PREDICTING INTELLIGIBILITY IN NOISY ROOMS

SM Jelfs Welsh School of Architecture, Cardiff University
M Fedeski Welsh School of Architecture, Cardiff University
JF Culling School of Psychology, Cardiff University

1 INTRODUCTION

Lavandier and Culling[1] have shown that the benefit of binaural hearing (the Binaural Masking Level Difference, BMLD) for understanding speech presented against a localised reverberant background noise can be predicted using equalisation-cancellation theory. Speech reception thresholds can be predicted by combining this with signal-to-noise ratio benefits arising from room colouration and a frequency weighting from the speech intelligibility index. Lavandier and Culling derived the necessary acoustic measurements by convolving Binaural Room Impulse Responses (BRIRs) with noise to create virtual acoustic waveforms, and performing the calculations on these waveforms. In principle, it is computationally more efficient to calculate all the necessary statistics directly from BRIRs. Since BRIRs can be produced by architectural acoustic software, one can predict from plan the intelligibility experienced in a given room for many spatial configurations of sound sources.
This paper replicates Lavandier and Cullings's method, working directly from BRIRs and uses the method to generate intelligibility maps of virtual rooms.

2 REPLICATION OF LAVANDIER & CULLING[1]

2.1 Overview

Lavandier and Culling measured speech reception thresholds (SRTs) for target speech within a localised interfering noise field in sixteen conditions. The target was always anechoic, located directly ahead of the listener, whilst the interferer position and level of reverberation was varied between conditions. The target material was taken from the IEEE recordings of the Harvard sentence list, using speaker DA. The interferer was white noise, filtered with the long-term spectrum of speech. Each condition was assigned a letter, from A to P, ordered in descending levels of interferer interaural coherence and was presented to the participants over headphones in an IAC sound attenuating booth..

Lavandier and Culling modelled their data directly from the stimulus material by first calculating the BMLD[2] and then combining this with the effects of room colouration. This was calculated as the band-by-band levels of the cochlear excitation patterns[3.] As all of their interferers were RMS equalised prior to presentation to the participants any differences in the level for each band was said to occur due to colouration arising from the room. The BMLD and room colouration values were weighted by frequency with respect to the Speech Intelligibility Index[4] band importance functions, and summed to give an overall, frequency dependant value. This value was then shifted in level relative to condition A. This was done for four sections of each target and interferer per condition, each 300ms long, and averaged to give a single value per condition. Both the measured SRTs and the Lavandier and Culling predicted values are shown in figure 1. Note that condition P is not shown in the figure as it was generated by Lavandier and Culling using anechoic independent noise sources for the interferer left and right channels, which could not be modelled by the following BRIR method.

2.2 Modelling Methodology

Whilst Lavandier and Culling were able to model their data with a high level of accuracy, accounting for approximately 94% of the variability in their data, computationally it is a very intensive process to work from the acoustic waveforms. As the interferers are made from a noise source, and therefore have an inherent variability within them, the calculations have to be done a number of time per condition and averaged to give an accurate answer. Due to this it is computationally more efficient to work directly from BRIRs, as the whole impulse response can be analysed as part of the process and be of a smaller size than using the waveforms (Lavandier and Culling required four sections of their stimuli, at 300ms, or 24000 samples of data, whilst the entire impulse responses are often less than 5000 samples at the same sampling frequency, 20000Hz).

The process of modelling the data from the BRIRs is essentially the same as that used by Lavandier and Culling. The BRIRs are generated using |wave[5], which is a UNIX software suite implementing the Peterson[6] image method. Individual BRIRs are generated for each target and interferer condition.

Firstly the BRIRs are passed through a gammatone filterbank[5] from 20Hz to 10kHz, with two bands for each Equivalent Rectangular Bandwidth (ERB)[3]. For each band the average interaural phase (Φ_{AVG}) is calculated from the instantaneous phase (Φ_L and Φ_R) and the instantaneous amplitude (A_L and A_R) using :

$$\Phi_{avg} = \frac{\sum_{i=0}^{N} (\Phi_{L(i)} - \Phi_{R(i)}) \times A_{L(i)} \times A_{R(i)}}{N \times mean} \quad \text{Eq. 1.} \qquad mean = \frac{\sum_{i=0}^{N} (A_{L(i)} \times A_{R(i)})}{N} \quad \text{Eq. 2.}$$

Along with the band-by-band interaural coherence of the interferer (ρ), taken as the maximum of the interferer interaural cross-correlation function, and the band-by-band phase of both target and interferer (Φ_t and Φ_i), the frequency dependent BMLD is calculated using :

$$BMLD = 10\log_{10}\left[\frac{k - \cos(\Phi_t - \Phi_i)}{k - \rho}\right] \quad \text{Eq. 3.}^{2}$$

where $k = (1+\sigma_\varepsilon^2)\exp(\omega_0^2 \sigma_\delta^2)$ and ω_0 = band centre frequency, $\sigma_\varepsilon, \sigma_\delta$ are 0.25 and 0.000105 respectively.

The resultant BMLD figures are weighted as per the SII, and summed to give an overall BMLD value for each condition. As with the Lavandier and Culling method the effect of room colouration was calculated from the cochlear excitation patterns generated from the interferer BRIRs and added to the BMLD figure. This was then shifted relative to condition A.

The results of our prediction method can be seen in figure 1. Our method predicts the variability in the data to approximately 92% accuracy. Whilst this may not be to the level of accuracy as the Lavandier and Culling method there is a considerable decrease in computation time using comparable hardware, from 1 hour 15 minutes, to little more than a minute for the fifteen conditions.

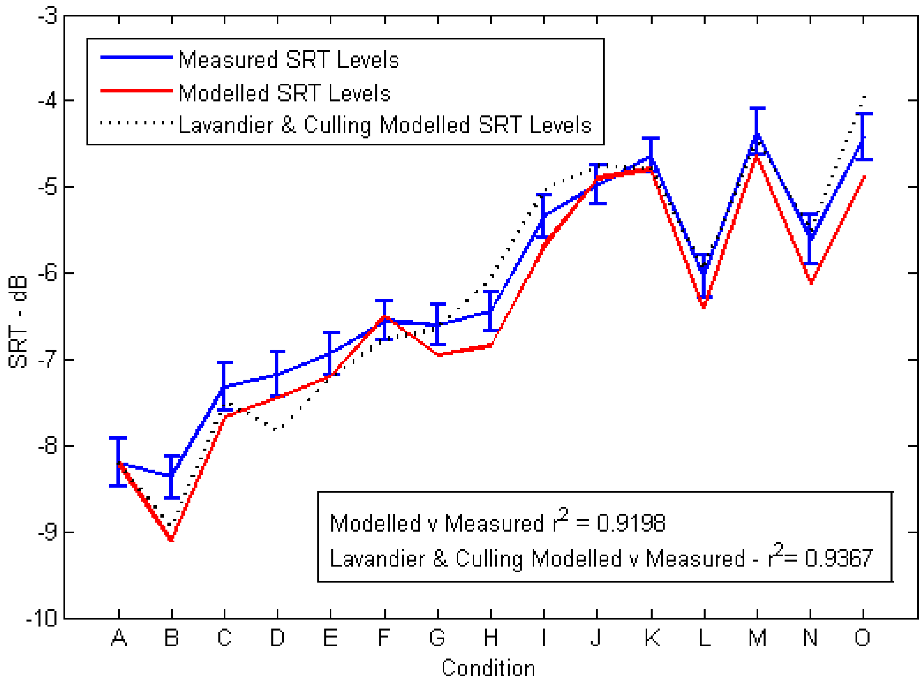

Figure 1 – Measured and modelled values of the SRTs.

3 INTELLIGIBILITY MAPS

The above replication shows that it is possible to predict the level of intelligibility a listener will receive for a given architectural and spatial configuration, when there is a localised target speech source and a localised target noise interfering source. For use in architectural acoustics it would be beneficial to be able to produce maps of rooms plotting the predicted intelligibility levels for any listener location for a given location. This can be done by applying the above method to a regularly sampled set of listener locations for any desired room configuration.

The process used to create the intelligibility maps is the same as that outlined in 2.2. There is one major difference, in that instead of calculating the level differences of the excitation patterns, the signal-to-noise ratio between the target and interferers at each location is calculated. This is because in the Lavandier and Culling experiment the target was always anechoic and in the same position, so the SNR differences between each condition only arose from the differences in the interferer. Calculating the SNR between the target and the interferer will take into account the effects of room colouration on both the target and the interferer.

In this paper a single room design is used. The room is 6.4m x 10m x 2.5m, with a target positioned at 4.8m x 7m x 1.5m. There is either a single interferer at 2m x 6.2m x 1.5m, or two interferers with the second at 3.5m x 2.5m x 1.5m. Both the target and the interferers are modelled as omnidirectional sources. The listening position is moved around the room in a 0.4m grid, always facing the target, and is modelled as two omnidirectional microphones, spaced at 0.25m. The room is either anechoic, with all surfaces having an α value of 1, or reverberant, with all surfaces having an α value of 0.5.

Figure 2 shows the room maps for an anechoic room, with a single interferer. The three panes show the best ear SNR alone, the BMLD alone, and the combined SNR and BMLD respectively, plotted for the location of the listener.. As one would expect, the SNR value decreases as the listener moves away from the target, with only the listener distance having an influence. The BMLD shows that there is a significant increase in perceived SNR for all areas apart from when on the axis between the target and the interferer. The reason for this is due to the lack of interaural phase difference between the target and the interferer. There is an overall increase in the level of intelligibility across the whole map, apart from the area directly behind the interferer with respect to the target. Whilst there is some increase of SNR in this area, the effect of the BMLD is negligible compared to the highly negative best ear SNR alone.

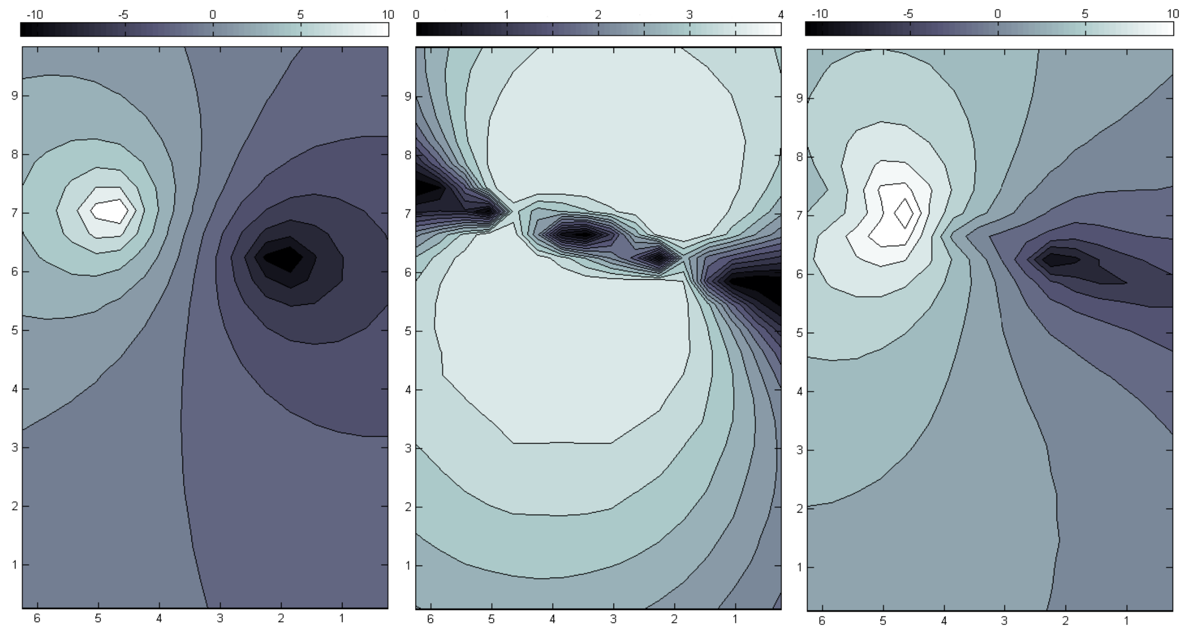

Figure 2. One Interferer, Anechoic
Left, SRT Only. Centre, BMLD Only. Right, SRT and BMLD combined.

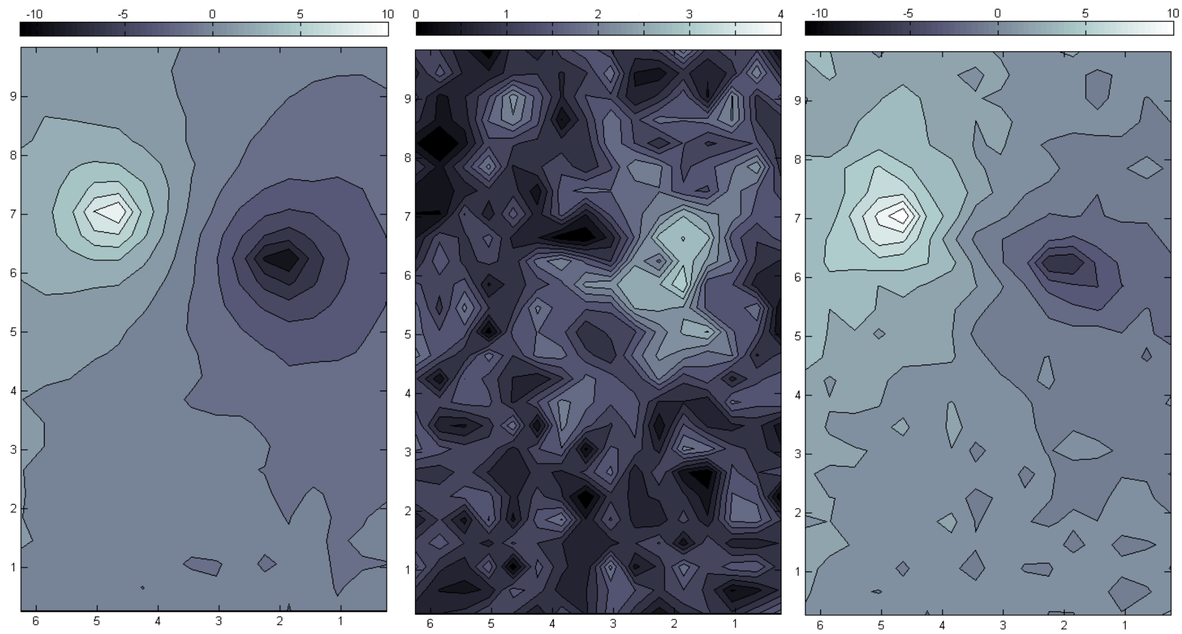

Figure 3. One Interferer, Reverberant
Left, SRT Only. Centre, BMLD Only. Right, SRT and BMLD combined.

Figure 3 shows the same room configuration as that in figure 2, except that the room is reverberant rather than anechoic. It can be seen that the listener will receive approximately the same SNR as we did in the anechoic case, with some areas having a better best ear SNR in the reverberant case than in the anechoic. The effect of the BMLD in reverberation is reduced compared to the anechoic case. This is likely to be primarily due to the effect of reverberation on the interaural coherence of the interferer, as well as to the effect of reverberation on the interaural phase of both target and masker. This would explain why there is a high BMLD value when the listener is close to the interfering sound source, as this will be where the interaural coherence will be at its greatest. Overall the listener is still predicted to receive an improved SNR when taking into account the BMLD compared to just best ear listening.

Figures 4 and 5 are similar to figures 2 and 3 respectively, but include two interfering sound sources rather than one. Again similar results are obtained as those in the single interferer conditions, with high BMLD values in the off-axis areas of the anechoic condition, and high BMLD values close to the interfering noise sources in the reverberant condition.

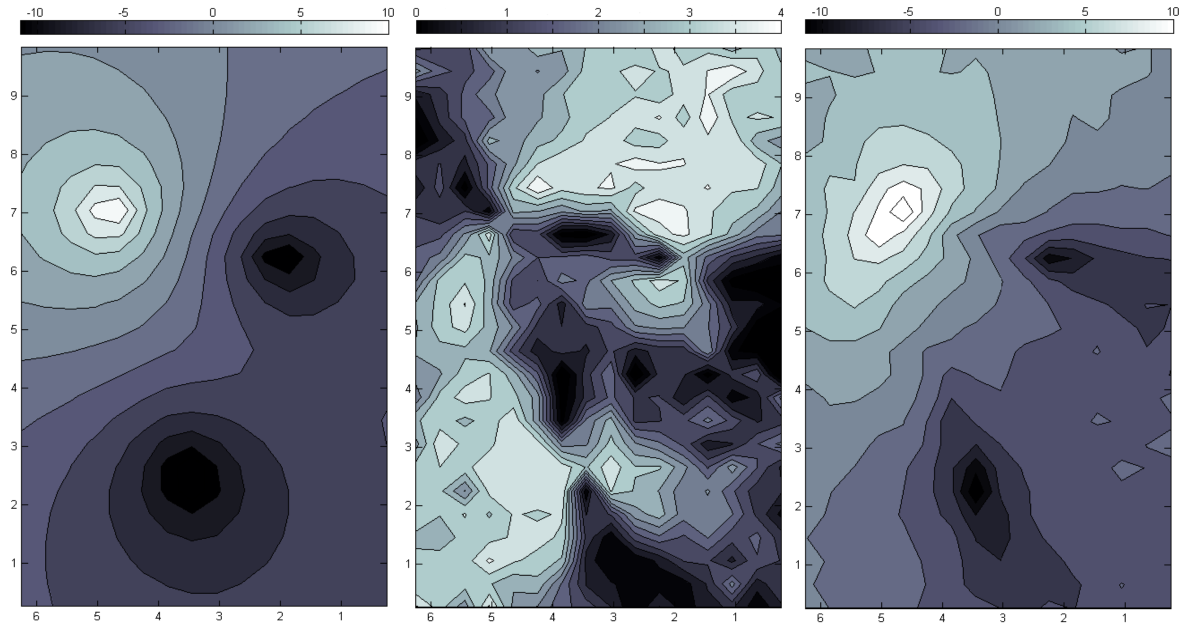

Figure 4. Two Interferers, Anechoic
Left, SRT Only. Centre, BMLD Only. Right, SRT and BMLD combined.

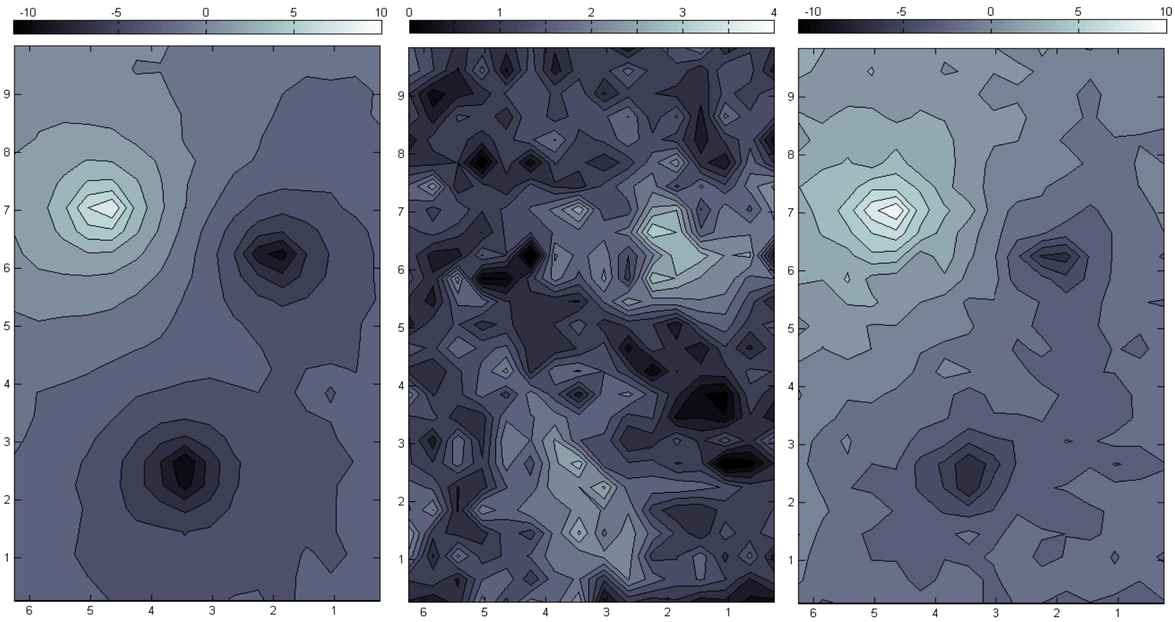

Figure 5. Two Interferers, Reverberant
Left, SRT Only. Centre, BMLD Only. Right, SRT and BMLD combined.

4 CONCLUSIONS

The method for calculating perceived signal-to-noise ratios, and their conversion to speech intelligibility data as outlined within this paper accurately predicts the speech reception threshold levels obtained by Lavandier and Culling. By applying this method to a number of theoretical room configurations and target and interferer configurations it is possible to predict the perceived signal-to-noise ratio at any number of locations, and to show the effects of reverberation on our ability to segregate a target voice from a localised interfering noise source. With further validation of the model to include a reverberant target as well as a reverberant interferer it will be possible to convert the combined BMLD and SNR maps into expected SRT maps.

5 REFERENCES

1. M. Lavandier, and J.F. Culling, Speech segregation in rooms: Importance of interferer interaual coherence. J.Acoust.Soc.Am. 123, 2977-78 (2008)
2. J.F. Culling, M.L. Hawley, and R.Y. Litovsky, The role of head-induced interaural time and level differences in speech reception threshold for multiple interfering sound sources. J.Acoust.Soc.Am. 116, 1057-56 (2004)
3. B.C.J. Moore and B.R. Glasberg, Formulae describing frequency selectivity as a function of frequency and level, and their use in calculating excitation patterns. Hear.Res. 28, 209-225 (1987)
4. American National Standard, Methods for Calculation of the Speech Intelligibility Index. American National Standards Institute. S3.5-1997. (1997)
5. J.F. Culling, Signal-processing software for teaching and research in psychoacoustics under UNIX and X-Windows. Behav.Res.Methods. 28, 376-382 (2006)
6. P.M. Peterson, Simulating the response of multiple microphones to a single acoustic source in a reverberant room. J.Acoust.Soc.Am. 80, 1527-29 (1986)