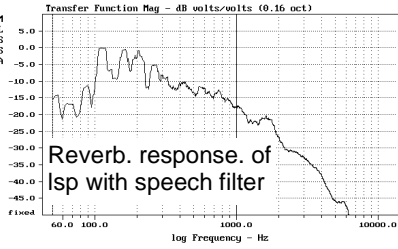
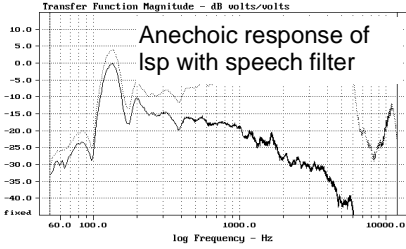
APPENDIX FREQUENCY RESPONSES OF FILTER SCENARIOS



Scenario 1

Scenario 2

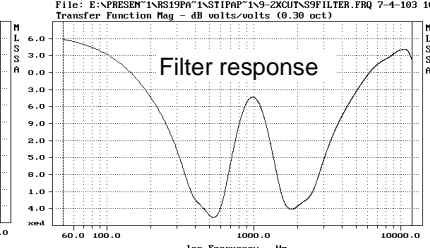
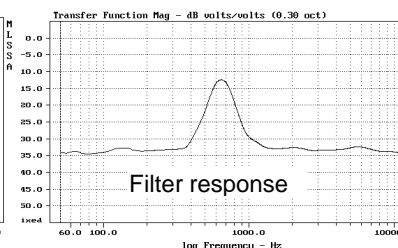
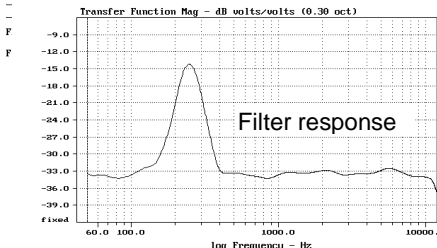
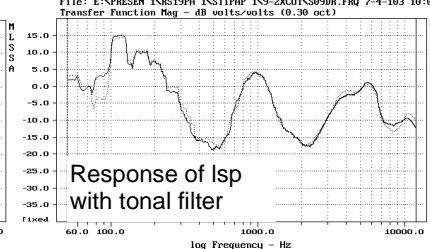
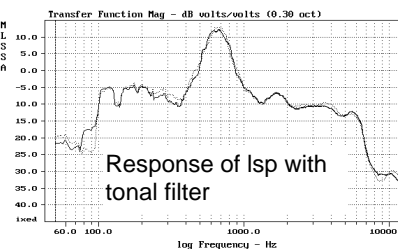
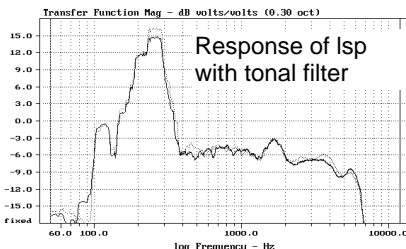
Scenario 3



Scenario 4

Scenario 5

Scenario 6



Scenario 7

Scenario 8

Scenario 9