



## Acoustics 2008

Geelong, Victoria, Australia 24 to 26 November 2008

### Acoustics and Sustainability:

How should acoustics adapt to meet future demands?

# Analysing Room Impulse Responses with Psychoacoustical Algorithms: A Preliminary Study

Doheon Lee and Densil Cabrera

Faculty of Architecture, Design and Planning, University of Sydney, NSW 2006, Australia

## ABSTRACT

Room impulse responses (RIRs) are used to characterise the acoustical conditions inside sound-critical rooms such as auditoria. The analysis of RIRs typically involves octave-band analysis, with parameters such as reverberation time, early decay time, temporal energy ratios and spatial parameters derived from this. This paper explores the potential for applying auditory models for the analysis of RIRs – incorporating auditory temporal integration (and masking), auditory filterbank analysis, and loudness calculation. The purpose of this is to produce analysis results that are closely related to the sound experienced by listeners. A preliminary step for such analysis is to filter RIRs so that their power spectrum is similar to that of typical material that would be listened to in the rooms (e.g. music or speech), and this paper proposes a music filter suitable for orchestral music, derived from long term power spectra of anechoic music recordings. Dynamic loudness analysis of RIRs yield a loudness decay functions that are approximately exponential, which should provide a useful analogy with conventional analysis methods applied to RIRs.

## INTRODUCTION

Room impulse responses (RIRs) are widely used to evaluate acoustical conditions of enclosed places. From the measured RIRs, a number of acoustical parameters are extracted – such as reverberation time, early decay time, strength factor and clarity index etc. – to predict various aspects of the acoustical quality of rooms. Although each parameter is used solely or combined with the other parameters to assess the acoustical qualities of auditoria, these parameters do not perfectly correlate with the actual human perception (for example, of reverberance, loudness or clarity). Furthermore, the details of the perceived reverberation are likely to differ from physical analysis – for example the roughly exponential decay curves obtained from RIRs may not correspond to the perceived decay pattern of the sound.

One issue in the discrepancy between the conventional acoustical parameters and the human perception is that the former does not sufficiently take into account characteristics of the auditory system, such as temporal integration and spectral masking. For instance, while the human auditory system emphasises and de-emphasises spectral or temporal components of sounds, compared to those measured by a microphone in the free-field (Moore *et al.*, 1997), the conventional acoustical parameters do little to reflect these characteristics. Moreover, although bandwidth of a source signal and spectral and temporal distances between source signals strongly influence the perceived loudness, those factors are not carefully considered in the conventional acoustical parameters.

In order to make more accurate predictions of human perception, psychoacoustical approaches have been developed, al-

though these are rarely applied to auditorium acoustics. In the Munich school of psychoacoustics, critical band rate, developed based on the vibrating area of the basilar membrane, is more often used than frequency in modelling perception (Zwicker & Fastl, 1999). Since the critical band rate has logarithmic relationship with frequency above 500 Hz, it may well explain why human perception mostly has a logarithmic response to frequency. For perceived loudness, the units *sones* or *phons* are widely used rather than decibel. Loudness models reflect the complex dependence of loudness on sound pressure level, frequency, bandwidth and time, and when the natural loudness unit is used (*sones*) they provide a ratio scale so that doubling or halving in loudness corresponds to a doubling or halving in units. A simple example of the deviation between loudness and sound pressure level is that a 1 kHz tone at 60 dB is perceived as equally loud to a 50 Hz tone at 85 dB and those two tones have the same loudness level value (*phon*) and the same loudness value (*sones*).

These aspects of the human perception are reflected in the time-varying or dynamic loudness models suggested by Zwicker (1977) and by Chalupper & Fastl (2002). However, analysing RIRs with such models raises some issues. Loudness models are designed for *signal* analysis, whereas a RIR is a *system* analysis. This distinction may seem subtle, since RIRs can be listened to like any audio signal, but the purpose of auditorium acoustics analysis is to assess how music or speech is affected by the room, not how a Dirac delta function sounds in the room. One clear difference between music and an impulse is the spectral distribution of the signal, and this is a theme explored the present paper.

An alternative approach to analysing an RIR in assessing room acoustics is to use anechoic music or speech convolved with the RIR. While this has several advantages, especially in auralisation, the results are biased by the particular selection of anechoic recording, making it more difficult to generalise beyond anechoic samples similar to that selected. The concept of the present paper is to retain the RIR in the analysis, but to adapt it so that it is more suitable for listener-oriented analysis; and also to use analysis methods based on an auditory model – in this case, the dynamic loudness model of Chalupper and Fastl (2002). A preliminary step in this process is to filter the RIR so that its power spectrum is similar to that which would be heard in the room acoustical context: for example, in a concert hall we are concerned with orchestral music; in a speech auditorium with speech. The following section examines how a filter might be developed based on orchestral music.

## POWER SPECTRAL CHARACTERISTICS OF ORCHESTRAL MUSIC

### Previous studies

An impulse excites a room with a ‘white’ power spectral distribution (equal power per linear spectral component), which is very different to the spectral distribution of music or speech, and also is very different to the distribution of filters in auditory spectral analysis (especially above 500 Hz). To put this in perspective, half of the power of a digital impulse is in the highest octave band (e.g., the 16 kHz octave band); three quarters of the power is in the highest two octave bands; and seven eighths of the power in the highest three octave bands. Therefore, listening to or analysing measured RIRs would be profoundly different to listening to or analysing the room response excited by music or speech. In conventional RIR analysis, this issue is ameliorated through extracting room acoustical parameters from octave band analysis (although, interestingly, each octave band retains a +3 dB spectral slope bias between its low and high cut-off frequencies, relative to the logarithmic frequency scale). For this reason, developing and applying useful filters, which are designed based on amplitude spectrum of music or speech, should provide a more appropriate signal for auditory modelling of RIRs. Hence, this section of the paper provides a survey of possible music filters based on measurements of the long term spectral qualities of music. Because the spectral distribution of music varies greatly depending on musical style, only orchestra performances are considered for this study – the focus of this study being on the analysis of impulse responses from concert halls.

With regard to the long term spectrum of music, Sivian *et al.* (1931) conducted a pioneering study of the spectral distribution of live music, and McKnight (1959), Bauer (1970) and Greiner and Eggers (1989) carried out the major studies of the spectral distribution of recorded music. McKnight (1959) investigated the highest peak amplitudes of music using VU meter readings. A number of music samples used in McKnight’s study had been recorded with a single condenser microphone excluding other studio equipment in order to record sounds close to the actual instruments. Bauer (1970) investigated the lowest peak amplitudes of music not exceeded more than 0.1% and 1% of the total length of the music. In Bauer’s study, amplitudes of music were represented relative to 1 kHz level-set tone. The study by Greiner and Eggers (1989) is similar to that of Bauer, except the researchers employed a larger number of percentile divisions; 1%, 5%, 10%, 20%, 30%, 40%, 50%, 60%, 70%, 80%, and 90% of the time. In their study, previously recorded music on compact disc was used for samples. With respect to bandwidth, McKnight used the one-third octave bands while

Bauer and Greiner & Eggers worked with one-octave bands. According to Bauer (1970), a bandwidth of one octave is the best compromise between a good amplitude response and a bandwidth certainty, although it introduces 3 dB errors in peak output amplitudes.

An average of the peak amplitudes of orchestra performances from McKnight’s study (1959) shows a dramatic increase at frequencies over 40 Hz to 100 Hz. Above that they are around -1 dB and -4 dB in VU meter readings before irregular peaks and dips appear in the high frequency range. For Bauer’s study (1970), an average of all the results yield a steep increase at frequencies from 30 Hz to 250 Hz and then the averaged value stays around the highest value until 4 kHz before it dramatically decreases. The difference between an average for 0.1% of the time and 1% of the time are almost constant over all the frequencies of interest by -3 dB. As would be expected, the average for 1% of the time has lower values. Since Greiner and Eggers (1989) worked with a large number of time divisions, amplitudes of orchestra performances only for 1%, 50% and 90% of the time are averaged to review their work on this paper. An averaged amplitude for 1% of the time mostly stays around -10 dB relative to 2 volts per octave band and those for 50% of the time stays around -20 dB at frequencies from 63 Hz to 2 kHz before a obvious decrease from 2 kHz to 16 kHz. For 90% of the time, a steep increase and decrease is found at frequencies from 32.5 Hz to 250 Hz and from 2 kHz to 16 kHz, respectively. Figure 1 shows the averages of the spectral amplitudes of orchestra performances from the three previous studies, except all the eight samples used in the Bauer’s study are averaged as the researcher does not provide details of performance styles of the samples.

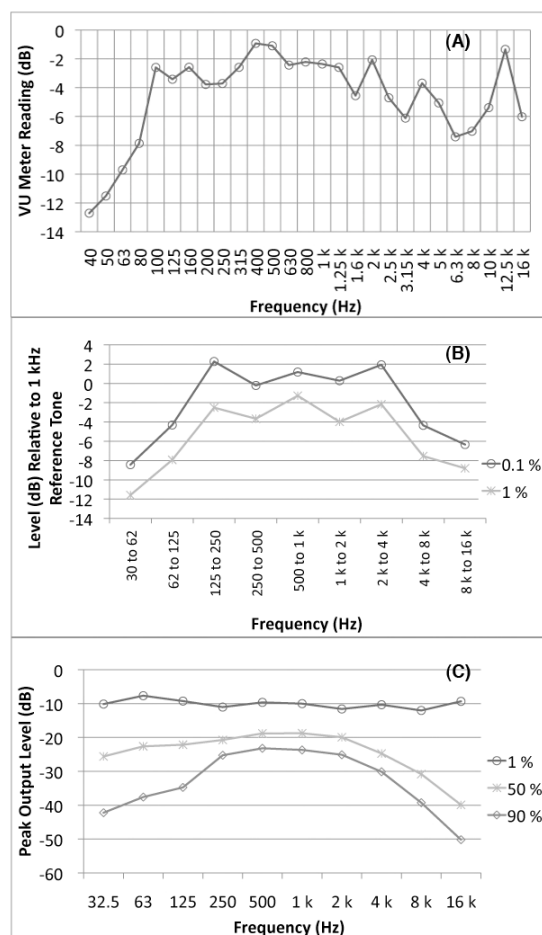


Figure 1. The averages of musical spectrum from the previous studies of; (A) McKnight, (B) Bauer and (C) Greiner and Eggers.

### Analysis of a selection of anechoic recordings

The power spectral characteristics of orchestral recordings in auditoria confound the spectral characteristics of the signal (orchestra) and system (room). Hence the previous studies cited are analyses of both the musical sources and the acoustic environments in which the recordings were made. In developing a weighting filter relevant to the orchestral signal alone it would be better to use data from anechoic recordings. In this section of the paper we examine seven anechoic recordings of music from Denon Test CD No. 2. Details of seven performances are given below. The first four performances last for around 30 seconds while the rest for around 90 seconds.

1. Bruckner, Symphony No. 4 in E-flat minor, *Romantic* (excerpt from first movement)
2. Handel, *Water Music* (Harty edition, excerpt from sixth piece. Allegro Deciso).
3. Mozart, *The Marriage of Figaro*, KV492 (excerpt from Overture)
4. Shostakovich, Symphony No. 5 in D minor, Op 47 (excerpt from first movement)
5. Johann and Josef Strauss, *Pizzicato Polka*
6. Bizet, *L'Arlésienne* Suite No.2 Minuet
7. Glinka, *Ruslan and Ludmilla* (excerpt from Overture)

Samples 1 and 4 contain brass and string sound playing *forte* and *fortissimo*. Sample 7 also has a loud brass and string part with an addition of timpani. Samples 5 and 6 are softer than the other samples. For sample 5, a small number of strings are played mostly at *pianissimo* and *piano*. Similarly to the sample 5, one flute and one piano are played at *mezzo piano* and *piano* for the most part of sample 6. For sample 2, a large number of strings are played at around *mezzo forte* and sample 3 has the widest dynamic range (from *pianissimo* to *forte* of strings and brasses) of the seven samples.

According to the original manufacturer of Denon Test CDs, all the performances were sampled in anechoic conditions, which met recommendations specified in ISO-3745 for anechoic chambers (Anechoic Orchestral Music Recordings, 1995). With respect to microphone positions, two omnidirectional microphones at positions above the head of conductor are mainly used. To sample instruments sounding weak, a number of spot microphones are used. Time differences, which occurred due to different microphone positions, are compensated in the recording process (Anechoic Orchestral Music Recordings, 1995).

All seven samples used in this study have a pair of stereophonic channels. Therefore, squared amplitudes of left and right channels were added to provide single values as done by the previous studies. The samples were fast-Fourier transformed with a window size of 65536 and the obtained FFT component power values were summed corresponding to the 1/3-octave bands before converting to level relative to full scale. The obtained single values represent  $L_{eq}$  at each 1/3-octave band.

Figure 2 shows amplitudes of all the seven samples. As seen in the figure, amplitudes of samples 1, 2, 3, 4 and 7 are between -20 dB to -40 dB for most one-third octave bands, and samples 5 and 6 are below -40 dB for most one-third octave bands. Differences between the two groups become greater as frequencies increase. Samples 1, 2, 3, 4 and 7 yield a similar pattern while samples 5 and 6 are distinctive.

Figure 3 plots the average of all the seven samples and of samples 5 & 6. Music Spectrum Loud refers to the former (since power averaging means that the spectra of samples 5

and 6 have little influence on the result) and Music Spectrum Soft to the latter. The two graphs shown in Figure 3 would be appropriate suggestions for orchestra performances playing at two different dynamics; soft and loud.

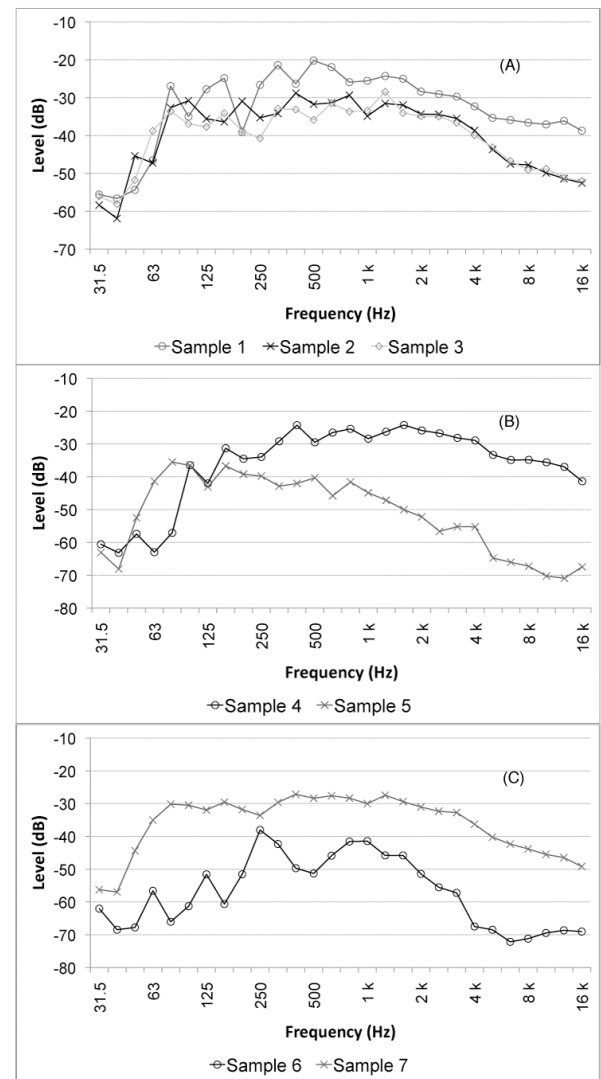


Figure 2. The spectral amplitudes of; (A) the samples 1, 2 & 3, (B) samples 4 & 5, and (C) samples 6 & 7 as a function of the one-octave bands.

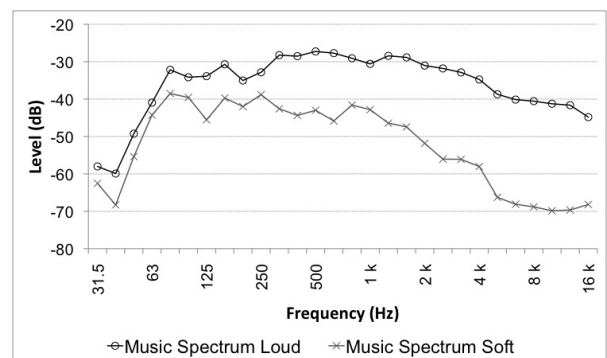


Figure 3. The averages of spectral amplitudes of all the samples and of samples 5 & 6 as a function of the one-octave bands.

Figure 4 shows how a music filter may be derived from the power spectrum of music. As mentioned previously, RIRs are measured using an initial stimulus possessing a 'white' spectrum which has a spectral slope of +3 dB per octave band or +1 dB per one-third octave band. The task of a music filter is to convert a 'white' spectrum to a music spectrum, and so it

is the product of the desired music spectrum and a ‘pink’ spectrum. Hence, in Figure 4, the Music Spectrum Loud is used to derive a music filter by multiplying with a -3 dB/octave function (pink), and the resulting filter function is smoothed. Note that, as a simple alternative to applying a music filter, a pink filter could be used to bring an impulse response somewhat closer to typical listening conditions, although it does deviate substantially from the music filter at the extremes of the frequency range.

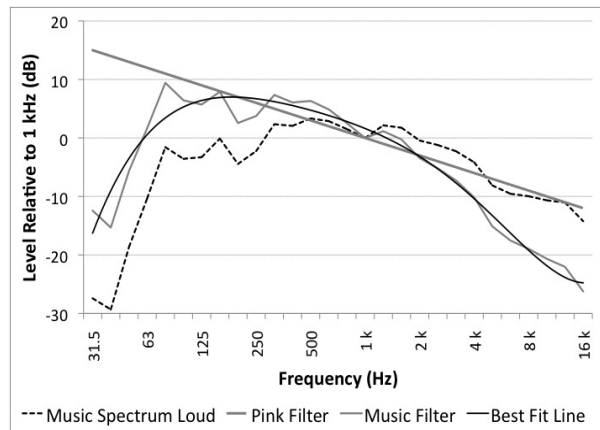


Figure 4. Pink filter and music filter derived from Music Spectrum Loud.

### APPLICATION OF A MUSIC FILTER AND PINK FILTER TO MEASURED RIRs

The music filter (best-fit line) and pink filter shown in Figure 4 were applied to measured RIRs from concert halls. The RIRs were measured in two halls, seating 700 and 2800, located in Rome, Italy. Those RIRs are named “Small-Close”, “Large-Close” and “Large-Distant”. Small-Close refers to the RIR measured in the hall with 700 seats at a receiver position close to the stage. Large-Close represents the RIR measured in the hall with 2800 seats at a receiver position close to the stage, and Large-Distant is for the RIR measured at a relatively long distance from the stage in the same hall for Large-Close. The RIRs were measured with fixed system gain, and so vary in level according to the acoustic conditions. In our analysis, we gave the Small-Close RIR has an instantaneous peak sound pressure level of about 85 dB, as seen in Figure 7 – and the cumulative power sum of the RIR at this gain would be substantially higher. Hence, we use the filter derived from loud (rather than soft) music in this analysis.

To provide a rough idea of acoustical conditions of those halls, spatially averaged reverberation times for the two halls are presented in Figure 5. Two receiver positions and three receiver positions were chosen in terms of different distances from the performing entity for the hall with 700 seats and 2800 seats, respectively. Those positions include the receiver positions for the three RIRs.

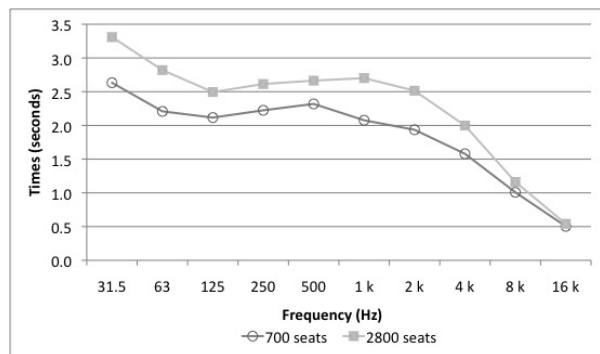


Figure 5. Octave band reverberation time of the two halls.

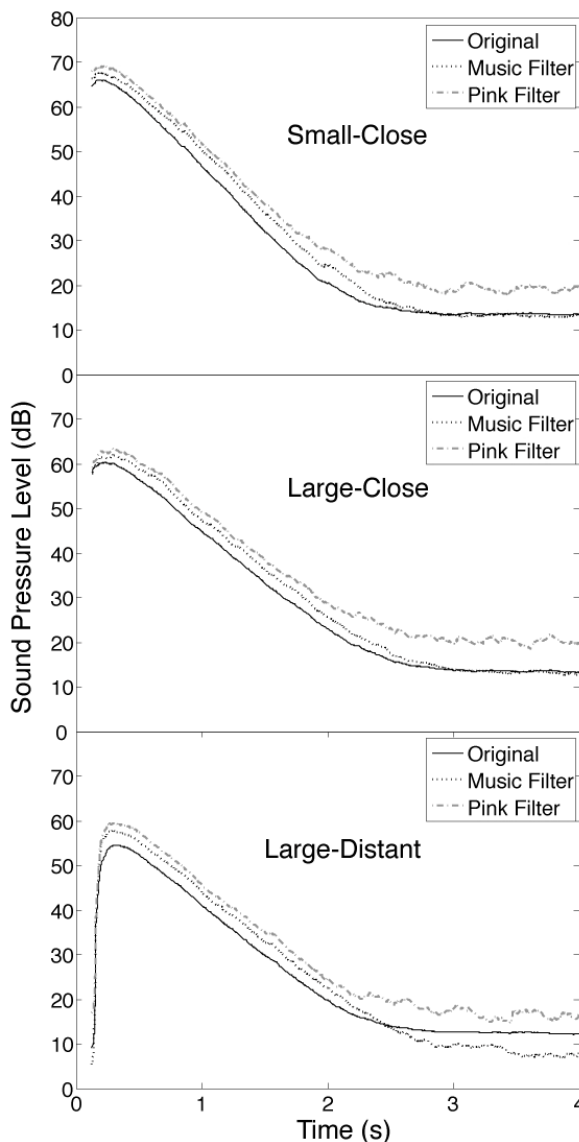


Figure 6. The sound pressure level (dB) of the three RIRs with an application of the two filters, as a function of time (seconds).

Figure 6 shows the sound pressure level of the three RIRs with the application of the two filters (music filter and pink filter) and without the filters, as a function of time. These sound pressure levels use exponential temporal integration with a 125 ms time constant (equivalent to the ‘fast’ setting of a sound level meter). The “Original” refers to the unfiltered RIR. As seen in the figure, as it would be expected, Small-Close shows the greatest sound pressure level, and Large-Distant the smallest sound pressure level. For all RIRs, the application of filters produces an overall gain. The gains produced by the pink filter are generally within 5 dB to 8 dB at the start of the decay curves, but increase towards the noise floor at the tail of the impulse response recordings. The gains produced by the music filter are within 5 dB at the start of the decay curves and decrease towards the noise floor.

Although the decay curves of Figure 6 are generated using the commonly used ‘fast’ integration time of 125 ms, it is unusual to analyse RIRs with this type of integration. Hence, by way of comparison, Figure 7 shows the decay curve of the small-close RIR compared to its instantaneous sound pressure level (derived from the Hilbert transform). This reveals the extent to which the decay curve has been smoothed by ‘fast’ integration, as well as the contrast between instantaneous and integrated sound pressure level at the start of the RIR. ‘Fast’ integration is intended to emulate auditory tem-

poral integration, and so makes an interesting comparison with the results of dynamic loudness modelling.

The reason why the vertical axes of Figures 6 and 7 are in sound pressure level units, rather than level with respect to some arbitrary reference (such as full scale amplitude of the medium) is that loudness modelling requires an assumption to be made about the sound pressure level received by the listener. Loudness models are non-linear, and will only yield useful results for reasonable listening levels. The gain that yielded the sound pressure levels selected for this analysis was chosen because these are of a similar order to levels that might be experienced in an auditorium.

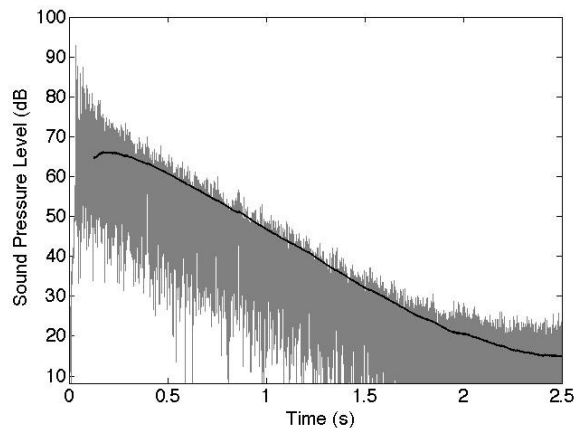


Figure 7. The sound pressure level of the Small-Close RIR, comparing instantaneous level (grey) with fast integration (black).

Figure 8 plots the modelled loudness of the three RIRs with an application of the two filters and without the filters. The model used is Chalupper and Fastl's (2002) dynamic loudness model, which is implemented in the computer program PsySound3 (Cabrera et al. 2008). As seen in the figure, all the initial and filtered RIRs in Small-Close show greater loudness than those in Large-Close and Large-Distant. In contrast to the sound pressure level comparisons shown in Figure 6, the loudness of the original and pink filtered signals yield similar results, while those of the music filtered analysis are slightly but obviously quieter. The fine temporal structure of the decay curves is similar, regardless of the application of a filter. Greater detail in the fine temporal structure is evident in the loudness decays than in the sound pressure level decays of Figure 6.

A striking feature of the loudness decay curves in Figure 8 is that they appear to exhibit approximately exponential decay, like the signal's decay curve prior to transformation to decibels. However, closer examination shows that while the first part of the loudness decay curves is approximately exponential, this is followed by faster loudness decay. Figure 9 compares the exponential decay rates (by using a logarithmic value scale) for the small-close RIR. In addition to showing the pressure and pressure-squared decays, it shows Stevens' power law (Stevens 1955) for loudness (where loudness is proportional to pressure raised to the power of 0.6). The comparison shows that the modelled loudness decay rate is similar to that expected from Stevens' power law, but with a faster decay rate once low sound pressure levels are encountered. This faster decay rate would be expected from steady state loudness theory from the fact that the fixed loudness exponent of 0.6 only applies to sounds of moderate loudness (for sound pressure levels roughly between 40 and 80 dB). The consistency of the modelled decay with steady state loudness theory suggests that temporal integration (and tem-

poral masking) is having little effect on the coarse structure of the loudness decay.

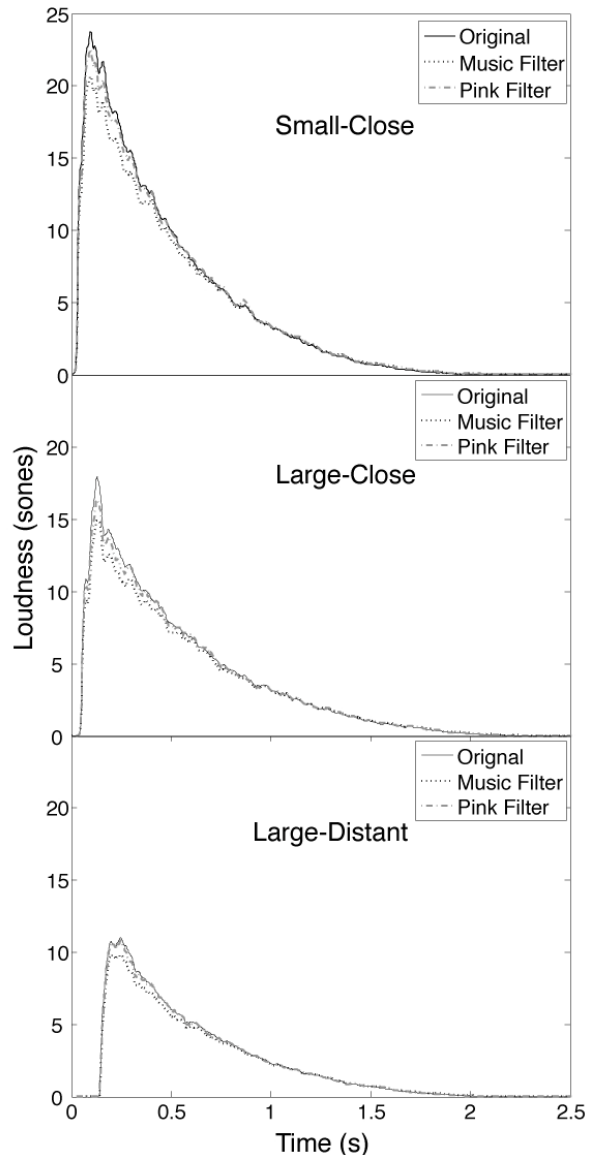


Figure 8. Loudness (sones) of the three RIRs with an application of the two filters, as a function of time.

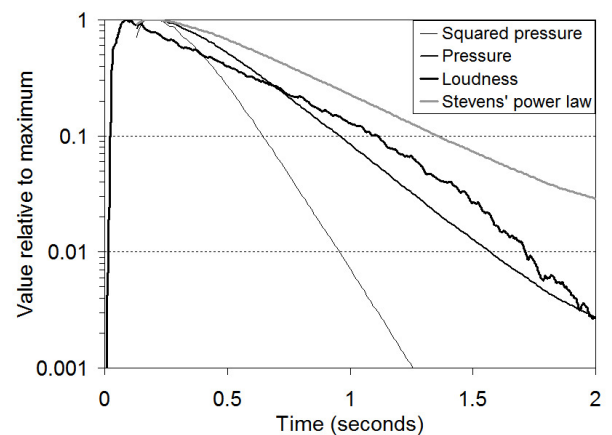


Figure 9. Comparison of decay rates on a logarithmic scale, for the small-close RIR (without music filtering). Normalised A-weighted squared pressure and pressure are shown, together with loudness, and the application of Stevens' power law to the pressure decay curve.

Figure 10 shows the averaged specific loudness (Sones/Bark) as a function of critical band rate (Bark). The specific loudness pattern can be thought of as a psychoacoustical spectrum, where values are the loudness attributable to the critical band rate units. As seen in the figure, Original yields the greatest loudness at critical band rates from approximately 11 Bark to 24 Bark and the music filtered RIRs attain the highest loudness at critical band rates from 3 Bark to 11 Bark. For the pink filter, a substantial increase in specific loudness below 3 Bark appears, which is probably due to the greater loudness growth function in the low frequency range (where the loudness exponent becomes greater than 0.6) and this is partly why the pink filter has the at the greatest loudness as seen in Figure 8. The charts show the importance of the peak in the outer ear transfer function above 15 Bark.

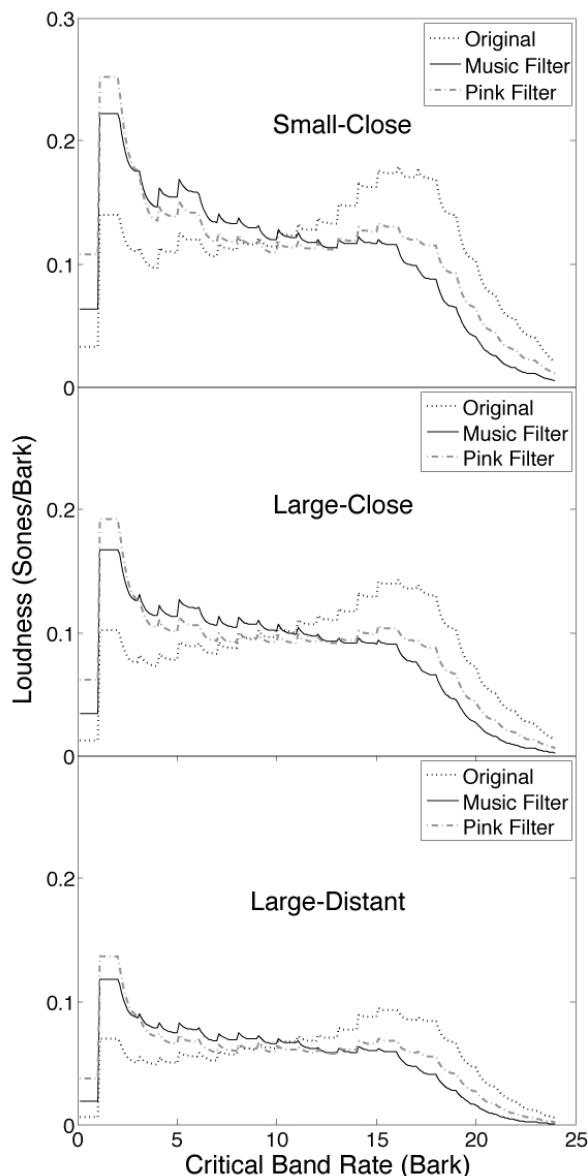


Figure 10. Average specific loudness (Sones/Bark) of the three RIRs with the two filters, as a function of critical band rate (Bark)

## DISCUSSION

This paper makes a preliminary examination of some possibilities in applying psychoacoustical models to RIR analysis. It has examined two aspects of this: (i) the application of spectral weighting to bring a RIR closer to the long term spectrum of orchestral music; and (ii) the application of a dynamic loudness model to filtered (and unfiltered) RIRs

from concert auditoria. The results indicate the type and extent of differences that might occur with applying these approaches to the analysis of RIRs from concert auditoria.

Perhaps there is no correct solution to the design of music filters because the spectral characteristics of music vary so much. The filters explored in this paper are taken as possible solutions, and are used by way of example. If more defensible music filters were to be derived for orchestral music, much more extensive anechoic recordings would be needed. Nevertheless, even if an ideal representative spectrum were derived, other factors – such as the directivity of sound radiated from the source – come in to play (presumably the direct sound is heard from the front of the orchestra, while the reverberation is heard from sound averaged over all radiation directions). On the other hand, auditory analysis of RIRs without applying a filter makes little sense because of the ‘white’ spectral bias of the excitation signal. The compromise solution, of a ‘pink’ filter provides some potential because of its simplicity, although it results in excessive energy in the very low and very high frequency ranges.

To use loudness models well, the signal should be calibrated to a realistic listening level. This could be done quite precisely if the sound power level of relevant music was known, and the strength factor associated with each RIR was known. For the present analysis, neither of these pieces of information is available, but nevertheless an approximate assumption can be made about listening level. Nevertheless, the problem remains that loudness models are non-linear with respect to sound pressure level (loudness growth and upward masking patterns change substantially with sound pressure level). An alternative solution to this might be to simplify the loudness model, to remove the non-linear gain dependence. However, it is not clear how this might be done while preserving a reasonable performance of the model over the full decay of a RIR. A more subtle aspect of this problem is that the dynamic characteristics of RIRs are very different to those of music, meaning that a dynamic loudness model will respond differently to RIRs than to music in auditoria. This, at least, will impinge on the process of applying realistic gain, and is likely also to be important in interpreting analysis results.

The results of applying a dynamic loudness model suggest that the temporal resolution of RIR decays may be finer than that produced by ‘fast’ temporal integration. The loudness decay function is exponential at first, and is consistent with the loudness that might be calculated from a steady state loudness model (although this is might not be so for very short reverberation times). The fact that the loudness decay has a relatively simple relationship to the sound pressure level could be helpful in simplifying the analysis model.

Analysis done subsequent to the preparation of this paper has explored the effect of gain on the loudness modelling of RIRs – and has shown that the RIR slope is gentler with increased gain. Hence the parallel slopes of loudness decay and Stevens’ power law decay in Figure 9 will not necessarily be found, although the loudness decay will still tend to be exponential in its first section. Furthermore, the application of a filter, such as the music filter, sensitises the decay to different parts of the spectrum, which will affect the slope if reverberation time varies with frequency.

The RIRs analysed here were made using an omnidirectional microphone. However, a more detailed approach could be taken using a binaural RIR, and possibly a binaural loudness model. The binaural summation procedure proposed by Sivonen and Ellermeier (2008) has some potential for this if a single time-varying specific loudness pattern is desired. That model performs binaural signal summation prior to input into

an arbitrary loudness model (and so could be applied to Chalupper and Fastl's dynamic model). Another approach could be to use the binaural summation procedure proposed by Moore and Glasberg (2007), which may be applied to the output of Glasberg and Moore's (2002) time-varying loudness model applied to each ear. That would allow an assessment of the loudness attributable to each ear, although since the analysis does not include phase information, it would not provide sufficient data for detailed auditory spatial modelling. Conventionally, binaural RIRs are analysed using the interaural cross correlation (and not the interaural level differences), but perhaps there is some prospect for integrating these approaches.

One question that arises from this approach is whether a loudness-based analysis of RIRs is in fact a good representation of auditory perception, and in a broader sense, low level cognition. Partly, this is to do with the question of whether loudness models are accurate. Beyond this, it might be that the important attributes of RIRs are not just related to loudness, but more to some measure of salience. For example, although there is a very dramatic decline in loudness as a RIR decays, a listener's attention may be drawn into listening to the details in the quieter parts of the reverberant tail.

## CONCLUSION

A psychoacoustical approach to RIR analysis has some possibilities, but there are considerable challenges to overcome in developing a practically useful analysis method. While such methods may draw on pre-existing psychoacoustical models, ultimately they should be validated and refined using subjective data – as pre-existing loudness models are derived from subjective data very different from RIRs. The reduction of single number parameters from loudness decay functions has not been explored here, but the fact that the decay functions are relatively simple suggests that this should be feasible. Again, the parameters would need to be based on subjective data (for example, assessments of reverberation period, overall loudness, clarity, and even spatial attributes for music sources convolved with RIRs). The work in this paper is the first step of a larger research project.

## REFERENCES

- Anechoic Orchestral Music Recordings* 1995. [CD-ROM]. Japan: Nippon Columbia.
- Bauer, B.B. 1970, "Octave-band spectral distribution of recorded music," *J. Audio Eng. Soc.* 18, 165-172.
- Cabrera, D., Ferguson, S., Rizwi, F. and Schubert, E. 2008a, "PsySound3: a program for the analysis of sound recordings," *Acoustics 2008*, Paris, France.
- Chalupper, J. & Fastl, H. 2002, "Dynamic Loudness Model (DLM) for normal and hearing-impaired listeners," *Acustica* 88, 378-386.
- Denon Professional Test CDs* [CD-ROM]. Japan: Nippon Columbia.
- Glasberg, B. R. and Moore, B.C.J. 2002, "A model of loudness applicable to time-varying sounds," *J. Audio Eng. Soc.* 50, 331-342.
- Greiner, R. A. & Eggers, J. 1989 "The spectral amplitude distribution of selected compact discs," *J. Audio Eng. Soc.* 37, 246-275.
- Moore, B. C. J., Glasberg, B.R., & Baer, T. 1997. "A model for the prediction of thresholds, loudness, and partial loudness," *J. Audio Eng. Soc.*, 45(4), 224-240.
- Moore, B. C. J. and Glasberg, B.R. 2007, "Modeling binaural loudness," *J. Acoust. Soc. Am.* 121(3), 1604-1612.
- McKnight, J. G. 1959 "The distribution of peak energy in recorded music, and its relation to magnetic recording system," *J. Audio Eng. Soc.* 7, 65-80.
- Sivian, L. J., Dunn, H. K. & White, S. D. 1931 "Absolute amplitudes and spectra of certain musical instruments and orchestras," *J. Acoust.Soc.Am.*, 2, 330-371.
- Sivonen, V. P. and Ellermeier, W. 2008 "Binaural loudness for artificial-head measurements in directional sound fields," *J. Audio Eng. Soc.* 56(6) 452-461.
- Stevens, S. S. 1955, "The measurement of loudness," *J. Acoust. Soc. Am.* 27(5) 815-829.
- Zwicker, E. 1977 "Procedure for calculating loudness of temporally variable sounds," *J. Acoust. Soc. Am.* 62, 675-682.
- Zwicker, E. & Fastl, H. 1999 *Psychoacoustics: facts and models*, Berlin; New York, Springer.