CAN YOU RELY ON THE STI?
A TEMPORAL ANALYSIS OF SEVEN SOUND SYSTEMS

Glenn Leembruggen ICE Design, Acoustic Directions, University of Sydney MIOA
David Gilfillan ICE Design, Gilfillan Soundwork MIOA
Lorenzo Morales London Southbank University MIOA

1 INTRODUCTION

The authors have long been aware that in many sound system situations, the measured Speech Transmission Index performance of the system does not fully indicate the degree of subjective speech intelligibility perceived by listeners to sound systems, or the degree to which speech can be easily or comfortably understood.

A number of authors[1,2,3,4] have investigated some of the areas in which STI appears to be mismatched to the real world of listening to sound systems. In particular, Leembruggen et al have been concerned about the inability of the STI to properly account for the loss of subjective intelligibility due to the combination of the ear's upward masking mechanism, poor frequency response and short-term speech spectra in reverberant situations.

This paper continues the investigation into the STI's limitations and explores differences in temporal behaviour of seven sound systems.

1.1 Background to the Paper

The authors have designed numerous speech sound systems[5,6,7,8] for important situations in which the acoustic environment presented strong challenges for speech intelligibility. In these situations, the authors have found that the combination of an extremely flat-frequency response for the direct field for all listeners and the highest possible directionality at all frequencies has been able to produce excellent outcomes, regardless of the measured (monaural) STI.

In earlier times, the Direct to Reverberant ratios in each octave band provided insight into the speech intelligibility produced by a sound system. Over the last fifteen years, clarity ratios C35 and C50 have often been used to assess the intelligibility. However, clarity ratios do not properly indicate differences in the fine-structure of the temporal responses between systems, and can also effectively de-rate the level of the direct field.

In presentations to the IOA, Griesinger[9] noted the importance of the direct field to the engagement of listeners with sound in concert halls. This engagement improves the ability of listeners to localise sound sources and discern the information content. Part of his hypothesis is that early reflections detract from engagement. Griesinger has developed a measure called LOC which relates the level of the direct field to the arrivals in the impulse response up to 100 ms, which is the integration time of the ear for loudness perception.

The combination of Griesinger's conclusions and our sound system experience has inspired the authors to explore the short-term temporal behaviour of speech. The fine structure of the early portion of the impulse response becomes important when transient-like signals such as short words and syllables are broadcast into reverberant spaces. Such transient signals do not allow the reverberant field to be fully excited (or "charged up"). This is in direct contrast to the use of a steady-state signal by the STI. STI is time-blind, and the arrival order of temporal information is ignored.

This paper work explores the early-time temporal performance of seven different loudspeaker systems with measured STI values ranging from 0.44 to 0.7 and different subjective sound qualities. It is hope that that this work will provide further insight into the mismatch of subjective speech intelligibility and their measured STIs.

2 INVESTIGATION PROCESS

Three aspects of the temporal behaviour of the sound systems were explored using the impulse response (IR) of each system:

a) Build-up (or integration) ratios

b) Griesinger's LOC parameter

c) Spectra of the speech envelope in five octave bands with three different phrases

2.1 Sound systems analysed

Table 1 lists pertinent aspects of the sound systems for which the IRs used in this study were measured. A rating is given for the subjective speech intelligibility, based on the robustness of word recognition by the authors. Word score tests were beyond the scope of this study.

Table 2 notes the authors' subjective listening impressions for each system.

No	**Location**	**Type**	**ave. RT_{60} 250 to 4k**	**Meas. STI**	**Subjective Intelligibility 0 to 10 scale**
1	Brisbane Banco Court	Distributed steered arrays covering public galleries.	1.8	0.53	3.5
2	Brisbane Airport Link Road Tunnel	Time sequenced horns at 40 m intervals.	3.8	0.52	4
3	City Of Sydney Council Chambers	Distributed overhead cardioid-like loudspeakers.	1.0	0.60	6
4	Northern Territory Parliament	Suspended steered array with downfill element	1.2	0.66	8
5	New Zealand Parliament	Suspended beam steered array with downfill element and distributed small gallery loudspeakers.	1.3	0.58	6.5
6	Grand Concourse at Central Station Sydney	Distributed beam steered Arrays (not time sequenced	3.7	0.46	3
7*	Small reverberant chamber*	Small test loudspeaker.	1.9	0.5	1.5

Table 1 List of sound systems providing the IRs used in this study.

* Reported at Reproduced Sound 19 in 2003

The Reverberant Chamber system is representative of a typical sound system in a moderately high reverberant environment. Typical of many situations, the reverberation times reduce with increasing frequency and the loudspeaker has low directionality.

System No	Location	Ease of Discerning Words	Subjective Listening Comments and
1	Brisbane Banco Court	Reasonable intelligibility but could be better. Some effort required.	Intimacy is not particularly high. Room RT is uncomfortably high and audience is located a long way from loudspeakers which have little horizontal directivity, The intelligibility is surprisingly good given these factors, but is not effortless.
2	Brisbane Airport Link Road Tunnel	Some difficulty but surprisingly easy given the RT.	Unexpectedly intimate given the RT. Long, delayed tunnel system with high background noise which masked some of the reverberance. Band limited by the nature of the loudspeakers, very loud listening levels used.
3	City Of Sydney Council Chambers	Relatively clear, but some effort required	Relatively clear but not particularly intimate and lacking crispness. A little muddy in quality due to reverberation.
4	Northern Territory Parliament	High intimacy, almost effortless intelligibility	Intimate and full, rich sounding,
5	New Zealand Parliament	Good intimacy, minor effort required on occasions	Reasonably intimate and rich sounding. Loudspeakers in gallery above the chamber add substantial reverberant sound and degrade intimacy. Direct field is excellent and intimate.
6	Rail Grand Concourse	Some difficulty but surprisingly easy given the RT.	Unexpectedly intimate given the RT. Long reverberant tail. Good tonal balance, large echoes at some positions resulting from reflections from the hard floor onto the curved roof and back to the listener.
7*	Small reverberant chamber*	Poor intelligibility and significant effort required to discern words.	Difficulties due to excessive reverberation at low mid-range frequencies.

Table 2: Subjective impressions of the intelligibility of the seven sound systems under test.

The impulse responses of the seven systems listed in Table 1 are given in the Appendix, along with the MTI, modulation transfer function (MTF) matrix and STI data computed from the IRs. Most of the MTF values decrease monotonically, indicating the absence of strong echoes.

The reverberation times (RT) in each of the spaces with the sound systems are given in Table 7 in the Appendix. These RTs were computed over the later sections of the linearised Schroeder decay curve, to allow the results to approach the RTs that would be measured with an omnidirectional source.

Banco Ceremonial Court, Brisbane Australia

City of Sydney Council Chambers, Sydney Australia

Nothern Territory Legislative Assembly, Darwin Australia

Debating Chamber in New Zealand Parliament, Wellington NZ

Grand Concourse – Central Station, Sydney `

Figure 1 Photos of some of the sounds systems that were used in this study.

2.2 Build Up Ratio

Early-to-late energy ratio (ELR) calculated by Eq. 1 has been historically used as a measure of a system's ability to deliver speech intelligibility. ELRs show the total level arriving before a split-time *t* relative to the total level arriving after *t*.

Build-up ratio (BUR) described by Eq. 2 shows the total level arriving before split-time interval *t* relative to the total level of sound. When a range of split-time intervals *t* is used, the BUR shows the temporal pattern of build-up or integration of the sound. With a split-time period of 50 ms, the BUR equates to $10*\log_{10}$ of the Definition or Deutlichkeit. The term "build-up ratio" is preferred by the authors over the term "integration ratio" as BUR reflects the physical process, and differentiates it from the term "integration time".

$$ELR = 10 * log_{10}\left\{\int_0^t p^2(t)dt / \int_t^{inf} p^2(t)dt\right\} \qquad \text{Eq. 1}$$

$$BUR = 10 * log_{10}\left\{\int_0^t p^2(t)dt / \int_0^{inf} p^2(t)dt\right\} \qquad \text{Eq. 2}$$

Figure 2 illustrates the differences between the ELR and BUR five octave bands computed from the IR of a road-tunnel sound system.

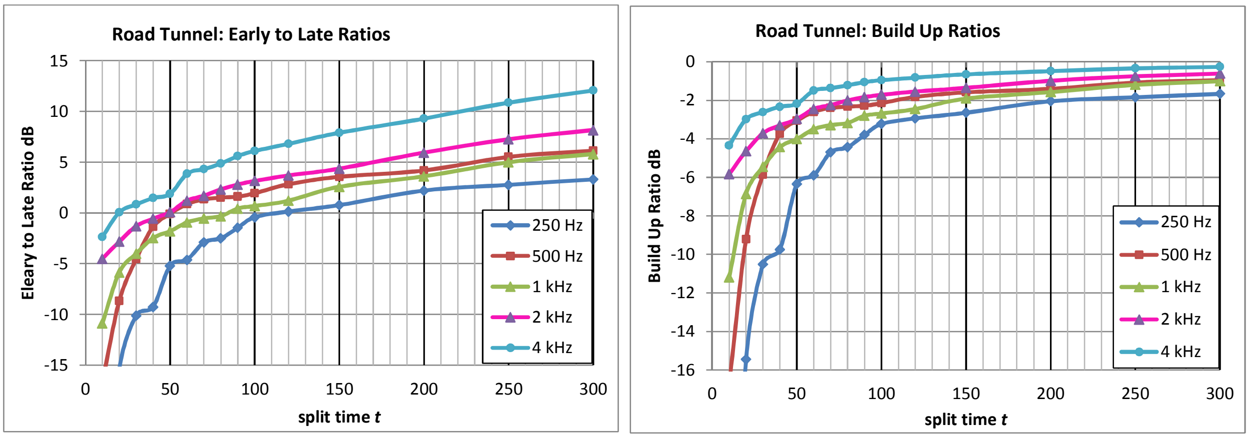


Figure 2 Example of build-up and early-to-late ratios with split times ranging between 10 ms and 300 ms

Compared to the ELR, the BUR has two advantages:-

- BUR is by its nature normalised and tends towards 1 with increasing split-time, allowing easier comparison between systems and frequencies, compared to the ELR which monotonically increases with split-time.
- The early sound field can be quantified relative to the total sound field.

In addition, the BUR has some similarity to the LOC parameter discussed in the next section.

Using the IRs for the sound systems, build-up ratios were computed in the octave bands 250 Hz to 4 kHz with the split-time t ranging from 10 ms to 300 ms.

2.3 LOC Parameter

It is beyond the scope of this paper to delve into the concepts behind the LOC parameter, and the reader is referred to [1] for comprehensive information. LOC is a measure of the ease of localisation, where LOC = 0 is assumed to be the threshold, and LOC = +3dB represents adequate perception for engagement and localisation.

Griesinger notes that experiments led to a measure based on the impulse response that predicts the threshold for horizontal localization for male speech. The measure counts the nerve firings above 700 Hz in a 100ms window that result from the onset of a continuous direct sound, and compares that count with the number of nerve firings that arise from the reflections in the same 100ms window.

Eq. 3 from (1) describes the calculation of the LOC parameter.

$$S + 1.5 + 10 * \log \int_{0}^{.005} p(t)^2 dt - (1/100) * \int_{.005}^{100} POS(S + 10 * \log \int_{.005}^{\tau} p(t)^2 dt) d\tau$$

$$\text{where } S = 20 - 10 * \log \int_{0}^{\infty} p(t)^2 dt \qquad \text{Eq. 3}$$

- *S* is a constant that establishes a sound pressure at which nerve firings cease, assumed to be 20 dB below the peak level of the sum of the direct and reverberant energy.
- *p(t)* is an impulse response measured in the near-side ear of a binaural head.
- *p(t)* is band limited to only include frequencies between 700 Hz and 4000 Hz.
- *POS* means positive values only.
- The calculation assumes the direct field is 5 ms long.
- The Direct/Reverberant ratio D/R is also computed for the 700 Hz and 4000 Hz bandwidth.

Binaural impulse responses should be used to calculate the LOC, and the strongest sound is assumed to arrive at the left ear.

LOC values were computed for the impulse responses of the seven systems. As binaural IRs were not available, IRs which were measured with a free-field microphone were used. Although this means the LOC values will not be in accordance with Griesinger’s use of the parameter, they are useful for comparison among the seven sound systems.

2.4 Modulation Spectra of Speech

Three short speech phrases, recorded anechoically by three different male talkers, were convolved with the impulse responses of the seven systems. The three phrases were:-

- “The birch canoe slid on the smooth planks.”
- “Glue the sheet to the dark blue background”
- “I left my heart to the sappers round Khe Sanh”

The modulation spectra of the original and convolved phrases were analysed using the following process:-

a) The speech was filtered into octave bands between 250 Hz and 4 kHz using 6th order Butterworth bandpass filters.

b) The Hilbert transform of each filtered file was computed, and the analytic signal formed using the original signal and the Hilbert transformed signal. (Note the analytic signal is a complex quantity with real and imaginary components).

c) With a carrier signal that has been amplitude-modulated, (such as occurs with speech), the magnitude of the analytic signal of the carrier contains the modulating signal. Calculating the absolute value of the analytic signal provides the envelope of the carrier, with a process that has similarity to full wave rectification.

d) The DC component of the envelope was removed by a high pass filter. With the short term, transient-like modulations in speech, the start and stop of the modulating signal introduces low frequency spectral components into the modulation. To minimise corruption of the speech spectrum by these components during the high pass filtering process, experiments were conducted to find the high pass filter that produced the least spectral leakage. A forward and reverse 2nd order Butterworth filter at 0.5 Hz was found to yield the best compromise between low frequency roll-off and settling time. However, the settling time of this filter is still sufficiently long that it effectively elongates the demodulated waveform.

e) To prevent spectral leakage with a rectangular window during the FFT process, each speech segment was padded with two seconds of leading and trailing zeros.

f) A 16th order FFT was then used to extract the spectrum and the spectrum smoothed to 1/6th octave, then the data was exported to Excel for comparison.

The modulation spectra of the three phrases are shown in Figure 3..Differences between the talkers and the phrases will have produced the visible differences in the spectra. Figure 4 compares the temporal pattern of modulations in three octave bands of the anechoic and reverberated “Khe Sanh” phrase.

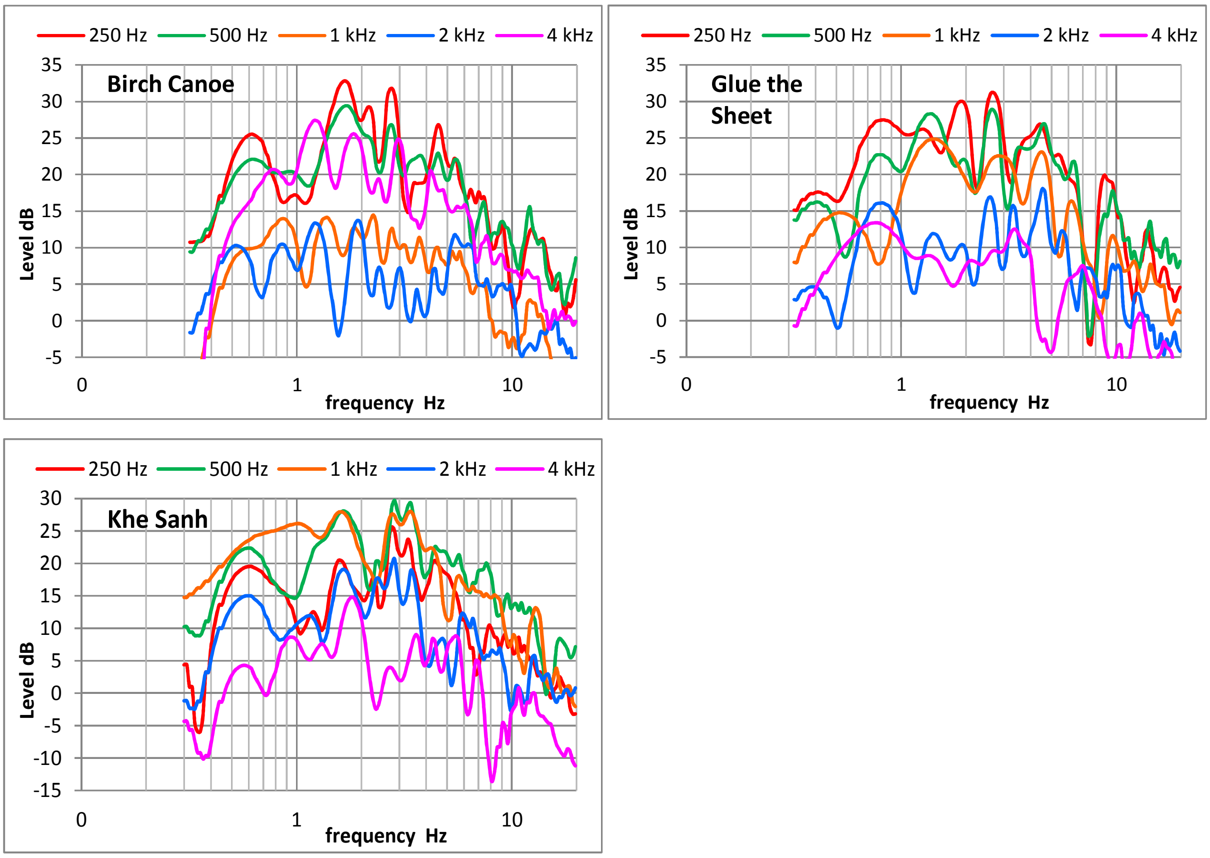
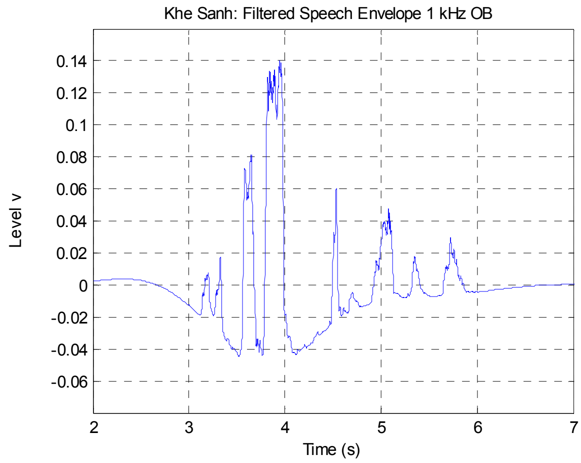
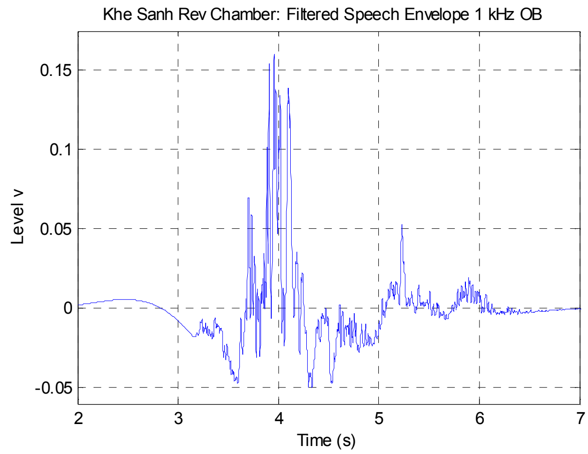


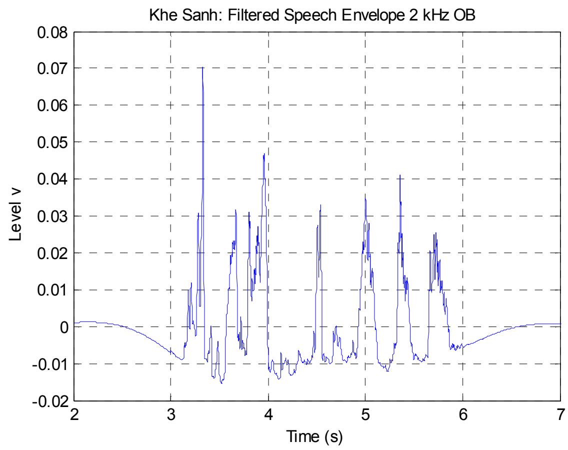
Figure 3 Modulation spectra of the three phrases



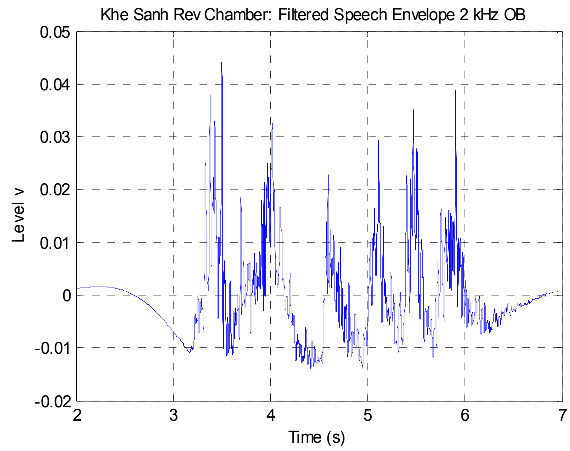
Anechoic envelope in 1 kHz band



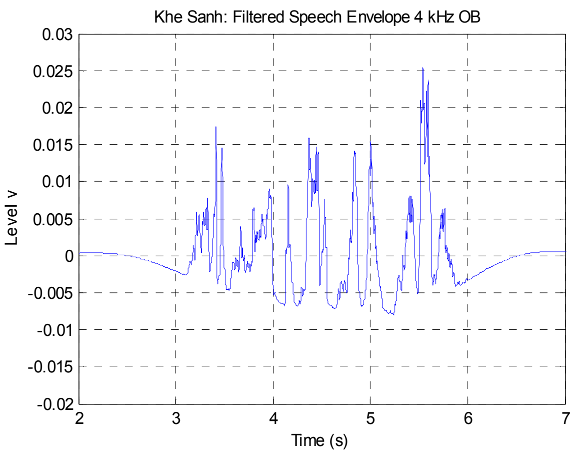
Reverberated envelope in 1 kHz band



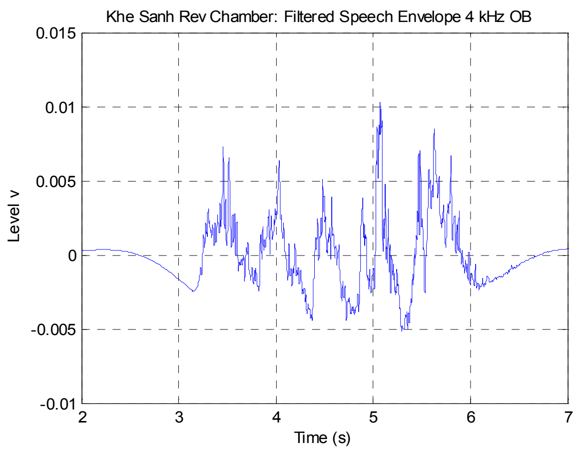
Anechoic envelope in 2 kHz band



Reverberated envelope in 2 kHz band



Anechoic envelope in 4 kHz band

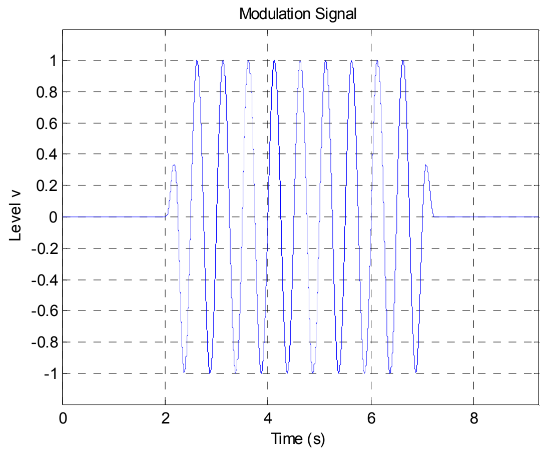


Reverberated envelope in 4 kHz band

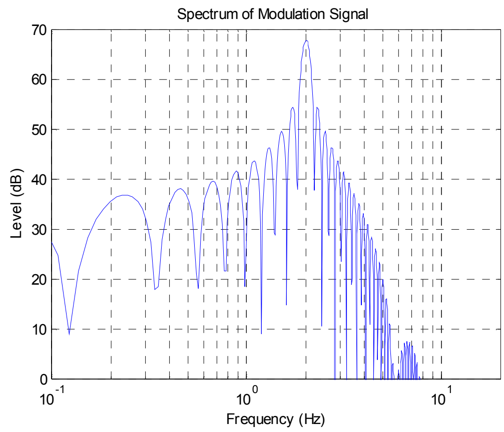
Figure 4 Examples of the temporal pattern of modulations in three octave bands of the anechoic and reverberated “Khe Sanh” phrase.

2.4.1 Calibration Process

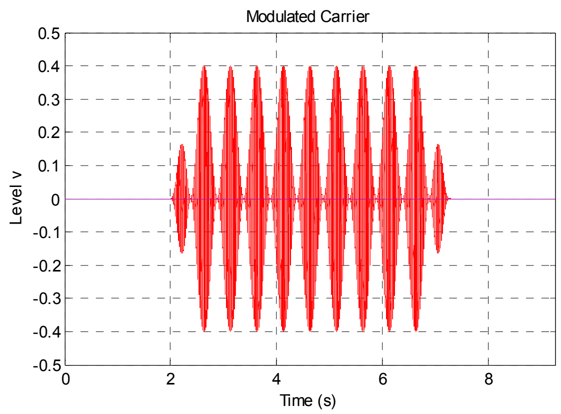
The process of extracting the modulation spectrum of speech was checked by feeding in known modulation values using both sine wave and noise carriers, and a STIPA signal.



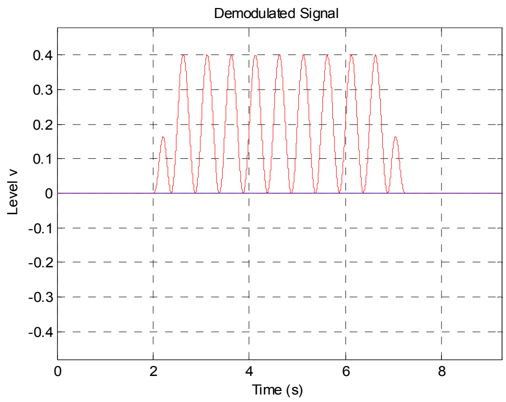
Waveform of modulating signal at 2 Hz



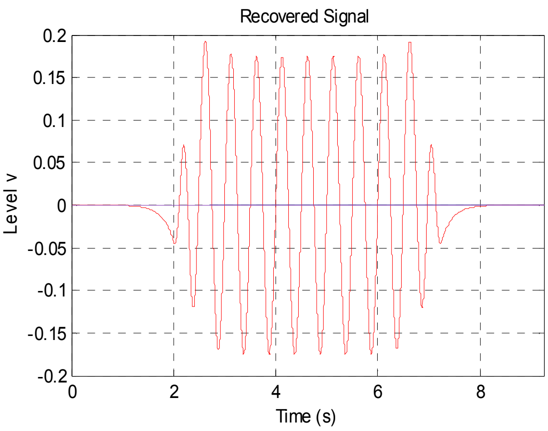
Spectrum of modulating signal at 2 Hz



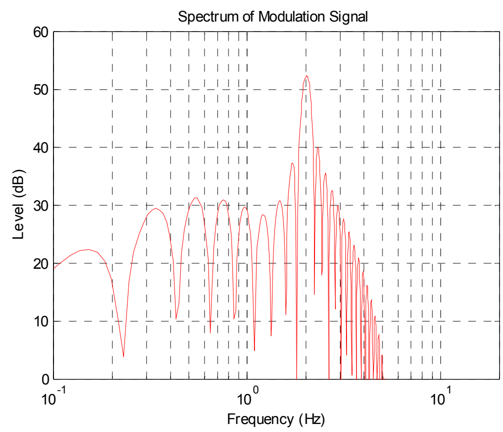
1 kHz carrier 100% modulated at 2 Hz



Demodulated carrier envelope



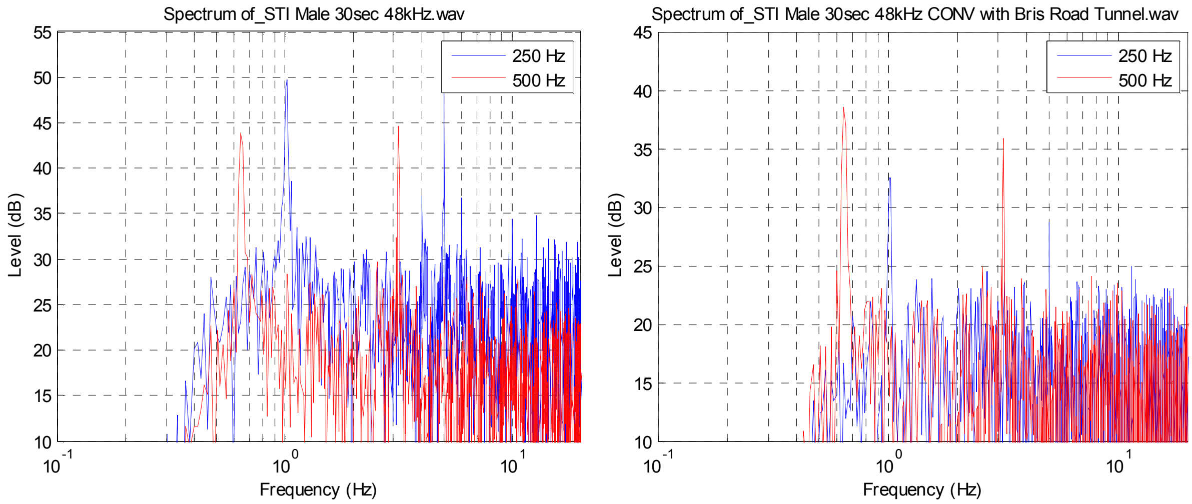
Recovered audio signal after high pass filtering



Spectrum of recovered audio

Figure 5 Illustration of the demodulation process

The demodulation process was also checked using a STIPA signal, which was reverberated with the Tunnel IR. Figure 6 compares the modulations of the STIPA test signal in 250 Hz and 500 Hz bands with the reverberated equivalents Noting that the STIPA signal consists of steady-state modulations, the anticipated low pass filter action of the reverberation is evident, with considerable loss in the modulation level occurring.



Raw STIPA signal

Reverberated STIPA signal with Tunnel IR

Figure 6 Comparison of raw and reverberated STIPA signals in 250 Hz and 500 Hz bands.

3 RESULTS – BUILD UP RATIOS

The build-up ratios in each octave band for each system are shown in Figure 23 in the Appendix, with the split time t relating to the start of each IR. The arithmetic (not energy) averages of the BURs over the range 250 Hz to 500 Hz and 1 kHz to 4 kHz were computed and are shown in Figure 7 and Figure 8.

We considered weighting the octave band BUR data in the averaging by the values used in the STI method to reflect the relative importance of each octave band to intelligibility. However, noting the effects of upward masking[2] on subjective intelligibility, we decided to not weight the BURs, as this would reduce the impact of the 250 Hz and 500 Hz octave bands on the average BUR.

The following observations are made:

- Given the vast differences in RT, the differences in BURs are relatively small; less than 5 dB separate all systems.
- Northern Territory Parliament has the highest BUR at each split time.
- Grand Concourse shows rapid rise of BUR up to 40 ms, followed by a slow rise to beyond 300 ms due to the high RT and strong echoes. Given the high RT and the room volume in which this system operates, its BUR performance is remarkably close to that of the other systems.
- Road tunnel shows a rapid rise of BUR up to 50 ms. After 70 ms, the BUR rise slows, due to the high RT.
- Reverberant chamber has the lowest BUR up to 50 ms, and second lowest up to 100 ms. This is surprisingly low given its STI performance, and probably relates to a reasonably omnidirectional loudspeaker in a high RT situation.

Hypothesising that the build-up ratios at 30 ms and 50 ms in specific frequency ranges are important to intimacy, we compared the relative BUR_{30} and BUR_{50} performances with respect to our subjective impressions. The comparisons are shown in in Table 3.

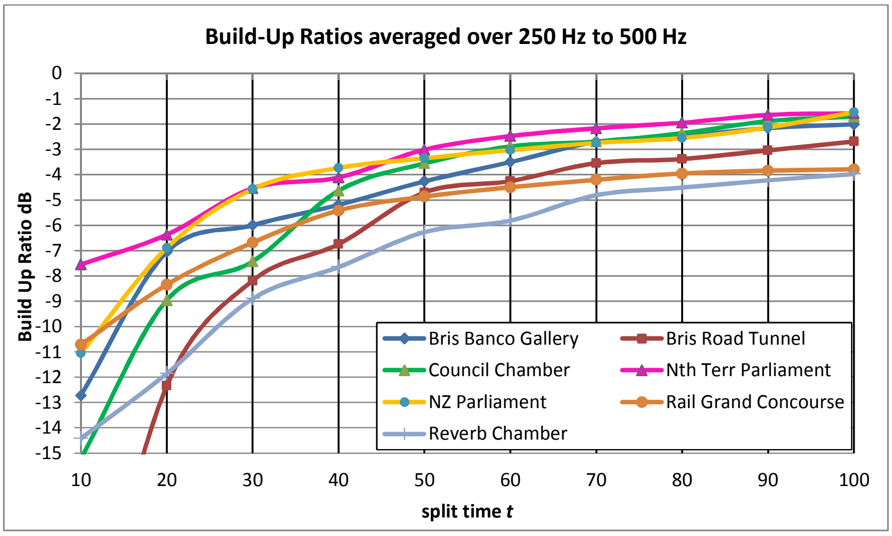


Figure 7 Build up-ratios averaged over 250 Hz to 500 Hz for the seven sound systems.

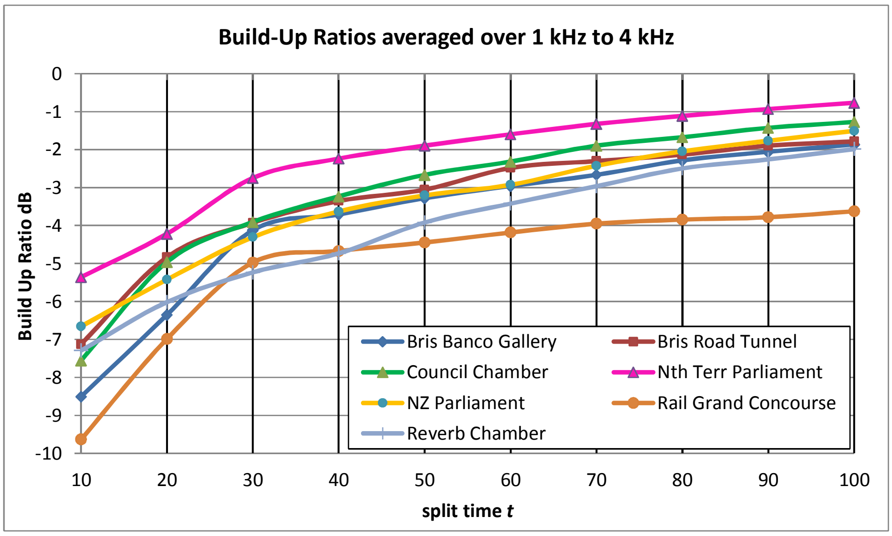


Figure 8 Build up-ratios averaged over 1 kHz to 4 kHz for the seven sound systems

30 ms		BUR_{30} value		BUR_{30} rating 7= highest 1= lowest		Subjective Rating
System No	**Location**	**250Hz to 500Hz**	**1kHz to 4kHz**	**250Hz to 500Hz**	**1kHz to 4kHz**	**7= highest 1= lowest**
1	Brisbane Banco Court	-6.0	-4.1	5	4	3
2	Brisbane Airport Link Road Tunnel	-8.2	-3.9	2	5	4
3	City Of Sydney Council Chambers	-7.4	-3.9	3	6	5
4	Northern Territory Parliament	-4.5	-2.7	7	7	7
5	New Zealand Parliament	-4.6	-4.3	6	3	6
6	Rail Grand Concourse	-6.7	-5.0	4	2	2
7*	Reverberant chamber	-8.9	-5.2	1	1	1

Table 3 BUR_{30} results in two bandwidths and rating of systems according in order of performance

50 ms		**BUR_{50} value**		**BUR_{30} rating 7= highest 1= lowest**		**Subjective Rating 7= highest 1= lowest**
System No	**Location**	**250Hz to 500Hz**	**1kHz to 4kHz**	**250Hz to 500Hz**	**1kHz to 4kHz**	
1	Brisbane Banco Court	-4.3	-3.3	4	3	3
2	Brisbane Airport Link Road Tunnel	-4.7	-3.1	3	5	4
3	City Of Sydney Council Chambers	-3.6	-2.7	5	6	5
4	Northern Territory Parliament	-3.0	-1.9	7	7	7
5	New Zealand Parliament	-3.4	-3.2	6	4	6
6	Rail Grand Concourse	-4.9	-4.4	2	2	2
7*	Reverberant chamber	-6.3	-3.9	1	2	1

Table 4 BUR_{50} results in two bandwidths and rating of systems according in order of performance

Noting that only a few dB separate the BUR results for both frequency ranges, and that our subjective tests were cursory, drawing conclusions is dangerous. However:

- The BUR 50 ratings seem to match our subjective ratings reasonably well
- Subjective intelligibility is improved with systems that have a high BUR in the 250 Hz to 500 Hz range.

3.1 LOC Metric

Table 5 partly re-states data in Table 1 and includes the computed LOC and Direct/Reverberant ratios in the 700 Hz to 4 kHz bandwidth. The RT values have been averaged over the 1 kHz to 4 kHz octave bands to reflect the bandwidth of the LOC metric.

System No	**Location**	**LOC**	**D/R**	**ave. RT60 1k-4k OB**	**Measured STI**	**Subjective Intelligibility 0 to10 scale**
1	Brisbane Banco Court	-5.7	-11.1	1.88	0.53	3.5
2	Brisbane Road Tunnel	-2.6	-7.5	3.01	0.52	4
3	Sydney Council Chambers	-5.4	-9.9	0.99	0.60	6
4	Northern Territory Parliament	-1.8	-5.9	1.09	0.66	8
5	New Zealand Parliament	-1.8	-7.2	1.28	0.58	6.5
6	Rail Grand Concourse	-1.9	-9.0	3.77	0.46	3
7*	Small reverberant chamber*	-0.8	-7.3	1.61	0.5	1.5

Table 5 Data for the seven sound systems with LOC and D/R values noted.

Observations of these results are:-

- The Council chamber (System 3) performs well with STI but scores a low LOC. Our subjective impressions are that the system does not provide the subjectively comfortable intelligibility expected of a system measuring STI 0.6.

- Concourse (System 6) has a poor STI but amongst the highest LOC. Our subjective impression of this system is that the system performs very well in the highly reverberant environment and large space.
- One surprise with the LOC result is the small reverberant chamber, as this sounds quite reverberant and requires concentration to extract the intelligibility. However, at frequencies between 250 Hz and 500 Hz at which the speech power is greatest, the reverberation time is 2.2 seconds and the combination of these two factors produces this perception, along with the contribution of upward masking effects. At higher frequencies, the average RT is 1.7 and this aids the LOC value.

To assist consideration of the data trends, an attempt was made to compare the measured parameters for the seven systems on one graph using a dimensionless scale of 1 to 5. The values of each parameter were scaled to fit this range. BURs with split times of 30 ms, 50 ms and 100 ms averaged over the 1 kHz to 4 kHz range were incorporated (corresponding to the LOC frequency range). Figure 9 shows this data for each system, with the scaled reverberation times.

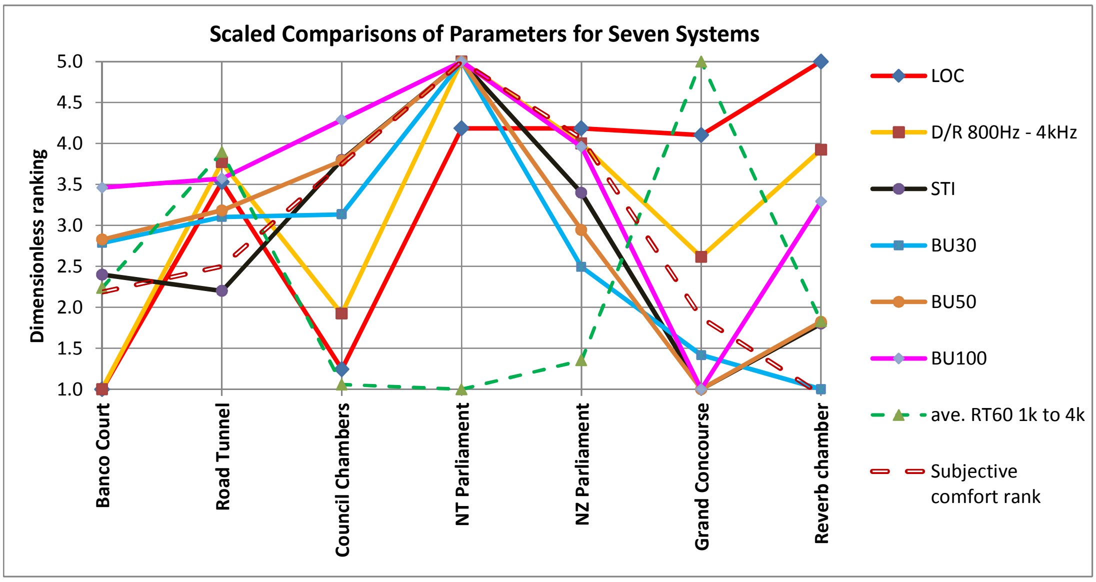


Figure 9: Comparison of measures for Seven Sound Systems .

We conclude that:-

- None of the parameters used in this study consistently reflect our subjective impressions of “comfortable” speech intelligibility performance.
- STI and (monaural) LOC do not appear to be related.
- LOC (calculated without a binaural IR) seems to be partly related (with two exceptions) to our perceptions of subjective intelligibility comfort for a given difficulty of acoustic environment.
- The high LOC value of the reverberant chamber is surprising giving the comparative difficulty of discerning words in this situation.
- The Council Chamber, despite a relatively low reverberation time shows a wide spread in metrics.
- The Railway Grand Concourse, with its high RT also presents a wide spread in metrics, and its perceptual ranking falls somewhere between the relatively high LOC and the relatively low STI.
- The Nth Territory Parliament scores highly on all metrics, including the subjective comfort rank.
- Simple band-limited D/R Ratio shows some consistency with subjective ranking.
- The road tunnel should perform better on STI than it does, according to the other metrics.

3.2 SPEECH SPECTRA

3.2.1 Results

Figure 10 gives an example of the analysis process and shows the modulation spectrum of the phrase “glue the sheet to the dark blue background” in the 1 kHz and 4 kHz octave bands for both the anechoic and reverberated speech with the seven systems. Figure 24 to Figure 26 in the Appendix provides the entire set of modulation spectra for the three phrases and five octave bands.

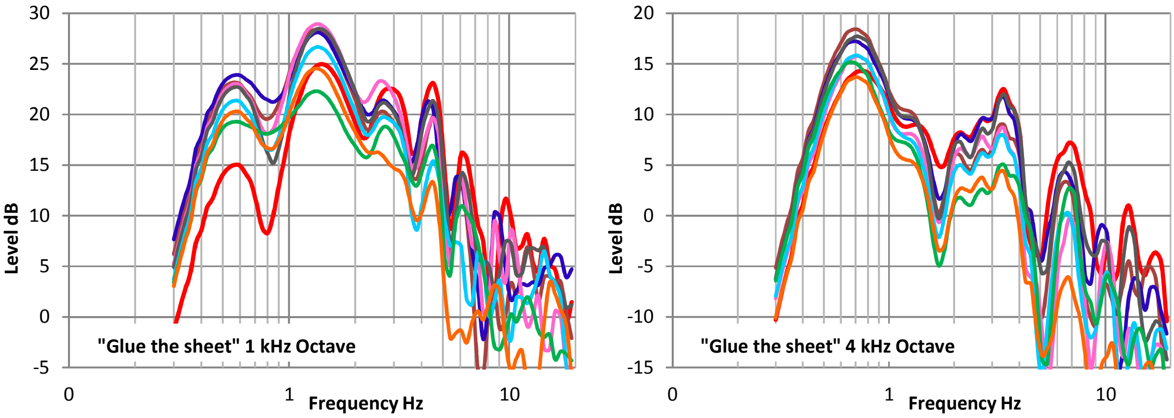


Figure 10 Modulation spectra for “Glue the sheet” in 1 kHz and 4 kHz octave bands

The reverberated modulation spectra shown in the Appendix were normalised to the octave-band levels of the anechoic speech and then referenced to the anechoic modulation spectra for analysis.

A wide range of amplitudes and unusual behaviours with respect to the original anechoic spectra is observed. Some of the more noteworthy spectra are shown in Figure 13 to Figure 20 and discussed in the table. No change to the modulation spectra would correspond to a horizontal line in the figures.

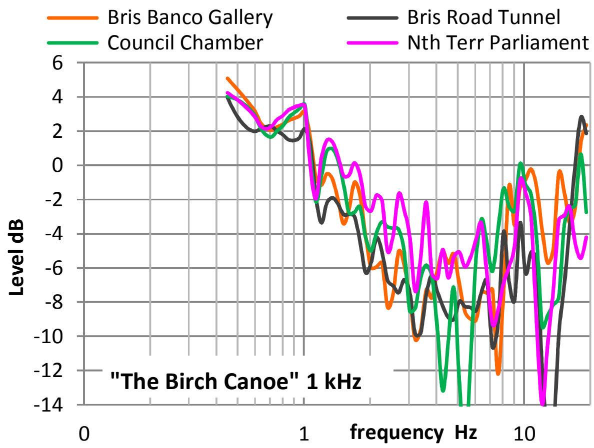


Figure 11. Low pass filtering action of reverberation is quite evident. The increase in modulation level above 8 Hz may be related to the inability of the reverberant field to charge up with short duration sounds.

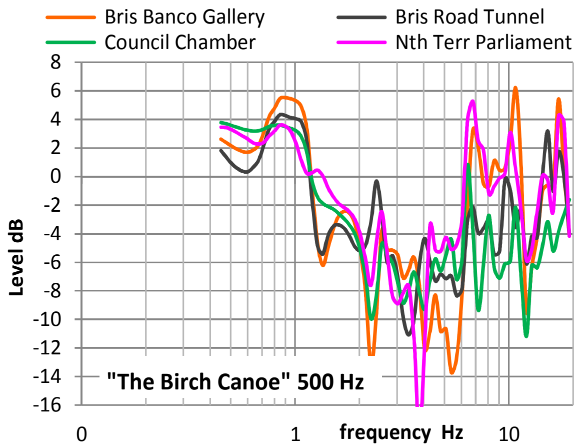


Figure 12. Modulation is actually increased at some frequencies. This is discussed in Section 3.2.2. Interesting that the NT Parliament, (the most intimate sounding system) produces this excess. This does not match the strong reflection in the IR at 15 ms.

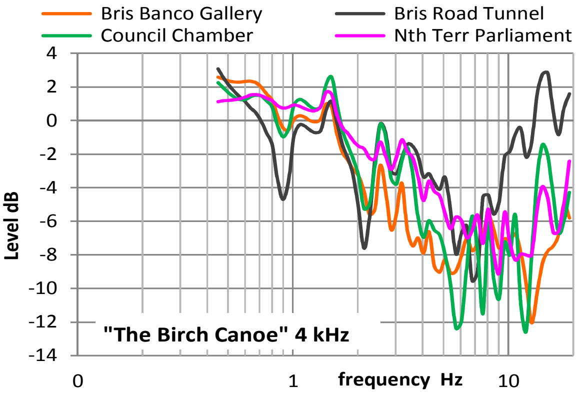


Figure 13 NT Parliament which is the most aurally comfortable system generally shows the smoothest and least roll off with increasing modulation frequency

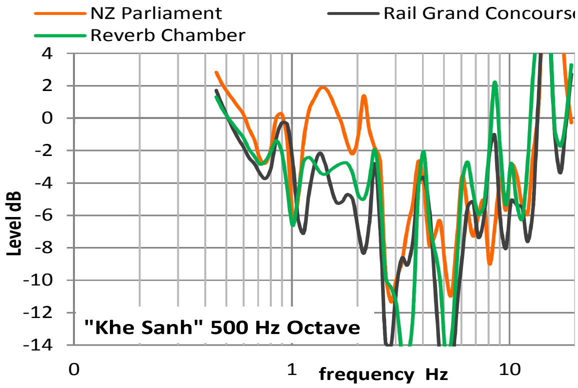


Figure 14 Although Reverb Chamber has lower RT and higher STI than the Concourse, it has similar modulation loss.

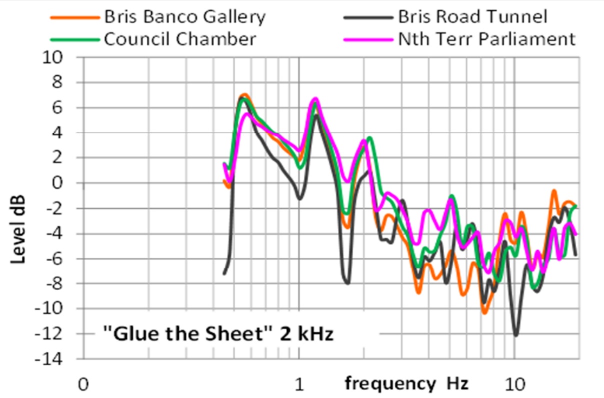


Figure 15 Although the RT of Tunnel (3 s) is longer than NT Parliament (1 s), loss of modulation is relatively minor & similar to Banco Court (2 sec)

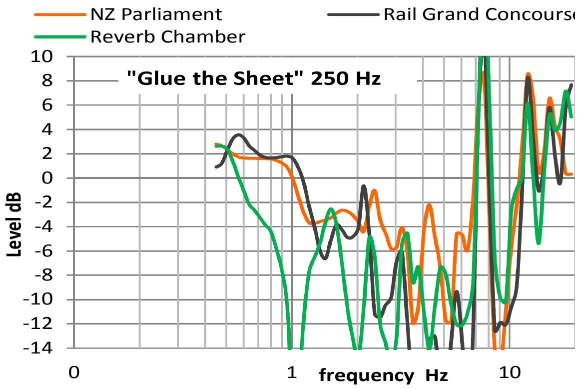


Figure 16. Although the RT of Grand Concourse (5.6 s) is longer than Reverb Chamber (2.4s), loss of modulation is much greater in the Reverb Chamber in this OB.

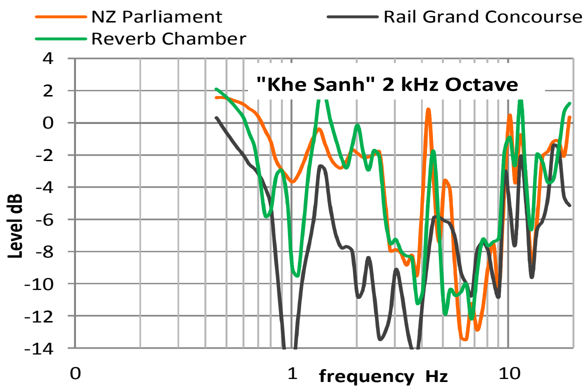


Figure 17. 2 kHz OB- Example 1

Significant loss of modulation in in Grand Concourse (RT of 3.8 s) from 1Hz to 4Hz compared to Rev Chamber (1.7 s), which probably contributes to lower STI score.

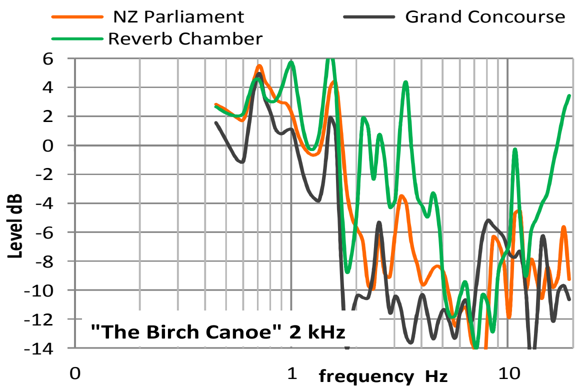


Figure 18. 2 kHz OB- Example 2

Similar result for Grand Concourse to Figure 17. The Reverb Chamber shows little loss in 1-4 Hz range.

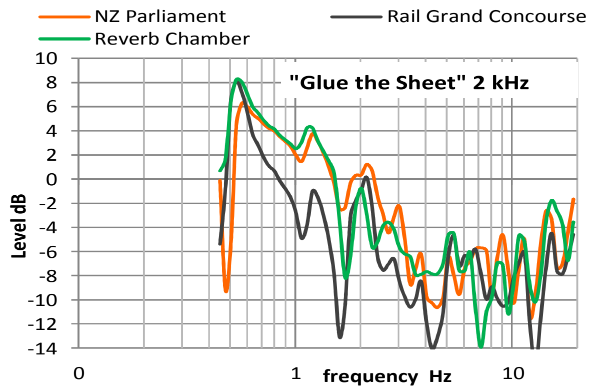


Figure 19. 2 kHz OB- Example 3

Concourse shows greatest loss but the actual modulation levels are remarkably different for the three phrases.

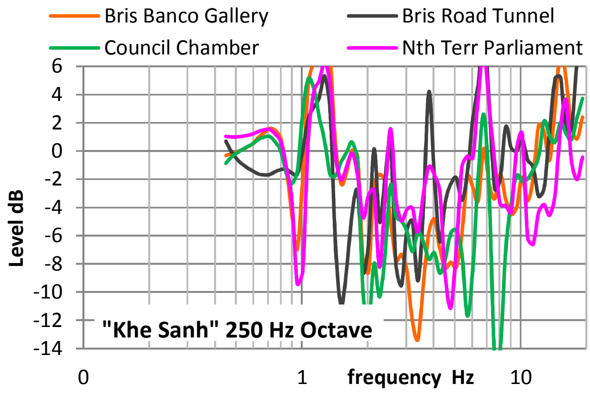


Figure 20. Significant level of additional modulation frequencies.

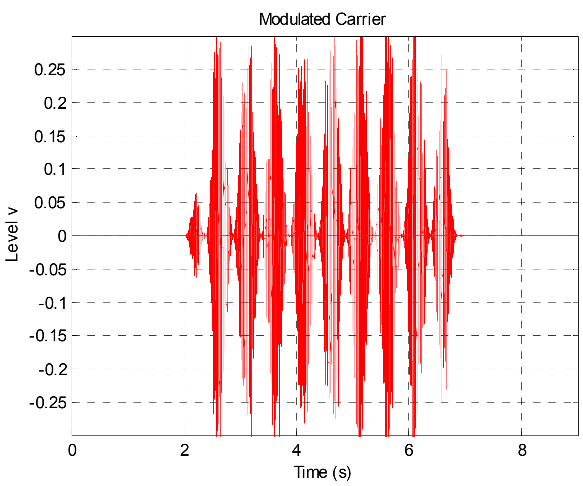
3.2.2 Discussion of additional modulation frequencies

As noted above, the spectra in figures in Section 3.2.1 above show amplification in modulation at some frequencies and in some cases, new modulation frequencies have appeared in the reverberated spectra.

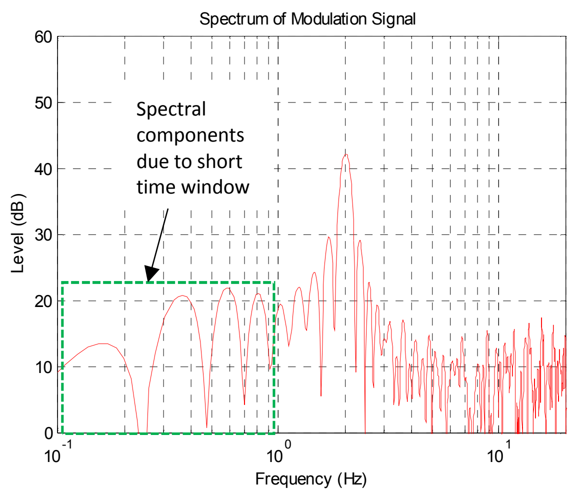
This was investigated using the following process:-

Noting that speech has a mixture of noise and sine wave carriers, simple modulated signals were constructed using both noise and sine wave carriers, and reverberated with the Tunnel IR. A 1 Hz sine wave was used to modulate a white noise carrier and a 1 kHz sine carrier at 100%. Figure 21 shows time waveforms and modulation spectra for anechoic and reverberated sine wave at 1 Hz with noise and sine carriers. Comparing the anechoic and reverberated spectra shows that modulation harmonics have been introduced by reverberation. The harmonics with noise occur essentially at 1 Hz intervals and are typically 14 dB below the fundamental of 2 Hz. With the sine carrier, the distortion is primarily 2nd harmonic, with a level only 11 dB below the fundamental. Inspection of the waveform of the sine carrier shows a flattened section of the modulating sine wave, which would be expected to produce even harmonics.

.

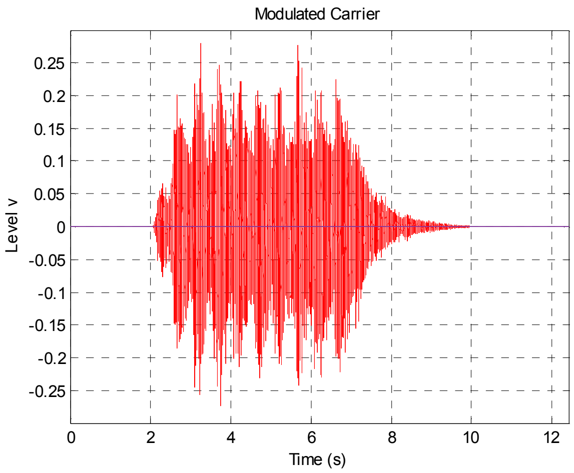


Waveform of anechoic modulation with noise

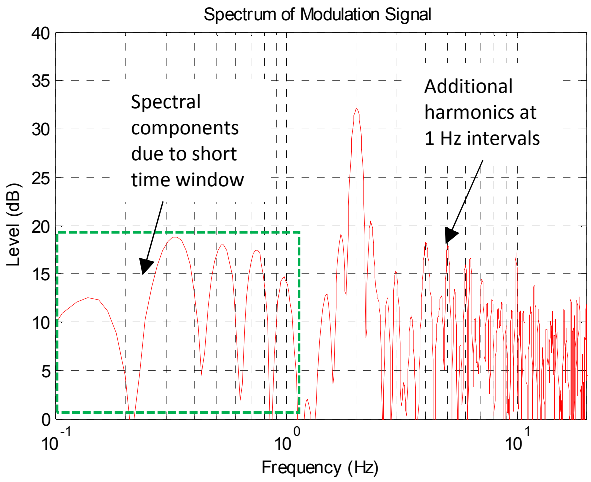


Modulation spectrum of signal at left.

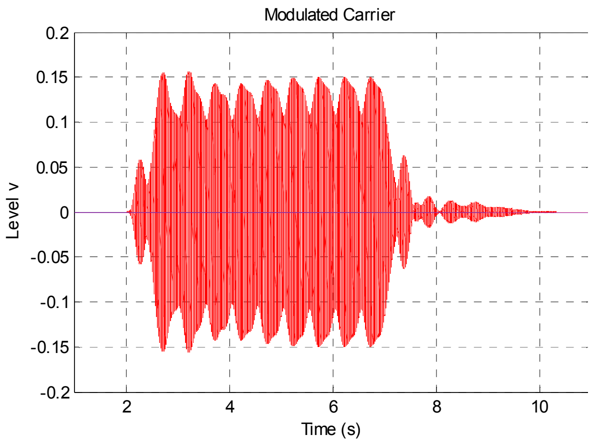
carrier 100% at 1 Hz



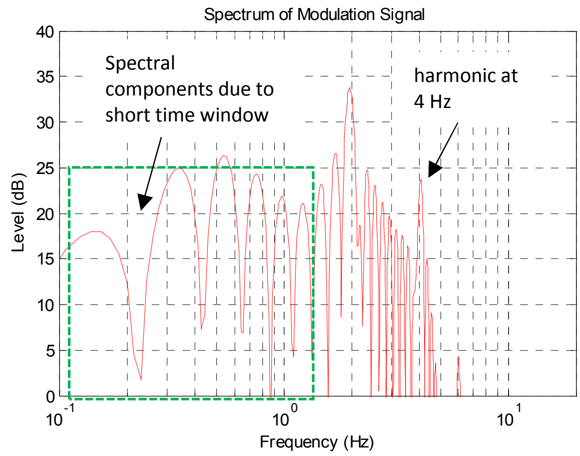
Waveform of noise carrier above convolved with Rail Concourse IR



Modulation spectrum of signal at left



Time waveform of sine carrier convolved with rail concourse



Modulation spectrum with sine carrier

Figure 21 Time waveforms and modulation spectra for anechoic and reverberated sine wave at 1 Hz with noise and sine carriers.

4 CONCLUSIONS

The following conclusions are made:-

a) The behaviour of seven different sound systems in different room acoustic environments has been explored with regard to aspects of their short-term temporal behaviour.

b) Systems with higher BUR_{30} values in a given frequency range generally show smoother and less change in the frequency content of the speech modulation in that frequency range. There are however significant exceptions to this generalisation.

c) The build-up ratio with a split time of 50 ms (BUR_{50}) in the 250 Hz to 500 Hz range appears to have a bearing on the subjective intimacy of amplified sound and the amount of effort that is required to discern words. As the long-term speech spectrum has a maximum level in the 250 Hz to 500 Hz range and this high level can produce upward masking in reverberant situations, higher BUR_{50} values in this frequency range suggest that there will be less upward masking than systems with lower BUR_{50} values.

d) The process of reverberation and reflection can produce harmonics in the modulation domain of the source speech signal. A number of modulation spectra showed unexpected results.

e) The degree of change in the modulation frequencies in reverberant environments (as distinct from the simple low pass filter action of reverberation) is surprisingly high.

f) Sound systems which sound subjectively intimate to the authors, or surprisingly intimate given the high reverberation times, show overall less difference between their speech modulation spectra and the anechoic modulation spectra.

g) Compared to speech, STI uses a steady state modulated test signal and this precludes a dynamic assessment of the modulation signals. The different speech phrases produced remarkably different spectra modulation when convolved with the various system responses. Steady-state measurements may be unsuitable for analysis of intelligibility performance.

h) The Reverberant Chamber system is representative of a typical unsophisticated sound system in a moderately reverberant environment. Although its STI is 0.5, it shows the lowest BUR_{30} and the greatest degradation on modulation spectra in the 250 Hz and 500 Hz bands. It is the least intimate-sounding of the systems and requires the highest degree of concentration to discern the words.

i) But can we rely on the STI? We consider that STI is a very useful metric, but it should not be relied on to provide a statement of a system's subjective speech intelligibility performance.

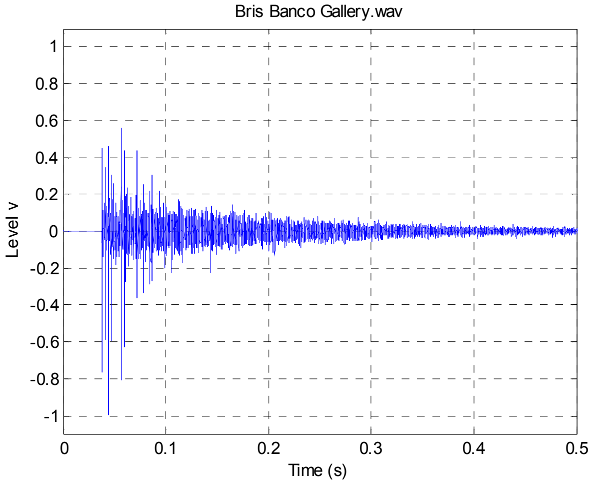
With speech, the complexities in the physical acoustic and the psycho-acoustic domains are complex and involve short-term, asymmetric signals with varying frequency content. The STI method is not able to account for this dynamic asymmetric structure.

j) Given the prevalence of specifications requiring a system STI of 0.5, we believe that there is some risk that the systems may not fulfil their intended purpose. In this context, we recommend the use of the performance categories in the current STI standard[10].

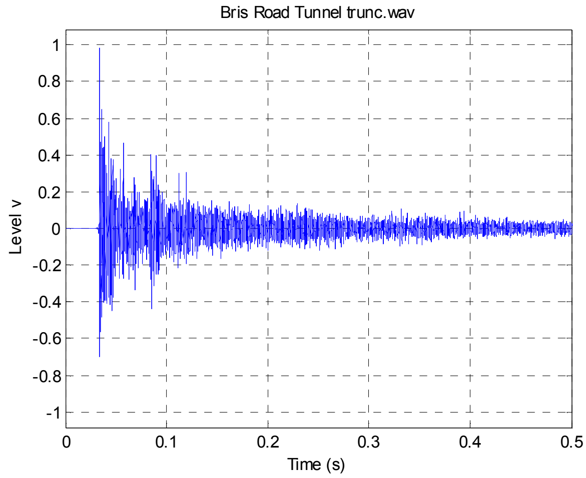
5 REFERENCES

1. G. Leembruggen, A. Stacy. Should the matrix be reloaded? Proceedings of the IOA. 2003.
2. G. Leembruggen, M. Hippler P. Mapp. Exploring ways to improve STI's recognition of the effects of poor spectral balance on subjective intelligibility. Proc. IOA. No 4, 2009, Vol. 31.
3. G. Leembruggen, M. Hippler and P. Mapp. Further investigations into improving sti's recognition of the effects of poor frequency response on subjective intelligibility. Convention Paper 128th AES Convention London, UK. 2010.
4. T. Steinbrecher. Speech Transmission Index: Too weak in time and frequency? Proc IOA. 2008, Vol. 30, Pt 6.
5. G. Leembruggen, D. Connor. Design and commissioning of sound reinforcement systems for the Australian Parliament; a holistic approach. J. Audio. Eng. Soc. October 1996, Vol. 44, No. 10.
6. G. Leembruggen. The design of three unusual loudspeakers for the High Court of Australia. J. Audio. Eng. Soc. 2000 May, Vol. 48, No. 5.
7. G. Leembruggen, M. Hanson, T. Hanson. The cone of an ancient New Zealand tree inspires the acoustic design for the new Zealand Supreme Court. Proc IOA. 2011, Vol. 33, Pt 6.
8. D. Gilfillan, G. Leembruggen. Raising the tone of the debate. Proc IOA. 2011, Vol. 33, Pt 6.
9. D. Griesinger. The relationship between audience engagement and the ability to perceive pitch, timbre, azimuth and envelopment of multiple sources. Proc IOA. 2011, Vol. 33, Pt 2.
10. International Standard No. 60268-16. Sound System Equipment: Objective rating of speech intelligibility by Speech Transmission Index. s.l. : IEC., 2011. 3rd Edition.

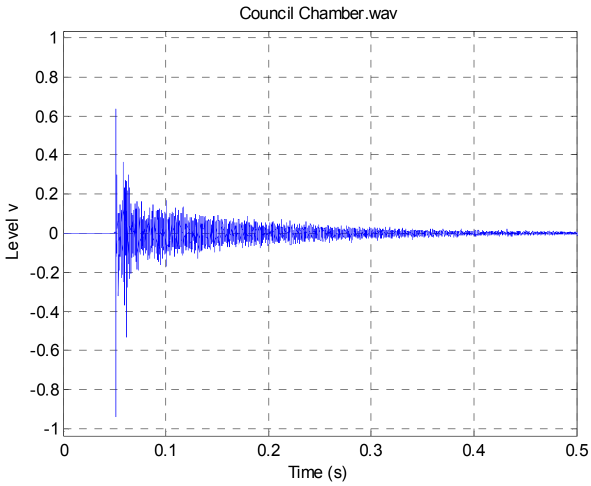
6 APPENDIX



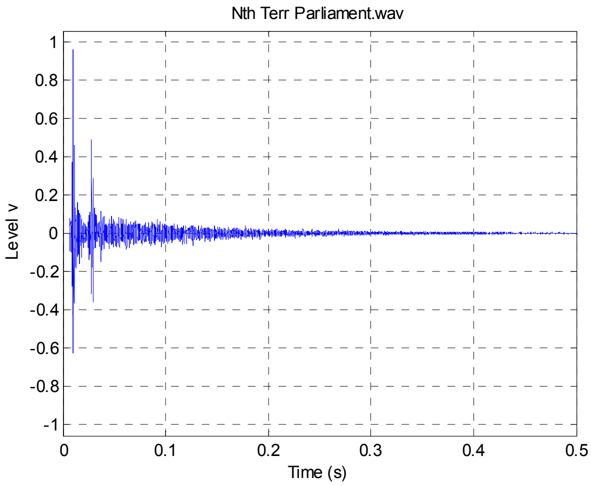
System 1 Brisbane Banco Court



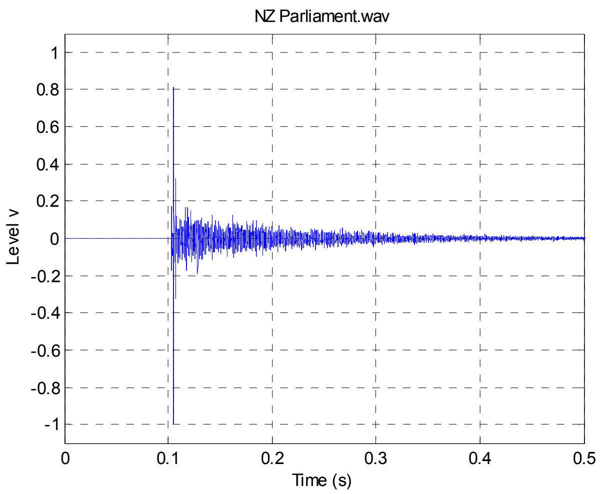
System 2 Brisbane Airport Link Tunnel



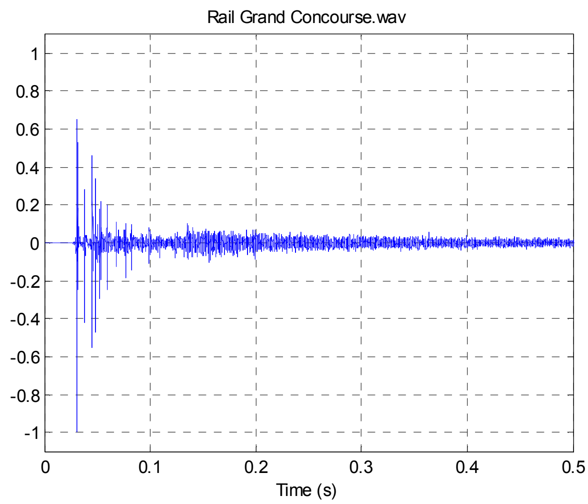
System 3 City Of Sydney Council Chambers



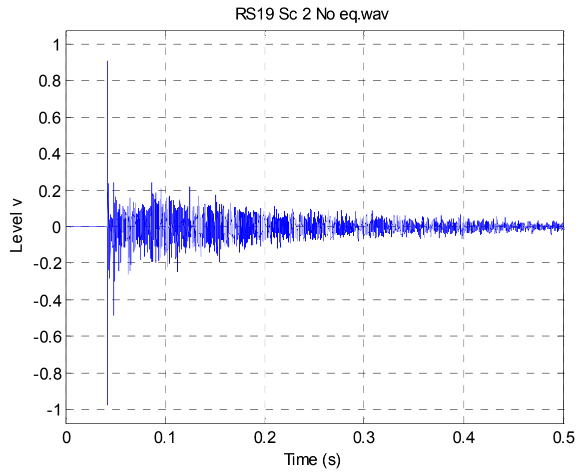
System 4 Northern Territory Parliament



System 5 New Zealand Parliament



System 6 Grand Concourse at Sydney Central Station



System 7 Rev Chamber

Figure 22 Impulse responses of the seven systems.

Bris Banco Court gallery							
	125	**250**	**500**	**1000**	**2000**	**4000**	**8000**
MTI	**0.41**	**0.54**	**0.49**	**0.53**	**0.51**	**0.57**	**0.64**
	0.89	0.90	0.90	0.90	0.88	0.92	0.96
	0.84	0.87	0.87	0.87	0.85	0.89	0.95
	0.77	0.83	0.82	0.82	0.79	0.85	0.92
	0.68	0.77	0.75	0.76	0.73	0.80	0.89
	0.56	0.70	0.67	0.68	0.64	0.72	0.84
	0.45	0.64	0.58	0.61	0.57	0.65	0.78
	0.34	0.57	0.50	0.54	0.49	0.58	0.72
	0.24	0.50	0.42	0.46	0.41	0.51	0.64
	0.18	0.41	0.35	0.40	0.34	0.46	0.56
	0.16	0.38	0.32	0.36	0.28	0.43	0.50
	0.16	0.36	0.27	0.32	0.24	0.39	0.45
	0.05	0.37	0.19	0.29	0.26	0.34	0.39
	0.13	0.26	0.08	0.22	0.30	0.32	0.34
	0.05	0.15	0.07	0.23	0.28	0.27	0.28
STI	**0.53**						

System 1 Brisbane Banco Court

Bris Road Tunnel							
	125	**250**	**500**	**1000**	**2000**	**4000**	**8000**
MTI	**0.28**	**0.43**	**0.52**	**0.46**	**0.52**	**0.6**	**0.66**
	0.394	0.639	0.737	0.737	0.824	0.91	0.959
	0.363	0.587	0.693	0.681	0.775	0.878	0.941
	0.331	0.559	0.665	0.638	0.729	0.845	0.919
	0.333	0.545	0.645	0.595	0.687	0.809	0.892
	0.27	0.528	0.612	0.54	0.638	0.768	0.855
	0.219	0.492	0.552	0.488	0.587	0.725	0.813
	0.233	0.416	0.521	0.428	0.541	0.679	0.763
	0.222	0.38	0.515	0.399	0.508	0.637	0.708
	0.195	0.397	0.487	0.365	0.474	0.589	0.643
	0.131	0.298	0.46	0.326	0.429	0.531	0.566
	0.096	0.305	0.417	0.296	0.386	0.468	0.483
	0.06	0.27	0.357	0.262	0.297	0.399	0.402
	0.018	0.076	0.383	0.26	0.268	0.389	0.382
	0.072	0.104	0.312	0.16	0.261	0.389	0.405
STIr	**0.52**						

System 2 Brisbane Airport Link Tunnel

City Sydney Council Chamber							
	125	**250**	**500**	**1000**	**2000**	**4000**	**8000**
MTI	**0.55**	**0.56**	**0.58**	**0.56**	**0.6**	**0.61**	**0.68**
	0.95	0.93	0.95	0.95	0.95	0.96	0.98
	0.93	0.90	0.92	0.92	0.93	0.94	0.96
	0.89	0.86	0.89	0.88	0.89	0.92	0.94
	0.85	0.80	0.84	0.84	0.85	0.88	0.92
	0.78	0.72	0.78	0.77	0.78	0.83	0.88
	0.70	0.65	0.70	0.69	0.72	0.76	0.83
	0.61	0.59	0.63	0.61	0.64	0.69	0.77
	0.51	0.53	0.54	0.52	0.57	0.61	0.70
	0.39	0.48	0.46	0.44	0.50	0.52	0.63
	0.26	0.42	0.39	0.39	0.45	0.45	0.56
	0.16	0.34	0.34	0.36	0.38	0.37	0.49
	0.22	0.28	0.28	0.30	0.34	0.31	0.42
	0.26	0.30	0.28	0.18	0.31	0.27	0.38
	0.20	0.22	0.25	0.16	0.29	0.29	0.38
STI	**0.60**						

System 3 City Of Sydney Council Chambers

Northern Terr Parliament							
	125	**250**	**500**	**1000**	**2000**	**4000**	**8000**
MTI	**0.61**	**0.78**	**0.59**	**0.61**	**0.63**	**0.72**	**0.81**
	0.95	1.00	0.95	0.96	0.96	0.98	0.99
	0.94	0.99	0.93	0.94	0.95	0.97	0.99
	0.91	0.99	0.90	0.91	0.92	0.95	0.98
	0.87	0.99	0.86	0.88	0.88	0.93	0.97
	0.82	0.98	0.80	0.82	0.83	0.90	0.95
	0.77	0.96	0.72	0.76	0.78	0.86	0.93
	0.72	0.94	0.64	0.69	0.71	0.82	0.91
	0.66	0.91	0.54	0.61	0.63	0.77	0.87
	0.59	0.86	0.43	0.52	0.55	0.71	0.84
	0.50	0.79	0.36	0.45	0.47	0.66	0.80
	0.35	0.69	0.34	0.40	0.41	0.62	0.76
	0.26	0.53	0.36	0.35	0.37	0.58	0.72
	0.24	0.34	0.30	0.27	0.35	0.52	0.68
	0.29	0.14	0.21	0.27	0.32	0.48	0.62
STI	**0.66**						

System 4 Northern Territory Parliament

NZ Parliament							
	125	**250**	**500**	**1000**	**2000**	**4000**	**8000**
MTI	**0.56**	**0.59**	**0.54**	**0.53**	**0.55**	**0.6**	**0.73**
	0.953	0.952	0.936	0.934	0.937	0.956	0.983
	0.93	0.929	0.905	0.903	0.907	0.934	0.974
	0.897	0.896	0.862	0.861	0.866	0.903	0.96
	0.854	0.853	0.805	0.807	0.813	0.861	0.942
	0.792	0.789	0.723	0.731	0.738	0.799	0.912
	0.726	0.719	0.634	0.651	0.657	0.73	0.876
	0.652	0.638	0.537	0.562	0.569	0.651	0.831
	0.567	0.548	0.444	0.465	0.474	0.564	0.776
	0.466	0.456	0.365	0.37	0.384	0.477	0.712
	0.366	0.384	0.294	0.294	0.325	0.412	0.65
	0.261	0.311	0.243	0.205	0.289	0.358	0.59
	0.101	0.29	0.221	0.158	0.274	0.3	0.547
	0.135	0.309	0.262	0.17	0.27	0.283	0.54
	0.229	0.419	0.305	0.203	0.24	0.324	0.529
STI	**0.58**						

System 5 New Zealand Parliament

Rail Grand Concourse							
	125	**250**	**500**	**1000**	**2000**	**4000**	**8000**
MTI	**0.48**	**0.41**	**0.44**	**0.43**	**0.41**	**0.48**	**0.64**
	0.82	0.74	0.70	0.69	0.70	0.79	0.93
	0.75	0.66	0.63	0.61	0.62	0.72	0.90
	0.68	0.60	0.57	0.53	0.55	0.66	0.86
	0.62	0.53	0.52	0.47	0.48	0.59	0.81
	0.56	0.44	0.46	0.43	0.41	0.51	0.75
	0.50	0.35	0.40	0.40	0.36	0.45	0.70
	0.45	0.27	0.35	0.37	0.30	0.40	0.66
	0.41	0.23	0.32	0.34	0.26	0.35	0.63
	0.37	0.21	0.29	0.32	0.23	0.33	0.61
	0.29	0.24	0.27	0.28	0.22	0.35	0.63
	0.27	0.26	0.31	0.27	0.24	0.41	0.65
	0.15	0.18	0.29	0.30	0.24	0.40	0.64
	0.28	0.23	0.32	0.29	0.26	0.35	0.58
	0.32	0.12	0.23	0.22	0.19	0.26	0.46
STI	**0.46**						

System 6 Grand Concourse at Sydney Central Station

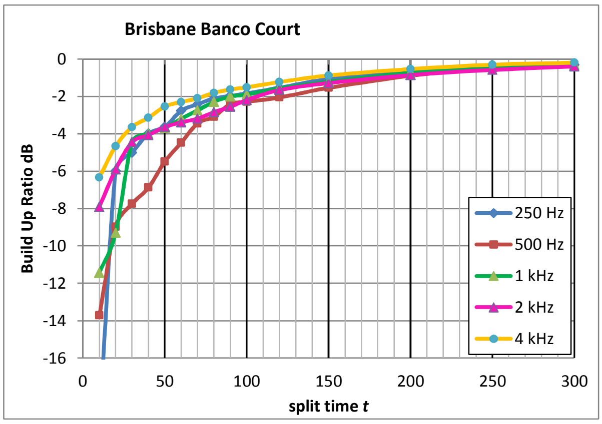
RS19 Sc 2 No eq							
	125	**250**	**500**	**1000**	**2000**	**4000**	**8000**
MTI	**0.35**	**0.4**	**0.44**	**0.47**	**0.5**	**0.56**	**0.61**
	0.701	0.803	0.88	0.9	0.917	0.938	0.959
	0.619	0.735	0.832	0.858	0.88	0.909	0.939
	0.542	0.659	0.776	0.803	0.833	0.87	0.91
	0.455	0.568	0.713	0.737	0.775	0.82	0.871
	0.354	0.456	0.638	0.651	0.701	0.751	0.813
	0.278	0.368	0.568	0.562	0.627	0.68	0.747
	0.232	0.308	0.488	0.466	0.548	0.603	0.67
	0.219	0.251	0.393	0.371	0.464	0.519	0.588
	0.217	0.157	0.3	0.291	0.364	0.432	0.511
	0.189	0.064	0.212	0.215	0.258	0.367	0.457
	0.057	0.085	0.142	0.144	0.181	0.304	0.404
	0.102	0.277	0.096	0.113	0.157	0.258	0.325
	0.1	0.183	0.058	0.133	0.122	0.268	0.264
	0.146	0.171	0.045	0.11	0.125	0.25	0.287
STI	**0.5**						

System 7 Rev Chamber

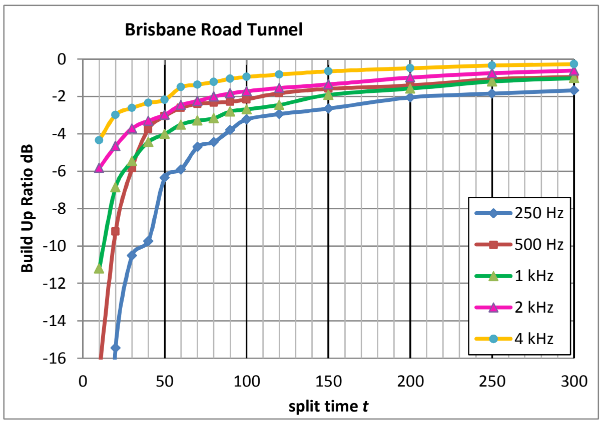
Table 6 STI,MTI and MTF values for seven systems.

	250	**500**	**1000**	**2000**	**4000**
Brisbane Banco Court	1.7	1.5	2.0	2.1	1.6
Brisbane Road Tunnel	5.7	4.3	3.6	3.0	2.4
Council Chamber	1.2	1.2	1.1	1.1	1.0
Nortthern Territory Parliament	1.6	1.1	1.0	1.1	1.0
NZ Parliament	1.3	1.3	1.4	1.3	1.1
Rail Grand Concourse	3.7	3.4	4.3	3.8	3.2
Reverb Chamber no filter	2.4	2.1	1.8	1.7	1.4

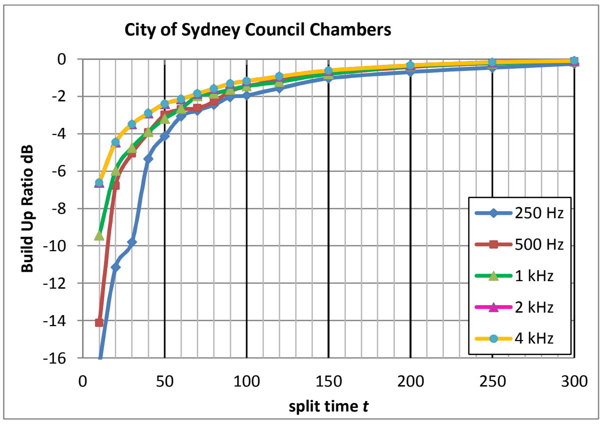
Table 7 Reverberation times measured in late linear section of the decay.



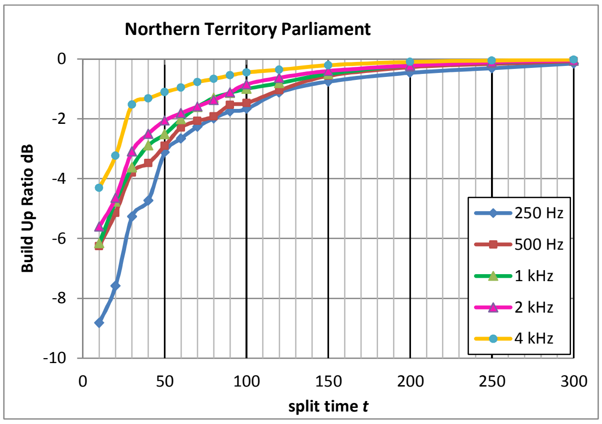
System 1 Brisbane Banco Court



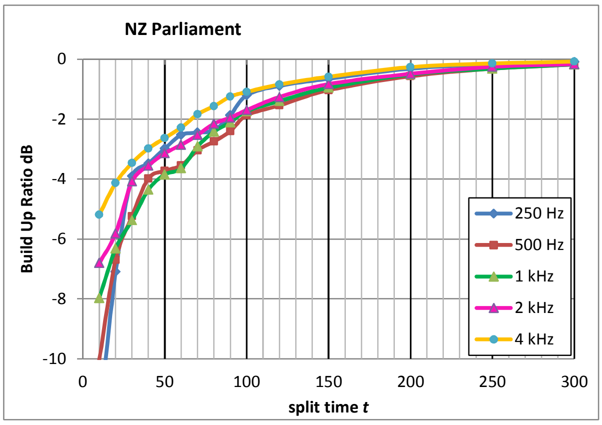
System 2 Brisbane Airport Link Tunnel



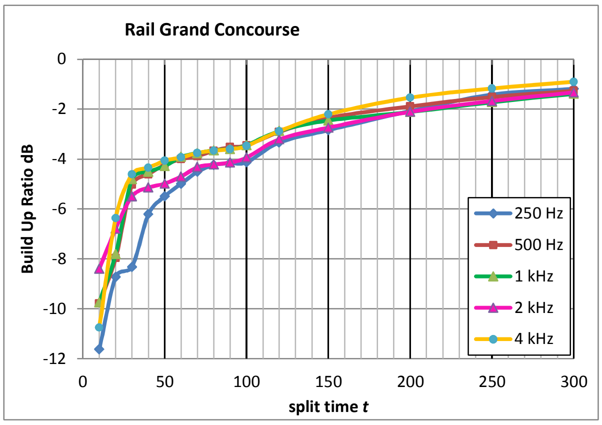
System 3 City Of Sydney Council Chambers



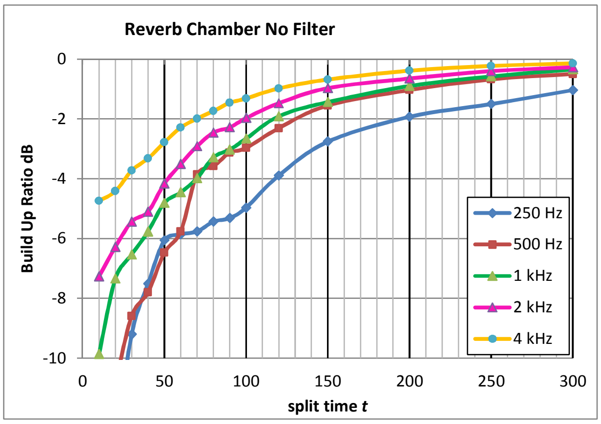
System 4 Northern Territory Parliament



System 5 New Zealand Parliament



System 6 Grand Concourse at Sydney Central Station



System 7 Reverb Chamber

Figure 23 Build-up ratios in octave bands for the seven sound systems.

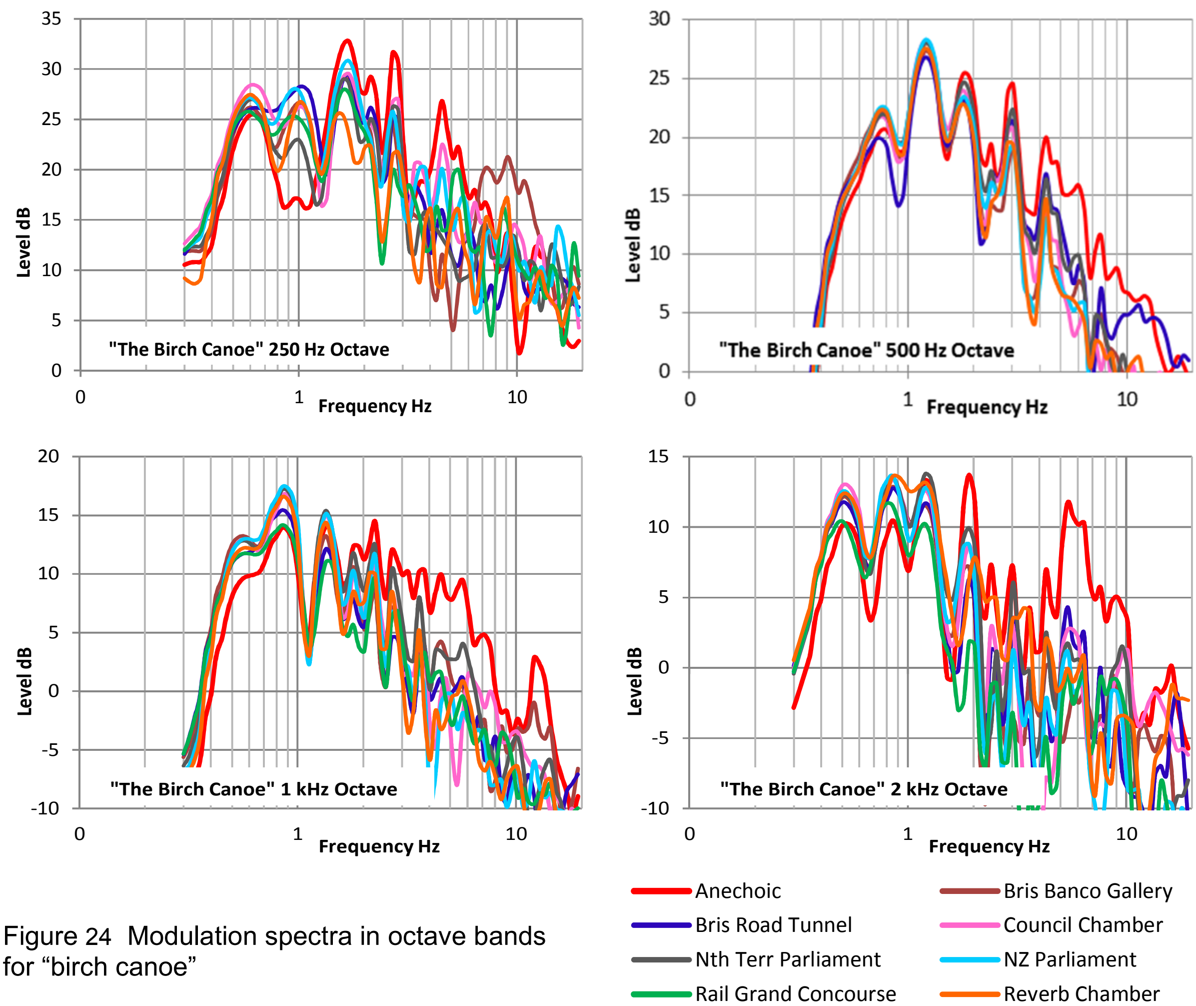


Figure 24 Modulation spectra in octave bands for “birch canoe”

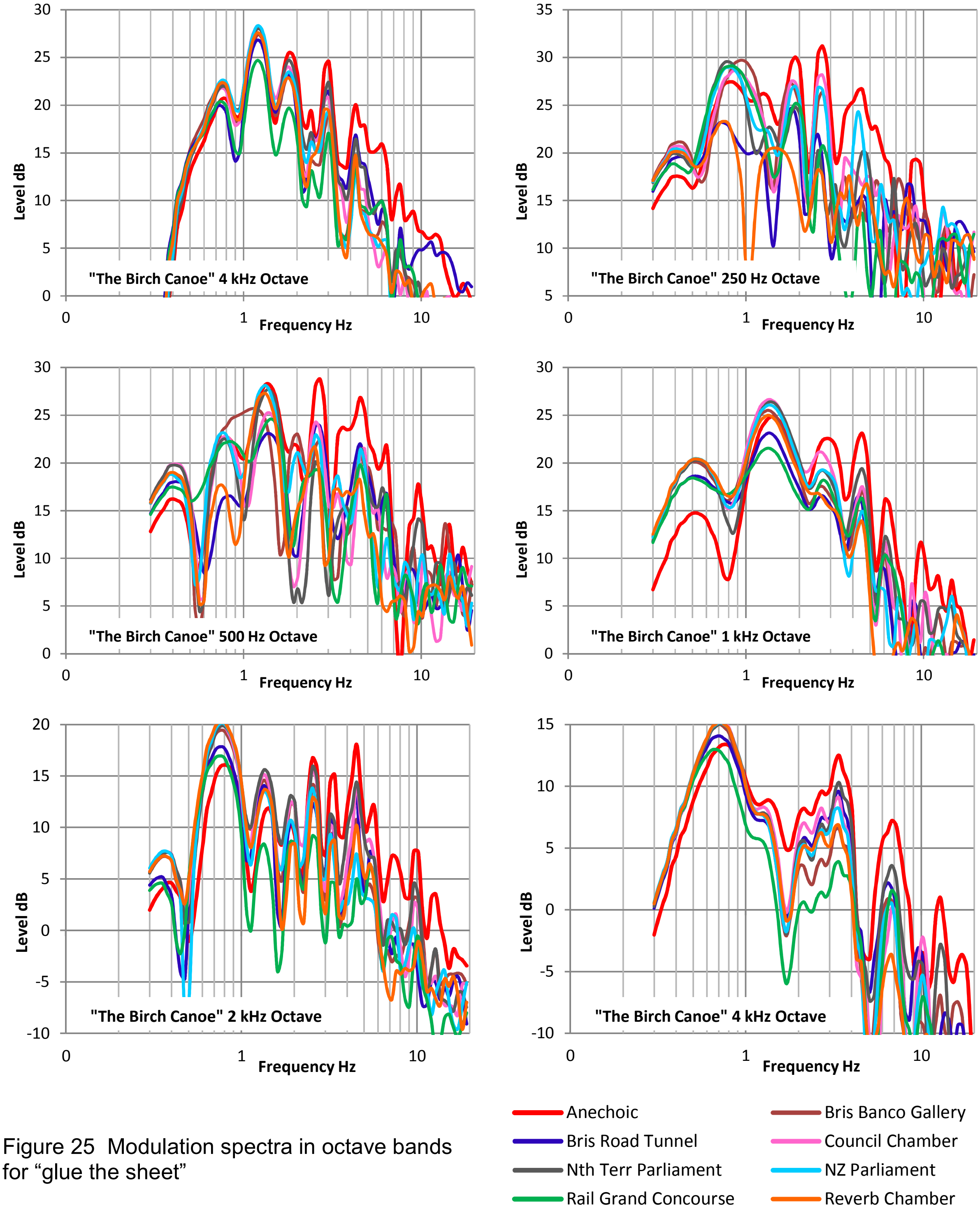


Figure 25 Modulation spectra in octave bands for “glue the sheet”

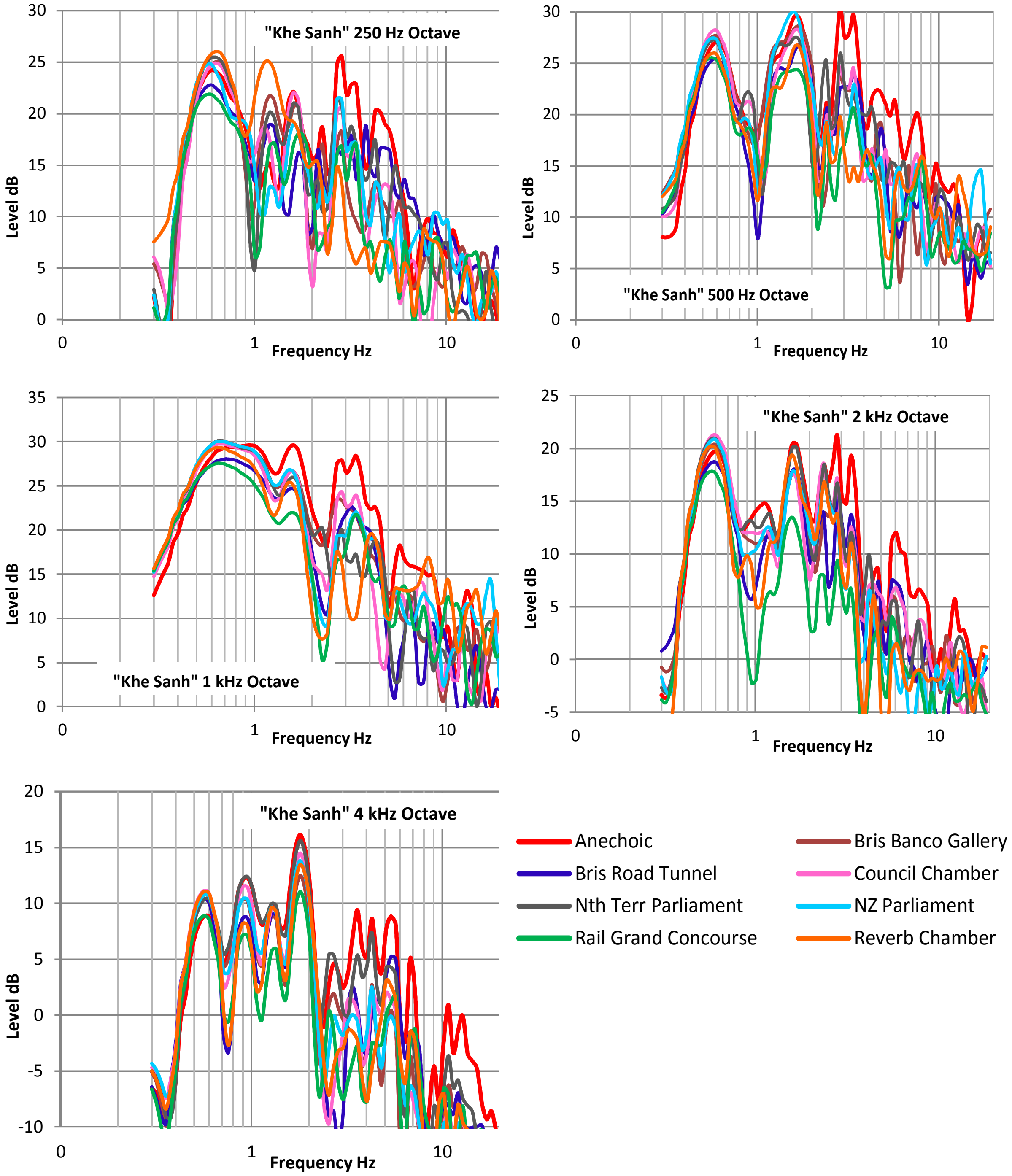


Figure 26 Modulation spectra in octave bands for "Khe Sanh"