Author(s)

First Name	Middle Name	Surname	Role	Email	SMPTE Member?
Glenn	Anthony	Leembruggen		glenn@acousticdirections.com	No

Affiliation

Organization	Address	Country
Acoustic Directions, ICE Design	6 Cary St Leichhardt	Australia

Author(s) – repeat Author and Affiliation boxes as needed--

First Name	Middle Name	Surname	Role	Email
Philip		Newell	SMPTE Member? No	philiprnewell@gmail.com

Affiliation

Organization	Address	Country
Consultant	Moana	Spain

Author

First Name	Middle Name	Surname	Role	Email
Joules		Newell	SMPTE Member? No	joules@newellacousticengineering.com

Affiliation

Organization	Address	Country
Newell Acoustic Engineering Company		UK

Author(s) – repeat Author and Affiliation boxes as needed--

First Name	Middle Name	Surname	Role	Email
David	(or initial)	Gilfillan	SMPTE Member? No	davidg@icedesign.net

Affiliation

Organization	Address	Country
ICE Design and Gilfillan SoundWork,	6 Cary St Leichhardt	Australia

Author(s) – repeat Author and Affiliation boxes as needed--

First Name	Middle Name	Surname	Role	Email
Dr Keith	(or initial)	Holland	SMPTE Member? No	krh@isvr.soton.ac.uk

Affiliation

Organization	Address	Country
University of Southampton		UK

Author(s) – repeat Author and Affiliation boxes as needed--

First Name	Middle Name	Surname	Role	Email
Brian	(or initial)	McCarty	SMPTE Member? Yes	bmccarty@coralseastudios.com

Affiliation

Organization	Address	Country
Coral Sea Studios		Australia

Publication Information

Pub ID	Pub Date
2010 Fall Conference Paper	

SMPTE Meeting Presentation

Is the X Curve Damaging Our Enjoyment of Cinema?

Glenn Leembruggen,
Principal Acoustic Directions, ICE Design Australia, Glenn@acousticdirections.com

Philip Newell
Consultant, philiprnewell@gmail.com

Joules Newell
Consultant Newell Acoustic Engineering Company, joules@newellacousticengineering.com

David Gilfillan,
Principal ICE Design and Gilfillan SoundWork, davidg@icedesign.net

Dr Keith Holland,
Senior Lecturer, University of Southampton, krh@isvr.soton.ac.uk

Brian McCarty
Coral Sea Studios, bmccarty@coralseastudios.com

Written for presentation at the

[SMPTE 2011 Australia]

Abstract.

Complaints about poor speech intelligibility and music fidelity in cinema sound tracks persist despite advancements in sound reproduction technology and accreditation of cinema sound systems. The X curve for the equalisation of cinema sound systems has been used for many years and is a foundational element for setting up the tonal balance and achieving accreditation. This paper challenges the acoustic theory behind the X curve, using analysis of acoustic and electro-acoustic measurements made in cinemas and in other spaces where high-fidelity and high intelligibility is required. Both the temporal and tonal developments of the sound field in a number of rooms are examined in detail and related to temporal properties of speech and music and human's perception of tonality.

Keywords. X-curve, equalization, room gain, reverberant build up, room gain, early and late sound fields, speech intelligibility, upward masking , frequency response measurements

Introduction

Complaints about poor speech intelligibility and music fidelity in cinema sound tracks persist despite advancements in sound reproduction technology and accreditation of cinema sound systems. The X curve for the equalisation of cinema sound systems has been used for many years and is a foundational element for setting up the tonal balance and achieving accreditation of cinemas.

This paper challenges the acoustic theory behind the X curve, using analysis of acoustic and electro-acoustic measurements made in cinemas and in other spaces where high-fidelity and high intelligibility is required. Both the temporal and tonal developments of the sound field in a number of rooms are examined in detail and related to temporal properties of speech and music and human's perception of tonality.

This work follows a series of conference papers that a number of the authors have presented about the sound quality in cinemas.

Note that this paper addresses only the shape of the X curve equalisation and does not discuss the general use of equalisation as required for certification of a cinema or dubbing room by Dolby Laboratories.

Background to the X Curve

The cinema reproduction chain is regarded as having two sections; the A and B chains. The B chain comprises the elements of interest for this paper: those that follow the switching of signal sources including equalisers, power amplifiers, loudspeakers, screen, and modifications to the sound caused by auditorium acoustics and distance to listeners.

The X Curve (X stands for Experimental) represents the target shape for the steady state frequency response of cinema sound systems when measured with pink noise at a reference position. The X curve, originally promulgated as ISO2969 and subsequently adopted by the SMPTE as S202, has been specified for many years and relates only to the B Chain and therefore has no role in pre-emphasis or other signal corrections.

Allen[1] describes the history of the X Curve equalisation and presents an explanation of the factors that are thought to be addressed by it. Figure 1 reproduced from Allen[1] shows the steady-state frequency response of the B Chain as specified by the X Curve.

Allen notes that in listening tests, an equalisation slope of around -3 dB per octave from about 2 kHz seemed to give the best aural results, along with a slight limitation to low-frequency bandwidth. The low-frequency limitation is easy to explain - more low-frequency energy would probably overload the loudspeaker and generate distortion components. Allen suggests that the reason for the apparently desirable HF droop is not easy to explain, but offers three possibilities, singly or in combination:

- Some psychoacoustic phenomena involving the integration of faraway sound and picture.
- Some distortion components in the loudspeaker, making HF objectionable.
- The result of reverberation build-up, as described below.

The bulk of the Allen’s explanation concerns the effect of reverberation, and it is this issue that this paper discusses.

Assumed Reverberant Build Up (as per Allen)

This section paraphrases the explanation given by Allen[1] for the effect of reverberation on steady-state measurements of frequency response in rooms.

Generally, the larger the volume of the room, the higher the reverberation time (RT) will be. More often than not, the RT will be greater at low frequencies and lower at high frequencies. For a room to reach a steady-state acoustic condition, the audio signal must be sustained for a period proportional to the reverberation times at all frequencies.

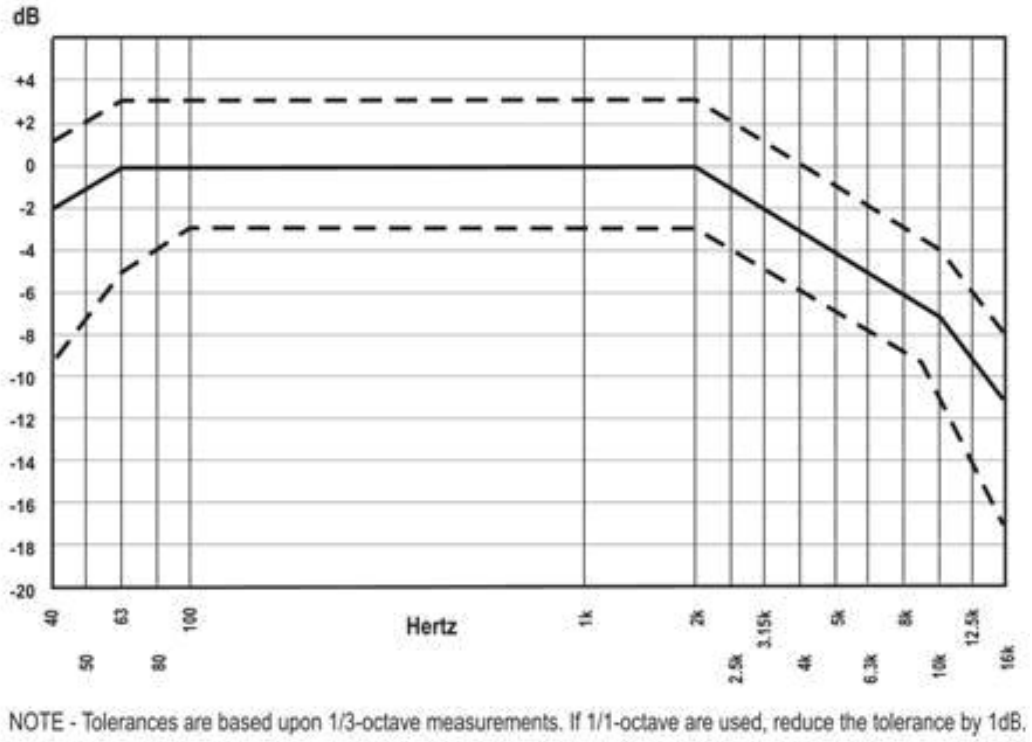


Figure 1 The X-curve. SMPTE 202M – 1998

Measurement of a steady-state signal (such as pink noise) will show the combination of the direct-field sound of the loudspeaker and the sound progressively added by room reverberation, which is termed “room gain”. Consequently, there will be a difference between the measured frequency response of the first sound-arrival (before reverberation build-up) and the response measured when the sound has reached the steady-state condition (after reverberation build-up).

Figure 2, also reproduced from reference[1] shows the increase in room gain as a function of frequency and time.

If the perceived spectral content of a sound is therefore determined by its duration, consideration must be given to the duration of cinematic content. Much of a typical movie sound track is dialogue, and as consonant sounds such as “t,” “p,” “d,” have a short duration, they may arrive and decay in an average-sized theatre before the first reflection arrives at listeners and before the onset of reverberation build-up.

So, according to Allen, if the requirement is to have the short-duration sounds reproduced with a flat frequency response, then the steady-state condition will show an apparent high-frequency droop. To achieve the X-curve characteristic, the steady-state frequency response as measured using a spectrum analyser with pink noise as a test signal, should be equalised to

show a -3 dB per octave slope above 2 kHz. Under this condition, the direct-field sound will be closer to flat than the -3 dB per octave slope would suggest.

.

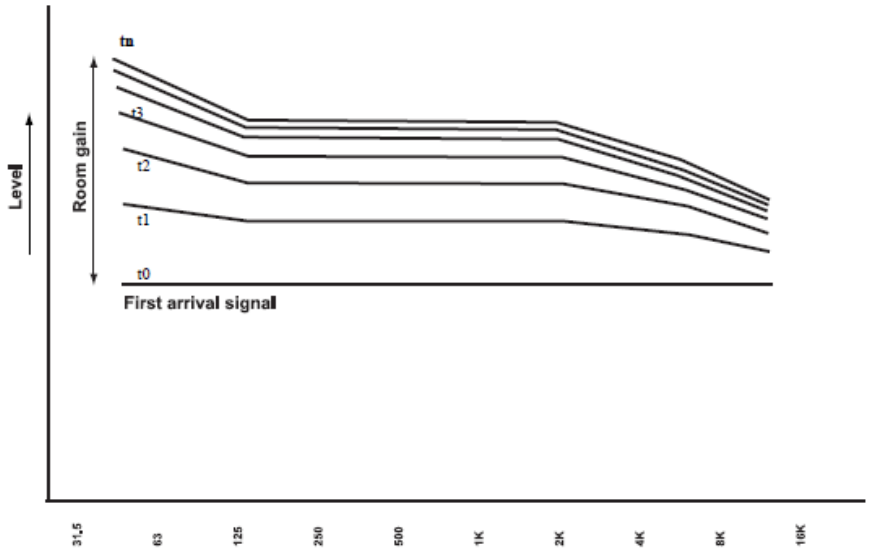


Figure 2 Frequency response of room gain vs time in medium to large size theatre (Allen[1])

Indeed, if a large room were to be equalised flat with pink noise, it is possible that the first-arrival signal would show a rising high-frequency characteristic, as shown in Figure 3.

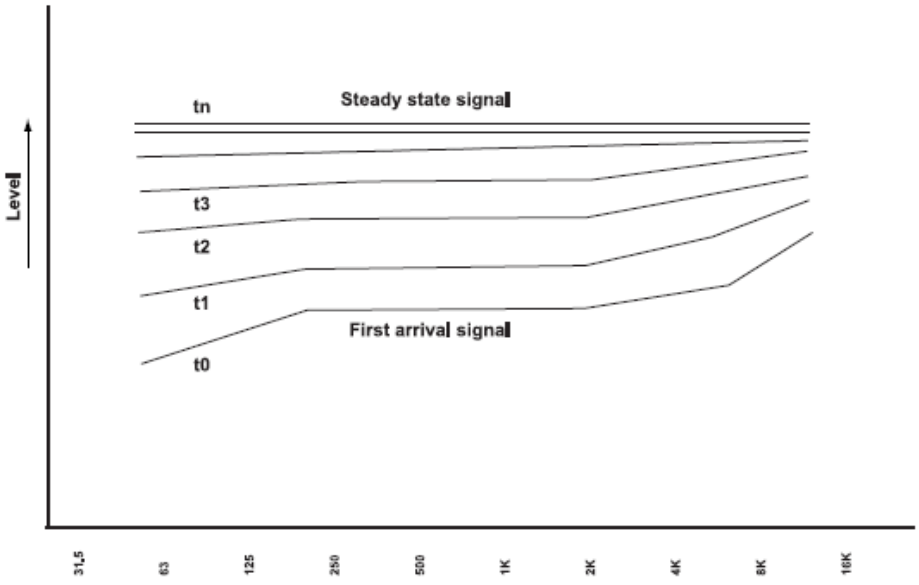


Figure 3 Build-up over time of frequency response measured with pink noise if steady-state response was equalised to be flat. (from Allen[1])

According to the standards, there is a family of "X" curves, which show flatter responses as rooms get smaller, however the response roll is still significant.

At present, dubbing theatres and commercial cinemas seeking Dolby certification are equalised as closely as possible to the X-curve at a point roughly two-thirds of the distance from the screen to the rear wall of the rooms. This region of the rooms has traditionally been considered to be the most representative average of the whole room and is generally where the mixing desk is located in a commercial dubbing facility

Acoustic Investigations

In work undertaken by two of the authors[2] in 2009, the impulse responses of 18 Dolby Digital and Dolby Surround rooms, (both studios and commercial cinemas), were measured at two positions:

- the standard calibration distance of two-thirds of the distance from the screen to the rear wall
- a distance of 2 metres from the central loudspeakers.

The size of the rooms was commensurate with a normal cinema and every room had Dolby certification.

The impulse responses at the standard calibration distance in the eighteen rooms were analysed using different time windows to assess the validity of Allen's hypothesis of reverberant build up shown in Figure 2.

Using Matlab, the frequency response of each measurement was computed from the impulse responses by the Fast Fourier Transform using rectangular and Tukey windows of different lengths. The resulting frequency responses were then energy-averaged over a 1/15th octave bandwidth and the values assigned to the associated frequency at the centre of each bandwidth.

Time Windows

The rectangular window has a benefit that all wanted data in an impulse response is weighted equally, but has the disadvantage that it generally produces spectral leakage (due to the infinitely fast truncation of the time data). To confirm that spectral leakage did not change the conclusions, a modified Tukey window was also used.

A Tukey window shape is also known as the "tapered cosine window", and can be regarded as a raised cosine window which has been convolved with a rectangular window. An example of the half Tukey window is given in Figure 4. The flat top of the window allows equal weighting of that section of time data, while the cosine section reduces the leakage due to the data truncation.

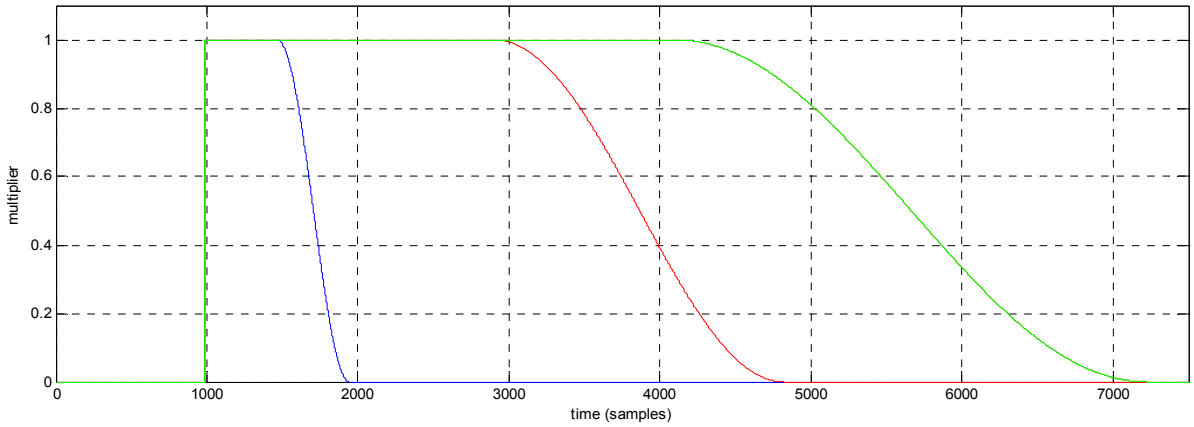


Figure 4 Example of half Tukey window

The following time data durations were used for the analysis:

10 ms	50 ms	80 ms	300 ms	All time data
Rectangular	Rectangular	Rectangular	Rectangular	Rectangular
Half Tukey	Half Tukey	Half Tukey	Half Tukey	N/A

The rationale for the selected time window lengths is:

- The 10 ms window includes the loudspeaker's direct field at mid frequencies and above and represents the likely lower limit of the psycho-acoustic temporal integration time[3].
- The 50 ms and 80 ms windows are mirrors of the C_{50} and C_{80} acoustic metrics discussed below.
- The 300 ms window is a reasonable time to integrate the majority of the room's discrete reflections, and will include reflections that are not useful for clarity.
- All time data represents the steady-state condition, which would be measured with pink noise if sufficient measurements were made to average out the stochastic variations in the noise.

Parallels with Intelligibility Metrics

Measures of the ratio of early-arriving sound to late-arriving sound are used as reasonably reliable indicators of the ability of a sound/room system to deliver speech intelligibility. The C_{50} and C_{80} metrics are based on the principle that clarity is determined by the relative strengths of useful and detrimental sound energy. Useful sound is the combined energy of the direct and early-reflected sounds, while "detrimental" sounds are the combined energy of late reflected sound, reverberant sound and ambient noise. A duration of 50 ms for speech and 80 ms for music is generally used for the time period dividing these two types of sound field.

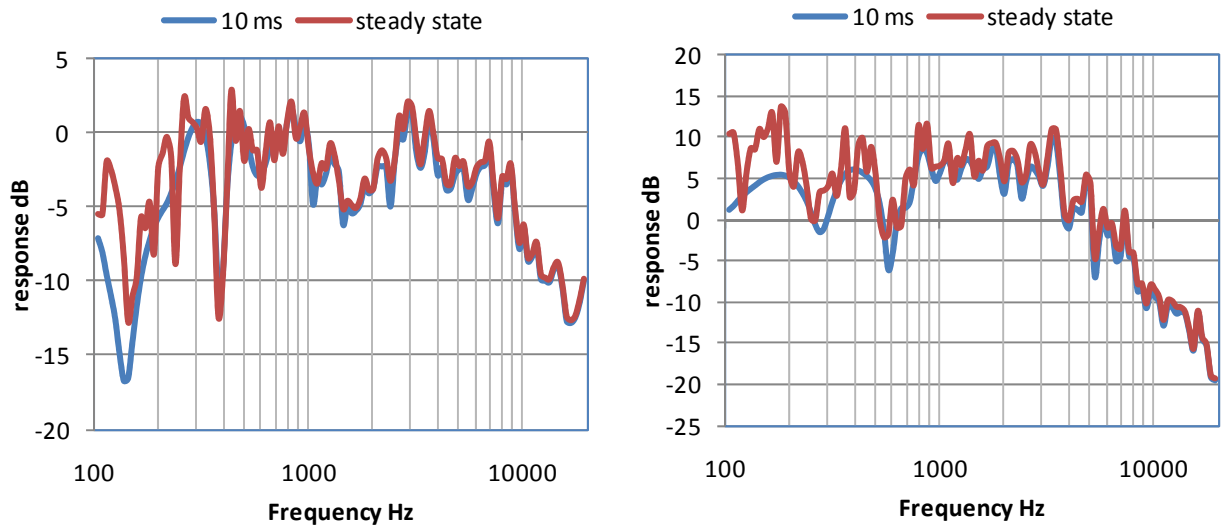
Both metrics are found by integrating appropriate portions of the room impulse response. It should also be recognised that the use of a sharp boundary division between early and late oversimplifies the situation.

The C_{50} is also loosely related to the direct to reverberant ratio (D/R) and includes the possible enhancement of speech sounds by strong early reflections.

Results and Discussion

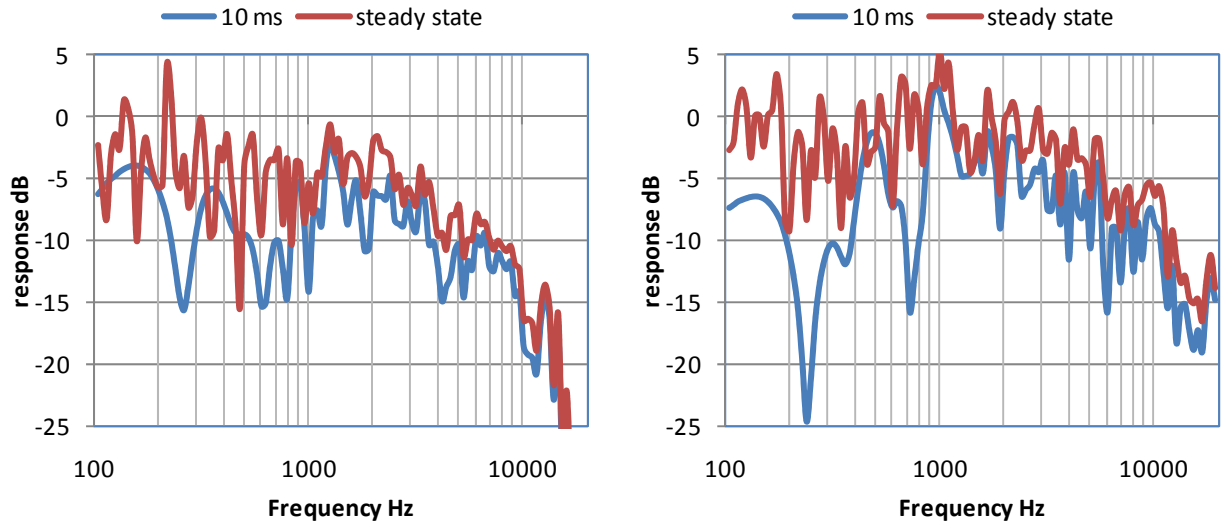
Analysis of Acoustic Measurements

Figure 5 and Figure 6 compare the frequency responses of four rooms with rectangular window lengths of 10 ms and all time data (steady-state) at the two-third's distance position.



Room 2 Room 9

Figure 5 Frequency responses of two rooms with 10 ms and all time data windows



Room 13 Room 15

Figure 6 Frequency responses of two rooms with 10 ms and all time data windows

The difference between the frequency responses of each of the four shorter time windows and the steady state response were computed for each room at the two locations. The differences were then averaged over the rooms.

Figure 7 shows the average differences in the frequency responses with four rectangular time windows with the steady state response at the two-thirds position, while Figure 8 shows the maximum positive and negative differences between each time-window response and the steady state response in each room. Note that the curves in Figure 8 do not show the response of any one room, but are comprised of data points from any number of rooms.

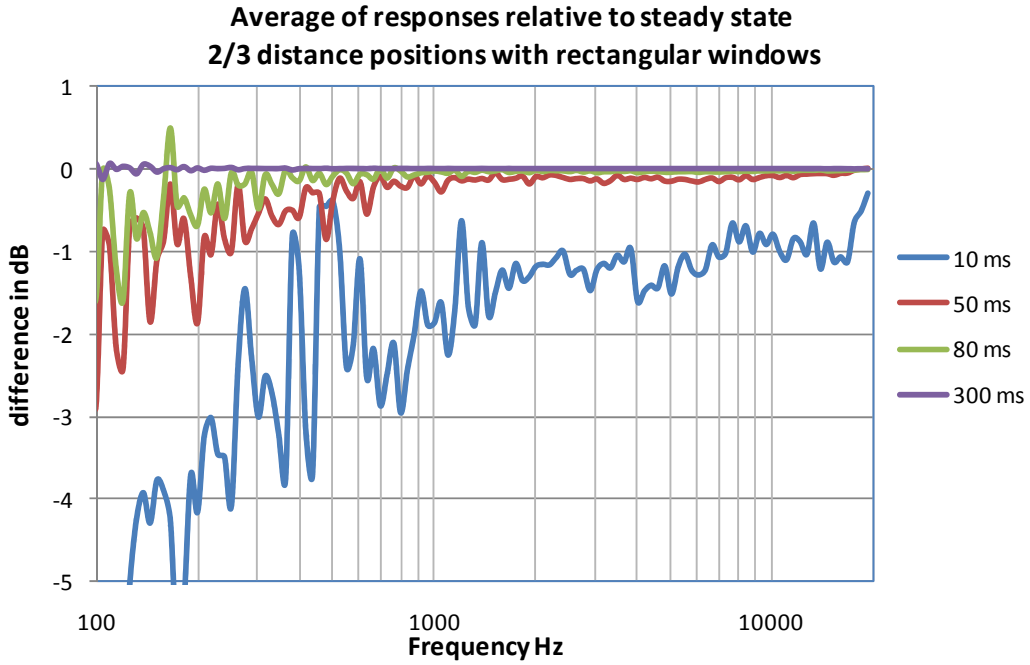


Figure 7 Average differences between time-window and the steady state responses

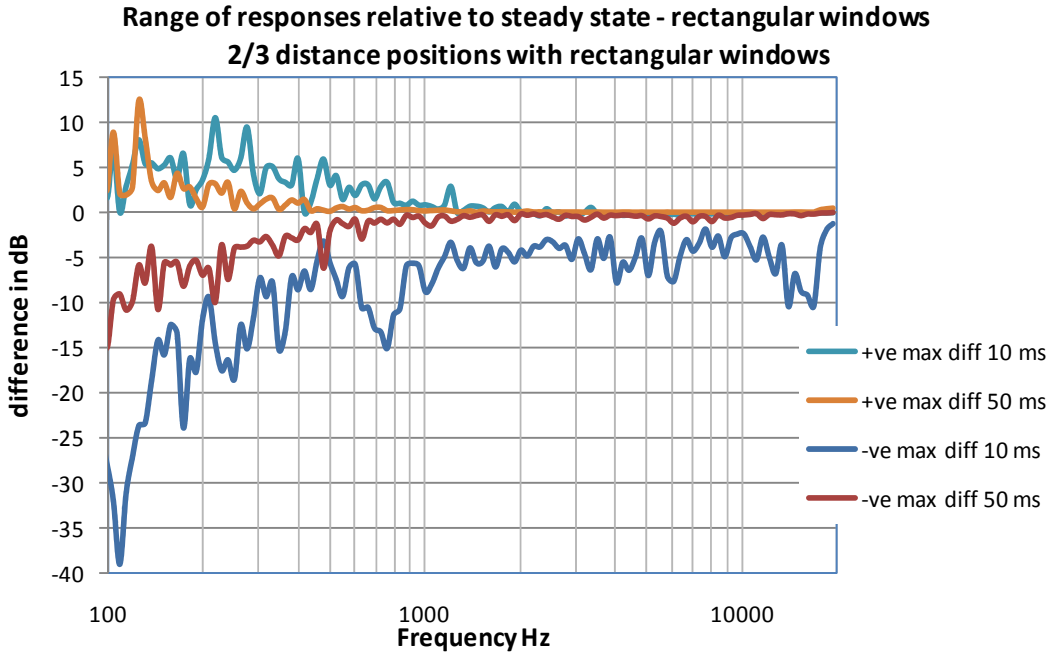


Figure 8 Maximum positive and negative differences between time-window and steady state responses.

Figure 9 shows the 10 ms and steady state responses of a suspended loudspeaker in the Debating Chamber in the New Zealand Parliament which has room volume of 3200 m^3. The reverberation times (RT) in this chamber lie between 1.4 and 1.1 seconds over the range of 125 Hz to 4 kHz, while at 8 kHz, the RT is 0.8 seconds.

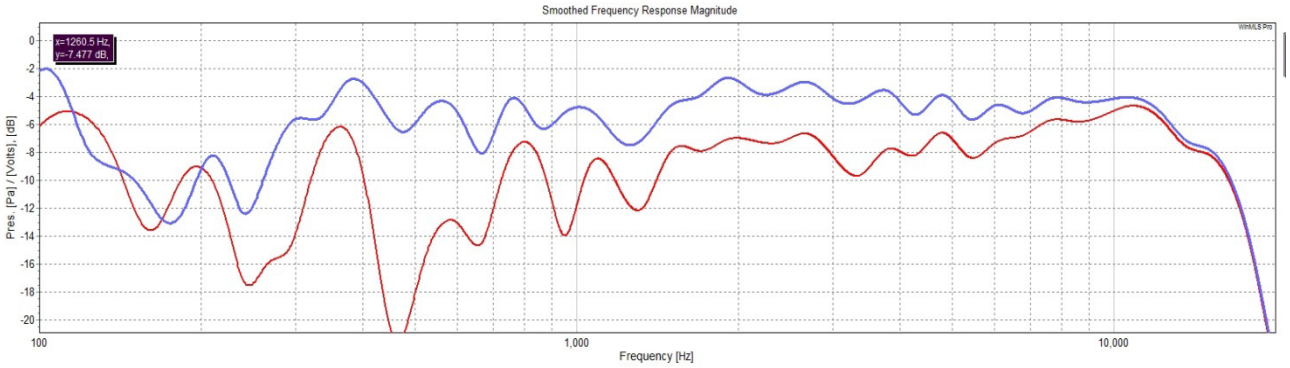


Figure 9 Frequency responses with 10 ms (red) window and steady-state (blue) for a loudspeaker in the New Zealand Parliament Debating Chamber.

Figure 10 shows the reverberation times in octave band intervals of the 18 dubbing and cinema rooms. Noting again that all these rooms have gained Dolby certification, the variety of reverberation characteristics is interesting. Above 1 kHz, the RTs are less than 0.4 secs., and this is commensurate with the upper and lower reverberation time limits recommended by Allen in the Dolby Standard[4], which are shown in Figure 11. Some of the rooms have lower RTs than the Dollby specified limit.

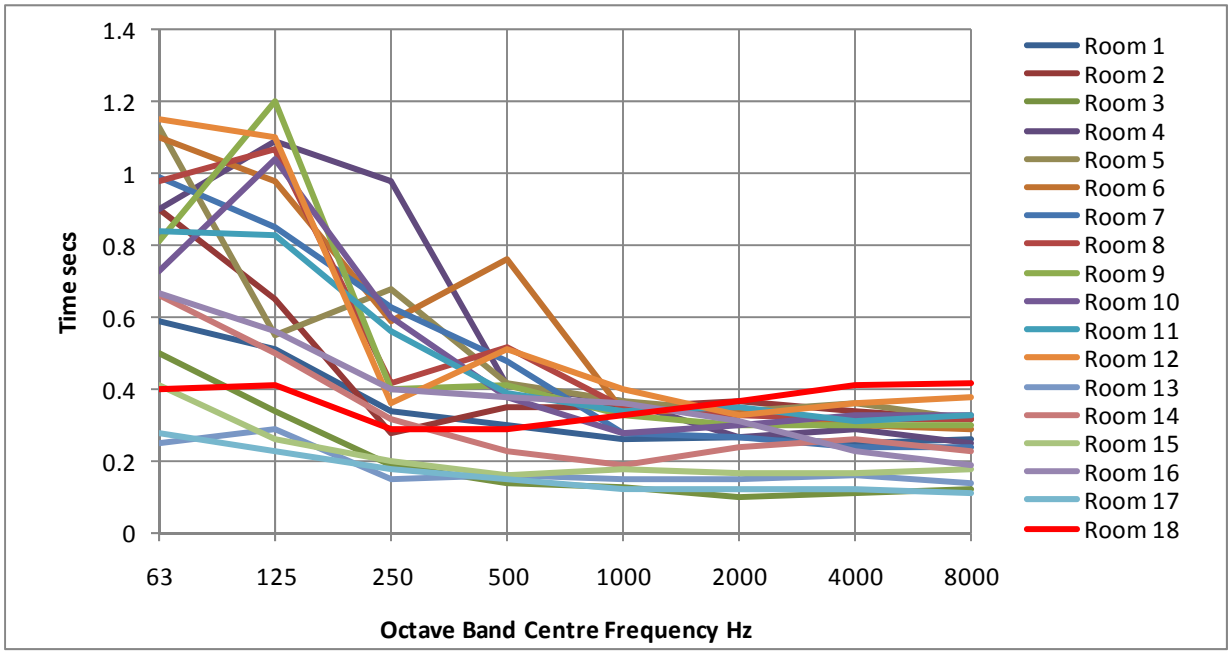


Figure 10 Reverberation times of dubbing and cinema rooms

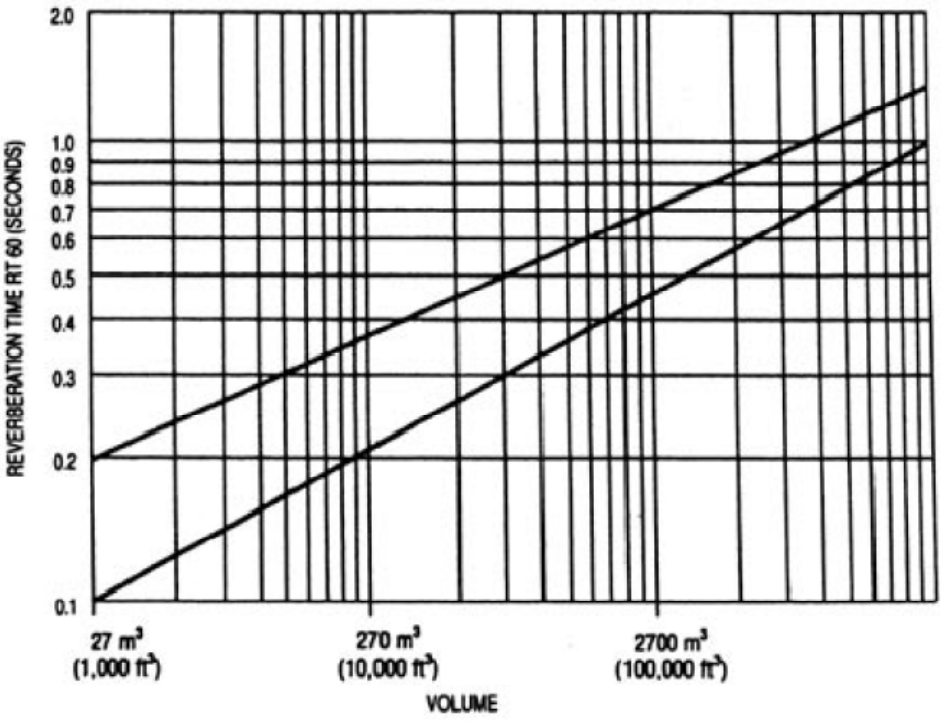


Figure 11 Upper and lower limits of recommended reverberation times for cinemas at 500 Hz[4]

Conclusions

The following conclusions have been drawn from the measured data.

- The average difference between the 10 ms response and the steady-state condition is approximately 1.3 dB above 2 kHz.. This indicates the typical steady-state response is 1.3 dB higher than the 10 ms windowed response in 18 Dolby Certified rooms.
- The average differences of the 50 ms, 80 ms and 300 ms responses with the steady-state responses are less than 0.2 dB.
- The lowest values of the 10 ms responses (-ve max differences) are approximately 4.5 dB below the average, and these lowest values are relatively constant from 1300 Hz to 12 kHz. Although the lowest response is not a response per se, the shape is dissimilar to the X curve.
- In summary, the reverberant build up described by Allen does occur to a small extent, however the change in room gain is much less than the X Curve shape suggests and the frequency response shape of the build-up is essentially flat.
- The results with the Tukey windows are similar to those of the rectangular window and are not presented here.
- Applying the X-curve (*or any other curve*) to shape the spectrum of the steady state sound field will resut in the direct field having a non flat frequency response, which is contrary to the intention of the X curve. This conclusion was illustrated by the measurements of cinema loudspeakers in the close field (2 m) presented by Newell et all[2].

- The 10 ms frequency response plots show that a large amount of energy above 1 kHz has been removed from the direct field. This must have repercussions on the tonal character of speech.
- The reverberation times of a number of the Dolby-accredited rooms above 1 kHz are commensurate with those recommended in 1994 Dolby Standard[4] and yet these rooms do not show the assumed frequency-response characteristic during reverberant build up. We therefore conclude that it is likely that the X curve has not been valid at least since 1994, and quite possibly earlier.
- Even in the comparatively hiogh RT environment of the NZ Parliament, the reverberant build-up is far less than hypothesised by the X-curve rationale.

Masking Issues

The use of the X curve is now examined with respect to issues affect speech intelligibility.

Self Masking

The loss of high frequency content can be looked at from another perspective, which is the reduction in intelligibility of speech sounds due to self-masking and extraneous sounds. The Speech Transmission Index (STI) metric attempts to address this issue and measures the loss of modulation as an effective signal to noise ratio.

The principal issue with the X curve regarding speech intelligibility is one of self masking in the ear, in which both upward and downward masking effects cause sounds in one frequency range to decrease the perceived signal-to-noise ratio in other frequency ranges. Self-speech masking is when the sounds in speech mask other sounds. This issue was explored by Leembruggen and Stacey[5], in which the relative intelligibilities of speech shaped by various frequency responses were examined under low-noise conditions. Considerable degradation in subjective intelligibility resulted from a number of responses, which was due to self masking effects.

In the situations examined by Leembruggen and Stacey, and later by Leembruggen, Hippler and Mapp[6,7], removing the high frequency content by use of the X Curve weighting would seriously degrade the early-time field (direct field up to 50 ms) and therefore degrade subjective intelligibility, particularly when the speech is recorded in a reverberant context.

The presence of music or Foley sounds during dialogue results in a decrease in the signal to noise ratio in a range of frequency bands, and when the early-time energy at high frequencies is reduced by the X Curve weighting, the masking caused by the lower frequency bands of speech becomes significantly more apparent.

Extraneous Noise

In a well designed cinema, background noise levels due to air-conditioning or break-in from adjoining cinemas are generally low. In this context, intimate dialogue that would be heard at say 50 dBA in real life, is presented to the audience at a significantly higher level, providing signal to noise ratios that are sufficiently high to allow good speech intelligibility.

However, there can be considerable occupational noise due to packets of chips, people whispering, or shuffling in their seat, and this noise can degrade the intelligibility of intimate speech by masking. Under these conditions, the removal of high frequency energy in speech by the X curve weighting will degrade speech clarity.

Listening Results

The use of the X curve was also examined using the subjective listening skills of the authors.

The authors have considerable experience in creating environments and sound systems for critical situations, such as recording studios, national parliaments and the highest courts. We have spent many years of deisgning and tuning speech sound systems in high profile situations and acoustic measurements.

Our listening work for these situations has universally supported the notion that a relatively flat high frequency response is critical for clarity, comfort and enjoyment of the sound. Our speech sound sytems invariably are equalised to be ostensibly flat up to 12 kHz approximately, when measured with any time window.

In contrast to the type of sound we deliver to our clients, is our perception of cinema sound, which is not hi-fi like and sometimes causes difficulties for us in understanding speech when there are accent differences and/or Foley effects.

Other Reasons

There are many additional reasons that a number of the authors have noted[8] about the inappropriateness of the X curve. These relate to:

- The spatial properties of the room that cannot be corrected using sound system equalisation.
- Whilst it is true that the perception of sound in large rooms is different from that in small rooms, the differences become less noticeable as the decay times decrease, and the reduction of the decay times in theatres for cinema reproduction is a growing trend.
- There is no 'one curve fits all' situation. Furthermore, the question remains as to whether *any* curve can even closely represent the different characteristics of different rooms. As Toole pointed out in a communication to the SMPTE (August 4th, 2010), and similarly in a recent book[9], '*Unlike a human, the microphone does not take any note of the angle of incidence of the direct and reflected sounds, nor does it make any allowance for the time of arrival of those sounds, nor does it acknowledge spectral variations among any of the sounds. The microphone simply adds them together. .....It is well known that two ears and a brain are vastly more analytical than a microphone and an analyser. Humans respond differently to sounds arriving from different directions at different times'*. No form of room equalisation can take all of this into account.
- If response correction is to be *reliably* applied, it should be done in the close field, otherwise it becomes convolved with the unequalisable non-minimum-phase characteristics of the room acoustics, and 'correction' then becomes an inappropriate word to use for the process. Any agreed curve could be more accurately applied within the electronics of a system, as a fixed equalisation, as with NAB equalisation for analogue tape recorders or RIAA equalisation for vinyl discs. However, this still leaves the question as to whether any such curve is necessary at all, these days, given the potential accuracy and efficiency of modern audio reproduction systems.
- When the overall X-curve room response is applied in the distant far field, the equalisation can even make a bad situation worse, especially in the poorer rooms, which are the ones which need the most 'help' with clarity.
- Essentially, whatever the source, the direct sound must be natural (i.e. the input must be the same as the output) if natural sound is to prevail overall. Would the industry ever dream of using a cloth to equalise a piano in a concert hall to an X curve ?

- Perhaps 40 years ago, when cinema responses were generally in a state of chaos, the X-curve was a viable means of achieving better standardisation. Indeed, it has been a reference which, over the subsequent years, has stood up well, but over those same years our methods of measurement have become so much more flexible and precise, our knowledge of psychoacoustics has developed greatly, and the introduction of digital recording has relieved the industry of many of the restrictions imposed by optical and magnetic analogue soundtracks. Loudspeaker and amplifier technology has also moved far ahead of where it was in 1971. Above all, however, there has been a trend towards reduced theatre sizes and drier acoustics, which reveal much more detail in the sound and render much more obvious the effects of inappropriate equalisation.

Conclusions

1. Measurements have shown that the reverberant build-up of sound over time essentially has an almost flat frequency characteristic above 1.3 kHz. This result, which was explored over 19 rooms is vastly different to the hypothethis associated with the X curve.
2. When averaged over 19 rooms, there is negligible difference between the responses of the early-time sound fields and the steady state and therefore there is no reasonable foundation for the explanation of reverberant build up expounded by Allen.
3. With the steady-state response equalised to the X-curve, the direct field of the loudspeakers in all the rooms assessed will not be flat, contrary to the X-curve hypothesis.
4. Working with a system equalised to the X curve makes the task of dubbing engineers much more difficult, as they need to compensate for an unknown amount of high frequency loss when mixing and mastering. It is as if someone has placed a blanket between the talker and the listener and the dubbing engineer must compensate for this blanket.
5. The use of the X Cuve exacerbates upward-masking effects for dialogue, thereby degrading the speech intelligibility of dialogue, particularly if it occurs during music or Foley effects. Although the engineer may try to compensate for this loss of intelligibility by equalising the dialogue, there is no reference point for that compensation. A consequence is that the cinema mix will be unsuitable for home cinema application, as the dialogue will have excessive high frequency content when played on a good quality home system.
6. The results do not depend greatly on the size of the rooms as would be expected from the Allen understanding of reverberant build up.
7. The reverberation times of a number of the Dolby-accredited rooms above 1 kHz are commensurate with those recommended in the 1994 Dolby Standard and yet these types of rooms do not show the assumed frequency-response characteristic during reverberant build up. We therefore conclude that it is likely that the X curve has not been valid at least since 1994, and quite possibly earlier.
8. The authors conclude that the use of the X curve is detrimental to the enjoyment of cinema.

.

References

1. Allen, I; 'The X-curve: Its Origins and History', SMPTE Journal, Vol.115, Nos. 7 & 8, pp. 264-275 (2006)
2. Newell, P; Holland, K; Torres-Guijarro, S; Castro, S; Valdigem, E; “Cinema Sound: A New Look at Old Concepts”, *Proc. of the Institute of Acoustics*, Vol. 32, Part 5, (Nov 2010)
3. Wallace, A; Blumstein, S. “Temporal integration in vowel perception” J Acoust Soc Am. 2009 March; 125(3): 1704–1711.
4. Allen, I. “Technical Guidelines for Dolby Stereo Theaters,” Dolby Laboratories, September 1994
5. Leembruggen, G., Stacy A. Should the Matrix be Reloaded? Proc IOA. 2003.
6. Leembruggen, G., Hippler, M., Mapp, P. Exploring ways to improve sti’s recognition of the effects of poor spectral balance on subjective intelligibility. *Proc. IOA.* 2009, Vol. 31, Pt 4.
7. Leembruggen, G., Hippler, M., Mapp, Further Investigations into Improving STI’s Recognition of the Effects of Poor Frequency Response on Subjective Intelligibility. Convention Paper 128th AES Convention 2010 May London, UK
8. Newell, P; Holland, K; Newell, J and Neskov, B , “New Proposals for the Calibration of Sound in Cinema Rooms” Convention Paper 130th AES Convention 2011.
9. Toole, Floyd E.; 'Sound Reproduction', Focal Press, Oxford, UK, (2008),