A New Method for Measuring and Calibrating Cinema Audio Systems for Optimal Sound Quality

Linda A. Gedemer

Acoustic Research Centre

School of Computing Science and Engineering

College of Science and Technology

University of Salford, Salford, UK

Submitted in partial fulfilment of the Requirements of the Degree of Doctor of Philosophy

2017

Contents

List of Figures

List of Tables

List of Abbreviations

2/3L	Two-thirds Length (of room)
AES	Audio Engineering Society
AMPAS	Academy of Motion Picture Arts and Sciences
ANOVA	Analysis of Variance
ANSI	American National Standards Institute
BRIR	Binaural Room Impulse Response
BRS	Binaural Room Scanning
dB	Decibel
F	Fisher's F Ratio (also referred to as the F Statistic)
FFT	Fast Fourier Transform
GUI	Graphical User Interface
HATS	Harman Audio Test Software
HRTF	Head Related Transfer Function
Hz	Hertz
IIR	Infinite Impulse Response
ISO	International Organization for Standardization
ITU	International Telecommunication Union
MPRC	Motion Picture Research Council
MUSHRA	Multiple Stimuli with Hidden Reference and Anchor
NATO	National Association of Theatre Owners
NFTU	Nor-disk Film/TV Union
NIDCD	National Institute on Deafness and Other Communication Disorders
OSHA	Occupational Safety & Health Administration
p	Probability
RT_{60}	Reverberation Time
SMPTE	Society of Motion Picture and Television Engineers
SPL	Sound Pressure Level
SQL	Structure Query Language (Database)

Acknowledgements

If at the beginning of my thesis writing you had told me that the 'Acknowledgements' section was going to be one of the toughest parts to write I would not have believed you. Until now... The issue is not about whom exactly to thank but the fact that there are so many.

Firstly, I must thank my supervisors, Drs. Trevor Cox and Sean Olive. It was a long road across many miles and I could not have succeeded without their valuable input.

I owe many thanks to Dr. Floyd Toole and my dear friend Dr. Ben Shirley for proof reading my thesis, especially since I'm so bad at spelling that even Spell Check gives up on me sometimes. I'm particularly grateful to Dr. Floyd Toole for serving as my "PhD Godfather" over the years; constantly providing guidance, encouragement, a shoulder to cry on and occasionally a really great glass of wine.

Speaking of friends... 5 years ago I came to Salford; a stranger in a strange land. And then I met the crew of G10/G11. Calum, Paul, Jenna, Gennaro, Charlie, Darius, Toby, James, Eulalia, Andy, Julian, Tony and Zbigniew (I'm sure I left someone out); we are not just friends, we are family. Of course I couldn't mention this group without also thanking the owner of The Crescent Pub for letting us take over their outdoor space and turn it into the official 'Salford Acoustic PGR Beer Garden and Smoking Lounge'. I miss those days.

But the two most important people in my life are the ones who deserve the most thanks; my husband Timothy and my daughter Bryce. To say I could not have done this without you would be a monumental understatement. I'm sure you would both get a kick out of knowing that this is one of the rare times I'm finding it hard to think of words to say. Timothy, my love for you and gratitude for your support through these crazy 5 years can never be fully expressed. But I don't mind taking the rest of my life to try ;-).

There were many times that I didn't think I had the strength to finish my PhD quest. Many hours of self-doubt and soul searching. But my daughter always gave me hope, as illustrated in figure A.1.

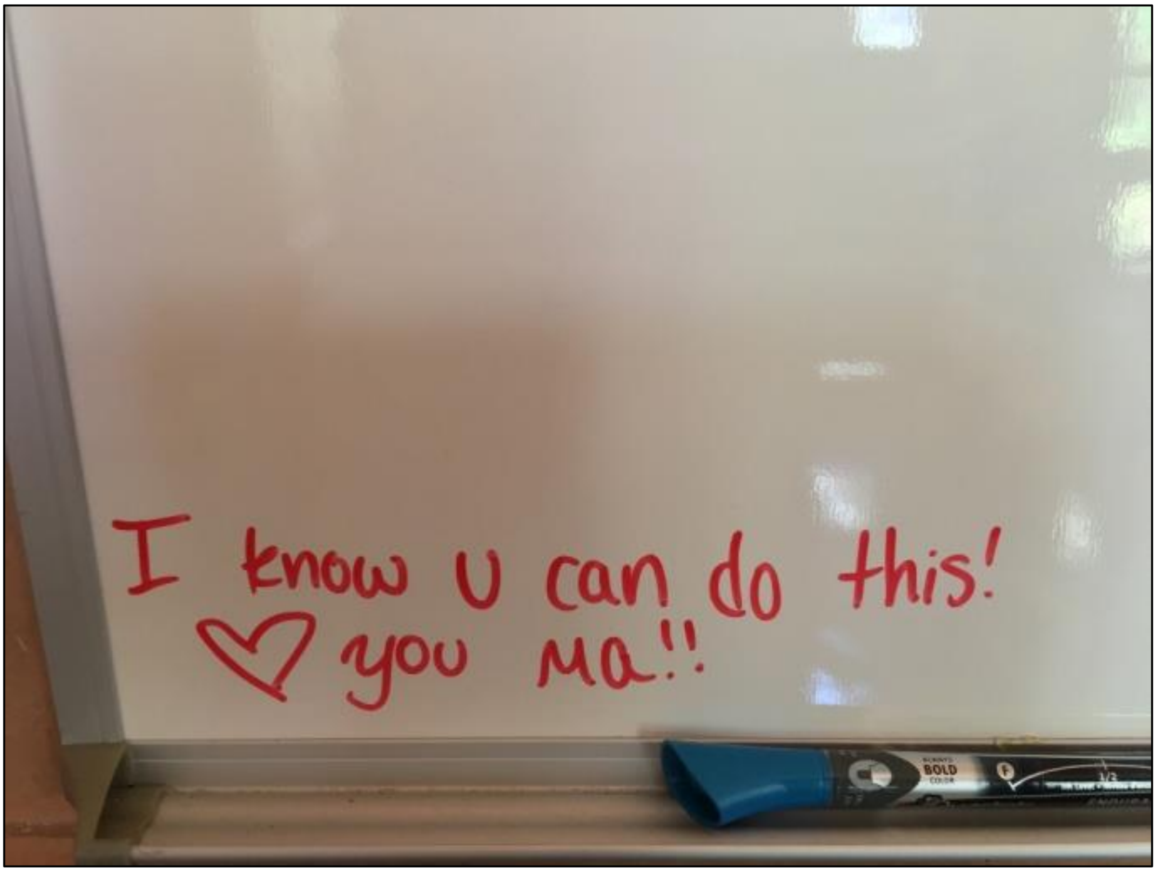


Figure A.1: Note of encouragement left by daughter on PhD whiteboard while in the depths of despair attempting to write thesis

I dedicate this thesis to you Bryce. You are my heart. Love you Bean!

Declaration of Originality

All of the data collection and analysis presented in this thesis are original works by the author, though select pieces of work were jointly researched. Aspects of this research have previously been presented at the following:

1. GEDEMER, L. 2013. Evaluation of the SMPTE X-curve Based on a Survey of Re-recording Mixers. *AES 135th Convention.* New York, NY: Audio Engineering Society.

2. GEDEMER, L. & WELTI, T. 2013. Validation of the Binaural Room Scanning Method for Cinema Audio Research. *AES 135th Convention.* New York, NY: Audio Engineering Society.

3. GEDEMER, L. & WELTI, T. 2013.Validation of the Binaural Room Scanning Method for Cinema Audio Research. Reproduced Sound 2013, Manchester, UK: Insitute of Acoustics.

4. GEDEMER, L. 2015. Predicting the In-room Response of Cinemas from Anechoic Loudspeaker Data. *AES 57th International Conference.* Hollywood, CA: Audio Engineering Society.

5. GEDEMER, L. 2015. Subjective Listening Test for Preferred Room Response in Cinemas - Part 1: System and Test Descriptions. *AES 139th Convention.* New York, NY: Audio Engineering Society.

6. GEDEMER, L. 2016. Subjective Listening Tests for Preferred Room Response in Cinemas - Part 2: Preference Test Results. *AES 140th Convention.* Paris, France: Audio Engineering Society.

Research presented in papers 1, 2, 4, and 6 is cited in TOOLE, F. 2016. *Sound Reproduction: The Acoustics and Psychoacoustics of Loudspeakers and Rooms*, 2nd Edition, Focal Press.

Abstract

The aim of this research is to utilize new methodologies and technology in order to gain insight into how the modern cinema audio system could be calibrated to provide improved audio performance. To this end, both objective and subjective measurements were developed to better understand the audio preferences of listeners, the requirements of the audio systems inclusive of the acoustic environment, and how the two are related.

Part of the data for this research was derived from a survey of re-recording mixers regarding their use and opinion of the current SMPTE standard. The survey confirmed anecdotal information suggesting that re-recording mixers use high-end pre-emphasis to compensate for the severe roll-off induced by the SMPTE X-curve. It is also noted that the re-recording mixers' opinions of how well their mix translates from dub-stage to cinema is correlated to how many years they have spent in the industry.

For the first time a binaural room scanning (BRS) system was validated via a series of listening tests for use in small, medium, and cinema-sized room listening tests. This system shows great promise as a tool for creating various sized room-based listening tests, as there were no statistical differences in ratings between the in situ and BRS playback conditions.

To further understand listener preference to in-room responses curves, a series of listening tests utilizing the BRS system were conducted using various sized cinemas, seating positions within the cinemas, audio tracks (including those mixed on a SMPTE calibrated dub-stage) and target curves. The overwhelming outcome was that regardless of cinema size, seating position or audio track utilized; the "curve" that listeners preferred is a relatively flat 0.9dB/octave slope with a 6.5dB bass boost below 105Hz and a -2.5dB roll off above 2.5kHz. Of the 5 target curves presented, the SMPTE X-curve place fourth with scores very near the low-rated perceptual anchor. This calls into question the notion of the X-curve providing "ideal" translation between dub-stage and cinema and in fact, challenges the concept of translation all together.

Research was completed in an effort to identifying the number of microphone positions required, along with their placement, in order to accurately capture a cinema's response for calibration purposes. A novel experiment utilizing anechoic loudspeaker data as a guideline for

analysis demonstrated that, with proper data, the number of microphones and their positions plays a less critical factor in determining the room response. The collected data shows that even with as few as 4 microphones at varied positions, the resultant room response will trend towards the anechoic data above 1kHz. From around 300Hz to 1kHz, there is evidence of seat effects that may be resolved through randomizing the microphone heights. Below 300Hz, the room becomes the dominating factor and more than 5 microphone positions will be required to properly identify any problems.

Chapter 1 - Introduction

1.1 General Introduction

In this section, the research area is introduced and the research aims are declared. The structure of the thesis is outlined and the contributions to the research field are identified.

1.2 Motivation

The public's concerns regarding poor cinema audio and the dwindling revenues in the film industry are providing a strong impetus amongst several standards organization to improve this situation. Standards organizations such as the Society of Motion Picture and Television Engineers (SMPTE), the Audio Engineering Society (AES), and InfoComm International are aiming to revise aged methods of calibrating cinema audio systems with the hopes of improving their performance. Their efforts are reflected in the SMPTE TC-25CSS (Technology Committee 25, Cinema Sound Systems), the now defunct AES-X218 (Measurement and Calibration of Sound Systems in Rooms) and InfoComm ANSI/INFOCOMM 1M-2009 (Audio Coverage Uniformity in Enclosed Listener Areas). However, these organizations are trending towards making subtle revisions to the existing standards which leads to the new standard still utilizing some of the old, out dated and ineffective methodologies. In addition, some of the research that is currently being conducted either to create new standards or to revise existing standards lacks scientific rigor.

1.3 Introduction to the SMPTE X-curve

The Society of Motion Picture and Television Engineers (SMPTE) 'X-Curve' is part of the existing SMPTE ST 202 standard which was first proposed in 1975 for calibrating dubbing stages and cinema loudspeaker systems. The goal is to provide a consistently high standard of sound quality in the creation and reproduction of film sound. The X-Curve aimed to provide the motion picture industry with a standard that ensured plausible inter-changeability of program material, from one studio to the next, from studio to theatre, and from film to film, which takes into account the different perceived spectral response of different room sizes.

The in situ calibration process consists of using pink noise as a source and measuring the loudspeaker system at up to 5 locations in the cinema and then equalizing it to a defined target curve in 1/3-octave bands. Correction factors to the target curve can be applied to account for different room sizes. Unfortunately, the SMPTE calibration has never been properly validated using scientific subjective and objective measurements. There is no scientific evidence that it actually delivers high quality audio translation from dub-stage to cinema.

In fact, there is anecdotal evidence showing that cinemas goers have become increasingly unhappy with the quality of audio in cinemas. The cinema-going population has become more sophisticated in their taste for high quality audio. This can be linked to the improvements in home cinema audio systems which now include items such as high-end Blu-ray DVD players, high-definition cable and satellite receivers, high-definition internet streaming devices, and a full complement of well designed loudspeakers. This audience has started to question the ever-increasing price of cinema tickets in light of the audio quality in the average cinema sounding worse than their home cinema system. Considering that in the United States, movie ticket sales in 2016 were at their lowest volume in 20 years (Miller, 2016), (Thompson, 2016), the industry cannot afford a continued downward trajectory of tickets sales for any reason.

1.4 Main Research Questions

The research presented in this thesis aims to add comprehensive scientific evidence to the growing body of research that is currently being conducted on cinema audio. One of the obvious potential benefits of this research to the cinema community is an improved cinema sound calibration method that could become the basis for revisions to the existing industry standard. Such a revised standard could provide a more consistent, higher quality of sound reproduction where films are mixed and later reproduced in the movie theatre and at home.

The questions that guided this research are:

- What percentages of film sound mixers (re-recording mixers) actually mix to the SMPTE standard? Do they feel it provides a consistent level of sound quality between the mixing/dubbing and theatre playback? In what ways does it fail or succeed in terms of sound quality? What alternative calibrations are employed in its place? Are they compensating for the SMPTE standard and if so, how?

- Does the SMPTE calibration method provide a consistently high sound quality across a variety of different cinema spaces based on objective and subjective sound quality measurements? How well does it perform compared to no calibration or a different calibration method?
- To what extent does the room acoustics with respect to cinema size and seat location influence the perceived sound quality of the different calibration methods?
- For the purposes of cinema calibration, what is the optimal number of microphone locations required in order to achieve accuracy in measurements? Can anechoic data from a loudspeaker better assist in the calibration process?
- Below what frequency does the room acoustics dominate what the listener hears and above which frequency does the direct sound play a dominate role?

1.5 Original Aspects of This Research

The research presented in the thesis is both original and of value to the scientific community, audio professionals and consumers involved in the production, reproduction and enjoyment of film sound. To date, cinema audio calibration has received comparatively little attention from the scientific community. Therefore, the hope is that the findings of this research represent a significant contribution towards a better scientific understanding of the perception and measurement of cinema loudspeakers, their acoustical interaction with the cinema space, and how to better calibrate the system for optimal sound quality. The following outcomes, which are presented throughout the thesis, are considered by the author to be novel contributions to the field of cinema audio calibration:

- A survey of re-recording mixers provided a clearer understanding of their work spaces (dub-stages), how these works spaces were calibrated and how this calibration affected the way which re-recording mixers process audio tracks. This survey is a novel approach to uncover information which can help guide authors of the new standards.
- The Binuaral Room Scanning system, though not a completely novel system, was

refined and validated for cinema-sized spaces. This system allows for researchers to conduct listening test on cinema and theatre audio systems and acoustics when in situ listening experiments prove logistically difficult.

- The first true double blind listening tests for cinema audio conducted to date. The results of these tests provide a basis for questioning the in-room target response curve currently utilized in the SMPTE standard and offers potential replacements.
- A process was developed for calibrating cinema audio systems which begins with the anechoic data from the loudspeakers. This process provides an easily scalable method of calibrating a cinema that allows the technician to ascertain what aspects of the room's response is created by the loudspeaker and which is created by the room itself.

1.6 Intended Audience and Applications for This Research

The author believes that scientific research should be judged not only on its scientific merit and originality, but also its real-world application and value to society. The research topic of this dissertation meets both criteria.

A review of the limited scientific literature available on calibration methods for cinema audio provides sufficient motivation and justification for this research based on the lack of robust scientific methodology applied to the standard currently in use. The potential practical value and application of this research is also quite clear; new knowledge may lead to improvements in the measurement, design, and performance of loudspeakers and listening rooms used in the production and reproduction of audio. Given that a large percentage of society spends a significant amount of time and money viewing movies in both their home and in commercial cinemas, there is a substantial audience that could potentially benefit from research of this kind.

Four different groups may potentially benefit from the author's research: (1) audio engineers involved in the production and reproduction of cinematic audio; (2) cinema-going consumers; (3) scientists working in the areas of product development for the cinema market and (4) cinema owners.

1.7 The Structure of the Thesis

The thesis is divided into seven chapters, and the main goals and themes of each are summarized below.

Chapter 1 introduces the research topic and author's motivation for investigating the history of cinema calibration and to form methods for improving upon the current standard calibration method. The main research objectives are defined, as well as potential applications and beneficiaries of the research. The overall structure of the dissertation and content of each chapter is described.

Chapter 2 provides the literature review for the thesis, summarizing previous and current research into the physical and psychological factors related to cinema calibration. The chapter begins with the origin and history of cinema calibration and surveys the evolution of cinema calibration through to today's standards.

Chapter 3 describes a survey conducted of re-recording mixers in order to obtain opinions of theatre calibration from one of the end-users' point of view. The survey gives an overview of the demographics and types of work spaces of the re-recording mixers surveyed. The survey also provides insight as to adjustments that are being made during the final mixing process.

Chapter 4 describes the BRS measurement and playback method used for the author's experiments, and explains why it is an important methodological tool for research on cinema calibration. Descriptions of the measurement and playback systems are given. Finally, the results of a validation test are presented in which listeners' calibration curve preference ratings, produced in situ, are compared to those produced using the BRS playback system.

Chapter 5 describes a listening test administered to trained listeners in order to obtain insight into preferences in room response target curves. A review of the BRS test system along with the test methodology is described. Results of the listening tests are presented and compared to the current target curve currently utilized for cinema calibration.

Chapter 6 describes a novel method for determining the number of microphone positions required to accurately capture a room's response by using a loudspeaker's anechoic data for comparison. Measurements across 7 different sized cinemas using this methodology are presented. The energy averages of room response measurements ranging from 216 positions to 4 positions were compared to the anechoic loudspeaker data and an analysis is presented.

Chapter 7 discusses the results of the experiments as a collective and the outcome as they relate to the research objective. It also summarizes the main conclusion of the research, identifies some limitations in its scope, and recommends additional areas of research that warrant further investigation.

Chapter 2 - Literature Review

2.1 Introduction

This section contains a review and summary which spans the historic through the current literature concerning cinema audio calibration. As this particular field has not been widely studied by the audio and acoustic communities in general, the literature review consists mainly of a historic perspective of cinema audio calibration and the early development of the SMPTE standard. The review begins with an overview of the most prevalent standard used in cinema audio; the Society of Motion Picture and Television Engineers' *SMPTE ST-202:2010 Standard for Motion-Picture - Dubbing Stages (Mixing Rooms), Screening Rooms and Indoor Theatres - B-Chain Electroacoustic Response*. The review moves into a breakdown, by time periods, of the evolution of the SMPTE cinema audio standard along with that of the allied International Organization for Standardization (ISO) standard.

2.2 Overview of SMPTE Cinema Audio Calibration

One of the largest challenges for film sound mixers (re-recording mixers) is creating a soundtrack that can translate from a dub-stage of one size into commercial cinemas of various sizes. In order to aid the mixer in this task, SMPTE created a calibration method and standard which includes a targeted room response curve for both dub-stages and cinemas that would help to ensure an accurate translation of a film's sound track from dub-stage to cinema.

With regards to the SMPTE cinema audio standard, the cinema reproduction chain is split into two sections; the A and B chains. The A chain makes up that part of a motion-picture audio system starting with the source and extending as far as the input source selector, as shown in Figures 2.1 and 2.2. The B chain comprises the elements that follow the switching of signal sources and includes equalizers, crossovers (if used), power amplifiers, loudspeakers, projection screens, and modifications to the sound caused by auditorium acoustics and distance to listeners, as depicted in Figure 2.3.

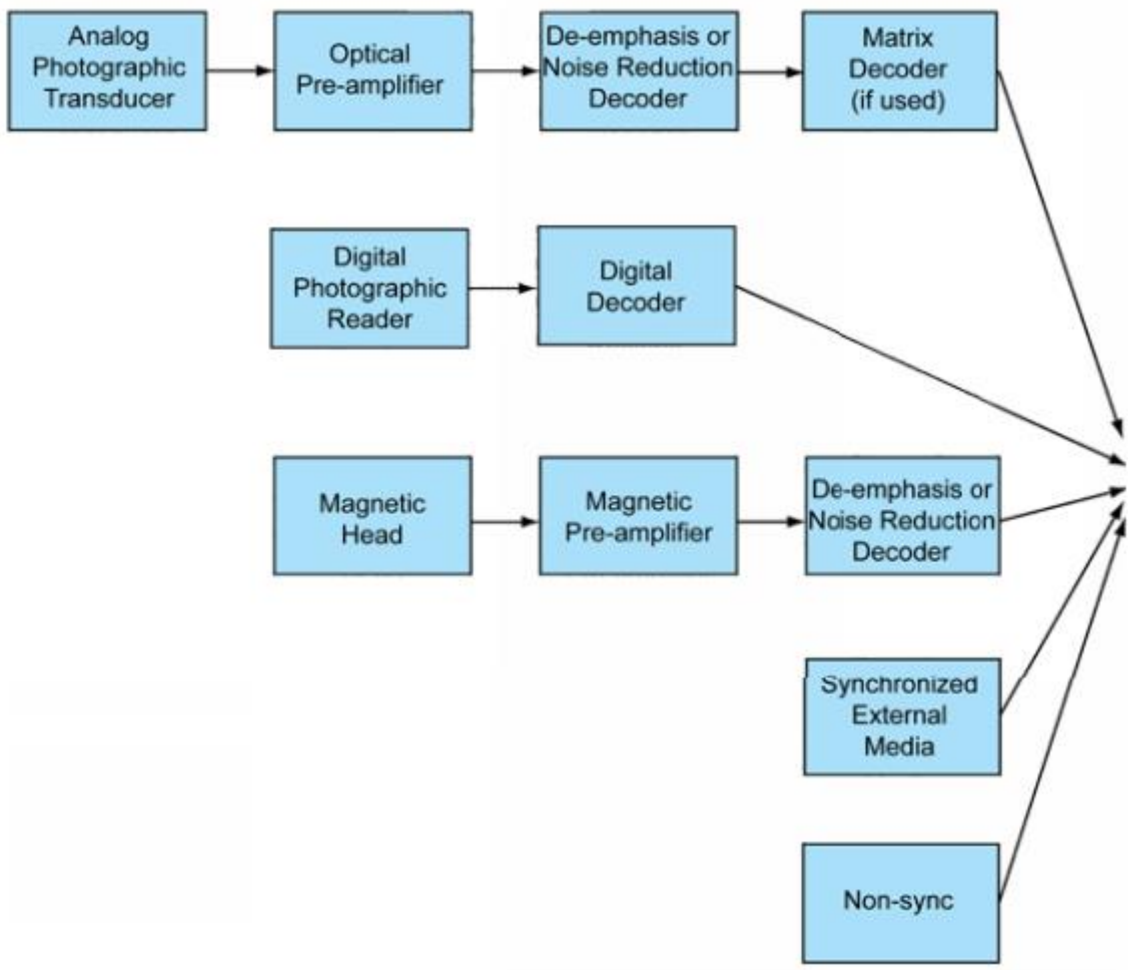


Figure 2.1: A-chain section of analogue theatrical audio reproduction chain; from (SMPTE, 2010)

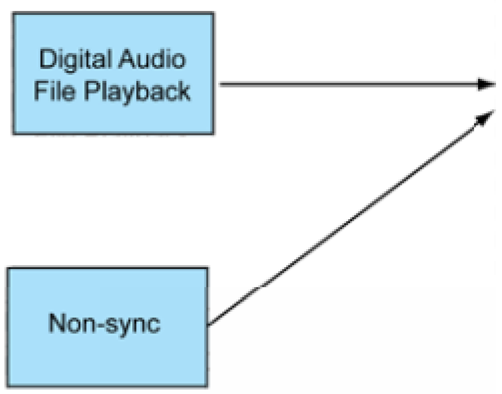


Figure 2.2: A-chain section of digital theatrical audio reproduction chain: from (SMPTE, 2010)

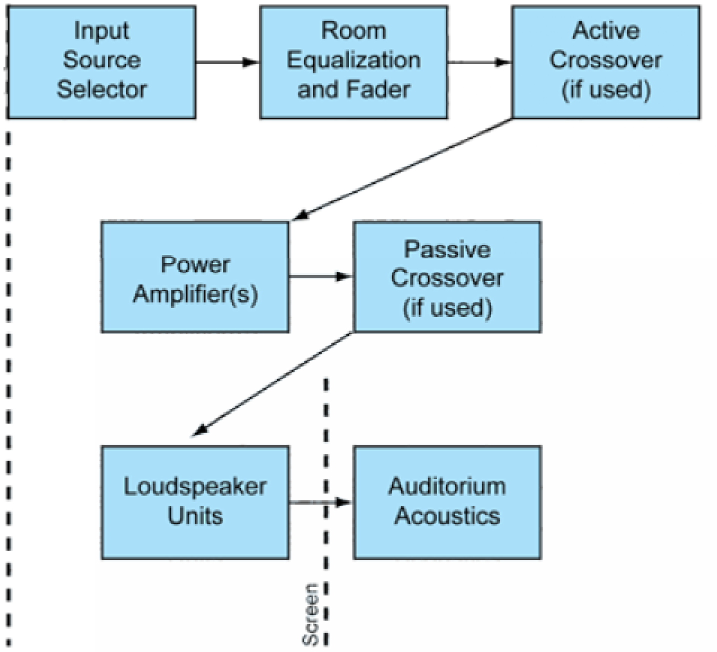


Figure 2.3: B-chain section of theatrical audio reproduction chain; from (SMPTE, 2010)

The SMPTE ‘X-curve’ represents the target shape for the steady state frequency response of cinema sound systems B-chain when measured with pink noise at a reference or series of reference positions using a 1/3rd octave real-time analyzer, as shown in Figure 2.4. The modern day X curve, originally disseminated in ISO 2969 (ISO, 1977) and shown in its current form in Figure 2.5, was subsequently adopted and revised over the years by SMPTE as standard ST202 (SMPTE, 2010) and has been specified for many years. The standard relates only to the B Chain and therefore has no role in electronic pre-emphasis equalization or other signal corrections within the A chain (Allen, 2006), (Rasmussen, 1976).

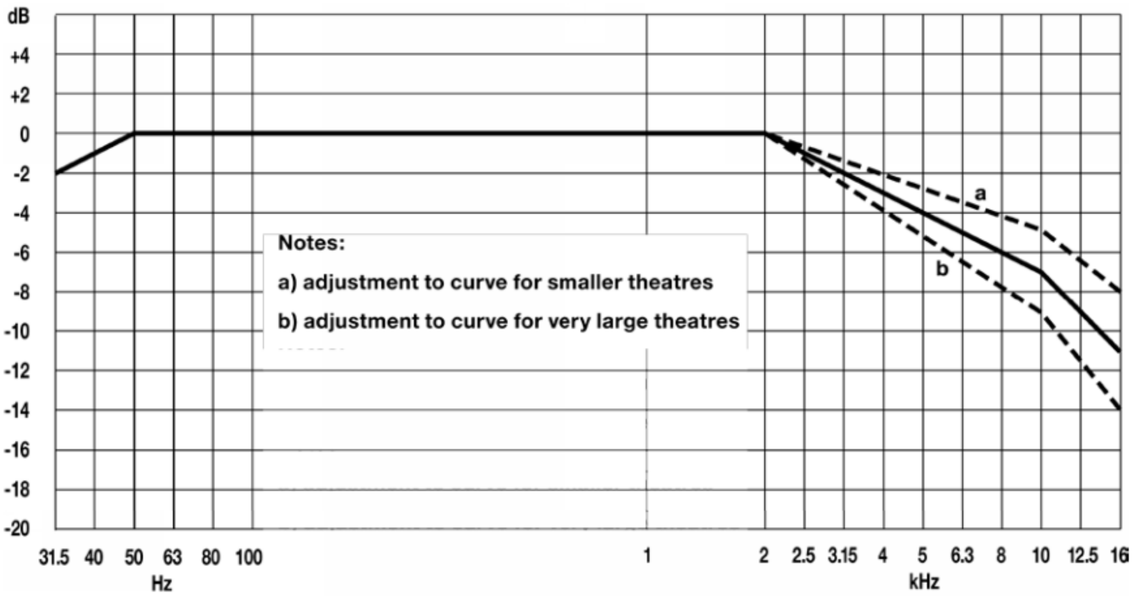


Figure 2.4: B-chain X-curve with modifications for theatre size; from (SMPTE, 2010)

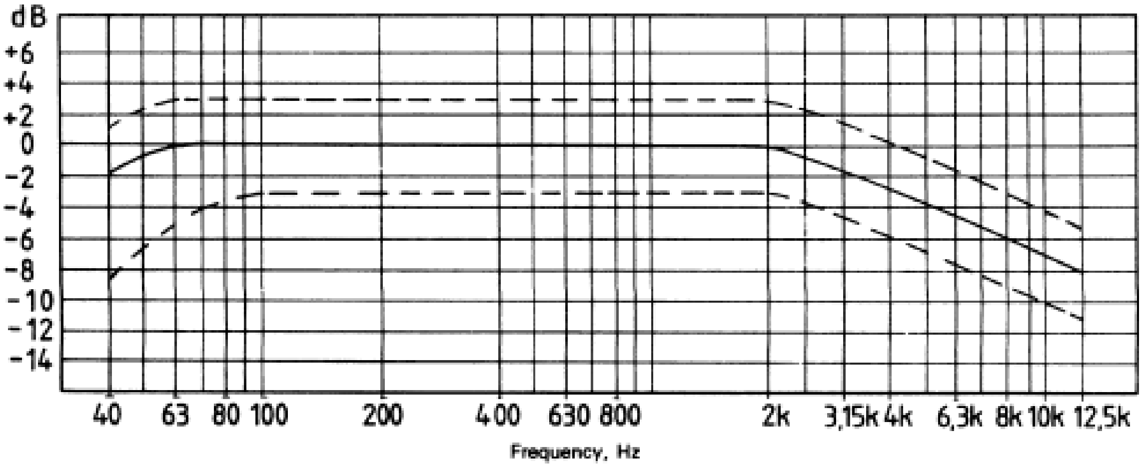


Figure 2.5: Curve X of B-chain characteristic including tolerances; from (ISO, 1987)

According to Allen (2006), the rationale behind the use of the X-curve is as follows:

- To compensate for some psychoacoustic phenomena involving the integration of faraway sound and picture.
- Distortion components in the loudspeaker (typically horn-loaded), making high frequencies objectionable.
- The result of excessive reverberation build-up.
- The lack of power handling in low frequency (subwoofer) speakers leading to audible

distortion.

Currently, dubbing stages and commercial cinemas seeking Dolby certification are equalized as closely as possible to the X-curve at a location roughly two-thirds of the distance from the screen to the rear wall. This area of the room has conventionally been considered the most representative average of the whole room and is generally where the mixing console is located in commercial dubbing stages. Before the advent of Digital Cinema Package (DCP) around 2011, Dolby approval was required before a studio was allowed to make 35mm print-masters[1] with Dolby encoding, which was the primary audio-encoding scheme used throughout the industry. Dolby approval dictated that dub-stages and cinemas be tuned to the X-curve in order to meet Dolby certification for encoding on the production side and decoding on the playback side (Dolby, 1994). The format that now requires Dolby certifications and in turn the use of the X-curve is ATMOS (Dolby, 2015) .

2.3 Early History of the Cinema Calibration

> *A new factor has come strongly into the picture, and I believe that it will call for some radical revisions of our criteria for best acoustics. I refer to the electrical reproduction of sound.* Edward W. Kellogg, General Electric Research Lab, 1930

There is a vast collection of documentation on the early developments of motion picture sound ranging from recording techniques, microphone developments, loudspeakers developments to encoding sound to film, and room acoustics. To write about all of the developments in the early history of sound for pictures would be somewhat superfluous for purpose of this thesis. Therefore, this early history review highlights selected pieces of literature that were deemed by the author to be most the most pertinent to the overall history of cinema calibration.

One of the earliest cinema audio systems of record was developed by Western Electric in 1924, shown in Figure 2.6. This system was comprised of an electrified phonograph player that the

[1] A printmaster is the final composite (dialogue, music and sound effects recorded together) mix of a film that can be transferred directly to a track negative or a magnetic stripe print with no further changes in level or equalization. It combines the various elements into a final composite soundtrack. When this is completed, an optical or digital sound track can be created for a feature film release print

Western Electric engineers were able to synchronize to a camera and projector, with audio playback being delivered by the company's public address loudspeakers. The only film company that showed interest in the system at that time was Warner Brothers, who went on to create a joint venture with Western Electric in 1926 named the Vitaphone Corporation (Thompson, 2004).

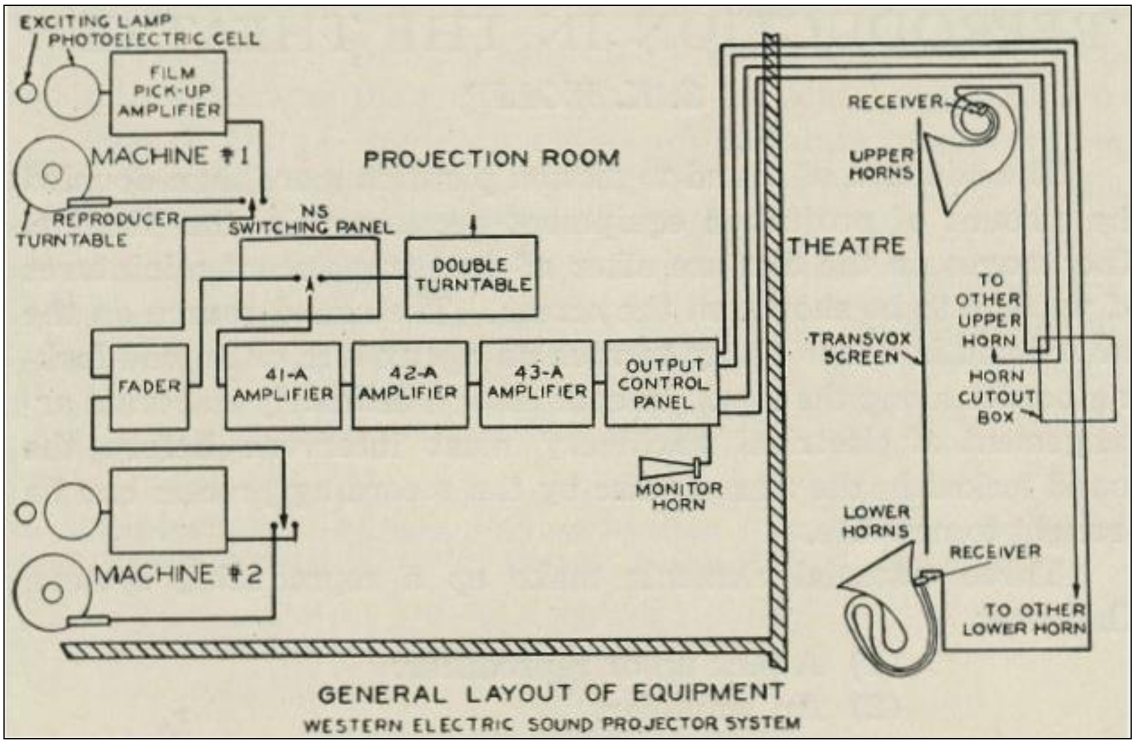


Figure 2.6: Early Western Electric sound projection system; from (Wolf, 1929a)

Scientific journals of the time show evidence that the first concerns with audio playback in cinemas were about room acoustics, as opposed to issues with the audio playback system itself. Many of the early pioneers of acoustic research, including Watson, Sabine, Swan, Wolf, Kellogg and Knudsen, took up the challenge of converting performance halls into "picture palaces" and in improving film stages (now sound stages) for audio recording. Reverberation times of 3 seconds or longer in both the theatres and the sound stages led to the expected loss of dialog clarity. The challenge posed to the acousticians of the day was to take performance halls that were originally designed for live music playback and prepare them for "talking pictures", and to take cavernous film stages and prepare them for dialog recording (Knudsen, 1932), (Kellogg, 1930), (Swan, 1929), (Wolf, 1929b), (Sabine, 1928), (Watson, 1927). Moving into the "Golden Age of Cinema", movie theatres were typically classified by three sizes: the 'average' theatre was 1000 seats, a 'medium' theatre was 2000 seats and a 'large' theatre was 4000 seats or more. These theatres usually had reverberation times lower than the earlier vaudeville theatres that had been converted for cinema, but still had reverberation times between 1 and 3 seconds throughout the frequency range (Loye, 1941), (Shortt et al., 1940), (Potwin and Schlanger, 1939).

As the acoustics of cinemas began to improve with regard to reverberation times, the focus shifted towards the effects that the combined room and loudspeaker performance had on the perception of film sound playback. Some of the earliest research in this area can be traced back to Wente, whose results were originally published in 1935 (Wente, 1935), (Genereux, 1992). During this time period, the quality of electronic signal transmission circuits could be tested by measuring the steady state transmission as a function of frequency. Wente felt that the same could hold true of the quality of sound transmission in a room. If the room were treated as if it were part of a transmission system much like an electronic circuit, then its quality could be evaluated by the same principals. Using a variable frequency oscillator and a high speed level recorder, Wente was able to record the transmission versus frequency characteristics between points in rooms having different acoustic characteristics. In measuring fluctuations in sound pressure at one point in a room by generating sound waves of various frequencies at another point in the same room, he was able to determine not only the reverberation time, but the 'quality' of the reverberation by octave and position-specific effects as well.

Parallel to Wente's work was that of Bedell and Kerney, whose work concentrated on the adaptation of the sound system in one concert hall (Constitution Hall in Washington D.C.) to faithfully reproduce an orchestral recording completed in another (the American Academy of Music in Philadelphia) (Bedell and Kerney, 1934). In essence, Bedell and Kerney's work was one of the first recorded experiments that attempted to achieve ideal sound 'translation' between rooms. In Bedell and Kerney's work, the playback hall was tuned using an oscillator which could reproduce frequencies from 35Hz to 15kHz. The same microphone used to record the orchestra was placed in several different positions throughout the hall with the acknowledgement that the frequency response of the microphone itself would be negated and there would be seat-to-seat differences due to varying acoustic conditions. Each of the three loudspeakers used to recreate the orchestra were checked independently at each measurement position. Using the curves recorded, the system was adjusted not only for uniform coverage but also to give an over-all 'flat' response characteristic by means of an equalization network permanently installed between the amplifier and the loudspeaker. Noting the variations between measurement position, Bedell and Kerney realized that compromises in the design of the corrective equalization network needed to be made in order to accommodate an acceptable response at as many of the seating positions as possible. The two researchers went on to observe that the most expeditious way to design an equalization network that would work for most of the seating area would be to use the frequency response curves obtained from the microphone positions closest to the playback loudspeakers. The final part of Bedell and Kerney's adaptation system was to design a set of "quality control networks", which put the final response of the sound system in the hands of a director seated in the theatre; the equivalent of our modern day front-of-house mixing engineer. The frequency response characteristics of these networks are shown in Figure 2.7.

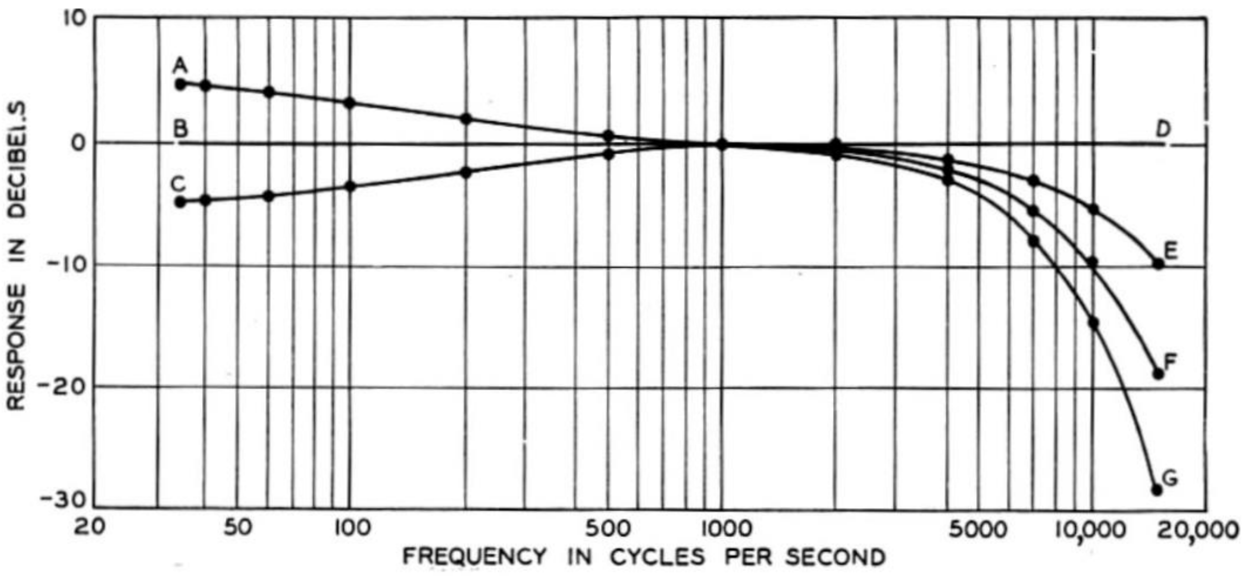


Figure 2.7: Frequency characteristics of quality control networks; from (Bedell and Kerney, 1934)

As the frequency response of film sound recordings was improving, attempts were made to also improve loudspeaker performance. The most notable advance in loudspeaker technology was the Shearer two-way system developed by Douglas Shearer, who was then head of the Metro-Goldwyn-Mayer sound department. MGM drove the design of this loudspeaker as the studio was unhappy with the Western Electric system that was widely in use and which could only produce a limited range of approximately 100Hz to 5kHz (Eargle, 2007), (Hilliard, 1985).

Hilliard wrote about the formulation of requirements to be met by sound systems in motion picture theatres and noted that the Shearer two-way horn system was engineered to meet the high standards which he proposed (Hilliard, 1936). Two of the criteria proposed were:

> *Flat Over-All Frequency Characteristic* - The system shall not deviate by more than +/-2 db from 50 to 8000 cycles over the entire angle of distribution within ten feet of the mouth of the horn.
>
> *Suitable Angular Distribution Characteristics* - The sound shall be radiated through a horizontal angle as great as 110 degrees and a vertical angle of 60 degrees, with nearly uniform response at all positions.

Figure 2.8 depicts the measured frequency response of the Shearer two-way horn system.

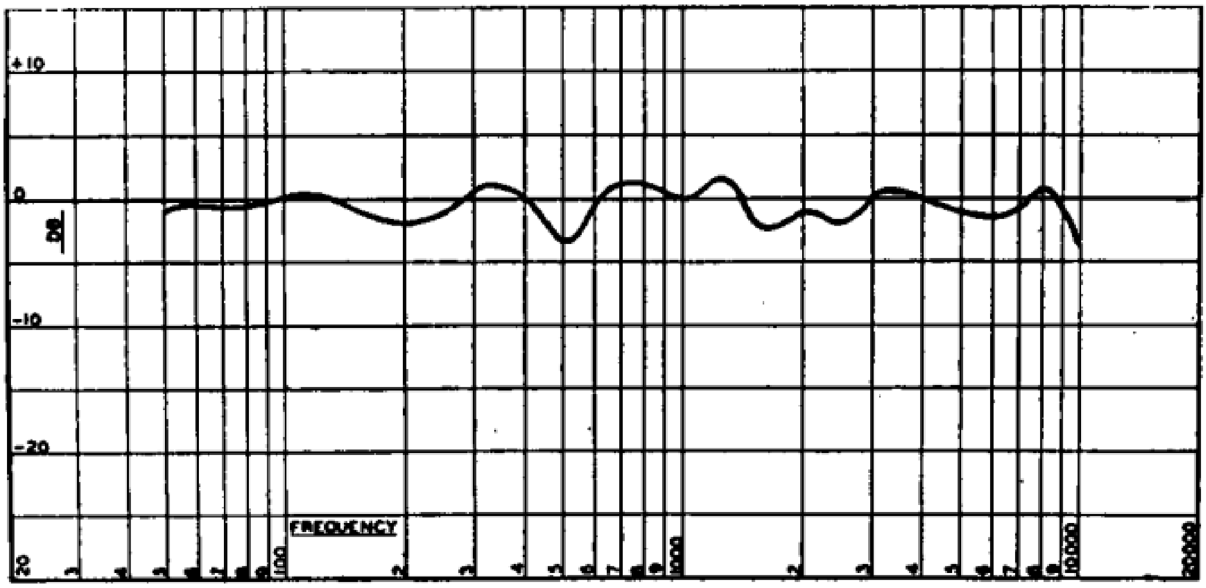


Figure 2.8: Output characteristic; Shearer two-way horn system, measured on normal axis, 10 inches from horn; from (Hilliard, 1936)

However, older existing loudspeakers were still deficient in frequency response especially in the high and low frequencies. To compensate for loudspeaker shortcomings, equalization much like the type used by Bedell and Kerney were introduced into the cinema amplifiers. Unfortunately, most cinemas had older amplifiers that were still performing satisfactorily and so cinema owners had to settle for an easy and inexpensive solution. This solution was to invite a seasoned audio engineer to sit in the auditorium while listening to a test reel of film and continuously adjust a set of variable frequency control networks. When the engineer was finally satisfied with the sound, the frequency corrections were measured and the necessary filter circuits were built permanently into the existing amplifier system (Aldred, 1997).

In 1937 the Motion Picture Research Council (MPRC), a division of the Academy of Motion Picture Arts and Sciences (AMPAS), convened a listening panel to audition various playback materials over typical cinema loudspeaker systems in an effort to determine the optimum playback criteria. The goal was to standardize sound across varying types of cinemas which often had different types of equipment and acoustic characteristics. The council also sought to determine improvement to the sound quality across all cinemas. Most notable of the determinations to be made was the amount of high-frequency attenuation required to compensate for the often shrill characteristics of older loudspeakers still installed at the time, the valve amplifiers, which had an audible hiss and the high-frequency noise (hiss) created by the film audio track itself (Hilliard, 1985).

Mono audio from the optical sound track resided on the edge the film stock (figure 2.9) and the thin space it was given limited the signal. Increasing the width of the optical track gave the signal a larger dynamic range but with less high frequency content, whereas a smaller optical track gave a higher frequency response but with a lower signal level and greater noise. As a compromise, the high frequency content of a film soundtrack was typically limited to 6kHz.

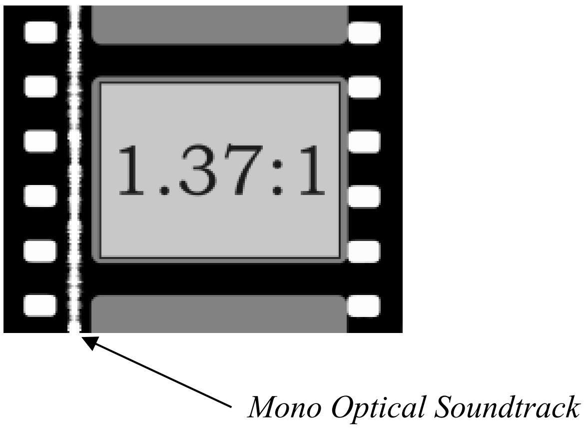


Figure 2.9: Standard Academy mono optical soundtrack; from (Bibra, 2008)

Additional issues with the high frequency content of film sound came from the slit-width of the optical reader used at the time, shown in figure 2.10. The slit plates were made from brass by hand and obtaining perfectly clean, straight edges on the sides of the slits were inherently difficult. Irregularities in the slit lead to further distortions of the high frequency portion of the optical track.

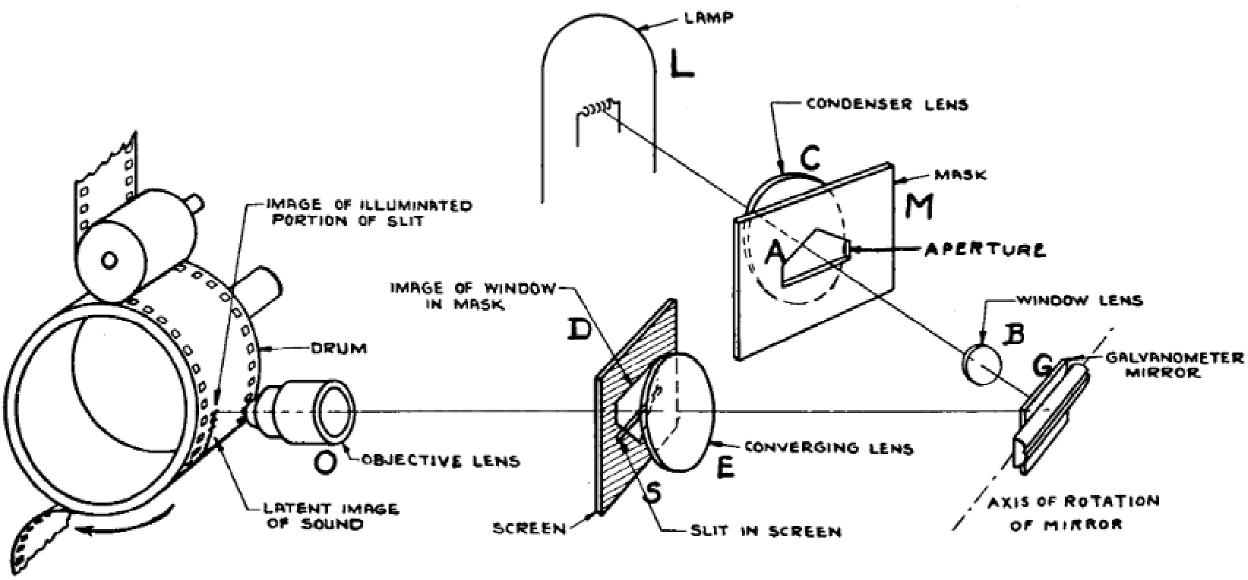


Figure 2.10: Optical mono audio track reader system, slit plate (screen) shown in 'D'; from (Kellogg, 1955)

Attempts to reduce hiss and other high frequency distortions included placing a filter at the output of the pre-amplifier, but this further reduced the high frequencies. The simplest solution developed at the time was to pre-emphasize the high frequencies when recording, and allow the play back noise reduction filter to achieve as close to flat response as possible to provide some consistency between the recording on film and playback in the cinema (Fielding, 1983), (Kellogg, 1955).

To aid the MPRC in determining the best solution for a filter response curve that could be used across all cinemas, a series of listening tests were convened. A test film reel was prepared containing samples of dialogue and music from each of the major studios which were made "under average and extreme conditions by each studio sound department". Members of the listening panel would repeatedly run the test reel in 6 Hollywood theatres which were all considered to have differing loudspeaker systems and acoustic characteristics. It was felt that any standard that would fit these 6 theatres could be applied to a majority of theatres throughout the country. Prior to the listening sessions, a set of test filters were installed at the input of the amplifiers in each of the theatres. During the sessions, these filters were varied and the committee determined an optimum electrical characteristic which would characterize this group of theatres. The process of documenting a typical electrical characteristic was carried out by playing a flat frequency test tone encoded onto the film and played over the projectors in each of the 6 theatres. The signals at the output of the amplifiers were then measured and the amalgamation of the noted frequency responses became the first response curve to be adopted as a standard; 'The Academy Curve'. This standard was first adopted in March of 1937, was revised in June of that same year, with a final version being published in October of 1938 (MPRC, 1937a, MPRC, 1937b, MPRC, 1938). The Academy Curve (also known as the "Normal Curve" or "N-Curve") was defined as flat 100 Hz - 1.6 kHz, down 7 dB at 40 Hz, down 10 dB at 5 kHz, and down 18 dB at 8 kHz. The Academy Curve was actually comprised of two slightly different curves, the sharper of the curves being used for systems with metal diaphragm loudspeakers and flatter curve to 5kHz for systems with Bakelite diaphragms, as shown in Figure 2.11.

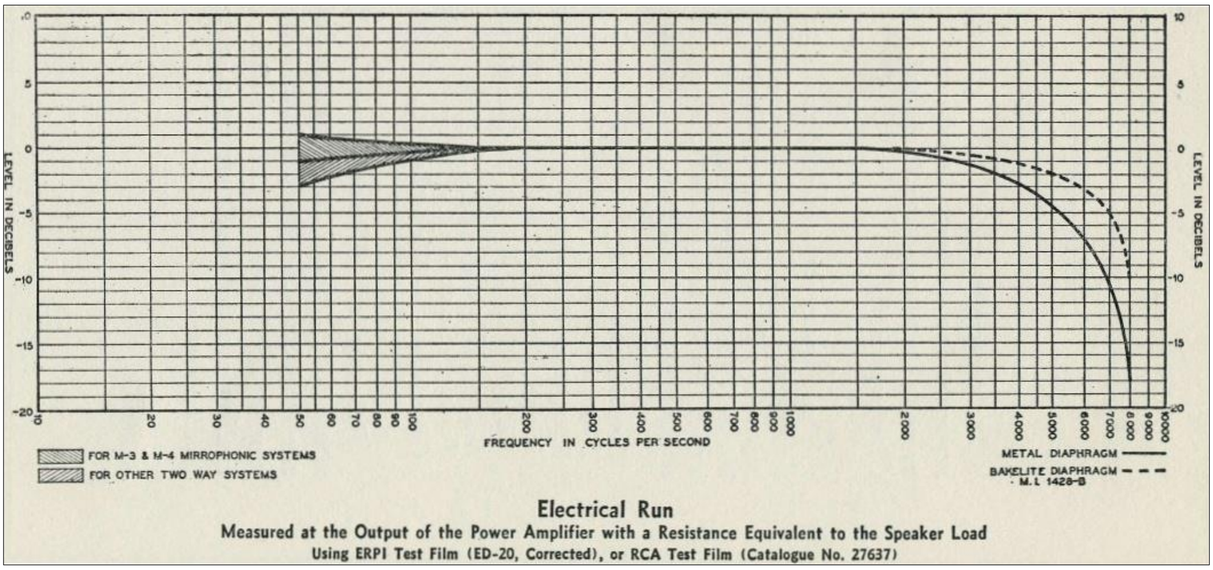


Figure 2.11: 1937 The first 'Academy Characteristic' - two curves, one for loudspeakers with Bakelite diaphragms (dashed line), one for metal diaphragms (solid line); from (MPRC, 1937a)

It was only an electrical curve measured at the amplifier output, and did not take into account loudspeaker performance or cinema acoustics. The Academy curve, though severely rolled off on the high end, imposed very little destruction of the original soundtrack as almost no information above 6kHz to 7KHz was typically recorded (Fielding, 1983).

In an effort to measure the electro-acoustic response of a cinema playback system, a method referred to as the "Academy" technique was used. This method measured the frequency response at the output of the amplifier and then used subjective judgment to assess the response in the audience seating area. Durst and Shortt felt that the newly published Academy standard, which only compensated for the playback system to the output of the amplifier, did not take into account important aspects, such as the entire production (recording) chain, the loudspeaker and the theatre acoustic (Durst and Shortt, 1939). They believed that the standard should cover the entire audio path, from the microphones used in the recording process through to the playback in the theatre itself, and should be defined by acoustical measurements in the theatre. During their research, Durst and Shortt experimented with several curves measured in situ and found that the curve shown in figure 2.12 was the most pleasing.

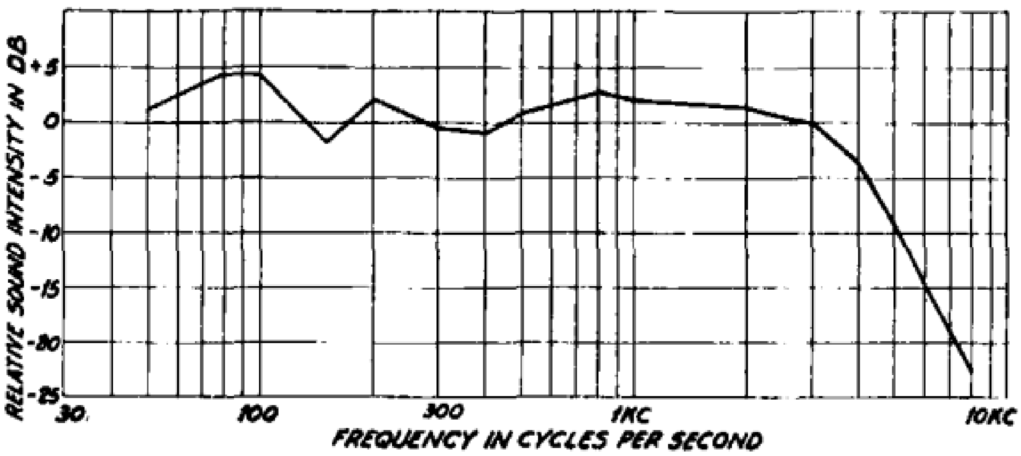


Figure 2.12: Overall acoustic characteristic adjusted for "more pleasing response"; from (Durst and Shortt, 1939)

In 1948 The Motion Picture Research Council issued a technical bulletin titled "Standard Electrical Characteristics for Theatre Sound Systems" (MPRC, 1948). This document provided seven different curves, each curve being optimized for a different group of loudspeakers. Figure

2.13 shows the curve associated with the Altec Lansing Systems, one of the most widely used theatre loudspeaker at that time.

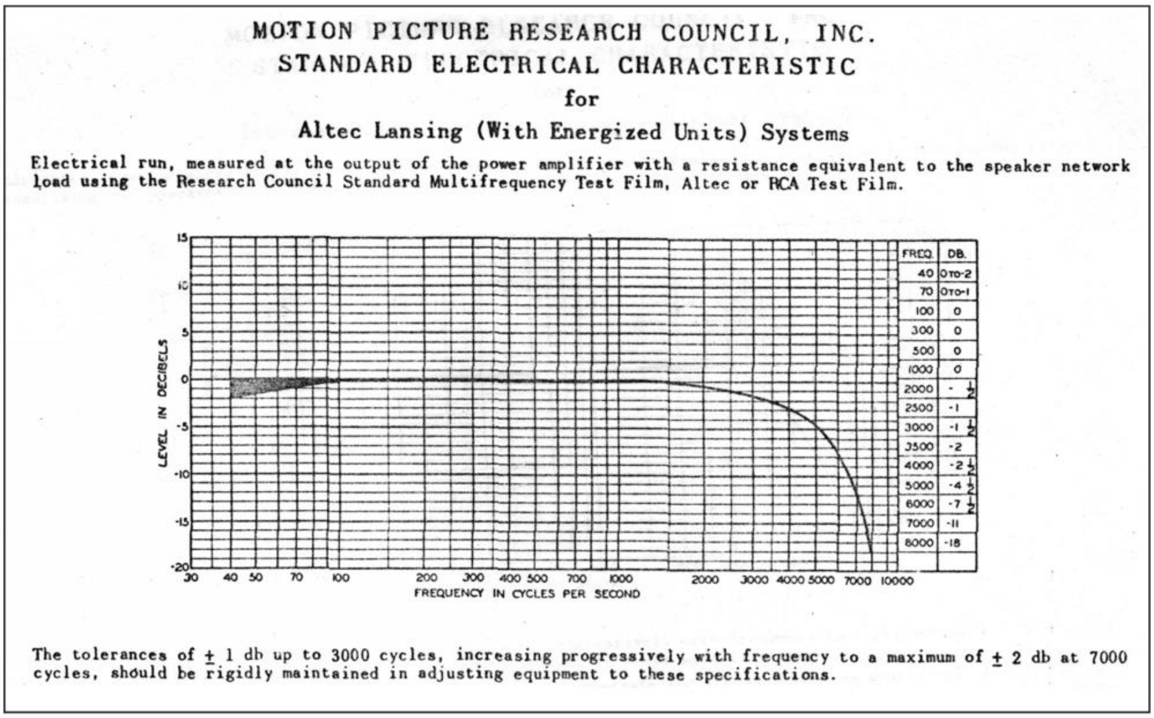


Figure 2.13: 1948 "Academy Characteristic" for the Altec Lansing Systems;

from (MPRC, 1948)

Each of the loudspeakers described in this document were typically two-way units with a passive crossover at 500Hz. Resister networks or "pads" were used to apply the low-frequency or high-frequency adjustments as needed. At the time, this was the only spectral correction available to help rectify a particular cinema's acoustic characteristics. Typically, the technician or engineer working in the cinema would use a piece of frequency response test-film and check the electrical response at the output of the amplifier. He would then listen to typical program material and check the tonal balance. If the technician felt that the tonal balance was askew, they would adjust the pads at the loudspeakers. The bulletin contained an interesting statement:

> Whenever such conditions exist that the particular characteristic recommended does not give satisfactory results, it is recommended that the acoustic characteristics of the auditorium be corrected. (MPRC, 1948)

It appears that even over 60 years ago there was an acknowledgement that equalization, in its primitive form, could not fully compensate for acoustic related issues.

2.4 Room Equalization and Calibration from 1950 through 1970's

From this early work in 1937, and again in 1948, very little additional research was conducted in cinema audio, and the early Academy response curves were utilized without question for over 20 years. Much of the lack of research during this time period was due to technical staff being reassigned to the war effort during World War II. Hilliard had written about the advancements in loudspeaker frequency response in his 1949 papers and noted that MPRC had labelled this increase in frequency range "desirable" (Hilliard, 1949). However, though the newer loudspeakers had a 'flat' frequency response extended out beyond 5kHz, the Academy curve remained with a roll-off starting at 2kHz.

Though research related specifically to cinema audio slowed down during this period, work regarding audio system calibration continued. Rudmose introduced the idea of using systematic measurements of a loudspeaker in its indoor environment in order to optimize system playback within a given room (Rudmose, 1958). He noted that systems were typically installed as-purchased, and then tuned by ear without taking additional measurements in situ. Rudmose felt that this lead to a less than desirable playback result, with little understanding of how the loudspeaker was performing and that more attention should be given to the overall response as measured in the room. Instead of using sine-wave oscillators, which were widely used for system tuning at that time, Rudmose recommended narrow-band noise (30Hz-50Hz) of which the centre frequency could be varied. He also proposed using wide band (white) noise and then to utilize a narrow band analyzer to assess the results or to use white noise that is 1/3rd octave filtered. Much of the equalizing of audio systems in this era was to remove room resonances, but Rudmose felt that his technique could also be used to shape the system response to correct for deficiencies in the room, such as excessive high frequency absorption. During the course of his research, Rudmose tested for the uniformity and distribution of sound energy from the loudspeaker though out the room. He conducted this work in an effort to better understand the effects of loudspeaker placement on in situ measurements.

Carlisle and Schwartz pushed Rudmose's work one step further by introducing a multiple microphone array that allowed sound to be measured in several places within a room simultaneously and added together to get a final result (Carlisle and Schwartz, 1959). These two researchers noted that in situ measurements with a multiple microphone array gave much better data at low frequencies than the anechoic data and correlated much closer to subjective listening evaluations. Snow continued with this line of research and worked with comparing anechoic (outdoor) loudspeaker measurements to indoor. He opined that testing loudspeakers under ideal conditions was difficult but "...when tests are made in room, test conditions become even worse and more ambiguous." (Snow, 1961) Snow concluded that both anechoic and in situ data were required in order to understand the loudspeaker / room combination. Anechoic data allowed loudspeaker deficiencies, which may have been masked by room effects, to be observed and in turn, in situ measurements provided data on room effects. Snow conceded that the use of in situ measurements was still clouded by the lack of knowledge as to which aspects of room effects had influence over how the sound was perceived. Part of the results from Snow's work is shown in figure 2.14.

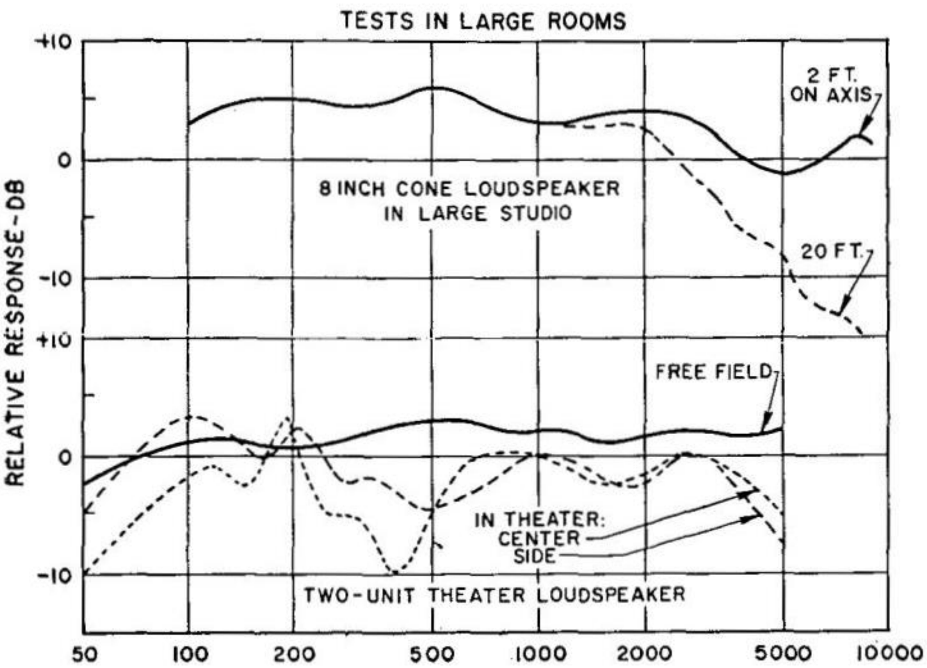


Figure 2.14: Effects of loudspeaker directivity, free-field (outdoor) on axis versus reverberant field; from (Snow, 1961)

Of interest to note is the apparent seat-dip effect shown in the centre test position which appears to be filled in by lateral reflections at the side position. The rise in low frequency content between the centre and side positions may also be evidence of a boundary condition.

Early attempts to equalize the electro acoustic chain from microphone to listener in varying types of spaces were performed and documented by several researchers (Boner and Boner, 1965), (Boner, 1966), (Connor, 1967), (Queen, 1971). These experiments were typically carried out in an effort to improve feedback in reverberant spaces such as churches, hotel ballrooms and airports. Some of these studies advocated for the use of equalization as a means to improve the overall characteristics of a loudspeaker within a room, such as to improve 'naturalness' and speech intelligibility. Though different from cinema audio systems, they did lay the groundwork for future cinema calibration by considering the entire audio playback chain, including the room as opposed to the just the electronic portion of the signal chain. Current literature on cinema audio still points towards these early works as justification for portions of the SMPTE standard that involve the 'shape' of the room response (Allen, 2006), (Holman, 1991).

The early works noted that, even though the electronic devices in the signal chain could be equalized with basic resistor / capacitor circuits to obtain a flat response at the output of the amplifier, the most complex and often erratic portion of the sound system was typically the room itself. With work centred on the desire to reduce room-ring modes and to lessen sound system feedback, there were discussions as to whether a perfectly flat overall room characteristic was desirable. Rudmose, Connor and Queen advocated for the use of equalization to create as flat of an overall room response as possible, however, their research did not involve verifying this was subjectively beneficial. Three researchers who conducted subjective research with regards to defining a house curve, and whose work is often cited when discussing the rationale behind the high frequency roll-off of the X-curve, are the Boners and Schulein (Allen, 2006), (Holman, 1993), (Holman, 1991).

The Boners observed that, within their research, a 'flat' room response did not render that best subjective playback environment. It was discovered that when the room response was shaped by basic equalization for a flat electro-acoustic frequency response, that the room was considered by listeners to be too bright when auditioning well-balanced program material. Experimentation by the Boners found that a high frequency roll-off by as much as 10dB at

10kHz in some rooms was preferred. It was during their experiments to minimize feedback and room ring modes that the Boners first established the idea of a 'house curve'. Of interest to note, the research they conducted was not focused on finding or creating a preferred house curve but was a consequence of experiments attempting to rid large reverberant spaces, such as churches, of feedback issues created by a combination of the sound system and the room acoustics. The Boners various works centred on highly reverberant rooms with volumes of 11,000 to 56,600 cubic meters and reverberation times of between 1.0 and 5.0 seconds. The rooms were 'voiced' by utilizing an oscillator driving the sound systems and each resonant mode which drove the system into feedback was noted. These frequencies were then carefully notched out of the systems by creating individually tuned narrow band custom resister / capacitor filters and inserted between the line amplifier and the power amplifier. It was found that when the problematic room ring modes were filtered, that the average sound pressure over the seating area was relatively uniform as shown in figure 2.15.

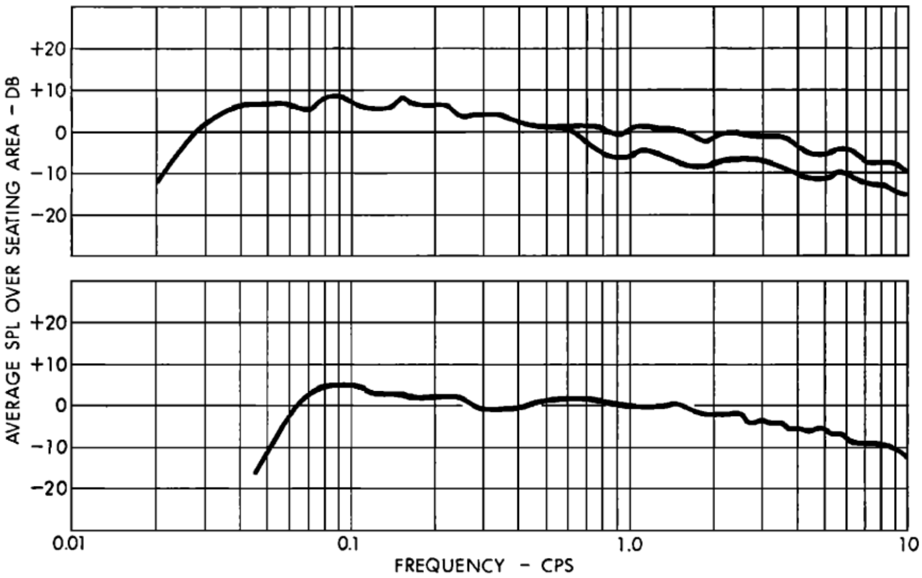


Figure 2.15: Typical house curves after equalization and filtering of major feedback modes; from (Boner and Boner, 1965)

During the Boners' work they made informal audience preference studies between house curves to determine which curve may be preferred. They found that an upward tilt towards high frequencies was not desired but that an upward tilt towards the low frequencies, such as the lower curve in figure 2.15, was preferred. The Boners continued with their preference survey with several of the systems they installed and found that the reaction to having the bass rolled of was negative; that "people seem to like big bass."

Though the Boners' practice was for speech reinforcement systems, which did not require theatre-to-theatre uniformity in the same way that cinema does, the concept of a house curve can be traced back to them. Further research by Robert Schulien (1975) somewhat concurred with the findings of the Boners.

Schulein's work considered not only sound reinforcement systems, but studio monitoring and home high-fidelity systems as well. His experiments questioned the influence of the listening room on perceived tonal balance and the psychoacoustic considerations necessary to correlate subjective and objective data with regards to a desired tonal balance. Schulein noted that the vast amount of research on loudspeakers had been towards producing a flat response, whether in an anechoic chamber, under reverberant conditions or in situ, and on the accuracy of the measurement techniques used. He opined that subjective ramifications of a flat loudspeaker response, especially those in situ, are not often covered to any degree and that they are rarely supported by experimentation due to the fact that such experiments are complex and time consuming. Schulein pointed towards the 'house curve' found in the Boners' research but did not represent it accurately, as shown in figure 2.16.

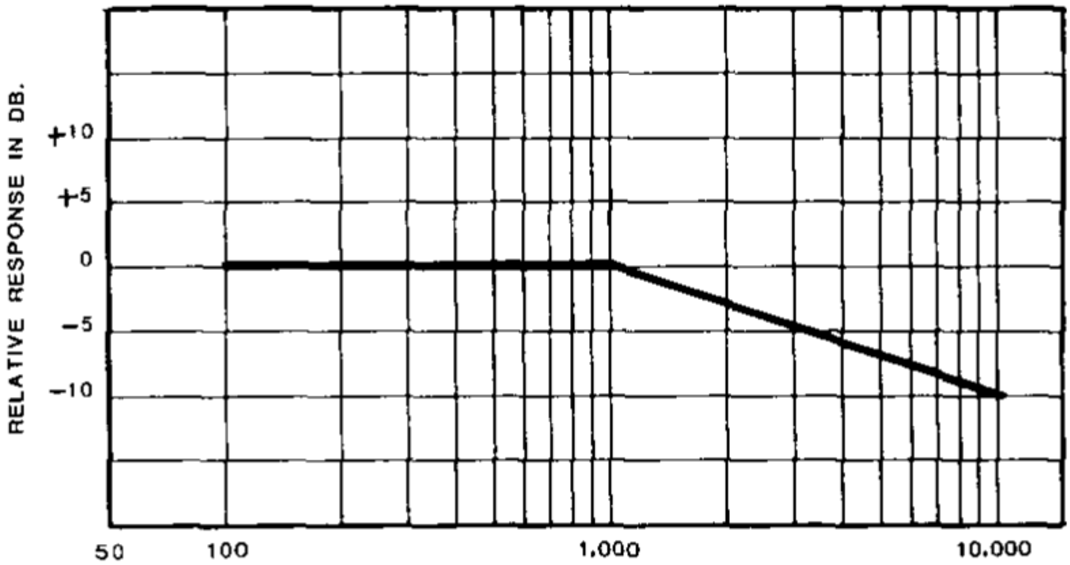


Figure 2.16: Schulein's interpretation of Boners' 'house curve'; from (Schulein, 1975)

The curve is figure 2.16 shows a sharp knee at 1kHz and no rise in level towards the low frequencies, which was present in the Boners' curve. Schulein conducted his own set of experiments which encompassed placing a loudspeaker at 1.2 meters from the listener (near-field) and the same loudspeaker at 15.2 meters (far-field). Listening to music and speech, the listeners were asked to adjust an equalizer in order to match the far-field loudspeaker to the near-field. The final results were obtained by playing pink noise through the far-field loudspeaker and recording the results by placing a microphone where the listener had been sitting. The results of this test are shown in figure 2.17.

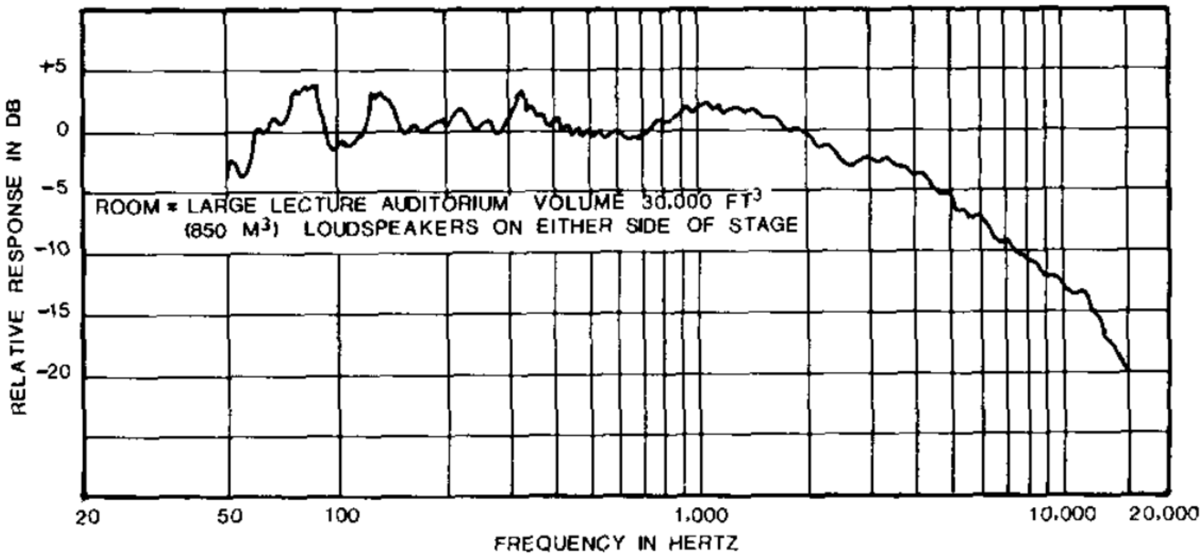


Figure 2.17: Example of a subjectively adjusted house curve normalized to remove the microphone; from (Schulein, 1975)

There are several issues with how these results were derived which include:

1. The reverberation characteristics of the test room were not reported
2. Only 4 listeners were used
3. Only one example of the 4 listeners was reported
4. The test was not double or single blind
5. The frequency response characteristics of the test loudspeaker were not reported
6. The spectral characteristics of the test source, music and speech, were not reported
7. The medium on which the test source resided (tape or vinyl) was not reported
8. The listeners were only given a one-octave equalizer to make adjustments on

9. The measurements of the final results were recorded using pink noise and 1/3 octave analysis
10. The normalization curve for the microphone was obtained by measuring the microphone 15 centimetres from the loudspeaker
11. Loudspeaker placement with respect to adjacent boundaries was not reported

Schulein repeated this experiment in presumably the same auditorium, but this time made the comparison between a single near-field loudspeaker and a pair of different model loudspeakers in the far-field. Instead of setting the near-field loudspeaker as the baseline with which to adjust the far-field loudspeaker to, the test subject was asked to adjust the level of 1/3rd octave noise while alternating between the two loudspeaker systems until they achieved an equal loudness match. After the test subject completed adjustments for the entire frequency band, the settings were recorded and a microphone was again placed where the subject's head had been and the level of each loudspeaker system was measured. Schulein noted that the result revealed that, as the test frequency was increased the level produced by the far-field system needed to be decreased in order to produce equal loudness, as shown in figure 2.18.

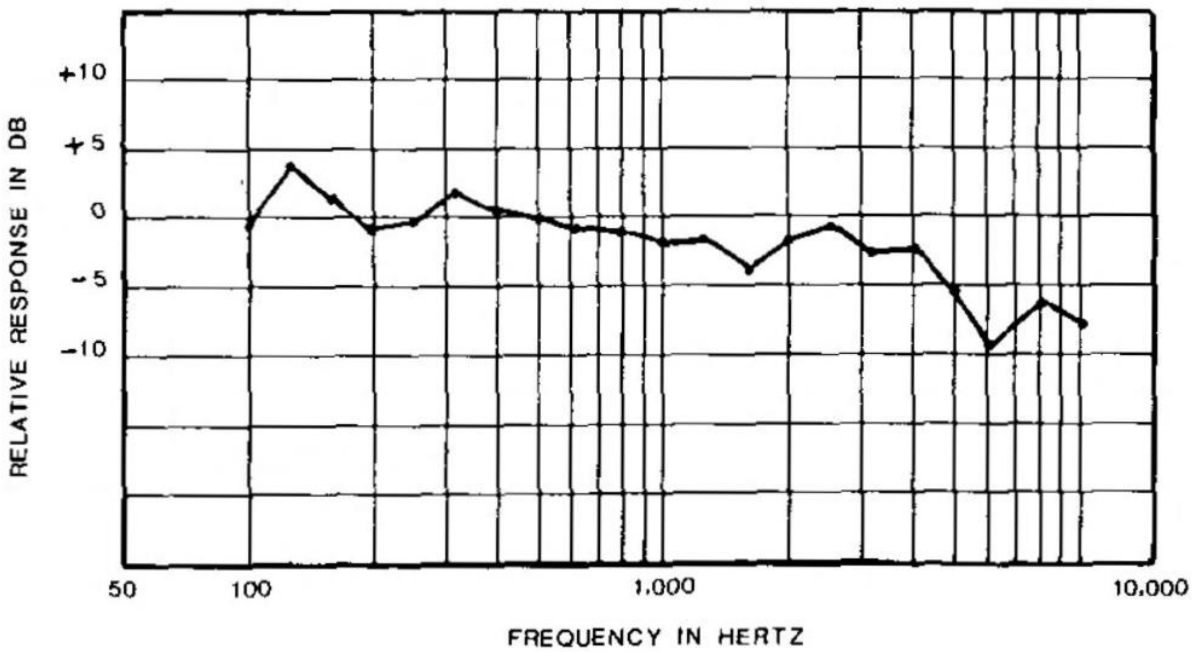


Figure 2.18: Sound pressure level of far source with respect of near source necessary for an equal loudness balance on 1/3rd octave basis; from (Schulein, 1975)

Note that, like the first experiment, this experiment lead to not only a high frequency roll off but a rising level towards the low frequencies.

Many of the same issues that were noted on the first experiment were still present in the second experiment. Schulein concluded that there were characteristics of the human auditory system that would partially account for the large amount of high frequency roll-off. The main characteristic pointed out was that of human hearing sensitivity to high frequencies in a diffuse-field. The data he used is depicted in the lower curve of figure 2.19.

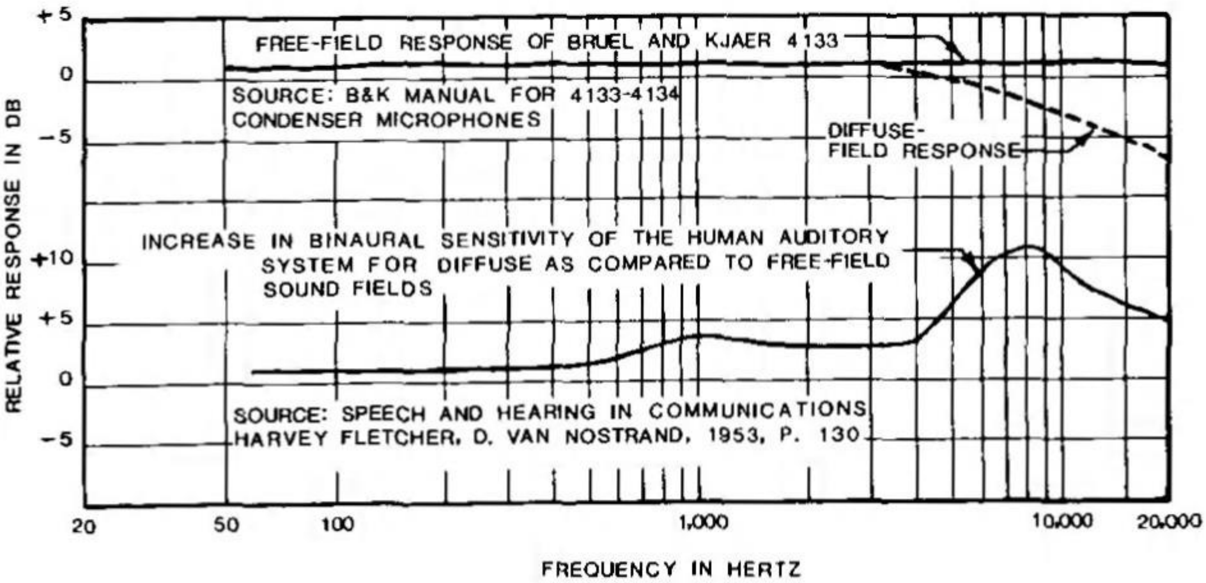


Figure 2.19: Diffuse-field versus free-field response characteristics of a typical human listener (lower curve); from (Schulein, 1975)

Note that Schulein references the diffuse versus free-field correction of Fletcher and Van Nostrand from 1953 though the ISO 454 standard available at that time depicts quite a different curve; figure 2.20.

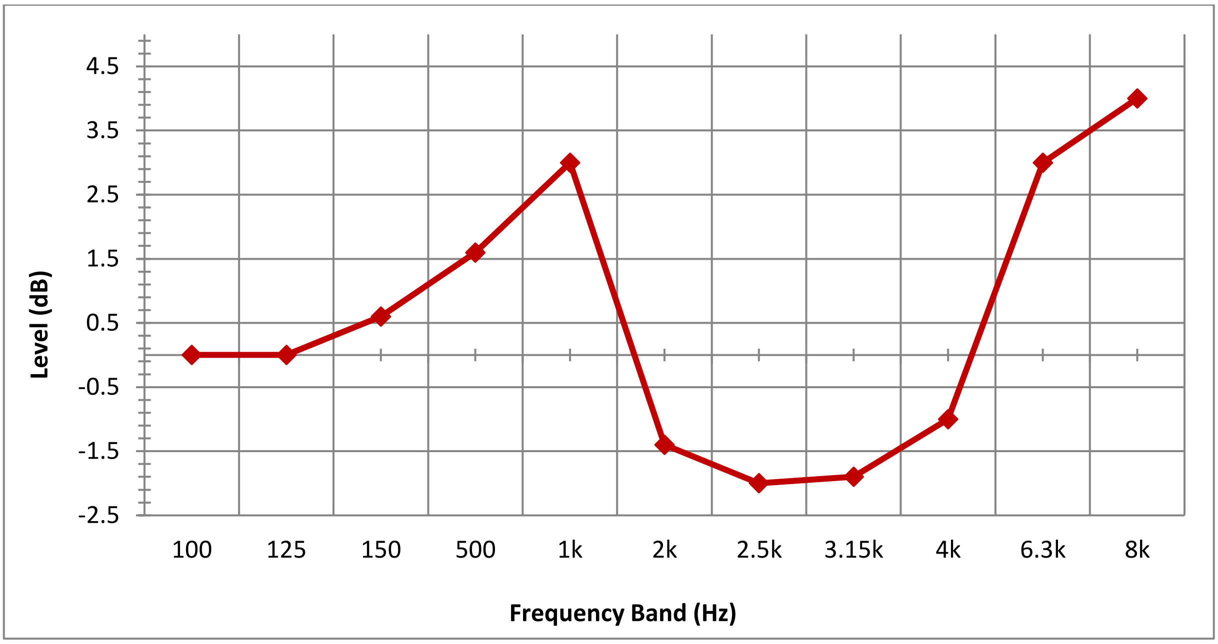


Figure 2.20: Diffuse-field correction; derived from (ISO, 1975a)

As an aside, if the diffuse-field sensitivity of the human auditory system is to be compensated for in the shape of a current calibration curve, the information shown in figure 2.21 should be considered.

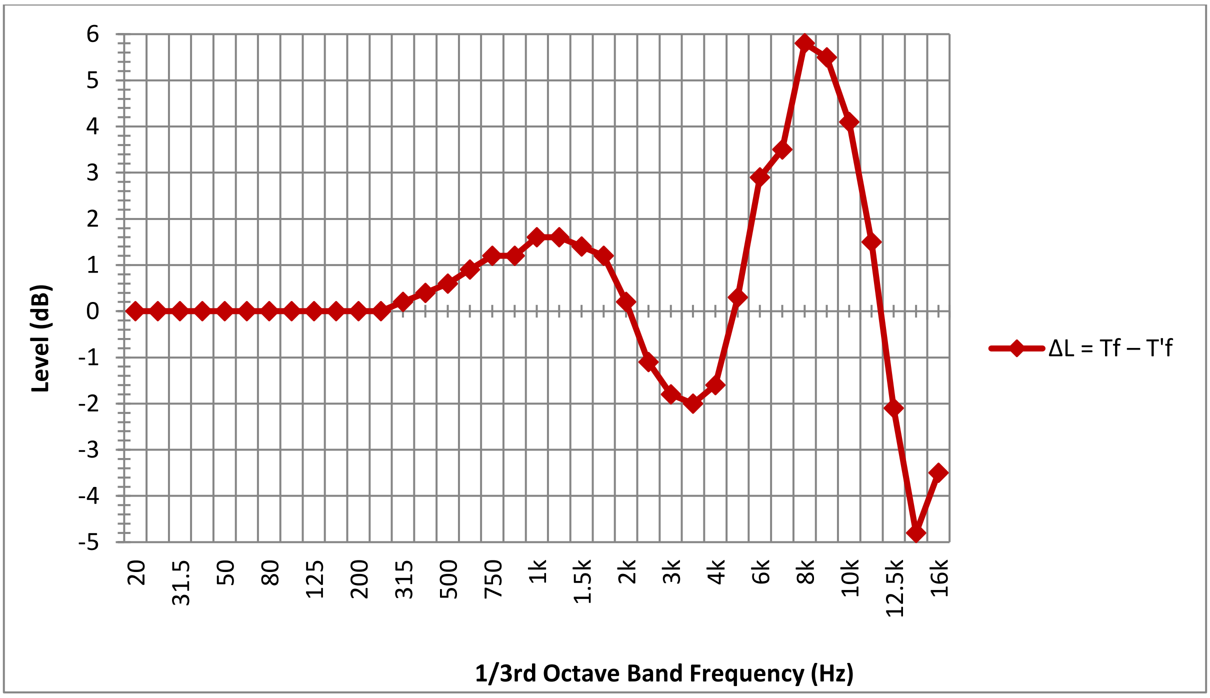


Figure 2.21: Diffuse-field correction $\Delta\mathbf{L}$; derived from (ISO, 2005), (ISO, 2016)

A similar experiment was completed by Dolby in 1971 and is referred to as the "Elstree experiment" named after the Elstree Studio in the U.K. Like Schulein's experiment, this experiment attempted to match near-field and far-field sound by placing one set of near-field monitors (KEF's) approximately 1.6 meters from the listeners and a different set of screen loudspeaker (Vitavox) approximately 12 meters from the listener. A collection of 'flat' recordings of dialog and music were played to verify that they required no additional equalization. The far-field loudspeaker were then equalized in an attempt to timbre match with the near-field loudspeaker (Allen, 2006). Allen wrote that a slope of 3dB per octave appeared to give the best match in addition to a 'slight' limitation in the lower frequencies. This experiment has many of the same experiment procedural errors as Schulein's work, including:

1. The experimental procedure has little to no documentation
2. The experimental results were never formally documented nor was the experiment ever published for peer review
3. The number of listeners in the experiment were never recorded
4. The test was not double or single blind
5. The frequency response characteristics of the test loudspeakers were not reported
6. The near-field and far-field loudspeakers were not matched nor were they equalized to be flat prior to the listening test
7. The spectral characteristics of the test source, music and speech, were not reported
8. The determination of what constituted a 'flat' test source was subjectively determined by an unknown number of listeners
9. How the listeners adjusted the loudspeakers to achieve a timbre match was not documented

In 1979 Bridges also created a similar experiment utilizing the same experimental procedure as Schulein in an attempt to find an objective reason why such a subjective roll off was required (Bridges, 1980). Bridges first calculated the hypothetical response of a known loudspeaker with a flat on-axis response using the room constants[2] of a 5,663m^3 room, and calculated attenuation of the room at a distance of 30.5m. He then plotted this response against a 'preferred' room curve that had a 3dB per octave slope starting at 1kHz. His findings were that there was a good fit between the two, as shown in figure 2.22.

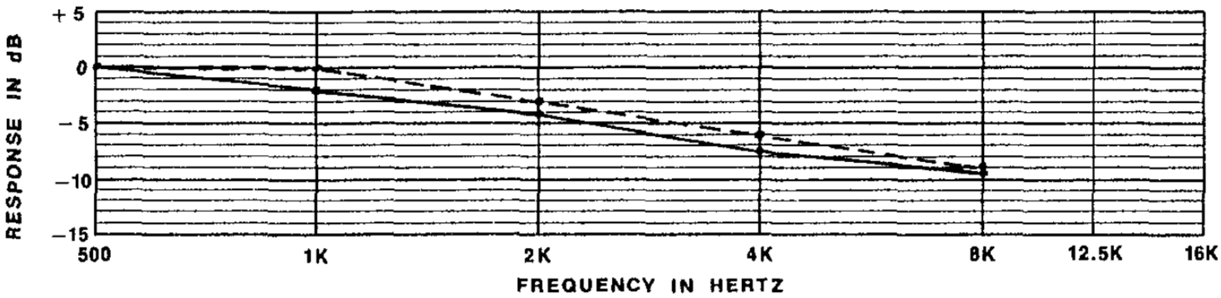


Figure 2.22: Calculated 'house curve' at 30.5m (solid line) and 'preferred house curve' (dashed line); from (Bridges, 1980)

The conclusion that Bridges drew from this part of his research was that listeners were making subjective frequency judgements from the direct sound of the loudspeaker. When he recreated Schulein's listening experiment with the same loudspeaker at a near-field position (0.76m) and the far-field position (18.3m), he found that the listener equalized both loudspeakers to have the same curve as the calculated findings, as shown in figures. The listen seemed to prefer to have the near-field loudspeaker exhibit the same response as the far-field loudspeaker, as shown in figure 2.23.

[2] Room constants: 500Hz / 4900, 1kHz / 5100, 2kHz / 6120, 4kHz / 7956, 8kHz / 10,600

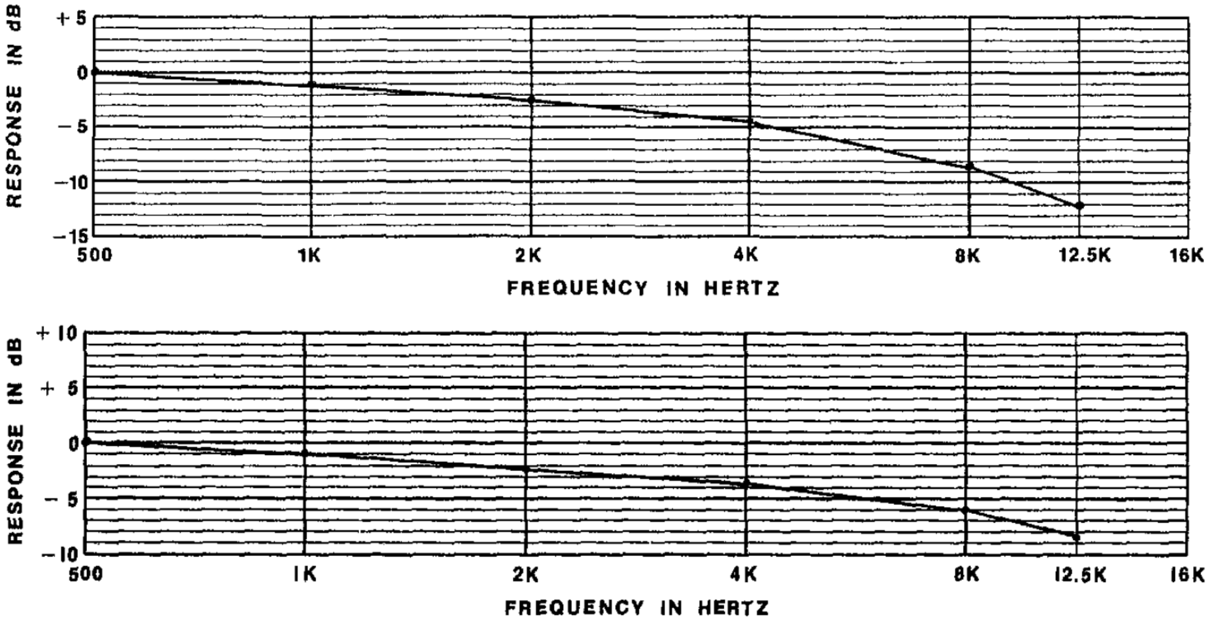


Figure 2.23: RTA trace of test loudspeaker after listener equalization, loudspeaker #1 - top, loudspeaker #2 - bottom; from (Bridges, 1980)

Unfortunately, Bridges did not document how many listeners he used for his experiment though he eluded that it was only one. In addition, he did not document information below 500Hz nor did he indicate which loudspeaker position was represented by which graph. However, his experiment did raise the notion that listeners make preference decisions based on the direct sound in the room and seem to prefer a downward tilt of the frequency response even in the near-field.

2.5 International Cinema Audio Standardization Research through the 1960's and 1970's

From the 1950's, many theatres throughout the United States and Europe were noted as deviating from the standard (Rasmussen, 1969). Hollywood was believed to be one of the few markets that attempted to keep a high level of uniformity between cinemas and dub-stages. This was probably due to the influence of The Academy of Motion Picture Arts and Sciences and the private 1000-seat theatre it maintained. The Academy Theatre, where press previews and other Academy previews were screened, was held as the standard by which all dub-stages were measured. If a film did not sound "correct", as deemed by an Academy committee, the studio was made aware of the deviation of its product and were asked to take corrective steps (Vlahos, 1969). Ljungberg further noted that the Hollywood theatres primarily had B-chain curves similar to the original Academy curve which had been specified in 1948 (Ljungberg, 1969). Around 1969, noting the lack of consistent sound quality in theatres, countries throughout Europe, along with the UK and the United States, began work on collecting data in order to study the response curves in various theatres and to document any similarities if they existed. Work in this area was lead most notably by Rasmussen in Denmark, Ljungberg in Sweden, Buckle and Lumkin of the UK with guidance from Vlahos in the US. Their work culminated in the formation of Working Group 3.1 of the ISO's Technical Committee 36 - Cinematography (Vlahos, 1969). The efforts of this working group anchored the first ISO standard on the monitor chain response in theatres titled *Electro-Acoustic Response of Motion-Picture Control Rooms and Indoor Theatres*.

Ljungberg's work began in 1965 with investigating complaints of poor dialog intelligibility in several theatres in Stockholm. Lungberg's company had already been working on the plans for a new sound department at a film company which included several mixing and screening rooms. The largest hurdle they had to overcome was to determine how to design these rooms so that they would sound correct in a theatre. They were faced with not knowing which theatre to match the film company's rooms to. Lungberg and his associates were frustrated by the fact that the only portion of the audio playback chain which had been standardized with objective measurements was that part of the chain up to the output of the amplifier. Since most loudspeaker / room system were tuned by ear, this left Ljungberg with very little data to work from (Ljungberg, 1969).

Ljungberg first introduced the concept of breaking the entire cinema playback system into the 'A-chain' and 'B-chain'. The A-chain consisted of the all equipment preceding the source switch with the B-chain consisting of everything after the switch. The B-chain includes not only the power amplifier and loudspeakers, but the screen and the auditorium acoustics as well. The rationale behind this decision was that the 'B' part of the chain remains constant and does not change regardless of what type of reproduction system is used in the 'A' part. The A-chain frequency response for optical playback was determined by the high-frequency slit loss whereas the A-chain for magnetic and non-synchronous sources (disk or tape based) were ostensibly 'flat'. The A chain already had a standard curve applied to it through the use of the Academy curve and therefore the A-chain measurements were typically consistent from cinema to cinema. However, the B-chain characteristics for multiple theatres where shown to have wide variations (Leembruggen et al., 2011), (Ljungberg, 1969), (Rasmussen, 1969), (Buckle and Lumkin, 1969).

As Lungberg completed his measurements on various theatres he noted that, not only were no two theatres alike, but that the situation was worse than he expected. He stated, "We slowly lost all ability of being astonished by any (B chain) curve and grew completely reckless in equalizing it." Lungberg set about developing a B-chain response recommendation through a series of listening test, which include several professionals who had worked on the listening test materials. Various types of test materials were used including domestic and foreign films, magnetic dubbing masters, music on tape and films and commercially pressed lacquers. The loudspeakers used were several different types typically used in theatres and monitor rooms. The results of these listening tests are shown in figure 2.24.

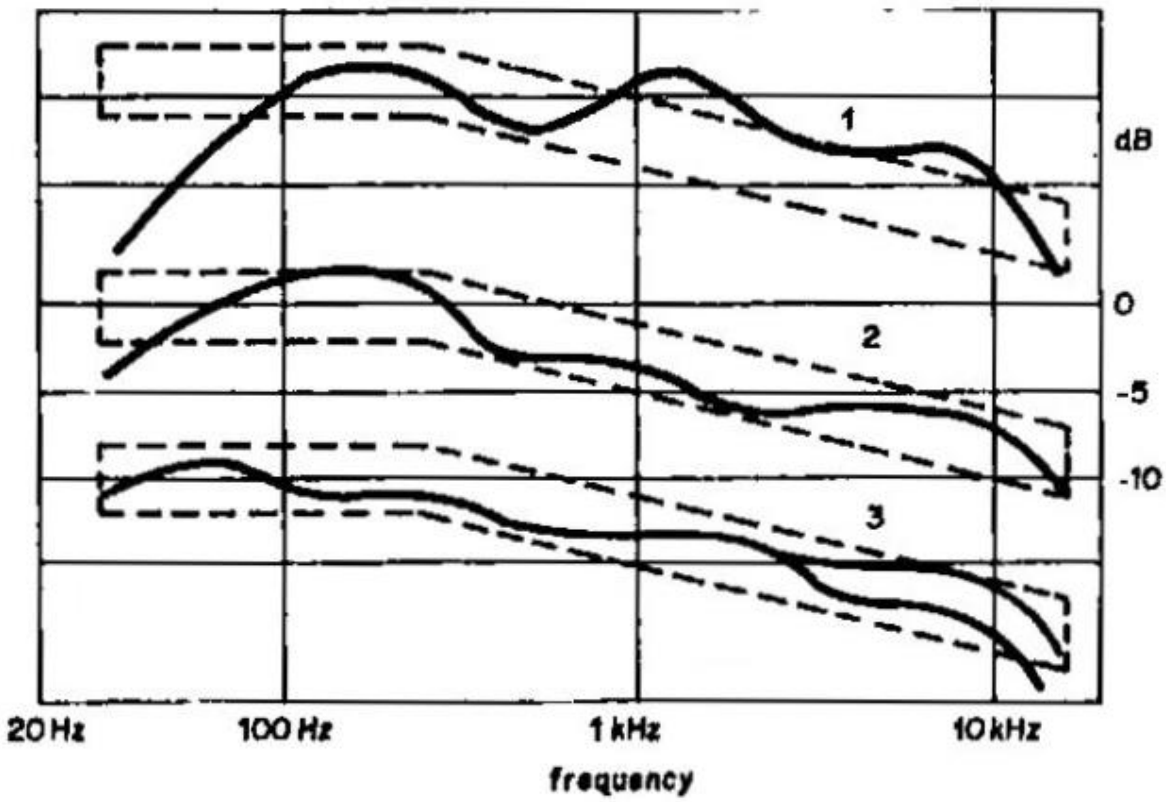


Figure 2.24: Monitor chain characteristics based on results of listening tests, with recommended tolerances (dotted box); from (Ljungberg, 1969)

In figure 2.24, curve 3 was found to be the most well-balanced when over the span of program material used while curve 2 was found too bass heavy and curve 1 was found to be too shrill. Rasmussen's work during this time period included the measuring of 25 theatres for the Danish Film Foundation. Under a program where free testing was provided to any theatre owner who was interested, Rasmussen ultimately measured an additional 100 theatres. The Danish effort primarily focused on measurement technique and the documentation of the results in the form of a typical curve and associated tolerances. The measurements of the first 25 theatres attempted to verify their performance against the current ISO draft proposal, shown in figure 2.25.

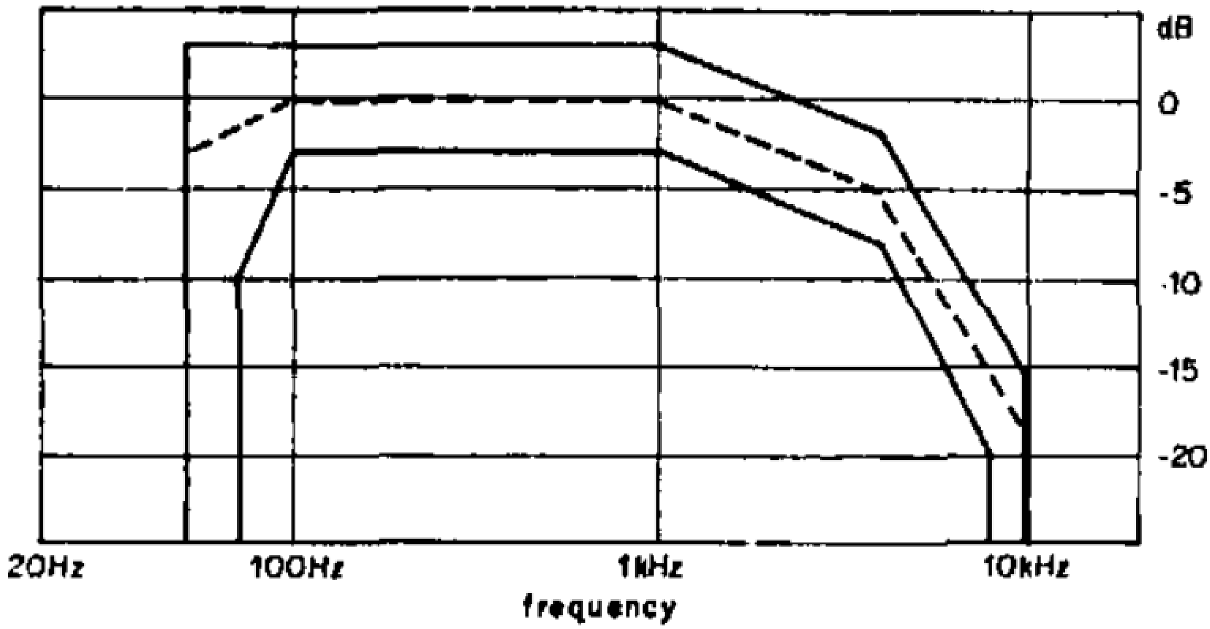


Figure 2.25: Sketch of ISO draft proposal, Moscow, May, 1969; from (Ljungberg, 1969)

This ISO draft proposal was created from measurements taken in various theatres in different countries, some of which included influences of the original Academy curve. Of the 25 theatres measured only 6 were within the tolerances of the standard while 9 need simple modifications and 10 were out of the tolerance bounds by large factors.

The testing of these first 25 theatres by Rasmussen and his team provided a basis for the 'recommended practice' section of the Scandinavian Film and Television Society (NFTU) Document NFTU-1069. This document specifies the test film, sound level meter, measurement positions, averaging method, overall acoustic response curve and acceptable tolerances

(Rasmussen, 1971), (Rasmussen, 1969). The testing loosely utilized the ISO standard as a basis and attempted to further define tolerances based on the first 25 theatres which were tested. The findings were so varied that the tolerances were ultimately set based on some of the test data and the listening experience of the research team. The curve and tolerances ultimately proposed are shown in figure 2.26.

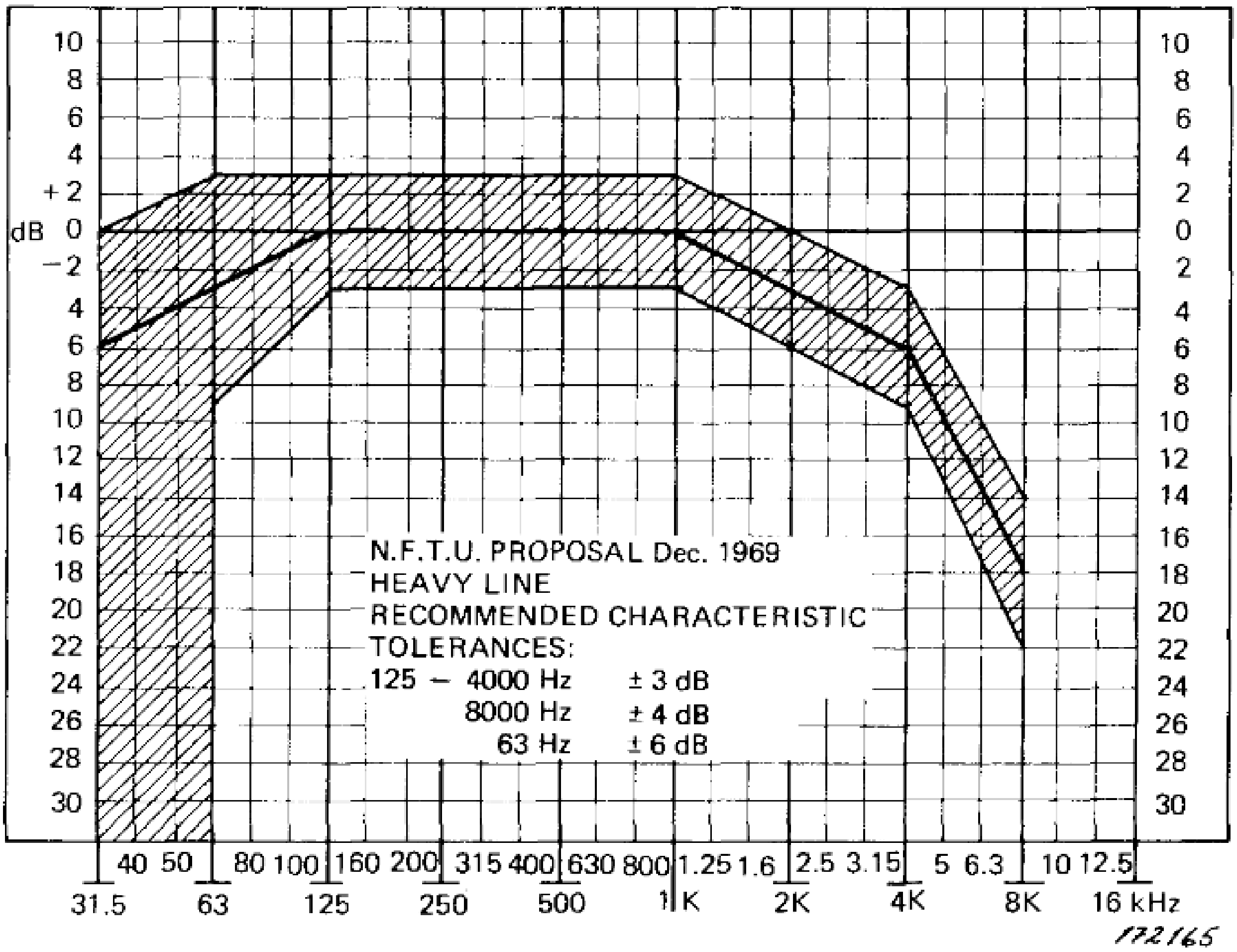


Figure 2.26: Optical-acoustic frequency response curve and tolerance limits of NFTU-Proposal 1069—1969; from (NFTU, 1969)

When testing the additional 100 theatres in comparison to this standard, Rasmussen found that only 27% were found to be inside the boundaries of the tolerances set, shown in figure 2.27.

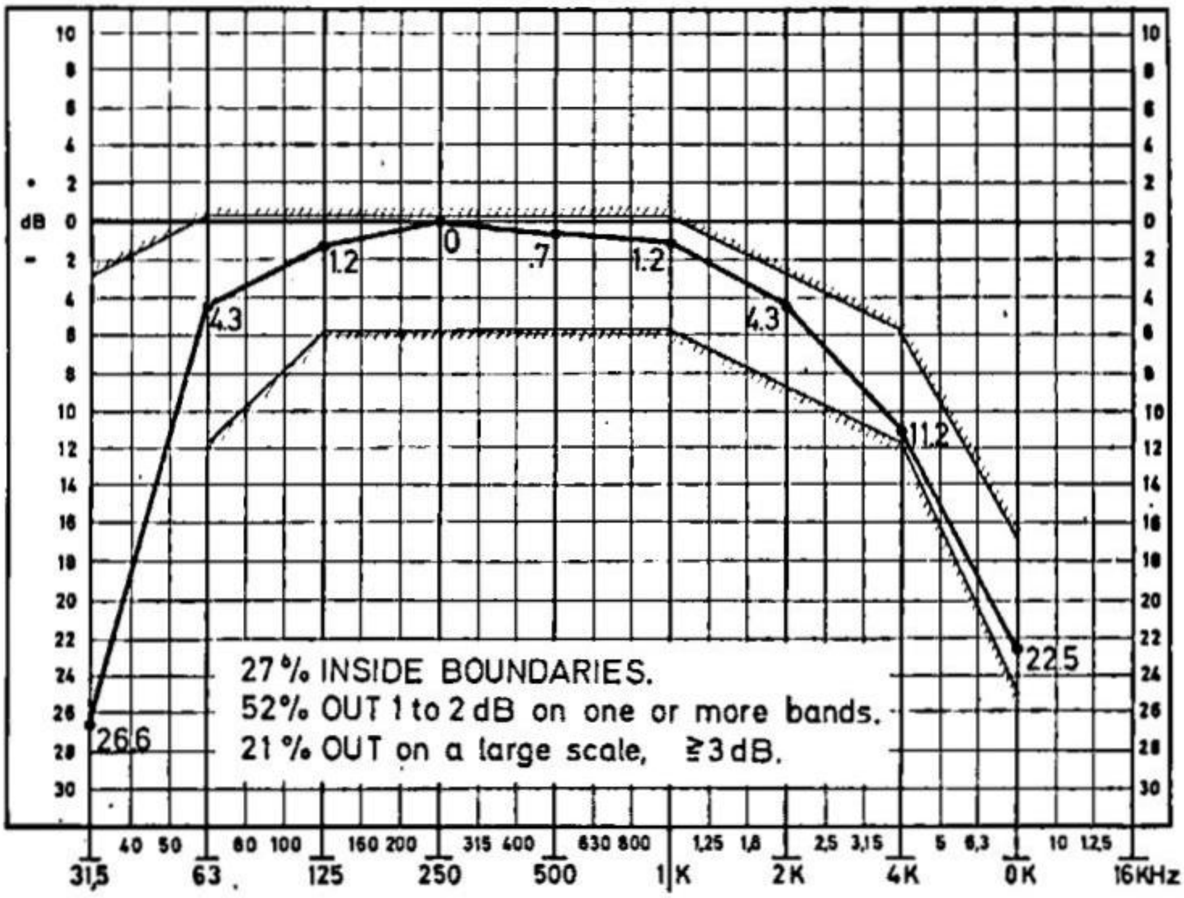


Figure 2.27: Arithmetic average of 100 theatres; from (Rasmussen, 1971)

Buckle and Lumkins' finding were similar to those of Ljungberg and Rasmussen's in that there was a wide range of measured curves across the 6 re-recording theatres at major film studios in London and 6 large cinemas they tested, some of which were built in the 1930's. Buckle and Lumkin made special note that any attempt to find separate envelopes to correlate with theatre size was futile (Buckle and Lumkin, 1969).

There are a series of similar processes and observations across all three pieces of research which are summarized as follows:

1. All surveys used test audio source that was limited to 8kHz, and many tests were taken using mono optical film which presented with its associated noise.
2. Some form of the Academy curve was in place, though not always followed closely.
3. The curves measured did not result from theatres being purposely calibrated this way

but simply reflected what was present.

4. The curves included old theatres with very little acoustic treatment and with long reverberation times.
5. The curves included projections screen from various manufacturers which all had unknown amounts of high frequency loss.
6. There was very little commonality in reverberation times across theatres.
7. The reverberation times in the re-recording (mix) theatres were much lower than that of the playback theatre.
8. All noted that the only way to truly bring theatres into agreement with each other was by way of acoustic / room modifications and that an acoustic match would not be obtained by adjusting electrical curves.

In the 1960's, audio on magnetic film strip was widely available and found to be superior in sonic quality and robustness. However, the predominant format for sound on film was still a mono optical track. Years later Rasmussen noted that, though it was predicted that optical sound would be replaced completely by magnetic sound, optical soundtracks remained just as the Academy curve had (Rasmussen, 1976). Because of this, there was still significant high frequency attenuation during playback due to constrains previously mentioned, in addition to the loudspeaker characteristics, the screen loss through the perforated projection screen and air absorption towards the rear of the larger cinemas. The dub-stages attempted to more closely mimic the commercial cinemas by using similar projectors, projection screens and loudspeakers. However, the dub-stage systems would still often incorporate additional filtering in the monitors in an attempt to further match the cinema systems' high frequency attenuation attributed to negative-to-positive print losses, variations in slit loss and electrical filters in the cinemas' loudspeaker systems. This would typically lead the mixer to equalize the soundtrack elements as they mixed in order to compensate for the high frequency roll-off. The high frequency boost, most often in the dialog track, would range around 6dB or more, centred in the 3kHz region. In addition, a low-pass filter would be included at approximately the 8kHz to 10kHz region in an effort to reduce the sibilant distortion often associated with boosting the dialog frequencies. Finally, a high-pass filter at around 100Hz was also incorporated in order to provide some balance to the sound; to keep the overall soundtrack from sounding too heavily weighted toward the low frequencies. It was further noted that if the monitor response in the dub-stage had been allowed to be flatter, then less equalization would be required and the distortion would have been lower (Holman, 2010), (Allen, 2006).

What ultimately occurred, from premix to playback, was a type of unintentional pre-emphasis and de-emphasis. This had the benefit of alleviating some of the high-frequency noise on the playback side. The downfall was that, working with a severely limited bandwidth gave rise to excessive pre-emphasis which in turn led to distortion. This distortion steadily increased through the magnetic premixes which were then transferred to the optical release print. In addition, there was no evidence provided that the balance between the low and high frequencies dictated by the excessive high frequency roll off would provide a better subjective impression of the final film sound mix.

2.6 The Development of the ISO and SMPTE Standards

One of the first efforts to create and internationally accepted theatre B-chain response curve happened at the 8th Congres del' Union Internationale des Associations Techniques Cinematographiques in Brussels in 1968 (UNIATEC, 1968). The curve reported is represented in figure 2.28.

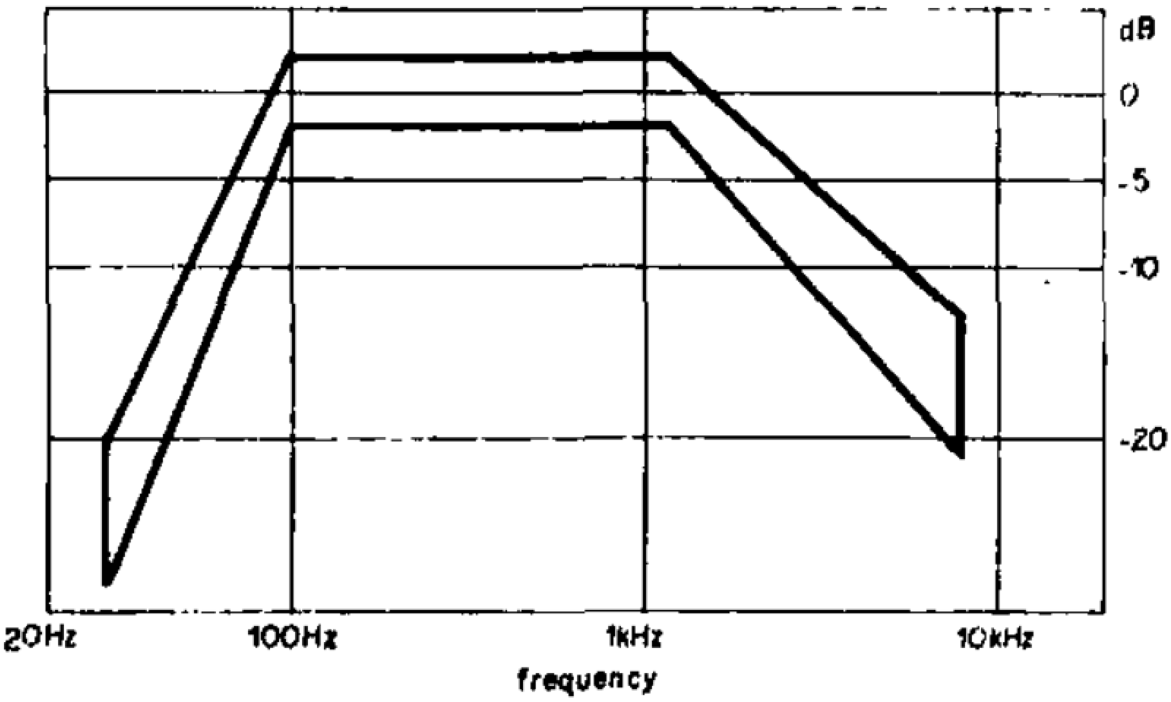


Figure 2.28: B-chain for mixing rooms from UNIATEC proposal September, 1968; from (Ljungberg, 1969)

The concern amongst researchers such as Lungberg was that this curve, when added to the Academy curve, produced a high frequency roll-off of approximately 30 to 38dB at 8kHz.

Work on an international standard had started with an ISO meeting in Moscow in May of 1969 which brought together studies from both the U.S. and the U.S.S.R. This meeting was an early attempt to standardize the response of the B-chain and its measurement processes. Initially it received a basic agreement from the U.S., the U.S.S.R, and most of the European countries who had been working on parallel studies of the B-chain. The three studies referenced in section 2.5 were published individually in the December 1969 issue of the SMPTE Journal along with a reprint of the draft international standard for the B-chain response prepared by ISO TC36 (SMPTE, 1969), (Vlahos, 1969).

The studies that appeared in the December 1969 SMPTE Journal all referenced the use of noise for B-chain measurements, typically played off film or tape. A tape recorder allowed direct insertion of noise into the B-chain while noise printed to film would require the subtraction of any A-chain characteristics from the measured response. Buckle and Lumkin utilized white noise in 1/3 octaves, Rasmussen and Ljungberg used both whole and 1/3 octaves, and the draft

ISO document cited the use of pink noise in sequential 1/3 octaves, typically measured with a sound level meter and an RMS millivoltmeter (Buckle and Lumkin, 1969), (Rasmussen, 1969), (Ljungberg, 1969), (SMPTE, 1969). The standard response curve presented in the ISO draft included a high frequency roll off of about 14dB at 8kHz, with a low frequency roll starting at 100Hz in an effort to provide spectral balance. However, this curve made up only the B-chain characteristic which would then need to be added to the old Academy Curve for the A-chain, as shown in figure 2.29.

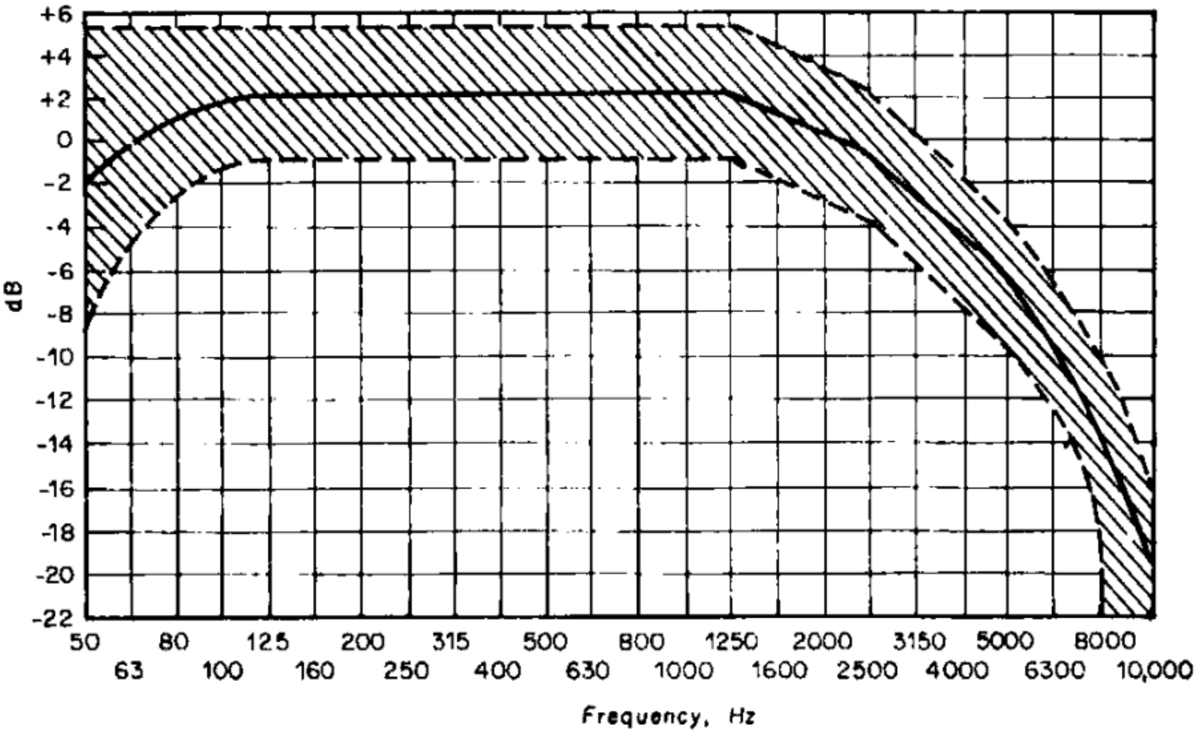


Figure 2.29: Standard curve of characteristics of monitoring chain, ISO/TC 36/WG 3 July, 1969; from (SMPTE, 1969)

During this exploratory phase of the first ISO draft it was revealed that not all of the countries involved had used the roll-off of the Academy curve, with some refusing to use it while other chose to use a 'flat' characteristic instead. The first draft of the ISO proposal was never approved as it was felt that the heavy Academy high frequency roll-off needed to be maintained to accommodate both films currently in production and those that were still in circulation and would be for several years (Rasmussen, 1976), (Vlahos, 1975). Vlahos additionally pointed out that, in practice, international standard committees attempt to reach agreements of some kind but, if aspects are not accepted by all parties involved, then sub-elements" are selected so that some agreement could be reached. He went on to acknowledge that compromise on certain aspects is an acceptable outcome.

After a meeting in London in June 1971, a modified draft was disseminated to member countries in April 1972 and an editorially modified version was circulated for voting in November 1972. Eleven countries voted yes, while three countries disapproved: Germany, Italy, and the U.S. Germany felt that its theatres had a much flatter characteristic than shown in the document. The primary reason for the U.S. to vote "no" was the belief that only the combined A and B-chain document would be of use. The argument from the U.S. was that one country could have a steeper A-chain roll-off and a flatter B-chain than another, yet demonstrate the same over all response. This case was further argued in a comment by Petro Vlahos, U.S. correspondent to ISO TC36, circulated in August 1973 (Vlahos, 1973). In an effort to solve this impasse, ISO TC36 then agreed to start work on an A-chain document to augment their work on the B-chain.

The U.K. drafted a new version of ISO DIS2969, which defined a 3dB per octave slope starting at 2kHz. It was apparent that in some countries, this curve was much brighter than current practice, and it was decided that ISO member countries should be asked to survey some of their typical theatres. A new draft of the document was circulated in February 1975, with two curves listed. Both were brighter than the previous 16 dB down at 8 kHz. One was 11 dB down at 8 kHz, and the other, the wide-range curve, down 6 dB at 8 kHz and were labeled 'Y' and 'X' respectively. A note in the draft stated: “This proposed draft includes two characteristics for the response beyond 4 kHz. The response and tolerances are identical up to 4 kHz. Depending on the replies by Preparatory Working Group 3 Specialists, only one of these characteristics will be chosen.”

During a meeting in London in June of 1975, a compromise was reached and it was decided that both curves should be in the next draft document. The 'Y' curve was to support films currently in circulation while the 'X' curve was to support new soundtrack formats such as the hue-modulated color soundtrack. These curves are shown in figure 2.30. This draft was refined and a new version was circulated for comments in September 1975 (Rasmussen, 1976), (ISO, 1975b).

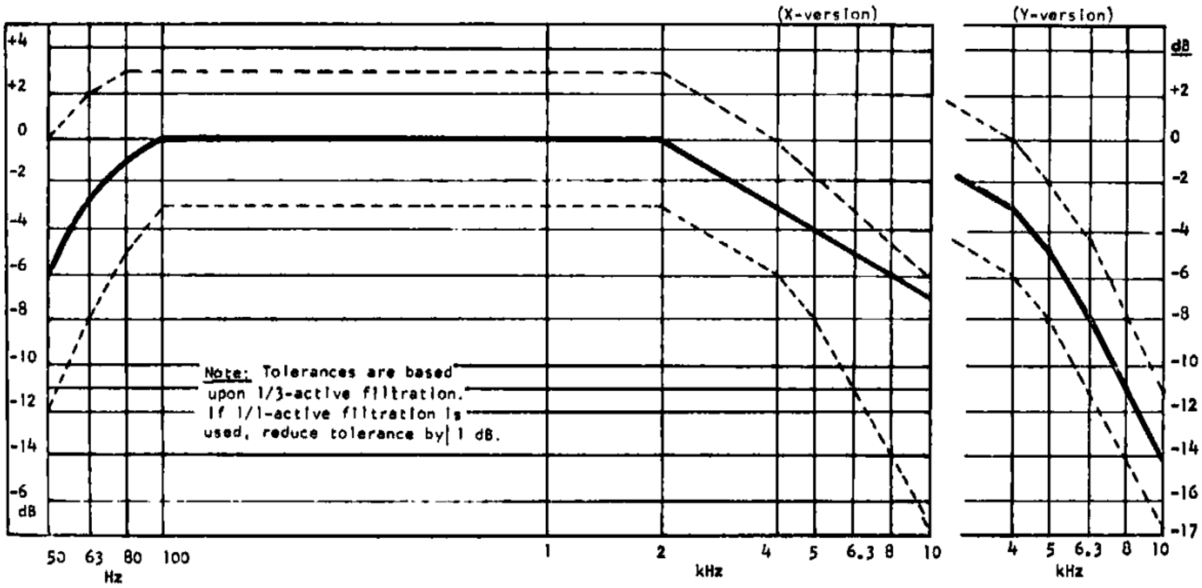


Figure 2.30: Proposal for B-chain response curve, agreed to by ISO preparatory working group at London meeting, 1975; from (Rasmussen, 1976)

Vlahos had raised the question of whether the old Academy curve should be abandoned and noted that the issue had been argued among member countries but countered that doing so would overlook pictures that were currently being made and that would be in theatres for years to come.

During this same time frame, a B-chain acoustic response curve for dub-stages was submitted to SMPTE for consideration as an American standard with the thought that it may become a U.S. proposal to the ISO (Vlahos, 1975). This proposed curve is shown in figure 2.31.

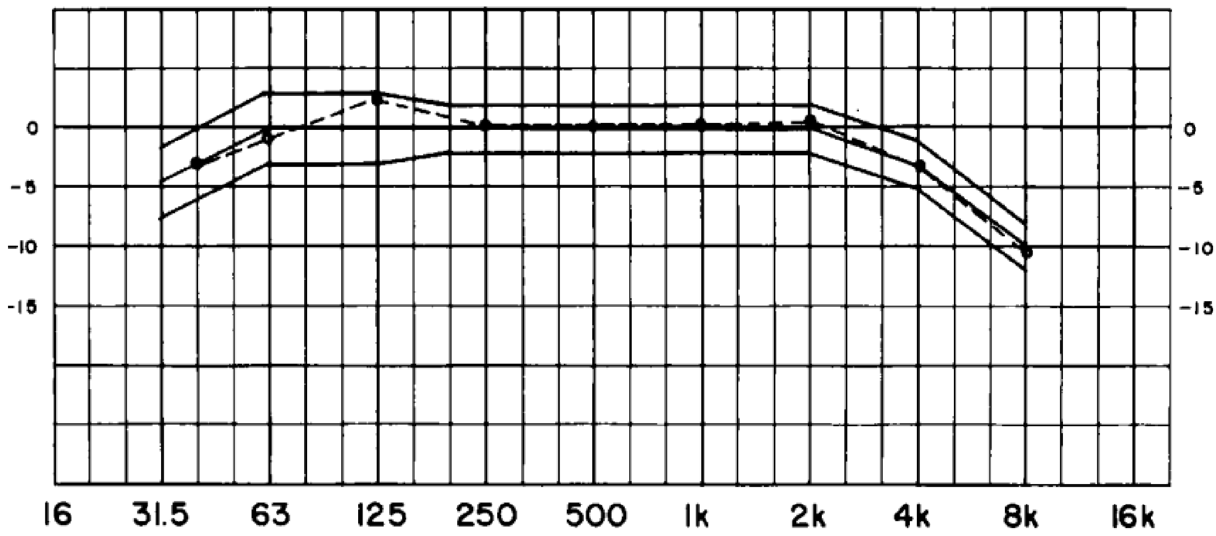


Figure 2.31: Dub-stage B-chain acoustic response, showing proposed standard and tolerance (solid lines) and average of 9 Hollywood dub-stages (dashed line); from (Vlahos, 1975)

Of interest to note, this curve was for dub-stages working with the extended frequency characteristics of magnetic prints and it was dependent on the mixer to lower the high frequencies level (follow the Academy curve) to compensate for optical releases. Also worth noting is the extended low frequency characteristics and the tighter tolerances in the mid and high frequencies of this curve. This curve never became part of a published SMPTE standard.

A possible reason that the curve depicted in figure 2.31 was not adopted was implied in a paper presented by Allen (1975). He commented that, though despite the improvement made by magnetic soundtracks, it seemed certain that optical soundtracks will continue to be used because of production cost savings.

With final ratification in 1977, the second ISO draft became the first published version (ISO, 1977). This document was mimicked in the U.S. and became SMPTE 202M Dubbing Theatres, Review Rooms, and Indoor Theatres – B-Chain Electro-Acoustic Response, in August of 1978.

2.7 The Current SMPTE Standard

On the scientific side, there are many physical variables (e.g. the directivity and quality of the loudspeakers, their mounting conditions, cinema screen attenuation and seat diffraction effects, the size and acoustical treatment of the cinema, etc.) that are still not adequately addressed within the SMPTE standard (Leembruggen et al., 2011), (Newell et al., 2011a). Steady-state measurements, followed by simple 1/3rd octave in situ equalization cannot adequately compensate or control for the psychoacoustic effects these variables have on the perception of cinema sound. In fact, the most recent version of the SMPTE standard, ST 202:2010, suggests an aural evaluation may be necessary to guide the final B-chain adjustment for the best quality sound. (SMPTE, 2010)

Recent anecdotal evidence from film sound mixers, cinema acousticians/installers, and moviegoers indicate that a high-quality loudspeaker system can sound worse after SMPTE calibration. This evidence is underscored by the survey of re-recording mixers presented in Chapter 3.

Since the official adoption of the SMPTE standard in 1978, very little has been changed about the standard except for some basic modifications made over the years to accommodate new playback methods such as surround sound and digital audio.

There were some minor changes to both the SMPTE and ISO standards around 1982, which entailed extending the frequency range from 10kHz to 12.5kHz. Additionally, it was believed that the B-chain response would vary with room size, and so additional curves in the upper frequencies were added which were correlated with seat count as opposed to room size (ISO, 1987). The next revisions of both the ISO and SMPTE standards came in 1991, when the X-curve was extended out to 16 kHz. At the same time, a second knee at 10kHz was added which resulted in a slightly steeper slope between 10 and 16kHz. In 1998 the SMPTE standard was once again revised but only to make a differentiation in how the surround sound loudspeakers were to be calibrated. It was acknowledged most of the audience is closer to the surround loudspeakers than the screen loudspeakers.

Recently, the SMPTE standard was revised in 2002 and once again in 2010. The revisions were mainly to remove older wording associated with analogue audio technology and to update it to include newer digital audio wording. SMPTE standards are typically reviewed every 5 years and are revised as needed.

2.8 Current Research on Cinema Calibration

Research on cinema calibration is currently being carried out by members of SMPTE and the Audio Engineering Society (AES). The author is on the cinema audio technical committees of both of these organizations as well as member of the Academy of Motion Picture Arts and Science Theatre Standards Committee.

In his 2006 paper, Allan had explained that high frequency roll-off of the steady state measurements were required so that short duration sounds would be perceived as flat. Leembruggen and his team found that in the mid to high frequencies (1.kHz to 12kHz) the early response (10ms) was only approximately 4.5dB below the steady state response and that the shape of this response does not resemble the X-curve (Leembruggen et al., 2011). In addition, they and Toole have shown that calibrating a cinema to the SMPTE standard creates a need to severely roll-off low frequencies in order for the steady-state measurements to meet the X-curve (Toole, 2015). The issue with this is that the X-curve overcompensates for a reverberant energy build-up in the low frequencies which was overestimated in the first place. In addition, the reverberation times in current cinemas are even less that when the standard was first written.

Additional related research that has recently been conducted is cited throughout this thesis in the appropriate chapters.

2.9 Summary

The literature review has spanned several decades of cinema calibration starting with the first known documentation through to the present day standards. To summarize, the SMPTE standard in its current form has evolved relatively little from its first inception. This is in light of the fact that almost every aspect of the cinema audio chain has improved vastly, from digital audio playback and better transducer technology, to improved room acoustics. In addition, methodology for conducting highly accurate listening tests has also come about since the early calibration standard was set forward. However, these new technologies and methodologies have either not been used or fully executed to prove the efficacy of the SMPTE standard.

The X-curve remains somewhat shrouded in confusion as to its origins and continued perpetuation. It is a combination of remnants from old calibration curves based on equipment and acoustic compromises and justified through somewhat questionable and outdated subjective research that has never been scientifically verified. The standard represented a means to reassure, at least to some degree, that a film would not be reproduced in a theatre whose acoustic response could not at least provide a minimal degree of playback quality. It was never meant to represent a situation of perfect translation from the dub-stage to the theatres, only a reasonable assurance that it wouldn't sound bad. It also didn't represent what was subjectively desired, as it was only partially based on psychoacoustic factors and those had only been marginally proven. The high frequency roll off was a remnant of past standards which were still in place during the creation of the original standard and are the natural effect of what happens to sound when limited by the equipment, passed through a projection screen and transmitted into a large, somewhat reverberant space.

The aim of this research is to examine the quality of the SMPTE calibration method as well as new calibration methods as applied in a variety of different-sized dubbing stages and cinemas.

Chapter 3 - Re-recording Mixers' Survey of Cinema Calibration Standard

3.1 Introduction

A re-recording mixer, also known as a dubbing mixer, is a member of a post-production sound team who works specifically with dialog, music and sound effects to create the final soundtrack for a film. Their work is performed in a mixing room called a "dubbing stage", or "dub-stage", which is essentially a studio that mimics the volume, shape and acoustic characteristics of a commercial cinema. They are responsible for ensuring that the sound in a film is technically correct, and creatively reflects the director's original idea. They are also responsible for ensuring that the final version of the film's soundtrack will be represented as accurately as possible in commercial cinemas.

In order to ascertain usage and preference information amongst re-recording mixers, a range of questions concerning cinema calibration was asked using quantitative and qualitative questions via an on-line survey of film re-recording mixers. The survey served to evaluate mixers' attitudes, preferences and mixing practices with regards to the SMPTE calibration standard.

Work in this chapter has previously been presented as the following:

> GEDEMER, L. 2013. Evaluation of the SMPTE X-curve Based on a Survey of Re-recording Mixers. AES 135th Convention, New York, NY: Audio Engineering Society.

3.2 Motivation

Currently, dubbing stages and commercial cinemas are equalized as closely as possible to the X-curve at a location roughly two-thirds of the distance from the screen to the rear wall of the rooms. This area of the room has conventionally been considered the most representative average of the whole room and is generally where the mixing console is located on commercial dubbing stages.

While researchers conduct objective and subjective tests on cinema audio, no known test or survey has been targeted at the community which works with cinema audio on a daily basis; the re-recording mixers. Survey questions regarding usage and preference can help lead towards more focused quantitative research.

The following questions serve as a basis in developing the survey:

- What percentages of film re-recording mixers are currently working on a dub-stage calibrated to the SMPTE standard?
- Are mixers cognitively compensating for the standard and if so, how?
- Do they feel it provides a consistent level of sound quality translation between the dub-stage and theatre playback?
- In what ways does it succeed or fail in terms of sound quality?
- What alternative calibrations are dub-stages using in its place and why?

3.3 Survey Development

The online survey was created and administered via So Go Survey, which also provided a basic level of data analysis and graphing (SoGoSurvey, 2013). The invitation to take the survey is shown in Appendix A.

A pilot survey was initially sent directly to five film re-recording mixers within the Hollywood film industry, so that questions could be vetted for content and accuracy prior to the final survey being administered. The on-line nature of the survey allowed the finalized version to be administered worldwide in an attempt to track regional preferences. An online request for survey participants was posted on several online forums catering to the audio post-production community such as the Post Sound Mixer Group on Linked In (which currently has over 8,000 members). Each participant who expressed interest was sent a unique link that allowed the participant to access the survey only once and did not allow the survey link to be shared with others. This secured the survey against "ballot box stuffing" (a person submitting more than one survey) or the survey being taken by someone outside the target group; both of which would have led to erroneous data. Survey volunteers were also pre-screened to verify their film mixing background and appropriateness for the survey. Screening was accomplished by using the Internet Movie Data Base (IMDB), a website that tracks peoples' participation on films, television shows, games and other types of media (IMDB, 2013). Participants remained anonymous throughout the survey and the only demographic information collected was that which was considered pertinent to the overall study (age, years in the industry, location, etc).

Survey participants were asked to answer questions of various types including interval scaled, open-ended, ratio scaled and multiple choice. Questions directly regarding the SMPTE X-curve were answered using a 10-point audio assessment scale modeled after qualities suggested by Gabrielsson (1979) such as clarity, fullness and brightness.

The survey was made of up 25 questions total with the survey logic being designed in such a way as to keep participants from having to answer questions that were incongruous with answers from previous questions. The survey was vetted by members of Harman International's marketing team who are adept at creating surveys in order to collect marketing data. The survey took most participants an average of 20 minutes to complete. The survey questions are shown in Appendix B with an example of answers from survey participant #1 shown in Appendix C.

3.4 Survey Results

The following section presents the survey results which were compiled and output directly from the So Go Survey program. In total, 35 re-recording mixers from 12 different countries participated. There are potential limitations to the final survey results when considering the sample size of 35 participants. Though the total number of currently working re-recording mixers in the film industry is unknow, a larger sample size may have provided a broader range of opinions and a more definitive portrayal of the use and preference of the SMPTE standard than those presented in this research. However, the 35 pre-screened participants do represent a viable cross section of the industry with regards to age and number of years working as a re-recording mixer, and so the results of this survey provide valuable insight into the use and preference of the SMPTE standard.

3.4.1 Participant Hearing

Of the 35 participants, 66% had their hearing checked within the last 5 years, 20% were checked more than 6 years ago and 14% had not had their hearing checked. 80% responded that their hearing was found to be normal, while 3% were not sure and 3% answered in the negative.

3.4.2 Demographics

Figures 3.1 through 3.3 represent the basic demographic information from the survey.

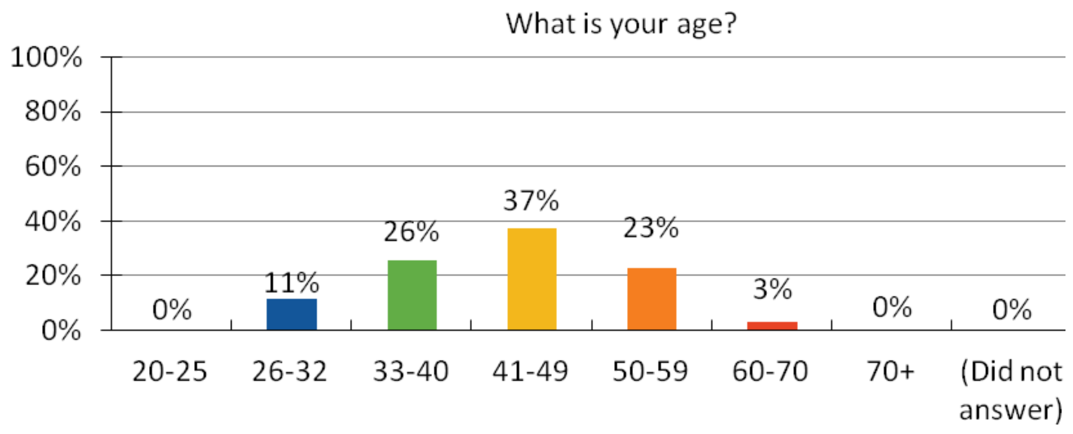


Figure 3.1: Survey participants' ages

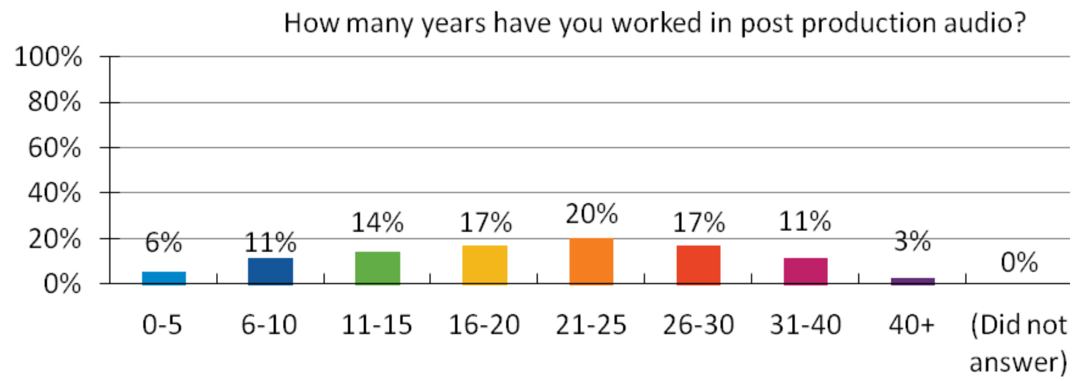


Figure 3.2: Number of years in post production

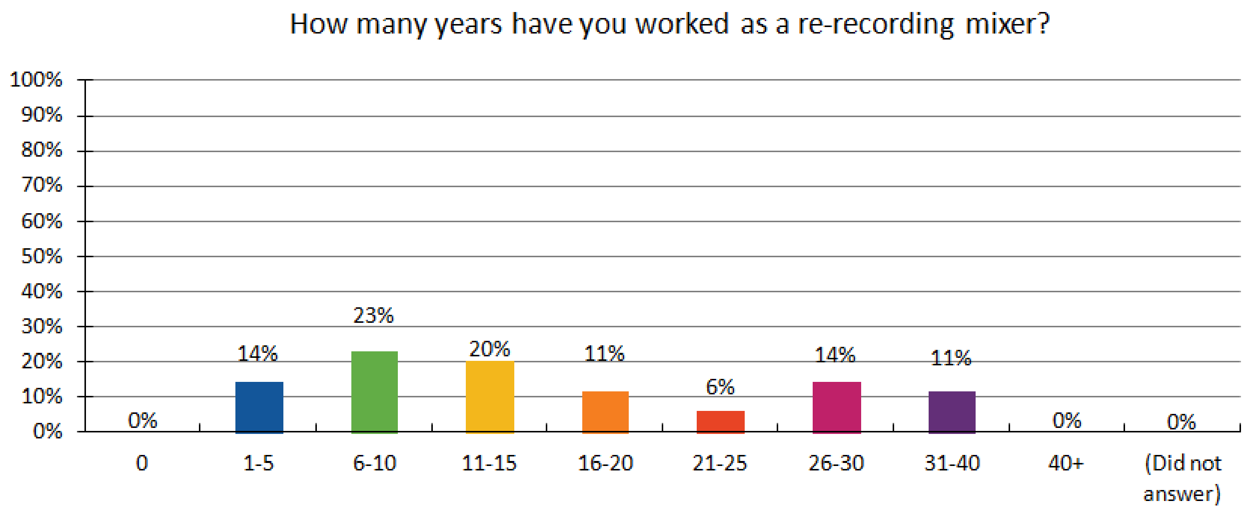


Figure 3.3: Number of years as a re-recording mixer

Survey participants were asked to identify which geographic region they were from and the breakdown was as follows:

- 66% USA
- 3% Canada
- 6% South America
- 6% UK
- 12% Europe
- 3% Africa
- 6% Australia / New Zealand

3.4.3 Work-place Information

Figures 3.4 through 3.6 represent the information about the survey participants' types of work and work-place.

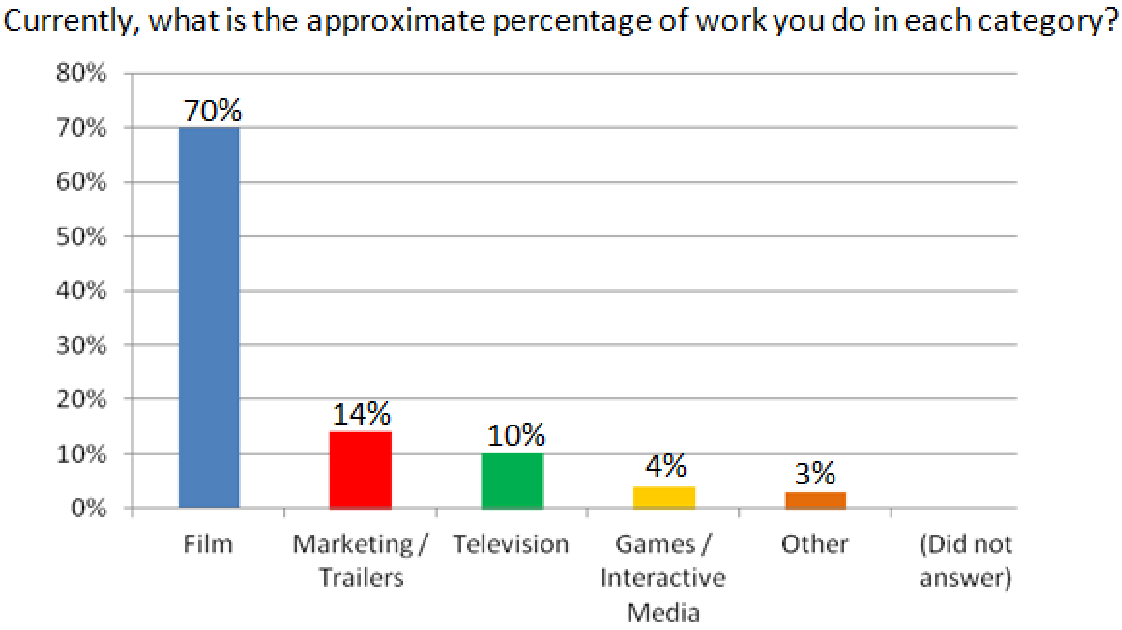


Figure 3.4: Percentage of work, by category, which participants typically work in

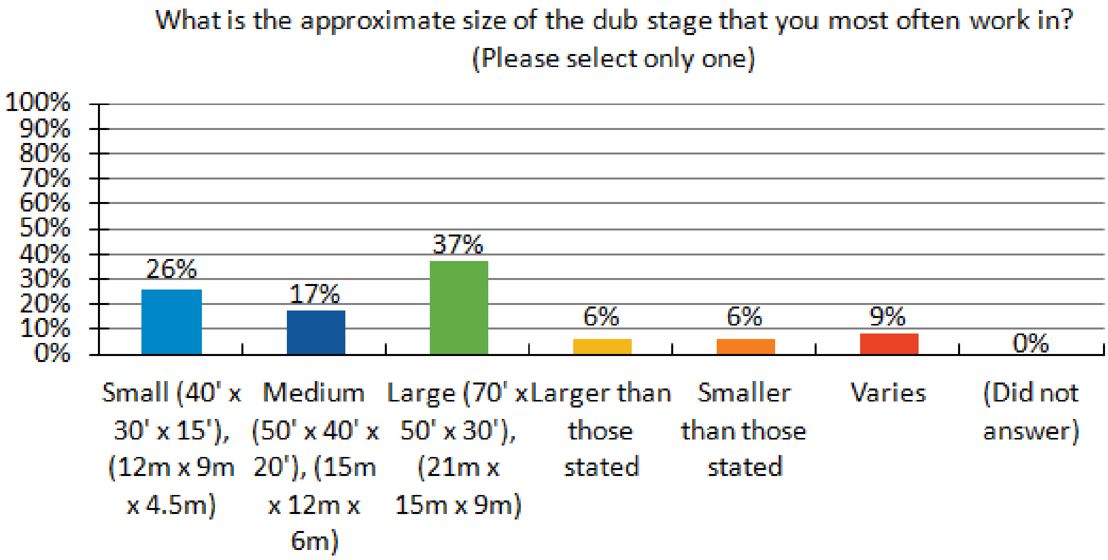


Figure 3.5: Dub-stage size participant works in

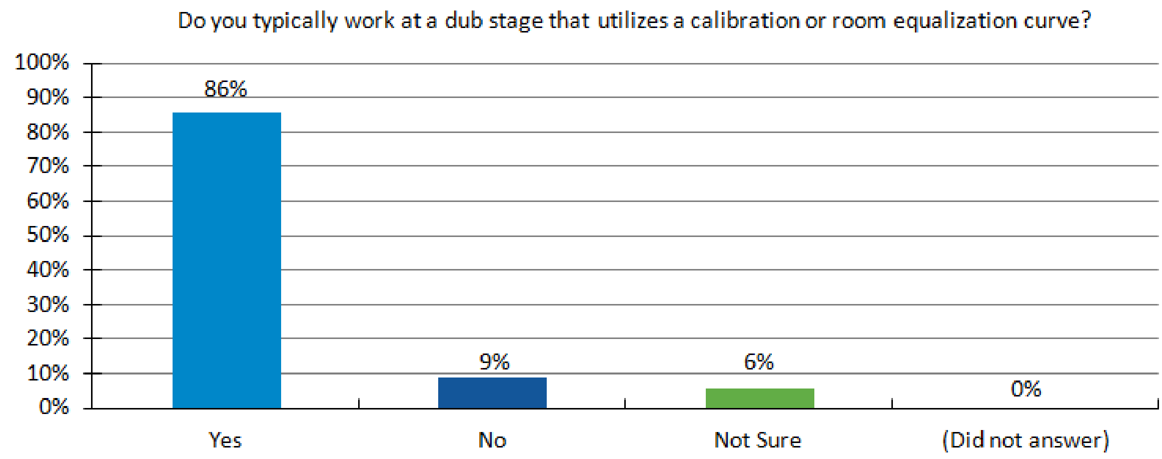


Figure 3.6: Percentage of participants who work on a calibrated dub-stage

Of the 3 participants who do not work on a calibrated dub-stage, all work outside the United States. When asked why their dub-stage does not use calibration, all 3 stated that they felt the translation of their mixes from stage to cinema was already good and that calibration was not required.

3.4.4 SMPTE X-Curve Ratings and Responses

Of the 86% (30 total) participating in the survey who work in a calibrated dub-stage, the following graphs depicted in Figures 3.7 through 3.14 represent their answers to questions specific to the SMPTE X-curve.

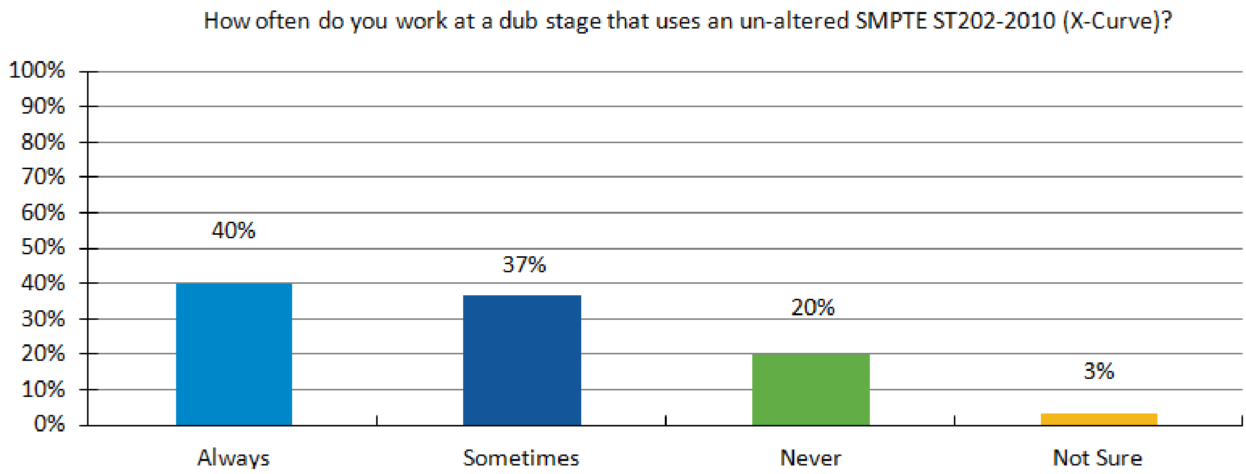


Figure 3.7: Amount of time participants work at a dub-stage calibrated to the SMPTE X-curve

Of the 37% of the participants (11 total) who responded that they only "sometimes" work on a SMPTE ST202 dub-stage, the approximate amount of time they work there was broken down as shown in Table 3.1:

Responses	Count	%
0-20%	4	36%
21-40%	2	18%
41-60%	1	9%
61-80%	1	9%
81-99%	3	27%
Total	11	

Table 3.1: Percentage of part time work spent on a SMPTE ST202 calibrated dub-stage

The data shown in Figure 3.7 and Table 3.1 show that half of all participants who work on a calibrated dub-stages (15 total) spend at least 80% of their time working on a SMPTE ST202 calibrated dub-stage.

Figure 3.8 shows how often the survey participants worked in a SMPTE ST202 calibrated dub-stage as compared to the size stage they are working in. Of interest to note is that those who work on the two largest sizes of stages spend a majority of their time on a SMPTE ST202 calibrated dub-stage.

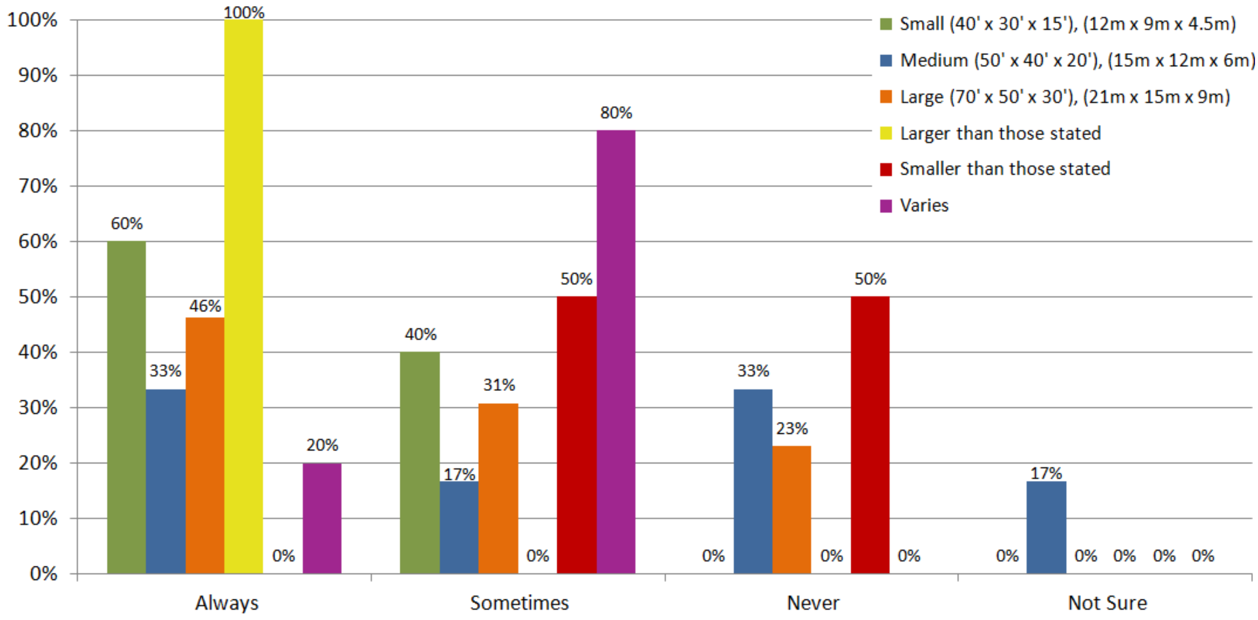


Figure 3.8: Percentage of work by stage size versus use of SMPTE ST202 calibration

The 77% of the participants (23 total) who responded that they "always" or "sometimes" work on a SMPTE ST202 calibrated dub-stage were asked to assess the quality sonic of their dub-stages. Figures 3.9 through 3.14 shows these results.

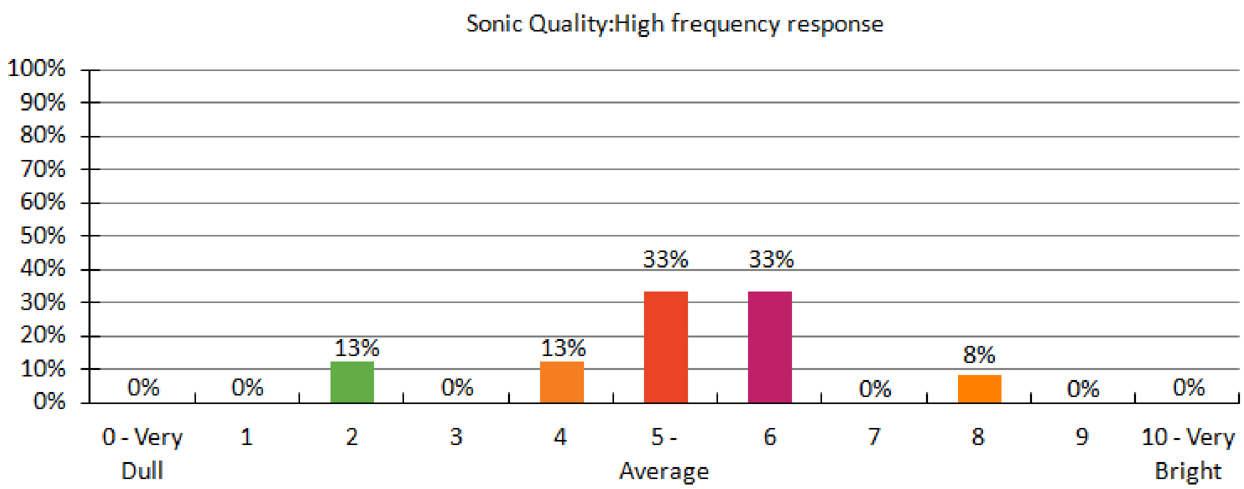


Figure 3.9: Quality of high frequency response

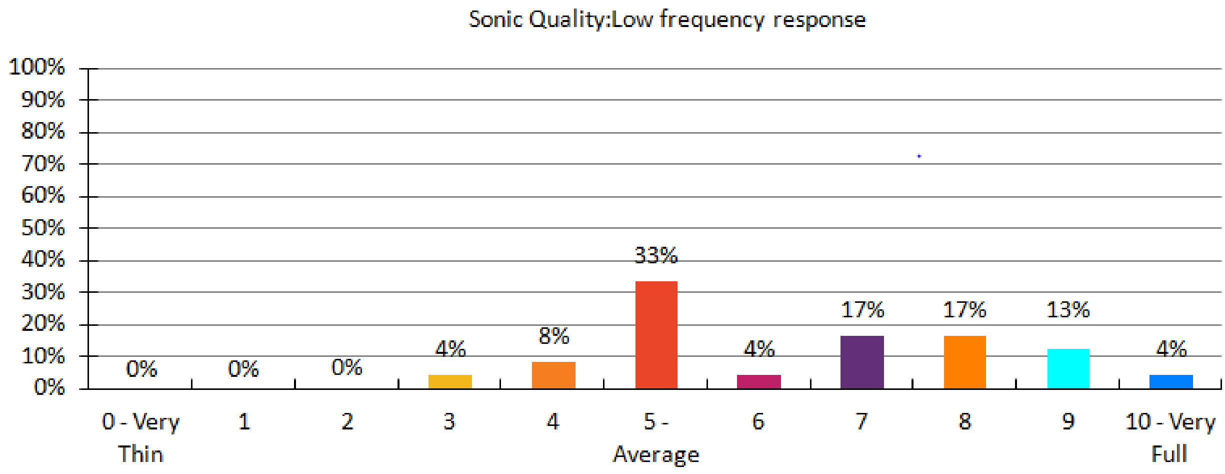


Figure 3.10: Quality of low frequency response

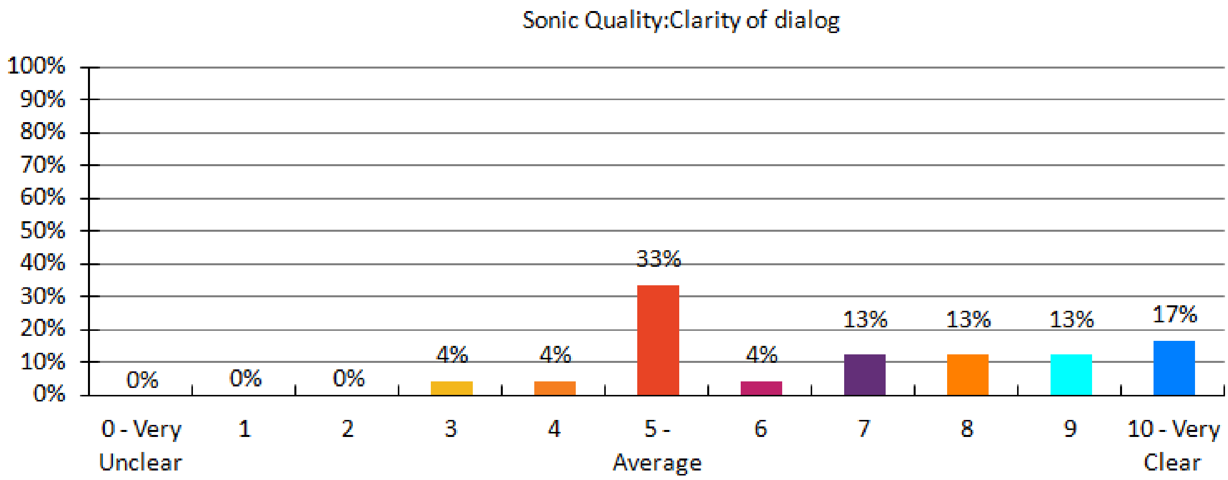


Figure 3.11: Clarity of dialog

During informal interviews conducted at the pilot survey stage, several participants expressed that room calibration can have an effect on overall spatial quality. For this reason, a question regarding spatial quality was added to the survey with the results depicted in Figure 3.12.

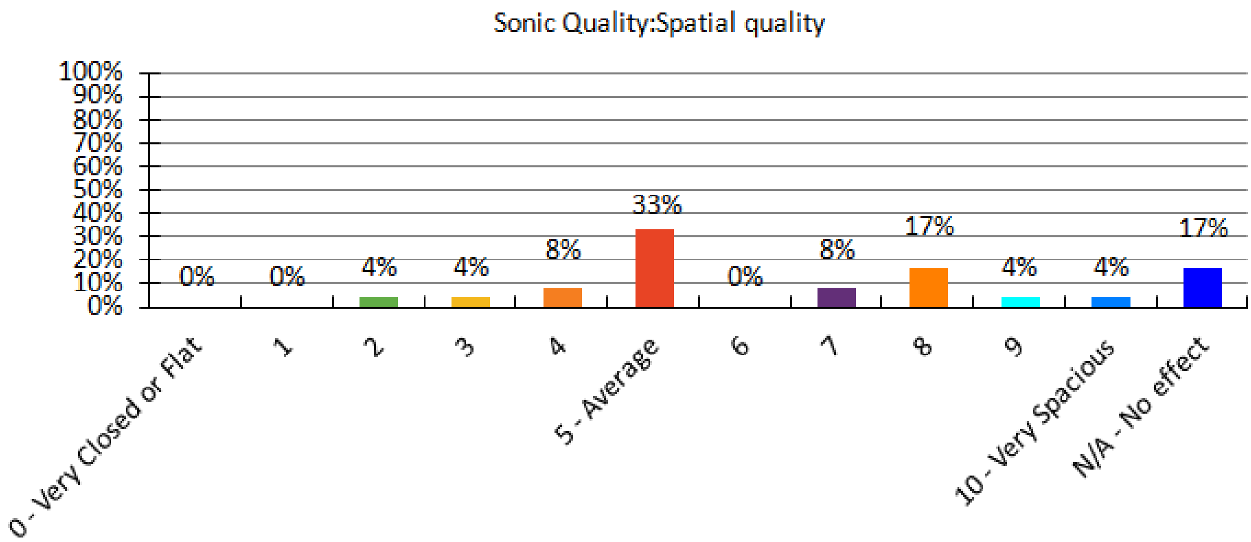


Figure 3.12: Spatial quality

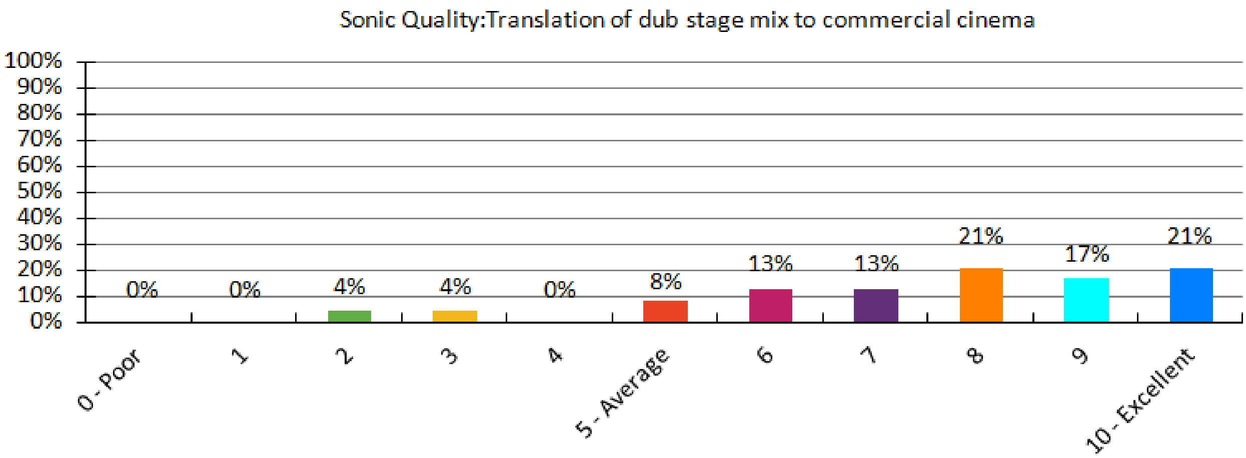


Figure 3.13: Translation of dub-stage mix to cinema

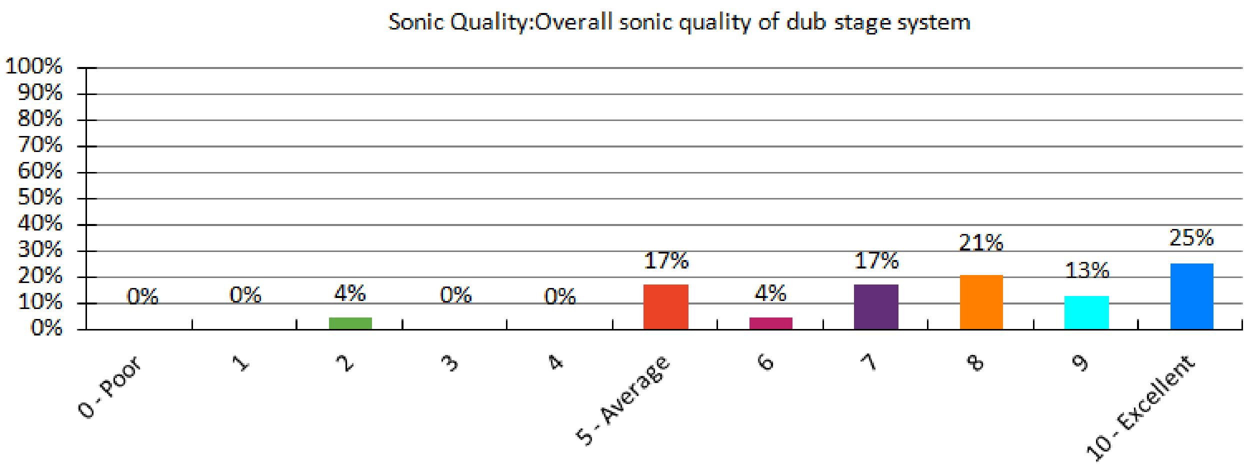


Figure 3.14: Overall quality of dub-stage system

3.4.5 Does SMPTE ST202 calibration effect how mixers work?

The 23 mixers who work on a SMPTE calibrated dub-stage (always and sometimes) were asked if they ever compensate for the X-curve in their mixes and if so, how. Five of the mixers stated that they did not use any compensation. Of the 19 mixers who do use some type of compensation, this usually takes the form of high frequency boost most often for music and dialog. Three mixers specifically stated the need to boost starting around 5k hertz. Only 1 mixer stated the need to use low frequency compensation, primarily below 150 hertz to aid the music track from sounding too thin.

3.4.6 Modified SMPTE and non SMPTE calibrated rooms

Survey participants were asked if they had ever worked in a room that was calibrated to a "modified" SMPTE ST202 X-curve or worked in a dub-stage that use a "house curve" that does not follow the SMPTE ST202 standard.

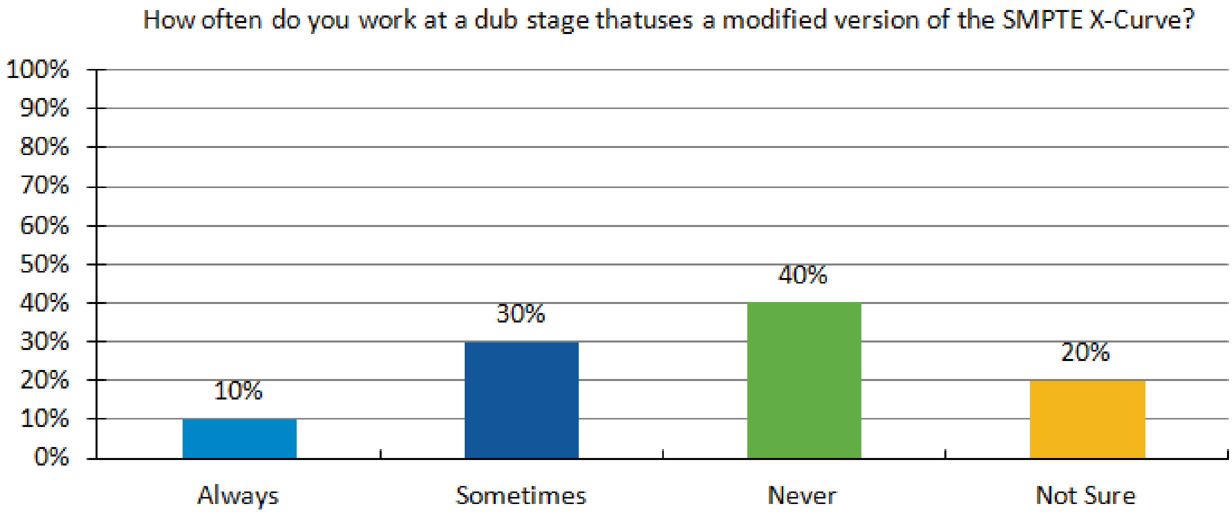


Figure 3.15: Percentage of instances where participants work in a "modified" SMPTE X-curve calibrated dub-stage

Of the 15 mixers who stated that they "sometimes" or were "not sure" how often they worked in a modified SMPTE X-curve room, 12 stated that it represented less than 20% of their overall work schedule.

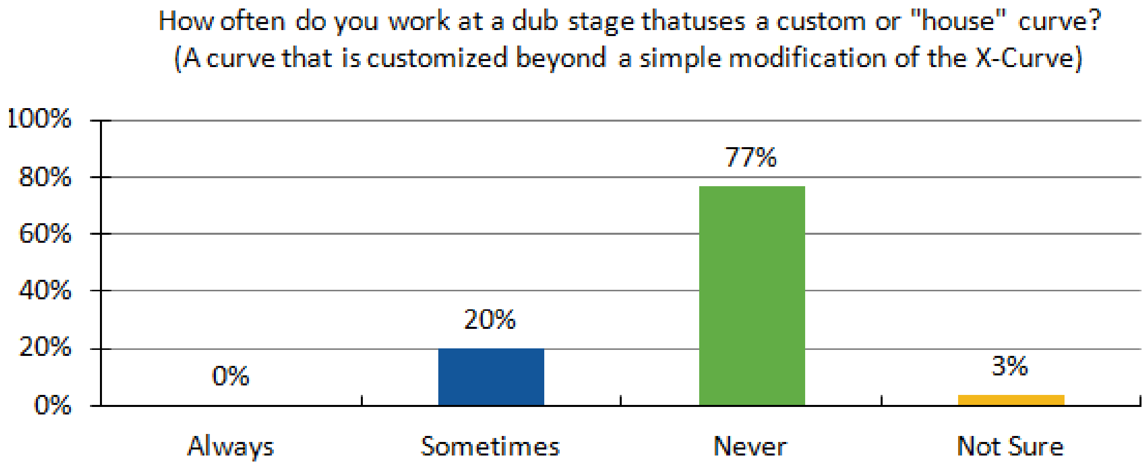


Figure 3.16: Percentage of instances where participants work in a dub-stage calibrated to a "house" curve

Only 6 mixers spend part of their time working in a dub-stage calibrated to a "house" curve and of those, 3 of those mixers work there less than 20% of their time.

3.5 Discussion

A series of open-ended questions were asked of the survey participants. The outcomes of those questions were compared to the survey data and are discussed in the next sections.

3.5.1 SMPTE X-curve Calibrated Dub-stages

The number of mixers compensating for the SMPTE X-curve in the higher frequencies coincides with Figure 3.9 and with the fact that 14 mixers rated high frequency response at or below average and 10 mixers rating clarity of dialog at or below average. Figure 3.10 shows that low frequency response was rated at or below average by 11 mixers, though only 1 mixer stated using any compensation in the lower frequencies. This may point towards a hesitancy to increase lower frequency which could make the mix "muddy" and further reduce dialog clarity. The high level of variability in X-curve compensation amongst mixers undermines translation, the corner stone of the SMPTE Standard.

Though a majority of mixers claim to use some type of compensation when mixing to the X-curve, a vast majority of them (83%) felt that the translation of their final mix from the dub-stage to a commercial cinema was above average to excellent. This number is mirrored by the quantity of mixers (80%) who rated the overall sonic quality of their dub-stage above average to excellent. This may be due to the mixers' years of experience in working in SMPTE calibrated rooms and learning to translate their work from dub-stage to cinema. There may also be some type of adaptation in process involved that allows mixers to adjust to the response curve their room has been tuned to.

From the survey data, it can be noted that 67% of mixers who had 26 to 40 years of experience rated their dub-stage to cinema translation above average to excellent, whereas 61% of mixers with 11 to 25 years and 46% of mixers with 1 to 10 years rated translation the same way. There is a question as to whether these descending experience-related percentages can be attributed to the skill of the mixer with regards to translation, adaptation or even hearing loss. Considering that mixers who have spent 26 to 40 years in a high volume dub-stage for long hours would make hearing loss an extremely plausible reason for being content with translation.

Half of the mixers surveyed spend at least 80% of their time working in a SMPTE ST202 calibrated dub-stage. Only 3 mixers spend the same percentage of work time in a "modified" SMPTE X-curve room. In stark contrast, of the few mixers who work in rooms calibrated to a "house" curve, working in such a dub-stage represents very little of their overall work time. However, of those who work primarily in SMPTE calibrated dub-stages, approximately half of them admitted that here are two-stages of equalization available in the B-chain; one for when the room is being calibrated and the other to be used at the discretion of the mixer. Again, this type of variability does not lend well to translation as the standard had intended and questions whether mixers feel that the X-curve is actually hindering translation in some way.

Figure 3.8 shows that for those who work on the two largest sized dub-stage (at or larger than 21m x 15m x 9m) spend a majority of their time on a stage calibrated to the SMPTE X-curve. This is not surprising as most, if not all, of the large stages started as film-based stages which typically necessitated Dolby Certification in order to print master in any of the Dolby encoding schemes. To obtain Dolby Certification, the dub-stage had to be tuned to the SMPTE X-curve by a Dolby licensed technician. The connection between Dolby and the SMPTE X-curve is covered in section 2.2.

It might seem counterintuitive to find that 60% of the work carried out on some of the smaller dub-stages was completed in X-curved tuned rooms. These rooms are typically too small to be film rooms and therefore do not share the same requirements of the larger stages. However, as the costs to build large stages went up as post-production budgets went down, there were requirements to build smaller dub-stages that could complete cinematic work. On some occasions, dub-stages reserved for television mixing were repurposed into film stages. Being built or repurposed for film work, these smaller rooms were required to be tuned to the X-curve.

3.5.2 "Modified" SMPTE X-curve Dub-stages

The mixers who work on "modified" X-curve dub-stages (not inclusive of those who were not sure) were asked if they knew how and why their stages were modified. Of these 12 mixers, 7 replied that the modifications were made to the high frequency range typically to create a "less severe" high end roll off than the standard SMPTE X-curve. The other 5 mixers knew only that their stage had been modified in order to improve sonic performance but were not sure how the modifications took form. When asked about translation, none of these mixers stated that they had any problems with their mixes playing back correctly in commercial cinemas.

3.5.3 "House" Curve Calibrated Dub-stages

Of the 6 mixers who work on a "house" calibrated dub-stage, 4 stated that less high frequency roll off than the X-curve and no low frequency roll off were used to calibrate the stages they work on. Two other mixers stated that the "house" curve used was client specific. These mixers offered no comments on the quality of translation between their dub-stage and the commercial cinemas.

3.5.4 Hearing Loss in the Professional Audio Community

As noted in section 3.4, only 66% of the survey participants stated that they had their hearing check with the 5 years prior to the survey while 34% either had their hearing check longer ago or not at all. Even with these percentages, 80% responded that their hearing was found to be normal, though it would be difficult for most to state that with complete certainty.

Hearing loss within the re-recording mixing community is a topic that is often avoided, especially in light of the number of years it takes to become a top mixer and the prestigious awards that these mixers are often given. With the multitude of years on a dub-stage and the often advanced age of the mixer, these combine to hasten hearing loss.

The House Ear Institute (HEI) in Los Angeles, California carried out large-scale audiological testing at four of the largest conventions in the US that attract audio engineers: Audio Engineering Society (AES), the National Sound Contractor Association (NSCA), the National Association of Music Merchandisers (NAMM), and Lighting Design International (LDI). This work was carried out by HEI over a 8 year period from 1997 to 2005 and included 6500 audiograms collected from 4816 attendees (some attendees had their hearing rechecked over successive years) all of whom identifies themselves as professional audio engineers or working in a related industry (Cruz, 2005).

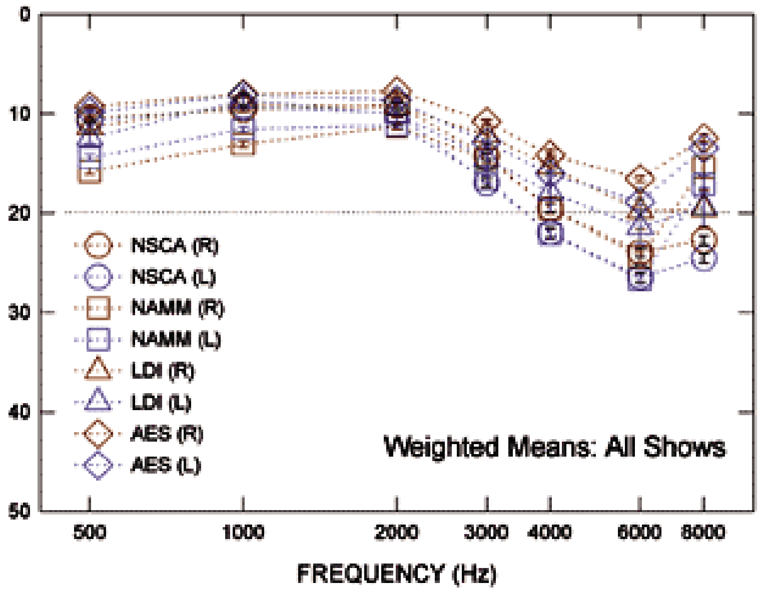


Figure 3.17: Weighted mean values of air-conduction (pure tone) thresholds for all attendees across the conventions, from (Cruz and Fisher, 2004)

Cruz and Fisher found that in comparison to the ISO 1999[3], the test participants have poorer air-conduction thresholds than their age and gender matched peers (Cruz and Fisher, 2004) . Their research went on to reveal that, as the test participants age goes to 40 years old and higher, the threshold shift compared to age and gender-matched peers grows even larger, with men having the largest discrepancy. Of interest to note, 63% of the participants in the re-recording survey were over the age of 40. The overarching finding from the audiological data was that all

[3] ISO 1999:1990 - Determination of occupational noise exposure and estimation of noise-induced hearing impairment

age and gender groups shared a distinct loss in sensitivity in the 4 to 6kHz range. Though these audiological tests covered many types of audio engineers, re-recording mixers are certainly part of this demographic.

The hearing tests conducted by Cruz and Fisher were standard pure tone audiometric tests and clearly unveiled tonal hearing loss in the audio community however, there is a question as to how this would translate to a re-recording mixers ability to clearly hear dialog. The ability to predict the loss of speech intelligibility from pure tone audiometry has been examined over years of research (Anjos et al., 2014), (Fletcher, 1950). Speech Recognition Thresholds (SRT) have been closely correlated to pure tone thresholds using various combinations of scores from the 0.25, 0.5, 1k, 2k, and 4kHz tone tests (Coren and Hakstian, 1994). In their research, Anjos et al found that SRT score has the strongest correlation with the average frequencies 500, 1k and 2kHz, while Speech Recognitions Scores (SRS) has the highest correlation with an average which include these three frequencies as well as 3k and 4kHz. Smoorenburg noted that hearing loss above 1kHz was typically related to the loss of speech reception in noise while hearing loss at and below 1kHz was related to the loss of speech reception in quiet (Smoorenburg, 1992). However, additional research has found that pure tone tests cannot give a complete picture of the loss of speech intelligibility. Researchers have observed that there is little to no correlation between pure tone test results and the ability for a listener to distinguish speech in background noise, as measured by the Speech in Noise (SiN) test or Hearing in Noise Test (HiNT) (Moore et al., 2014), (Vermiglio et al., 2012).

Dubbing stages are required to follow the Occupational Safety & Health Administrations (OSHA) guidelines for work place noise exposure, however, anecdotal information has shown that these guidelines are rarely followed or enforced. Table 3.2 shows OSHA's permissible levels and exposure times for work-place noise. OSHA considers noises that have variations in maxima at intervals of 1 second or less to be continuous.

Duration per day, hours	Sound level dBA slow response
8	90
6	92
4	95
3	97
2	100
1 1/2	102
1	105
1/2	110
1/4 or less	115

Table 3.2: Permissible noise exposures; from OSHA (1983)

OSHA recommendations are based on an 8-hour work day and allow for a 5dB increase in permissible level for each instance that the exposure time is halved (known as an "exchange rate"). The latest ANSI standard, ANSI/ASA S3.44-2016/Part 1, prescribes a 3dB exchange rate as a general guideline (ANSI, 2016).

Temporary threshold shift (TTS), where a listener experiences a rise in the auditory threshold after exposure to noise, can take hours to days to resolve completely (16 to 24 hours is the time typically used in hearing research) (Mathur and Roland, 2016). Chronic exposure to noise without sufficient time between exposures to allow for recovery can cause threshold shifts to become permanent. With demanding schedules and long hours in the studios, re-recording mixers typical exceed an 8-hour work day and often work in continuous exposure to 90dBA of "noise" or higher on a daily basis without adequate recovery time. Further to this point, it has been shown that often hearing loss is caused by repeated exposure to noise above 85dBA over long periods (NIDCD, 2016).

The OSHA guidelines do not provide for any specific age-related corrections when taking into account appropriate noise level and exposure durations, though this practice is recommended by the American College of Occupational and Environmental Medicine (ACOEM, 2007). One of the reasons given by ACOEM is shown in Figure 3.18, which depicts the progression of hearing loss following exposure to 90dBA of noise averaged across the work day. The data shown illustrates hearing loss for white males at ages 20, 30, 40, 50 and 60 years with 0 - 40 years of exposure.

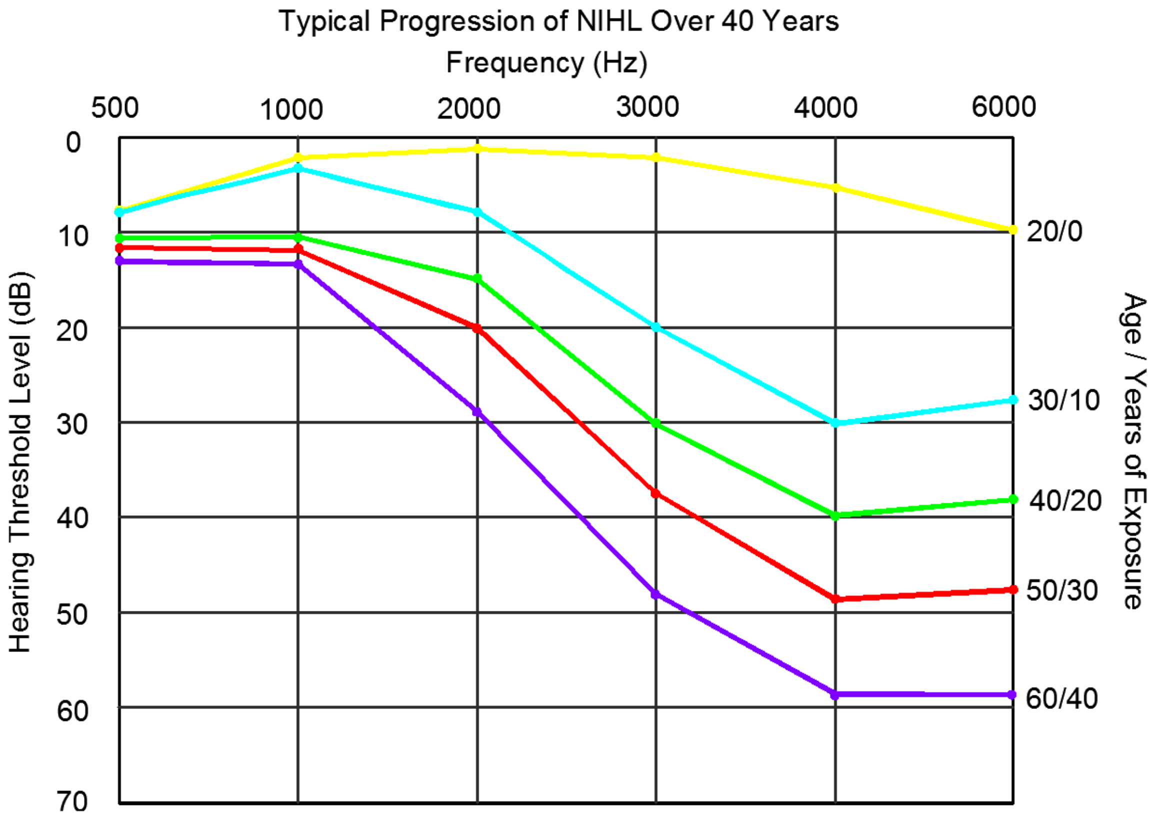


Figure 3.18: Progression of hearing loss, white males aged 20, 30, 40, 50 and 60 when exposed to 90dBA over a work day for 0 to 40 years; derived from (ANSI, 1996a)

Many of the survey participants had worked for several years in the movie sound industry with 82% having work in production audio for over 10 years and 52% having worked specifically as a re-recording mixer for over 10 years. This places a majority of the survey participants in a noisy work environment for many years and well into advanced years of age. With the 4kHz to 6kHz dip in hearing shown by the Cruz and Fisher data, added to that of typical age related hearing loss, and the impact of years in a noisy work environment, one cannot help but to wonder what influence this has on the mixer's ability to judge spectral quality well enough to effective mix full bandwidth audio.

Several researchers have linked both noise-induced threshold shift and age related hearing loss to the loss of speech intelligibility (Pulkki and Karjalainen, 2015), (Humes, 2015), (Dubno et al., 1984), (Skinner, 1980). Pulkki and Karjalainen further noted that music listening can suffer even with small hearing impairments. They went on to recommended that the current A-weighted levels used should be dropped down to 80dB to 75dB in order to assure hearing health.

3.5.5 The Possible Effects of Learning

Re-recording mixers are exposed to their listening environments for very long periods of time on a daily basis, possibly leading to this having the strongest effect on their overall audio preferences. The concept of 'mere exposure effect', where repetitious exposure to a stimulus can lead to a positive effect, has been well covered in psychological research literature (Baddeley et al., 2014), (Zajonc, 2001), (Bierley et al., 1985).

In 1956, Kirk undertook an experiment to understand a similar concept, 'implicit learning'. In this experiment, Kirk repeatedly exposed 210 college students to music which had been band-limited into one of four frequency ranges, as shown in figure 3.19.

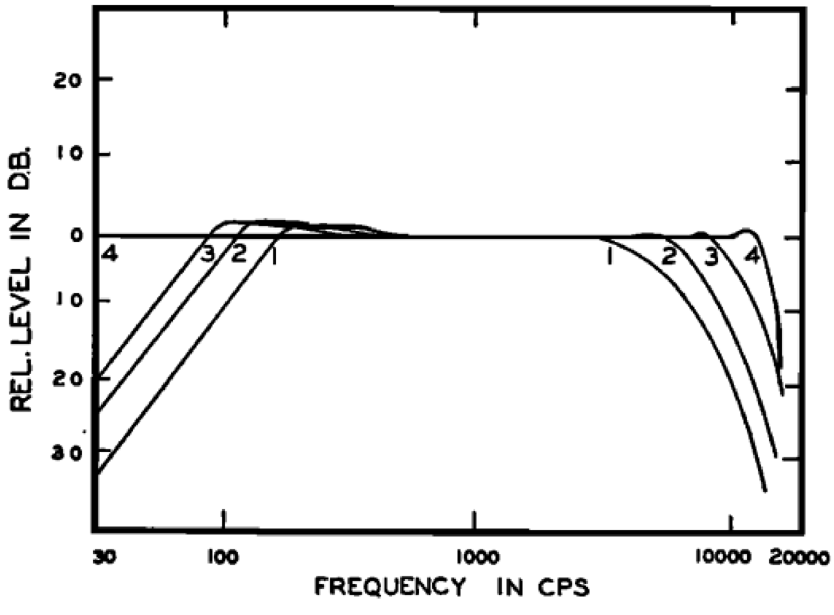


Figure 3.19: Four response curves used in frequency range preference test from (Kirk, 1956)

The results of this experiment showed evidence that students who repeatedly listened to music reproduced over a restricted frequency range came to prefer this range when given a choice.

The results of Kirk's experiment are compelling, though conclusions should not be drawn solely from a single experiment; especially one involving the complex processes of human learning.

Research in the area of audio preference based on exposure time has produced conflicting results (Olive, 2012), (Dougherty, 2009). However, the concepts of mere exposure effects and implicit learning should not be overlooked when consider what influence these may have on re-recording mixers adaptation to thse SMPTE X-curve and how it affects the way they mix.

3.6 Conclusion

Re-recording mixers compensating the for high frequency roll-off is far from being a new concept, as was noted by Allen with regards to the old Academy curve (Allen, 1975). Allen further noted that the ways in which compensation was applied varied from film to film and from studio to studio.

From the overall results of the survey, it appears that re-recording mixers have become acclimated to working with the SMPTE X-curve or some form of it. However, it is clear that when they do, there is a prominent trend to compensate for it by increasing the overall level of the mix in the higher frequencies. Given the average age of the mixers surveyed and the numbers of years that they have worked in sound, some of this compensation may be due to hearing loss. With either of these issues in mind, there is a strong question as to whether the audience will agree with the mixer that the final soundtrack has effectively translated from the dub-stage to the commercial cinema.

The overall theory of translation is continuously called into question by the variable means in which some mixers compensate for the X-Curve, the difference in calibrations carried out on the dub-stages they work on, and the perceived level of translation that they feel they are achieving.

3.7 Summary

In this chapter, a survey of re-recoding mixers was undertaken in an effort to understand their usage and sonic preferences with regards to working on a dubbing stage. First, data was collected on demographics such as age and the numbers of years mixers had worked in the production / pro-production sound industry, and what country they are working in. Next, the survey participants were asked about their work and work-place. This included questions on the types of projects they worked on, the average size of the dub-stage that they most often worked on and how that stage is typically calibrated. Only mixers who worked on calibrated dub-stages were ask the next series of questions regarding how the rooms they work in were calibrated and how they perceived the resultant sonic quality. Finally, a series of open-ended questions were asked of the former set of mixers in order to gain further insight to the previous answers they had given and to see how the data may be related.

A comparative analysis was carried out on various sections of data in an effort to get a complete understanding of how the mixers' demographics, work environments and mixing techniques may influence the way their mixes are created within a calibrated dub-stage. Data from various sources on hearing loss were also considered in this analysis.

A brief review of occupational noise exposure and hearing loss as it pertains to re-recording mixers was discussed. In addition, the concept of a learning effect with regards to mixer adaptation to the X-curve was explored.

Chapter 4 - Validating the Binaural Room Scanning (BRS) Method for Cinema Research

4.1 Introduction

Even in the early years of cinema calibration, there was a belief that any method of assessing a cinema loudspeaker system should include the effects of subjective factors. As early as 1946 the Theatre Sound Standardization Committee of the Motion Picture Research Council undertook efforts to correlate subjective listening tests with cinema loudspeaker measurements (Hilliard, 1949).

Around the time that the early SMPTE and ISO standards were being written in the late 1960's, Vlahos somewhat ironically noted that new techniques and technology for acoustic measurements should ultimately replace subjective listening test methods for determining cinema acoustic response (Vlahos, 1969). However, as Toole has noted on various occasions, the human hearing mechanism should not be replaced by an omni-directional microphone taking steady-state measurements, such as the one required by the SMPTE standard. This system simply cannot make allowances for sounds arriving from different directions at different times and the spectral variations among these sounds. The human hearing system, which includes the brain, is much more analytical and responds differently to sound than a microphone connected to an analyzer (Toole, 2015), (2008), (2006).

For this reason, assessments of target response curves, regardless of which curve is in question, should not ignore the use of subjective testing as a valuable research tool. However, conducting subjective studies in situ is not always possible for logistical reasons. The work in this chapter involved validating a Binaural Room Scanning (BRS) system as a possible replacement for in situ testing for full-scale scientific studies on listener preference for room target response curves.

The first stage of this research investigated the effect of playback method, BRS and in-situ listening, on two smaller rooms which both model spaces where listeners watch movies; a living room and a purpose-built home theatre. The second stage moved this research into a cinema-sized lab in order to study the effectiveness of the BRS system for cinema research.

This chapter reports the results of validation tests which made use of a custom BRS system that was originally developed for research and evaluation of different loudspeakers and different listening spaces (Olive et al., 2007). To validate the performance of the BRS system, listening evaluations of different in-room equalizations of a 5.1 loudspeaker system were made both in situ and via the BRS system. This was repeated using three different loudspeaker systems in three different sized listening rooms.

Work in this chapter has previously been presented as the following:

GEDEMER, L. & WELTI, T. 2013. Validation of the Binaural Room Scanning Method for Cinema Audio Research. AES 135th Convention, New York, NY: Audio Engineering Society. (Gedemer and Welti, 2013b)

GEDEMER, L. & WELTI, T. 2013.Validation of the Binaural Room Scanning Method for Cinema Audio Research. Reproduced Sound 2013, Manchester, UK: Insitute of Acoustics. (Gedemer and Welti, 2013a)

GEDEMER, L. 2015. Subjective Listening Tests for Preferred Room Response in Cinemas - Part 1: System and Test Descriptions. AES 139th Convention, New York, NY: Audio Engineering Society. (Gedemer, 2015b)

4.2 Motivation

Conducting subjective listening tests in situ within a commercial cinema or professional dub-stage can prove logistically difficult. Binaural Room Scanning (BRS) affords many practical, logistical and methodological benefits over in situ listening tests, including the elimination of sighted biases related to the objects being tested (Mackensen et al., 1999). BRS provides the benefit of allowing larger samples of listeners to participate in testing, something that professional dub-stages and public cinemas cannot easily accommodate. It allows many of the acoustic factors of the loudspeaker and/or listening space to be captured, reproduced and evaluated in a controlled and reliable manner. BRS also allows different experimental variables to be easily manipulated and presented to listeners over different time periods to study how they learn and adapt to varying loudspeakers and acoustic environments (Olive et al., 2007), (Pellegrini et al., 2007), (Bech et al., 2005), (Christensen et al., 2005), (Pellegrini, 2001), (Horbach et al., 1999). For the experiments presented in this chapter, this particular benefit allowed the control of independent variables such as different rooms, seating positions and target room curves, with the ability to have listeners switch at-will between target room curves being a key aspect.

BRS systems have shown a significant positive correlation with responses to in situ listening tests for loudspeaker preference (Olive et al., 2007), loudspeaker setups (Koehl et al., 2011), dub-stage and cinema reproduction over headphones (Smyth et al., 2010), and automotive audio tests (Olive and Welti, 2009), (Bech et al., 2005). The work of Koehl and his team further demonstrated that headphones enabled better consistency between the listeners than the loudspeaker setups used. However, in order to move forward with utilizing a BRS system for cinema research the system needed to be validated and shown that it was appropriate for such use.

4.3 Overview of Binaural Room Scanning

In past decades, psychoacoustic studies have utilized binaural capture and playback systems for a wide variety of experiments including those of loudspeakers (Olive et al., 1994) listening rooms (Olive et al., 1995) concert halls (Schroeder et al., 1974) and automotive audio systems (Granier, 1996). Lacking the equipment or ability to track the listeners' head movements, these earlier studies utilized single-position Binaural Room Impulse Responses (BRIRs) but noted the limitations of static position playback, such as localization errors. Several previous experiments (Mackensen et al., 1999), (Thurlow and Runge, 1967), (Thurlow et al., 1967), (Wallach, 1940), showed the importance of head rotation in the horizontal plane for accurate sound rendering. With consideration of these findings, the audio capture and playback in a BRS system is reliant on, and must be performed with, the actual head orientation angle. In doing so, confusions that can lead to unreliable test results, such as front-back reversals and internal-external localization as described by Blauert (1997), can be widely avoided.

In subsequent years head-tracking technology evolved and, when teamed with a rotatable binaural microphone, became a cost-effective tool for listening tests now known as Binaural Room Scanning. BRS can trace its origins back to 1999 when Ulrich Horbach, then a researcher at Studer working on digital mixing consoles, set out to develop a system that would offer professional sound engineers the ability to remotely monitor their mixes via headphones through a scanned copy of their studio environment. BRS is defined as a method of capturing a binaural representation of a sound source and the room it is located in using a dummy head with binaural microphones in the ears and then later reproducing it over a pair of calibrated headphones. Capture is undertaken by using a manikin equipped with microphones placed at its ears in an effort to mimic a human listener. The manikin head is equipped with a stepper motor which allows the capture of BRIR at various rotational angles. With this system, multiple BRIRs are made at differing head angles which are then stored separately as data files. During playback, music or other types of audio signals are processed through the appropriate head-oriented BRIR filters using a real-time convolution engine. The BRIR processed audio signal is reproduced through high-quality headphones equipped with a low-latency head-tracking system (Horbach et al., 1999). This method is aimed at ensuring that the in-room signals are presented in a spatially-stabilized manner as listeners turn their heads, thereby avoiding most localization errors.

Figure 4.1 shows simplified block diagrams of the BRS scanning and playback systems.

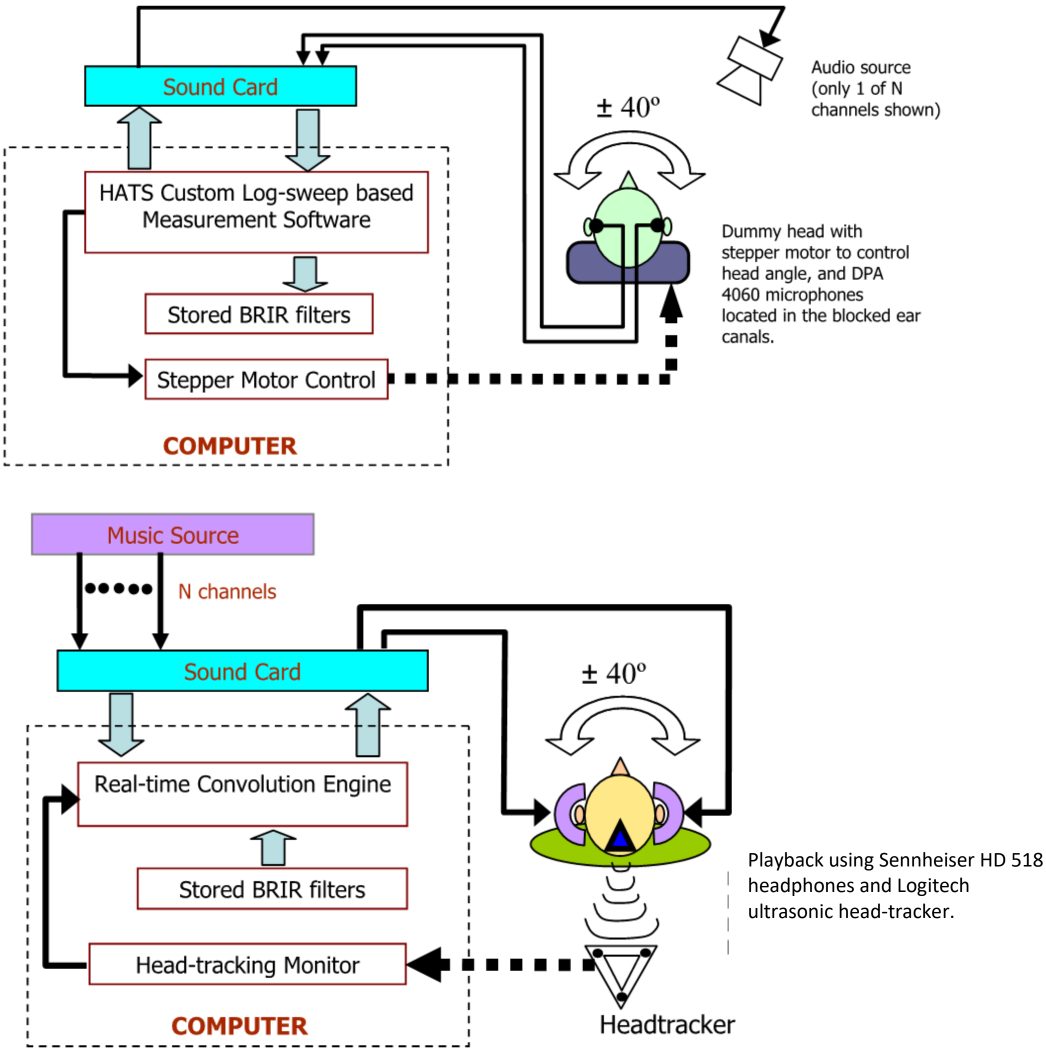


Figure 4.1: Block diagrams of the binaural room scanning (BRS) measurement and playback system

4.4 Potential Sources of BRS Errors

Olive and his team (2007) identified the source of potential errors with the BRS system and defined them as:

1) Measurement errors – These are related to the repeatability and accuracy of BRS measurements and the playback of binaural signals.

2) Anatomical errors – These are related to the differences in the shape and size of the manikin's head/torso/pinna versus listeners'.

3) Positional errors – These are related to positional differences between the manikin's head at the time of the BRS measurement, and the listener's head observed in situ.

4) Cognitive-related errors – These are related to errors from inaccurate BRS reproduction of non-auditory (e.g. tactile and visual) stimuli.

These errors were categorized as directionally-dependent and directionally-independent with sub-categories of non-individualized and individualized, as shown in figure 4.2.

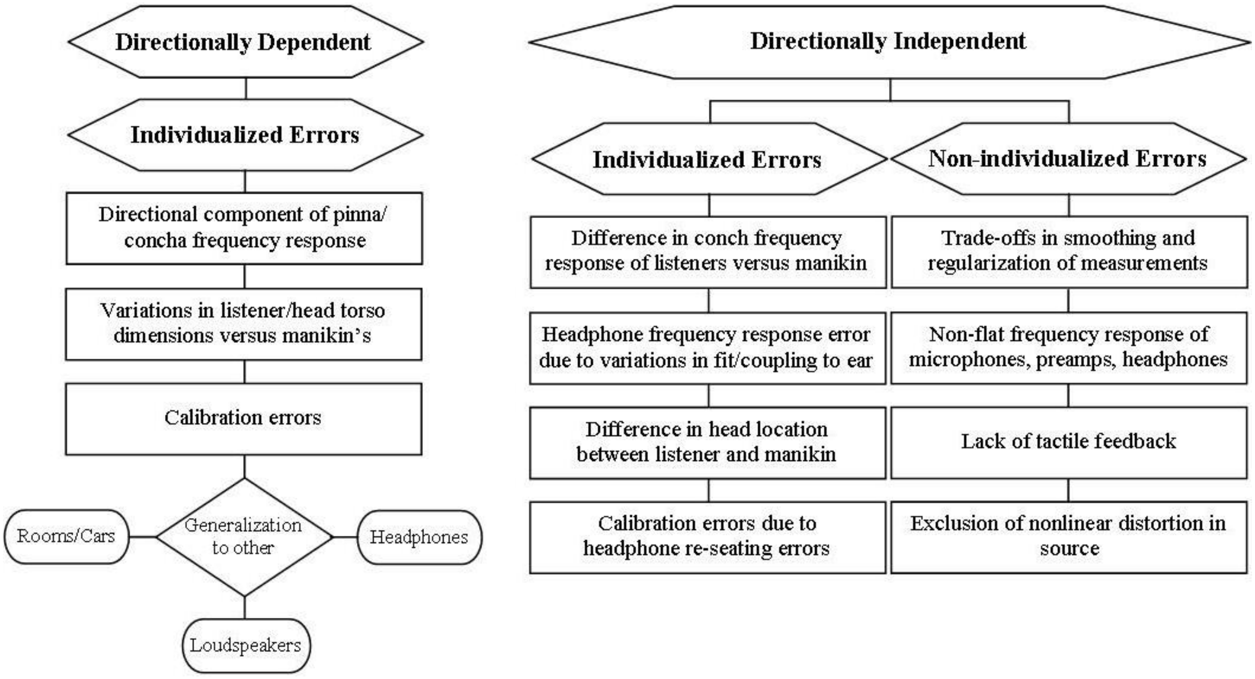


Figure 4.2: Source of BRS errors; derived from (Olive et al., 2007)

The directionally dependent errors are those measurement errors which change as the angle of the sound arriving at the binaural manikin's ears changes. These types of errors can all be classified as "individual". In contrast, directionally independent errors are not influenced by the location of the listener's ears or the sound source. Directionally independent errors can be further broken down into individualized and non-individualized errors.

4.4.1 Directionally Dependent Errors

The first two sources of directionally dependent errors are directly related to the physical differences between the listener and the manikin. One error is based on the differences in pinna / concha shape while the other is due to differences in head / torso shape and size. These errors change with each individual and therefore cannot be corrected through the use of a globally applied filter. The third source of directionally dependent errors stems from the BRS calibration process itself, which is further explained in section 4.5.2. This error is based on the idea that the calibration of the BRS system to one loudspeaker, room or headphone may not be appropriate to apply to other like objects. The reason that this error falls into the individual classification is that it will change from one object to another, such as calibrating to a particular set of headphones and then applying the calibration filter to a different set of headphones.

The first error is difficult and time consuming to correct as it requires that each test subject's pinna / concha be measured and taken into consideration during calibration. There are also questions as to whether individual pinna measurements are required. Researchers have found that, when using head-tracked binaural playback, listening test results on localization accuracy are the same as when the listener sits in the manikin's position (Romigh et al., 2015), (Begault et al., 2001), (Horbach et al., 1999). Research has also shown that listeners tend to adapt to non-individualized binaural recording and, with repeated listening tests, further adaptation helped to resolve many of the initially perceived errors (Zahorik, 2002), (Minnaar et al., 2001).

The second error has been identified as changes in arch-shaped notches in the mid-band frequencies that lessen as the frequencies get higher (Algazi et al., 2001). However, more perceptual research on this type of error needs to be conducted to determine if it is significant. In the case of the BRS experiments conducted for this thesis, the third error is of no consequence as the headphones utilized in the calibration procedure were the same ones used for the listening tests.

4.4.2 Directionally Independent Individualized Errors

The first source of directionally independent individualized error is caused by the differences in transfer function between the manikin's and the listeners' conchae. This error is somewhat directionally independent, with the non-directional (frequency response) part of this error being corrected during calibration. The second source of error is related to the change in frequency response of the headphones due to varying fit between listeners and repeated re-seatings of the headphones. Compensating for varying fit is difficult, as it once again requires measurements for each individual listener and there remains a question as to the magnitude and perception of this error. Informal listening tests by the author and her colleagues have indicated that this is not mandatory, although further research is needed. Attempts are made to correct for a degree of re-seating errors by undertaking the energy averaging of several re-seated measurements on the binaural manikin prior to calibration.

Errors caused by the difference in manikin and listener seating locations in these experiments were minimized by carefully measuring and marking the locations by using a hand-held laser distance measurer (Bosch GLM 30 MP). Though exact placements of the listeners can be difficult, it is believed that any residual error is probably unnoticeably small.

4.4.2 Directionally Independent Non-individualized Errors

Directionally independent non-individualized errors are typically the easiest and least costly to correct. Non-linear frequency responses from the manikin microphones, microphone pre-amps and headphones can all be corrected with a single filter set and in a one-step process, as is shown in section 4.5.2. Regularization and smoothing of the signals during the calibration process, as described in section 4.5.2, will also introduce some amount of error which is difficult to avoid but a necessary tradeoff.

The lack of tactile and/or visual cues can create another source of directionally independent non-individualized errors. These errors were present in the experiments reported in this thesis; notably the lack of tactile feedback from the subwoofers which were not present during BRS playback and the absence of visual feedback with respect to the rooms which were scanned.

Concerns regarding visual errors brought on by a type of cognitive dissonance when comparing BRS and in situ playback have been shown to be unwarranted in similar tests for automotive audio (Postel et al., 2011), and loudspeaker auditioning in varying rooms (Olive and Martens, 2007). In their research, Gros and Chateau showed that environmental and/or visual cues have little influence on audio quality scores (Gros and Chateu, 2002). Beresford et al. further look at the significance of differences in listening environments when comparing a listening room to a car with final results of their listening tests broken into two groups: trained and untrained listeners. For untrained listeners, the lack of visual accord did not have a significant effect on quality scores. For trained listeners, there was an interaction between stimulus and visual context but it proved to be of a very small magnitude ($\eta^2 = 0.017$) (Beresford et al., 2006a), (Beresford et al., 2006b). These studies also lacked the tactile feedback of a subwoofer, which did not appear to have an influence on listener's ratings. However, this issue warrants further research.

4.5 BRS System Validation

The performance target for the BRS system is relatively straightforward: the system should accurately capture a listening space and deliver the identical signals to the listeners' ears as if they had been receiving those signals in situ.

The specific method of validating the BRS system as covered in this chapter was chosen due to its relationship to how the author ultimately intended to use it. Its primary application was to be for conducting research into cinema in-room target response preferences with respect to varying program material, room size and seating positions within rooms. The experiments in this chapter addresses how well the BRS system performs when used as a substitute for in situ listening, specifically when preferences to room response curves are in question. The main research question was: "Are there significant differences in listeners' room response curve preferences using BRS playback compared to in situ playback?"

4.5.1 Binaural Room Scanning

The BRS system utilizes a custom-built full-body manikin built by Harman Becker Automotive and outfitted with DPA 4060 omni-directional microphones placed at the blocked canals of the binaural manikin. The location of the microphones was chosen because it produces the least inter-listener variability in head related transfer function (HRTF) measurements (Blauert, 2005), (Hammershøi and Möller, 1996). The neck of the manikin was equipped with a stepper motor so that its head could be rotated within the horizontal plane at selectable resolutions starting at 1º. Figure 4.3 shows the binaural room scanning manikin used for these experiments; "Sidney", which was named after Sidney Harman.

Figure 4.3: Full-body binaural room scanning manikin 'Sidney'

The performance of Sidney is comparable to commercially available head and torso simulators, and falls within the IEC 60318-7 tolerances[4] (IEC, 2011). The HRTF responses of Sidney and 3 commercially available head and torso simulators are shown in figures 4.4 and 4.5.

[4] Electroacoustics - Simulators of human head and ear - Part 7: Head and torso simulator for acoustic measurement of hearing aids

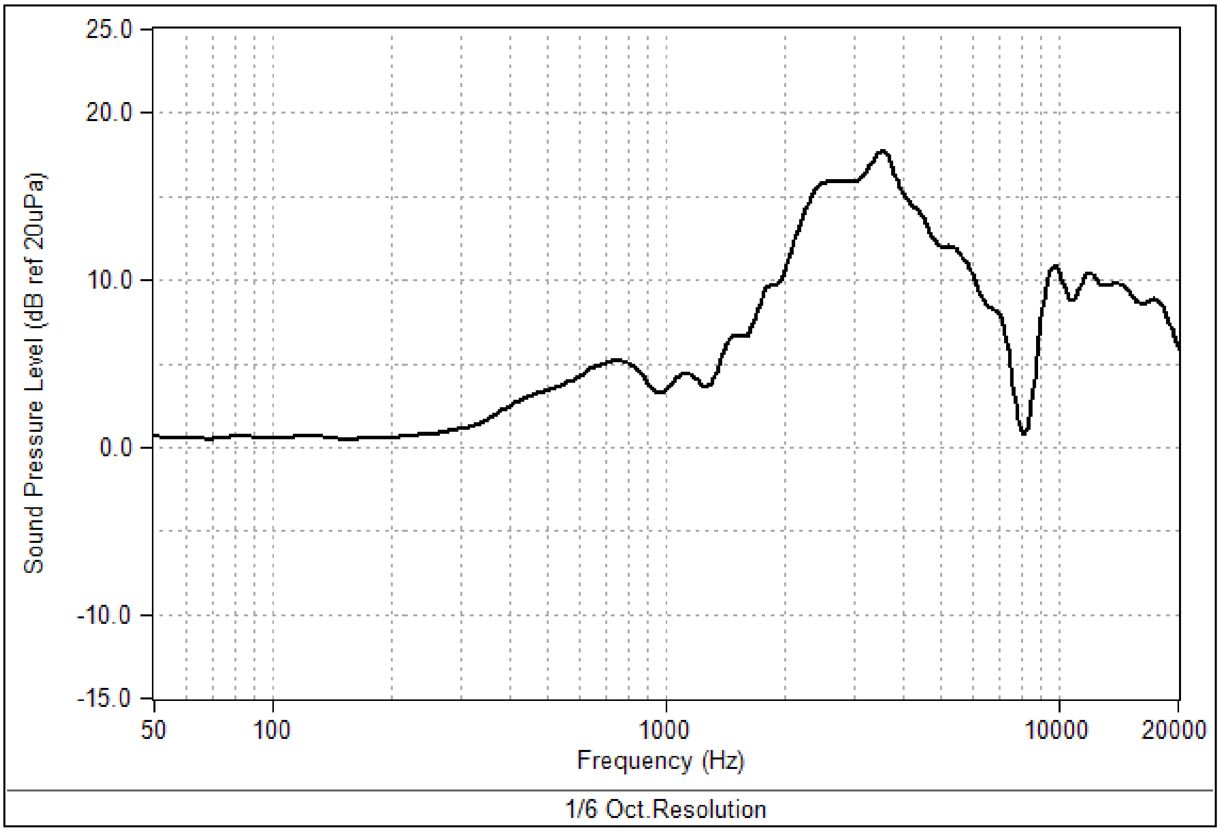


Figure 4.4: Free-field HRTF of Sidney Head and Torso Simulator at 0° Azimuth, Left Ear Presented

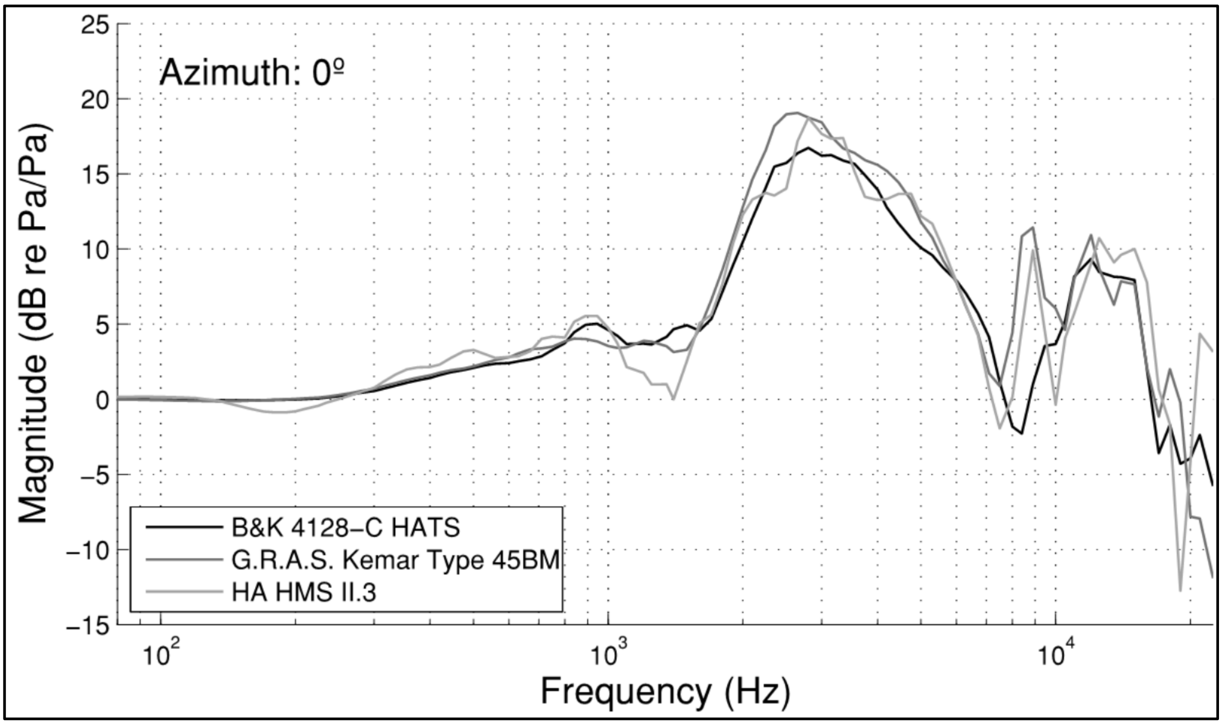


Figure 4.5: Free-field HRTFs of Brüel & Kjær, G.R.A.S. and Head Acoustics Head and Torso Simulators at 0° Azimuth, Left Ear Presented; from (Snaidero et al., 2011)

The system computer (Lenovo Thinkpad T420s) ran a custom Matlab program[5] which generated the test signals and recorded the measured sets of BRIRs. The test interface for this program is shown in appendix E. The test signal used was a swept sine wave which gave an optimal signal-to-noise ratio at low frequencies (Farina, 2000). The signal-to-noise was further improved by averaging several repeated measurements. The playback levels for the swept sine waves were adjusted to create an SPL of approximately 80dBA at the listening position.

The BRIRs were measured for a range of horizontal head orientations at the listener's location in three different rooms using a 5º angular resolution. Research has shown that sensitivity to horizontal rotation depends highly on the audio material utilized. In the case of pink noise, it has been noted that a source rotation of 2.5º could make an audible difference (Hiekkanen et al., 2009). This observation is congruent with the findings of Blauert, who observed that localization blur in the horizontal plane can be less than 2.5º during listening tests with pink noise (Blauert, 1997). However, Hiekkanen noted further that even a change in source direction of 10º was found difficult to notice with certain music signals. In fact, the angular resolution for a BRS scan is not directly comparable to localization blur as they are two different issues. The latter relates to a stationary head and moving source, the former relates to a stationary source and moving head. In work by Welti and Zhang (2010), it was found that listeners had difficulty discriminating between 5º, 10º and 15º angular resolution for BRS playback of music signals. For these reasons, a horizontal resolution of 5º was deemed to be acceptable, especially in light of the fact that source localization was not a parameter under test. Binaural impulse responses were taken in 5º increments from -40º to +40º to render a total set of 17 BRIRs. The horizontal limits of +/-40 º was chosen based on previous studies having shown that head-tracking should not be restricted to under +/-30º from center line, in order to achieve a sense of naturalness (Laitinen et al., 2012).

During playback, program material was convolved with the BRIR filter sets using a Matlab-based real-time convolution engine. The head-tracker monitored the angular position of the listener’s head and sent the updated coordinates to the BRS playback software that switched to the appropriate BRIR filters corresponding to that angle. Interpolation algorithms embedded

---

[5] The program is owned and copyrighted by Harman International and therefore the author, who is under non-disclosure agreements with Harman, is not allowed to release any details of this program.

into the Matlab code provided smooth transitions between the different binaural filters. Correction filters for the combined manikin head microphones and playback headphones, as well as the in situ loudspeaker responses were also applied, as described below.

4.5.2 Calibration of Manikin Microphones and Playback Headphones

For the BRS tests, Sennheiser HD518 circumaural headphones were chosen as the reference headphones. These headphones are free-air equivalent, with open-backs and low acoustic source impedance, which make them suitable for the reproduction of binaural signals captured at the blocked meatus of the manikin's ears (Møller et al., 1995). There were a few criteria that further made these headphones a good choice for this task including their consistent fit and seal on the listeners' head and ears, which are necessary for repeatable sound reproduction. In addition, these headphones have a relatively wide bandwidth and smooth frequency response (not flat, but smooth). Møller et al noted that, ideally headphones for binaural playback should have a linear or easily correctable transfer function.

The Sennheiser HD518 headphones were first measured on a GRAS 45CA test fixture equipped with an IEC 711 coupler in order to verify their performance. The headphones were then re-measured on the binaural room scanning manikin. This calibration technique attempts to correct the entire BRS measurement system and playback chain, including the headphones, using a set of filters as shown in figure 4.6.

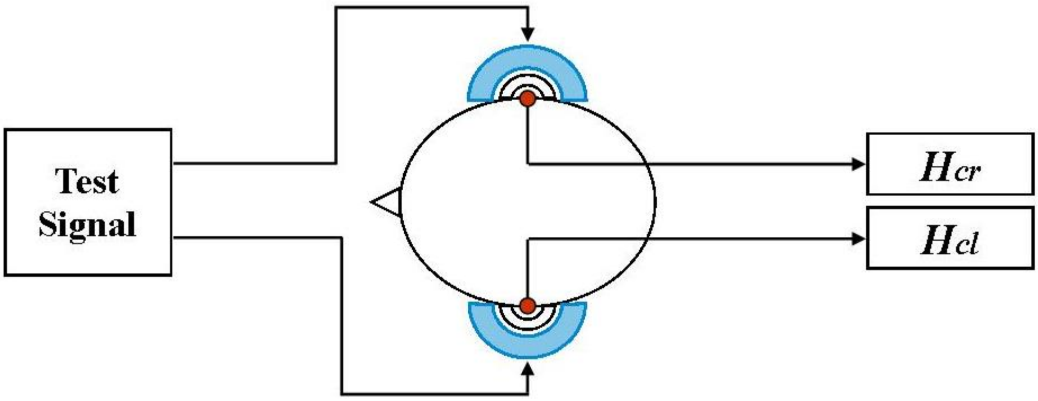


Figure 4.6: Simultaneous calibration of manikin microphones and headphones; stored as filters ***Hcr*** and ***Hcl***

A study by Schärer and Lindau observed that, when using non-individualized BRIRs and non-individualized headphone correction filters, the process of correcting the manikin microphones and headphones simultaneously provided the best possible frequency response compensation (Schärer and Lindau, 2009).

The correction filters (***Hcr*** and ***Hcl***) were created in Matlab and began by smoothing the notches in the frequency domain prior to creating an inverse transfer function. Smoothing deep notches in the frequency response serves many purposes including the avoidance of ringing artifacts caused by excessive peaks in the inverse filter, prevention of adding additional noise, and reductions in computational load. During processing, a nonlinear regularizing function is used to give priority to peaks. With this approach, there will still be some deep notches in the resultant post-calibrated frequency response; however, it has been found that these types of dips are less audible than peaks and are therefore less of a concern than the peaks that would be created had the notches been left unsmoothed (Toole and Olive, 1988), (Bücklein, 1981).

The correction filter is then band limited to minimize errors in the upper frequencies where deep notches are most prominent, and to avoid excessive boosting in the lower frequencies where the response of the headphones rolls off. The accuracy of correction (or error minimization) is thereby concentrated in a specific all pass region, in this case the region between 25 Hz to 14k Hz. The frequency response of the filter is then smoothed once again to correct for discontinuities created by abruptly band limiting the all pass region. A cepstral processing technique (Matlab rceps(x)) was used to derive a minimum phase impulse response from the magnitude only correction filter (Oppenheim and Schafer, 1975).

Figure 4.7 shows the resultant measured frequency response of the left and right channels. These curves were based on an average of six reseats of the headphones with the goal of minimizing errors related to headphone positioning on the manikin. Figure 4.7 shows the response of the headphone before (dotted curves) and after flattening (solid curves).

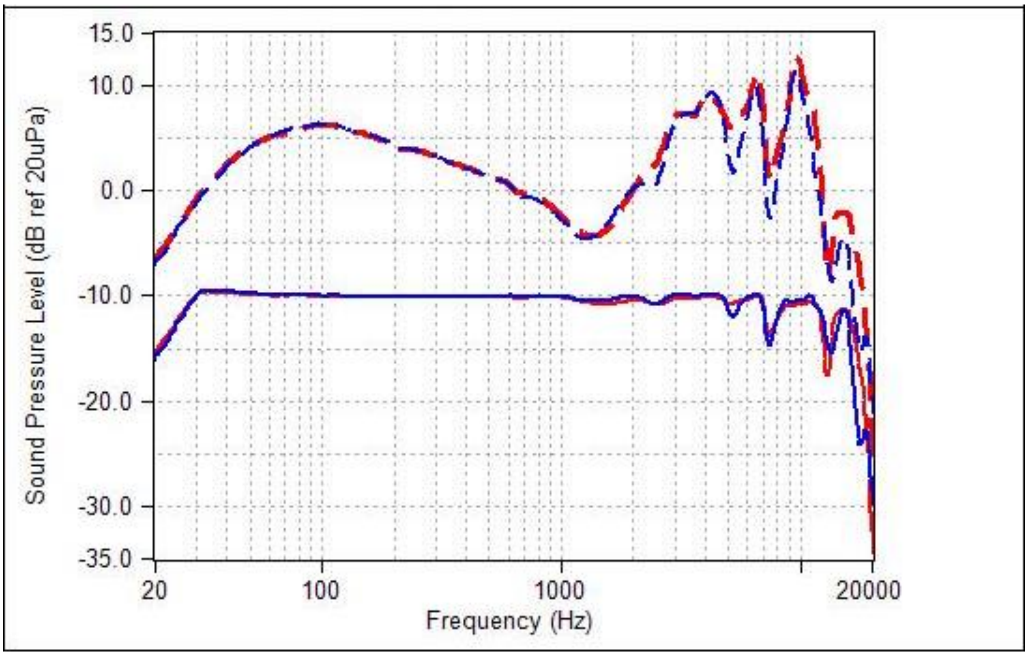


Figure 4.7: Measured frequency response of the left and right channels: dotted curves show the frequency response of the left and right channels of the Sennheiser HD518 headphones, the solid lines show the same after equalization

Riederer investigated the characteristics and repeatability of headphone-to-ear transfer functions (PTFs) and noted that deviations due to repetitions in headphone re-seatings (5 resets) was less than 3dB below 7k Hz and around 3dB up to 13k Hz (Riederer, 2005, Riederer, 1998). His results were obtained using circumaural headphones (Sennheiser HD580), similar to the ones used in these experiments. This underlined the need to band limit the equalization of the HD518 headphones to a range of 25 Hz to 14k Hz; taking into consideration that above 14k Hz can be problematic due to the increasingly variable response of headphone re-seatings.

4.5.3 The Selected Listening Rooms

Three different sized rooms were utilized for these experiments; the Living Room, Eargle Theatre, and Arena Lab, all of which are demonstration rooms at Harman International's Northridge campus. These rooms were specifically chosen as they are representative of typical environments in which people watch movies. One room, the "Living Room", with dimensions of 2.4m (H) x 4.3m (W) x 6.5m (L), approximates the size and acoustical properties of the average home living room. There are heavy curtains on the back wall, fabric wrapped acoustic panels on the sidewalls, and the floor was covered with low pile loop carpet. Other furnishings include metal frame office chairs and equipment shelving. The second room, the "Eargle Theater" which measures 2.7m (H) x 5.3m (W) x 7m (L), is representative of a high-end, purpose-built home theatre. Acoustical treatment includes cloth covered wall cavities with absorptive material, leather theatre seating, heavy curtains, and wall-mounted diffusive panels. The third room, the "Arena Lab" is a cinema sized space that is acoustically treated with acoustic padding on the walls and spray-on acoustic fiberglass insulation on the ceiling. This room is typically used for large audio system testing and has dimensions which measures 8.5m (H) x 14.6m (W) x 14m (L) with a reverberation time of 0.62 seconds. Photos of all three listening rooms are shown in appendix D.

4.5.4 Flattening of In Situ Loudspeaker Responses

For both the BRS and the in situ tests, the loudspeakers were individually measured at the listening position in the three different sized room using the Harman Audio Test System (HATS) and nine dbx RTA-M omnidirectional microphones. First, a correction filter was created for each individual microphone by calibrating them in an anechoic chamber applying the IEC 61094-8 free-field substitution method using a JBL M2 as the source and a Bruel and Kjaer 4136 as the reference microphone (IEC, 2012). The microphones were then placed in an array centred around the listening position on a 47.5 cm (18 in) grid. The spatial average of the nine microphones was used to create a "flattening" filter for the in situ loudspeaker response which was applied to both the in-site and the BRS playback systems. This flattening filter was created using the "AutoEQ" function in HATS and is implemented using a BSS Audio BLU-800 processor for playback. A simplified block diagram of this system is shown in figure 4.8.

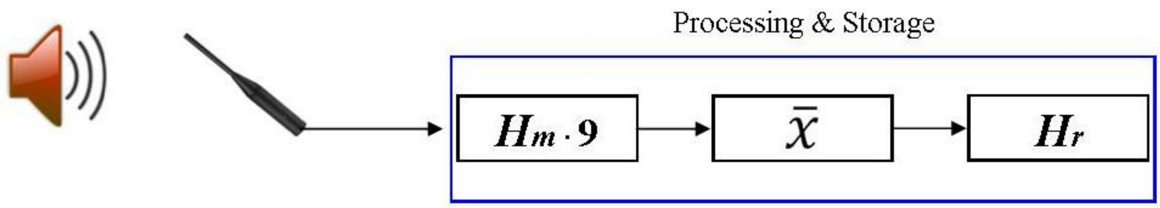


Figure 4.8: Calibration of the in-room loudspeaker response to create the flattening filter ***Hr***. ***Hm*** represents the individual microphone correction filter

AutoEQ allows for a measured curve to be matched to a target curve stored within HATS via a set of automatically assigned filters whose general parameters are set by the user. These parameters include the number of bi-quads assigned, start and stop frequencies, sample rate, level matching, peak to dip ratio, and input curve smoothing, amongst others (Devantier, 2011). An example of the AutoEQ assignments for a single loudspeaker channel is shown in appendix F. Figure 4.9 gives an example of a centre loudspeaker before and after it has been equalized.

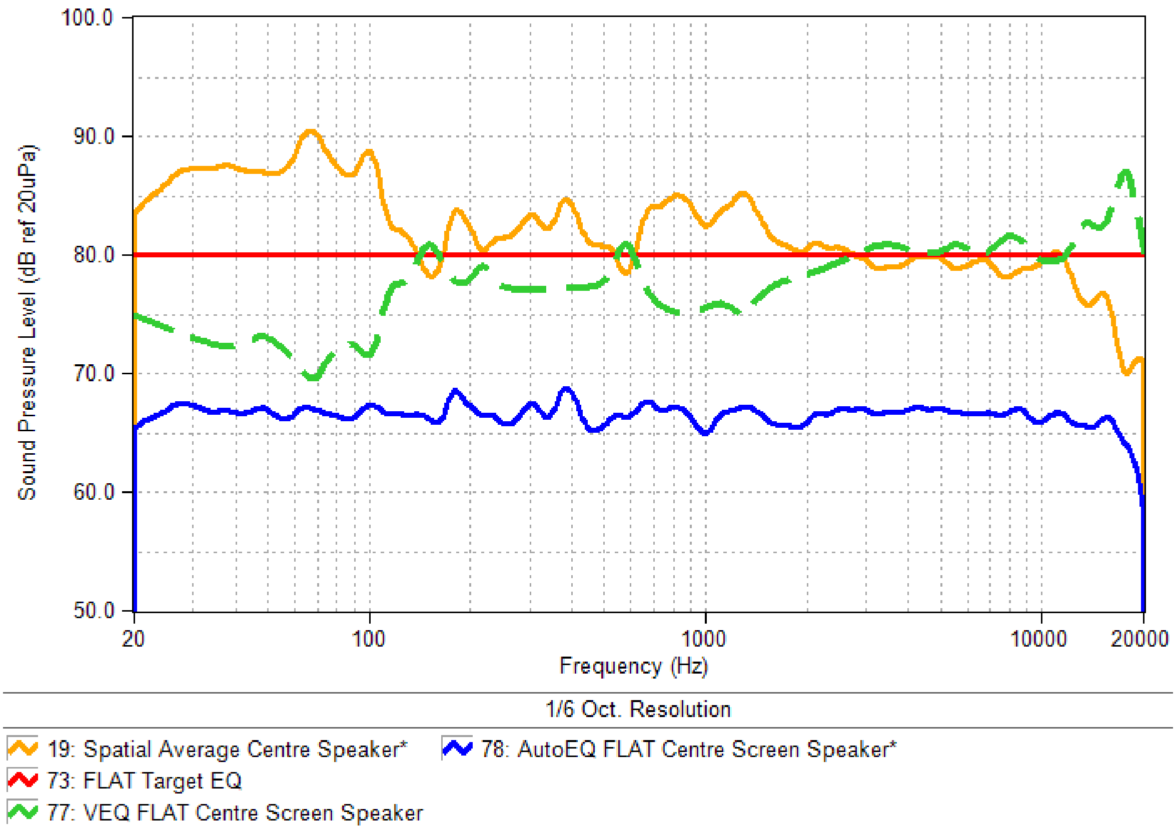


Figure 4.9: Example of HATS AutoEQ. The spatial average (orange) is equalized to match the flat target curve (red). The correction filter, ***Hr,*** is shown in green. The resultant measured in situ loudspeaker response is shown in blue

The final collective "flat" baseline response was used in conjunction with each of the 4 target response curves, thus ensuring that any audible differences among curves were likely related to differences in their frequency responses, and not the response of the manikin microphones, playback headphones, calibration microphones or in situ loudspeaker response.

4.5.5 Creation of In-room Target Response Curves

The validation experiment outlined in this chapter was based on analyzing listeners' preferences between four in-room target response curves (A-D) using the BRS playback method and comparing them to those of the in situ method.

ITU-R BS.1534 recommends that, when using the Multiple Stimuli with Hidden Reference and Anchor (MUSHRA) test methodology, a well-defined anchor should be used as a hidden reference (ITU, 2015). This anchor is used to provide a baseline, or reference, for comparison and to help identify "outliers" within the group of listeners. The use of a hidden reference aids in the assessment of listener reliability and subjective bias (Harris and Holland, 2009). For the anchor to be considered an effective baseline, it is recommended that its characteristics are similar to one of the other signals under testing (Zielinski, 2016).

Curve A simulated the SMPTE X-curve with a high frequency roll off of 4dB between 2kHz and 8kHz and a low frequency roll off starting at 50Hz. Curve B represents the anchor used and was roughly based on the SMPTE X-curve but with an added rise at 1kHz. Curve D was derived from research conducted by Olive et al on target response curves for listening rooms and headphones. This in-room target curve is flat with a 7dB boost starting below 100 Hz, and a gentle treble cut of -1dB starting at 235Hz. The 8dB slope of Curve D is similar to the target curve that was preferred in recent studies on different room correction products and headphones (Olive et al., 2013a), (Olive et al., 2013b), (Olive et al., 2009). Curve E utilized no equalization, thus leaving a flat in-room response. Curve C was omitted from this experiment but was utilized in further experiments and is described in Chapter 5.

The in-room target responses applied to both the in situ and BRS playback systems were designed using the AutoEQ function in HATS, and implemented using the BSS Audio BLU-800, and are shown in figures 4.10 through 4.12.

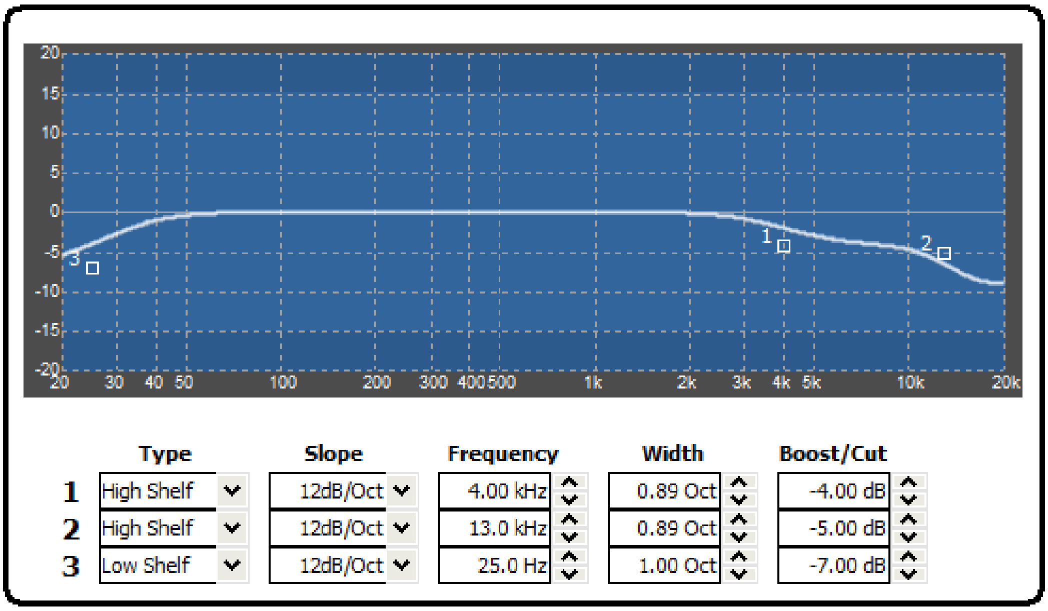


Figure 4.10 Target room response curve 'A'

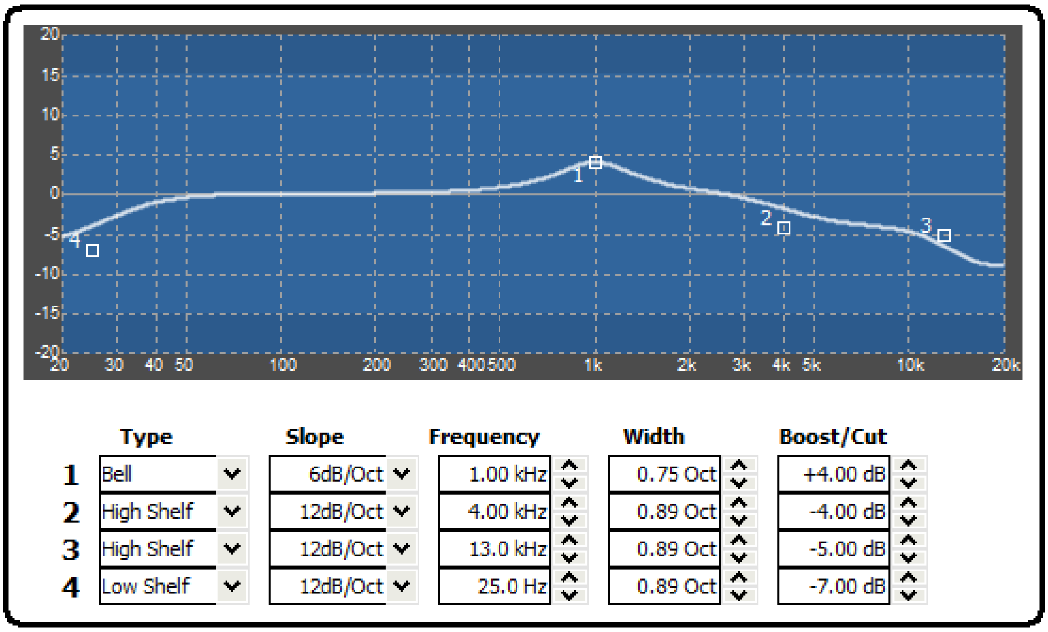


Figure 4.11: Target room response curve 'B'

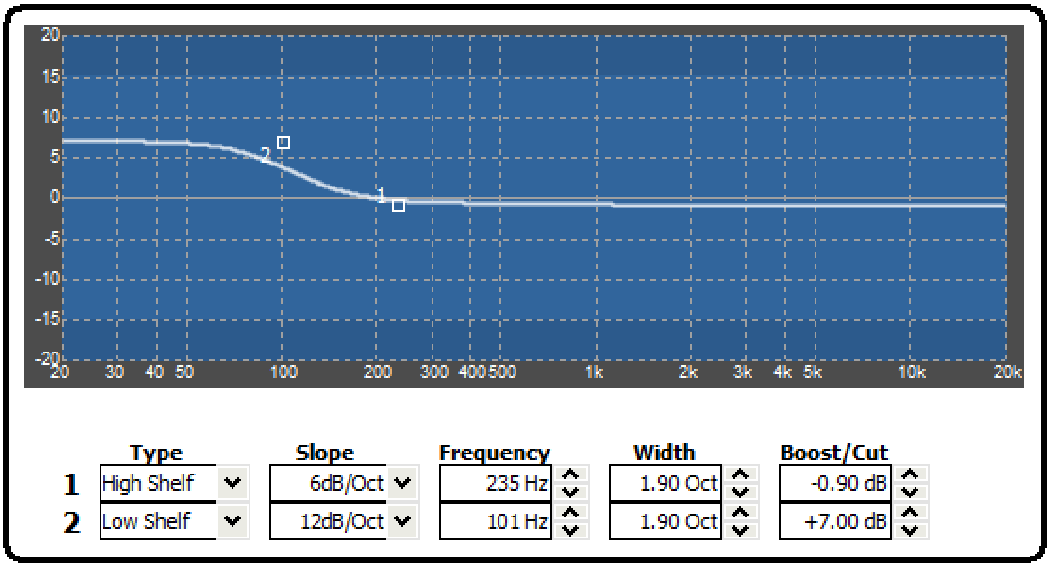


Figure 4.12: Target room response curve 'D'

4.5.6 Program Material

Four different 5.1 music programs were selected including an orchestral film score (***RS***), Steely Dan's "Gaslighting Abbie" (***SD***), Blue Man Group's "Sing Along" (***BM***) and Toy Matinee's "Last Plane Out" (***TM***). The pop music tracks were digitally transferred from a DVD-Audio compact disc, and saved as 6-channel waveform audio file (*.wav). The orchestral score which was obtained as 6 discrete mono wave files directly from the engineer who mixed this track. This track was then imported into ProTools, and converted to a single 6-channel interleaved .wav file. All files were levels matched and edited into short 20-30 second loops using ProTools digital editing software and saved as 6-channel (16 bit, 48 kHz) waveform audio file (*.wav). All four programs have broad and smooth long term spectra indicating they contained signals that were sufficiently spectrally dense and broadband to facilitate reliable judgments of bass and treble balance, in addition to being revealing of spectral features based on previous tests and listener training exercises.

Figure 4.13 shows spectral analysis of the program material. For each program loop long term power spectral density using the Welch method (Welch, 1967) was calculated for each channel, and the results summed to represent overall content.

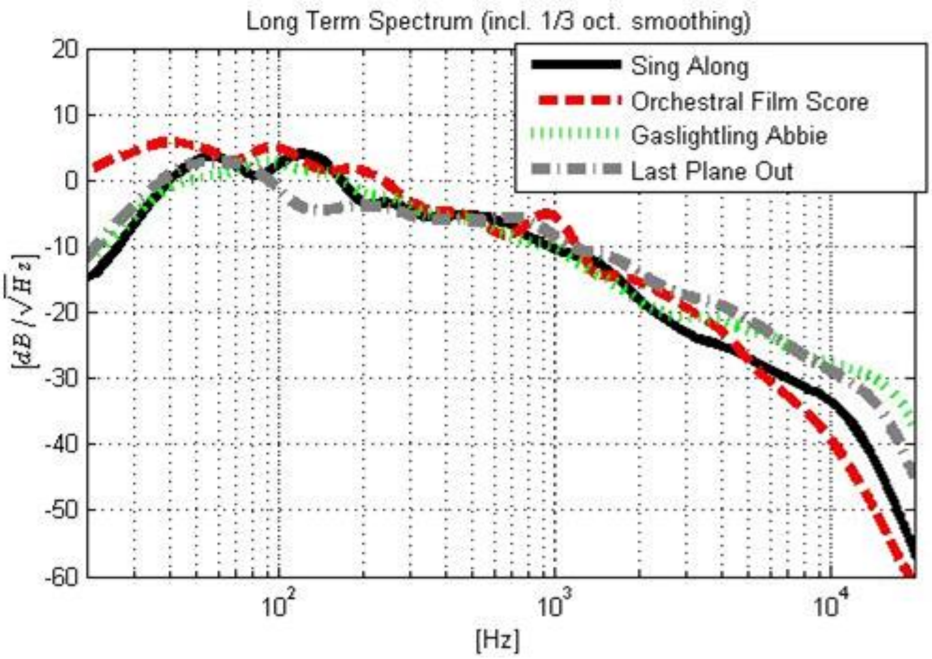


Figure 4.13: Long-term spectral power density of the four pieces of program material, levels normalized at 500Hz

In previous research, Olive found that pink noise elicited the highest percentage of correct responses (88%) in listening tests matching the timbre of 8 different equalization curves to their measured frequency response. A piece of pop music used in the same test scored extremely close to pink noise (87.5%). It was believed that this occurred due to the pop music's extended bass when compared to solo vocal music, jazz music, solo or small instrument ensembles, or spoken voice. This piece of music was also linear, though not flat, throughout the vocal range (100Hz to 1kHz). Both Toole and Olive identified full orchestra and popular music as ideal test signals for spectral based listening tests (Toole, 1982), (Olive, 1994). This further influenced the choice of programs used in this research.

4.5.7 Selection of Listeners

A total of nine listeners: three females and six males were selected for the tests, all of whom were Harman employees paid for their participation. Subjects had their hearing tested using an Earscan® ES3, ANSI S3.6 compliant pure tone audiometer, and were found to have normal hearing per ISO standard 389-1 (Earscan, 2013), (ANSI, 1996b), (ISO, 1998). All of the test subjects were appropriately briefed by use of a standard script in order to avoid inadvertent experimenter bias and were trained using example tests prior to recorded experiments commencing. In addition, the subjects needed to have passed listener training tasks that required them to reliably identify and discriminate among different spectral distortions added to different music programs, achieving a skill level of eight out of ten or higher in all tasks. This training was carried out using Harman's "How to Listen" software (Harman, 2011). The listeners' experience with formal listening tests ranged from 1 to 20 years.

With listening tests there are always questions whether to use trained or untrained listeners. Can trained listeners effectively represent a general section of the population? Are trained listeners really more selective and reliable? Isn't a larger group of untrained listeners better than a small group of trained listeners? The decision to use trained versus untrained listeners was based partially on the requirements of the ITU-R BS.1116 standard and with the recognition that untrained subjects typically do not provide reliable data (ITU, 1997). This coincides with the findings of Gabrielsson and his team, who observed that trained listeners tend to provide more consistent ratings across varying tests and objects under observation (Gabrielsson, 1979, Gabrielsson et al., 1974). Previous research has also found that trained listeners tend to use a smaller portion of the rating scale in addition to a tendency to use the lower part of a preference scale. Trained listeners do not necessarily have different preferences than untrained listeners per se, but they appear to be more critical and difficult to please. Untrained listeners, on the other hand, have a tendency to use much larger portions of the ratings scale which in turn has been shown to lead to a larger error variance (Olive, 2003). Trained listeners also show stability in the adoption and use of the rating scale; the range of ratings remains confined to an area and does not to vary as the listeners "learn" the test. In opposition, untrained listeners will typically use a larger portion of the rating scale at the beginning of the test and then move towards using a smaller portion as they adjust to the test, with this process tending to be repeated with each new test that they take. In his research, Bech also found an "economy of scale" when using trained listeners in the fact that a particular confidence interval from 1 trained listener was equivalent to that of averaging up to 7 untrained listeners (Bech, 1992).

4.5.8 Test Design

The listening test was administered using a MUSHRA-style test where subjects were asked to rate their preference amongst four different in-room target response curves applied to three different cinemas. During the tests, the in-room target curves were comparatively rated on an 11-point preference scale that had semantic differentials on every second interval labeled:

Strong Dislike (1), *Dislike* (3), *Neither Like/Dislike* (5), *Like* (7), and *Strong Like* (9). A strong preference between target curves was indicated by a separation in preference ratings of ≥ 2 points, a moderate preference ≥1 points, and a slight preference ≤ 0.5 rating.

For the two smaller listening rooms, the Living Room and the Eargle Theatre, validity of the BRS system was tested using a 2 x 4 x 4 x 2 x 2 repeated measures ANOVA design where the between-subjects factor was playback method (in situ or BRS playback method). The independent variables were: two test methods (BRS, in situ), four room response curves (A through D), four programs (film score and three popular music pieces), two rooms, and two observations while the dependent variable was preference rating. Due to scheduling constraints and lab availability, not all of the same listeners were available for testing on the Arena Lab and so it was required that this room be validated in a separate experiment. The same ANOVA design as the smaller listening room was used with the exception of the independent variable "room" set to 1 (2 x 4 x 4 x 1 x 2).

The experimental context in which auditory stimuli are presented has been shown to influence preference choices or ratings (Rumsey, 2006). In a similar manner, it has been demonstrated that the trial ordering of music programs has influenced preference choices for different multichannel microphone techniques (Kim et al., 2006), (Martens et al., 2006). For these reasons, all presentations of the stimuli were performed using a double-blind, multiple comparison method (four target curves at a time) that utilized a randomized presentation order for both the target curves and programs. The order in which the BRS and in situ tests were completed was equally distributed among the subjects to minimize possible order biases. In order to directly compare and measure the effects of the two different test methods as accurately as possible, other aspects of the experimental design and listening test conditions were held constant. For both BRS and in situ methods, the stimuli were presented the same number of times, at the same playback level.

Each listener participated in two listening sessions conducted on different days lasting approximately 20-30 minutes each. This structure was based on the recommendations of Bech, that listeners should only participate in one session per day and that each session should last no longer than 30-40 minutes (Bech, 1990). In one session, the listener evaluated the room response curve using the BRS playback system, and in another session, they evaluated the room response curve in situ. The order of the test sessions was randomly assigned to each subject in a balanced way, so that any learning or order effects were equally distributed between the two test methods. In each test session, the listener completed 8 trials (4 programs x 2 observations). In each trial, the listener made comparative judgments between the four in-room target response curves until their final preference ratings were recorded. The listening test software application then automatically loaded the next trial, randomly reassigning a letter (A through D) to each room response curve. This test procedure was administered for each of the three listening rooms in random order. In total, 216 trials were carried out in (8 trials x 3 rooms x 9 test subjects).

A custom software application known simply as Harman Listening Test Software (LTS) was used to administer the experiments. This included control and switching of sound files, control of the BSS audio DSP for the selection of the target curves, as well as the collection and storage of the listener response data to a Structured Query Language (SQL) data base. This software allows the participants to take control of the order of playback conditions and to audition conditions several times at will, which is useful for reducing the impact of the "recency" effect as noted by Zielinski et al. (2008). The software also automatically checks for tied ratings and prompts the listener to make a forced choice between the test objects. The graphical user interface (GUI) utilized for the listening tests is shown in figure 4.14.

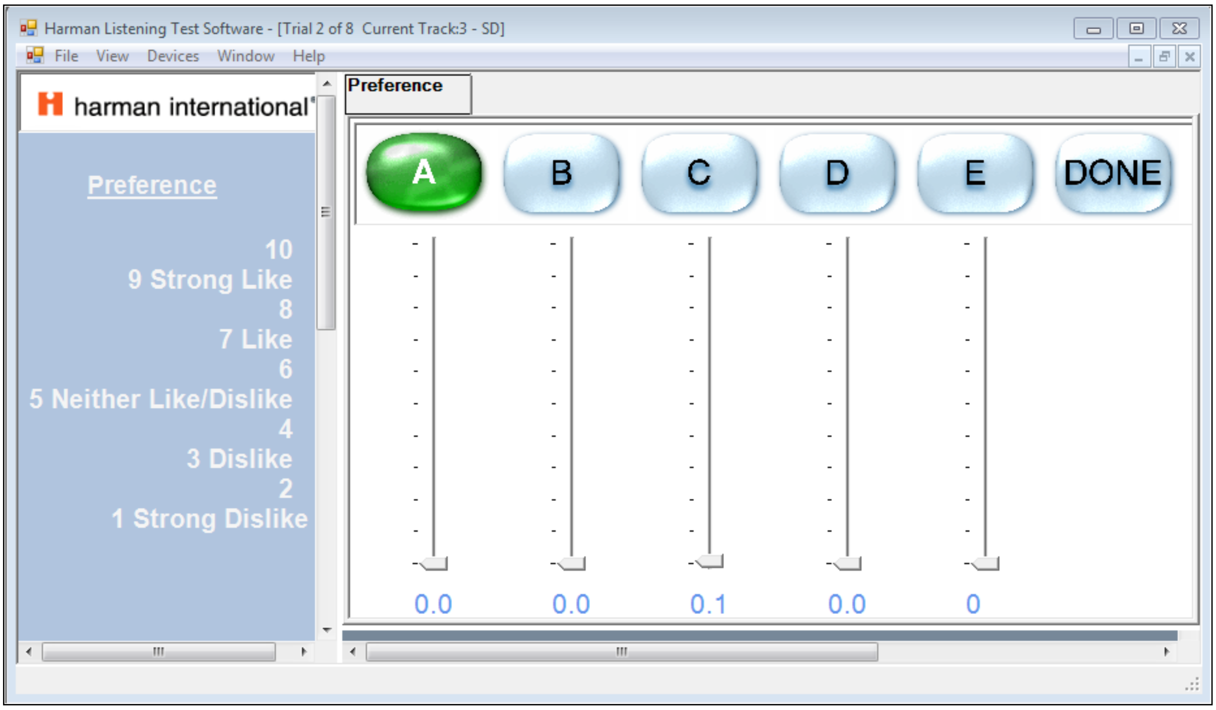


Figure 4.14: GUI for the preference listening test

4.5.9 Test Setup and Hardware

The hardware for the BRS listening test setup consisted of a Windows laptop (Lenovo Thinkpad T420s), a digital sound card (RME Fireface UC) a programmable digital signal processor (BSS Audio BLU-800 and 80) and ultrasonic head-tracker (Logitech 3DH-311). The digital signal processor is used to store the room flattening filter (***Hr***) and the four in-room target response curves.

A key piece of the BRS playback system is the head-tracker which constantly monitors the position of the listener's head, and sends the angular coordinates to the playback engine, which in turn switches to the corresponding set of measured BRIRs. The ultrasonic Logitech head-tracker utilized has an orientation resolution of 0.1°, an update rate of 50Hz and maximum latency of 30ms. The latency of the entire BRS playback system was found to be approximately 50ms which, inclusive of the update rate of the head-tracker, does not cause significant audible effects under normal listening conditions (Stitt et al., 2016), (Lindau, 2009), (Yairi et al., 2008), (Brungart et al., 2006), (Wenzel, 1999), (Bronkhorst, 1995), (Sandvad, 1996).

The BRS portions of the experiments were conducted with the listeners sitting in a loudspeaker lab at Harman, where the background noise level was measured to be between 30 and 35dBA.

The hardware used for the Living Room in situ listening test consisted of the same laptop, digital sound card and programmable digital signal processor in addition to a Harman Kardon 7550HD AV receiver / amplifier, and Harman Kardon HKTS 60 5.1 loudspeaker system. The Eargle Theater in situ listening test equipment was similar to that of the Living Room with the difference of a three JBL Project Everest DD66000 left, center and right loudspeakers, six JBL S4Ai THX Ultra2 surround loudspeakers and four JBL S1S-EX 18-inch THX Ultra2 subwoofers. The electronics in the system include a Mark Levinson No. 502 media console/surround processor, seven bridged JBL Synthesis S820 amplifiers (for the L, C, R speakers and subwoofers), and a JBL Synthesis 7 x 160-watt seven-channel amplifier (for the surrounds). In addition to the front-end equipment used in the other two rooms, the Arena Lab setup utilized a 5.1 loudspeaker system consisting of five JBL M2 full-range loudspeakers for the left, centre, right and surround channel and a JBL 4645C subwoofer all powered by two Crown I-Tech 4x3500HD amplifiers.

For the in situ portion of the experiments, listeners were positioned centrally in a standard 5.1 setup based on ITU-R BS.1116-1 listening room recommendations (ITU, 1997). This position matched the position of the BRS manikin during the scanning process. The background noise levels were measured with a Brüel & Kjær Type 2239A sound level meter and were found to be 34dBA in the Living Room, 25dBA in the Eargle Theater, and 42dBA in the Arena Lab.

The relative playback levels of the different target response curves and program material were adjusted for equal loudness objectively using an ITU-compliant loudness meter (B & K Type 2239A) and based on the ITU-R1770.3 loudness standard (ITU, 2012b). Levels were further verified by informal listening to assure that there were no discernible loudness differences between test curves. The average playback level of the music was adjusted to a comfortable level of 80dBA for both in situ and BRS listening tests.

4.6 Results

The following sections report on the statistical analysis utilized and the results of the listening tests.

4.6.1 Statistical Analysis

The data was analyzed using SAS Stat View software running a repeated measures analysis of variance (ANOVA). Additional analyses were completed using Matlab and custom designed Excel spreadsheets. Since the same participants experienced all conditions that were assessed in the Living Room and Eargle Theater tests, a one-way repeated measures ANOVAs was deemed appropriate in order to identify whether there were any significant differences of the mean ratings for preference between BRS and in situ playback conditions. A one-way repeated measures ANOVA was also utilized for the Arena Lab test; which again employed a different set of listeners and therefore was treated as a separate test for the purpose of analysis. A complete factorial analysis was used in the ANOVA model with a significance level of 0.05 utilized for all statistical tests. Per the MUSHRA test standard, data should be normalized if no anchor is utilized (ITU, 2001). However, since this experiment employed an anchor point, it was not necessary to normalize the data prior to statistical analysis.

To effectively apply a repeated measured ANOVA the data must meet the following assumptions:

1. Independence: Observations are independent. In the case of repeated measures, this assumption means that the subjects should be mutually exclusive. The measurements made within each subject are, of course, not independent from each other in repeated measures designs.
2. Normality: Data should be sampled from a population that is normally distributed.
3. Homoscedasticity (homogeneity of variance): The population variances for the various levels within a factor should be similar.

Sphericity, often considered a type of homogeneity of variance, refers to the equality of the population variances calculated for the differences between all pairs of within-subject conditions.

In order to check the data for conformation to the assumption of homoscedasticity (the variances are not significantly different), a Mauchly's test of sphericity with a significance level of 0.05 was employed using Matlab (mausphercnst(x,alpha)); the code is shown in appendix G (Mauchly, 1940), (Trujillo-Ortiz et al., 2008). It should be noted that omnibus sphericity cannot be tested if the number of participants (listeners) in the study is not greater than or equal to the product of the levels of all factors; which in the case of these experiments, it is not. In addition, omnibus sphericity is rarely achieved. For these reasons, multiple comparisons within common groups in order to observe "local sphericity" were undertaken using the above cited Matlab code. For any multiple comparisons where local sphericity has been attained, it is considered that the overall sphericity assumption has been met (Kirk, 2013), (Field, 2013), (Rouanet and Lepine, 1970). Of interest to note is that statistical program such as SPSS only provide local tests of sphericity as they do not provide a true omnibus test. Since groupings of the data from both listening tests were found to be non-significant ($p>0.05$), and are therefore considered locally spherical, they could be used for the ANOVA analyses.

Once the ANOVA analyses were completed, a post hoc analysis was undertaken in an effort to explore the difference among means in order to provide specific information on which means were significantly different from each other. The post hoc test utilized was the Scheffe test, which was chosen due to its conservative nature and flexibility in testing all possible contrasts amongst the various test factor means and not just the pairwise differences (Scheffe, 1999). This flexibility is beneficial in uncovering which factors, or combinations of factors, produced the significant differences found in the original ANOVA test, especially those differences that were not originally planned for during the design of the experiment. However, since Scheffe's test is prone to type I errors where the Bonferroni/Dunn test is not, an exploratory comparison between the two tests was completed to validate the results (McHugh, 2011) (Jaccard et al., 1984).

Several different types of effect size analyses were also used in an effort to understand the magnitude of the differences between and within factors. The explanation of the different effect size tests and their use are broken into categories and explained briefly (Kirk, 1994), (McGraw and Wong, 1992), (Cohen, 1988).

From the correlation family of effect size tests:

- η^2 - Reflects the proportion of the total differences in the scores that is associated with differences between the conditions. In addition, it measures the variance of the sample but not the population so it will tend to be biased and overestimates the effect size with smaller sample sizes.
  The formula for η^2 is given as:

$$\eta^2 = \frac{SS_{ef}}{SS_t} \qquad \text{Eq. 4.1}$$

  Where SS_t is the total sum of squares and SS_{ef} is the sum of squares for the effect.

From the difference family of effect size test:

- d (Cohen's d) - Cohen's d is a standardized difference between two means and therefore the calculated d can be greater than 1.00 or less than -1.00, as it is an estimate of the degree of overlap between the two group's distributions.
  The formula for d is given as:

$$d = \frac{\bar{x}_1 - \bar{x}_2}{S_p} \qquad \text{Eq. 4.2}$$

Where $\bar{x}_1$ and $\bar{x}_2$ are the means of the two factors being compared and S_p is the "pooled" or average standard deviation.

The pool standard deviation is typically calculated by:

$$S_p = \frac{\sqrt{(\sigma_1)^2 + (\sigma_2)^2}}{2} \qquad \text{Eq. 4.3}$$

Cohen's ***d*** is a straight forward method for analyzing the effect size between two factors as the value of ***d*** represents the difference, in standard deviations, between the two factors. Cohen's ***d*** can be negative or positive depending on the relationship between the two means (factors) involved. If ***d*** is positive, then the mean of the first factor is higher than the second; if ***d*** is negative then this relationship is reversed.

Common Language effect size:

- ***CL*** - Considered one of the easiest of the effect tests to interpret, ***CL*** is the probability that a randomly selected score from the one factor or group will be greater than a randomly sampled score from the other factor or group. ***CL*** is given in a range from 0.50 to 0.98, which is then typically converted into a percentage.

$$CL = \frac{0-(\bar{x}_1 - \bar{x}_2)}{\sqrt{(s_1)^2 + (s_2)^2}} \qquad \text{Eq. 4.4}$$

The interpretation of these effect sizes and their relationships to each other are shown in table 4.1.

d	***η*²**	***CL (convert to %)***	**Description**
< 0	-	-	Adverse Effect or No Effect
0.0	.000	50%	No Effect
0.1	.003	53%	
0.2	.010	56%	Small Effect
0.3	.022	58%	
0.4	.039	61%	
0.5	.060	64%	Intermediate Effect
0.6	.083	66%	
0.7	.110	69%	
0.8	.140	71%	Large Effect
0.9	.168	74%	
≥ 1.0	≥.200	≥76%	

Table 4.1: Interpretation of effect sizes, sensu (Kirk, 1994), (McGraw and Wong, 1992), (Cohen, 1988)

As each of the tests use different parameters to determine effect size, which in turn leads to varying strengths and weaknesses, the author felt it prudent to run all the tests described to verify effect sizes.

4.6.2 Results from Eargle and Living Room Test

As the intention of these tests was to prove the validity of using BRS in place of in situ listening test, to determine if there was a statistical difference between ratings for the in situ and BRS playback conditions, data from the listening tests were checked for Gaussian distribution and subjected to a repeated measures ANOVA test. The following were observed:

- ***Target Curve*** was significant: F (3,24) = 96.781, p < .0001
- ***Observation*** was not significant: F (1,8) = 0.091, p = .7711
- ***Program*** was not significant: F (3,24) = 0.905, p = .4532
- ***Room*** was not significant: F (3,24) = 2.488, p = .1534
- ***Method*** was not significant: F (1,8) = 3.34, p = .0916
- ***Room*Target Curve*** was significant: F (3,24) = 5.682, p = .0044

The full ANOVA table for this experiment can be found in appendix H.

To verify that homoscedasticity between the two playback methods was met, box plots of the conditions were created, as shown in figure 4.15.

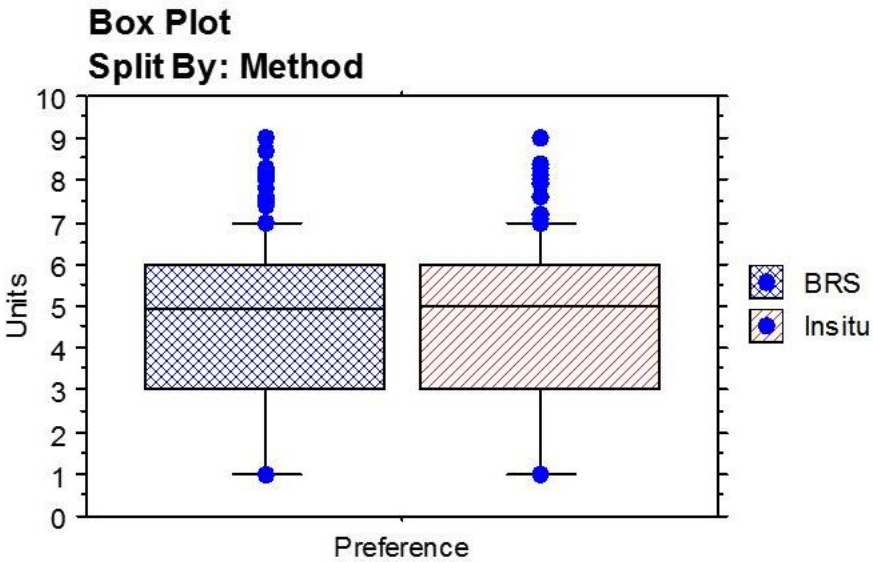


Figure 4.15: Homoscedasticity between *playback methods*

From figure 4:13 one can observe that the distribution of scores and the means between the two playback conditions were approximately equal.

Table 4.2 shows the extent of the differences between the two playback methods' score distributions.

	Preference	BRS	In situ
Mean	4.439	4.386	4.492
Std. Dev.	2.035	2.022	2.048
Std. Error	0.06	0.084	0.085
Count	1152	576	576
Minimum	1	1	1
Maximum	9	9	9
Skewness	-0.08	-0.099	-0.064
Kurtosis	-0.772	-0.802	-0.75

Table 4.2: Descriptive statistics split by *method*

The skewness for each of the methods is well between 1.0 and -1.0, showing a fairly symmetrical distribution of the data. In addition, the kurtosis for each method shows that the data exhibits a raised cosine to semicircle distribution of -1.0 < $\boldsymbol{k}$ < -05 (DeCarlo, 1997).

Figure 4.16 shows the combined target curve preference data for the in situ and BRS playback conditions for all conditions in both the Eargle and the Living Room with 95% confidence intervals shown.

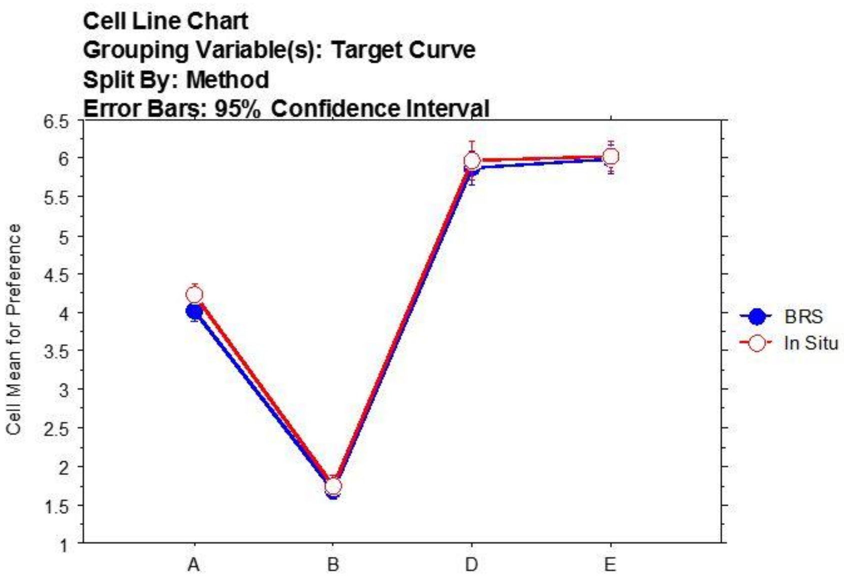


Figure 4.16: Preference data per *target curve* and *playback method*

The graph in figure 4.16 depicts that the preference scores for the BRS and in situ playbacks were nearly identical. It had been noted in a cursory analysis of the data that the preference scores for curves D and E appeared to be very close, and so a split in target curve scores between *rooms* and *methods* was warranted. In addition, the ANOVA results showed that not only were the target curves found to be significant but also the interaction between the rooms and the target curves. For this reason, the data was re-plotted, separating the Living Room from the Eargle Theatre, as shown in figure 4.17.

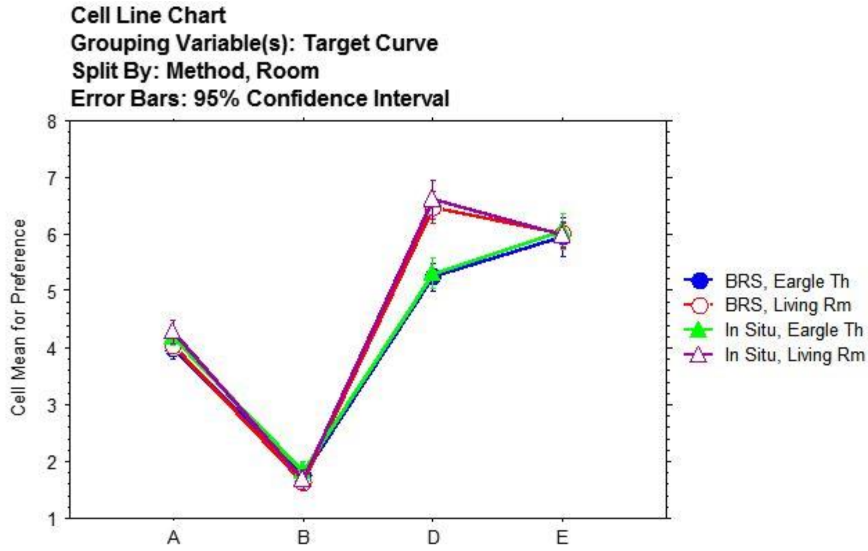


Figure 4.17: Combined data for the two playback conditions for all programs and both rooms

Figure 4.17 clearly shows that the preference in target curves was partially determined by the playback room, but not the playback condition; once again the BRS and in situ scores are nearly identical.

The reliability of listeners is difficult to determine directly from the ANOVA test results and ideally, a subject should repeat approximately the same score for repeated ratings of the same stimuli. Figure 4.18 shows averaged ratings for observations Ob1 and Ob2 for each listener.

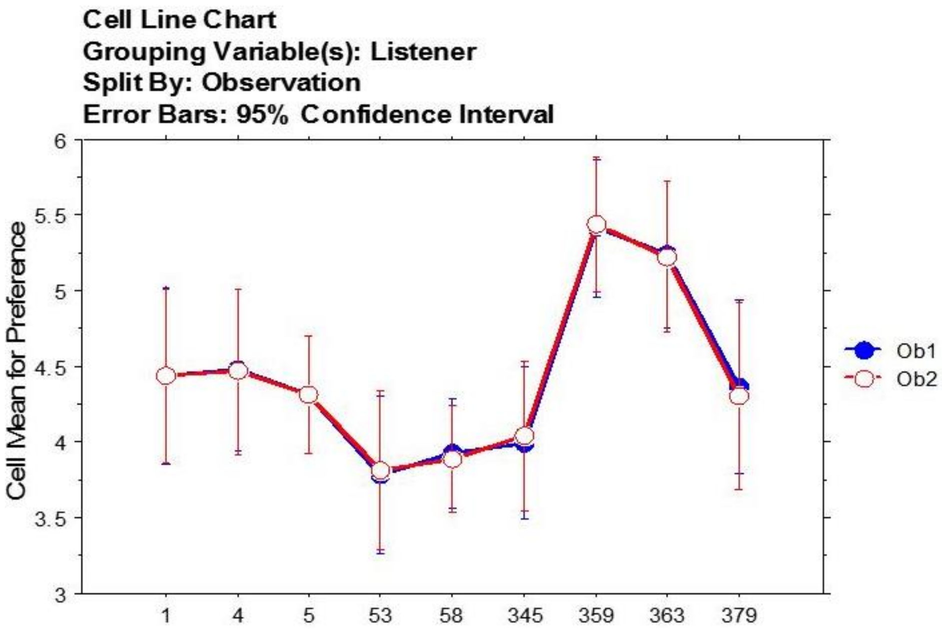


Figure 4.18: Repeatability of test subjects for repeated observations of the same stimuli.

Though there were the expected variations between listeners scores, there was little to no variation in the individuals' scores between the two observations.

Post hoc analyses were completed for *method* and *target curves* with the results shown in tables 4.3 and 4.4.

Scheffe for Preference
Effect: Method
Significance Level: 5 %

	Mean Diff.	Crit. Diff.	P-Value
BRS, Insitu	-.106	.128	.0916

Bonferroni/Dunn for Preference
Effect: Method
Significance Level: 5 %

	Mean Diff.	Crit. Diff.	P-Value
BRS, Insitu	-.106	.128	.0916

Tables 4.3a & 4.3b: Scheffe and Bonferroni / Dunn post hoc analyses for *method*

Scheffe for Preference
Effect: Target Curve
Significance Level: 5 %

	Mean Diff.	Crit. Diff.	P-Value	
A, B	2.398	.866	<.0001	S
A, D	-1.788	.866	<.0001	S
A, E	-1.878	.866	<.0001	S
B, D	-4.186	.866	<.0001	S
B, E	-4.276	.866	<.0001	S
D, E	-.090	.866	.9920	

Bonferroni/Dunn for Preference
Effect: Target Curve
Significance Level: 5 %

	Mean Diff.	Crit. Diff.	P-Value	
A, B	2.398	.829	<.0001	S
A, D	-1.788	.829	<.0001	S
A, E	-1.878	.829	<.0001	S
B, D	-4.186	.829	<.0001	S
B, E	-4.276	.829	<.0001	S
D, E	-.090	.829	.7588	

Tables 4.4a & 4.4b: Scheffe and Bonferroni / Dunn post hoc analyses for *target curve*

Tables 4.3a and 4.3b show agreement between the Scheffe and Bonferroni / Dunn post hoc analyses and that two playback methods were not found to be significantly different. Tables 4.4a and 4.4b shows that the Scheffe and Bonferroni test have slight differences in the ***p***-value between target curves D and E. However, both of these tests demonstrate that there is not a significant difference between curves D and E.

Several different types of effects sizes analyses were completed for the various factors with the results shown in table 4.5.

	d	η^2	CL
Observation	0.00	0.000	50.04%
Method	0.03	0.001	50.73%
Room	0.15	0.005	54.13%
Program	-	0.000	-
BM/RS	-0.01	-	50.18%
BM/SD	0.03	-	50.97%
BM/TM	0.05	-	51.33%
RS/SD	0.04	-	51.11%
RS/TM	0.05	-	51.47%
SD/TM	0.01	-	50.35%
Curve	-	0.730	-
A/B	-3.16	-	98.74%
A/D	1.57	-	85.56%
A/E	1.89	-	90.96%
B/D	-3.73	-	99.58%
B/E	-4.39	-	99.91%
D/E	0.07	-	51.96%
$D/E_{LivingRm}$	-0.48	-	63.20%
$D/E_{EargleTh}$	0.57	-	65.63%

Table 4.5: Effect size scores for all comparisons with data from both Living Room and Eargle Theatre; effect sizes between *target curves* highlighted in blue for small difference and pink for intermediate differences

The most important effect size to recognize is the one involving the two methods; BRS and in situ. Once again, it can be seen that there is no statistical difference between the two as the effect size is very small (no effect) and the ***CL*** score is nearly 50%. The effect size between the rooms, when viewed by themselves, is on the cusp of being considered a 'small effect', and confirms the findings of the ANOVA test. However, when considered against the target curve data, the effect of the different rooms comes into play. As expected, the effect size between the two observations was negligible with a ***d*** score at the 'no effect" level and a ***CL*** of almost exactly 50%. The program data showed no effect on the preference scores and the pair-wise comparisons between the individual pieces of program material found that no one particular piece scored higher than another. The most significant factor in the ANOVA was *target curve*, as its effect size confirmed, and most of the target curves showed significant differences between them. However, when compared across the entire data set, target curves D and E were found to be very similar, as is shown in the low effect scores between them and a ***CL*** of close to 50%. This may account for the effect size of the target curves being rated at 'intermediate' as opposed to 'large'. To understand the degree by which these two curves are related the data was split by *room*, which was indicated by the ANOVA score of ***p***=.0044 for *room*target curve*. When the data for these two curves was separated by *room* and the effects tests were run again, there was an improvement in the effect size score with the pair showing a lower 'intermediate' score in both rooms. The residual similarities in effect size between D and E are due to variances in individual listeners' preference, which is shown in figure 4.19.

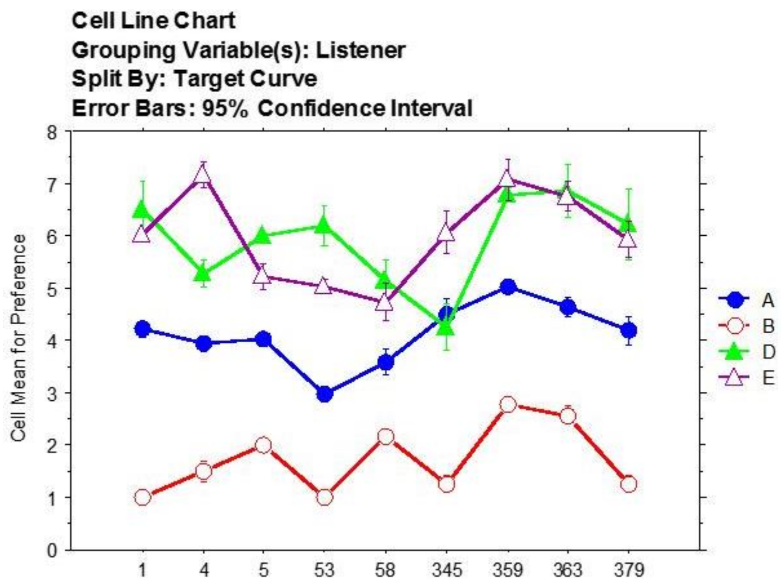


Figure 4.19: Target curve preference by *listener*

Once a factorial post hoc test was conducted excluding the factor *listener*, these two curves became statistically different as shown in tables 4.6a and 4.6b.

Scheffe for Preference
Effect: Target Curve
Significance Level: 5 %
Split By: Room
Cell: Eargle Th

	Mean Diff.	Crit. Diff.	P-Value	
A, B	2.303	.356	<.0001	S
A, D	-1.185	.356	<.0001	S
A, E	-1.911	.356	<.0001	S
B, D	-3.487	.356	<.0001	S
B, E	-4.214	.356	<.0001	S
D, E	-.726	.356	<.0001	S

Scheffe for Preference
Effect: Target Curve
Significance Level: 5 %
Split By: Room
Cell: Living Rm

	Mean Diff.	Crit. Diff.	P-Value	
A, B	2.494	.337	<.0001	S
A, D	-2.392	.337	<.0001	S
A, E	-1.844	.337	<.0001	S
B, D	-4.885	.337	<.0001	S
B, E	-4.338	.337	<.0001	S
D, E	.547	.337	.0001	S

Tables 4.6a & 4.6b: Scheffe post hoc factorial analyses for *target curve* split by *room*; a) Eargle Theatre, b) Living Room

The Bonferroni / Dunn test results are not shown as they were in agreement with the Scheffe test results.

4.6.3 Results Arena Lab Test

Like the Living Room and Eargle Theatre, the Arena Lab data was analyzed in the same way to determine if there was a statistical difference between ratings for the in situ and BRS playback conditions. From the repeated measures ANOVA the following were observed:

- ***Target Curve*** was significant: $\boldsymbol{F}\,(3,24) = 53.854$, $\boldsymbol{p} < .0001$
- ***Observation*** was not significant: $\boldsymbol{F}\,(1,8) = 0.013$, $\boldsymbol{p} = .9112$
- ***Program*** was not significant: $\boldsymbol{F}\,(3,24) = 1.110$, $\boldsymbol{p} = .3646$
- ***Method*** was not significant: $\boldsymbol{F}\,(1,8) = 0.008$, $\boldsymbol{p} = .9298$

The full ANOVA table for this experiment can be found in appendix I.

To verify that homoscedasticity between the two playback methods was met, box plots of the conditions were created, as shown in figure 4.20.

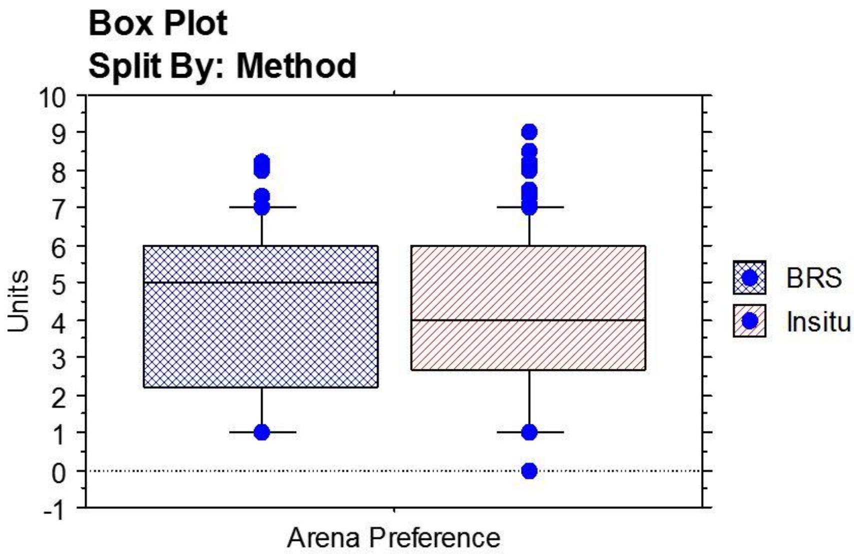


Figure 4.20: Homoscedasticity between *playback methods*

Figure 4:20 shows that the distribution of scores between the two playback conditions is not quite equal and are skewed in different directions and to different degrees.

Table 4.7 shows the extent of the differences between the two playback methods' score distributions.

	Preference	BRS	In situ
Mean	4.345	4.351	4.340
Std. Dev.	2.207	2.177	2.240
Std. Error	0.092	0.128	0.132
Count	576	288	288
Minimum	0	1	0
Maximum	9	8.2	9
Skewness	-0.02	-0.11	0.063
Kurtosis	-1.062	-1.198	-0.942

Table 4.7: Descriptive statistics for both *playback methods*

Though the two methods lack perfect homogeneity, the means, standard deviations and standard errors were similar in values and therefore the data is considered statistically comparable. The data between the methods is slightly skewed and in different directions but this finding is marginal. The kurtosis shows a semicircle to uniform distribution of scores with $-1.2 < \boldsymbol{k} < -1.0$.

Figure 4.21 shows the combined target curve preference data for the in situ and BRS playback conditions for all conditions with 95% confidence intervals shown.

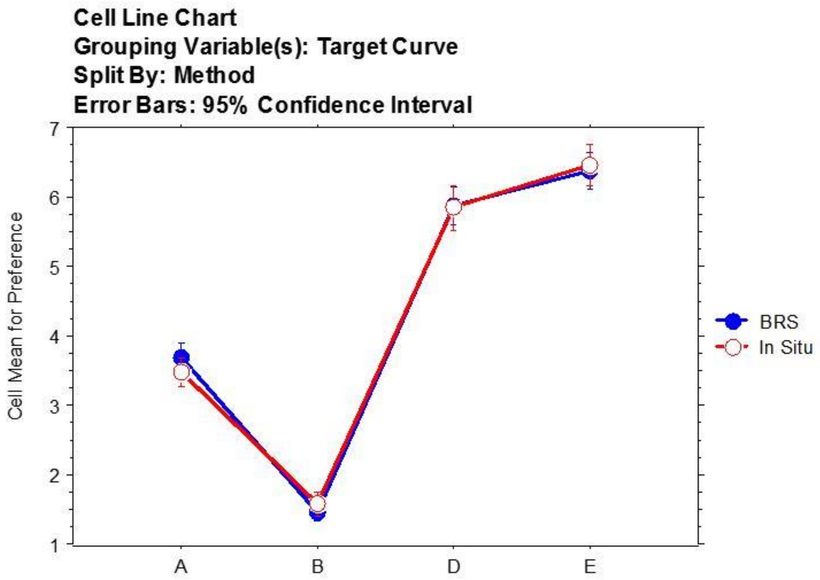


Figure 4.21: Preference data per *target curve* and *playback method*

Once again, in an effort to determine listener repeatability, the averaged ratings for observations Ob1 and Ob2 for each listener was graphed and is shown in figure 4.22.

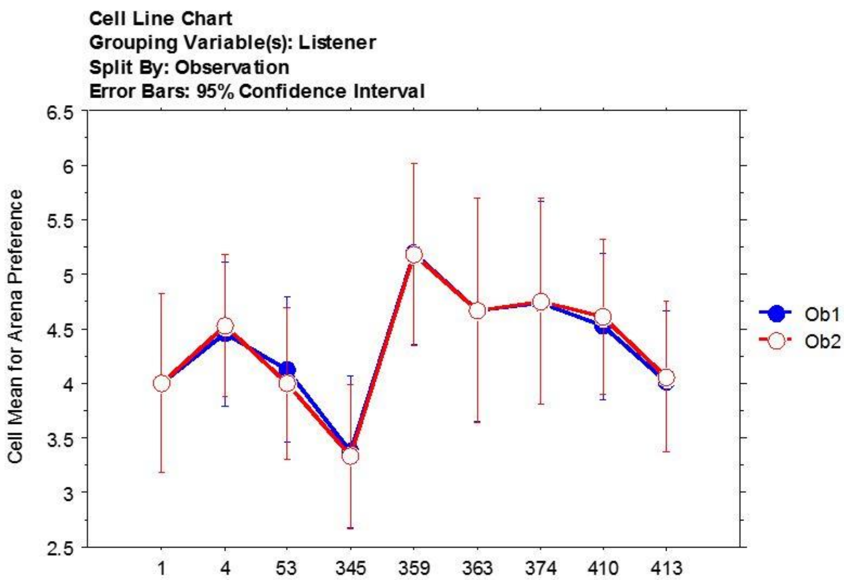


Figure 4.22: Repeatability of test subjects for repeated observations of the same stimuli.

Post hoc analyses were completed for *playback method* and *target curves* with the results shown in tables 4.8(a-b) and 4.9(a-b).

Scheffe for Arena Preference
Effect: Method
Significance Level: 5 %

	Mean Diff.	Crit. Diff.	P-Value
BRS, Insitu	.011	.362	.9504

Bonferroni/Dunn for Arena Preference
Effect: Method
Significance Level: 5 %

	Mean Diff.	Crit. Diff.	P-Value
BRS, Insitu	.011	.362	.9504

Tables 4.8a & 4.8b: Scheffe and Bonferroni / Dunn post hoc analyses for *playback method*

Scheffe for Preference
Effect: Target Curve
Significance Level: 5 %

	Mean Diff.	Crit. Diff.	P-Value	
A, B	2.065	1.302	.0010	S
A, D	-2.284	1.302	.0003	S
A, E	-2.837	1.302	<.0001	S
B, D	-4.349	1.302	<.0001	S
B, E	-4.901	1.302	<.0001	S
D, E	-.553	1.302	.6578	

Bonferroni/Dunn for Preference
Effect: Target Curve
Significance Level: 5 %

	Mean Diff.	Crit. Diff.	P-Value	
A, B	2.065	1.246	<.0001	S
A, D	-2.284	1.246	<.0001	S
A, E	-2.837	1.246	<.0001	S
B, D	-4.349	1.246	<.0001	S
B, E	-4.901	1.246	<.0001	S
D, E	-.553	1.246	.2142	

Tables 4.9a & 4.9b: Scheffe and Bonferroni / Dunn post hoc analyses for *target curve*

Tables 4.8a and 4.8b demonstrate agreement between the Scheffe and Bonferroni / Dunn post hoc analyses and that two playback methods were not found to be significantly different. Tables 4.9a and 4.9b display large difference in the p-values between the Scheffe and Bonferroni tests for target curves D and E, however, both of these tests still show that there is not a significant difference between these two curves.

The different types of effects sizes analyses were completed for the various factors with the results shown in table 4.10.

	d	η^2	*CL*
Observation	0.00	0.000	50.04%
Method	0.00	0.000	50.14%
Program	-	0.000	-
BM/RS	-0.04	-	51.05%
BM/SD	-0.01	-	50.32%
BM/TM	0.02	-	50.48%
RS/SD	0.03	-	50.74%
RS/TM	0.05	-	51.53%
SD/TM	0.03	-	50.81%
Curve	-	0.781	-
A/B	-2.64	-	92.93%
A/D	2.08	-	92.93%
A/E	2.70	-	97.18%
B/D	-4.27	-	99.87%
B/E	-5.06	-	99.98%
D/E	0.45	-	62.37%

Table 4.10: Effect size scores for all general comparisons; effect sizes between *target curves* highlighted in pink for intermediate differences

Again, no statistical difference was found between *methods*, *observations* or *program*. The pair-wise comparisons between the individual pieces of program material found that no one particular piece scored higher than another. The most significant factor in the ANOVA was *target curve*, as is reflected in their effect size scores. Almost all of the target curves showed significant differences between them with the exception of curves D and E. Table 4.9 shows that these two differ by only a small to intermediate level. As with test results for the Living Room and Eargle Theatre, the remaining similarities in effect size between D and E can be attributed to variances in individual listener preference as shown in figure 4.23.

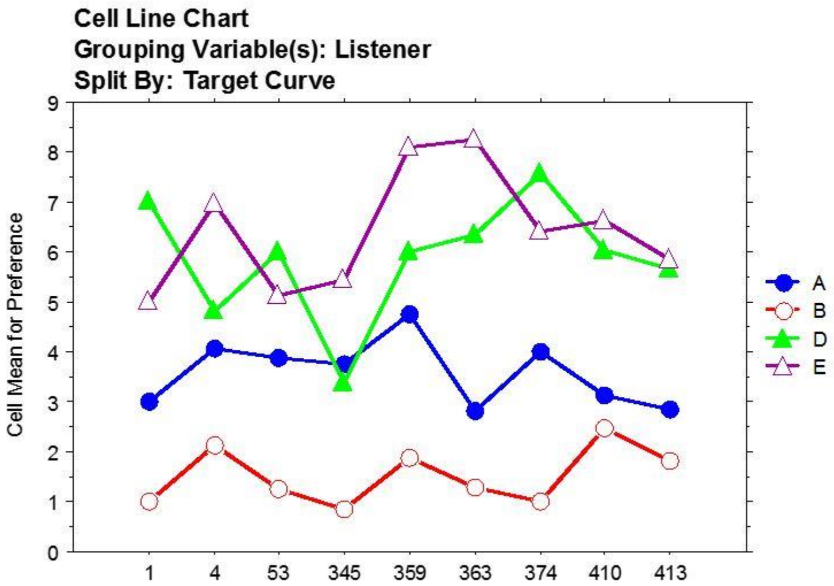


Figure 4.23: Target curve preference by *listener*

Once a factorial post hoc test was conducted that excluded the factor *listener*, these two curves became statistically different as shown in tables 4.11a and 4.11b.

Scheffe for Preference
Effect: Target Curve
Significance Level: 5 %

	Mean Diff.	Crit. Diff.	P-Value	
A, B	2.065	.358	<.0001	S
A, D	-2.284	.358	<.0001	S
A, E	-2.837	.358	<.0001	S
B, D	-4.349	.358	<.0001	S
B, E	-4.901	.358	<.0001	S
D, E	-.553	.358	.0004	S

Bonferroni/Dunn for Preference
Effect: Target Curve
Significance Level: 5 %

	Mean Diff.	Crit. Diff.	P-Value	
A, B	2.065	.338	<.0001	S
A, D	-2.284	.338	<.0001	S
A, E	-2.837	.338	<.0001	S
B, D	-4.349	.338	<.0001	S
B, E	-4.901	.338	<.0001	S
D, E	-.553	.338	<.0001	S

Tables 4.11a & 4.11b: Scheffe and Bonferroni / Dunn post hoc factorial analyses for *target curve* with factor *listener* removed

4.7 Discussion

The statistical observations in the previous section are somewhat expected, but do help to confirm the validity of the BRS test method. The room response curve should have the largest influence on ratings, or the test would have been ill-designed and / or poorly executed. Since the orders of the trials were randomized at all levels, a statistical difference across repeated observations would not be expected, unless there was a learning effect. Program should not be a factor, since subjects were asked to judge preference of the playback quality and not the program. However, program can be factor if it is not carefully selected, which was not the case here. Furthermore, in these types of tests, it sometimes occurs that there is an interaction between room response curve and program, but again, this was not the case here.

In the first test involving the Living Room and Eargle Theatre, the factor *room* was not found to be statistically significant in and of itself. The *room*target curve* interaction was significant, however. This is interesting but not totally unexpected. This simply means that subjects had a different room response curve preference depending on which room they were listening in, which is depicted in figure 4.15. It can be seen that there was a marked difference in overall preference for the C target curve between the two rooms, for both playback methods. Since the sound systems in each room were equalized to the same measured target curves at the listeners seat, the difference in preference ratings is likely attributable to differences in directivity in the loudspeakers and/or acoustical properties of the rooms. The Eargle Theater had JBL Project Everest speakers, which had waveguides for mid/high frequency drivers and thus controlled directivity over a larger bandwidth, and the room itself is well damped acoustically. The Living Room had Harman Kardon HKTS consumer speakers with small direct radiators and no directivity control. This room was smaller and with less controlled acoustics (curtains, carpet). It would not be surprising if these differences caused an interaction with the preferred target curve. It is possible that differences in perception of direct and reflected sound had an effect on perceived spectral balance that was not reflected in the steady state measurements.

For all three rooms there appeared to be similarities in preferences for target curves D and E, as they were initially found to be statistically equal. However, when analyzed with individual listener preference in mind, it became apparent that some listeners had not only differed in their preference between the rooms but between each other as well. Once again, this is not completely unexpected but the degree of this effect size was a bit surprising. Though individual listener preference did effect the overall target curve scores, it should be emphasised that determining a preference for target score was not the objective of this experiment. It is important to note that regardless of listener target curve preference, the listeners' preferences did not change with observation or playback methodology. The most important finding of this validation experiment was that there was no statistical effect of *method* on preference in any of the three rooms under testing.

The question of whether individualized HRTF calibrations are required for BRS based evaluations of in-room target response preferences can be answered, in part, by comparing the individual listener target curve ratings for each of the two playback methods to see if they are in agreement. Here, the individual listener preference results are essentially the same for both in situ and BRS playback method in all three rooms under testing. Further to this point, though individualized HRTFs are found to be superior to even the “best fit” non-individualized HRTFs in terms of localization (Andreopoulou and Katz, 2015), research has found poor correlation between spatial and timbral aspects of binaural playback (Le Bagousse et al., 2011).

It should be noted that you cannot prove the null hypothesis to be true simply because there is no evidence to prove it false. The lack of an effect due to the BRS method in this study doesn’t preclude there will not be an effect if the test is repeated using a different set of experimental conditions or listeners. Nonetheless, these results are encouraging and indicate that the BRS system has sufficient accuracy to allow listeners to discern differences in room target curves to the point where they can make repeatable preference ratings that are virtually the same as those made in situ. Of interest to note is that these test results, like others before (Olive et al., 2007), were created using a generalized BRS system that was not calibrated to any particular listener participating in the experiment. The results show a high level of listener reliability between the two methods in absence of a more complex or individually calibrated system. Considered together, the results of the statistical analysis points towards the "robustness" of the BRS system used in these experiments and support it as a viable option for listening tests in lieu of impractical in situ listening tests.

4.8 Limitations of the Binaural Room Scanning System

Limitations of the BRS system have been summarized in Figure 4.2 and further described in section 4.4. For applications related to testing room response curves, the main limitation of the BRS system is that it does not take into account tactile information, such as low frequency reproduction. The results of the validation study described in this chapter indicate that the absence of this information did not play a significant role in the experimental results. However, for testing that aims at recreating a cinema environment, where significantly more tactile information is present (subwoofers), failure to include such tactile information in the playback experience may produce different results. This lack of tactile information may have led to the similarity in scores between curves D and E. This is an area that requires additional research.

With regards to applications of the BRS system beyond those identified in this research, the BRS system should be further verified utilizing additional attributes. A potential avenue of additional research in this area is the use of more detailed attributes in which to compare in situ listening to BRS listening. Beyond a simple comparison of target room response curves, Nicol et al suggests the use of timbral attributes typically associated with instruments, speech and spatial sound reproduction systems (Nicol et al., 2014), (Elliott et al., 2013), (Mattila, 2001), (Zacharov and Koivuniemi, 2001)

Nicol et al further recommend the attribute of 'naturalness' since binaural playback typically aims to mimic natural listening inclusive of the realism, naturalness, and the fidelity of the original sound scene. The attribute of naturalness also takes on importance as its ratings have be strongly correlated with preference (Berg and Rumsey, 2000).

4.9 Summary

This chapter reports the results of a validation experiment performed using a BRS system that captures and stores binaural room impulse responses of sound sources within a given listening space and then reproduces them via a head-tracking headphone system.

The experiments investigated whether a listener's room response curve preference ratings made in situ were similar to ratings made using a BRS reproduction of the same room. Three rooms, which mimicked typical movie viewing rooms, were utilized during these tests. They included a living room, a purpose-built home theatre and large cinema sized space.

The main conclusions of these experiments are as follows:

1. There were no significant differences in preference between the BRS and In situ methods.
2. The main effect was due to the room response curve.
3. There was an interaction between the different listening rooms, and the target curves. This interaction was confined to curve D.
4. There was an interaction between target curves D and E which could be partially attributed to individual preference.

4.10 Further Research Using the BRS System

The validated BRS system is used as an aid in determining the preferred in-room target response and how its perceived sound quality is influenced by program, cinema size and listening position within a cinema. Refer to Chapter 5 for a description of this research.

Chapter 5 - Subjective Listening Test for Preferred Room Response

5.1 Introduction

> *The use of a theatre loudspeaker system is the adaptation of objective equipment to an art. Any method of rating such a tool must include the effects of the subjective factors.* John K. Hilliard, Altec Lansing Corporation, 1949 (Hilliard, 1949)

Current cinema audio research has focused on room calibration standards, such as SMPTE ST202:2010, and has mainly been centred on room response measurement techniques and evaluations. The SMPTE X-curve stands largely unchanged for its origins, though improvements in cinema equipment and dramatic changes in cinema acoustics have occurred. The initial shape of the X-curve was driven by the need to compensate for poor high frequency response of the playback equipment and a belief that a large amount of high frequency roll-off was further required in order to model a flat response in the far-field of a reverberant space. Even as playback equipment improved, the argument to maintain the shape of the X-curve was based on somewhat speculative and often non-scientific subjective experiments which were typically conducted in highly reverberant spaced such as churches. These experiments, which continue to anchored the explanation for the shape of the X-curve, were conducted in the 1960's and 1970's in facilities with equipment and room acoustics that do not represent today's modern cinemas.

Further, the results of the survey in chapter 3 calls into question the validity of dub-stage to cinema translation in light of the variability between mixers' preferences and their hearing abilities. The theory proposed in this chapter is that, in calibrating a room to any specified target response curve, the cinemagoers' preferences should be taken into account when establishing an ideal target curve to tune a cinema to and that these preferences must be obtained with scientific rigor.

Conducting in situ subjective listening tests in more than one commercial cinema utilizing multiple listeners over a given time period can become logistically challenging. As previously discussed in Chapter 4, the Binaural Room Scanning (BRS) system utilized afforded the opportunity to overcome many logistical issues as well as providing methodological benefits over in situ listening tests. The validity and efficacy of the BRS system utilized for these experiments has been borne out in previous research (Gedemer and Welti, 2013b), (Olive and Welti, 2009), (Olive et al., 2007). The BRS measurement and test setup were originally described in the author's previous work (Gedemer, 2015b) and in Chapter 4 but will be summarized in this chapter.

This chapter focussed on obtaining listener preference data in three different sized cinemas and at three different listening positions within those cinemas. This effort was undertaken to not only review overall listener preference but to also see if there are preference changes with cinema size or listening position.

Work in this chapter has previously been presented as the following:

> GEDEMER, L. 2016. Subjective Listening Tests for Preferred Room Response in Cinemas - Part 2: Preference Test Results. AES 140th Convention. Paris, France: Audio Engineering Society

5.2 Motivation

The published standard for cinema calibration depicts recommended target response curves against which to tune dub-stages and commercial cinemas. These target curves are an effort to provide the best possible translation between what the re-recording mixer creates on the dub-stage and what the audience in the cinema ultimately hears. However, a review of relevant literature on this subject has not produced a definitive study on whether the published, recommended target curves have been properly vetted through scientific method. The question remains as to whether these target curves are considered ideal and are preferred by listeners.

The following questions serve as a basis for the developed subjective listening tests:

- What is the preferred spectral balance (in-room target response curve)
  - For conventional music tracks?
  - For X-curve monitored film sound tracks?
  - For non X-curve monitored film sound tracks?
- Do listeners' perceptions of spectral balance change with venue size?

- Do listeners' perceptions of spectral balance change with seating location?

5.3 Methodology

5.3.1 Binaural Room Scanning

The system computer (Lenovo Thinkpad T420s) ran a custom Matlab program which generated the test signals and recorded the measured sets of BRIRs. A 4-second swept sine wave ranging from 20 Hz to 20kHz was use as the measurement signal with several repeated measurements being averaged together to improve signal-to-noise. The playback levels for the sine sweeps were adjusted to create an SPL of approximately 85dB C-weighed at a position 2/3rds of the length of the cinema from the screen, in the center seat (2/3L position) as per the SMPTE standard.

In each cinema, a single JBL M2 loudspeaker was placed at the front and centre of the cinema in front of the projection screen and lifted to a position aimed at the 2/3L seat. The aiming of the loudspeaker was confirmed using a Bosch GLL 2-45 self-levelling long-range cross line laser. The decision to place the loudspeaker in front of the screen instead of behind it was twofold; first was due to the inconsistencies in the amount of high frequency roll-off in projection screens and the desire to negate this variable, the second was the lack of access to the space behind the screen in certain theatres. A mono playback configuration was chosen deliberately as previous research has found that test subjects tend to be more discerning when listening to audio in mono versus multi-channel format (Olive and Welti, 2009), (Olive et al., 2008), (Toole, 1983).

The Binaural Room Impulse Responses (BRIR's) were measured using the custom-built, full manikin described in chapter 4. BRIR's were taken in 5º increments from -40º to +40º to render a total set of 17 BRIRs at each of three listener locations within the three cinemas; the second row (front), a location approximately 2/3rd the length of the cinema (2/3L) and the second to last row (back). These positions allowed not only a comparison between different sized cinemas, but a comparison of listening positions within a cinema as well.

During playback, program material was convolved with the BRIR filter sets using a Pure Data based real-time convolution engine, shown in appendix J. The head-tracker monitored the angular position of the listener's head and sent the updated coordinates to the BRS playback software that switched to the appropriate BRIR filters corresponding to that angle.

5.3.2 Calibration of Manikin Microphones and Playback Headphones

For the listening tests, the same set of Sennheiser HD518 circumaural headphones as used in the experiment described in chapter 4 were chosen as the reference headphone. The headphones were measured on the binaural room scanning manikin, as described in chapter 4, and the resultant calibration filters were implemented in the Pure Data real-time convolution engine.

5.3.3 The Selected Listening Rooms

Professional screening rooms located within Los Angeles, CA were utilized for this research as they provided an opportunity to take measurements in a cinema setting without the operational issues of using a commercial cinema. A total of three different sized screening rooms were selected in order to represent the varying sizes of professional cinemas, screening rooms and dub-stages. BRS measurements were taken in the cinemas described in table 5.1 and shown in figured 5.1 through 5.3.

Cinema	**Seat Count**	**Seating Area (in meters)**	**RT_{60} (in seconds)**
A	60	12.8L x 8W	0.26
B	161	17L x 14W	0.44
C	516	46L x 20W	0.36

Table 5.1: Research cinemas utilized

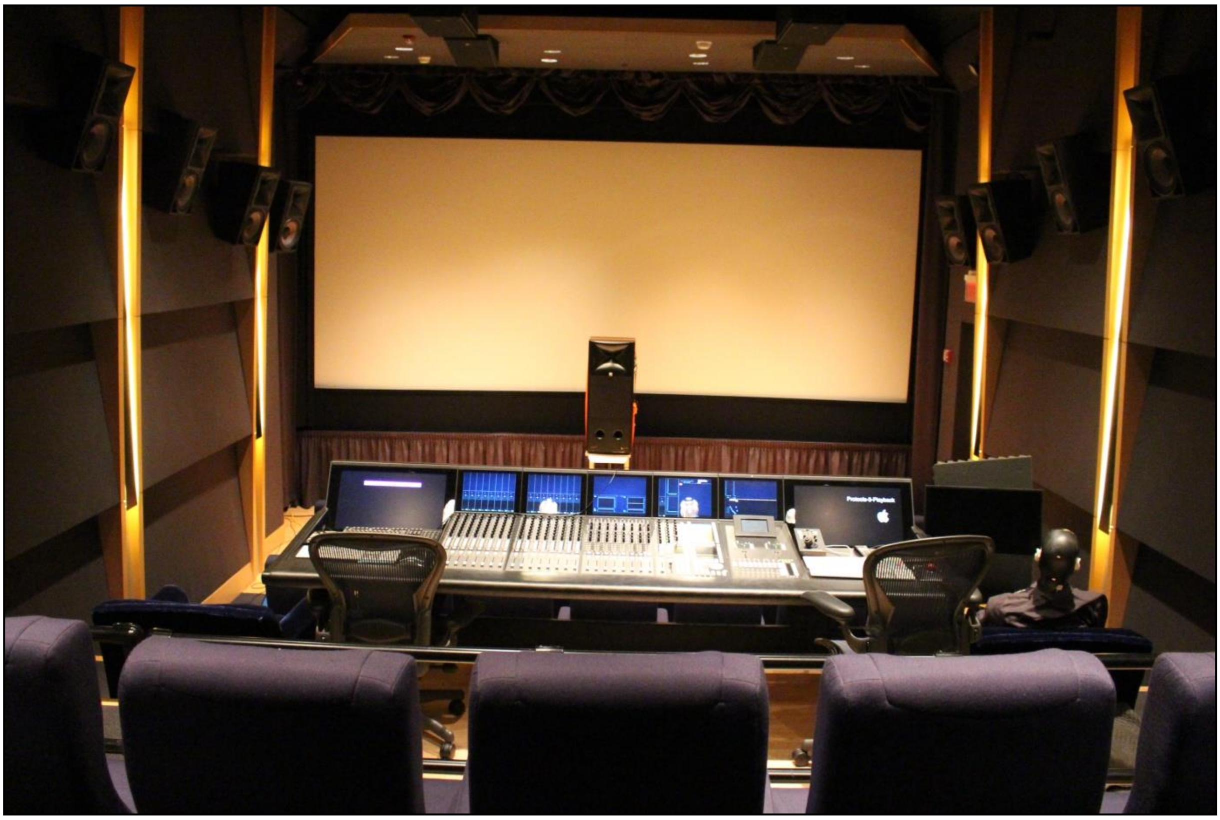


Figure 5.1: Cinema A

Figure 5.2: Cinema B

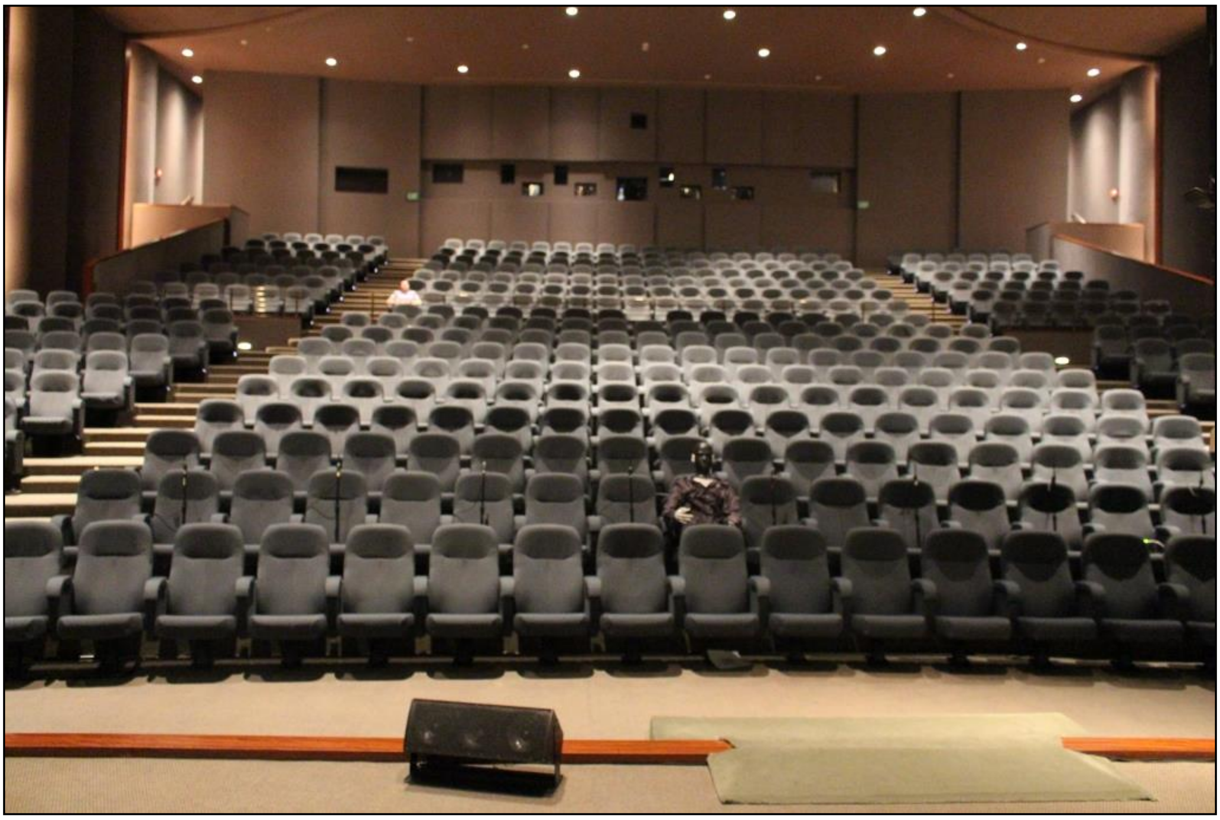


Figure 5.3: Cinema C

5.3.4 Flattening of In Situ Loudspeaker Responses

The JBL M2 loudspeaker was measured at the three listening positions in each cinema using the Harman Audio Test System (HATS) and nine calibrated dbx RTA-M omnidirectional microphones. The microphones were then placed in an array centred around the listening positions, an example of which is shown in figure 5.4.

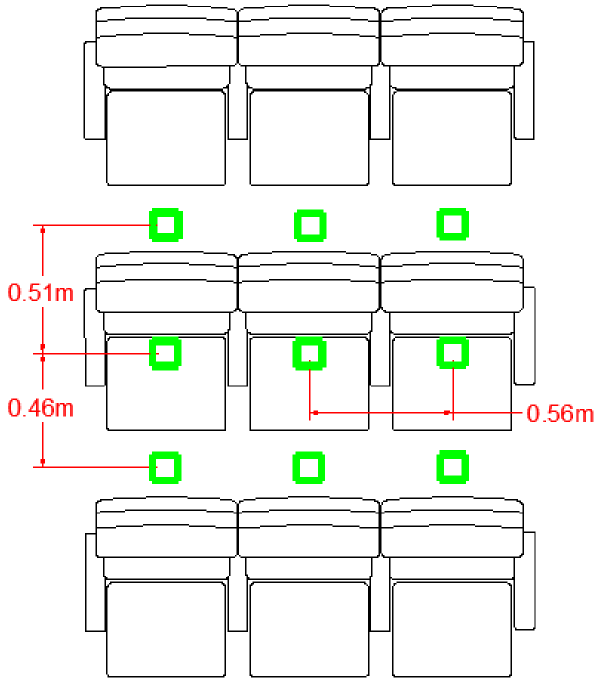


Figure 5.4: 9-microphone array for spatial averaging

The spatial averages were then equalized from 1kHz and below for each cinema listening position in order to create a "flattening" filter. Equalization above 1kHz was not carried out in order to leave high frequency attenuation effects, such as air loss, intact. This decision was driven by the desire to observe if naturally occurring high frequency attenuation could affect listeners' preference of target curve while positioned in the back of the larger cinemas. In addition, research has shown that what the listener hears in a cinema above 1kHz is primarily driven by the direct sound field (Toole, 2015), (Gedemer, 2015a). Any anomalies not attributable to the loudspeaker are very likely cause by the seats and / or other treatment in the cinema. These are often of non-minimum phase and can be difficult to equalized without the use of specialized filters (Maamar et al., 2006), (Marques and Freitas, 2005) which are typically not implemented in commercially available cinema audio processors. In addition, these anomalies vary from seat to seat and any attempt to correct them at one seat will most like cause additional issues at another.

The flattening filter was created using the "AutoEQ" function in HATS and was implemented using a BSS Audio BLU-800 processor for playback and applied to the BRS playback system. This filter, along with the calibration filters of the manikin microphones and headphones, was used to create a final collective "flat" baseline response which was used in conjunction with each of the 5 in-room target response curves. This ensured that any audible differences among curves were likely related to differences in their frequency responses, and not the response of the manikin microphones, playback headphones, calibration microphones or in situ loudspeaker response.

5.3.5 Creation of In-room Target Response Curves

Four of the in-room target curves were created and implement within a BSS Audio Blu-800 DSP processor. These curves, notated simply as A, B, C, D, are shown in Figures 5.5 through 5.8. Curve A simulated the SMPTE X-curve with a high frequency roll off starting at 2kHz that goes down by 4dB at 8kHz and a low frequency roll off starting at 50Hz. Curve B represents what was used as an audible anchor and was roughly based on the SMPTE X-curve but with an added rise at 1kHz. Curve C is relatively flat with a 6.5dB bass boost below 105Hz and a -2.5dB roll off above 2.5kHz. This overall 9dB slope was derived from research by Olive et al on headphone target curves as well as loudspeaker preference curves (Olive and Welti, 2015), (Olive et al., 2013b). Curve D was derived from research conducted by Olive et al. on room correction products (2009) and further headphone research by Fleischmann et al. (2012). This in-room target curve was flat with a 7dB boost starting below 100Hz, and a gentle treble cut of -1dB starting at 235Hz. This 8dB slope or delta of the preferred target curve is similar to the target curve that was preferred in a recent study on headphone target curves (Olive et al., 2013a). The fifth "curve", E, utilized no equalization thus leaving a flat room response. The decision to forgo a curve with an upward tilting slope towards the high frequencies was based on the earlier research by the Boners, which found that listeners did not prefer room responses with this type of shape (Boner and Boner, 1965). This finding was later reinforced by the 'Method of Adjustment' test utilized by Olive (Olive et al., 2013b).

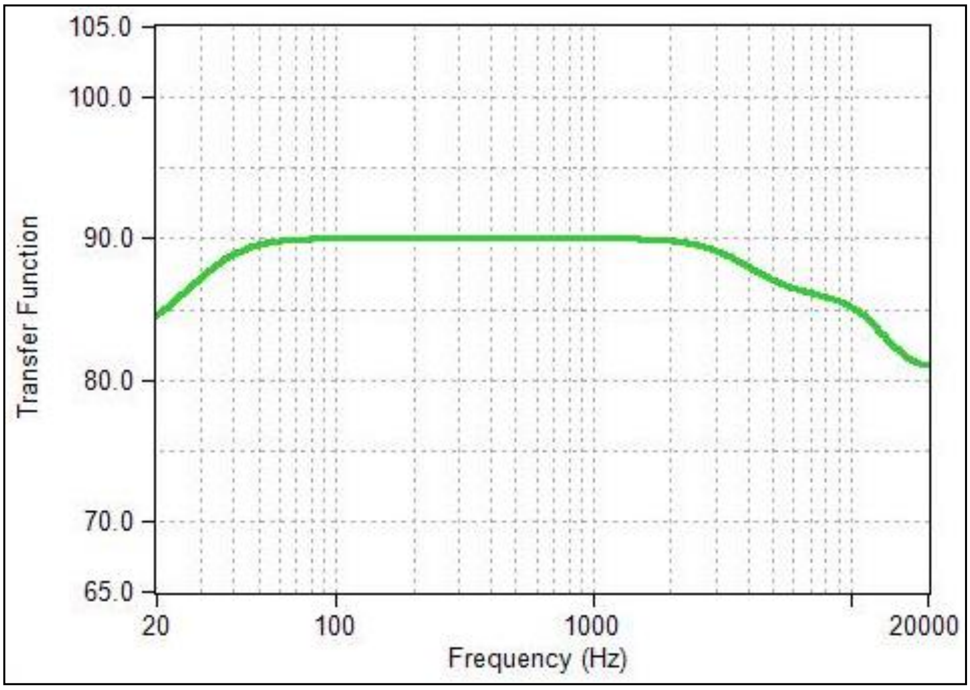


Figure 5.5: Target room response curve 'A'

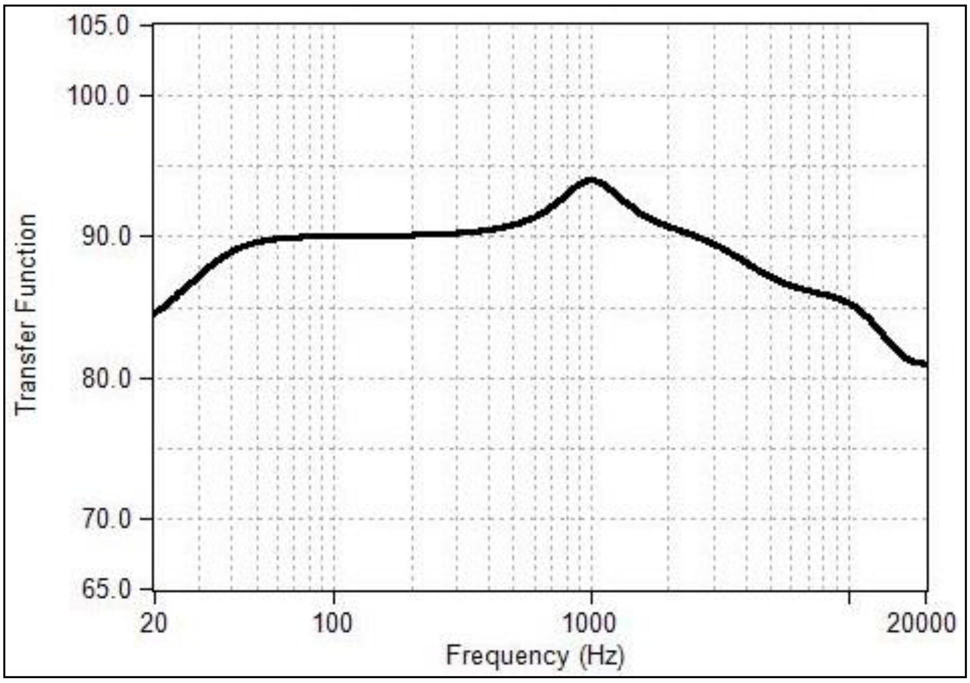


Figure 5.6: Target room response curve 'B

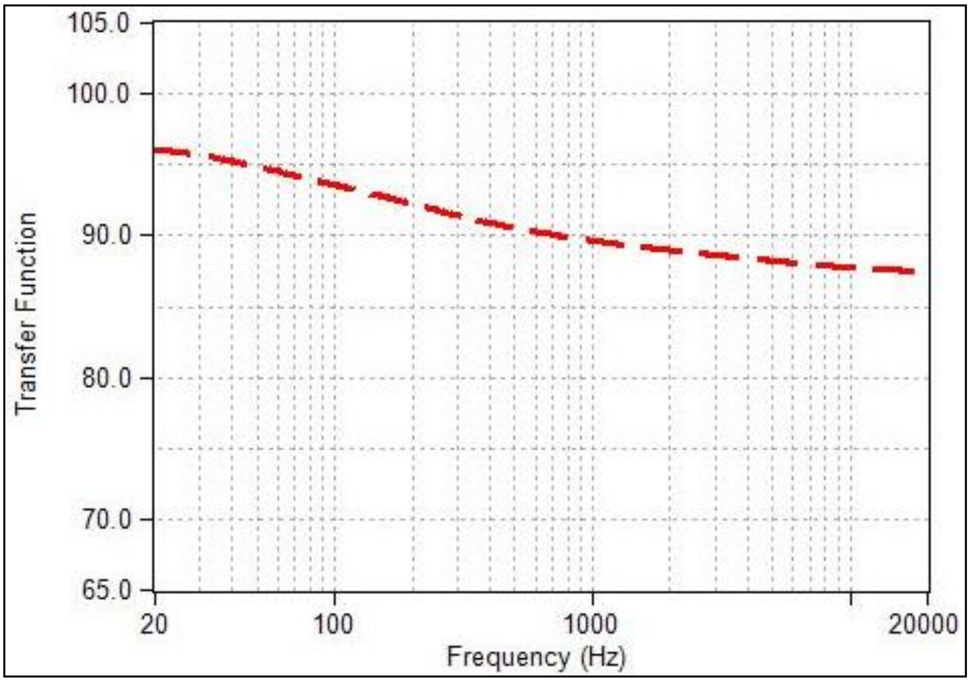


Figure 5.7: Target room response curve 'C'

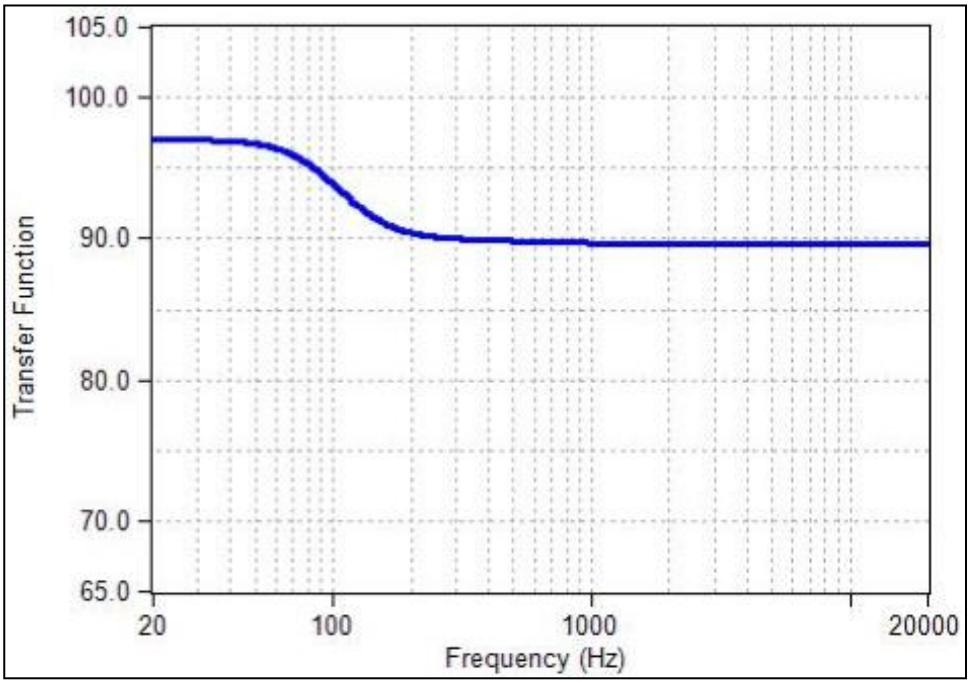


Figure 5.8: Target room response curve 'D'

Correction filters for the combined manikin head microphones and playback headphones, as well as the in situ loudspeaker responses were applied to each of the 5 target response curves. This ensured that any audible differences among curves were likely related to differences in their frequency responses, and not the response of the manikin microphones, playback headphones, calibration microphones or in situ loudspeaker response. The resultant curves, as heard by the listeners during testing, are described and shown in section 5.3.6

5.3.6. Target Curve Creation and Implementation

Figures 5.9(a-e) through 5.11(a-e) illustrate the target room response curves and the listener curves at each position. The target room response was added to the measured in-room response, the in situ loudspeaker flattening curve in each cinema and position, in addition to the flattened headphone response. These curves depict what the listeners actually heard during the tests given their choice of curve and their listening position in the cinema. The curves for each listener position are shown offset for clarity and in 1/6 octave resolution.

Figures 5.9 a-d: Target room response curves and the listener curves at each position in cinema A

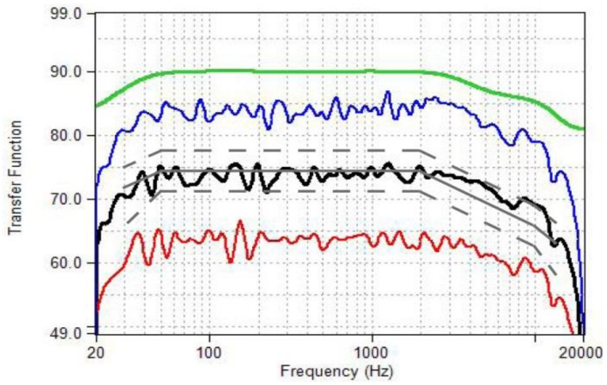


Figure 5.9(a): Cinema A: Green - Curve A, Blue - Front Row, Black - 2/3L, Red - Back Row. SMPTE X-curve and tolerances in grey

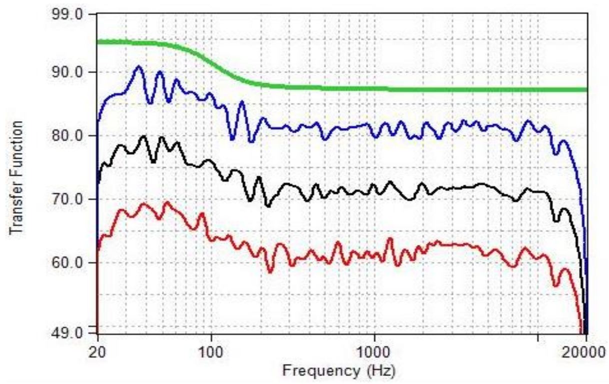


Figure 5.9(d): Cinema A: Green - Curve D, Blue - Front Row, Black - 2/3L, Red - Back Row

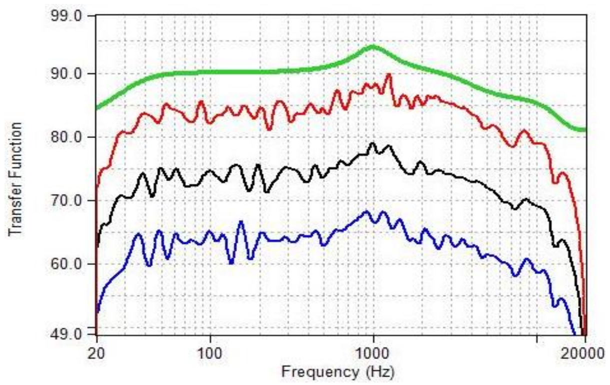


Figure 5.9(b): Cinema A: Green - Curve B, Blue - Front Row, Black - 2/3L, Red - Back Row

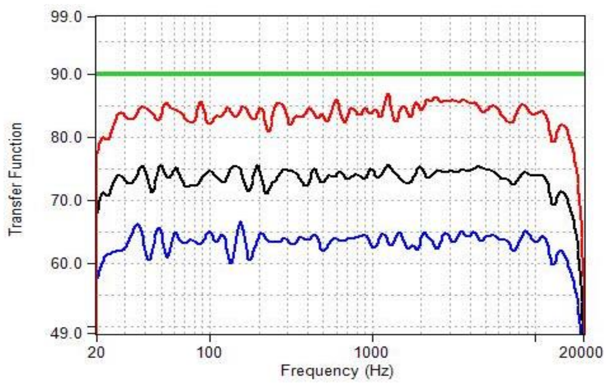


Figure 5.9(e): Cinema A: Green - Curve E, Blue - Front Row, Black - 2/3L, Red - Back Row

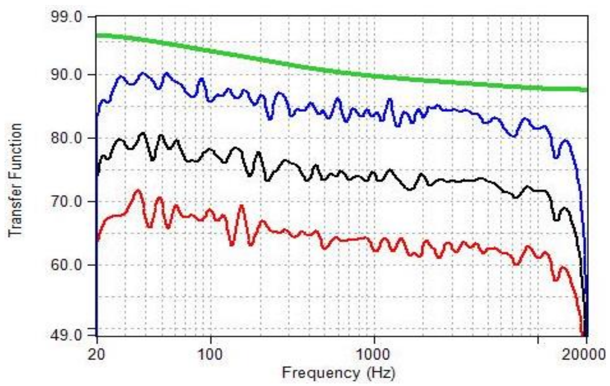


Figure 5.9(c): Cinema A: Green - Curve C, Blue - Front Row, Black - 2/3L, Red - Back Row

Figures 5.10 a-d: Target room response curves and the listener curves at each position in cinema B

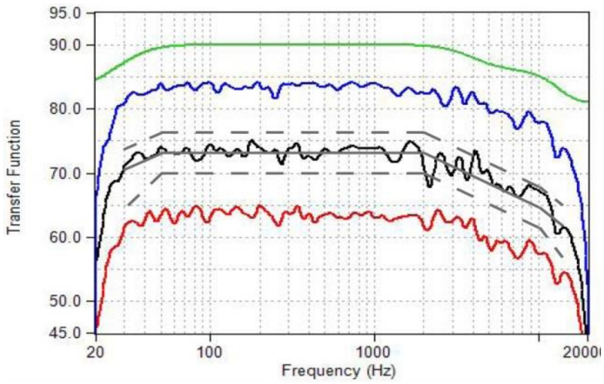


Figure 5.10(a): Cinema B: Green - Curve A, Blue - Front Row, Black - 2/3L, Red - Back Row. SMPTE X-curve tolerances in grey

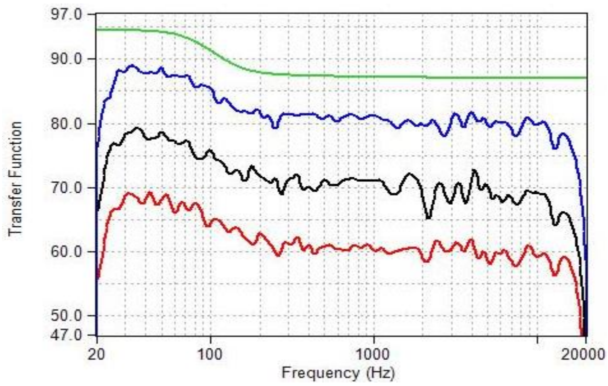


Figure 5.10(d): Cinema B: Green - Curve D, Blue - Front Row, Black - 2/3L, Red - Back Row

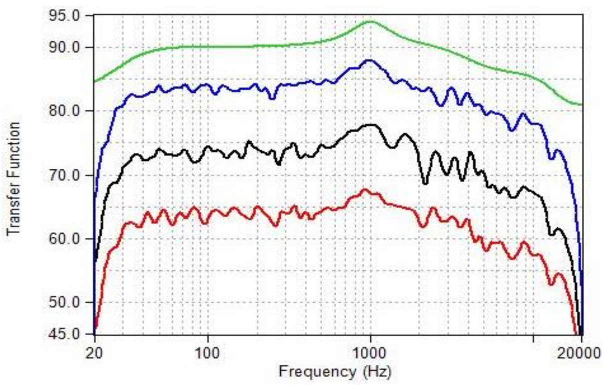


Figure 5.10(b): Cinema B: Green - Curve B, Blue - Front Row, Black - 2/3L, Red - Back Row

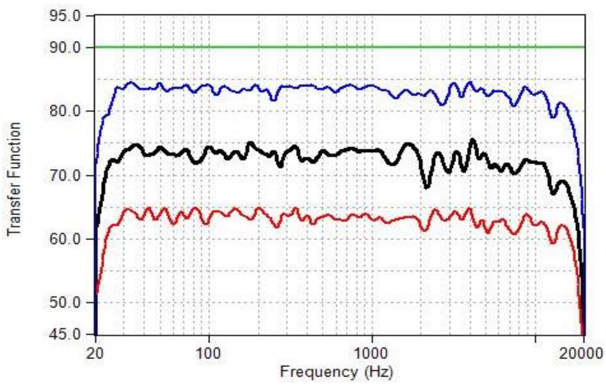


Figure 5.10(e): Cinema B: Green - Curve E, Blue - Front Row, Black - 2/3L, Red - Back Row

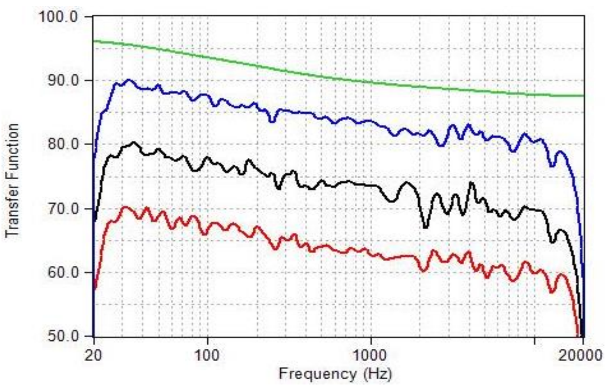


Figure 5.10(c): Cinema B: Green - Curve C, Blue - Front Row, Black - 2/3L, Red - Back Row

Figures 5.11a-d: Target room response curves and the listener curves at each position in cinema C

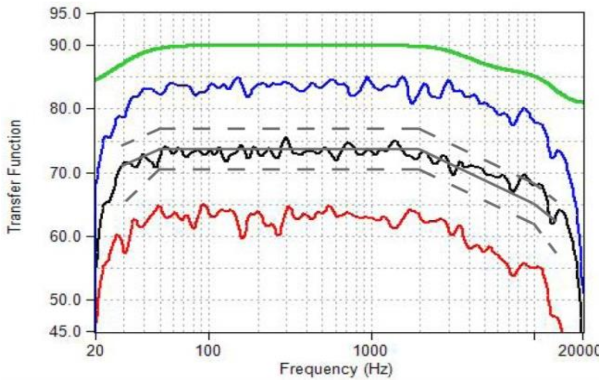


Figure 5.11(a): Cinema C: Green - Curve A, Blue - Front Row, Black - 2/3L, Red - Back Row. SMPTE X-curve and tolerances in grey

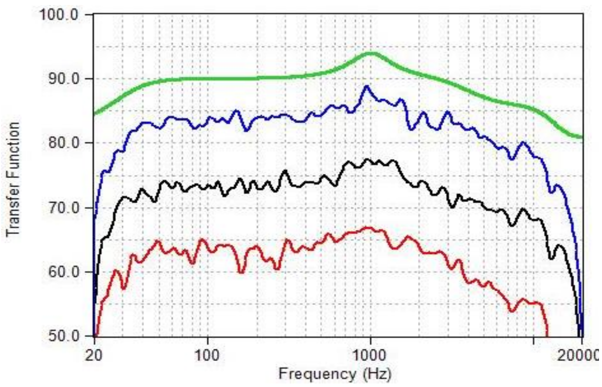


Figure 5.11(b): Cinema C: Green - Curve B, Blue - Front Row, Black - 2/3L, Red - Back Row

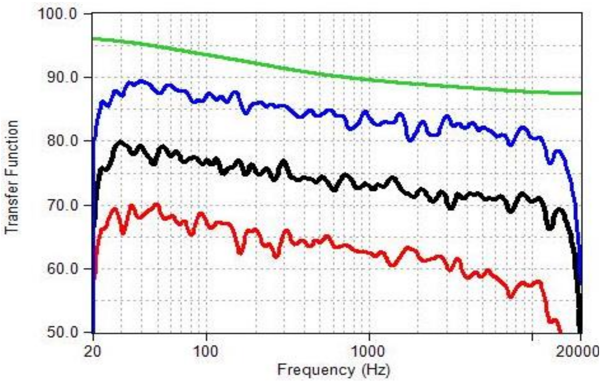


Figure 5.11(c): Cinema C: Green - Curve C, Blue - Front Row, Black - 2/3L, Red - Back Row

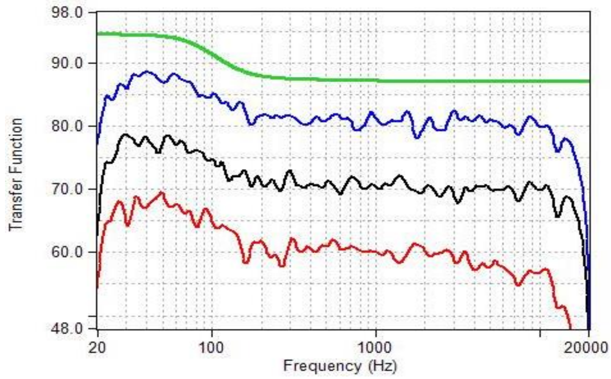


Figure 5.11(d): Cinema C: Green - Curve D, Blue - Front Row, Black - 2/3L, Red - Back Row

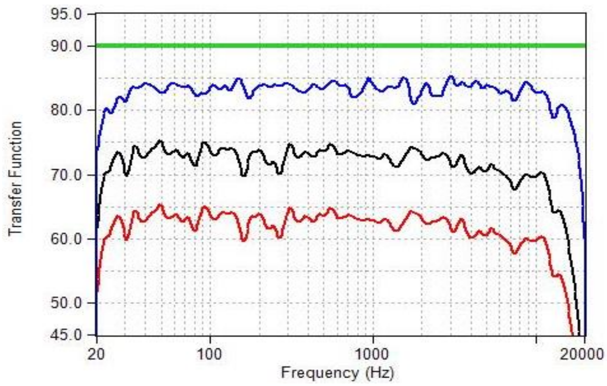


Figure 5.11(e): Cinema C: Green - Curve E, Blue - Front Row, Black - 2/3L, Red - Back Row

5.3.7 Program Material

Five different mono music programs were selected, level matched and edited into short 20-30 second loops in ProTools. They included a piece of orchestral score mixed in a room that was not calibrated to the X-curve (***Sc***), the same piece of score mixed for a film trailer in an X-curve calibrated room (***Sc-T***), a piece of score that was mixed for a theatrically released film in an X-curve calibrated room (***Sc-F***), Steely Dan's "Cousin Dupree" (***SD***), and Toy Matinee's "Last Plane Out" (***TM***). All five programs have broad and smooth long term spectra, and have been found to be revealing of spectral features based on previous tests and listener training exercises.

The mono program from the original 5.1 channel music selections were down mixed according to recommended ITU practices as shown in table 5.2.

Mono 1/0 Format	**Left**	**Right**	**Centre**	**Left Surround**	**Right Surround**
C'	0.7071	0.7071	1.0000	0.5000	0.5000

Table 5.2: Downward mixing equations for 3/2 source materials, from (ITU, 2012a)

This process was accomplished using Waves M360° Surround Manager software, which has the ITU standard mix-down levels as a selectable option, shown in figure 5.12 (Waves, 2005).

Figure 5.12: Screen capture of Waves M360° Surround Manager software showing mono ITU mix-down option

Critical listening assessments conducted by five professional sound designers revealed no audible artefacts from the mix-down process.

The ITU standard does not state a specific method for the down-mix of the low frequency enhancement (LFE) channel. However, it does state that the low frequency content of a mix, from 20 to 120Hz, should not be carried solely by the LFE channel. Therefore, the channels listed in the above table all carry signals in this frequency range. However, the LFE channels of the audio tracks were part of the final mono mix and were added at a ratio of approximately 0.30. The exact amount of LFE added to the mono mix varied by program. However, the levels were verified against the original 5.1 audio tracks in ProTools, through informal listening tests and via spectral analysis as shown in figure 5.13.

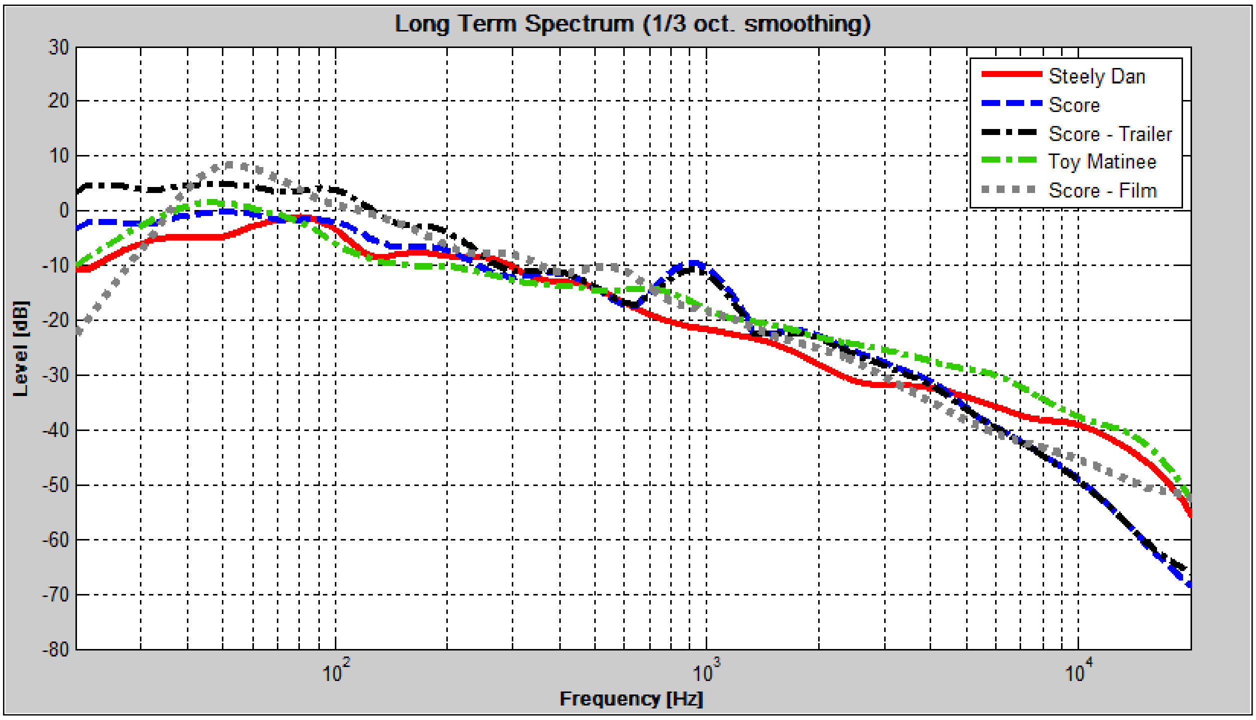


Figure 5.13: Long-term spectral power density of the five pieces of program material as calculated using the Welch method (Welch, 1967). Levels normalized at 500Hz

5.3.8 Selection of Listeners

A total of 14 trained listeners were used for these tests, 10 of whom are Harman employees paid for their participation. The other 4 listeners were audio professionals working in cinema sound design and editing with a range of experience from 4 to 30 years in the industry. These listeners were remunerated for their participation. The age range of the listeners was 24 to 56 years, with a media age of 32 years. The listeners were tested to confirm that they have audiometric normal hearing. In addition, all listeners needed to have passed listener training tasks that required them to reliably identify and discriminate among different spectral distortions added to different music programs, achieving a skill level of eight out of ten or higher in all tasks. This training was carried out using Harman's "How to Listen" software (Harman, 2011). The listeners' experience with formal listening tests ranged from 1 to 23 years.

5.3.9 Test Design

The listening test was administered using a MUSHRA-style test where subjects were asked to rate their preference amongst five different in-room target response curves applied to three different seating locations within three different cinemas.

This listening test utilized a 5 x 5 x 3 x 3 x 2 repeated measures ANOVA design, with the following independent variables: five room response curves (A through E), five programs (three film score and two popular music pieces), three rooms, three listening locations and two observations. The dependent variable was preference rating. For the listening tests, the target curves were comparatively rated on an 11-point preference scale (0-10) that had semantic differentials on every second interval labelled: *Strong Dislike* (1), *Dislike* (3), *Neither Like/Dislike* (5), *Like* (7), and *Strong Like* (9). A strong preference between target curves was indicated by a separation in preference ratings of ≥ 2 points, a moderate preference ≥1 points, and a slight preference ≤ 0.5 rating. The graphical user interface (GUI) for the preference test is shown in figure 4.14 in chapter 4.

All presentations of the stimuli were performed using a double-blind, multiple comparison method (five target curves at a time) that utilized a randomized presentation order for both the target curves and programs. The order in which the various cinema / position tests were completed was equally distributed among the subjects to minimize possible order biases. For all tests the stimuli were presented the same number of times and at the same playback level.

Each listener participated in a listening session lasting approximately 20-30 minutes each. In each test session, the listener completed 10 trials (5 programs x 2 observations). 1,260 trials were carried out in total (10 trials x 3 cinemas x 3 listening positions x 14 test subjects). In each trial, the test subject made comparative judgments between the five in-room target response curves until their final preference ratings were recorded. The listening tests were administered in random order across all of the listeners over a 6 month period with no test subject taking more than 2 listening test within a given day. In some cases, listening tests were administered weeks apart.

5.3.10 Test Setup and Hardware

The hardware for the listening test setup consisted of the Sennheiser HD518 headphones, two Windows laptops (both Lenovo Thinkpad T420s), two digital sound cards (RME Fireface UC and MOTU) a programmable digital signal processor (BSS Audio BLU-800) and magnetic field head-tracker (Razer Hydra). The Razer Hydra head-tracker utilized has a latency of 10ms. The latency of the entire BRS playback system is less than 50ms which, as previously referenced in chapter 4, does not cause audible effects under normal listening conditions.

The relative playback levels of the different target response curves and program material were adjusted for equal loudness based on the ITU-R1770.3 loudness standard (ITU, 2012b) and verified by informal listening to assure that there were no discernible loudness differences between test curves. The average playback level of the music was adjusted to a comfortable level of 80dBA. The experiments were conducted with the listeners sitting in one of the loudspeaker labs at Harman, where the background noise level was measured to be between 30 and 35dBA.

5.4 Results

The following sections report on the statistical analysis utilized and the results of the listening tests.

5.4.1 Statistical Analysis

The data was analyzed using SAS Stat View software running a one-way repeated measures analysis of variance (ANOVA). Additional analyses were completing using Matlab and custom designed Excel spreadsheets. Since the same participants experienced all conditions that were assessed in the tests, a one-way repeated measures ANOVAs was deemed appropriate in order to identify whether there were any significant differences of the mean ratings for preference between the five presented target curves. A complete factorial analysis was used in the ANOVA model with a significance level of 0.05 utilized for all statistical tests. Since an audible anchor was included in the target curves, normalizing the data prior to statistical analysis was deemed unnecessary. In addition, noting that outliers can provide important information during analyses, no outliers were removed from the data sets.

A Mauchly's test of sphericity with a significance level of 0.05 was utilized on various groups of data (grouped by curve, room, program, listening position and cinema) to confirm local sphericity. The grouped data from the listening tests were found to be non-significant ($p>0.05$) are therefore were considered locally spherical.

After the ANOVA analyses were completed, a repeated measures Scheffe post hoc analysis with a significance level of 0.05 was completed in order to define which means were significantly different from each other. Following the Scheffe analysis, several effect size analyses were complete in an effort to identify the degree of the differences between and within factors.

5.4.1. Results

The intention of these tests was to study for preferences among cinema target curves based on cinema size and listening position within a cinema. Data from the listening tests were checked for Gaussian distribution and subjected to a repeated measures ANOVA test. The following were observed:

- ***Observation*** was not significant: $\boldsymbol{F}$ (1,13) = 0.186, $\boldsymbol{p}$ = .6731
- ***Listening Position*** was not significant: $\boldsymbol{F}$ (2,26) = 1.04, $\boldsymbol{p}$ = .3676
- ***Target Curve*** was significant: $\boldsymbol{F}$ (4,52) = 547.123, $\boldsymbol{p}$ < .0001
- ***Cinema*** was significant: $\boldsymbol{F}$ (2,26) = 15.639, $\boldsymbol{p}$ = .0001
- ***Program*** was significant: $\boldsymbol{F}$ (4,52) = 12.607, $\boldsymbol{p}$ = .0001
- ***Target Curve*Program*** was significant: $\boldsymbol{F}$ (16,208) = 34.868, $\boldsymbol{p}$ = .0001
- ***Cinema*Target Curve*** was significant: $\boldsymbol{F}$ (8,104) = 7.873, $\boldsymbol{p}$ = .0001
- ***Cinema*Program*** was significant: $\boldsymbol{F}$ (8,104) = 5.608, $\boldsymbol{p}$ = .0001
- ***Cinema*Listening Position*** was significant: $\boldsymbol{F}$ (4,52) = 2.744, $\boldsymbol{p}$ = .038

The full ANOVA table for this experiment can be found in appendix K.

Only those higher-order factors or combination of factors found to be significant were analyzed separately and are elaborated upon in the following sections.

5.4.2 Results Based on Target Curve

To check for homoscedasticity within and between the 5 target curves, box plots of the conditions were created, as shown in figure 5.14.

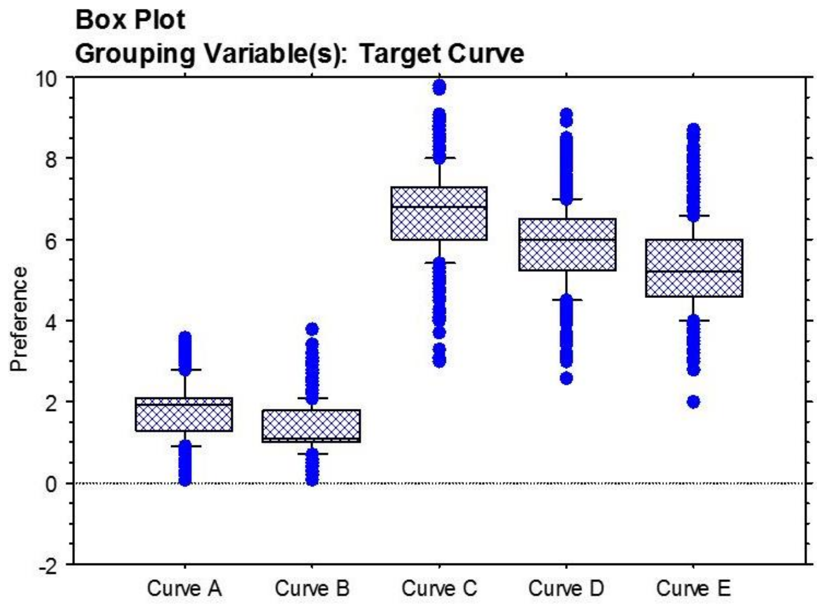


Figure 5.14: Homoscedasticity between *target curves*

From figure 5.14 one can observe that the dispersion of scores is more or less comparable for each of the 5 target curves and there exists a distinct level of agreement within the scores for any one of the given curves. Most of the outliers shown can be considered "suspected outliers", as they are within two to three times the interquartile range and show distributions centred around the interquartile range. Curves A and B have a large proportion of their interquartile ranges overlapping while curves C, D and E also display some degree of overlap among their interquartile ranges. Both of these observations warranted further analyses as to the nature of these relationships.

The skewness for each of the target curves is well between 1.0 and -1.0, showing a fairly symmetrical distribution of the data; reference table 5.3. In addition, the kurtosis for each curve shows that the data is near a normal (Gaussian) distribution of $\boldsymbol{k} \approx 0$.

	Curve A	Curve B	Curve C	Curve D	Curve E
Mean	1.778	1.336	6.682	5.893	5.275
Std. Dev.	0.717	0.618	0.92	0.964	1.065
Std. Error	0.02	0.017	0.026	0.027	0.03
Count	1260	1260	1260	1260	1260
Minimum	0.1	0.1	3	2.6	2
Maximum	3.6	3.8	9.8	9.1	8.7
Skewness	-0.127	0.688	-0.354	-0.274	0.119
Kurtosis	-0.279	0.249	0.443	0.032	-0.02

Table 5.3: Descriptive statistics split by *target curve*

A post hoc analysis, shown in table 5.4, was completed for the target curves and indicates that the differences in target curves were significantly different with the exception of curves A and B.

Scheffe for Preference
Effect: Target Curve
Significance Level: 5 %

	Mean Diff.	Crit. Diff.	P-Value	
Curve A, Curve B	.442	.475	.0818	
Curve A, Curve C	-4.904	.475	<.0001	S
Curve A, Curve D	-4.115	.475	<.0001	S
Curve A, Curve E	-3.498	.475	<.0001	S
Curve B, Curve C	-5.346	.475	<.0001	S
Curve B, Curve D	-4.557	.475	<.0001	S
Curve B, Curve E	-3.939	.475	<.0001	S
Curve C, Curve D	.789	.475	.0001	S
Curve C, Curve E	1.406	.475	<.0001	S
Curve D, Curve E	.618	.475	.0044	S

Table 5.4: Scheffe post hoc analysis for *target curve*

Of interest to note is that a Bonferroni / Dunn test was conducted on the same data set and curves A and B were found to be statistically different. This is most likely due to the Scheffe test being more conservative and thus being prone to type II errors.

Further information about the differences or similarities between the target curves is explored through effect size analyses, which are shown in table 5.5.

	d	η^2	*CL*
Curve	-	0.864	-
A/B	0.66		68.01%
A/C	-5.94	-	100.00%
A/D	-4.84	-	99.70%
A/E	-3.85	-	99.68%
B/C	-6.82	-	100.00%
B/D	-5.63	-	100.00%
B/E	-4.52	-	99.93%
C/D	0.84	-	72.31%
C/E	1.41	-	84.13%
D/E	0.61	-	66.65%

Table 5.5: Effect size scores for *target curves;* effect sizes between *target curves* highlighted in pink for intermediate differences

Large differences are shown between most of the pairs of target curves, though curves A and B show an intermediate difference. This result is in partial agreement with the results of the Scheffe post hoc analysis but does not completely explain why those results were shown as statistically similar. Ideally, an intermediate difference between curves should lead to a statistical difference in the post hoc test. Further to this, curves D and E also show the same intermediate level of difference though they were found to be statistical difference in the post hoc tests. This calls into question the results of the Scheffe test and points towards the results of the Bonferroni / Dunn test, where all pairs were found to be different, as being more accurate.

With a significance level of $\boldsymbol{F}$ (4,52) = 547.123, $\boldsymbol{p}$ < .0001 and effect size score of $\boldsymbol{\eta^2}$=0.864, the factor *target curve* has the largest influence of all factors on preference scores. There are, however, slight interactions between pairs of curves which points towards the influence of other factors. The relationships between all of the target curves and other factors are further evaluated in sections 5.4.5 and 5.4.7.

5.4.3 Results Based on Cinema

Box plots for all three cinemas, shown in figure 5.15, show a near identical distribution of scores.

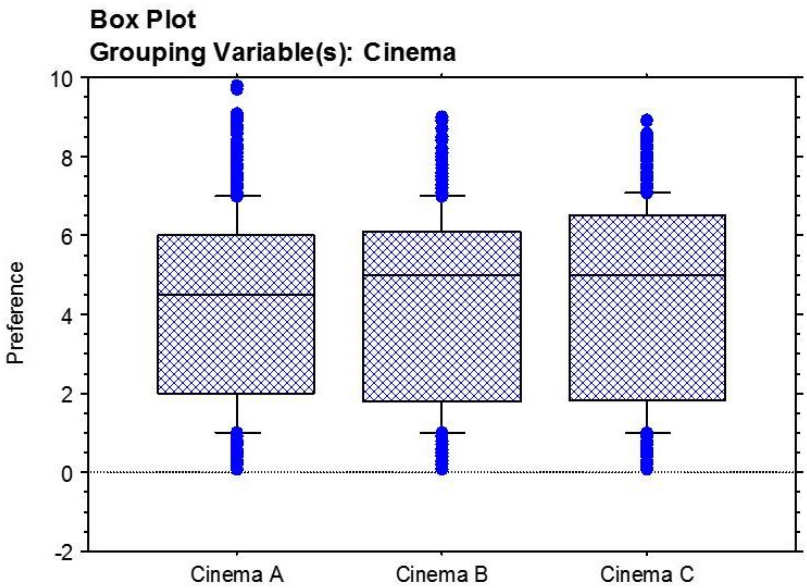


Figure 5.15: Homoscedasticity between *cinemas*

The skewness figures in table 5.6 show that the distribution of the data for each of the cinemas is nearly symmetrical and that the kurtosis for each cinema depicts the data following a uniform distribution for all three cinemas ($\boldsymbol{k} \leq -1.2$).

	Cinema A	Cinema B	Cinema C
Mean	4.108	4.159	4.312
Std. Dev.	2.277	2.37	2.453
Std. Error	0.05	0.052	0.054
Count	2100	2100	2100
Minimum	0.1	0.1	0.1
Maximum	9.8	9	8.9
Skewness	-0.055	-0.123	-0.2
Kurtosis	-1.339	-1.456	-1.496

Table 5.6: Descriptive statistics split by *cinema*

Though there is good agreement in the distribution of data across the three cinemas, the factor *cinema* was found to be of statistical significance in the ANOVA. A post hoc analysis (table 5.7) shows the statistical differences between the three cinemas.

Scheffe for Preference
Effect: Cinema
Significance Level: 5 %

	Mean Diff.	Crit. Diff.	P-Value	
Cinema A, Cinema B	-.051	.098	.4205	
Cinema A, Cinema C	-.204	.098	<.0001	S
Cinema B, Cinema C	-.153	.098	.0018	S

Table 5.7: Scheffe post hoc analysis for *cinema*

Cinemas A and B show results that are statistically similar per the Scheffe test, which is mirrored in the effect size analysis shown in table 5.8.

	d	η^2	***CL***
Cinemas	-	0.001	-
A/B	-0.02	-	50.62%
A/C	-0.09	-	52.43%
B/C	-0.06	-	51.79%

Table 5.8: Effect size scores for *cinemas*

However, the post hoc and effect sizes tests disagree with regard to the relationship between cinemas A and C, as well as B and C, where table 5.8 pair-wise comparisons shows little to no difference between these cinemas.

The ANOVA analysis shows that the cinemas are significant in their overall effect on preference scores (F (2,26) = 15.639, p = .0001) though the effect size of η^2=0.001 shows that the effect is negligible. However, further observations of the cinemas with respect to the other factors shows interactions among them. These interactions are explored further in sections 5.4.5, 5.4.6 and 5.4.8.

5.4.4 Results Based on Program

The box plots in figure 5.16 display nearly identical distribution of scores for all five programs.

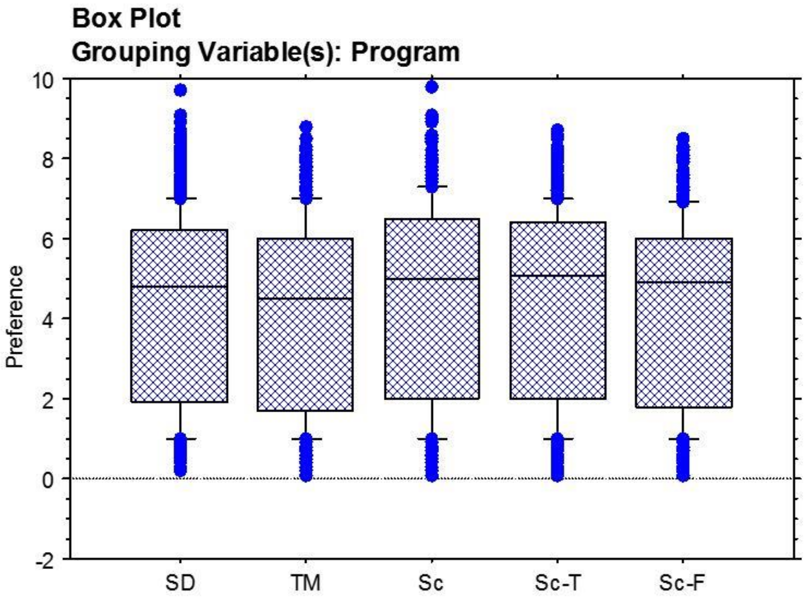


Figure 5.16: Homoscedasticity between *programs*

The skewness figures in table 5.9 show that the distribution of the data for programs SD and TM are symmetrical while Sc, Sc-T and Sc-F are slightly skewed. The kurtosis for each program depicts the data following a uniform distribution ($\boldsymbol{k} \leq -1.2$).

	SD	TM	Sc	Sc-T	Sc-F
Mean	4.175	4.045	4.338	4.309	4.097
Std. Dev.	2.386	2.324	2.448	2.363	2.311
Std. Error	0.067	0.065	0.069	0.067	0.065
Count	1260	1260	1260	1260	1260
Minimum	0.2	0.1	0.1	0.1	0.1
Maximum	9.7	8.8	9.8	8.7	8.5
Skewness	-0.067	-0.056	-0.128	-0.228	-0.174
Kurtosis	-1.443	-1.414	-1.426	-1.466	-1.455

Table 5.9: Descriptive statistics split by *program*

As with the cinemas, there is good agreement in the distribution of data across the five programs. However, since the factor *program* was found to be of statistical importance in the ANOVA, a post hoc analysis (table 5.10) was completed to review the statistical differences between the five programs.

Scheffe for Preference
Effect: Program
Significance Level: 5 %

	Mean Diff.	Crit. Diff.	P-Value	
SD, TM	.131	.164	.1796	
SD, Sc	-.163	.164	.0516	
SD, Sc-T	-.133	.164	.1649	
SD, Sc-F	.079	.164	.6725	
TM, Sc	-.294	.164	<.0001	S
TM, Sc-T	-.264	.164	.0002	S
TM, Sc-F	-.052	.164	.9013	
Sc, Sc-T	.029	.164	.9874	
Sc, Sc-F	.241	.164	.0008	S
Sc-T, Sc-F	.212	.164	.0045	S

Table 5.10: Scheffe post hoc analysis for *program*

Table 5.10 shows significant differences between four pairs of programs, while their direct, pair-wise comparison effect sizes show those differences to be negligible (table 5.11).

	d	η^2	*CL*
Program	-	0.002	-
SD/TM	0.06	-	51.56%
SD/Sc	0.07	-	51.90%
SD/Sc-T	0.06	-	51.59%
SD/Sc-F	0.03	-	50.94%
TM/Sc	0.12	-	53.46%
TM/Sc-T	0.11	-	53.17%
TM/Sc-F	0.02	-	50.63%
Sc/Sc-T	0.01	-	50.34%
Sc/Sc-F	0.10	-	52.85%
Sc-T/Sc-F	0.09	-	52.56%

Table 5.11: Effect size scores for *programs*

The factor *program* was found to be statistically significant, as shown by the ANOVA (F (4,52) = 12.607, p = .0001), though the effect size of η^2=0.002 shows that the overall effect was negligible. However, additional interactions with other factors were observed and therefore further analyzed was warranted, which is presented in sections 5.4.6 and 5.4.7.

5.4.5 Results Based on Target Curve*Program

Figure 5.17 shows the combined target curve preference data as it relates to program material; noting that the 95% confidence intervals are too small to be depicted clearly and +/- 1 standard deviation intervals are not used in order to maintain clarity in the graph.

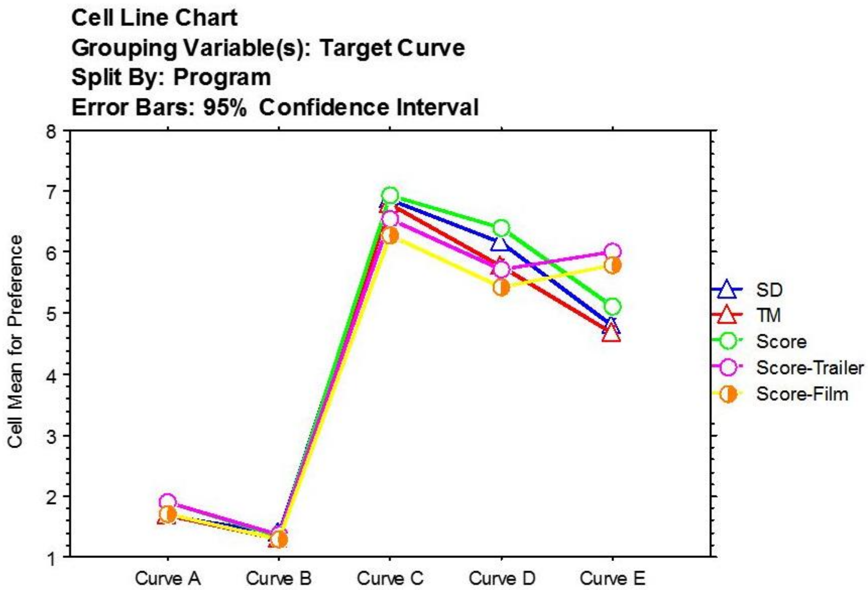


Figure 5.17: *Target curve* preference based on *program*

The ANOVA analysis found that the combined interaction between *target curve* and *program* produced significant results: F (16,208) = 34.868, p = .0001. While it has been previously shown that *target curve* alone produced the largest significant result, there are interactions with *program* which produced changes in preference within curves C, D and E. These changes are most noticeable for programs Sc-T and Sc-F, as reflected in table 5.12.

For the purpose of brevity, the effect size tables for *program* with respect to *target curve* and *target curve* with respect to *program* have been truncated to show only those effects that are of intermediate or large in nature. The full effect size tables for these two interacting factors are given in appendixes L and M.

	d	*CL*
Curve C - Program		
SD/Sc-F	0.63	67.30%
TM/Sc-F	0.56	65.41%
Sc/Sc-F	0.67	68.12%
Curve D - Program		
SD/Sc-T	0.50	63.72%
SD/Sc-F	0.83	72.11%
TM/Sc	-0.68	68.44%
Sc/Sc-T	0.75	70.19%
Sc/Sc-F	1.07	77.51%
Curve E - Program		
SD/Sc-T	-1.26	81.38%
SD/Sc-F	-1.10	78.18%
TM/Sc	-0.49	63.62%
TM/Sc-T	-1.40	83.82%
TM/Sc-F	-1.24	81.01%
Sc/Sc-T	-0.93	74.46%
Sc/Sc-F	-0.75	70.28%

Table 5.12: Effect size scores between *programs* with respect to *target curves;* effect sizes between programs highlighted in pink for intermediate differences and yellow for large difference

Within target curves A and B, there were little to no differences in preference scores between pairs of programs (i.e., program had no effect on preference). However, for curve C there were intermediate differences in scores that were affected by the program material, primarily with program Sc-F, which scored lower than programs SD, TM and Sc. For curve D, higher preference scores were given for programs SD and Sc compared to the other programs, as shown by the intermediate to large effect size differences between scores. The preference scores for between curves D and E essentially "switched places" for programs Sc-T and Sc-F, showing that listeners preferred curve E over D for these two pieces of program. This is reflected in table 5.12 with the ***d*** scores changing from positive to negative between the two curves in the pairwise comparisons for these programs.

Upon further review of the relationships between target curve preference and program material, it was found that nearly all target curves showed effectively large differences between them with respect to program. However, certain curves scored somewhat similarly (intermediate effects size differences) with regard to program, as depicted in table 5.13.

	d	*CL*
SD - Curve		
A/B	0.50	63.80%
TM - Curve		
A/B	0.60	66.36%
Sc - Curves		
C/D	0.59	66.18%
Sc-T - Curves		
C/E	0.60	66.31%
D/E	-0.28	57.89%
Sc-F - Curves		
A/B	0.63	67.17%
C/E	0.48	63.37%
D/E	-0.40	61.15%

Table 5.13: Effect size scores between *target curves* with respect to *programs;* effect sizes between target curves highlighted in pink for intermediate differences

Curves A and B scored similarly on 4 out of the 5 programs while curve pairs C/E and D/E scored similarly for Sc-T and Sc-F; the two pieces of program material which were mixed on a SMPTE X-curve calibrated dub-stage. Of these pairing, the ***d*** scores show that curve C was preferred slightly over E where curve E was preferred marginally over D.

A factorial post hoc test was conducted with factors *target curve* and *program*, in which all of the target curves' preference scores were found to be significantly different with respect to program. However, the post hoc analysis for *program* also uncovered differences in preference scores influenced by the type of program, as revealed in tables 5.14(a) and 5.14(b).

Scheffe for Preference
Effect: Target Curve
Significance Level: 5 %

	Mean Diff.	Crit. Diff.	P-Value	
Curve A, Curve B	.442	.100	<.0001	S
Curve A, Curve C	-4.904	.100	<.0001	S
Curve A, Curve D	-4.115	.100	<.0001	S
Curve A, Curve E	-3.498	.100	<.0001	S
Curve B, Curve C	-5.346	.100	<.0001	S
Curve B, Curve D	-4.557	.100	<.0001	S
Curve B, Curve E	-3.939	.100	<.0001	S
Curve C, Curve D	.789	.100	<.0001	S
Curve C, Curve E	1.406	.100	<.0001	S
Curve D, Curve E	.618	.100	<.0001	S

Scheffe for Preference
Effect: Program
Significance Level: 5 %

	Mean Diff.	Crit. Diff.	P-Value	
SD, TM	.131	.100	.0028	S
SD, Sc	-.163	.100	<.0001	S
SD, Sc-T	-.133	.100	.0022	S
SD, Sc-F	.079	.100	.2131	
TM, Sc	-.294	.100	<.0001	S
TM, Sc-T	-.264	.100	<.0001	S
TM, Sc-F	-.052	.100	.6292	
Sc, Sc-T	.029	.100	.9361	
Sc, Sc-F	.241	.100	<.0001	S
Sc-T, Sc-F	.212	.100	<.0001	S

Table 5.14(a) & 14(b) Scheffe post hoc factorial analyses for *target curve* and *program*

5.4.6 Results Based on Cinema*Target Curve

Figure 5.18 shows the combined target curve preference data for all conditions in each cinema; noting that the 95% confidence intervals were too small to be depicted clearly and so +/- 1 standard deviation intervals were shown instead.

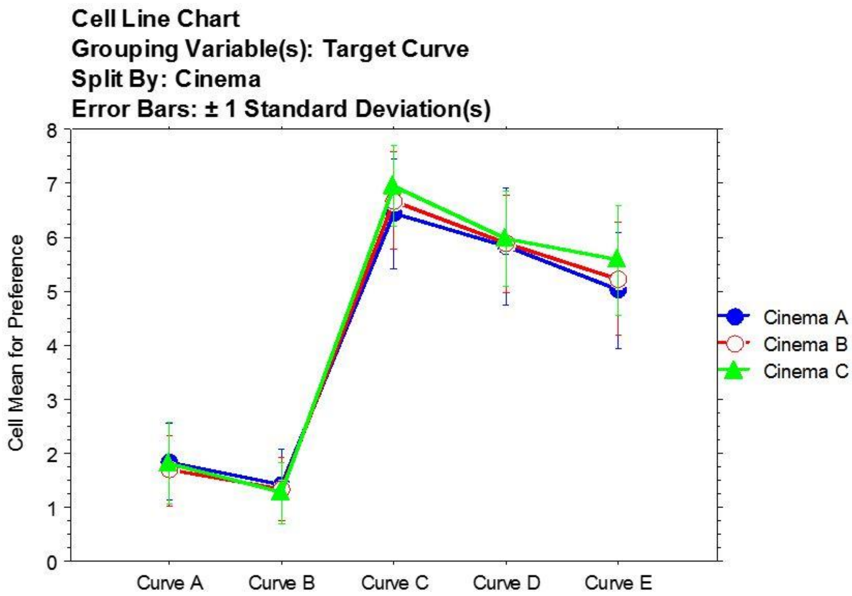


Figure 5.18: *Target curve* preference based on *cinema*

The ANOVA analysis found that the combined interaction between *cinema* and *target curve* produced significant results: $\boldsymbol{F}$ (8,104) = 7.873, $\boldsymbol{p}$ = .0001. This interaction was somewhat expected due to the large variance in cinemas size, especially between cinema A (60 seats) and cinema C (516 seats). From figure 5.18 it can be seen that the preference scores for three of the curves, A, B and D, differ very little between the three cinemas while curves C and E display moderate differences.

In reviewing effect size scores between *cinemas* for each *target curve*, there is either no effect size or only very small effect size differences between the cinemas for curves A, B and D. This indicates that the cinema which the target curve is played back in has little to no effect on the preference scores for these three curves. In comparison, an intermediate effect size exists between cinemas for curves C and E. The largest effect sizes for these two curves are between cinemas A and C, which are the smallest and largest cinemas, respectively. In addition, there is a small effect size difference between these two curves for cinemas B and C, as shown in table 5.15.

	d	*CL*
Curve A - Cinemas		
A/B	0.23	56.53%
A/C	0.04	51.26%
B/C	-0.18	54.99%
Curve B - Cinemas		
A/B	0.13	53.75%
A/C	0.24	56.67%
B/C	0.11	53.13%
Curve C - Cinemas		
A/B	-0.24	56.85%
A/C	-0.56	65.41%
B/C	-0.32	59.00%
Curve D - Cinemas		
A/B	-0.05	51.41%
A/C	-0.14	53.96%
B/C	-0.10	52.79%
Curve E - Cinemas		
A/B	-0.20	55.62%
A/C	-0.54	64.77%
B/C	-0.34	59.46%

Table 5.15: Effect size scores between *target curves* with respect to *cinemas*; effect sizes between cinemas highlighted in blue for small differences and pink for intermediate differences

Table 5.15 shows that curves C and E were preferred more in the larger cinema C than the smaller cinema A.

The data was also analyzed between curves in each cinema in order to review the affect that the cinemas had on pair-wise comparisons of curves (table 5.14). Curves A and B displayed an intermediate level of differences in both cinemas A and B where the same was found between curves C and D only in cinema A. Curves D and E showed an intermediate to small level of difference in all three cinemas, with the smallest difference displayed in cinema C.

	d	*CL*
Cinema A - Curves		
A/B	0.62	66.93%
A/C	-5.23	99.90%
A/D	-4.41	99.91%
A/E	-3.56	99.41%
B/C	-5.82	99.99%
B/D	-4.91	100.00%
B/E	-4.04	99.79%
C/D	0.58	65.82%
C/E	1.36	83.11%
D/E	0.75	70.30%
Cinema B - Curves		
A/B	0.56	65.41%
A/C	-6.29	100.00%
A/D	-5.28	99.90%
A/E	-4.07	99.80%
B/C	-7.00	100.00%
B/D	-5.94	100.00%
B/E	-4.61	99.94%
C/D	-0.87	73.16%
C/E	1.48	85.23%
D/E	0.67	68.11%
Cinema C - Curves		
A/B	0.80	71.46%
A/C	-6.78	100.00%
A/D	-5.02	99.98%
A/E	-4.21	99.85%
B/C	-8.46	100.00%
B/D	-6.27	100.00%
B/E	-5.23	99.90%
C/D	1.18	79.78%
C/E	1.53	86.03%
D/E	0.41	61.47%

Table 5.16: Effect size scores between *cinemas* with respect to *target curve;* effect sizes between target curves highlighted in blue for small difference and pink for intermediate differences

From table 5.16, it can be seen that for cinemas A and B, curve B is only preferred slightly more than the anchor curve A. In the largest of the cinemas (C), curves D and E scored very similarly.

A factorial post hoc test was conducted with factors *target curve* and *cinema*, in which cinemas A and B were found to be similar with respect to target curve scores but that all curves proved to be statistically different within each cinema, as shown in tables 5.17(a) and 5.17(b).

Scheffe for Preference
Effect: Cinema
Significance Level: 5 %

	Mean Diff.	Crit. Diff.	P-Value	
Cinema A, Cinema B	-.051	.065	.1612	
Cinema A, Cinema C	-.204	.065	<.0001	S
Cinema B, Cinema C	-.153	.065	<.0001	S

Scheffe for Preference
Effect: Target Curve
Significance Level: 5 %

	Mean Diff.	Crit. Diff.	P-Value	
Curve A, Curve B	.442	.106	<.0001	S
Curve A, Curve C	-4.904	.106	<.0001	S
Curve A, Curve D	-4.115	.106	<.0001	S
Curve A, Curve E	-3.498	.106	<.0001	S
Curve B, Curve C	-5.346	.106	<.0001	S
Curve B, Curve D	-4.557	.106	<.0001	S
Curve B, Curve E	-3.939	.106	<.0001	S
Curve C, Curve D	.789	.106	<.0001	S
Curve C, Curve E	1.406	.106	<.0001	S
Curve D, Curve E	.618	.106	<.0001	S

Table 5.17(a) & 17(b) Scheffe post hoc factorial analyses for *cinema* and *target curve*

5.4.6 Results Based on Cinema*Program

Figure 5.19 depicts the preference data as related to cinema and program material.

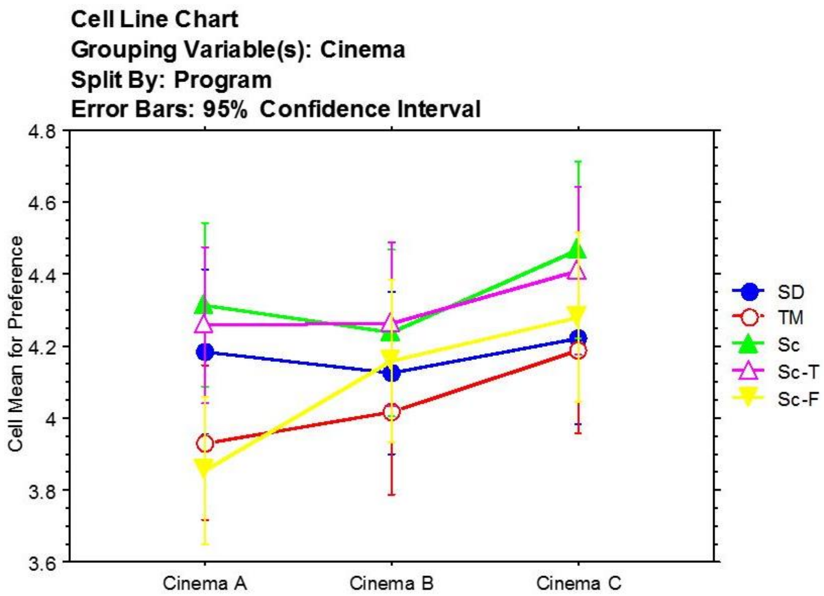


Figure 5.19: Preference scores as based on *cinema* and *program*

The ANOVA analysis found that the combined interaction between *cinema* and *program* produced significant results (F (8,104) = 5.608, p = .0001), which is primarily due to cinema A displaying a wider variance in scores between the 5 different pieces of program material when compared to cinemas B and C.

In reviewing effect size scores between the pieces of program material in a given cinema, no effect size exists between program materials in cinemas B and C. However, in cinema A it can be seen that small effects exist between three pairs of program material, as shown in table 5.18.

	d	*CL*
Cinema A - Program		
SD/TM	0.11	53.08%
SD/Sc	-0.06	51.56%
SD/Sc-T	-0.03	50.91%
SD/Sc-F	0.15	54.12%
TM/Sc	-0.17	54.68%
TM/Sc-T	-0.15	54.09%
TM/Sc-F	0.04	51.00%
Sc/Sc-T	0.02	50.68%
Sc/Sc-F	0.21	55.78%
Sc-T/Sc-F	0.18	55.18%
Cinema B - Program		
SD/TM	0.05	51.29%
SD/Sc	-0.05	51.33%
SD/Sc-T	-0.06	51.63%
SD/Sc-F	-0.01	50.40%
TM/Sc	-0.09	52.59%
TM/Sc-T	-0.10	52.89%
TM/Sc-F	-0.06	51.69%
Sc/Sc-T	-0.01	50.28%
Sc/Sc-F	0.03	50.94%
Sc-T/Sc-F	0.04	51.23%
Cinema C - Program		
SD/TM	0.01	50.40%
SD/Sc	-0.10	52.76%
SD/Sc-T	-0.08	52.17%
SD/Sc-F	-0.02	50.69%
TM/Sc	-0.11	53.19%
TM/Sc-T	-0.09	52.61%
TM/Sc-F	-0.04	51.10%
Sc/Sc-T	0.02	50.64%
Sc/Sc-F	0.07	52.08%
Sc-T/Sc-F	0.05	51.48%

Table 5.18: Effect size scores between *programs* with respect to *cinema*; small effect sizes for cinema A between various programs highlighted in blue

The data was then analyzed to observe effect sizes between cinemas for a given piece of program material; shown in figure 5.19.

	d	*CL*
SD - Cinemas		
A/B	0.03	50.71%
A/C	-0.02	50.43%
B/C	-0.04	51.13%
TM - Cinemas		
A/B	-0.04	51.04%
A/C	-0.11	53.12%
B/C	-0.07	52.03%
Sc - Cinemas		
A/B	0.03	50.92%
A/C	-0.06	51.73%
B/C	-0.09	52.59%
Sc-T - Cinemas		
A/B	-0.00	50.02%
A/C	-0.14	53.96%
B/C	-0.10	52.79%
Sc-F - Cinemas		
A/B	-0.14	53.83%
A/C	-0.19	55.24%
B/C	-0.05	51.45%

Table 5.19: Effect size scores between *cinemas* with respect to *program*; small effect size for program Sc-F between cinemas A and C highlighted in blue

From this data it can seen that there is a small difference in preference for program Sc-F between cinemas A and C, with the preference weighing slightly more towards the larger cinema.

A factorial post hoc test was conducted with factors *cinema* and *program*, in which cinemas A and C were found to be statistically different with respect to program. In addition, the post hoc analyses found that all programs were found to be statistically similar within each cinema, with the exception of programs TM and Sc, as shown in tables 5.20(a) and 5.20(b). However, the difference between these two programs was marginal, as $\boldsymbol{p}$ =.0457 with a significance level of $\boldsymbol{\alpha}$ =.05.

Scheffe for Preference
Effect: Cinema
Significance Level: 5 %

	Mean Diff.	Crit. Diff.	P-Value	
Cinema A, Cinema B	-.051	.179	.7854	
Cinema A, Cinema C	-.204	.179	.0205	S
Cinema B, Cinema C	-.153	.179	.1116	

Scheffe for Preference
Effect: Program
Significance Level: 5 %

	Mean Diff.	Crit. Diff.	P-Value	
SD, TM	.131	.290	.7487	
SD, Sc	-.163	.290	.5605	
SD, Sc-T	-.133	.290	.7353	
SD, Sc-F	.079	.290	.9520	
TM, Sc	-.294	.290	.0457	S
TM, Sc-T	-.264	.290	.0968	
TM, Sc-F	-.052	.290	.9892	
Sc, Sc-T	.029	.290	.9988	
Sc, Sc-F	.241	.290	.1613	
Sc-T, Sc-F	.212	.290	.2817	

Table 5.20(a) & 20(b) Scheffe post hoc factorial analyses for *cinema* and *program*

5.4.8 Results Based on Cinema*Listening Position

Though *listening position* was not found to be significant in the ANOVA analysis when considered as a stand-alone factor, there was an interaction between *cinema***listening position* that was found to be significant, albeit with a low ***F*** value of ***F*** (4,52) = 2.744 and a marginal ***p*** value of ***p*** =.038. Figure 5.20 displays preference data as related to cinema and listening position.

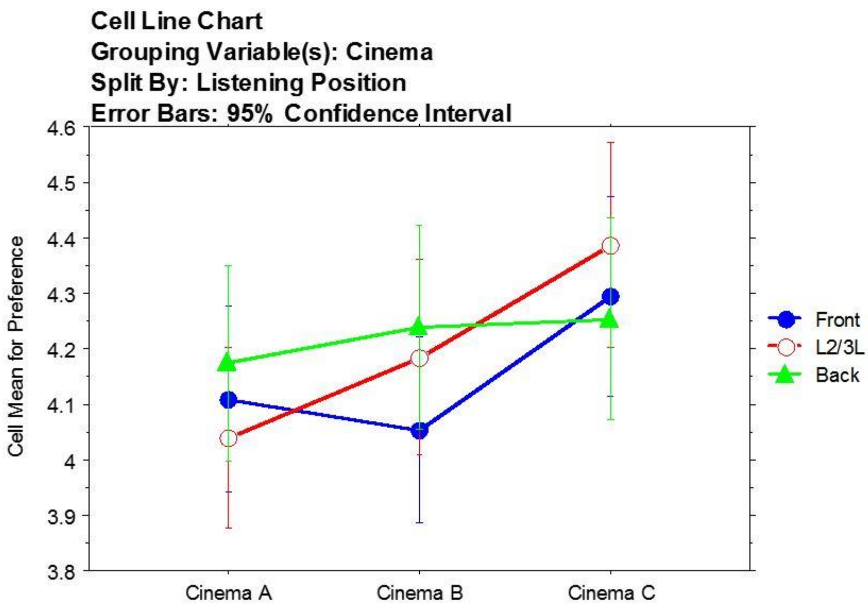


Figure 5.20: Preference scores based on *cinema* versus *program*

The 95% confidence intervals shown in figure 5.20 display a high degree of overlap for all listening position and cinemas with the exception of the 2/3L position between cinema A and C.

In reviewing effect size scores between the cinemas for a given listening position, there was only a small effect between cinemas A and C for the 2/3L position as shown in table 5.21.

	d	*CL*
Front - Cinemas		
A/B	0.02	50.70%
A/C	0.08	52.20%
B/C	0.10	52.88%
2/3L - Cinemas		
A/B	0.06	51.77%
A/C	0.15	54.18%
B/C	0.08	52.35%
Back - Cinemas		
A/B	0.03	50.76%
A/C	0.03	50.94%
B/C	0.01	50.17%

Table 5.21: Effect size scores between *cinemas* with respect to *listening position*; small effect size for position 2/3L between cinemas A and C highlighted in blue

Effect size scores in table 5.22 depict that within cinemas there is no difference in preferences between listening positions.

	d	*CL*
Cinema A - Position		
Front / 2/3L	0.03	50.89%
Front / Back	0.03	50.78%
2/3L / Back	0.06	51.66%
Cinema B - Position		
Front / 2/3L	0.06	51.58%
Front / Back	0.08	52.21%
2/3L / Back	0.02	50.64%
Cinema C - Position		
Front / 2/3L	0.04	51.07%
Front / Back	0.02	50.46%
2/3L / Back	0.05	51.52%

Table 5.22: Effect size scores between *listening positions* with respect to *cinema*

These results are reflected in the factorial post hoc analysis (tables 5.23 a & b) between *cinema* and *listening position*.

Scheffe for Preference
Effect: Cinema
Significance Level: 5 %

	Mean Diff.	Crit. Diff.	P-Value	
Cinema A, Cinema B	-.051	.179	.7856	
Cinema A, Cinema C	-.204	.179	.0206	S
Cinema B, Cinema C	-.153	.179	.1120	

Scheffe for Preference
Effect: Listening Position
Significance Level: 5 %

	Mean Diff.	Crit. Diff.	P-Value
Front, L2/3L	-.051	.179	.7817
Front, Back	-.070	.179	.6340
L2/3L, Back	-.018	.179	.9685

Table 5.23(a) & 23(b) Scheffe post hoc factorial analyses for *cinema* and *listening position*

5.5 Discussion

The results presented in the previous sections show a preference for target curve C when considering factors *cinema*, *program*, *listening position*, and *subject*, analyzed collectively and individually. Curves A and B scored low across all tests regardless of cinema, program, listening position, subject or combination thereof.

The factor *program* did have an effect on preference scores between curves D and E, as shown in figure 5.16. Further analysis demonstrates that for programs Sc-T (Score-Trailer) and Sc-F (Score-Film), listeners tended to exhibit a higher level of preference towards the flat target curve (E), than the curve with the most noticeable bass boost (D), especially in cinema C, the largest of the three cinemas (figure 5.21).

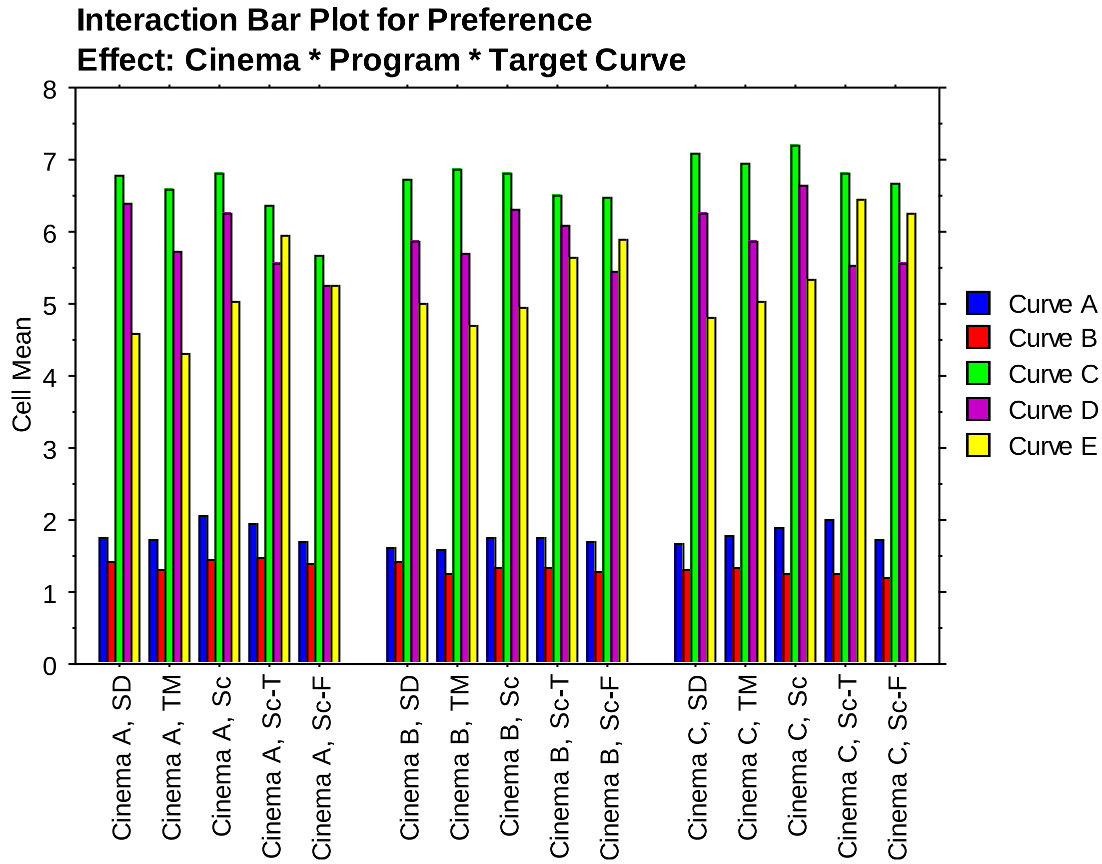


Figure 5.21: *Target curve* preference based on *program* and *cinema*

This is possibly due to these programs having more low frequency content than the others; referring to the long-term spectral power density of the programs shown in figure 5.12. Further to this point, the program Sc-T has elevated lower frequency levels over the film version of the same content (Sc-F). Unlike, feature-length films, trailers must meet a standard level of loudness in order for the trailer to receive a film rating from the Motion Picture Association of America's (MPSS) Classification and Rating Administration (CARA). This arose out of the increasing levels of trailer audio in what had become a "loudness battle" between the movie studios who were attempting to garner additional market share (louder perceived as better). To answer this problem, the Trailer Audio Standards Association was formed and created what has become known as the TASA standard. This standard attempts to maintain trailer program levels within the parameter of $L_{eq}(m)$ = 85dB, using a metering system based on various weightings at octave frequency bands. In order for a trailer to receive an MPAA rating, it must meet the TASA standard level. This standard is based on the ITU-486 noise level for sound broadcasting, both of which are shown in figure 5.22.

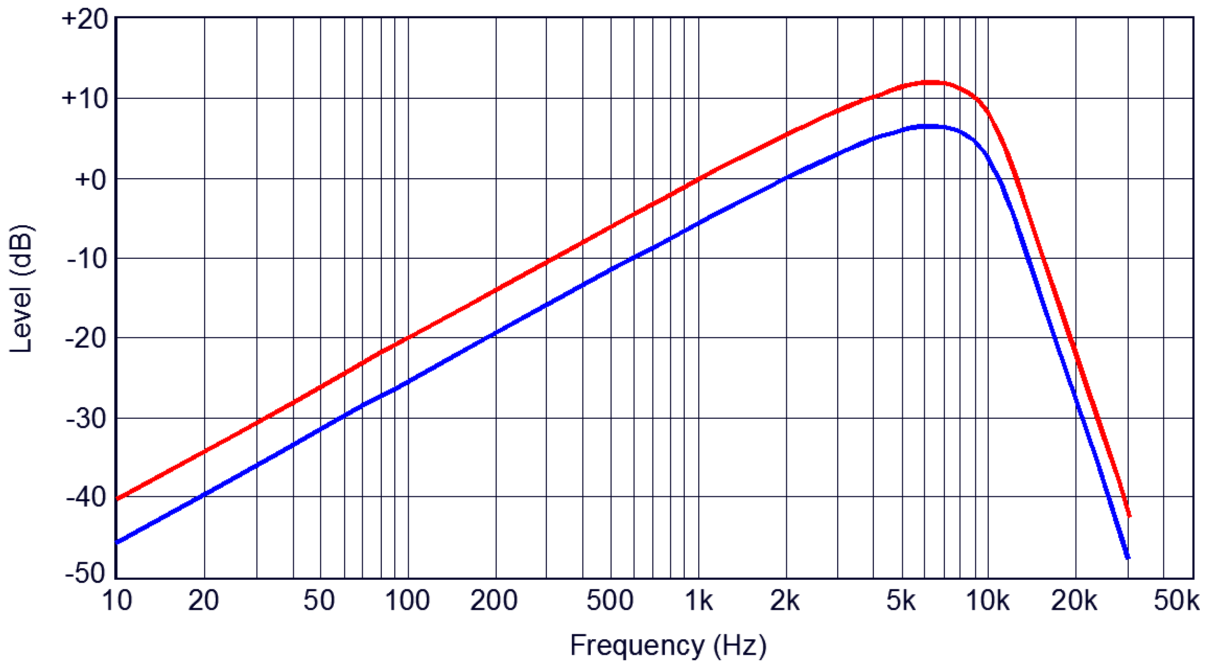


Figure 5.22: Trailer frequency response weighting curve, $L_{eq}(m)$, shown in blue with ITU-486 shown in red; derived from (TASA, 2013) and (ITU, 1987)

$L_{eq}(m)$ is an integrated measurement of level over the duration of the trailer, and it's weighting emphasize mid and upper frequencies while rolling off the effect of the bass. Essentially this allows sound editors and mixers to balance loud passages against quiet passages and to add a large amount of low frequency information while still meeting the TASA standard. This typically leads many trailers to have louder low frequency information than their associated feature-length films.

It was expected that programs Sc-T and Sc-F would have had a larger affect when interacting with curve A. One of the basic premises for the SMPTE X-curve is to provide good translation between the dub-stage and the cinema. However, the results found that curve A was not the preferred curve for program material mixed on a SMPTE calibrated dub-stage. This implies that ideal translation was not occurring during these experiments. If the theory of translation is valid for all circumstances of programs mixed on a SMPTE calibrated dub-stage, then curve A should have been the preferred curve for programs Sc-T and Sc-F.

The factor *cinema* does appear to have a small effect, with curves C and E rating slightly higher in cinema C than in cinema A; referring to figure 5.17. This may be partially due to cinema C being much larger than cinema A, with longer reverberation times, especially in the lower frequencies (figure 5.23).

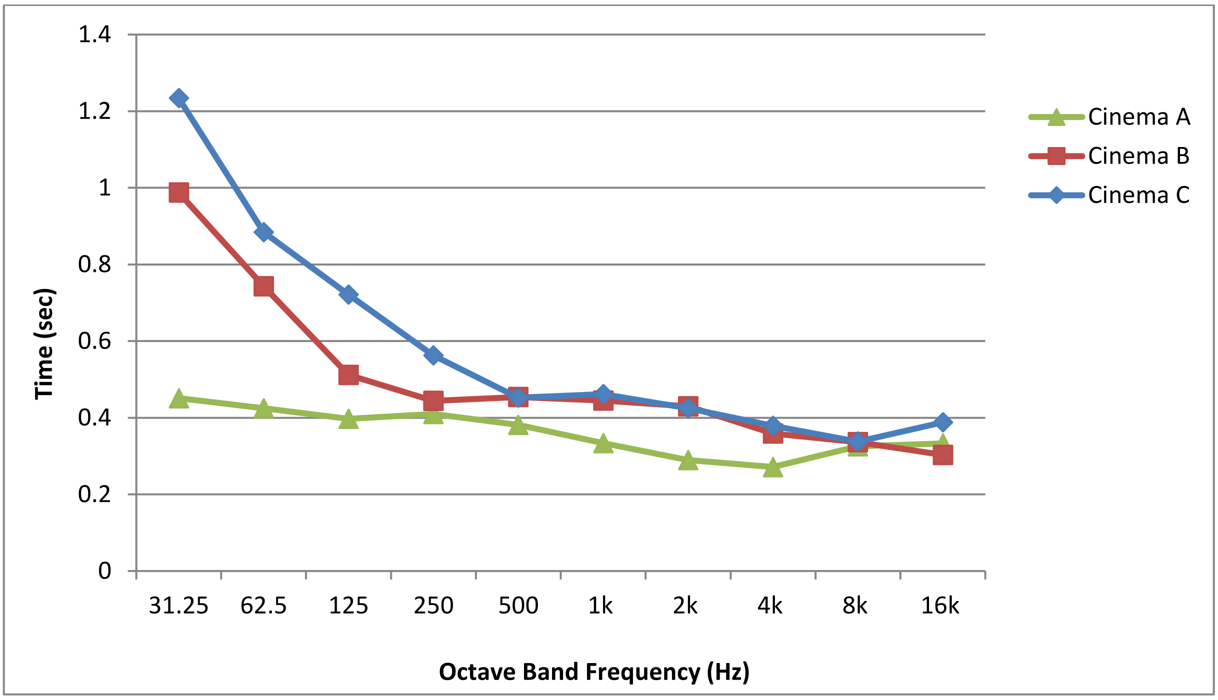


Figure 5.23: Reverberation time by octave for cinemas A, B and C

This may have led test subjects to perceive cinema C as being more "theatrical" sounding, rendering flat or "flattish" curves with somewhat enhanced low frequency playback more preferred for this size cinema.

The listening position of the test subjects within any of the cinemas did not have a significant effect on preference. However, in comparing the test results between cinemas, ranking of the preference mean did change from cinema to cinema (figure 5.19) although not in a significant manner. The only noted difference that *listening position* had on preference was that the 2/3L position was preferred more in cinema A than C, but this finding was so marginal that it did not warrant further analysis. These results were somewhat unexpected in light of the differences in high frequency air loss between cinemas, especially when comparing the frequency response of the back rows of cinemas A (row 6) and C (row 17), as seen in figure 5.24.

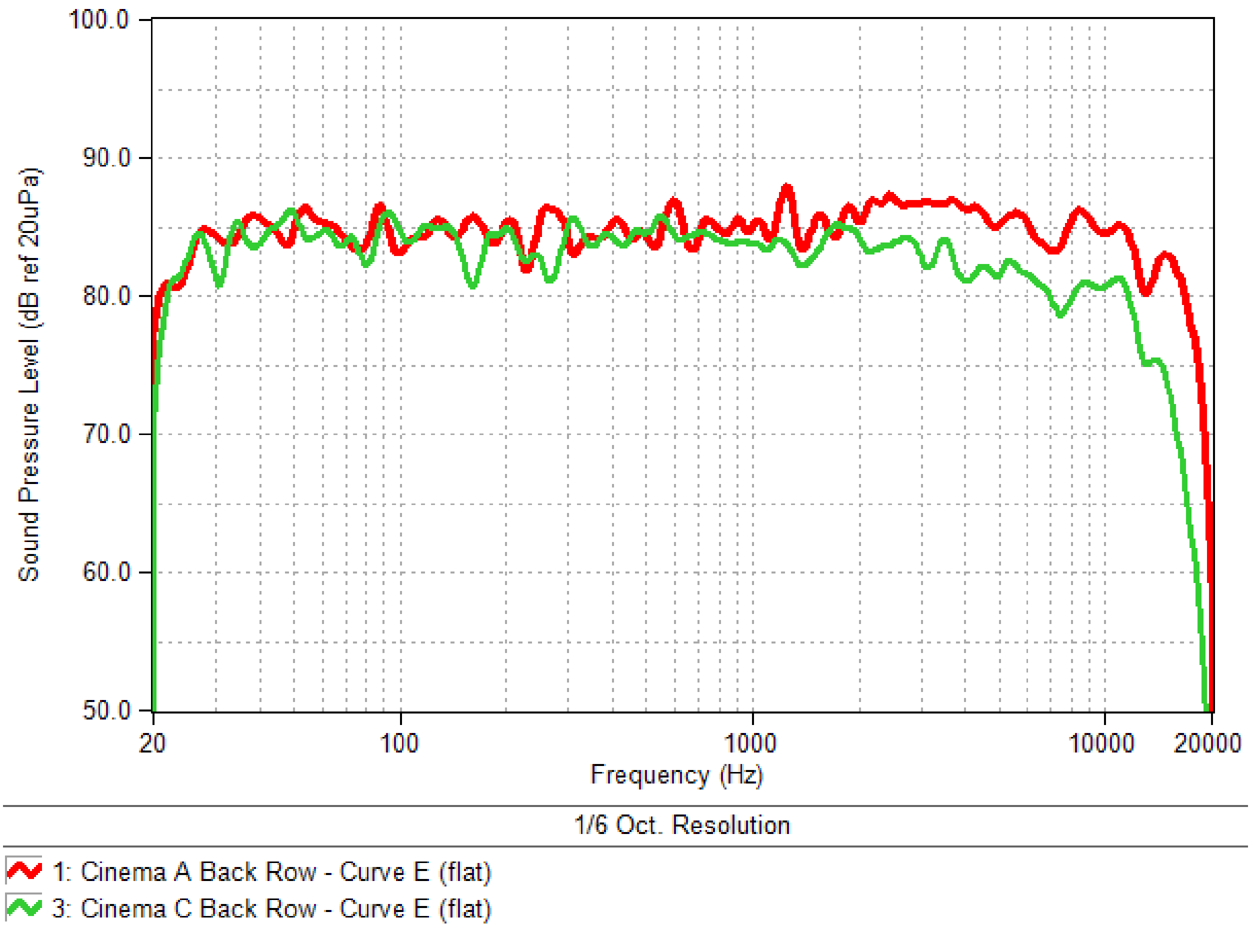


Figure 5.24: Frequency response for curve E (flat) for the back row in cinemas A and C

A possible explanation for these finding may be listener adaptation over the successive trials. The listening test described in this chapter were administered in a successive treatment condition, as described by Keppel and Wicken (2004), where a variable (or set of variables) is held constant throughout a experimental session. In this case, the room and listening position were held constant throughout a single session and only changed between subsequent listening sessions. All other aspects of the listening tests, including trials and session orders were randomized in a balanced way across listeners. Previous studies, where successive treatments of room were used, have shown that the successive treatment condition can produce room acoustic adaptation due to the longer exposure time to the same room acoustics (Olive et al., 1995).

An effect was noted between listeners, shown in figure 5.25, as two of the listeners had higher preferences for curve E over curve D and one listener scored these two curves very similarly overall. These three listeners all fall within the median age of the listeners and have previous experience in taking listening tests; therefore, there is nothing notable to explain this difference.

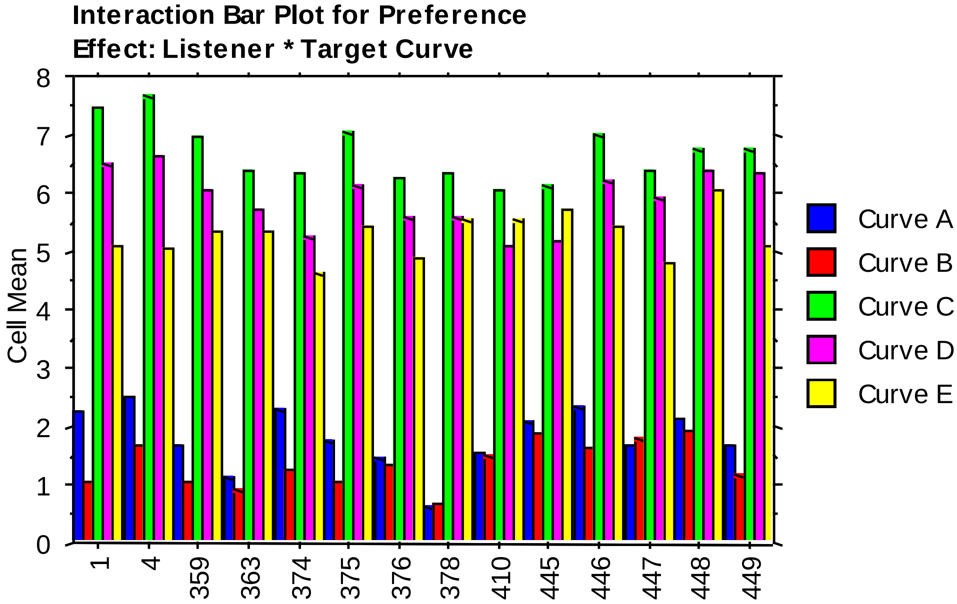


Figure 5.25: *Target curve* preference based on *subject (listener)*

5.6 Preference Curves in Prior Research

Prior research has shown preferences similar to the finding reported in this chapter. In various circumstances, listeners have preferred a curve that displays an elevated bass response with a nearly flat transitional "tilt" into a lowered high frequency response.

Even in the early days of cinema tuning it was found that a flat response was not desirable but instead, a tilted response curve gave a "more pleasing response" as noted by Durst and Shortt (1939) and shown in figure 5.26.

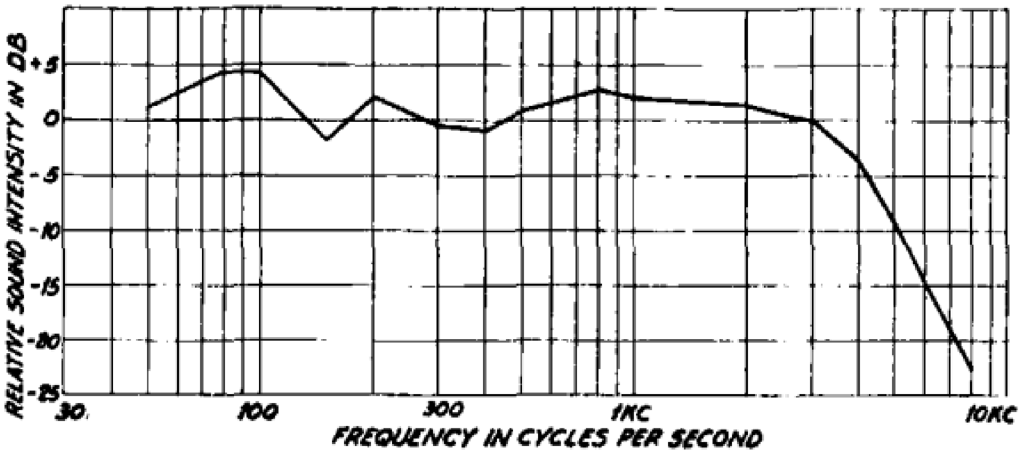


Figure 5.26: Overall acoustic characteristic adjusted for "more pleasing response" from (Durst and Shortt, 1939)

Note that the extreme high frequency roll-off in figure 5.26 was to compensate for noise from the recording methods and equipment and was not derived solely by preference (other than the preference not to hear noise).

In their work on equalizing rooms to a "house curve", Boner and Boner also found that listeners preferred linear, sloping curves that rose towards to the lower frequencies, as shown in figure 5.27.

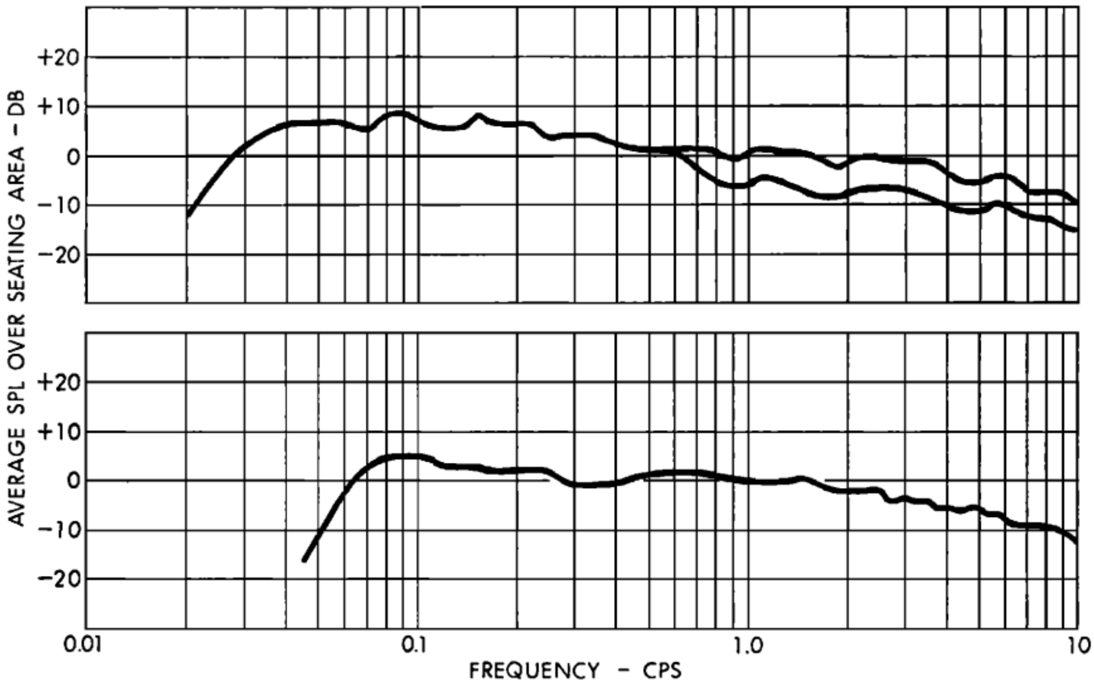


Figure 5.27: Typical house curves after equalization and filtering of major feedback modes, (Boner and Boner, 1965)

The Boners further found that listeners were more willing to accept a flat horizontal curve with some roll-off from 1kHz to 10kHz, but that a curve which tilted up towards the higher frequencies or rolled off lower frequencies was not acceptable.

In 1969 Ljungberg undertook a series of listening tests which involved using sound engineers as test subjects; the same sound engineers who were responsible for the recordings used during the listening tests. This experiment was as part of his ongoing work towards a cinema and control room standard. What Ljungberg had found was a set of curves which he mapped against an empirical curve tolerance that was flat from 30Hz to 250Hz and then followed a 1.5dB/octave roll off with a +/-2 tolerance, as shown in figure 5.28 (Ljungberg, 1969).

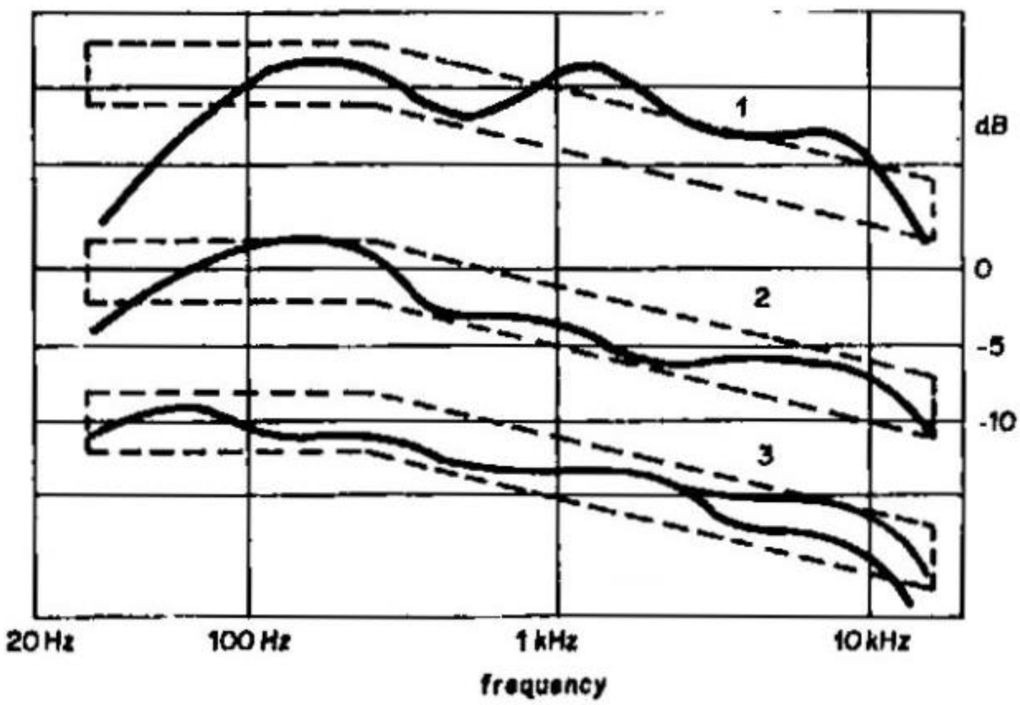


Figure 5.28: Results of listening tests conducted with high-quality dialog and music program material in a re-recording theatre of 300m^3 with good diffusion and a 0.5s reverberation time, flat; from (Ljungberg, 1969)

The comments received by Ljungberg were as follows:

Curve 1: Too thin on music, a little nasal and shrill on dialog.

Curve 2: Well-balanced on all sorts of program, possible a trifle too heavy in bass.

Curve 3: Very well balanced on all types of program, both treble characters [sic] quite acceptable.

Møller noted that an ideal curve for listening to "hi-fi" equipment in a listening room is one that is as smooth and straight as possible, signifying that all frequencies are reproduced at approximately equal level with a primary emphasis being given to the 60Hz to 6kHz range. (Møller, 1974). Møller further reported that recordings made of a combination of near-field and far-field information should be reproduced using a curve with a boost in the low frequencies and a small amount of cut in the high frequencies. The curve shown in figure 5.29 was partially based on listening tests conducted by Brüel & Kjaer and partially from curves taken from average concert halls.

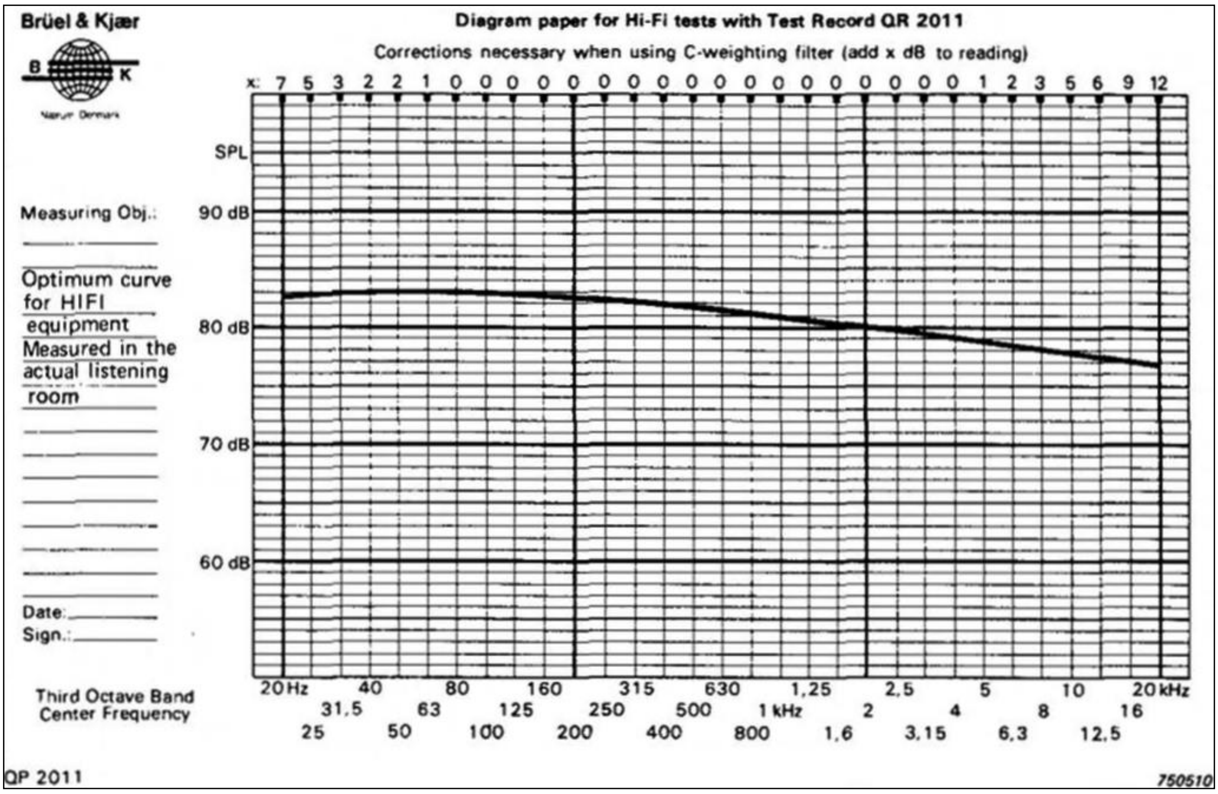


Figure 5.29: Optimum curve for hifi equipment measured in the actual listening room; from (Møller, 1974)

Holman revealed a similar sloping curve while studying 9 commercial cinemas, all presumably calibrated using the SMPTE standard (2007). The average of the data is shown in figure 5.30.

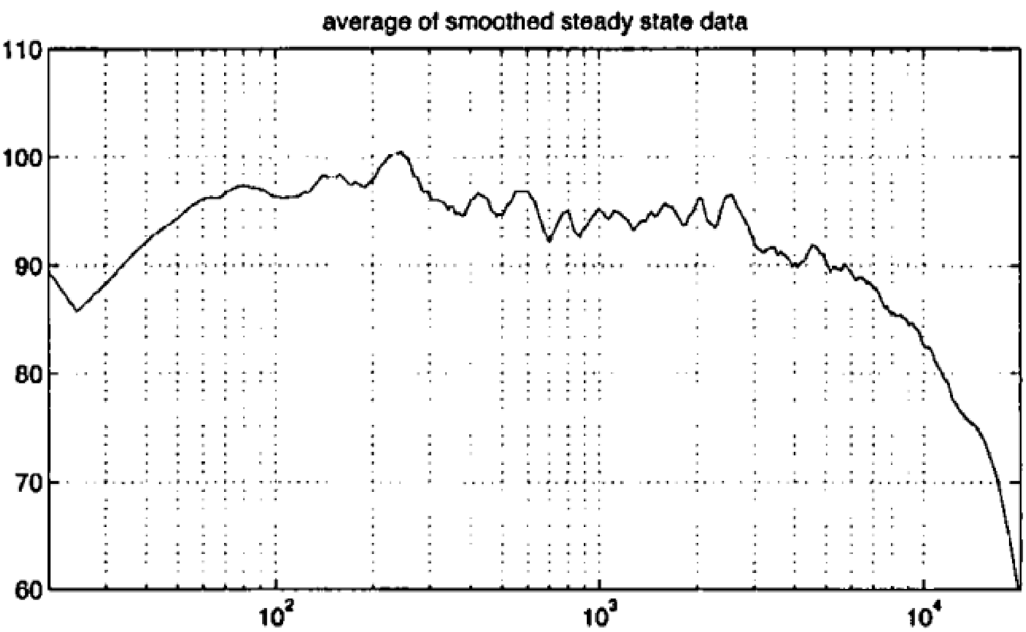


Figure 5.30: Steady state (160ms record) average frequency response; from (Holman, 2007)

Research on five room correction products utilized for non-cinema listening spaces (i.e. small to mid-sized studios and living rooms) have shown that the preferred room correction product had shown a similar smooth in-room target response with a 10dB negative slope from 20Hz to 20kHz, while the least prefer product had specified a flat in-room target curve as shown in figure 5.31 (Olive et al., 2009).

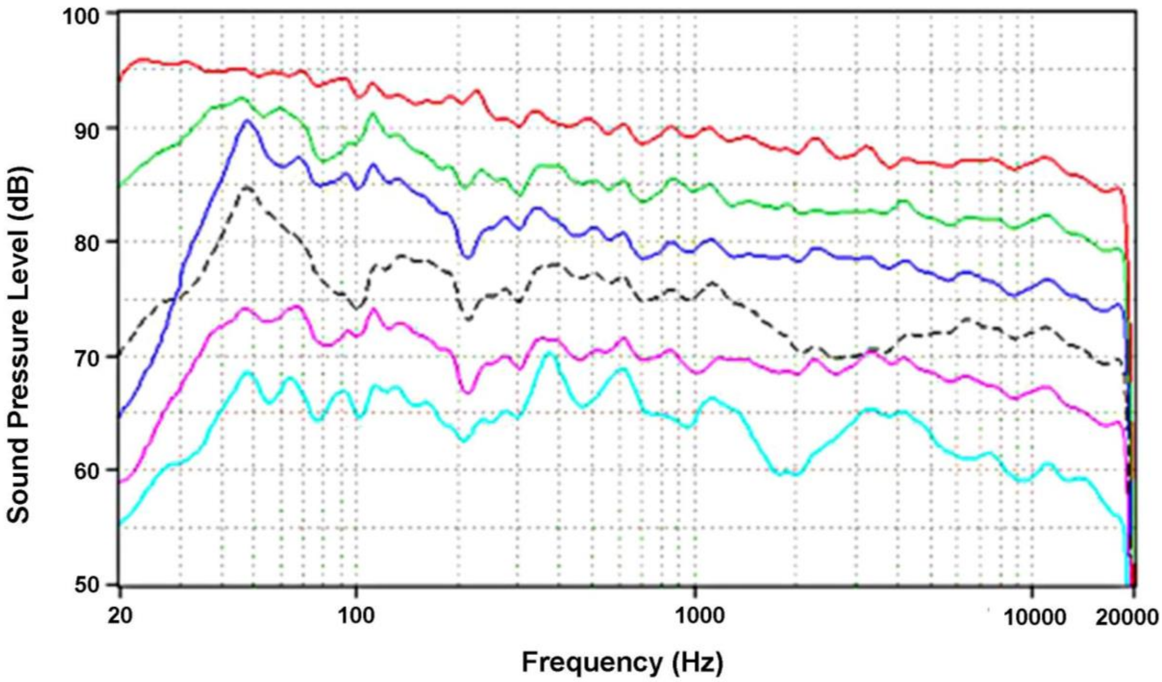


Figure 5.31: In-room frequency response measurements for room corrections products, shown in descending order based on preference ratings and offset by 5dB for clarity; from (Olive et al., 2009)

Further research by Olive et al. found that the preferred in-room and headphone target response curves were not flat but also had an equivalent sloping shape and were extremely similar in characteristics (Olive and Welti, 2015, Olive et al., 2013a, Olive et al., 2013b), as shown in figure 5.32.

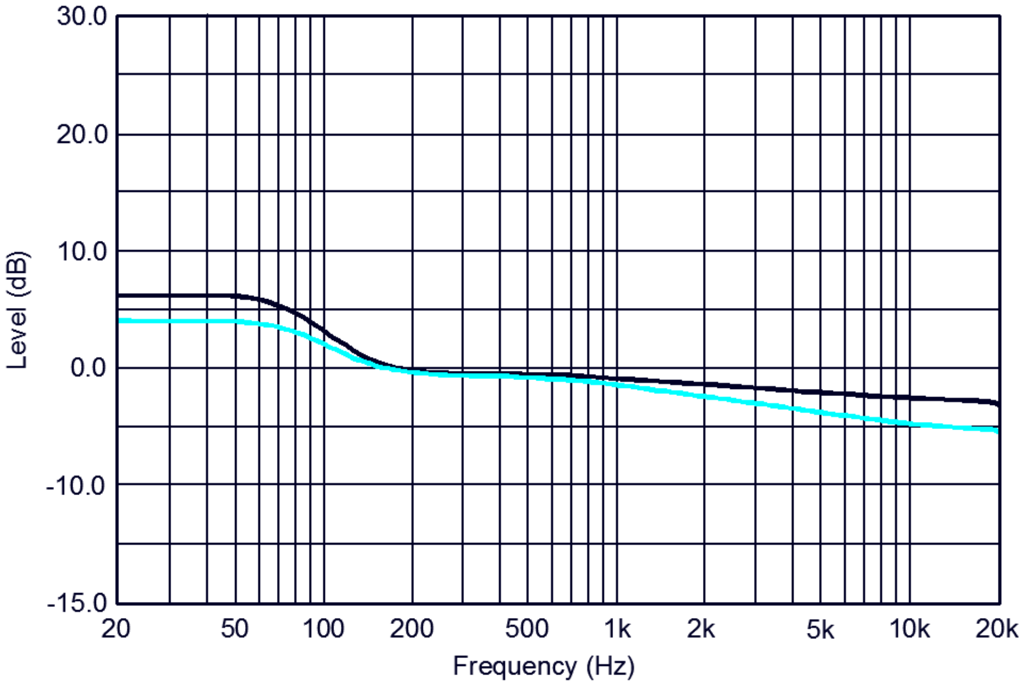


Figure 5.32: Preferred in-room loudspeaker (black) and headphone (cyan) target response curves, derived from (Olive and Welti, 2015, Olive et al., 2013a, Olive et al., 2013b)

Based on research by Olive et al, Dennis considered if comparable curves would be found in tests for automotive audio. Using the same method-of-adjustment test as Olive and his team, Dennis had 8 expert and 95 non-expert listeners adjust bass and treble filters until they achieved their preferred frequency response (Dennis, 2015). The results of this test are shown in figure 5.33.

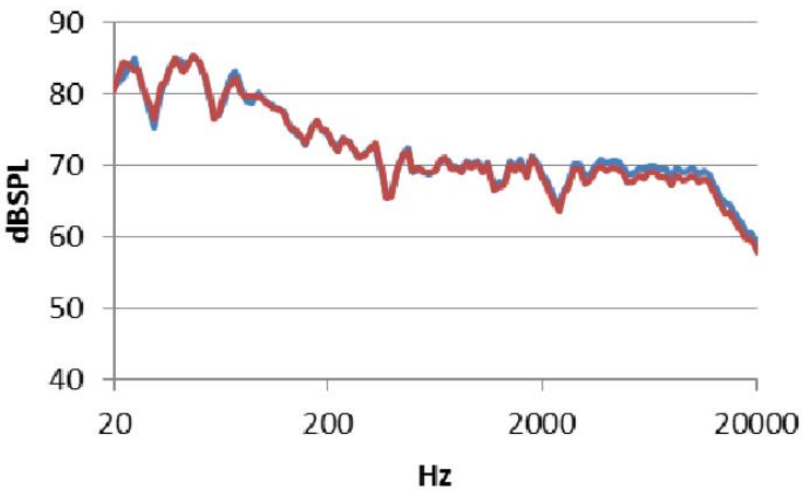


Figure 5.33: Bass and treble adjusted independently (blue), bass and treble adjusted together (red); from (Dennis, 2015)

These tests rendered a much higher preferred bass response than the in-room or headphone test, but a similar sloping shaped preference curve was found.

Research involving the variability in how cinemas are being calibrated uncovered not only a wide range of calibrated responses but also a trend towards calibration curves that roughly follows the same sloping response as found in prior research (figure 5.34).

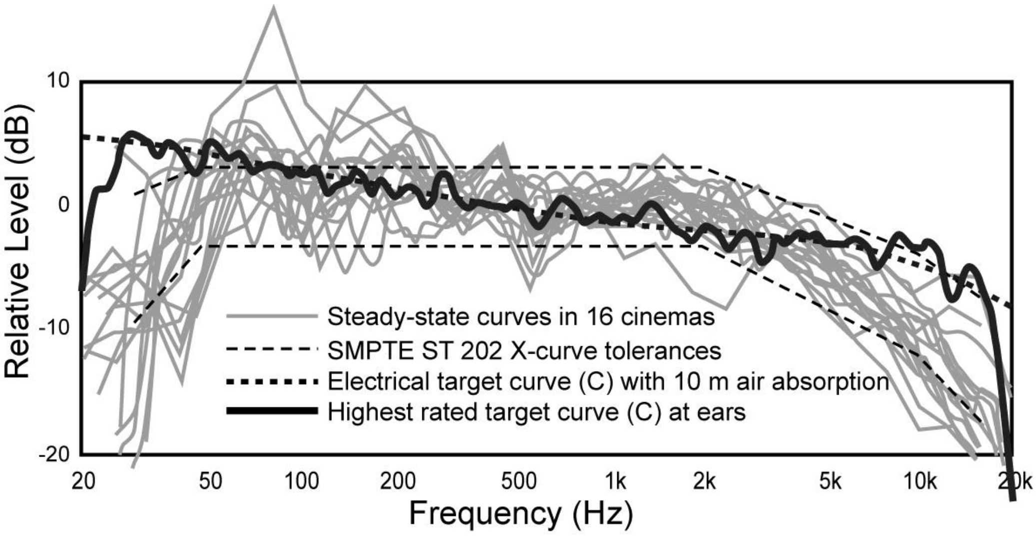


Figure 5.34: Comparison of SMPTE calibrated cinemas responses and curve C in cinema Ca the 2/3L listening position; the cinema data are from Figures 10, 16(a) and 17 in (Toole, 2015); from (Toole, 2016)

Recent research by the SMPTE B-chain study group found similar responses in 4 cinemas (figure 5.35) and 2 dub-stages (figure 5.36). The following graphs show the averaged measured response for the centre channel loudspeaker taken at the reference (2/3L) position, smoothed over 1/3rd octave bandwidth with 48 PPO window with the responses normalized before averaging (SMPTE, 2014).

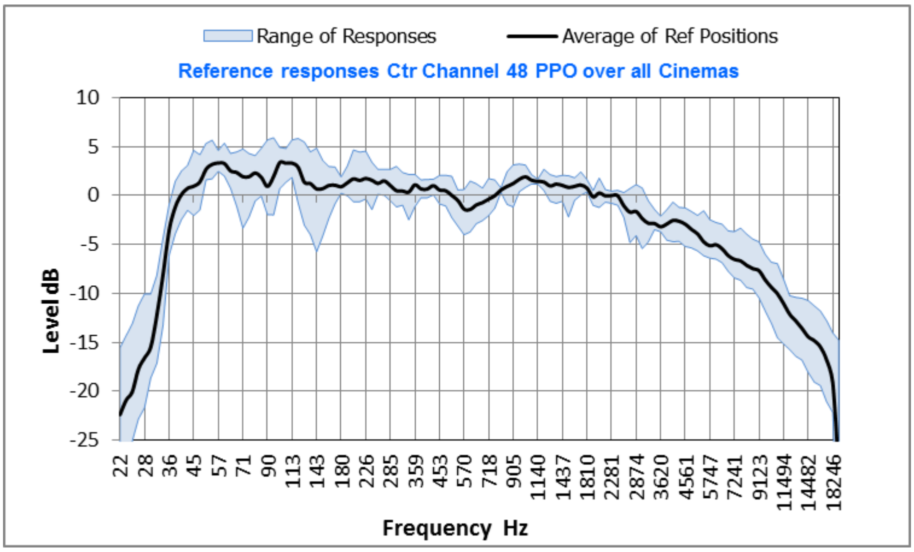


Figure 5.35: Overall average and range of responses for 4 cinemas; from (SMPTE, 2014)

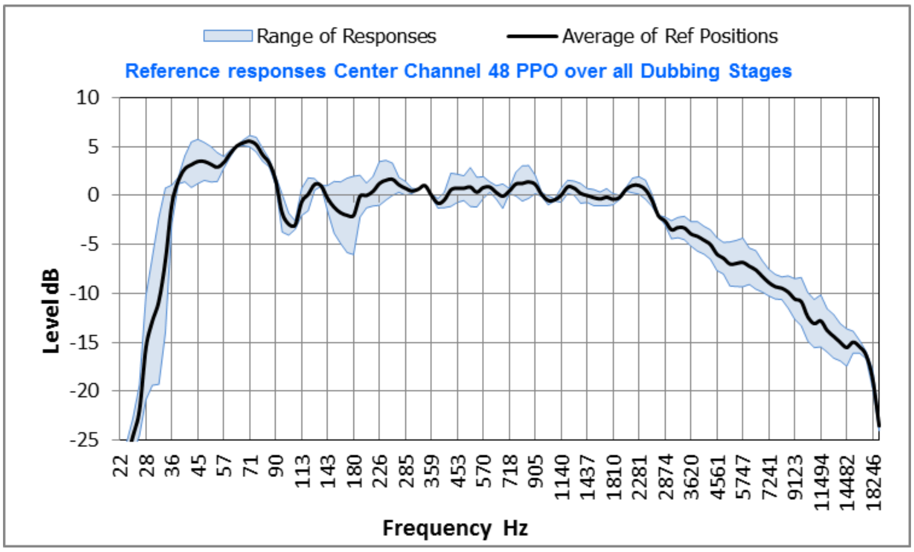


Figure 5.36: Overall average and range of responses for 2 dub-stages; from (SMPTE, 2014)

The report release by the B-chain study group also noted, when compared to the X-curve and its tolerances, the alignments of the real-world examples typically fell within a few dB with the exception of the low frequencies. The low frequencies were found to be, on average, 5dB above the X-curve.

It is interesting to note that for these cinemas and dub-stages shown in figures 5.34 through 5.36, at low frequencies, the technicians tended to tune the room towards a similar curve as that of curve C when attempting to calibrate to the SMPTE X-curve. The calibrations drift towards the upper limit of the tolerance at the lower frequencies while drifting to the lower limit of the tolerance at the highest frequencies. Given that the final tuning in cinemas and dub-stages is typically done by ear and at the discretion of the technician, it appears that in the case of the above given references, these technicians preferred a tilted curve with more low frequency content.

5.7 Limitations of This Study

When conducting subjective listening tests, it is important to remember that there is unlikely to be a perfect experiment that will give conclusive, yes/no answers. Inevitably, variables other than those being tested or the lack of certain variables will influence results. While this study was conducted in working cinemas, certain aspects of the cinemas were not employed, such as utilizing the installed loudspeakers as a test source, placing the test loudspeaker behind the projection screen and not utilizing a multi-channel playback system. The decision to design the experiment as described was an effort to control variables. However, the affect that these variables might have on listener preference cannot be overlooked. With this in mind, the use of the "house" loudspeaker system, placing the loudspeakers behind the screen and utilizing a multi-channel system all provide paths for future research.

5.8 Summary

This chapter reports the results of a series of listening test for preferred in-room response target curves for cinemas. A BRS system was used to capture the BRIR in three different cinemas at three different seating locations. The three cinemas used for these tests included a 60-seat, 161-seat and 516-seat cinema. BRIR were convolved with 5 different types of programs and filtered through 5 different target curves.

These experiments investigated not only target response curve preference but also any potential influence that the cinema, listening position or program material may have on that preference. The main conclusions of these experiments are as follows:

1. The main effect was due to the factor *target curve*.
2. The overall preferred target curve was curve C.

3. Listeners rated curve A extremely low, even for the program material mixed on a dub-stage calibrated to the SMPTE X-curve. This calls into question the validity of dub-stage to cinema translation.

4. There was an interaction between the different cinemas and target curves, though this mainly appears to occur with curves C and E in cinema C.

5. There was an interaction between the different programs and target curves. This interaction appears to be confined to programs Sc-T and Sc-F and is most apparent in cinema C.

6. There was no evidence of listeners being sensitive to the varying amounts of high frequency air attenuation in seats at the front, 2/3L and rear locations within the venues. This is important because there were no visual cues as to their locations which would allow contextual influences.

Chapter 6 - Predicting the In-room Response of Cinemas from Anechoic Loudspeaker Data

6.1 Introduction

In recent years there has been a good deal of research undertaken regarding the in-room responses of cinemas, most notably on how they are measured and calibrated. Much of this research has been centred on measurement procedures such as the number of microphone positions to be spatially averaged, the measurement equipment itself and the assessment of the resultant room curve.

There has been a small collection of research published in recent years that addresses the number of microphone positions to use for cinema calibration. Much of this research does not look to differentiate between what part of the measured response is the room and what part is the loudspeaker (Cengarle and Mateos, 2014) while some has looked at using close-field data taken from the loudspeaker in situ (Newell et al., 2011a). Noting that the high Q and non-minimum-phase nature of loudspeaker-room interactions at any given listener position creates issues for thorough equalization, some researcher have sought out a solution through de-reverberation based subtractive equalization (Fielder, 2001). In Fielder's work he describes a method of breaking down the process into three steps: separating the "undesirable" transfer function components, de-reverberation of the undesirable components and then adding a new "desired" component. The effort requires the ability to create an inversion filter that cancels the combined loudspeaker-room transfer function leaving only the loudspeaker without any room effect. Though this process shows promise over the traditional loudspeaker-room equalization methods, Fielder notes the difficulties and limitations of this method.

The research in this chapter attempts to find a compromise between the questionable practice of equalization based on combined loudspeaker-room measurements in absence of loudspeaker data, and the difficult subtractive equalization method based on de-reverberation. It is proposed that the calibration (equalization) of a cinema be based on measurements that place increased focus on the anechoic data of the loudspeaker being utilized and the quality of that loudspeaker. In this research, a well-defined loudspeaker which displays constant directivity was utilized and the playback system was held constant in each cinema so that the interaction between only the loudspeaker and the room could be observed.

Work in this chapter has been peer reviewed and previously presented at the following:

GEDEMER, L. 2015. Predicting the In-room Response of Cinemas from Anechoic Loudspeaker Data. AES 57th International Conference, Hollywood, CA: Audio Engineering Society. (Gedemer, 2015a)

6.2 Motivation

Current calibration work being completely on cinema audio system centres around the 5 microphone positions as stated in the current SMPTE standard ST-202-2010 (SMPTE, 2010) and forwarded by the most recently published SMPTE report (SMPTE, 2014); refer to figure 6.1.

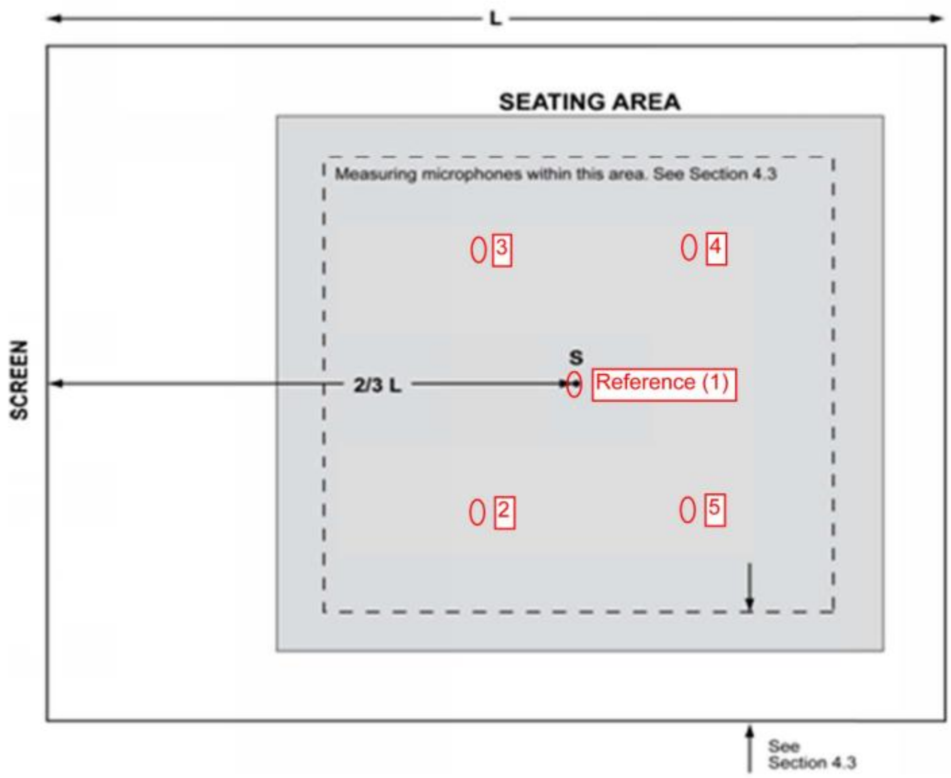


Figure 6.1: Microphone layout, commonly utilized positions shown in red; from (SMPTE, 2014)

The hypothesis behind the research described in this chapter is that the number of microphones required to calibrate a cinema is that number which, when spatially averaged, will trend towards the anechoic data of the loudspeakers used in the cinema. A secondary hypothesis is that a loudspeaker with smooth off-axis frequency response will allow fewer microphones to be needed to capture the in-room response. In an acoustically well controlled room equipped with loudspeakers that have uniform directivity, such as a well designed cinema, screening room or dub-stage, the room curve may be predicted in the lower frequencies from the anechoic total sound power measurement while at the higher frequencies it may be predicted by the anechoic on-axis data (Toole, 2015, 2012), (Engebretson and Eargle, 1982).

The following questions serve as a basis for the research in this chapter:

- How well can a cinema's steady-state, in-room amplitude response be predicted from anechoic loudspeaker data?
- How much does room size affect the in-room response of the loudspeaker and how?
- How well does the in-room response "track" the anechoic data and how frequency dependant is this?
- How many microphones positions should be averaged in order to produce an acceptable deviation from the loudspeaker data?
- Does randomizing the heights of the measurement microphones effect the data measured and how?
- How does the number of microphones which are averaged together effect the data measured?

The research presented in this chapter utilized a loudspeaker that was characterized by a comprehensive set of anechoic measurements and then measured in a number of different sized cinemas. In this way, the acoustical effects of the room and its interaction with the loudspeaker could be separated and identified from the effects produced by the loudspeaker alone. To the author's knowledge, this is the first such experiment on cinema sound that has done so.

Earlier studies have pointed out the risks of assessing the in-room response of a cinema sound system given the insufficient number of measurement positions (between 1 and 5) required by the standard and the wide variability of equalizing the room by technicians (SMPTE, 2014), (Newell et al., 2012), (Newell et al., 2011a). These few positions cannot provide any meaningful separation between the performance of the sound source and the effects of the room itself, especially in absence of the anechoic data on the loudspeakers being utilized. This could lead technicians to possibly equalize for a room effect in one seat only to make the performance inferior in another. Measured curves can be ambiguous because they include evidence of acoustical interference which can be strictly location dependant. One of the more pronounced effects of this type is comb filtering, which is most apparent in a frequency response graph at middle and high frequencies when the delayed signals are acoustically summed with that of the direct sound of the loudspeaker at a single microphone. Though comb filtering can be quite audible, in certain circumstances it can be inaudible or perceived as a form of 'spaciousness' (Brunner et al., 2007) and, in these cases, may be consider somewhat benign. Spatial averaging attempts to attenuate the visual evidence of comb filtering. However, the amount of comb filtering that is attenuated by spatial averaging would be partially dependent on the number of microphones used, their spacings, and the frequency range in question.

It has been shown that high Q resonances in a loudspeaker's response can be audible. In most cases the effect may be small to moderate. Medium to low Q resonances by contrast are very audible and should be addressed (Toole and Olive, 1988), (Toole, 1986a). Additional phenomena which are not resonances, such as directivity mismatch or crossover effects are also audible. Minimum phase anomalies such as resonances can be equalized, however, non-minimum phase issues such as crossover interference will not benefit from equalization. Without the necessary loudspeaker data, the ability to tell the difference using solely in-room measurements would be difficult.

Technicians may also be tempted to equalized acoustic phenomena which are perceptually unimportant and/or change with head movement, as research has shown that listeners are capable of perceptually sorting out the direct and reflected sounds in a way that minimizes the timbral colouration (Lindau, 2014), (Watkins, 2005). This is counterintuitive to what the in-room frequency response measurement made with a microphone at the listener’s ear position would seem to indicate. Effectively, listeners are able to learn and adapt to the invariant acoustic features in rooms, such as the reflection patterns, so that they can better hear the true features of the sound source. However, it would be impossible for a person examining response curves from a small set of stationary microphones spatially averaged together to discern which effects are perceptible or even correctable. Attempts to equalize these irregularities can, in effect, add resonances to the playback system, thus degrading the sound quality from what may have been an ideal or, at the very least, a neutral sound source (Newell et al., 2011b). Without sufficient information on the acoustic performance of the loudspeaker, it is difficult for a cinema technician to know the difference between anomalies in the frequency response caused by room effects and those caused by shortcomings of the loudspeaker.

6.3 Methodology

The use of spatial averaging utilizing a given number of microphone positions within a cinema or dubbing stage to determine an in-room response is a commonly practiced method. In this chapter, such a method was employed at several cinema-like venues of various seating capacities. Professional screening rooms located within Los Angeles, California were utilized for this purpose. The spatial averages of these rooms were then compared to the anechoic data of the loudspeaker used.

6.3.1 FFT Time Window

Early in the experiment it was determined that the measurement data could be time windowed to eliminate reflections though this could lead to issues regarding elements of the sound that the listener hears being eliminated from the measured signal. The direct sound and some early reflections contain most of the information in a cinema while reverberation tends to be a lesser factor as modern cinemas naturally have low reverberation times.

It is known that longer FFT windows render more accurate data at lower frequencies as a result of improved frequency resolution, as given be equation 6.1:

$$\boldsymbol{df} = 1/\boldsymbol{T} \qquad \text{Eq. 6.1}$$

$\boldsymbol{df}$ is the frequency resolution

$\boldsymbol{T}$ is the duration of the window.

To observe the difference in window lengths first-hand, a sample measurement was reviewed using various time windows, 10ms, 50ms and 500ms along with two different window types, Rectangular and Tukey. The 10ms window was chosen as it shows almost solely the first arrival and very early reflections but has a resolution of only 100Hz. The 50ms window was chosen as it is a rough approximation of the precedence effect interval for speech (Blauert, 1997) and displays a frequency resolution of 20Hz, equivalent to 1/3 octave at 60Hz. The 500ms window was chosen as it is long enough to include the reverberation time of the room and closely approximated a steady state response in cinemas. The 500ms window duration displays a frequency resolution of 2Hz. The two window shapes used, rectangular and Tukey, provide for good frequency resolution at the trade-off of spectral leakage, though rectangular displays more spectral leakage than Tukey. Rectangular is considered ideal where spectral analysis is the main requirement and the signal duration is shorter than the length of the window. Tukey provides for good spectral analysis with lower spectral leakage than rectangular (Benjamin, 2004), (Dactron, 2003). However, rectangular is preferred when the signal is broadband in nature, when the frequencies in question are closely spaced and of equal amplitude, or when accuracy in low frequency measurements are important (Benjamin, 2004), (Cerna and Harvey, 2000). In comparison, the Hanning or similar shaped window attenuates the direct field and very early time information with shorter window lengths as the tapering commences at the start time of the window. This could lead to a loss of frequency resolution when evaluating transient signals and also displays errors at low frequencies (Poularikas, 1998), (Harris, 1978). For these reasons, data was processed for comparison using the rectangular window for the longer window of 500ms and Tukey for the shorter 50ms and 10ms windows, as shown in figure 6.2.

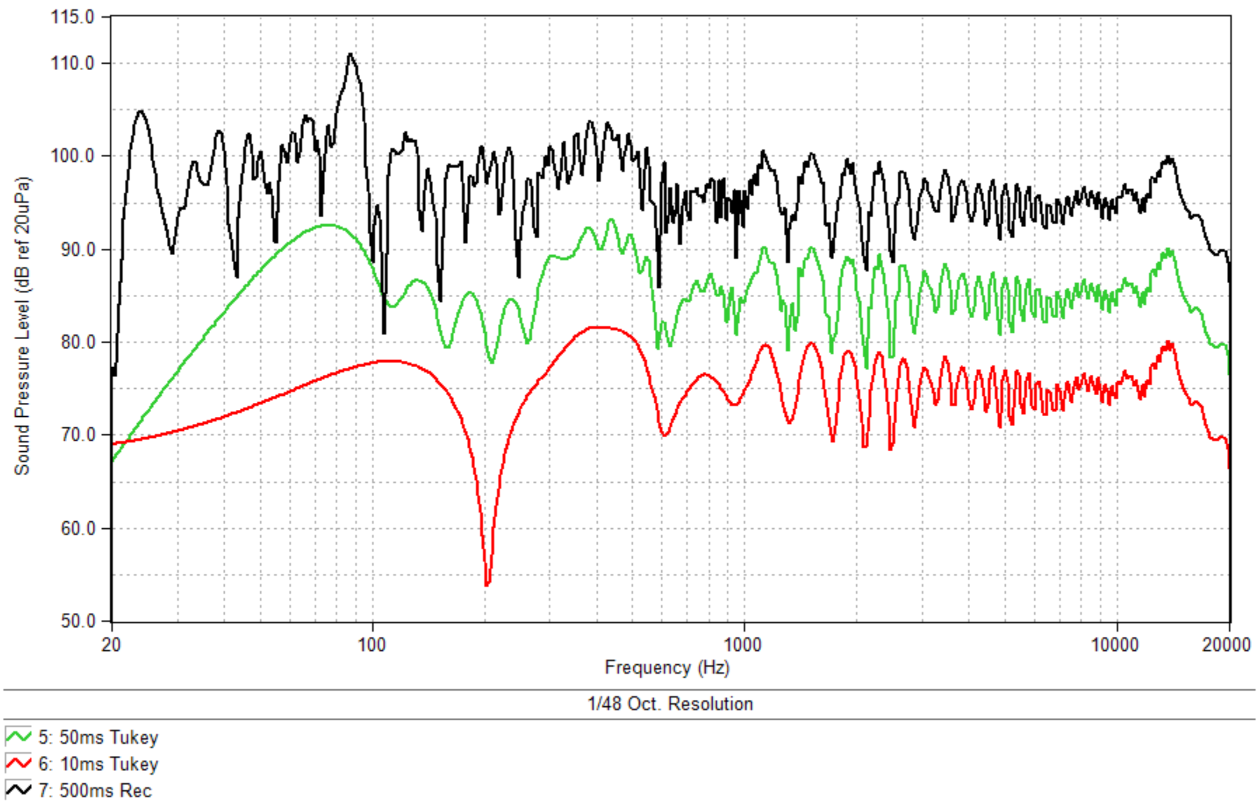


Figure 6.2: A single microphone position at 2/3L; 500ms Rectangular, 50ms and 10ms Tukey windows (shown offset by 10dB)

Figure 6.2 shows a clear difference in window lengths' resolution from 20Hz to 3kHz. However, above 3kHz, very little difference is seen between the two window types and durations when examined with 1/48 octave resolution.

It has been observed that above approximately 500Hz, when considering window lengths between 10ms and 2000ms, the resolution of the data does not change and that the direct sound is indeed dominant (SMPTE, 2014), (Fielder, 2012), (Newell et al., 2010), (Munro, 2004).

Ultimately a 500ms rectangular window was chosen in order to obtain good frequency resolution across the entire bandwidth and to provide a reasonable representation of a steady state response such as used by the SMPTE standard.

6.3.2 Description of the rooms under research

A total of seven different sized screening rooms were selected in order to represent the varying sizes of professional cinemas, screening rooms and dub-stages. The author felt that the inclusion of the smaller, 24-seat room was viable given the trend towards building smaller dub-stages and screening rooms. This body of research was conducted in the cinemas shown in table 6.1 and pictures in appendix N.

Cinema	**Seat Count**	**Number of Microphone Positions**	**Floor Area (in meters)**	**Approximate Volume (in meters^3)**	**RT_{60}**
A	24	24	11L x 5.3W	232	0.27
B	30	30	12.2L x 5.5W	335	0.29
C	60	30	12.8L x 8W	N/A	0.35
D	114	59	18.3L x 9.5W	1043	0.40
E	161	79	17L x 14W	1708	0.44
F	211	103	22.3L x 10.7W	1258	0.41
G	516	109	46L x 20W	N/A	0.46

Table 6.1: Research cinemas utilized; RT_{60} is an average of the 500 Hz and 1 kHz bands

As shown in table 6.1, the measured reverberation time (RT_{60}) in all of these rooms was between .27s and .46s. The rooms' reverberation times were measured using ISO 3382-1 recommendations (ISO, 2009) with the JBL M2 speaker placed centred in front of the screen aimed at the 2/3L position and measured by the BRS system described in chapter 4. The analysis of the impulse responses was undertaken using Matlab ([RT,DRR,CTE,CFS,EDT] = irstats(filename,varargin)); the code is shown in appendix O (Hummersone, 2013). The RT_{60} measurements of the larger cinemas E, F and G, fall within the guidelines provided in table 6.2.

Volume [m^3]	1,000	5,000	10,000	20,000	50,000
RT_{max}	0.55	0.7	1.0	1.3	1.4
RT_{min}	0.35	0.45	0.7	0.9	1.0

Table 6.2: Recommended reverberation times at 500Hz for various theatre sizes; from (Dolby, 1994)

The reverberation times per octave are shown in figure 6.3.

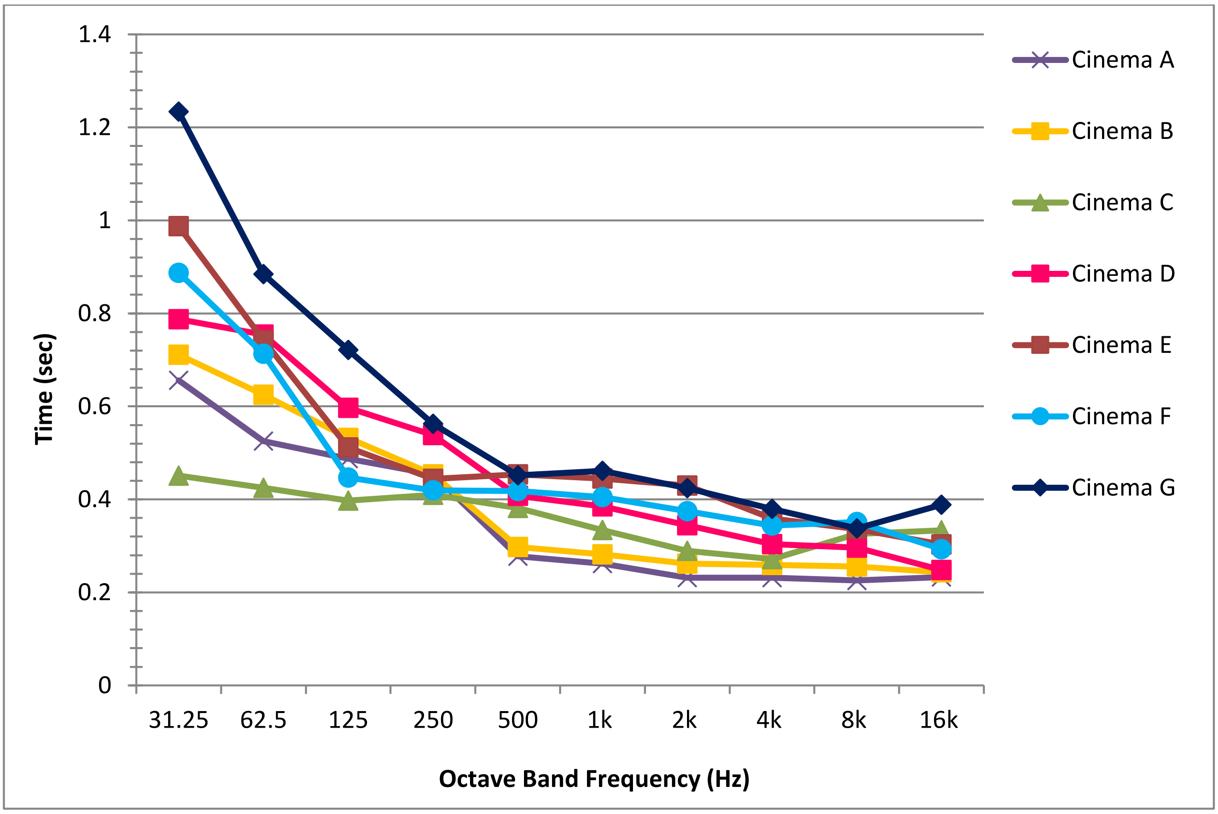


Figure 6.3: Reverberation times by octave for all cinemas

It was found that 3 of the 7 cinemas under testing (D, E and F) fell within the 80 to 300 seat range, which is typical of a multiplex cinema as depicted in figure 6.4.

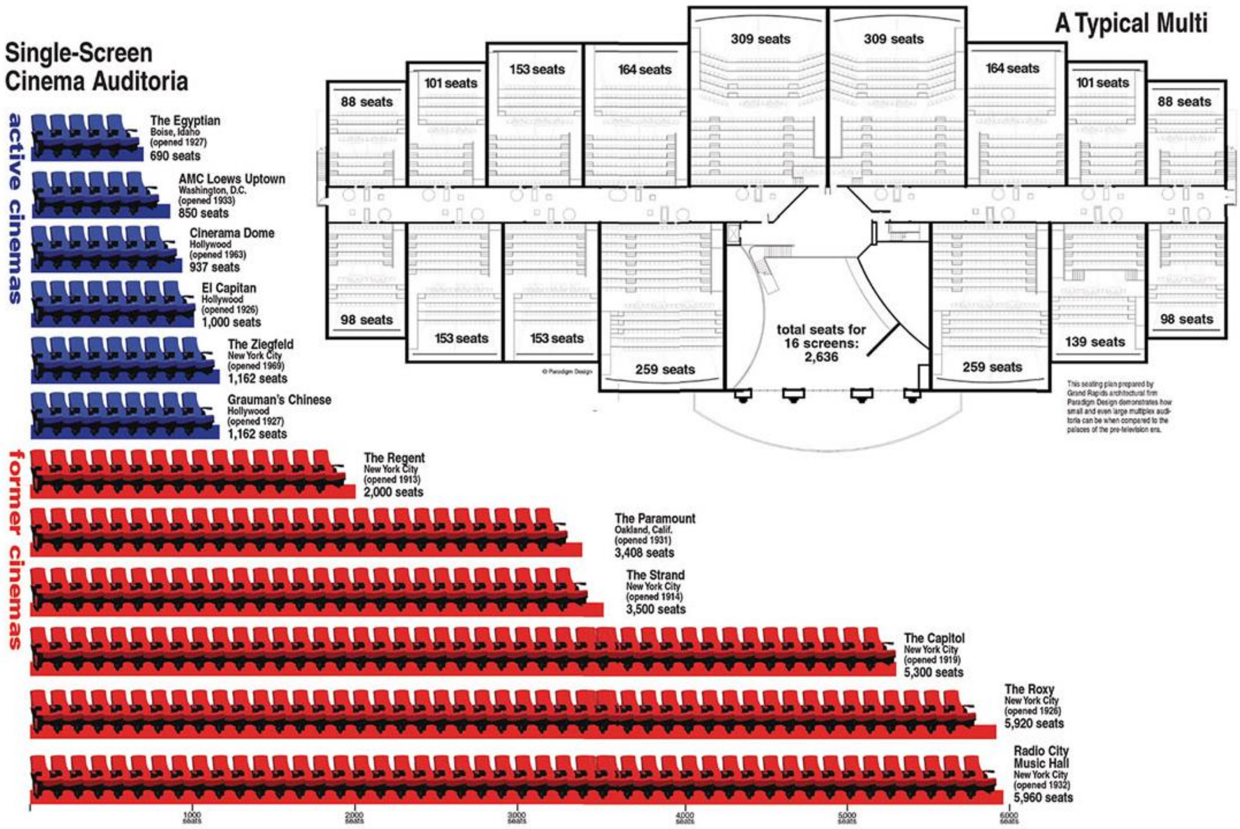


Figure 6.4: Typical multiplex theatre sizes; from (NATO, 2013)

6.3.3 The playback system

The JBL M2 was chosen as the sound source for these measurements because of its exemplary anechoic performance and power handling capabilities required to produce a sufficient sound pressure level in the various sized cinemas. In addition, the M2 displays constant directivity. Figure 6.5 shows its frequency response measured in a calibrated anechoic chamber at a distance of 2 meters. The four curves from top to bottom represent its on-axis, listening window (+/-30° horizontal, +/-10° vertical), sound power shown as a normalized amplitude response curve, and the directivity index (Devantier, 2002). The importance of constant directivity is the relative shape of the frequency response remains unchanged as one moves off-axis. Sound power hypothetically represents all of the sounds arriving at the listening position after any number of reflections from any direction. The total sound power is determined by measuring the loudspeaker at equal angular increments, in the case of the JBL M2 this was 5°, and then applying an appropriate weighting value at each angular increment. The weighting value corresponds to the area of the spherical quadrangle centred at the microphone position for a particular angular position (ANSI/CEA, 2015). The JBL M2 was measured in 283m^3 chamber, which is fitted with 1.2m fibreglass wedges and dual steel acoustic doors. The chamber is anechoic, +/- 0.5dB at 1/24th octave, from 60Hz to above 20kHz and measure to be +/-0.5dB at 1/12th octave from 60Hz to 20Hz.

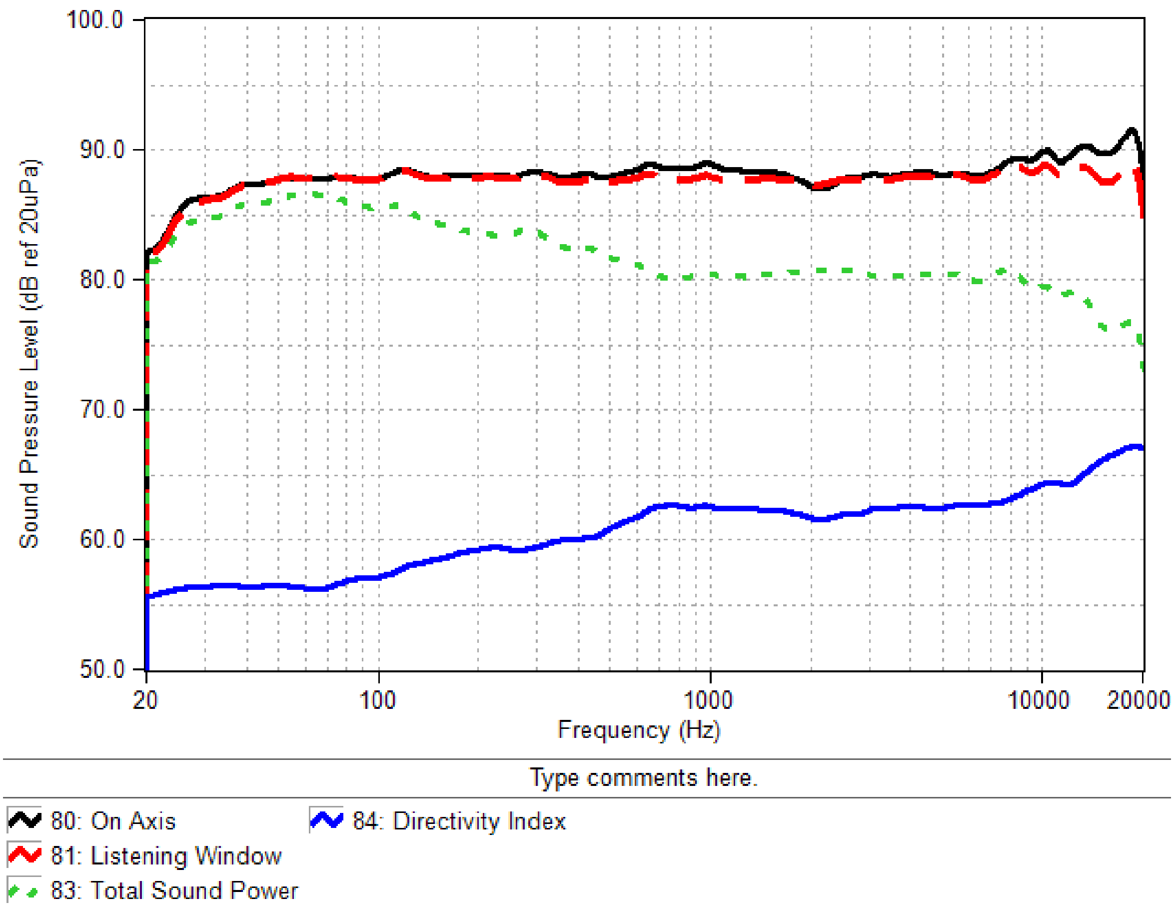


Figure 6.5: Anechoic data of JBL M2 loudspeaker combined with the Crown iTech 9000 amplifier. On-axis (solid black), listening window (dashed red), sound power (dotted green) and directivity index (solid blue)

The only processing utilized within the playback system was the crossover and equalization required by the JBL M2 loudspeaker, which is available as a DSP pre-set in the Crown iTech 9000 amplifier.

The JBL M2 loudspeaker was placed at centre screen position in each cinema approximately 0.3 meter in front of the screen. It was raised on a lift and aimed at the 2/3L position in the seating area using a Bosch GLL 2-45 self-levelling long-range cross line laser. There are no published industry standards for the mounting height of screen loudspeakers but the general guideline that many cinemas adhere to is orienting the screen loudspeakers so that the reference axis of the loudspeaker is aimed at a location on the centreline of the main seating area at a distance about 2/3L the length of the house (JBL, 2003).

The loudspeaker was placed in front of the screen instead of behind it in order to minimize the effects of the screen and the loudspeaker mounting conditions behind the screen. By removing the screen and mounting effect from the measurements, efforts were able to focus on how the cinema acoustics interact with the loudspeaker. Research on the acoustic properties of different projection screen materials along with variations in loudspeaker mounting behind a screen, and their combined, wide variable effects on loudspeaker performance has been well researched and documented by others (Newell et al., 2013), (Long et al., 2012). It has been duly noted that a projection screen can have a large influence on the in-room response of a loudspeaker; an influence which should not be ignored during room measurement and calibration. However, the decision to avoid placement of the loudspeaker behind that screen was an attempt to limit the number of variables which would have been present with 7 different screens. In addition, access to placement of the test loudspeaker behind the screen in all cinemas was highly restricted.

6.3.4 The microphones and microphone positions

Measurements were carried out using 12, omnidirectional dbx RTA-M microphones that were individually calibrated in an anechoic chamber applying the free-field substitution method using a JBL M2 as the source and a Bruel and Kjaer 4136 as the reference microphone.

Within each cinema, the microphones' calibrated, flat frequency response axes were aimed at the loudspeaker. Microphones were placed at every other seat location, with an offset of one seat in every other row in order to avoid symmetry and positional redundancy. Though there has been recommendations for the placement of microphones for spatial avergaing by other researchers based on wavelength (Lubman, 1971), this configuration was chosen as it represents one that can easily be replicated in a commercial cinema by an audio technician; figure 6.6 shows an example of this layout.

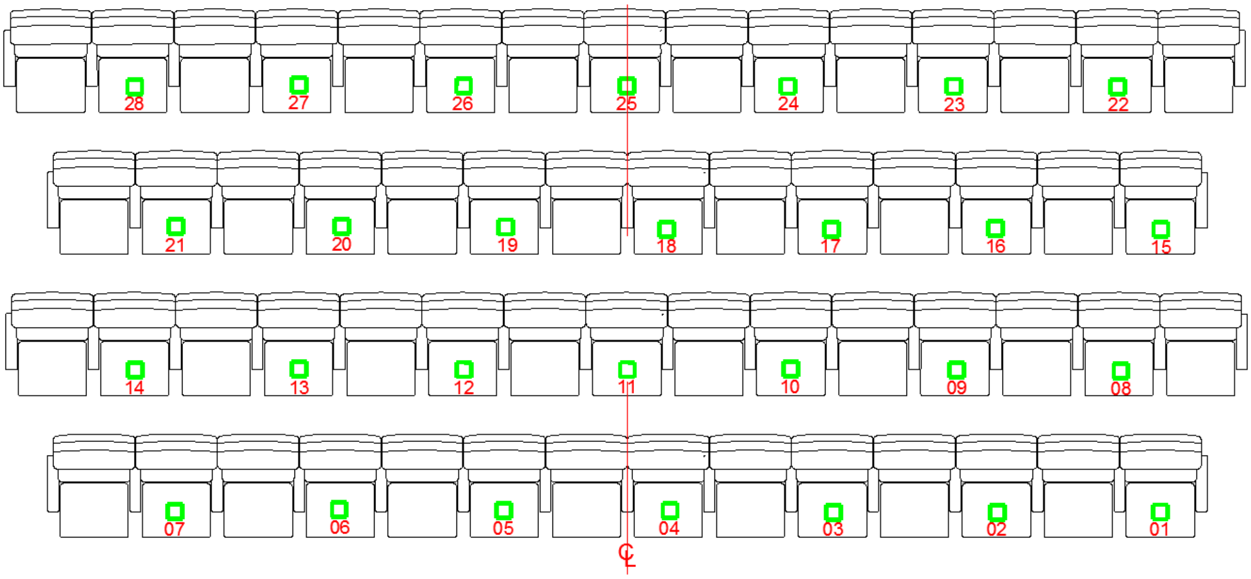


Figure 6.6: Example microphone layout

The exception to placing a microphone in every other seat was Cinema G, where it was thought that measuring 258 locations would be excessive and that 109 positions would adequately represent the cinema. In this case, a microphone was placed in every other seat in every other row with the exception of the left side seating area which was mirrored the right side seating area.

Figure 6.7 depicts the average distance between microphones in the cinemas. Slight variations in microphone distances between cinemas were due to different styles of theatre seats being used, along with varying seating level depths. A microphone was placed at a position 2/3rds of the length of the cinema from the screen, in the center seat (2/3L position), and all microphones were positioned so that the capsules were 1.0 to 1.2 meters above the finished floor, both as recommended in the SMPTE 202-2010 standard. In selected cinemas the microphones were placed at random heights above the seats in order to isolate the seat-dip effect, as further described in 6.4.2.

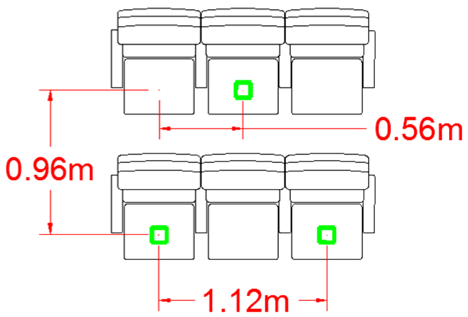


Figure 6.7: Average distance between microphones

6.3.5 The measurement procedure

Frequency response measurements were taken using Harman Audio Test Software (HATS). A 4s swept sine wave ranging from 20Hz to 20kHz was utilized. Given that each cinema displayed a reverberation time of approximately 0.45s or less, an FFT stop gate time of 500ms was used to capture the entire impulse response in an effort to best represent the system's steady state response and to provide improved resolution in the lower frequencies. The playback levels for the sine sweeps were adjusted to create an SPL of approximately 85dB C-weighed at the 2/3L position as per the SMPTE standard.

6.3.6 Data post-processing

A gain factor was added (or subtracted) in order to normalize the levels between the measurement positions. This process provided an averaged response that did not weigh the power of one position more than another. The normalized level was based on the C-weighted level from 2k to 10kHz at the reference position (85dB).

6.3.7 Spatial averaging of the microphone positions

Spatial averaging was used to determine the mean value of sound pressure level over the various positions within the cinemas, as determined by the power averging shown in Equation 6.2:

$$SPL_{ave} = 10 * \log\left(10^{\left(\frac{SPL_1}{10}\right)} + 10^{\left(\frac{SPL_2}{10}\right)} + 10^{\left(\frac{SPL_3}{10}\right)} ... 10^{\left(\frac{SPL_N}{10}\right)}\right) - 10 * \log(N) \qquad \text{Eq. 6.2}$$

N is the number of positions

Microphones were placed in predetermined groups for averaging. For each cinema, one group was comprised of all the measurement positions. Subsequent groupings were made up of sets of 4, 8, 12 and 16 positions, with larger cinemas including a grouping of 36 positions.

Within each cinema, every grouping had to meet the following requirements:

- Include the 2/3L position
- Within the groupings of 4, the only common position between groupings is the 2/3L position
- Avoidance of symmetry between the left and right sides
- Avoid using adjacent positions, with the exception of the smaller cinemas
- Avoid permutations between groups of similar sizes
- No group will consist of only extreme outer positions (front / back rows, sides)
- Avoid boundary condition when ever possible

Six sets of each grouping size were created for each cinema. From each set of grouping sizes, only the set that displayed the maxium deviation from the loudspeaker data was considered for further review. This allows an observation based on a worst-case scenario.

6.4 Results

Section 6.4 reports on the analysis of the measurements and the results.

6.4.1 Overall in-room response compared to loudspeaker data

Figures 6.8 and 6.9 show the in-room response of the JBL M2 spatially averaged across all microphone positions and 4 positions, respectively, versus its anechoic response for each cinema. The anechoically measured response of the JBL M2 is shown in solid black. This response curve is an amalgam of the sound power data below 300Hz, the listening window data above 300Hz and 10m of air attenuation beginning at 1kHz. The decision to use this curve is threefold:

1. Below 300Hz the sound power data gives a rough approximation of a diffuse-field response.
2. As the cinemas have relatively low reverberation times, especially in the mid to upper frequencies and are devoid of early lateral reflections, these rooms are dominated by the listening window data above approximately 300Hz.

3. 10m of air attenuation represents a good compromise for describing the high frequency loss across all 7 cinemas

The heavy dotted black curves represent a +/-3dB "tolerance", similar to SMPTE ST202-2010. These overall responses, along with the response of the M2, have been normalized to 85dB between 2k and 10kHz.

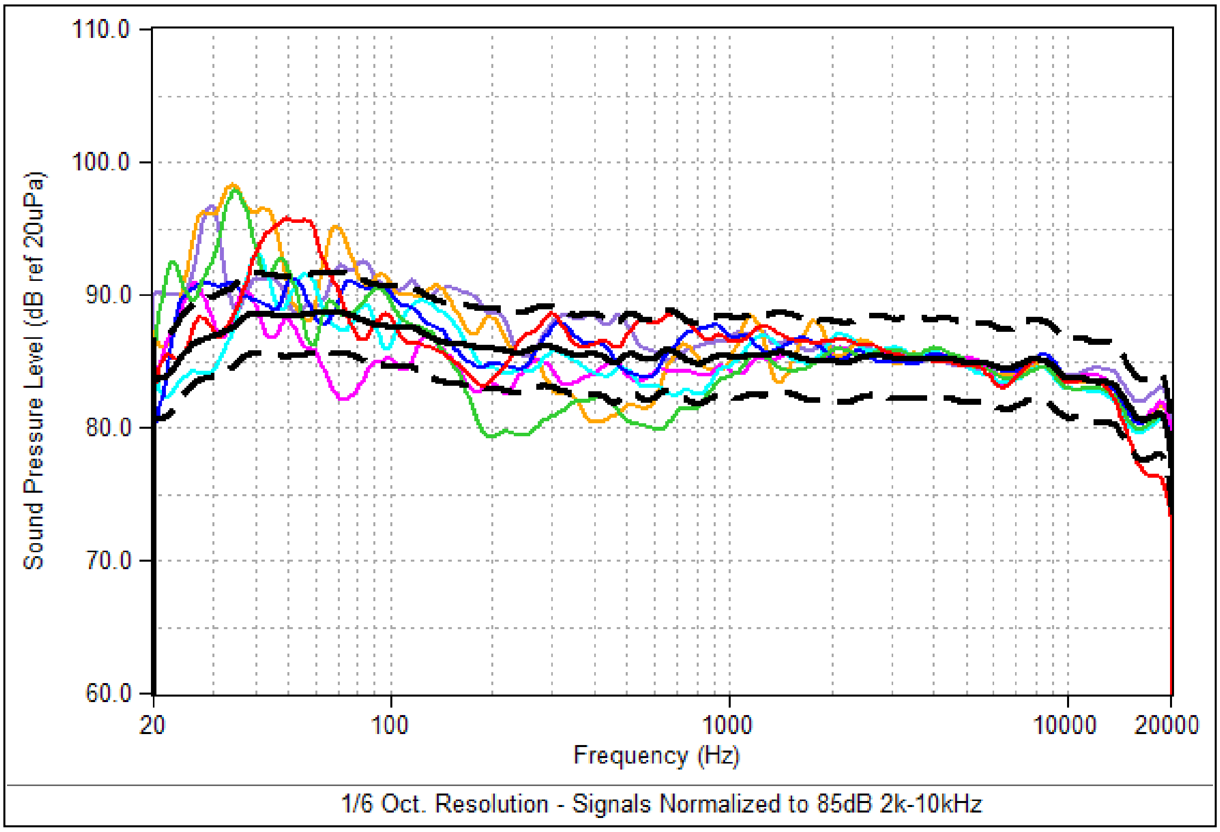


Figure 6.8: In-room response using all positions spatially averaged for each cinema compared to the JBL M2 sound power / listening window curve, shown in solid black; the black dashed black lines show the +/- 3dB tolerance (Cinema A - purple, B - gold, C - magenta, D - cyan, E - blue, F - green, G - red)

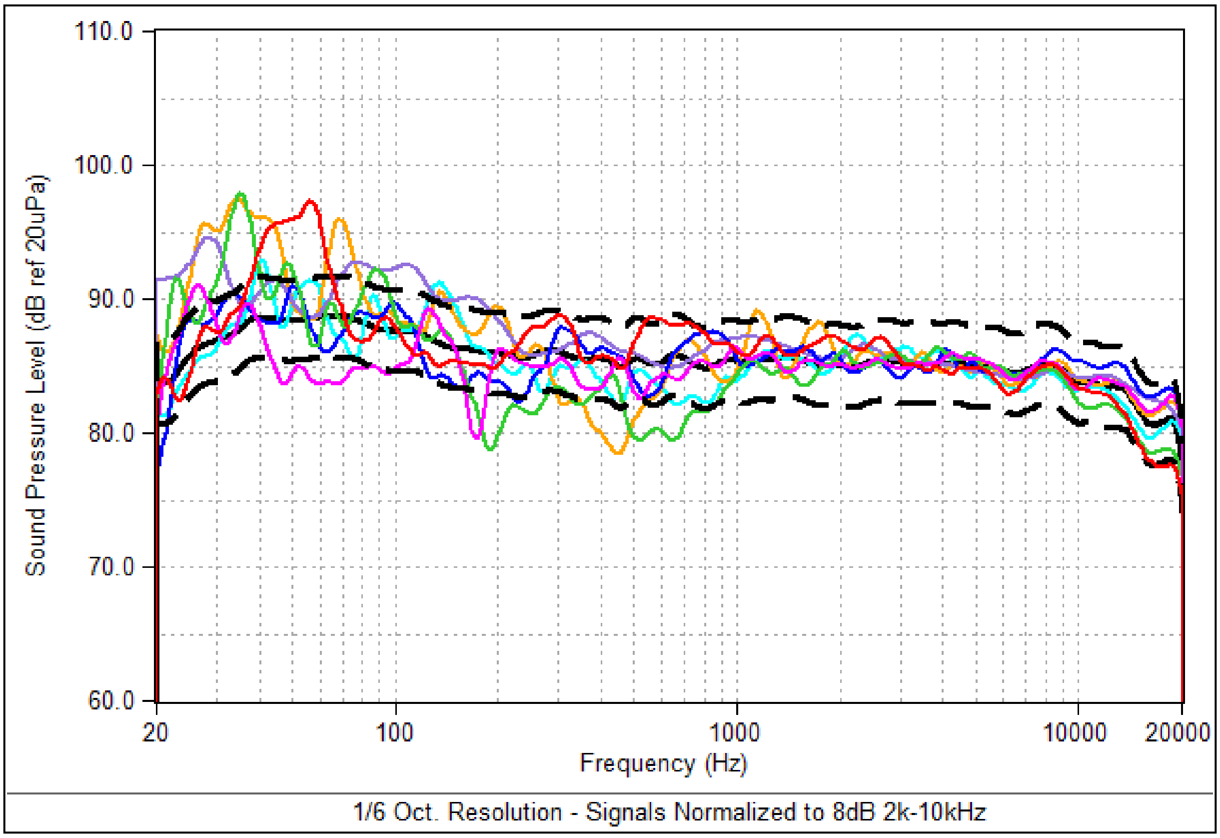


Figure 6.9: In-room response using 4 positions spatially averaged for each cinema compared to the JBL M2 sound power / listening window curve, shown in solid black; the black dashed black lines show the +/- 3dB tolerance (Cinema A - purple, B - gold, C - magenta, D - cyan, E - blue, F - green, G - red)

The amount of air attenuation at 10m was calculated using ISO standards and is shown in figure 6.10.

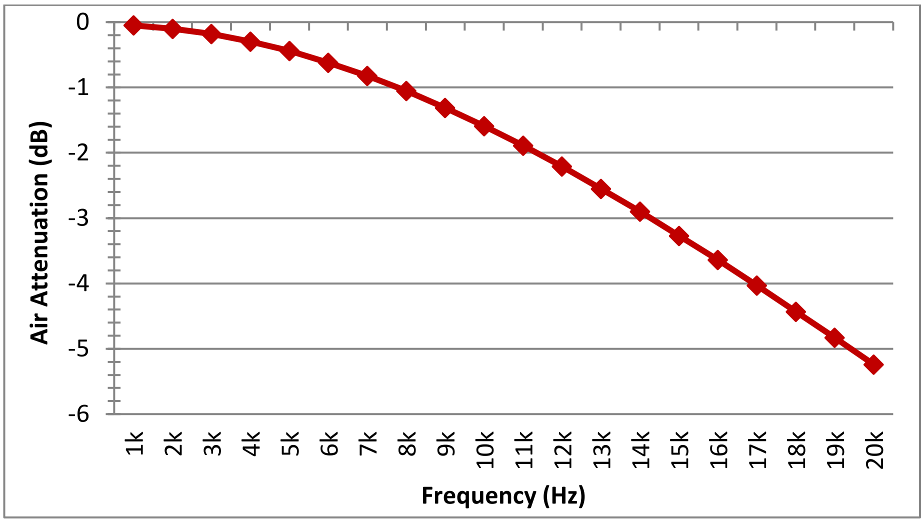


Figure 6.10: Air attenuation per frequency at 10 meters, 20°C and 50% RH; derived from (ISO, 1993, ISO, 1996)

From figures 6.8 and 6.9 it can be seen that beginning at approximately 200Hz and continuing upward to approximately 16kHz, all seven cinemas follow the listening window curve within the +/-3dB tolerance with the exception of cinemas B and F. Above 16kHz, the in-room curve for cinema G start to fall below the tolerance window by an amount related to the additional air absorption posed in this larger cinema. From 1kHz down to 200Hz the in-room curves begin to diverge from each other due to room-specific acoustic interferences. This is almost certainly due to the off-axis sound not being adequately reflected in certain cinemas, and most likely related to the variance in frequency-dependent absorption across the cinemas that continues to be at play below 1kHz. In this region, the direct sound still acts as a significant, though not dominant, effect. Issues such as seat dip become apparent in this frequency range; which is a topic explored further in section 6.4.2. However, below approximately 200 Hz there is little to absorb sound and therefore the curves rise to meet the sound power curve.

Four very different sized cinemas were reviewed separately in order to observe variations in response with respect to volume and reverberation. Figure 6.11 shows the spatially averaged response of every microphone position in cinemas A, C, D and G in comparison to the sound power / listening window and the listening window curves of the M2 loudspeaker (both with air attenuation added). Once again, these overall responses, along with the response of the M2, have been normalized to have the same 85dB SPL level between 2k and 10k Hz.

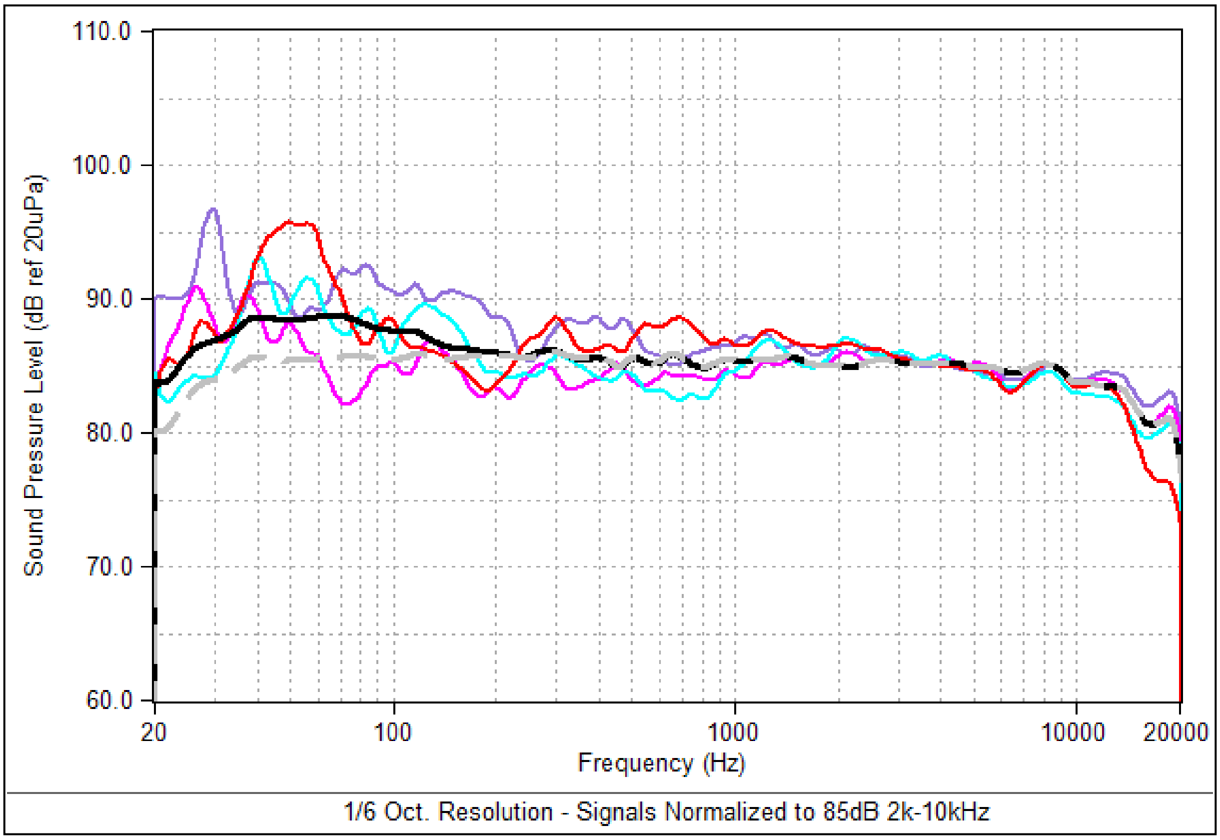


Figure 6.11: Overall in-room response of cinemas A, C, D and G (purple, magenta, cyan and red), compared to the JBL M2 sound power / listening window (black), listening window (grey)

The response in all four of these cinemas follow the sound power / listening curve very closely between approximately 200Hz and 16kHz. Cinema C had the lowest measured low-frequency reverberation time of all the cinemas and so it follows that its in-room response curve would parallel that of the listening window data below 200Hz, as shown in figure 6.11. Of interest to note, cinema C is the only cinema in this group which is used as a dub-stage as well as for screenings and it demonstrates similar low frequency reverberation times found in the dub-stages of the SMPTE B-chain report (2014). Cinema A in contrast, though a smaller room than C, has longer reverberation times in the lower frequencies, which provides for an energy build-up that more closely follows the sound power curve. In addition, this room (as well as cinema B) is more "box" shaped causing it to be prone to room modes. Cinemas D and G, which are much larger in size comparatively, also have longer reverberations times and in turn also follow the sound power curve.

6.4.2 Randomizing microphone heights

In an effort to observe seat-dip effect, three of the seven cinemas, C, E and F, had measurements taken a second time with microphones again placed in every other seat but with their heights randomized between 1.3 and 2 meters. These three cinemas were chosen for their varying sizes and seating layouts.

Initially there were concerns about how empty seats during measurements might render invalid results with respect to seat dip. However, research into the seat dip effect has shown that there is little difference in the attenuation whether an audience is present or not (Choi et al., 2015), (Schultz and Watters, 1964), (Sessler and West, 1964).

Figure 6.12 shows the differences in the spatially averaged responses between ear and random heights of all the microphone positions for the three selected cinemas. All of the responses have the same overall C-weighted level. The pairs of response curves from each cinema have been shown separated by 10 dB for clarity.

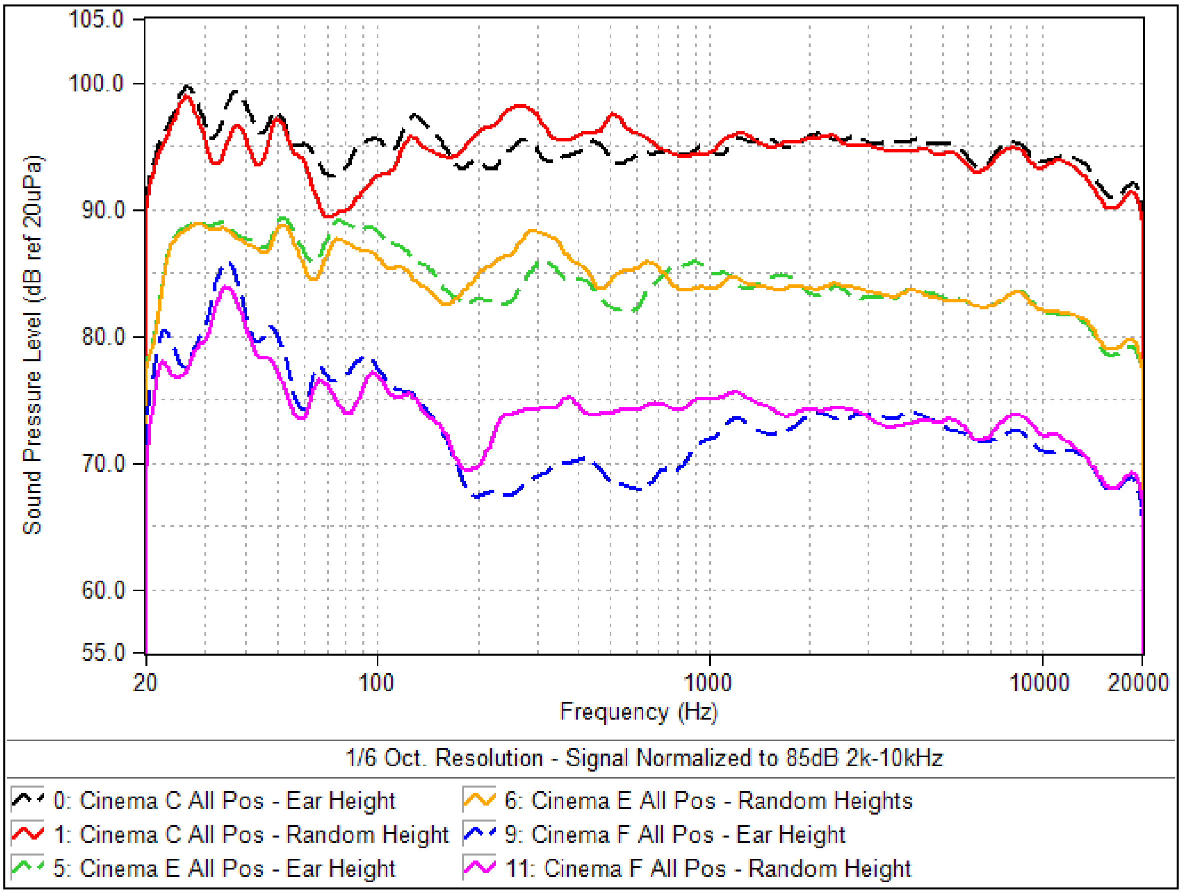


Figure 6.12: Cinema C, E and F comparisons of averaged responses all measurement positions, ear heights (dashed line) versus random heights (solid lines)

Figure 6.13 depicts the differences in responses between ear and random heights at the 2/3L microphone position for the three selected cinemas. The pairs of response curves from each cinema have been shown separated by 15dB for clarity.

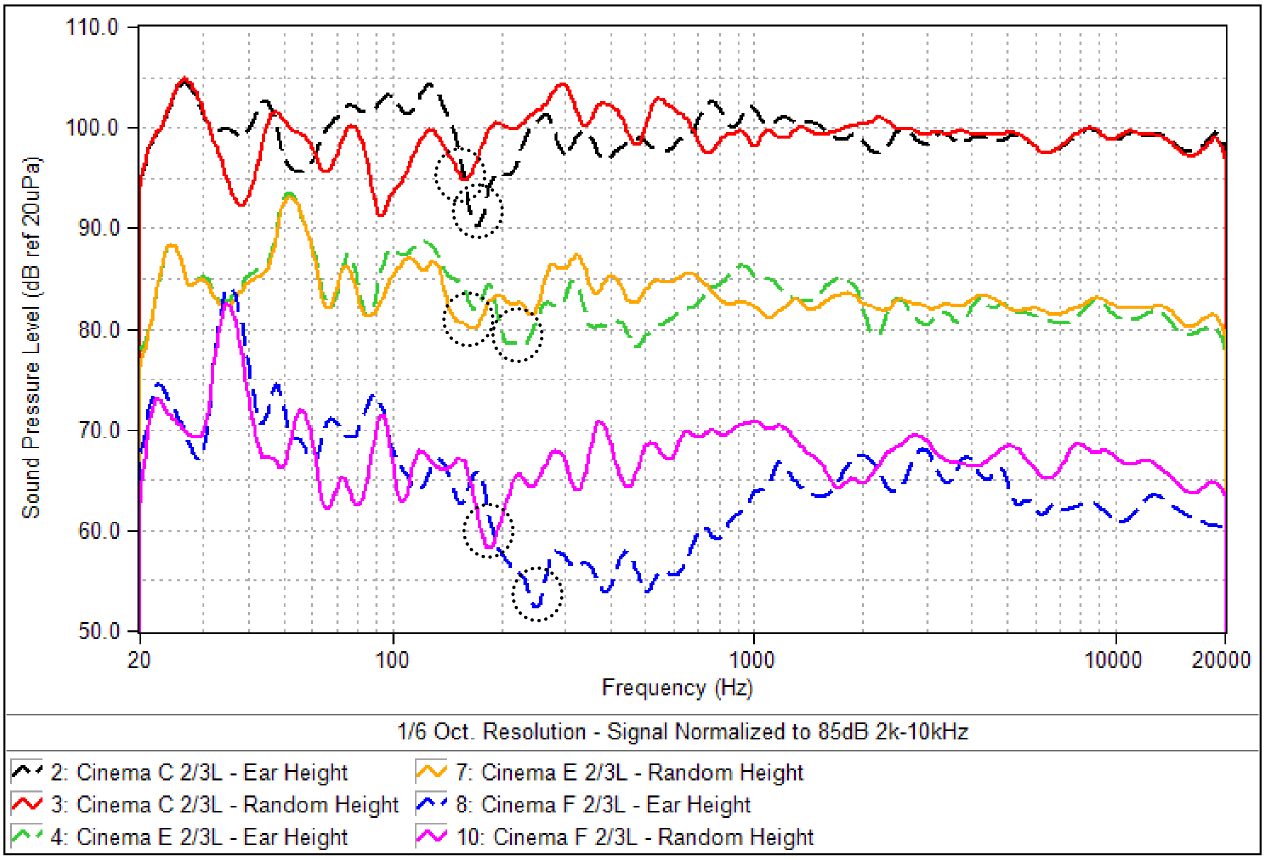


Figure 6.13: Cinema C, E and F comparisons of responses at the 2/3L positions, ear heights (dashed line) versus random heights (solid lines)

All cinemas displayed some amount of seat effect, as expected. The larger cinemas, especially those with a very gradual rake to the seating area, such as cinema F, displayed the largest amount of seat effect. Research has found, however, that even stadium seating can display a seat effect even though it has an elevated angle of incidence (Holman, 2007), (Bradley, 1991). Both sets of figures show that randomizing the microphone heights, essentially a type of vertical spatial-averaging, appears to fill in some of this effect but does not eradicate it all together. Prior research by Davies and Lam (1994) has shown that raising microphone heights (above ear height, between 1.56 meters to 2.38 meters) caused the frequency of the dip and its associated attenuation to decrease but did not eliminate it completely, as shown in the circled portions of figure 6.13. Raising the microphones also appears to have a slight smoothing effect on the frequency response above 1k Hz, but it is difficult to state with certainty if this has any real beneficial effect subjectively.

Figure 6.14 shows that with 4 spatially average microphone positions placed at random heights, the in-room response measurements are within +/-3dB of the loudspeaker data starting at approximately 200Hz to 300Hz.

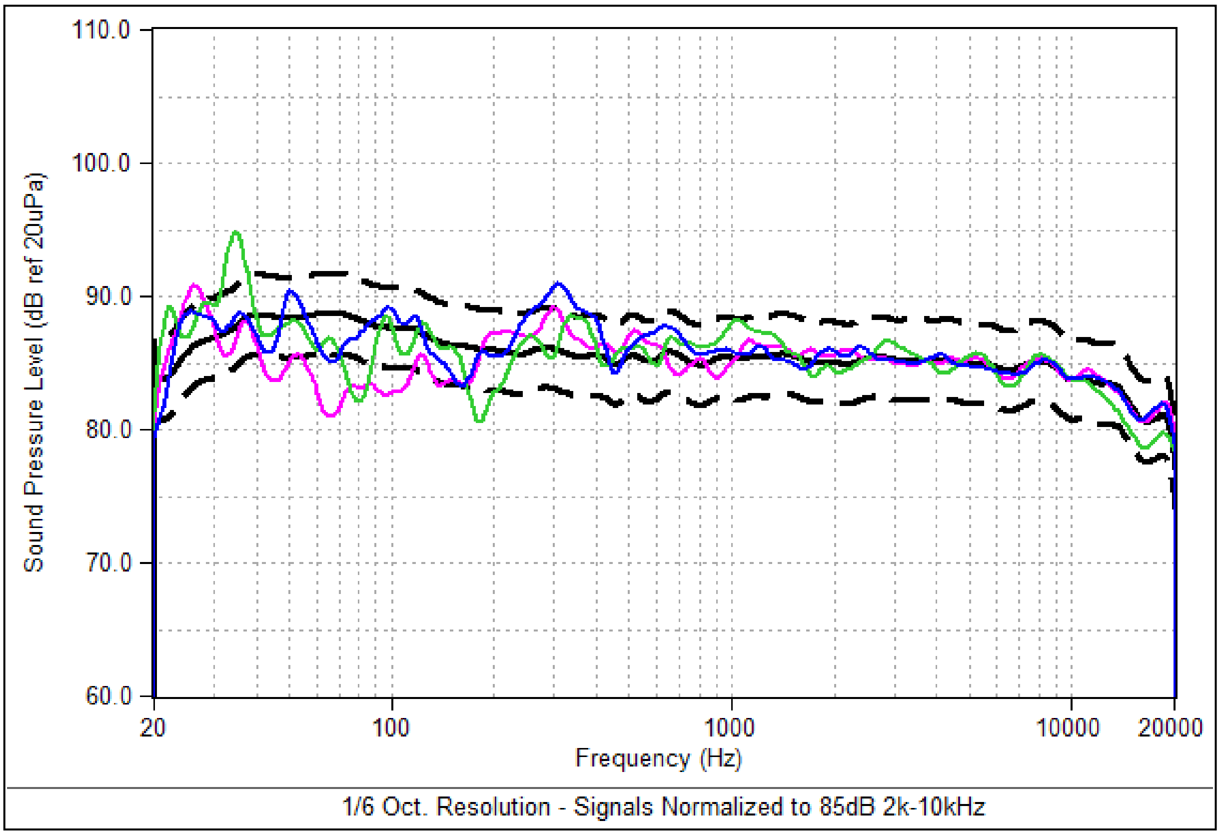


Figure 6.14: Cinema C, E and F spatial average of 4 positions (magenta, blue and green) compared to the M2 sound power / listening window and tolerance (black)

Overall, filling in seat dip and other seat effects by raising the microphones can bring the in-room response closer to the sound power curve, especially in the region between 200 and 800Hz.

The perception of seat dip in cinemas is a topic that has been researched very little to date. One recent research project found that many listeners were not able to hear discernible differences between sitting and standing but that some listeners noted a rise in bass upon standing (Newell et al., 2015b). This observation corroborates earlier findings in research which utilized a simulated concert hall, where it was found that listeners were able to detect a change in attenuation which they attributed to a change in bass tone or warmth (Davies et al., 1996).

Further research in this area is warranted to ascertain if seat-dip in a cinema has an effect on listener preference.

6.4.3 Comparing microphone groupings to loudspeaker data using frequency-by-frequency normalization

The anechoic data of the loudspeaker is held as the baseline by which the in-room measurements, divided out into groups of spatially averaged positions, are compared. A novel approach to this comparison is to directly observe the difference between the anechoic data and the measured data.

When comparing a spatially averaged signal to a signal measured in an anechoic chamber, diffuse-field methods are difficult to apply as neither signal was obtained in a truly diffuse-field. As noted in sections 6.3.3 and 6.4.1, a curve created from the M2 sound power and listening window data was used as it provided the best approximation of the in-room response. However, this curve is derived from anechoic measurements, is theoretically in its approximation and cannot be considered an exact representation of a diffuse-field measurement. As noted in prior research entailing dub-stages and cinemas which were mentioned in section 6.3.1, the total energy in these rooms from 500Hz to 20kHz consist almost entirely of the energy at 50ms or less. The reverberant energy therefore has little or no effect on the signal that is received by the re-recording mixer or most of the audience. For these reasons, a straight forward method of frequency-by-frequency normalization using the magnitude of each signal at a given frequency was utilized.

The power in the difference spectrum, or 'room signal', for a given spatial averaging and within a given frequency band is derived from:

$$|Z_{ni}(f)|^2 = |Y_{ni}(f)|^2 - |X_i(f)|^2 \quad b_i \le f \le e_i \qquad \text{Eq. 6.3}$$

$\boldsymbol{Z_{ni}(f)}$ is the room signal, $\boldsymbol{Y_{ni}(f)}$ is the spatial average of $\boldsymbol{n}$ microphones, $\boldsymbol{X_i(f)}$ is the loudspeaker sound power / listening window curve and $\boldsymbol{i}$ represents an octave

The initial analysis was undertaken using 48 points per octave and so:

$$i = \sum_{p=1}^{48} i_p \quad \text{where} \quad p \in \mathbb{N} \qquad \text{Eq. 6.4}$$

Since the process in equation 6.4 is iterative for each of these points within a given octave $\boldsymbol{i}$, the term $\boldsymbol{i}$ is used in place of $\boldsymbol{i_p}$ for ease of notation.

When expressed in decibels, the difference spectrum can be expressed as:

$$\boldsymbol{Z_{ni}}(f)dB = 10\,log_{10}\frac{|Y_{ni}(f)|^2}{|X_i(f)|^2} \qquad \text{Eq. 6.5}$$

An example graphical representation of the difference spectrum is shown in figure 6.15.

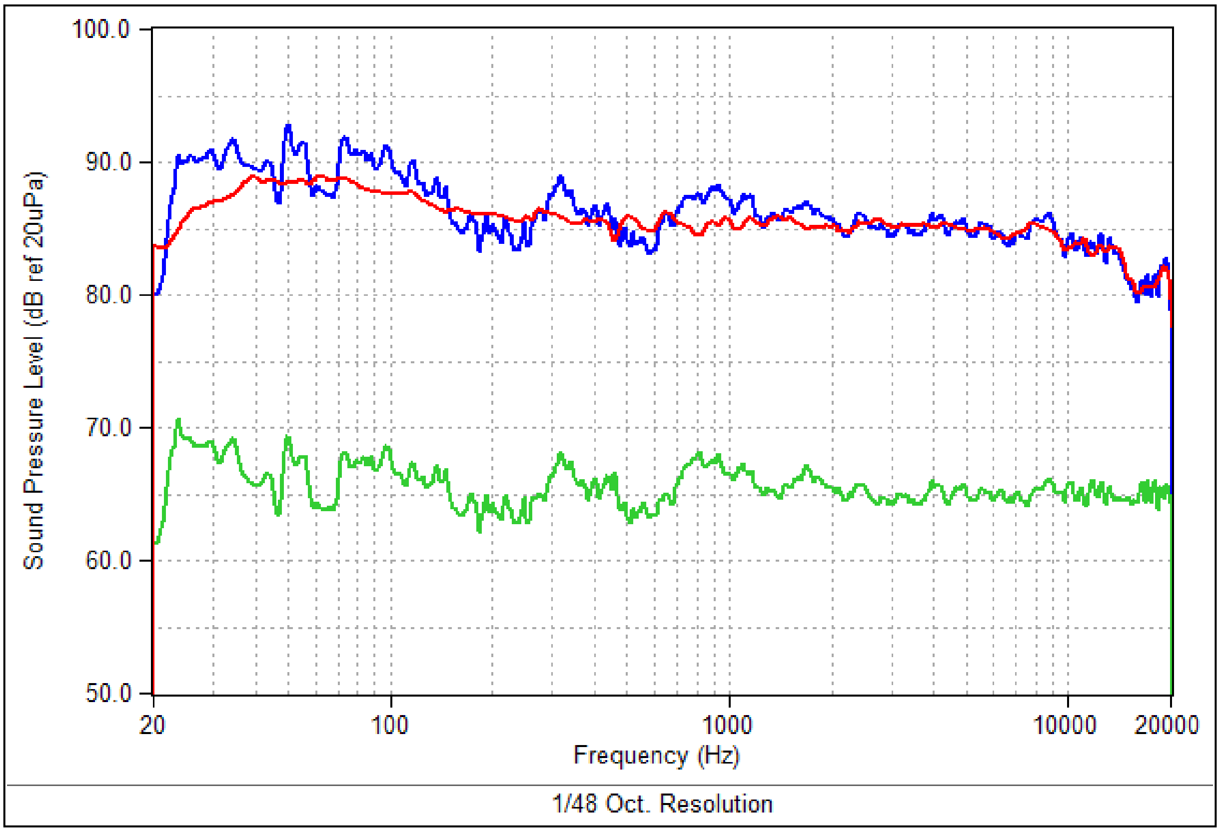


Figure 6.15: Difference spectrum ***Z*** (green) derived from a 36 position spatial average ***Y*** (blue) and the M2 sound power / listening window with air loss ***X*** (red); the difference spectrum is offset by +65dB for illustration purposes

The mean of the difference spectrum expressed in dB can be computed by:

$$\bar{\boldsymbol{z}}_{ni}dB = 10\, log_{10} \sum_{f=b_i}^{e_i} \overline{|Z_{ni}(f)|^2} \qquad \text{Eq. 6.6}$$

The standard deviation of the difference spectrum expressed in dB is expressed using:

$$\boldsymbol{\sigma}_{ni}dB = 10log_{10}\sqrt{\left(\frac{1}{e_i}\right)\sum_{f=b_i}^{e_i}|Z_{ni}(f)|^2 - \overline{|Z_{ni}(f)|^2}} \qquad \text{Eq. 6.7}$$

An analysis of the range between the maximum and minimum peak values (peak-to-peak) of the difference spectrum by octave band provides useful information on the amount of 'spectral ripple' in the difference spectrum within any given octave. Smaller peak-to-peak values indicate that the difference spectrum has a relatively smooth frequency response while larger values indicate peaks and notches which may be audible. The expression for the peak-to-peak value in dB for a given octave can be expressed as such:

$$\boldsymbol{Z}_{ni}p\_pdB = 10\, log_{10}\left(|Z_{ni}(f)|^2{}_{max} - |Z_{ni}(f)|^2{}_{min}\right) \qquad \text{Eq. 6.8}$$

The peak-to-peak values along with the means and standard deviations provide metrics to observe how similar the measured data is to the loudspeaker data and the level of variance between the spatially averaged groups within any given octave.

6.4.4 Comparing microphone groupings to loudspeaker data by cinema

In this section, the difference spectrum $\boldsymbol{Z}$ from 4 microphone position groupings (*all*, 16, 8 and 4) were chosen to compare to the loudspeaker data, which is shown as a flat baseline with +/- 3dB tolerances. These groupings were chosen as they represent a wide variation; starting with an 'extreme' grouping of *all* positions to smaller groupings which could be utilized in actual cinemas calibrations. The groupings of 36 and 12 positions displayed data similar to *all* positions and 16 positions respectively and were therefore not used.

Once the difference spectrum was derived, the data was smoothed to 1/6 octave, as this resolution is more revealing of peaks and notches than the 1/3 octave used by the SMPTE standard and is closely related to the Moore et al critical-bandwidth estimates (Moore, 2012), (Moore and Tan, 2004), (Moore and Glasberg, 1983).

Figures 6.16 through 6.19 show the comparisons between the position groupings and their relationships to the loudspeaker data. Note that figures 6.16 through 6.18 are for measurements taken at approximate ear height of 1.0 to 1.2 meters above the finished floor while figure 6.19 is for measurements taken at random heights. Cinemas are offset by 20dB for clarity.

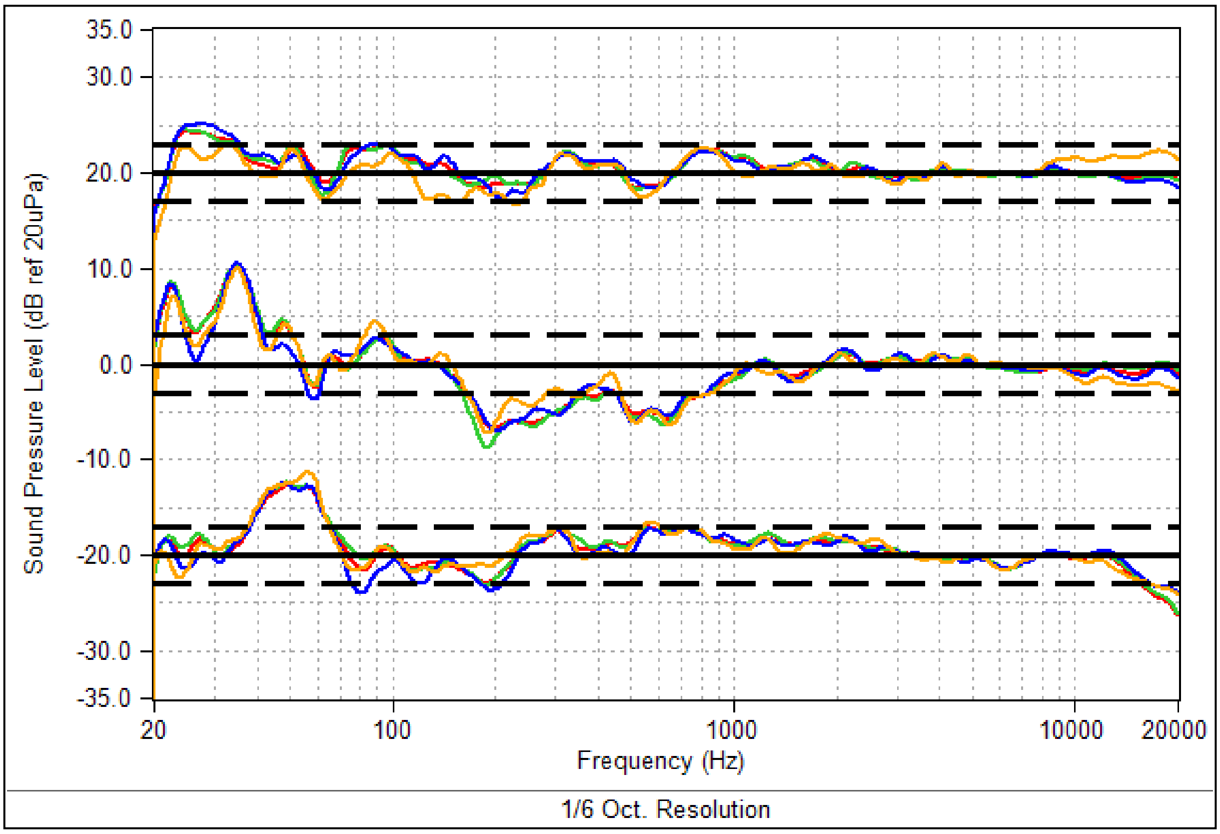


Figure 6.16: Cinemas E, F and G (from top) showing spatial averages of all positions (red), 16 positions (green), 8 positions (blue), 4 positions (gold), and loudspeaker data with +/- 3dB tolerance window (black)

In figure 6.16, cinemas E and G show good agreement with the loudspeaker data, being within the +/- 3dB tolerance from approximately 90Hz to 16kHz. Cinema F however, does not fall within the loudspeaker tolerance between approximately 160Hz and 800Hz. As discussed in section 6.4.2, cinema F has a very gentle floor slope and therefore exhibits a seat-dip effect in this region. In all three of these cinemas there is very little difference between the 4 position groupings within the 500Hz to 9kHz frequency range, with divergence only appearing at the lower and upper frequencies.

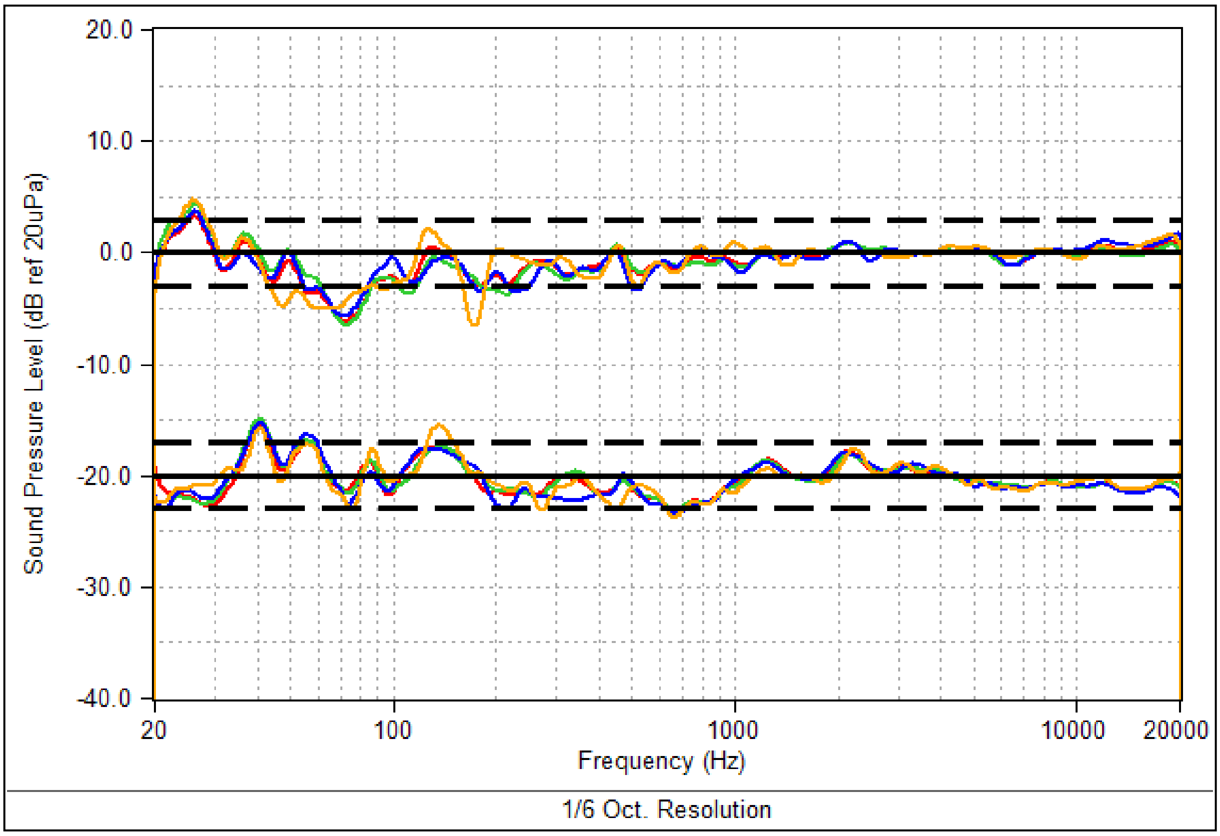


Figure 6.17: Cinemas C and D (from top) showing spatial average of all positions (red), 16 positions (green), 8 positions (blue), 4 positions (gold), loudspeaker data with +/- 3dB tolerance window (black)

From figure 6.17 it can be seen that both cinemas C and D show conformity with the loudspeaker data in the 200Hz to 20kHz range. Cinema D shows a shallow dip in the region between 500Hz and 1kHz, which may possibly be some form of seat effect. This cinema has a gentle sloping floor like cinema F, which lends itself to a higher potential for seat dip and other seat effects. In these two cinemas there is very little difference between the 4 position groupings within the 400Hz to 20kHz frequency range, with deviations only appearing at the lower frequencies.

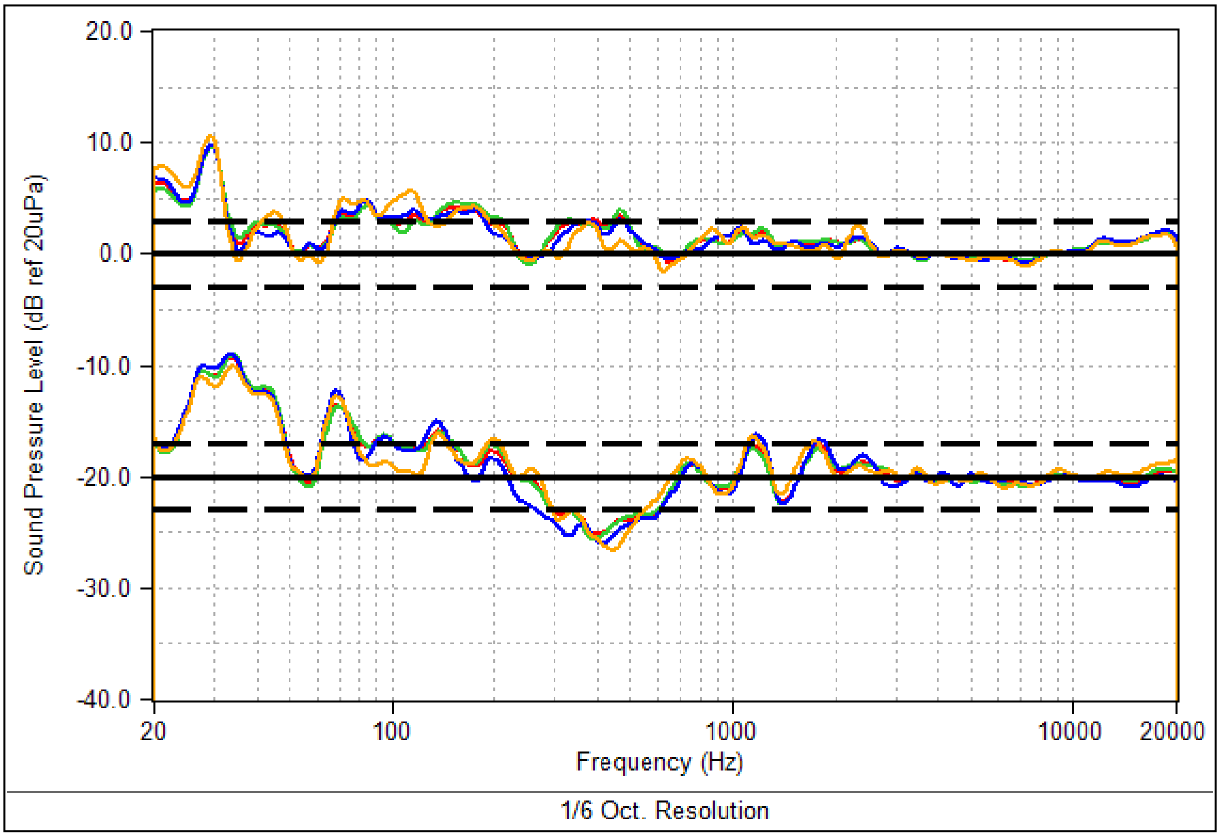


Figure 6.18: Cinemas A and B (from top) showing spatial average of all positions (red), 16 positions (green), 8 positions (blue), 4 positions (gold), loudspeaker data with +/- 3dB tolerance window (black)

Figure 6.18 shows the difference spectrums from the two smallest cinemas, A and B. These two cinemas display very different responses, as cinema A falls within the +/- 3dB tolerance from approximately 200Hz through 20kHz while cinema B displays a large dip between 200Hz and 600Hz. Cinema A also shows a much smoother response than cinema B in the 700Hz to 3kHz region. Both cinemas show a rise in bass energy starting at 200Hz, though cinema A's bass response is much smoother than B's and remains close to the +3 tolerance of the loudspeaker data. These dissimilarities are certainly caused by the major acoustical differences between the two cinemas. Cinema A is purpose-built for audio playback and quality control where cinema B was built to quality control picture only and therefore acoustics were not a priority in the design and building of this room. However, it can still be seen that there is very little difference between the 4 position groupings within the 500Hz to 20kHz frequency range, with deviations only appearing at the lower frequencies. This result was not unexpected given the small size of the cinemas.

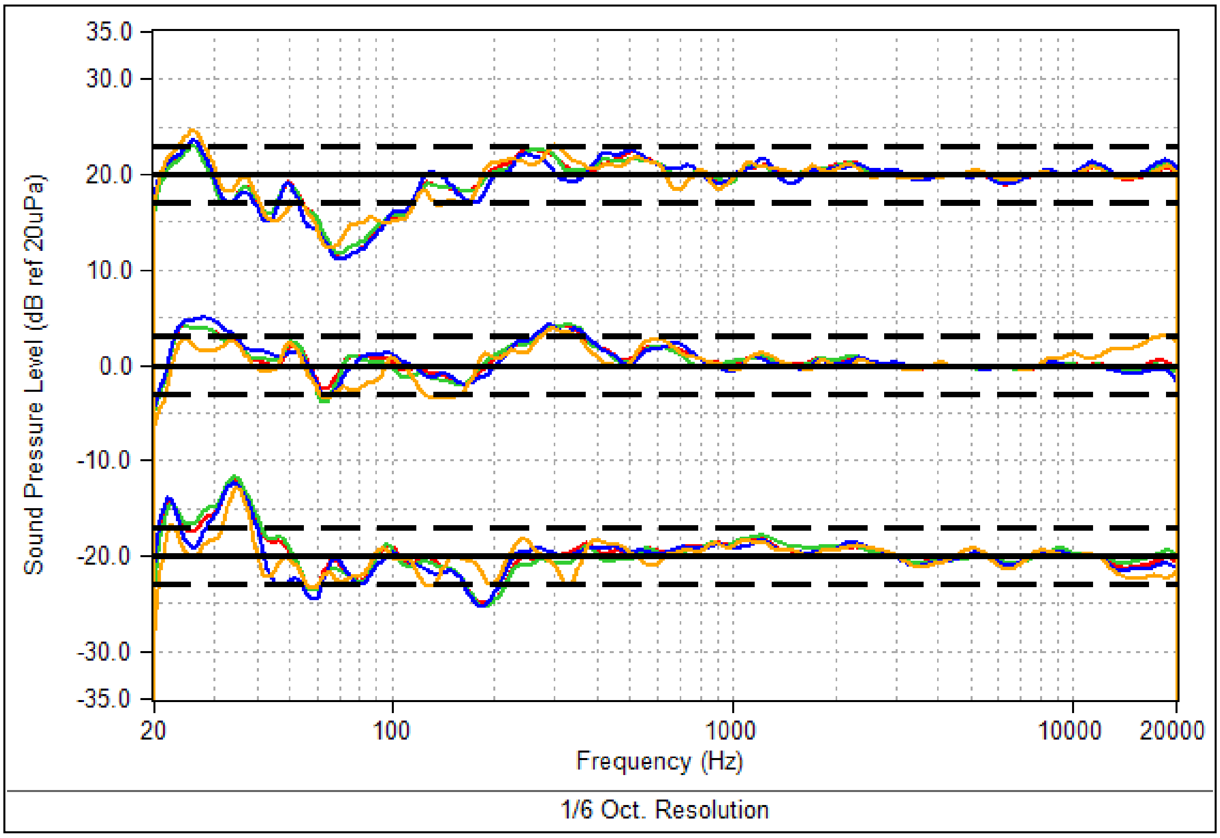


Figure 6.19: Cinemas C, E, and F (from top) showing spatial average of all positions (red), 16 positions (green), 8 positions (blue), 4 positions (gold), loudspeaker data with +/- 3dB tolerance window (black)

Figure 6.19 shows the data from the three cinemas where measurements were taken at random heights. In comparison to the measurements that were taken at ear height, the random height measurements bring the difference spectrum more in line with the loudspeaker data, especially in cinema F. Of interest to note is that the dip below 100Hz in cinema C has actually become deeper, indicating that randomizing the microphone heights in this cinema might have uncovered an issue that perhaps reflections were filling in when the microphones were at ear height.

6.4.5 Comparing microphone groupings using means, standard deviations and peak-to-peak data of the difference spectrum

For those cinemas where measurements were made at both ear and random height, the random height measurements were used in the following comparative analysis.

The peak-to-peak value of the difference spectrum from each of the 4 position groupings was averaged together and is displayed by cinema in figure 6.20.

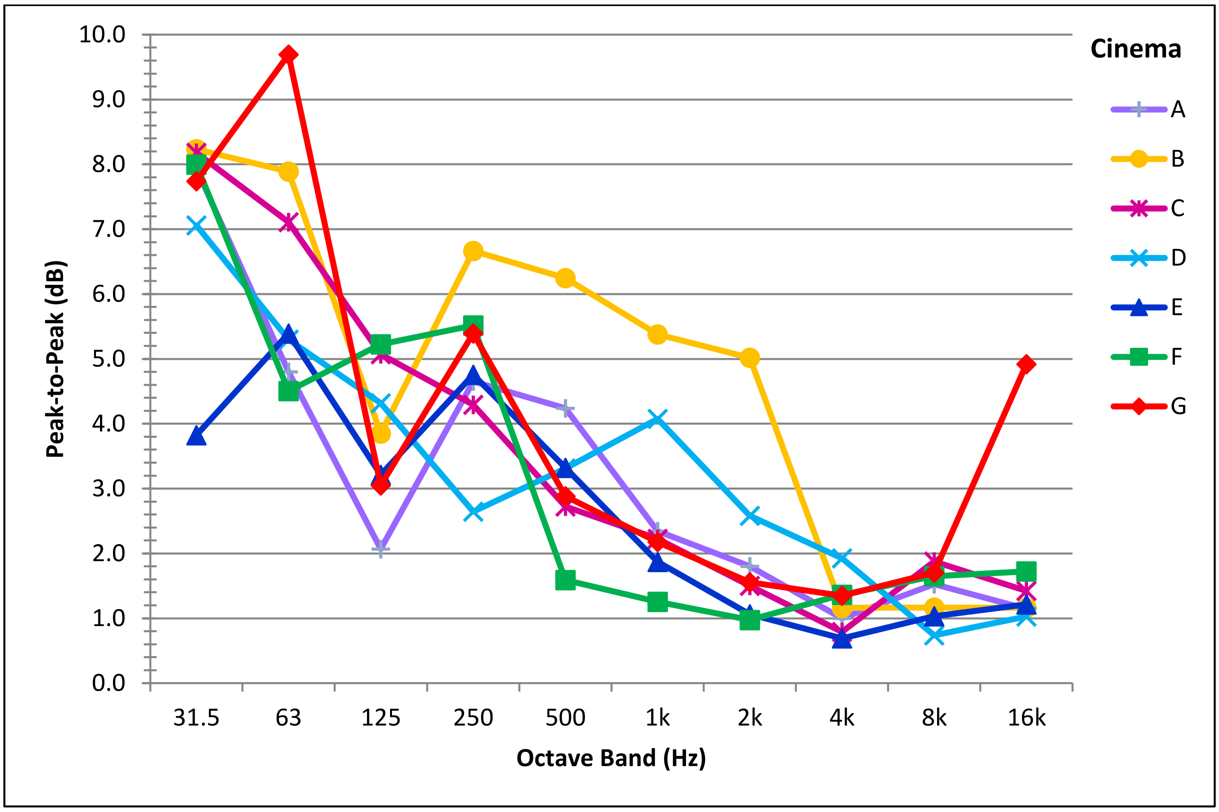


Figure 6.20: Average peak-to-peak value of the difference spectrum by cinema

When considering the average peak-to-peak value in each cinema, it can be seen that the difference spectrums in all cinemas have relatively smooth frequency responses above the 1kHz octave band with the exception of cinema B and to a lesser extent, cinema D. Below 1kHz, the room starts to become more of an issue as seat effects come into play. Below 250Hz, the peak-to-peak values in the cinemas steadily increase and start to widely diverge from each other as additional acoustic interference becomes a factor. This illustrates that differences between cinemas are mainly relegated to the lower frequencies and the extreme upper frequencies, which is not unexpected.

Since the average of the peak-to-peak values was calculated from the 4 different position groupings, the standard deviations give an indication as to the level of difference, or range of differences, between the groupings within any given cinema. This data is shown in figure 6.21.

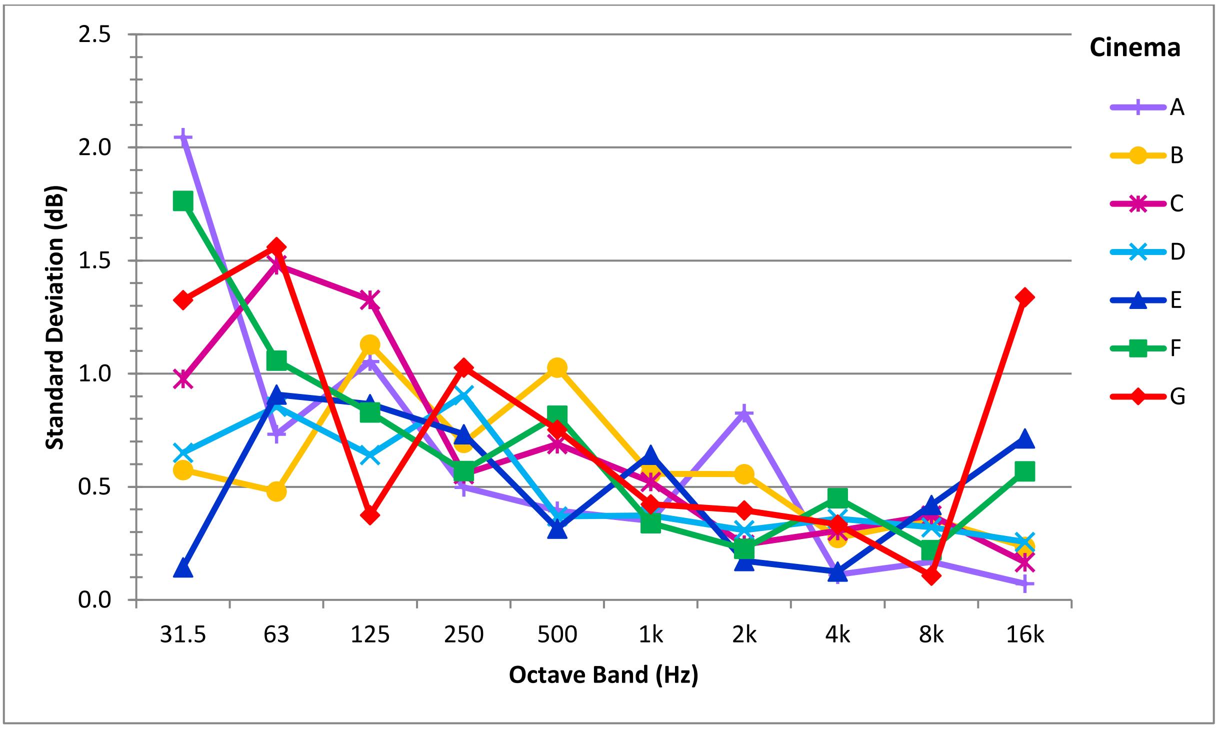


Figure 6.21: Standard deviations of the difference spectrums values by cinema

Within all of the cinemas, the 4 position groupings differed by only 1dB standard deviation above 250Hz with most displaying only 0.5dB standard deviation or less above 1kHz, with the exception of cinema G in the 16kHz octave band. This larger standard deviation in cinema G was most likely caused by differences in the amount of air attenuation measured between groupings. The lower standard deviations indicate that, within these frequency bands, increasing the number of positions in the spatial average will have little to no effect on the difference spectrum. Below 250Hz the position groupings start to display larger standard deviations, though most are kept within 1.5dB.

To further study the differences between the position groupings, the peak-to-peak values of the difference spectrums from each of the 7 cinemas was averaged together and is displayed by position grouping in figure 6.22.

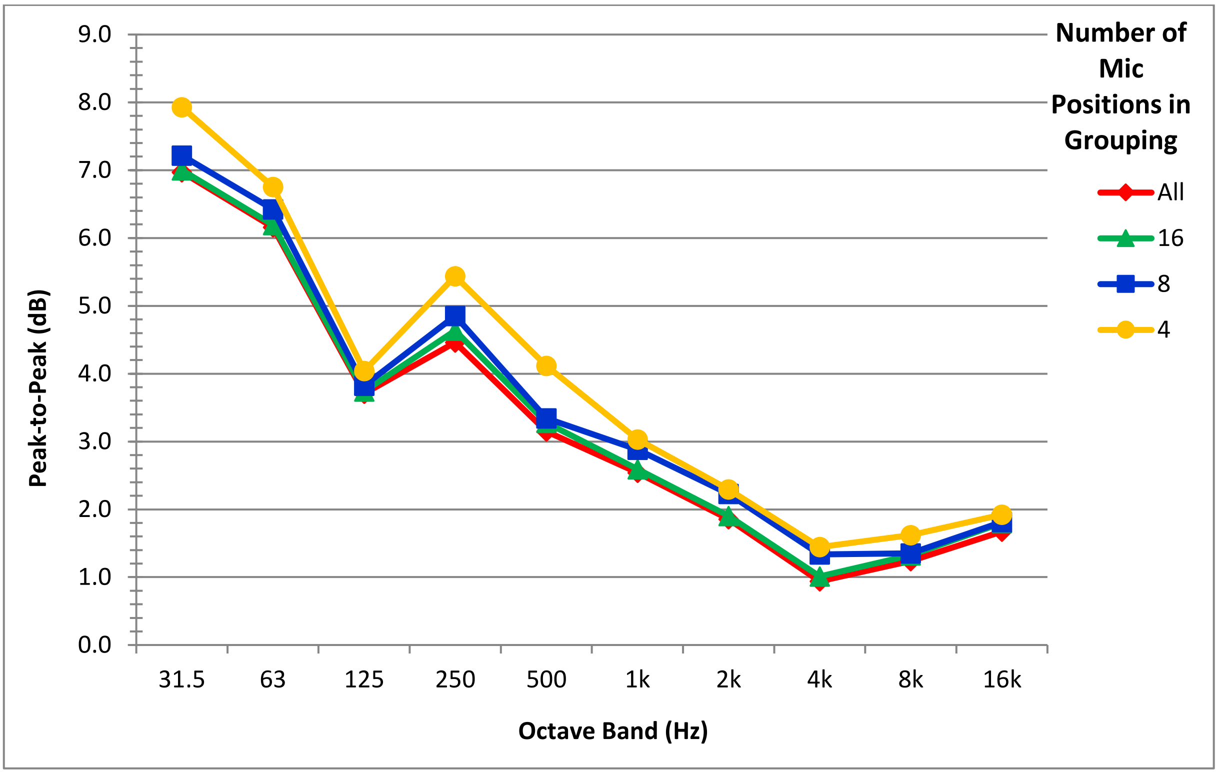


Figure 6.22: Average peak-to-peak values of the difference spectrum for each position grouping

Figure 6.22 displays that the 4 different position groupings, when averaged across the cinemas, exhibit peak-to-peak values that only differ by approximately 1dB. The peak-to-peak values of the 4 position groupings by individual cinema are shown in appendix P.

As the averages of the peak-to-peak values were calculated from the 7 different cinemas, the standard deviations give an indication as to the level of difference between the cinemas for each of the position groupings, as shown in figure 6.23.

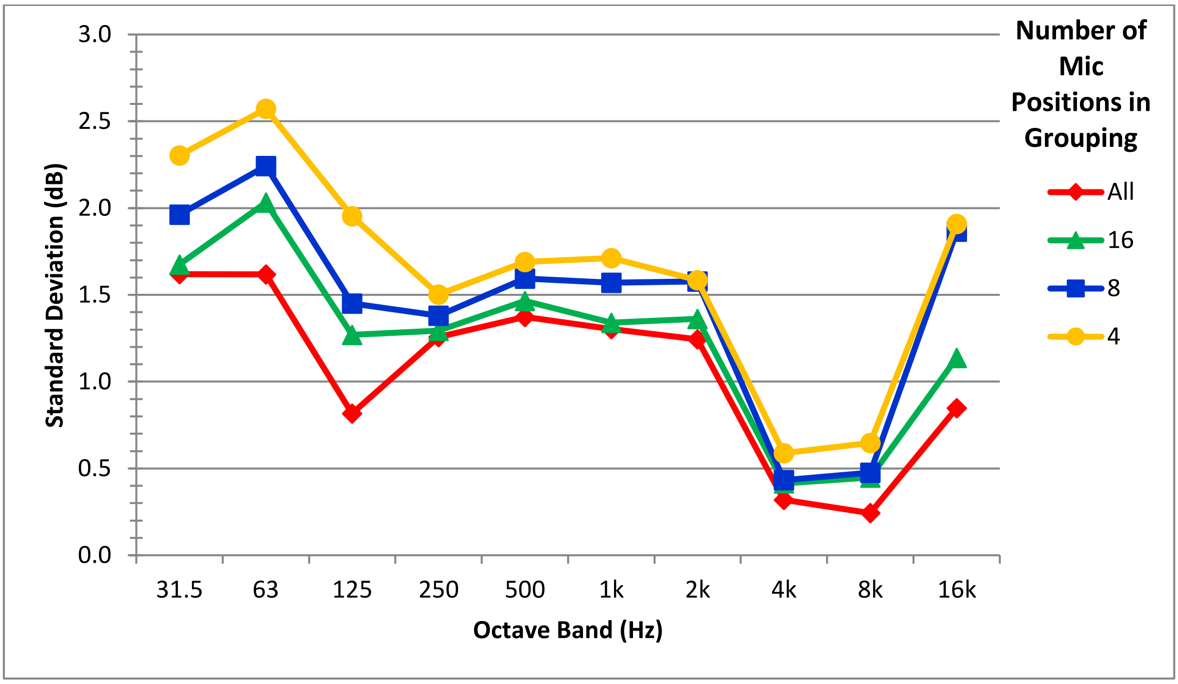


Figure 6.23: Standard deviations of the difference spectrums by position grouping

The standard deviations show that, though the position groupings means are within 1dB of each other across all octaves, their standard deviations start to vary below 250Hz and above 8kHz. Once again, this is expected as the largest difference between cinemas is in the lower and extreme upper frequencies as shown in figures 6.20 and 6.21. The standard deviations displayed in figure 6.23 demonstrate that, from 250Hz to 12kHz, the range of peak-to-peak values for any given grouping is held to approximately 1.5dB. Further noted in figure 2.23, the standard deviation, regardless of frequency band, lowers with an increase in the number of microphones positions used which is to be expected.

6.5 Discussion

Completed anechoic data can provide very useful information to a cinema technician when calibrating a dub-stage or cinema. In 1949, Hilliard inadvertently showed the connection between anechoic (open air) data and measured in-room response when he presented information on the performance of a new cinema loudspeaker, shown in figure 6.24.

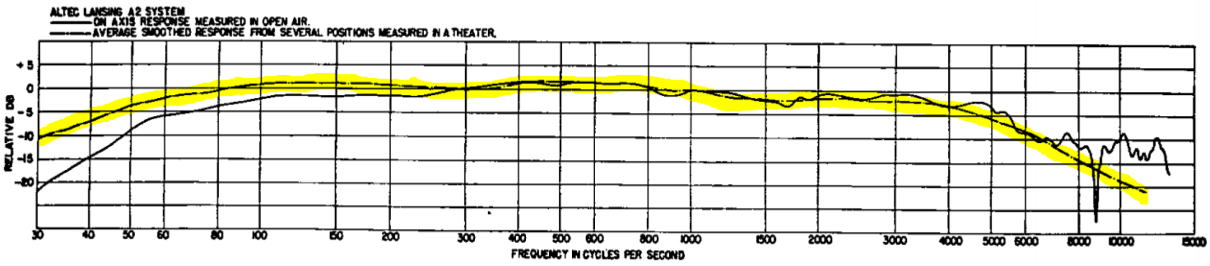


Figure 6.24: Altec Lansing A2 cinema loudspeaker, averaged response from on-axis response measured in the Academy Award Theatre (highlighted in yellow) versus the measured in open air; from (Hilliard, 1949)

Note from figure 6.24 it can be seen that the differences between the open air (anechoic) and in-room averaged response is in the lower and upper frequencies, with good agreement from 300Hz through 8kHz.

Early research also discussed the importance of having anechoic data available when testing a loudspeaker in a room, especially if the loudspeaker does not have good off-axis characteristics (Snow, 1961). Some early researchers also used the range of differences between the anechoic signal and the signal measured in-situ as a basis for studying the effects of a room on a specific loudspeaker. Korn and Hougardy called the range of differences been these two signals the "gain of the room" and assigned the variable ***K*** (measured by 1/3rd octave) to its value (Korn and Hougardy, 1959). These researchers felt that the ***K*** value could provide a modification factor to bring the room more in line with the anechoic data as well as providing information on the ratio of direct to reverberant energy.

Schulein (1975) found similar results to those in this chapter; when placing the same loudspeaker in three different rooms with the same microphone placement, there was little variation in the mid to high frequency response between the three rooms. Schulein's results are shown in figure 6.25.

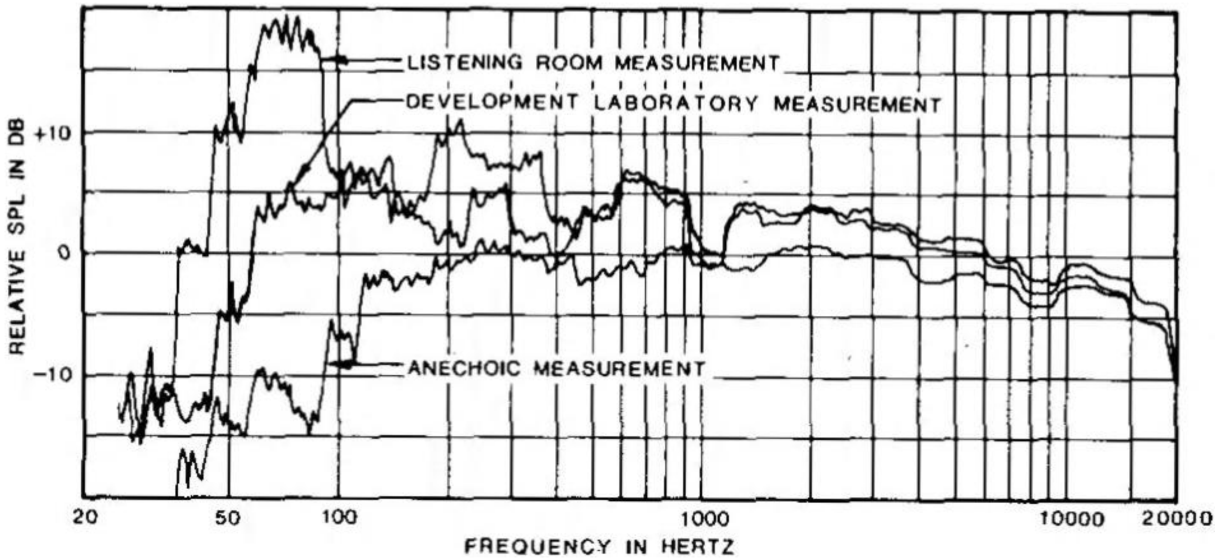


Figure 6.25: Influence of various rooms on a particular loudspeaker or a fixed loudspeaker - to - microphone spacing; from (Schulein, 1975)

Note the level of agreement between the data from 900Hz to 20,000Hz, which are all within approximately 3dB of each other. Schulein specifically commented that the variations were held to the low frequency range and were a result of the room acoustics.

Mäkivirta and Anet also found similar results when testing like models of studio reference monitors (Genelec) across 164 professional studios. They observed that measurements showed decreasing variations towards the high frequencies, most notably above 1kHz, as shown in figure 6.26 (Mäkivirta and Anet, 2001).

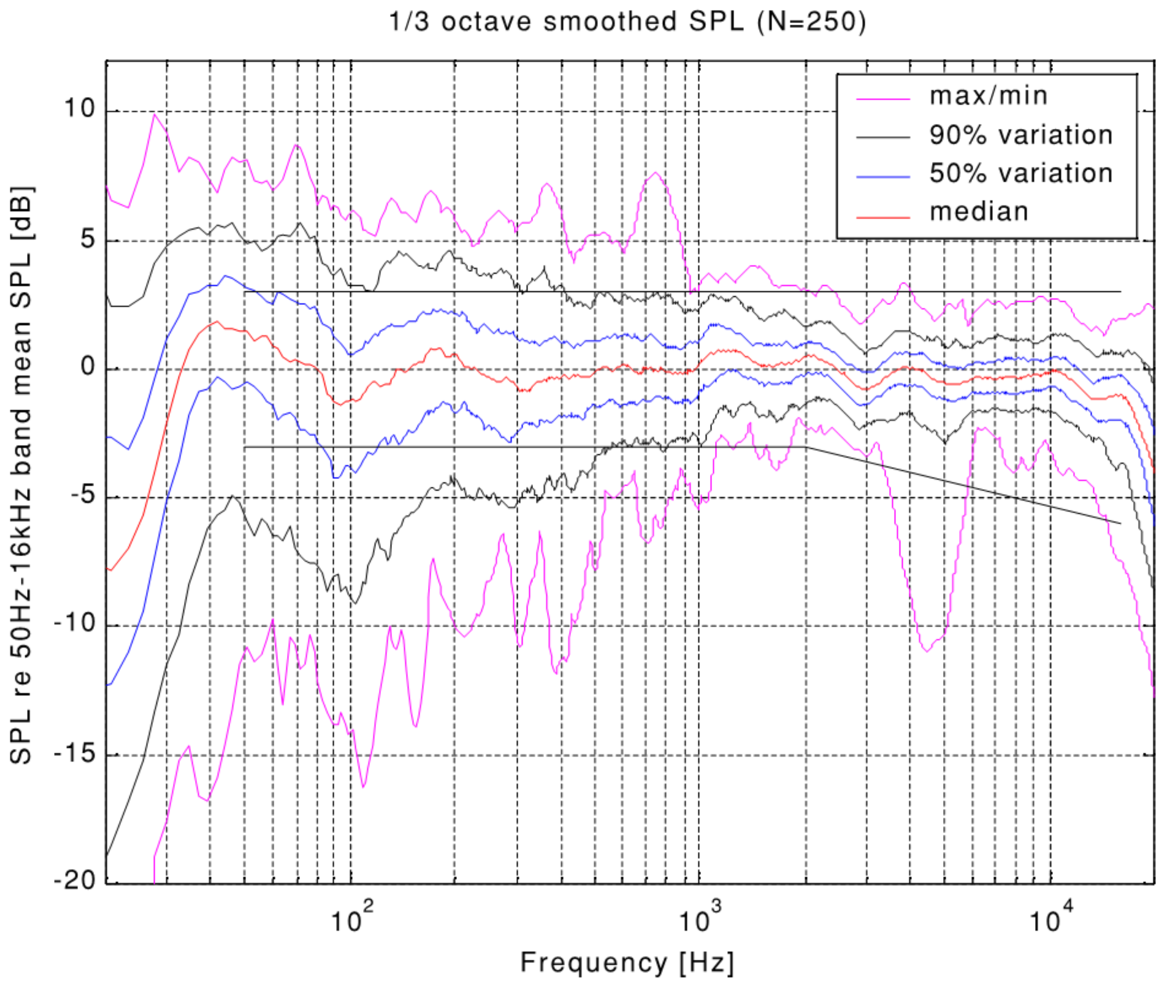


Figure 6.26: 1/3 octave smoothed frequency response averaged over 250 measurements in 164 studios showing the median, 50% variation, 90% variation and max / min variations; from (Mäkivirta and Anet, 2001)

The research in this chapter has shown that if the loudspeaker has constant directivity, a smooth off-axis response, then spatially averaging more than 4 in-room measurements to verify how it will perform in the room may not be necessary, especially if anechoic data is available. The response of the loudspeaker above the region of 800Hz, and in some cases starting as low as 300Hz, will translate directly to the cinema.

At frequencies above 1kHz, equalization will be mainly to compensate for any limitation in the loudspeaker. The high frequency roll-off dictated by the SMPTE X-curve is typically achieved simply by placing the loudspeaker behind a perforated screen. In larger cinemas, additional high frequency roll-off will naturally occur from air loss, as was observed with the measurements in this chapter. It was between 8kHz and 16kHz that deviations between the loudspeaker data and the measured data occurred; a phenomenon that was mostly observed in cinema G. However, the various position groupings showed very little difference between each other from 1kHz through 20kHz and so any additional number of microphone positions above 4 will not alleviate this deviation.

At mid frequencies, approximately between 200Hz and 1kHz, there will be diffraction from nearby seats, early reflections from walls, floors and ceilings, along with possible cabinet diffraction of the loudspeakers. There may also be crossover/directivity effects from the loudspeaker. Most of these effects will vary from seat to seat and thus at mid frequencies, a larger number of averages may be beneficial if the response and directivity of the loudspeaker is unknown. If effects from the seats are a concern, then raising the microphones above seated ear height can mitigate some mid-frequency anomalies. Newell at al recommends raising the microphones to a level of 60cm or more above the seat backs to mitigate the issue all together (Newell et al., 2015b). Even after the spatial averaging of several locations, mid-frequency irregularities will either be in the loudspeaker response, or will be some diffraction or room absorption effect that is occurring at all seats. Once again, it will be difficult to differentiate between the two without loudspeaker data. If the loudspeaker has constant directivity, the sound power data is known and the microphone heights are randomized especially in cinemas with gently sloped floors, then 4 carefully chosen positions may be adequate to calibrate the cinema in the range between 200Hz and 1kHz.

At low frequencies (below 200Hz), one could make a rough estimate of the general in-room response from the sound power data, if it is available. However, as shown in this chapter, measurements are highly variable and room-dependent below 200Hz and increasing the number of positions which are spatially averaged with not necessarily bring the in-room measurement into line with the loudspeaker data. The ability to accurately predict low frequency behavior in a room from the loudspeaker data is relegated to those rooms whose potential acoustic interferences in this frequency range have been well controlled, such as appears to have been the case with cinema A. These findings are in agreement with other research that has described the difficulty of measuring the low frequency response in such a way that meaningfully represents all seating positions (Hill et al., 2016), (Newell et al., 2015a). In the lower frequencies, mismatches between the spatially averaged measurements and the loudspeaker response would be due to issues such as modal behavior, acoustical absorption, loudspeaker boundary effects, and seat effects. Some of these effects are position related and therefore any attempt to equalize them from spatially average data, no matter how many positions were used, could lead to degrading the sound quality in areas throughout the cinema as opposed to improving it. A better understanding of how the potential performance of a room in the lower frequencies may be obtained by reviewing the anechoic data in conjunction with the low frequency reverberation time data. Cinemas G and F had comparatively longer low frequency reverberation times (as shown in figure 6.3) and also displayed larger low frequency rises above that of the anechoic data (as shown in figure 6.16).

Also revealed during this research, it was found that in absence of anechoic data, useful data about the loudspeaker may be obtained from the in-room response measurements, but only above 500Hz to 1kHz in typical cinemas. Additionally, if there is no information on the loudspeaker, one can measure the high frequency direct response of the loudspeaker or the loudspeaker / screen combination without windowing the data. The effects in the room at high frequencies are all due to the loudspeaker and should not be seat dependent, with the exception of directivity/coverage. Direct-field measurements of the loudspeaker can also be utilized, as suggested by SMPTE B-chain report (2014) and Newell et al (2011a), (2011b), though a balance must be reached between a short enough window length to avoid detrimental early reflections in the data and a long enough window to provide adequate frequency resolution.

With either method, this may only require a few measurements to characterize; one directly on-axis to verify or characterize the loudspeaker and 4 throughout the seating area to identify seat interaction and boundary affects. If the loudspeaker's directivity is not constant, then additional measurements across the seating area may be required.

If the loudspeaker has a flat on-axis amplitude response and constant directivity it will deliver a predictable direct sound to most listeners in the audience. With a typical cinema-perforated screen in place and a given amount of air loss, the direct sound will closely approximate the SMPTE ST202 high-frequency roll-off without the need for additional equalization (Engebretson and Eargle, 1982). If required, compensation for the screen loss will result in a flat direct sound, and equalization can create any other choice of target curve. If the anechoic performance is known, and screen loss is defined, then it may not be necessary to make any in situ measurements above 500Hz to 1kHz other than to verify that the loudspeaker is performing correctly. It appears that regardless of the room size, and to some extent reverberation time, if the loudspeaker displays constant directivity, then the number of measurement points above 4 becomes moot, especially above 1khz. However, it is still recommended that the measurement positions be carefully considered so that they best represent the overall layout of the cinema and are not weighed heavily towards extreme boundary positions.

6.6 Limitations of This Study

These measurements were all carried out with the loudspeaker placed at center position. If the loudspeaker had been placed at either side position, one could assume that boundary effects might come further into play, especially at low frequencies. At mid frequencies, there is a possibility of early lateral reflections from the side walls attenuating some of the seat interaction effects (Davies et al., 1996), but this phenomena would be different at different seats because of the horizontal plane reflections. However, since most cinema loudspeakers are much more directional than the JBL M2, most of the effects created by interactions with the side walls would be relegated to the low frequencies. The data in the chapter must be regarded with the understanding that the JBL M2 is not a cinema loudspeaker and that the centre position chosen only represents a portion of how a loudspeaker interacts with a cinema.

Measurements were made with the loudspeaker on a lift which could not be raised to the same height as the in-room loudspeaker had been installed. The lower height of the test loudspeaker undoubtedly added to the seat effect seen in several of the measurements. With this in mind, it is believed that if the test loudspeaker could have been raised higher, there is a chance that the measured data would have tracked the loudspeaker data closer.

Throughout this chapter a +/-3dB 'tolerance' was utilized for the various analyses of the measured signals. This tolerance was derived from the current SMPTE standard (SMPTE, 2010) and to date, this is the only known standardized tolerance window for this type of analysis that has be published. Much research has been undertaken in an effort to define the perceptual thresholds of resonances and modes that arise in loudspeakers and loudspeaker / room combinations. This research may help to establish a more defined tolerance window to be used when calibrating a cinema. However, it has been noted in prior research that the detection threshold of resonances and modes is highly dependent on Q, program material (continuous versus transient), whether the coloration is a resonance or antiresonance, the level of reverberation at the resonant / modal frequency, and modal decay (Fazenda et al., 2015), (Avis et al., 2007), (Olive et al., 1997), (Olive and Toole, 1989), (Toole and Olive, 1988), (Toole and Olive, 1986), (Bücklein, 1981), (Fryer, 1975), (Moulana, 1975). With these factors in mind, establishing a well-defined tolerance window that is acceptable for the tuning of commercial cinemas and professional dub-stages will required careful consider and further research.

6.7 Summary

In this chapter, a method of determining the number of microphone positions necessary for calibrating a cinema by utilizing anechoic data from a loudspeaker was investigated. A series of measurements using an anechoically measured and well defined loudspeaker was completed in 7 different cinemas ranging from 24 to 516 seats. Microphone positions were spatially averaged into groups of 4, 8, 12, 16, 36 and all tested positions. Various permutations of these groups were analyzed and those groups which showed the largest deviation from the anechoic data of the loudspeaker were then analyzed further. Using a frequency-by-frequency normalization method, the anechoic loudspeaker data was 'removed' from the various groupings and the resultant 'difference spectrums' were then studied individually and in comparisons between groups.

The main conclusions of these experiments are as follows:

1. If the loudspeaker has a flat on-axis amplitude response and constant directivity it will deliver a predictable direct sound to most of the seating area above 300Hz to 500Hz, with air attenuations as a function of listening distance being the main variable.
2. The room size has an effect on the loudspeaker response at the extreme high frequencies due to varying amounts of air loss, and at low frequencies due to modal behavior, acoustical absorption, loudspeaker boundary effects, and seat effects.
3. In the frequency range between 200Hz to 16kHz, the in-room responses tracks the anechoic data of the loudspeaker within +/-3dB (less above 1kHz) with the exceptions of cinemas B and F. This range can be extended lower and higher depending on room

size and how well the acoustic interferences in the low frequencies are controlled. In addition, conducting a vertical spatial averaging by raising the microphones can bring the mid-frequency data in the 200Hz and 1kHz range in better agreement with the loudspeaker data.

4. With respect to the peak-to-peak values of the difference spectrum; within any given cinema, the various position groupings were within a 1dB standard deviation between 250Hz and 20kHz, with the exception of cinema G. Between cinemas, the various position groupings' average peak-to-peak values were within 1dB from each other across all octaves though their standard deviations diverge below 250Hz and above 8kHz. The data implies that, within the frequency range of 250Hz to 8kHz and potentially higher, spatially averaging more than 4 carefully chosen microphone positions will produce only marginally more accurate data.

Chapter 7 - Summary, Conclusions and Further Work

7.1 Summary and Conclusions

The quest to standardize cinema audio began with the advent of sound accompanying picture, and continues through to today. It is a complex issue from both objective and subject standpoints; involving everything from the production through the delivery chain; equipment manufactures and content creators, to the audience itself. Though much work has been done to improve the existing standard, there are many reasons to question the basis and ongoing practices in this evolving standard. The primary aim of this research was to explore and develop alternatives to the existing practices through a better understanding of how the current standards are viewed, how they are utilized and how their end results are perceived. In this chapter, the main outcomes and conclusions of the research are summarised and potential avenues for further research are presented.

7.1.1 Re-recording Mixers' Survey of Cinema Calibration Standard

In chapter 3 a survey of re-recording mixers' use and opinions on cinema calibration standards was documented. This survey uncovered a large range of opinions on the use and efficacy of the SMPTE standard. While a majority of mixers rated the high frequency response and clarity of dialog for their dub-stage at or below average, many of them were content with the translation of their final mix to the cinema. However, the percentage of mixers who rated translated above average decreased with age. In addition, 83% of the mixers who work on SMPTE calibrated dub-stages confirmed that they used some type of compensation for the curve, typically in the high frequencies.

The advancing age and long hours in the studio raised questions regarding mixers' hearing abilities, especially in light of 34% of the mixers stating that they have not had their hearing checked for more than 5 years, or not at all. A brief overview of occupational noise standards and associated hearing research was presented along with a discussion on the potential for hearing loss among re-recording mixers. In addition, the theory of implicit learning and the possible adaptation of mixers to the X-curve were introduced.

7.1.2 Validating the Binaural Room Scanning (BRS) Method for Cinema Research

The experiments presented in chapter 3 entailed the validation of a binaural room scanning system for use in cinema research. The system, its calibration and potential sources of errors were discussed. A series of subjective listening tests were conducted to verify that the BRS system could be utilized for further experiment on preferred in-room target response curves.

Within the subjective listening test, three different sized rooms, each representative of a "real world" listening environment, were tested. Four different audio programs were chosen; three popular music tracks and one piece of cinematic score. Of the four target curves created for the test, one was derived from prior research, one mimicked the SMPTE X-curve, one served as an audible anchor and the remaining curve was flat. The findings of the listening tests were as follows:

- There were no significant differences in preference between the BRS and In situ methods in any of the three rooms.
- The main effect was due to the target curve, with curve A scoring the lowest consistently across the rooms and the ranks order of curves B, C and D remaining the same between the Eargle Theatre and the Arena Lab.
- There was an interaction between the different listening rooms, and the target curves. This interaction was confined to curve C in a comparison between the Living Room and the Eargle Theatre.
- There was an interaction between target curves B and C which could be partially attributed to individual preference.

In consideration of these results, the BRS system was deemed valid for further perceptual research on in-room response curve preference.

7.1.3 Subjective Listening Test for Preferred Room Response

In chapter 5 the BRS system validated in chapter 4 was utilized to conduct double-blind listening test on in-room target curves for 3 different seating locations each within 3 different sized cinemas.

Five different audio programs were chosen; two popular music tracks and three pieces of cinematic score, two of which were mixed on a SMPTE X-curve calibrated dub-stage. Of the five target curves created for the test, two were derived from prior research, one mimicked the

SMPTE X-curve, one served as an audible anchor and the remaining curve was flat. The findings of the listening tests were as follows:

- The main effect was due to the factor *target curve*.
- The overall preferred target curve was curve C while curve A scored extremely low even for those pieces of program material which were mixed on an X-curve calibrated dub-stage. This casts doubt on the theory of 'translation' between dub-stages and cinemas.
- There was an interaction between the different cinemas and target curves, though this mainly appears to occur with curves C and E in cinema C, which is the largest of the 3 cinemas.
- There was an interaction between the different programs and target curves, which was not completely unexpected. This interaction appears to be confined to programs Sc-T and Sc-F and is most apparent in cinema C.
- The listening position had little to no effect on preference scores. This finding was somewhat unexpected due to differences in high frequency air loss between the smaller and larger cinemas, A and C respectively. It is suspected that a type of adaptation may be occurring due to repeated exposure of the listeners to the acoustic conditions of a particular position.

A review of prior research revealed similarities between the preferred curve C from this experiment and those publicized by other researchers.

7.1.4 Predicting the In-room Response of Cinemas from Anechoic Loudspeaker Data

Chapter 6 described an alternative method of cinema calibration which utilized anechoic data from a loudspeaker to serve as the basis for comparing a cinema's in-room response. Part of this research reviewed the number of microphone positions that should be spatially averaged together in order to best represent the in-room response.

In-room measurements of 7 various sized cinemas was compared to the anechoic data from the loudspeaker which was used to make these measurements. The measurements were taken at both ear and random heights in an effort to quantify any seat effect that was occurring in the cinemas. Data was spatially averaged over all measurement positions as well as grouping of 4, 8, 12, 16 and 36 positions. An analysis was undertaken between the in-room measurements and the loudspeaker data, as well as between measurement position groupings, utilized a multi-band spectral subtraction method in order to isolate the signal created by the interaction between the loudspeaker and each cinema.

The findings of this research were as follows:

- Using anechoic loudspeaker data that includes its sound power and listening window data, along with a reasonable amount of compensation for air attenuation, provides a viable comparison point for in-room measurements. This data allows a technician who is calibrating a cinema to differentiate between frequency anomalies with the loudspeaker and those of the cinema
- A loudspeaker with flat on-axis amplitude response and constant directivity will translate directly into the cinema, allowing for a predictable response across the listening area above 300Hz to 500Hz. In this frequency range, the main variations between the in-room response and the loudspeaker data will be air attenuation which is a function of cinema size and listening distance.
- Between 200Hz to 16kHz, most of the in-room responses tracked the anechoic data of the loudspeaker within +/-3dB (less above 1kHz). The only exceptions were cinemas B, which was not built specifically for auditioning audio, and cinema F which has a very gently sloping floor. This observation held true regardless of whether all the measurement positions or only 4 positions were spatially averaged together. Smaller rooms, as well as larger rooms with well controlled low frequency responses, can extend this range beyond 200Hz and 16kHz.
- Between 200Hz and 1kHz, there may be variations between the loudspeaker data and the measured in-room response due to seat effects and other acoustic interactions. However, these variations can be lessened by raising and/or randomizing the microphone heights.
- With respect to peak-to-peak values of the difference spectrum, at frequencies below approximately 250Hz, the microphone groupings begin to diverge from each other, though only marginally. In addition, below this frequency, comparisons between

cinemas also start to diverge. Below 500Hz all of the peak-to-peak values across cinemas and position groupings rise steadily. This signifies that the relationship between the loudspeaker and the room becomes generally less correlated below this frequency.

- The main effect on the loudspeaker response at the upper frequencies in any given cinema is due to varying amounts of air loss and is seat and room specific. At low frequencies, the main effect is due to modal behavior, acoustical absorption, loudspeaker boundary effects, and seat effects, which are also room specific.

7.2 Final Summary

There were several questions that guided this research and which also tie the various chapters together. These included:

- How is the SMPTE X-curve calibration being employed in dub-stages and are re-recording mixers content with the results? Are re-recording mixers compensating for the X-curve calibration in any way and if so, how? Could mixers hearing abilities possibly affect the quality of their mixes and should they be the best judge of audio playback quality and translation between the dub-stage on the cinema?
- If the cinema-going public was given a choice as to which target response curve a cinema was calibrated to, what would be their preference? How does the size of cinema, seating position within the cinema or program affect this preference?
- To what extent do room acoustics with respect to cinema size and seat location influence the perceived sound quality of the different calibration methods?
- Can anechoic data be utilized as a tool in cinema calibration of if so, how precise is it at predicting the in-room response? When using anechoic data, how many microphone positions need to be spatially averaged together to get accurate results? Does the number of microphones that are spatially averaged together need to be scaled with cinema size?
- When calibrating a cinema, at what frequency is the room typically dominant and at what frequency is the direct sound dominant? Are these frequencies affected by room size and if so, how?

This research has demonstrated some of the weaknesses of the SMPTE ST 202 standard, its recommended calibration procedures and the use of the X-curve as an in-room target response curve.

In this thesis it has been shown that the SMPTE X-curve is compensated for by some re-recording mixers and is not preferred by listeners, even when previewing program material that was mixed on an X-curve calibrated dub-stage. The basic premise behind the standard, one of providing ideal translation, is impractical. The goal of having a film which is mixed by one person in a single dub-stage accurately translate into hundreds of cinemas, calibrated by just as many people for an audience of thousands is unachievable.

The shape of a calibration curve is often discussed with questions arising as to why any particular shape should matter; if the dub-stage and the cinema are calibrated to the same curve, good translation will prevail. However, the shape of a calibration curve should not severely limit the frequency response of a room in such a way that causes the re-recording mixer to add undue amounts of pre-emphasis during the mixing process to compensate for it. In addition, the calibration curve used should not cause the technicians to push its tolerance envelopes in an attempt to make the playback cinema sound better. Both of these issues can create too much variability in the final sound which is delivered to the listeners.

Another question that arises is the need to change the X-curve at all; stating that it has been in use for decades and there is no need to change it. As discussed in chapter 2, the shape of the X-curve first came about through compensation for the poor frequency response of playback equipment and later perpetuated through the belief that this particular shape was required to deliver a flat, neutral direct sound to the audience. Advancements in playback equipment have negated the early need to severely roll-off high and low frequencies and research in the past decade has shown that the X-curve does not deliver a flat neutral sound which is desired.

The research within this thesis has shown that a possible replacement for the X-curve is one that is preferred by listeners, which also closely mimics what naturally occurs to sound in enclosed spaces; the high frequencies dip and the low frequencies rise. In chapters 4 and 5 it was shown that this "curve", or tilted response, was preferred across cinemas sizes, seating positions and program material in addition to being preferred in headphone response, loudspeakers in smaller rooms and in automotive audio. Chapter 6 showed that a loudspeaker with a flat on-axis response and constant directivity follows the same basic curve across cinema sizes with variations in the degree of tilt below 200Hz and above 16kHz. In the 200Hz to 16kHz frequency range, a general tilting target response can be easily applied across varying cinema sizes and calibration to this response can be accomplished with as few as 4 microphone position if anechoic data on the loudspeaker is available. Below 200Hz it would be extremely difficult to use equalization on all cinemas to conform them to a target curve regardless of what shape it is; the in-room measurements are too dominated by room-specific acoustical interference in this range. To begin shaping the frequency response in this range would require that cinemas be designed or re-designed to lessen acoustical interferences. The best way forward may be to create a standard that defines a loudspeaker curve as opposed to a room curve since the room is dominated by direct sound from 200Hz to 16kHz. The frequency response and directional characteristics of the loudspeaker should also be part of this standard.

7.3 Further Work

The work in this thesis demonstrates a few of the underlying issues with the current standard for cinema calibration and makes recommendations for possible improvements. A survey of how the standards are applying by re-recording mixers has provided insight into the variability of application that may render as less than ideal final listening experience. Subjective analysis of the current target response curve recommended by the standard and widely used by the film industry showed that listeners may prefer a different curve, one which has been preferred in other listening environments. And finally, a methodology for calibrating a cinema which may lead to more accurate calibration of the audio playback system has been presented. However, there is a great deal of further work that can be performed in order to further improve cinema audio playback. Some potential avenues of further research are suggested in this section.

7.3.1 Future Research Mixer Survey

The survey presented encompasses the input of 35 participants, which is a very small representation of the world-wide re-recording mixer community. Further work can expand on this survey by administering the survey questions through local professional organizations, online forums and collaborative efforts with various universities. The hope would be to collect larger amounts of data from several countries which could be used to create a data base which tracks re-recording mixers' opinions and comments and categorizes them by demographic and work-related factors.

Additional research could be carried out to better understand the correlations between dub-stage playback levels, the amount of hearing loss in the re-recording mixers who work on these dub-stages and samples of the power spectra of mixes they create. This could give valuable insight into the effect that monitoring levels play on a mixers' hearing and the resulting spectral balance of their mixes.

7.3.2 Future Research Subjective Listening Tests

Given the amount of low frequency information in modern cinemas, a consideration for future refinement of this series of tests would be the addition of a subwoofer to the playback system. This could allow for a more "real world" simulation of a cinema, with an extended low frequency response beyond that of headphones alone.

Conducting a method of adjustment test can provide further insight into listener preference, giving the listener the opportunity to shape their own target response. The difficulty with this process is that a subjectively preferred target curve is a reflection not only of the playback system but of the program being auditioned. A useful investigation needs to include evidence of how the calibration of the dub-stage affects the spectral content of a mix. In the presented experiments the spectra of “score” (non- X-curve monitored) and the same music “score – trailer” (X-curve monitored) reveal that the only difference is the elevated low-frequency content of the trailer. Was this a result of adding impact to the trailer, or of it sounding better, or both? The fact that the high frequencies were untouched is notable. These are important issues which could drive future research.

In situ tests in carefully selected cinemas would also permit investigation of the perceptual importance of different amounts of bass build up in venues having different reflectivities, as discussed by Toole (2015). This would be helpful in leading to refinements to the target curve at low frequencies.

The decision to design the experiment as described was an effort to control variables. However, the affect that these variables might have on listener preference cannot be overlooked. With this in mind, the use of the "house" loudspeaker system, placing the loudspeakers behind the screen and utilizing a multi-channel system all provide paths for future research.

7.3.3 Future Research in Cinema Calibration

The anechoic data utilized in the research presented does not provide information on the extra loading on the loudspeaker diaphragm such as when the loudspeaker is mounted in a baffle wall, which is common in most cinemas. In addition, information on high frequency screen loss was also not provided in the anechoic data used. Obtaining anechoic data with the loudspeaker mounted in a 2pi space behind a projection screen would provide an opportunity to observe correlations between the measured in room data and anechoic data with these additional parameters added. As loudspeaker mounting techniques and screen materials vary from cinema to cinema, care must be used when determining a representative model for collecting such anechoic data.

Future tests could also identify the ideal properties of cinema loudspeakers and how these characteristics tie to subjective data. Research by Toole, Olive and Devantier have shown that listener preferences for loudspeakers can be correlated to their anechoic measurements (Olive, 2004b, Olive, 2004a), (Devantier, 2002), (Toole, 1986a, Toole, 1986b). In addition, the research in this thesis has revealed that the in-room response of a loudspeaker can be predicted from its anechoic data within +/-3dB (or less) down to approximately 200Hz. Therefore, one should be able to predict listeners' preference of a loudspeaker within a given cinema from its anechoic data.

References

ACOEM 2007. The Role of the Professional Supervisor in the Audiometric Testing Component of Hearing Conservation Programs. *Policies & Position Statements.* Elk Grove Village, IL USA: Amercian College of Occupational and Environmental Medicine.

ALDRED, J. 1997. 100 Years of Cinema Loudspeakers. *The Newsletter of the Association of Motion Picture Sound,* 3**,** 9-14.

ALGAZI, V., DUDA, R., MORRISON, R. & THOMPSON, D. 2001. Structural Composition and Decomposition of HRTFs. *IEEE Workshop on Applications of Signal Processing to Audio and Acoustics.* New Paltz, New York: Institute of Electrical and Electronics Engineers.

ALLEN, I. 1975. The Production of Wide-Range, Low-Distortion Optical Soundtracks Utilizing the Dolby Noise Reduction System. *SMPTE Motion Imaging Journal,* 84**,** 720-729.

ALLEN, I. 2006. The X-Curve: Its Origins and History: Electro-Acoustic Characteristics in the Cinema and the Mix-Room, the Large Room and the Small. *SMPTE Motion Imaging Journal,* 115**,** 264-275.

ANDREOPOULOU, A. & KATZ, B. F. On the Use of Subjective HRTF Evaluations for Creating Global Perceptual Similarity Metrics of Assessors and Assessees. The 21st International Conference on Auditory Display - ICAD, 2015 Graz, Austria. Georgia Institute of Technology.

ANJOS, W. T. D., LUDIMILA, L., RESENDE, L. M. D. & COSTA-GUARISCO, L. P. 2014. Correlation Between the Hearing Loss Classifications and Speech Recognition. *Revista CEFAC,* 16**,** 1109-1116.

ANSI 1996a. ANSI 3.44-1996 Determination of Occupational Noise Exposure and Estimation of Noise-Induced Hearing Impairment. *Acoustics.* Washington, DC USA: American National Standards Insitute.

ANSI 1996b. ANSI S3.6-1996 American National Standard Specification for Audiometers. Washington, DC USA: American National Standards Insitute.

ANSI 2016. ANSI/ASA S3.44-2016 Estimation of Noise-Induced Hearing Loss – Part 1: Method for Calculating Expected Noise-induced Permanent Threshold Shift *Acoustics.* Washington, DC USA: American National Standards Institute.

ANSI/CEA 2015. ANSI/CEA-2034-A Standard Method of Measurement for In-Home Loudspeakers. Arlington, VA USA: Consumer Technology Association.

AVIS, M. R., FAZENDA, B. M. & DAVIES, W. J. 2007. Thresholds of Detection for Changes to the Q Factor of Low-Frequency Modes in Listening Environments. *Journal of the Audio Engineering Society,* 55**,** 611-622.

BADDELEY, A. D., EYSENCK, M. W. & ANDERSON, M. C. 2014. *Memory,* Hove, UK, Psychology Press.

BECH, S. 1990. Listening Tests on Loudspeakers: A Discussion of Experimental Procedures and Evaluation of the Response Data. *AES 8th International Conference: The Sound of Audio.* Washington, DC USA: Audio Engineering Society.

BECH, S. 1992. Selection and Training of Subjects for Listening Tests on Sound-Reproducing Equipment. *Journal of the Audio Engineering Society,* 40**,** 590-610.

BECH, S., GULBOL, M. A., MARTIN, G., GHANO, J. & ELLERMEIER, W. A Listening Test System for Automotive Audio - Part 2: System Initial Verification. AES 118th Convention, May 28-31, 2005 2005 Barcelona, Spain. Audio Engineering Society.

BEDELL, E. H. & KERNEY, I. 1934. System Adaptation. *Bell System Technical Journal,* 13**,** 301-308.

BEGAULT, D., WENZEL, E. & ANDERSON, M. 2001. Direct Comparison of the Impact of Head Tracking, Reveberation, and Individualized Head-Related Transder Functions on the Spatial Perception of a Virtual Speech Source. *Journal of the Audio Engineering Society,* 49**,** 904-916.

BENJAMIN, E. Extending Quasi-Anechoic Measurements to Low Frequencies. AES 117th Convention, 2004 San Francisco, CA USA. Audio Engineering Society.

BERESFORD, K., FORD, N., RUMSEY, F. & ZIELINSKI, S. Contextual Effects on Sound Quality Judgements: Listening Room and Automotive Environments AES 120th Convention, 2006a Paris, France. Audio Engineering Society.

BERESFORD, K., FORD, N., RUMSEY, F. & ZIELINSKI, S. Contextual Effects on Sound Quality Judgements: Part II – Multi-Stimulus vs. Single Stimulus Method. AES 121st Convention, 2006b San Francisco, CA USA. Audio Engineering Society.

BERG, J. & RUMSEY, F. Correlation Between Emotive, Descriptive and Naturalness Attributes in Subjective Data Relating to Spatial Sound Reproduction. Audio Engineering Society Convention 109, 2000. Audio Engineering Society.

BIBRA, G. 2008. *Film Format Guide* [Online]. Available: http://www.dvdlog.de/filmformate/filmformate-en.htm [Accessed June 20th, 2013].

BIERLEY, C., MCSWEENEY, F. K. & VANNIEUWKERK, R. 1985. Classical Conditioning of Preferences for Stimuli. *Journal of Consumer Research,* 12**,** 316-323.

BLAUERT, J. 1997. *Spatial Hearing,* Cambridge, MA USA, The MIT Press.

BLAUERT, J. 2005. *Communication Acoustics,* New York, NY USA, Springer Publishing.

BONER, C. P. & BONER, C. R. 1965. Minimizing Feedback in Sound Systems and Room-Ring Modes with Passive Networks. *The Journal of the Acoustical Society of America,* 37**,** 131-135.

BONER, C. R. Some Examples of Sound-system Correction of Acoustically Difficult Rooms. AES 31st Convention, 1966 New York, NY USA. Audio Engineering Society.

BRADLEY, J. S. 1991. Some Further Investigations of the Seat Dip Effect. *The Journal of the Acoustical Society of America,* 90**,** 324-333.

BRIDGES, S. Effects of Direct Sound on Perceived Frequency Response of a Sound System.pdf. AES 66th Convention, 1980 Los Angeles, CA USA. Audio Engineering Sociery.

BRONKHORST, A. W. 1995. Localization of real and virtual sound sources. *The Journal of the Acoustical Society of America,* 98**,** 2542-2553.

BRUNGART, D. S., A.J., K. & SIMPSON, B. D. 2006. Effects of Headtracker Latency in Virtual Audio Displays. *The Journal of the Acoustical Society of America,* 54**,** 32-44

BRUNNER, S., MAEMPEL, H.-J. & WEINZIERL, S. On the Audibility of Comb Filter Distortions. Audio Engineering Society Convention 122, 2007. Audio Engineering Society.

BUCKLE, C. G. & LUMKIN, A. W. 1969. The Evaluation and Standardization of the Loudspeaker-Acoustics Link in Motion-Picture Theaters. *SMPTE Motion Imaging Journal,* 78**,** 1058-1063.

BÜCKLEIN, R. 1981. The Audibility of Frequency Response Irregularities. *Journal of the Audio Engineering Society,* 29**,** 126-131.

CARLISLE, R. W. & SCHWARTZ, A. 1959. Evaluation of a Stereophonic Loudspeaker by Multiple Microphone Arrays. *The Journal of the Acoustical Society of America,* 31**,** 1348-1351.

CENGARLE, G. & MATEOS, T. Effect of Microphone Number and Positioning on the Average of Frequency Responses in Cinema Calibration. AES 136th Convention, 2014 Berlin, Germany. Audio Engineering Society.

CERNA, M. & HARVEY, A. F. 2000. The Fundamentals of FFT-Based Signal Analysis and Measurement. Austin, Texas USA: National Instruments.

CHOI, Y.-J., BRADLEY, J. S. & JEONG, D.-U. 2015. Experimental Investigation of Chair Type, Row Spacing, Occupants, and Carpet on Theatre Chair Absorption. *Journal of the Acoustical Society of America,* 137**,** 105-116.

CHRISTENSEN, F., LYDOLF, M., MARTIN, G., MINNAAR, P., PEDERSEN, B. & SONG, W.-K. A Listening Test System for Automotive Audio - Part 1: System Description. AES 118th Convention, 2005 Barcelona, Spain.

COHEN, J. 1988. *Statistical Power Analysis for the Behavioural Sciences* New York, NY USA, Routledge.

CONNOR, W. K. 1967. Theoretical and Practical Considerations in the Equalization of Sound Systems. *Journal of the Audio Engineering Society,* 15**,** 194-198.

COREN, S. & HAKSTIAN, A. R. 1994. Predicting Speech Recognition Thresholds From Pure Tone Hearing Thresholds. *Perceptual and motor skills,* 79**,** 1003-1008.

CRUZ, R. 2005. *Hearing Conservation for Audio Industry Professionals* [Online]. The Hearing Review. Available: http://www.hearingreview.com/2005/10/hearing-conservation-for-audio-industry-professionals/ [Accessed August 26th 2013].

CRUZ, R. & FISHER, L. 2004. Audiometric Profiles of Professional Musicians, Audio Engineers, & Music Industry Personnel. *American Auditory Society Scientific Meeting.* Scottsdale, AZ USA.

DACTRON 2003. Understanding FFT windows. *Application Note AN014.* Milpitas, CA USA: LDS Group.

DAVIES, W. J., COX, T. J. & LAM, Y. W. 1996. Subjective Perception of Seat Dip Attenuation. *Acta Acustica united with Acustica* 82**,** 784-792.

DAVIES, W. J. & LAM, Y. W. 1994. New Attributes of Seat Dip Attenuation. *Applied Acoustics,* 41**,** 1-23.

DECARLO, L. T. 1997. On the Meaning and Use of Kurtosis. *Psychological Methods,* 2**,** 292.

DENNIS, P. In-Vehicle Audio System Sound Quality Preference Study. AES 139th Convention, 2015 New York, NY USA. Audio Engineering Society.

DEVANTIER, A. Charaterizing the Amplitude Response of Loudspeaker Systems. AES 113th Convention, 2002 Los Angeles, CA USA. Audio Engineering Society.

DEVANTIER, A. 2011. HATS User Manual. Northridge, CA USA: Harman International.

DOLBY 1994. Technical Guidelines for Dolby Stereo Theatres. San Francisco, CA USA: Dolby Laboratories.

DOLBY 2015. Dolby Atmos® Specifications. San Francisco, CA USA: Dolby Laboritories, Inc.

DOUGHERTY, D. 2009. The Sizzling Sound of Music. *Radar* [Online]. Available: http://radar.oreilly.com/2009/03/the-sizzling-sound-of-music.html [Accessed October 21st, 2015].

DUBNO, J. R., DIRKS, D. D. & MORGAN, D. E. 1984. Effects of Age and Mild Hearing Loss on Speech Recognition in Noise. *The Journal of the Acoustical Society of America,* 76**,** 87-96.

DURST, F. & SHORTT, E. J. 1939. Characteristics of Film Reproducer Systems. *SMPTE Motion Imaging Journal,* 32**,** 169-187.

EARGLE, J. 2007. *The JBL Story - 60 Years of Audio Innovation,* Milwakee, WI USA, Hal Leonard Corp.

EARSCAN. 2013. *Earscan 3 Automatic Threshold Audiometer* [Online]. Murphy, NC USA: Micro Audiometrics Corporation. Available: http://www.microaud.com/pdf/brochures/Earscan3_ES3.pdf [Accessed January 4th, 2014].

ELLIOTT, T. M., HAMILTON, L. S. & THEUNISSEN, F. E. 2013. Acoustic Structure of the Five Perceptual Dimensions of Timbre in Orchestral Instrument Tones. *The Journal of the Acoustical Society of America,* 133**,** 389-404.

ENGEBRETSON, M. & EARGLE, J. 1982. Cinema Sound Reproduction Systems: Technology Advances and System Design Considerations. *SMPTE Motion Imaging Journal,* 91**,** 1046-1057.

FARINA, A. Simultaneous Measurement of Impulse Response and Distortion with a Swept-Sine Technique AES 108th Convention, 2000 Paris, France. Audio Engineering Society.

FAZENDA, B. M., STEPHENSON, M. & GOLDBERG, A. 2015. Perceptual Thresholds for the Effects of Room Modes as a Function of Modal Decay. *The Journal of the Acoustical Society of America,* 137**,** 1088-1098.

FIELD, A. 2013. *Discovering Statistics Using IBM SPSS Statistics,* Thousand Oaks, CA USA, Sage Publishing.

FIELDER, L. D. Practical Limits for Room Equalization. AES 111th Convention, 2001 New York, NY USA. Audio Engineering Society.

FIELDER, L. D. 2012. Frequency Response Versus Time-of-Arrival for Typical Cinemas. *SMPTE Annual Technical Conference & Exhibition.* Hollywood, CA USA: SMPTE.

FIELDING, R. 1983. *A Technological History of Motion Pictures and Television,* Berkeley, CA USA, University of California Press.

FLEISCHMANN, F., SILZLE, A. & PLOGSTIES, J. Identification and Evaluation of Target Curves for Headphones. AES 133rd Convention, 2012 San Francisco, CA USA. Audio Engineering Society.

FLETCHER, H. 1950. A Method of Calculating Hearing Loss for Speech from an Audiogram. *Acta Oto-Laryngologica,* 38**,** 26-37.

FRYER, P. A. Intermodulation Distortion Listening Tests. Audio Engineering Society Convention 50, 1975. Audio Engineering Society.

GABRIELSSON, A. 1979. Dimension Analyses of Perceived Sound Quality of Sound-Reproducing Systems. *Scandinavian Journal of Psychology,* 20**,** 159-169.

GABRIELSSON, A., ROSENBERG, U. & SJÖGREN, H. 1974. Judgments and Dimension Analyses of Perceived Sound Quality of Sound-Reproducing Systems. *The Journal of the Acoustical Society of America,* 55**,** 854-861.

GEDEMER, L. 2015a. Predicting the In-room Response of Cinemas from Anechoic Loudspeaker Data. *AES 57th International Conference on The Future of Audio Entertainment Technology – Cinema, Television and the Internet.* Hollywood, CA, USA Audio Engineering Society.

GEDEMER, L. Subjective Listening Test for Preferred Room Response in Cinemas - Part 1: System and Test Descriptions. AES 139th Convention, 2015b New York, NY USA. Audio Engineering Society.

GEDEMER, L. & WELTI, T. 2013a. Validation of the Binaural Room Scanning Method for Cinema Audio Research *Reproduced Sound: Delivering the Experience - From Imagination to Realisation in Audio.* Manchester, UK: Insitute of Acoustics.

GEDEMER, L. & WELTI, T. Validation of the Binaural Room Scanning Method for Cinema Audio Research. AES 135th Convention, 2013b New York, NY USA. Audio Engineering Society.

GENEREUX, R. 1992. Signal Processing Considerations for Acoustic Environment Correction. *AES 7th Conference: Digital Signal Processing.* London, UK: Audio Engineering Society.

GRANIER, E. Comparing and Optimizing Audio Systems in Cars. AES 100th Convention, 1996 Copenhagen, Denmark. Audio Engineering Society.

GROS, L. & CHATEU, N. 2002. The Impact of Listening and Conversational Situations on Speech Perceived Quality for Time-Varying Impairments. *Measurement of Speech and Audio Quality In Networks - International On-Line Workshop.* Prague, Czech Repulic: Czech Technical University.

HAMMERSHØI, D. & MÖLLER, H. 1996. Sound Transmission To and Within the Human Ear Canal. *The Journal of the Acoustical Society of America,* 100**,** 408-427.

HARMAN. 2011. *How to Listen* [Online]. Available: http://harmanhowtolisten.blogspot.com/ [Accessed June 6 2014].

HARRIS, F. 1978. On the Use of Windows for Harmonic Analysis with the Discrete Fourier Transform. *Proceedings of the IEEE,* 66**,** 51-83.

HARRIS, L. & HOLLAND, K. 2009. Using Statistics to Analyse Listening Test Data: Some Sources and Advice for Non-Statisticians. *Reproduced Sound: The Audio Explosion.* Brighton, UK: Institute of Acoustics.

HIEKKANEN, T., MAKIVIRTA, A. & KARJALAINEN, M. 2009. Virtualized Listening Tests for Loudpeakers. *Journal of the Audio Engineering Society,* 57**,** 237-251.

HILL, A. J., HAWKSFORD, M. O. & NEWELL, P. 2016. Enhanced Wide-Area Low-Frequency Sound Reproduction in Cinemas: Effective and Practical Alternatives to Current Calibration Strategies. *Journal of the Audio Engineering Society,* 64**,** 280-298.

HILLIARD, J. K. 1936. A Study of Theater Loud Speakers and the Resultant Development of the Shearer Two-Way Horn System. *SMPTE Motion Imaging Journal,* 27**,** 45-60.

HILLIARD, J. K. 1949. Theater Loudspeaker Design, Performance, and Measurement. *SMPTE Motion Imaging Journal,* 52**,** 629-640.

HILLIARD, J. K. 1985. A Brief History of Early Motion Picture Sound Recording and Reproducing Practices. *Journal of the Audio Engineering Society,* 33**,** 271-278.

HOLMAN, T. 1991. New Factors in Sound for Cinema and Television. *Journal of the Audio Engineering Society,* 39**,** 529-539.

HOLMAN, T. 1993. *Timbre Correction Units for Use in Sound Systems*. 707,116.

HOLMAN, T. 2007. Cinema Electro-Acoustic Quality Redux. *SMPTE Motion Imaging Journal,* 116**,** 220-233.

HOLMAN, T. 2010. *Sound for Film and Television,* New York, NY USA, Focal Press.

HORBACH, U., KARAMUSTAFAOGLU, A., PELLEGRINI, R., MACKENSEN, P. & THEILE, G. Design and Applications of a Data-Based Auralization System for Surround Sound. AES 106th Convention, 1999 Muich, Germany. Audio Engineering Society.

HUMES, L. E. 2015. Age-Related Hearing Loss, Noise-Induced Hearing Loss, and Speech-Understanding Performance. *The Journal of the Acoustical Society of America,* 137**,** 2248-2248.

HUMMERSONE, C. 2013. *Irstats: Calculate RT, DRR, CTE, and EDT for Impulse Response File. A MATLAB File* [Online]. Available: https://www.mathworks.com/matlabcentral/fileexchange/42566-impulse-response-acoustic-information-calculator [Accessed July 2nd, 2015].

IEC 2012. IEC 61094-8:2012 Measurement Microphones - Part 8: Methods for Detemining the Free-Field Sensitivity of Working Standard Microphones by Comparison *Measurement Microphones.* Geneva, Switzerland: International Electrotechnical Commission.

IEC, D. 2011. TS 60318–7: 2011, Electroacoustics—Simulators of Human Head and Ear—Part 7: Head and Torso Simulator for the Measurement of Hearing Aids. *Berlin, Germany: Beuth*.

IMDB. 2013. *Internet Movie Database* [Online]. Available: www.imdb.com [Accessed June 6th, 2013].

ISO 1975a. ISO 454 Relation Between Sound Pressure Level of Narrow Bands of Noise in Diffuse Field and in a Frontally Incident Free Field for Equal Loudness. *Acoustics.* Geneva, Switzerland: International Organisation for Standardisation.

ISO 1975b. ISO DIS2629.2 Specification and Measurements for the Electro-Acoustic Response of Motion-Picture Control Rooms and Indoor Theatres. *Sound.* Geneva, Switzerland: International Organization for Standardization.

ISO 1977. ISO 2969 Specification and Measurements for the Electro-Acoustic Response of Motion-Picture Control Rooms and Indoor Theatres. *Sound.* Geneva, Switzerland: International Organization for Standardization.

ISO 1987. ISO 2969:1987 B-chain Electro-Acoustic Response of Motion-Picture Control Rooms and Indoor Theaters. *Sound.* Sound: International Organization for Standardization.

ISO 1993. 9613–1:1993 Attenuation of Sound During Propagation Outdoors. Part 1: Calculation of the Absorption of Sound by the Atmosphere. *Acoustics* Geneva, Switzerland: International Organization for Standardization.

ISO 1996. 9613-2:1996 Attenuation of Sound During Propagation Outdoors–Part 2: General Method of Calculation. *Acoustics.* Geneva, Switzerland: International Organization for Standardization.

ISO 1998. ISO 389-1:1998 Reference Equivalent Threshold Sound Pressure Levels for Pure Tones and Supra-Aural Earphones. *Acoustics-Reference Zero for the Calibration of*

Audiometric Equipment Geneva, Switzerland: International Organization for Standardization.

ISO 2005. ISO 389-7:2005 Reference Threshold of Hearing Under Free-Field and Diffuse-Field Listening Conditions. *Acoustics-Reference Zero for the Calibration of Audiometric Equipment.* Geneva, Switzerland: International Organization for Standardization.

ISO 2009. ISO 3382-1:2009 Acoustics - Measurement of Room Acoustic Parameters. *Part 1: Performance Spaces.* Geneva, Switzerland: International Organization for Standardization.

ISO 2016. ISO 389-7 2006 Reference Threshold of Hearing Under Free-Field and Diffuse-Field Listening Conditions - Ammendment 1. *Acoustics-Reference Zero for the Calibration of Audiometric Equipment.* Geneva, Switzerland: International Organization for Standardization.

ITU 1987. ITU-R BS.468-4 Measurement of Audio-Frequency Noise Voltage Level in Sound Broadcasting *Broadcasting Service (Sound).* Geneva, Switzerland: International Telecommunication Union.

ITU 1997. ITU-R BS.1116-1 Methods for the Subjective Assessment of Small Impairments in Audio Systems Including Multichannel Sound Systems. *Broadcasting Service (Sound).* Geneva, Switzerland: International Telecommunication Union.

ITU 2001. ITU-R BS.1534-1 Method for the Subjective Assessment of Intermediate Quality Level of Audio Systems. *Broadcasting Service (Sound).* Geneva, Switzerland: International Telecommunication Union.

ITU 2012a. ITU-R BS.775-3 Multichannel Stereophonic Sound System With and Withoug Accompanying Picture. *Broadcasting Service (Sound).* Geneva, Switzerland: International Telecommunication Union.

ITU 2012b. ITU-R BS.1770-3 Algorithms to Measure Audio Programme Loudness and True-Peak Audio Level. *Broadcasting Service (Sound).* Geneva, Switzerland: International Telecommunication Union.

ITU 2015. ITU-R BS. 1534-3 Method for Subjective Assessment of Intermediate Quality Level of Coding Systems. *Broadcasting Service (Sound).* Geneva, Switzerland: International Telecommunications Union.

JACCARD, J., BECKER, M. A. & WOOD, G. 1984. Pairwise Multiple Comparison Procedures: A Review. *Psychological Bulletin,* 96**,** 589.

JBL 2003. Cinema Sound System Manual. Northridge, CA USA: Harman International.

KELLOGG, E. W. 1930. Some New Aspects of Reverberation. *Journal of the Society of Motion Picture Engineers,* 14**,** 96-107.

KELLOGG, E. W. 1955. Second Installment: History of Sound Motion Pictures. *SMPTE Motion Imaging Journal,* 64**,** 356-374.

KEPPEL, G. & WICKENS, T. D. 2004. *Design and Analysis: A Researcher's Handbook,* Englewood Cliffs, NJ USA, Prentice Hall.

KIM, S., DE FRANCISCO, M., WALKER, K., MARUI, A. & MARTENS, W. L. 2006. An Examination of the Influence of Musical Selection on Listener Preferences for Multichannel Microphone Technique. *AES 28th International Conference: The Future of Audio Technology--Surround and Beyond.* Pitea, Sweden: Audio Engineering Society.

KIRK, R. 2013. *Experimental Design: Procedures for the Behavioral Sciences,* Thousand Oaks, CA USA, Sage Publication.

KIRK, R. E. 1956. Learning, a Major Factor Influencing Preferences for High-Fidelity Reproducing Systems. *The Journal of the Acoustical Society of America,* 28**,** 1113-1116.

KIRK, R. E. 1994. *Experimental Design: Procedures for the Social Sciences,* Belmont, CA USA, Wadsworth Publishing.

KNUDSEN, V. O. 1932. *Architectural Acoustics,* London, UK, Chapman & Hall.

KOEHL, V., PAQUIER, M. & DELIKARIS-MANIAS, S. Comparison of Subjective Assessments Obtained from Listening Tests Through Headphones and Loudspeaker Setups. AES 131st Convention, 2011 New York, New York USA. Audio Engineering Society.

KORN, T. S. & HOUGARDY, J. 1959. Mesures des Haut-Parleurs dans les Pieux d'Utilisation. *Acta Acustica united with Acustica,* 9**,** 121-126.

LAITINEN, M. V., PIHLAJAMAKI, T., LOSLER, S. & VILLE PULKKI, V. Influence of Resolution of Head Tracking in Synthesis of Binaural Audio. AES 126th Convention, 2012 Budapest, Hungary. Audio Engineering Society.

LE BAGOUSSE, S., PAQUIER, M., COLOMES, C. & MOULIN, S. Sound Quality Evaluation Based on Attributes-Application to Binaural Contents. Audio Engineering Society Convention 131, 2011. Audio Engineering Society.

LEEMBRUGGEN, G., NEWELL, P., NEWELL, J., GILFILLAN, D., HOLLAND, K. & MCCARTY, B. 2011. Is X Curve Damaging Our Enjoyment of Cinema. *SMPTE Annual Technical Conference & Exhibition.* Hollywood, CA USA: SMPTE.

LINDAU, A. 2009. The Perception of System Latency in Dynamic Binaural Synthesis. *Proc. of 35th DAGA***,** 1063-1066.

LINDAU, A. 2014. *Binaural Resynthesis of Acoustical Environments. Technology and Perceptual Evaluation.* PhD Thesis, Technical University Berlin.

LJUNGBERG, L. 1969. Standardized Sound Reproduction in Cinemas and Control Rooms. *SMPTE Motion Imaging Journal,* 78**,** 1046-1053.

LONG, B., SCHWENKE, R., SOPER, P. & LEEMBRUGGEN, G. 2012. Further Investigations Into the Interactions Between Cinema Loudspeakers and Screens. *SMPTE Motion Imaging Journal,* 121**,** 46-62.

LOYE, D. P. 1941. Acoustic Design Features of Studio Stages, Monitor Rooms, and Review Rooms. *SMPTE Motion Imaging Journal,* 36**,** 593-621.

LUBMAN, D. 1971. Spatial Averaging in Sound Power Measurements. *Journal of Sound and Vibration,* 16**,** 43-58.

MAAMAR, A., KALE, I., KRUKOWSKI, A. & DAOUD, B. 2006. Partial Equalization of Non-Minimum-Phase Impulse Responses. *EURASIP Journal on Advances in Signal Processing,* 2006**,** 1-9.

MACKENSEN, P., FLEDERHOF, U., THEILE, G. & PELLEGRINI, J. 1999. Binaural Room Scanning - A New Tool for Acoustic and Psychoacoustic Research. *The Journal of the Acoustical Society of America,* 105**,** 1343-1344.

MÄKIVIRTA, A. & ANET, C. A Survey Study of In-Situ Stereo and Multi-Channel Monitoring Conditions. AES 111th Convention, 2001 New York, NY USA. Audio Engineering Society.

MARQUES, A. & FREITAS, D. 2005. Infinite Impulse Response (IIR) Inverse Filter Design for the Equalization of Non-Minimum Phase Loudspeaker Systems. *IEEE Workshop on Applications of Signal Processing to Audio and Acoustics.* New Paltz, NY USA: IEEE.

MARTENS, W. L., MARUI, A. & KIM, S. 2006. Investigating Contextual Dependency in a Pairwise Preference Choice Task. *AES 28th International Conference: The Future of Audio Technology--Surround and Beyond.* Pitea, Sweden: Audio Engineering Society.

MATHUR, N. & ROLAND, P. 2016. Noise-Induced Hearing Loss Clinical Presentation. *Medscape* [Online]. Available: http://emedicine.medscape.com/article/857813-clinical#b1 [Accessed July 23rd, 2016].

MATTILA, V.-V. Descriptive Analysis of Speech Quality in Mobile Communications: Descriptive Language Development and External Preference Mapping. Audio Engineering Society Convention 111, 2001. Audio Engineering Society.

MAUCHLY, J. W. 1940. Significance Test for Sphericity of a Normal N-Variate Distribution. *The Annals of Mathematical Statistics,* 11**,** 204-209.

MCGRAW, K. O. & WONG, S. 1992. A Common Language Effect Size Statistic. *Psychological Bulletin,* 111**,** 361.

MCHUGH, M. L. 2011. Multiple Comparison Analysis Testing in ANOVA. *Biochemia Medica,* 21**,** 203-209.

MILLER, J. 2016. *2016 Is on Track to Be Hollywood's Worst Year for Ticket Sales in a Century* [Online]. Vanity Fair Hollywood. Available: http://www.vanityfair.com/hollywood/2016/06/hollywood-movie-ticket-sales [Accessed July 27th 2016].

MINNAAR, P., OLESEN, S., CHRISTENSEN, F. & MØLLER, H. 2001. Localization With Binaural Recordings from Artificial and Human Heads. *Journal of the Audio Engineering Society,* 49**,** 323-336.

MØLLER, H. 1974. Relevant Loudspeaker Tests in Studios, in Hi-Fi Dealers' Demo Rooms, in the Home etc. - Using 1/3 Octave, Pink Weighted, Random Noise. *AES 47th Convention* Copenhagen, Denmark: Audio Engineering Society.

MØLLER, H., HAMMERSHØI, D., JENSEN, C. B. & SØRENSEN, M. F. 1995. Transfer Characteristics of Headphones Measured on Human Ears. *Journal of the Audio Engineering Society,* 43**,** 203-217.

MOORE, B. C. 2012. *An Introduction to the Psychology of Hearing,* Bingley, UK, Emerald Group Publishing Ltd.

MOORE, B. C. & GLASBERG, B. R. 1983. Suggested Formulae for Calculating Auditory-Filter Bandwidths and Excitation Patterns. *The Journal of the Acoustical Society of America,* 74**,** 750-753.

MOORE, B. C. & TAN, C. T. 2004. Development and Validation of a Method for Predicting the Perceived Naturalness of Sounds Subjected to Spectral Distortion. *Journal of the Audio Engineering Society,* 52**,** 900-914.

MOORE, D. R., EDMONDSON-JONES, M., DAWES, P., FORTNUM, H., MCCORMACK, A., PIERZYCKI, R. H. & MUNRO, K. J. 2014. Relation Between Speech-in-Noise Threshold, Hearing Loss and Cognition from 40–69 Years of Age. *PloS one,* 9**,** e107720.

MOULANA, K. 1975. *Tonal Colouration Caused by a Single Resonance.* University of Surrey.

MPRC 1937a. Research Council Adopts Standard Electrical Characteristic for Two Way Reproducing System in Theatres. *Technical Bulletin of the Academy of Motion Picture Arts & Sciences.* Hollywood, CA USA: Academy of Motion Picture Arts & Sciences.

MPRC 1937b. Revised Standard Electrical Characteristic for Two Way Reproducing Systems in Theatres. *Technincal Bulletin of the Academy of Motion Picture Arts & Sciences.* Hollywood, CA USA: Academy of Motion Picture Arts & Sciences.

MPRC 1938. Revised Standard Electrical Characteristics for Two-Way Reproducing Systems in Theatres. *Technical Bulletin of the Academy of Motion Picture Arts & Sciences.* Hollywood, CA USA: Academy of Motion Picture Arts & Sciences.

MPRC 1948. Standard Electrical Characteristics for Theatre Sound Systems. *Technical Bulletin of the Academy of Motion Picture Arts & Sciences* Hollywood, CA USA: Academy of Motion Picture Arts & Sciences.

MUNRO, A. Room Acoustics and Equalisation of Speaker Systems for Multipurpose Theatres. 116th AES Convention 2004 Berlin, Germany. Audio Engineering Society.

NATO. 2013. National Association of Theatre Owners. Available: http://www.natoonline.org/2013/eoe [Accessed June 26th 2013].

NEWELL, J., NEWELL, P. & HOLLAND, K. 2015a. Considerations for the Generation and Measurement of Low Frequency Effects in Cinema Rooms. *AES 57th International Conference: The Future of Audio Entertainment Technology–Cinema, Television and the Internet.* Hollywood, CA USA: Audio Engineering Society.

NEWELL, P., DOMINGUEZ, D. S., HOLLAND, K., TORRES-GUIJARRO, S. & CASTRO, S. 2015b. The Relationship Between Subjective and Objective Response Differences at Different Heights Above Cinema Seating. *Reproduced Sound: Playing with Fire, Sound for Art, Entertainment and Emergency Purposes.* Moreton-in-Marsh, UK: Institute of Acoustics.

NEWELL, P., GARCIA, J. G. & HOLLAND, K. 2013. An Investigation Into the Acoustic Effect of Cinema Screens on Loudspeaker Performance. *Reproduced Sound: Delivering the Experience - From Imagination to Realisation in Audio.* Manchester, UK: Institute of Acoustics.

NEWELL, P., HOLLAND, K., NEWELL, J. & NESKOV, B. 2011a. New Proposals for the Calibration of Sound in Cinema Rooms. *AES 130th Convention.* London, UK: Audio Engineering Society.

NEWELL, P., HOLLAND, K., TORRES-GUIJARRO, S., CASTRO, S. & VALDIGEM, E. 2010. Cinema Sound: A New Look at Old Concepts. *Reproduced Sound: Performing Arts Venue: Balancing the Design* Cardiff, Wales UK: Institute of Acoustics.

NEWELL, P., LEEMBRUGGEN, G., HOLLAND, K., NEWELL, J., TORRES-GUIJARRO, S., GILFILLAN, D., SANTOS-DOMINGUEZ, D. & CASTRO, S. 2011b. Does 1/3 Octave Equalisation Improve the Sound in a Typical Cinema?'. *Reproduced Sound: Sound Systems: Engineering or Art?* Brighton, UK: Institute of Acoustics.

NFTU 1969. 1069-1969 Recommended Practice for Measuring Overall Frequency Response in Cinemas for Optical Sound 35mm. Copenhagen, Denmark: Scandinavian Film and TV Society - Nordic Film / TV Union.

NICOL, R., GROS, L., COLOMES, C., NOISTERNIG, M., WARUSFEL, O., BAHU, H., KATZ, B. F. & SIMON, L. S. 2014. A Roadmap for Assessing the Quality of Experience of 3d Audio Binaural Rendering.

NIDCD. 2016. *Quick Statistics About Hearing* [Online]. Bethesda, MD USA: U.S. Department of Health & Human Service. Available: https://www.nidcd.nih.gov/health/statistics/quick-statistics-hearing [Accessed July 6th, 2016].

OLIVE, S. A Method For Training Listeners and Selecting Program Material for Listening Tests. AES 97th Convention, 1994 San Francisco, CA USA. Audio Engineering Society.

OLIVE, S. 2003. Differences in Performance and Preference of Trained Versus Untrained Listeners in Loudspeaker Tests A Case Study. *Journal of the Audio Engineering Society,* 51**,** 806-825.

OLIVE, S. A Multi Regression Model for Predicting Loudspeaker Preference Using Objective Measurements: Part 2 - Development of the Model. AES 117th Convention, 2004a San Francisco, CA, USA. Audio Engineering Society.

OLIVE, S. A Multiple Regression Model for Predicting Loudspeaker Preference Using Objective Measurements: Part 1 - Listening Test Results. AES 116th Convention, 2004b Berlin, Germany. Audio Engineering Society.

OLIVE, S., HESS, S. & DEVANTIER, A. Comparison of Loudspeaker-Room Equalization Preference for Multichannel, Streo and Mono Reproductions - Are Listeners More Discriminating in Mono. AES 124th Convention, 2008 Amsterdam, The Netherlands. Audio Engineering Society.

OLIVE, S., JACKSON, J., DEVANTIER, A., HUNT, A. & HESS, S. The Subjective and Objective Evaluation of Room Correction Products. AES 127th Convention, 2009 New York, NY. Audio Engineering Society.

OLIVE, S. & MARTENS, W. Interaction Between Loudspeakers and Room Acoustics Influences Loudspeaker Preferences in Multichannel Audio Reproduction AES 123rd Convention, October 5-8 2007 New York, NY USA. Audio Engineering Society.

OLIVE, S., SCHUCK, P., SALLY, S. & BONNEVILLE, M. 1994. The Effects of Loudspeaker Placement on Listener Preference Ratings. *Journal of the Audio Engineering Society,* 42**,** 651-669.

OLIVE, S., SCHUCK, P., SALLY, S. & BONNEVILLE, M. The Variability of Loudspeaker Sound Quality Amoung Four Domestic-Sized Rooms. AES 99th Convention, 1995 New York, NY USA. Audio Engineering Society.

OLIVE, S. & WELTI, T. 2009. Validation of A Binaural Car Scanning System for Subjective Evaluation of Automotive Audio Systems. *AES 36th International Conference, Automotive Audio - Sound in Motion.* Dearborn, MI USA: Audio Engineering Society.

OLIVE, S. & WELTI, T. Factors that Influence Listeners Preferred Bass and Treble Balance in Headphones. AES 139th Convention, October 2015 New York, NY USA. Audio Engineering Society.

OLIVE, S., WELTI, T. & MARTENS, W. Listener Loudspeaker Preference Ratings Obtained In Situ Match Those Obtained via a Binaural Room Scanning Measurement and Playback System. AES 122nd Convention, 2007 Vienna, Austria.

OLIVE, S., WELTI, T. & MCMULLIN, E. Listener Preference For Different Headphone Target Response Curves. AES 134th Convention, 2013a Rome, Italy. Audio Engineering Society.

OLIVE, S., WELTI, T. & MCMULLIN, E. Listener Preferences for In-Room Loudspeaker and Headphone Target Responses. AES 135th Convention, 2013b New York, NY, USA. Audio Engineering Society.

OLIVE, S. E. Some New Evidence that Teenagers and College Students May Prefer Accurate Sound Reproduction. AES 132nd Convention, 2012 Budapest, Hungary. Audio Engineering Society.

OLIVE, S. E., SCHUCK, P. L., RYAN, J. G., SALLY, S. L. & BONNEVILLE, M. E. 1997. The Detection Thresholds of Resonances at Low Frequencies. *Journal of the Audio Engineering Society,* 45**,** 116-128.

OLIVE, S. E. & TOOLE, F. E. 1989. The Detection of Reflections in Typical Rooms. *Journal of the Audio Engineering Society,* 37**,** 539-553.

OPPENHEIM, A. V. & SCHAFER, R. W. 1975. *Digital Signal Processing,* London, UK, Printice-Hall.

OSHA 1983. Occupational Health and Environmental Control. *2-2.35 A-29 CFR 1910.95 (b)(1) Guidelines for Noise Enforcement: Appendix A, US Department of Labor.* Washington, DC USA: Occupational Safety and Health Administration.

PELLEGRINI, R. 2001. *A Virtual Reference Listening Room as an Application of Auditory Virtual Environments.* PhD Thesis, Ruhr-Universitat Bochum.

PELLEGRINI, R., KUHN, C. & GEBHARDY, M. Heaphone Technology for Surround Sound Monitoring - A Virtual 5.1 Listening Room. AES 122nd Convention, 2007 Vienna, Austria. Audio Engineering Society.

POSTEL, F., HEGARTY, P. & BECH, S. A Listening Test System for Automotive Audio - Part 5: The Influence of Listening Environment on the Realism of Binaural Reproduction. AES 130th Convention, 2011 London, UK. Audio Engineering Society.

POTWIN, C. & SCHLANGER, B. 1939. Coördinating Acoustics and Architecture in the Design of the Motion Picture Theater. *Journal of the Society of Motion Picture Engineers,* 32**,** 156-168.

POULARIKAS, A. D. 1998. Windows. *The Handbook of Formulas and Tables for Signal Processing.* Boca Raton, FL USA: CRC Press.

PULKKI, V. & KARJALAINEN, M. 2015. *Communication Acoustics: An Introduction to Speech, Audio and Psychoacoustics,* West Sussex, UK, John Wiley & Sons.

QUEEN, D. 1971. Discete Field Equalization for Sound Reinforcement Systems. *Journal of the Audio Engineering Society,* 19**,** 32-35.

RASMUSSEN, E. 1969. A Report on Listening Characteristics in 25 Danish Cinemas. *SMPTE Motion Imaging Journal,* 78**,** 1054-1057.

RASMUSSEN, E. 1971. Scandinavian Efforts to Standardize Acoustic Response in Theaters and Dubbing Rooms. *SMPTE Motion Imaging Journal,* 80**,** 896-899.

RASMUSSEN, E. 1976. Acoustic Response Measurements and Standards for Motion-Picture Theaters. *SMPTE Motion Imaging Journal,* 85**,** 164-169.

RIEDERER, K. Repeatability Analysis of Head-Related Transfer Function Measurements. AES 105th Convention, 1998 San Francisco, CA USA. Audio Engineering Society.

RIEDERER, K. 2005. *HRTF Analysis: Objective and Subjective Evaluation of Measured Head-Related Transfer Functions.* PhD Thesis, Helsinki University of Technology.

ROMIGH, G. D., BRUNGART, D. S. & SIMPSON, B. D. 2015. Free-Field Localization Performance With a Head-Tracked Virtual Auditory Display. *IEEE Journal of Selected Topics in Signal Processing,* 9**,** 943-954.

ROUANET, H. & LEPINE, D. 1970. Comparison Between Treatments in a Repeated-Measurement Design: ANOVA and Multivariate Methods. *British Journal of Mathematical and Statistical Psychology,* 23**,** 147-163.

RUDMOSE, W. 1958. Equalization of Sound Systems. *Noise Control,* 4**,** 24.

RUMSEY, F. Spatial Audio and Sensory Evaluation Techniques-Context, History and Aims. 1st International Workshop on Spatial Audio & Sensory Evaluation Techniques, 2006 Guildford, England.

SABINE, P. E. 1928. The Acoustics of Sound Recording Rooms. *Transactions of the Society of Motion Picture Engineers,* 12**,** 809-822.

SANDVAD, J. Dynamic Aspects of Auditory Virtual Environments. AES 100th Convention, May 1996 Copenhagen, Denmark. Audio Engineering Society.

SCHÄRER, Z. & LINDAU, A. Evaluation of Equalization Methods for Binaural Signals. AES 126th Convention, 2009 Munich, Germany. Audio Engineering Society.

SCHEFFE, H. 1999. *The Analysis of Variance,* Hoboken, NJ USA, John Wiley & Sons.

SCHROEDER, M. R., GOTTLOB, D. & SIEBRASSE, K. F. 1974. Comparative Study of European Concert Halls: Correlation of Subjective Preference with Geometric and Acoustic Parameters. *The Journal of the Acoustical Society of America,* 56**,** 1195-1201.

SCHULEIN, R. B. 1975. In Situ Measurement and Equalization of Sound Reproduction Systems. *Journal of the Audio Engineering Society,* 23**,** 178-186.

SCHULTZ, T. J. & WATTERS, B. G. 1964. Propagation of Sound across Audience Seating. *The Journal of the Acoustical Society of America,* 36**,** 885-896.

SESSLER, G. M. & WEST, J. E. 1964. Sound Transmission over Theatre Seats. *The Journal of the Acoustical Society of America,* 36**,** 1725-1732.

SHORTT, E. J., FRIEDL, G. & BARNETT, H. 1940. A New High-Quality Sound System. *SMPTE Motion Imaging Journal,* 34**,** 212-219.

SKINNER, M. W. 1980. Speech Intelligibility in Noise-Induced Hearing Loss: Effects of High-Frequency Compensation. *The Journal of the Acoustical Society of America,* 67**,** 306-317.

SMOORENBURG, G. F. 1992. Speech Reception In Quiet and In noisy Conditions by Individuals With Noise-Induced Hearing Loss In Relation to Their Tone Audiogram. *The journal of the acoustical society of America,* 91**,** 421-437.

SMPTE 1969. Draft Proposal for Standardizing Acoustic Response of a Monitoring Chain in Motion-Picture Control Rooms. *SMPTE Motion Imaging Journal.* SMPTE.

SMPTE 2010. SMPTE Standard 202-2010 For Motion-Pictures -Dubbing Stages (Mixing Rooms), Screening Rooms and Indoor Theaters - B-Chain Electroacoustic Response. White Plains, NY USA: Society of Motion Picture and Television Engineers.

SMPTE 2014. Frequency and Temporal Response Analysis of Theatres and Dubbing Stages. *25CSS-B Chain.* 2.31 ed. White Plains, NY USA: Society of Motion Picture and Television Engineers.

SMYTH, S., SMYTH, M., CHEUNG, S. & KRAMER, L. 2010. A Virtual Acoustic Film Dubbing Stage. *AES 40th International Conference Spatial Audio - Sense the Space of Sound.* Tokyo, Japan: Audio Engineering Society.

SNAIDERO, T., JACOBSEN, F. & BUCHHOLZ, J. 2011. Measuring HRTFs of Brüel & Kjær Type 4128-C, GRAS KEMAR Type 45BM, and Head Acoustics HMS II. 3 Head and Torso Simulators. Lyngby, Denmark: Technical University of Denmark, Department of Electrical Engineering.

SNOW, W. B. 1961. Loudspeaker Testing in Rooms. *Journal of the Audio Engineering Society,* 9**,** 54-60.

SOGOSURVEY. 2013. *So Go Survey* [Online]. Available: www.sogosurvey.com [Accessed].

STITT, P., HENDRICKX, E., MESSONNIER, J.-C. & KATZ, B. The Influence of Head Tracking Latency on Binaural Rendering in Simple and Complex Sound Scenes. Audio Engineering Society Convention 140, 2016. Audio Engineering Society.

SWAN, C. M. 1929. Acoustics of Picture Theaters. *Architectural Forum,* 51**,** 545-548.

TASA 2013. The TASA Standard. Burbank, California USA: Trailer Audio Standards Association.

THOMPSON, D. 2016. *Hollywood Has a Huge Millennial Problem* [Online]. The Atlantic. Available: http://www.theatlantic.com/business/archive/2016/06/hollywood-has-a-huge-millennial-problem/486209/?single_page=true [Accessed July 27th 2016].

THOMPSON, E. A. 2004. *The Soundscape of Modernity: Architectural Acoustics and the Culture of Listening in America, 1900-1933*, Mit Press.

THURLOW, W., MANGELS, J. & RUNGE, P. 1967. Head Movements During Sound Localization. *The Journal of the Acoustical Society of America,* 42**,** 489-493.

THURLOW, W. & RUNGE, P. 1967. Effect of Induced Head Movements on Localization of Sounds. *The Journal of the Acoustical Society of America,* 42**,** 480-488.

TOOLE, F. 1982. Listening Tests - Turning Opinion into Fact. *Journal of the Audio Engineering Society,* 30**,** 431-445.

TOOLE, F. Subjective Measurements of Loudspeakers - A Comparison of Stereo and Mono Listening. AES 74th Convention, 1983 New York, NY. Audio Engineering Society.

TOOLE, F. 1986a. Loudspeaker Measurements and Their Relationship to Listener Preferences - Part 1. *Journal of the Audio Engineering Society,* 34**,** 227-235.

TOOLE, F. 1986b. Loudspeaker Measurements and Their Relationship to Listener Preferences - Part 2. *Journal of the Audio Engineering Society,* 34**,** 323-348.

TOOLE, F. 2006. Loudspeakers and Rooms for Sound Reproduction - A Scientific Review. *Journal of the Audio Engineering Society,* 54**,** 451-476.

TOOLE, F. 2008. *Sound Reproduction: The Acoustics and Psychoacoustics of Loudspeakers and Rooms,* New York, NY USA, Focal Press.

TOOLE, F. Home Theaters to Cinema, Sound Reproduction in Small and Large Rooms - Lecture. AES 133rd Convention, 2012 San Francisco, CA, USA. Audio Engineering Society.

TOOLE, F. 2015. The Measurement and Calibration of Sound Reproducing Systems. *Journal of the Audio Engineering Society,* 63**,** 512-541.

TOOLE, F. 2016. Sound Reproduction: The Acoustics and Psychoachoustics of Loudspeakers and Rooms. 2nd Edition. *Unpublished Manuscript.* Focal Press.

TOOLE, F. & OLIVE, S. 1988. The Modification of Timbre by Resonances: Perception and Measurement. *Journal of the Audio Engineering Society,* 36**,** 122-142.

TOOLE, F. E. & OLIVE, S. E. The Perception of Sound Coloration Due to Resonances in Loudspeakers and Other Audio Components. Audio Engineering Society Convention 81, 1986. Audio Engineering Society.

TRUJILLO-ORTIZ, A., HERNANDEZ-WALLS, R. & BARBA-ROJO, K. 2008. *Mauspehercnst: Mauchly's Sphericity Test for Orthogonal Contrasts. A MATLAB File.* [Online]. Available: http://www.mathworks.com/matlabcentral/fileexchange/loadFile.do?objectId=19648 [Accessed September 1st, 2015].

UNIATEC. Actes-Résumés du 8éme Congrès UNIATEC 8th Congres del' Union Internationale des Associations Techniques Cinematographiques 1968 Bruxelles, Belguim.

Commission Superieure Technique du Cinéma, International Audio-Visual Technical Centre.

VERMIGLIO, A. J., SOLI, S. D., FREED, D. J. & FISHER, L. M. 2012. The Relationship Between High-Frequency Pure-Tone Hearing Loss, Hearing in Noise Test (HINT) Thresholds, and the Articulation Index. *Journal of the American Academy of Audiology,* 23**,** 779-788.

VLAHOS, P. 1969. An Acoustic Response Standard. *SMPTE Motion Imaging Journal,* 78**,** 1043-1044.

VLAHOS, P. 1973. *RE: Comment on DIS 2969—Electro-Acoustic Response of Motion Picture Control Rooms.* Type to TC36, I.

VLAHOS, P. 1975. An Optical/Acoustic Response Standard for Theaters and Its Correlation to Dubbing Rooms. *SMPTE Motion Imaging Journal,* 84**,** 3-7.

WALLACH, H. 1940. The Role of Head Movements and Vestibular and Visual Cues in Sound Localization. *Experimental Psychology,* 27**,** 339-368.

WATKINS, A. J. 2005. Perceptual Compensation for Effects of Echo and of Reverberation on Speech Identification. *Acta Acustica united with Acustica,* 91**,** 892-901.

WATSON, F. R. 1927. Acoustics of Motion Picture Theaters. *Transactions of the Society of Motion Picture Engineers,* 11**,** 641-650.

WAVES 2005. Waves M360° Surround Manager Software Audio Processors User's Guide Tel-Aviv, Israel: Waves Audio Ltd.

WELCH, P. D. 1967. The Use of Fast Fourier Transform for the Estimation of Power Spectra: A Method Based on Time Averaging Over Short, Modified Periodograms. *IEEE Transactions on Audio and Electroacoustics,* 15**,** 70-73.

WELTI, T. & ZHANG, X. Angular Resolution Requirements for Binaural Room Scanning. AES 129th Convention, 2010 San Francisco, CA, USA. Audio Engineering Society.

WENTE, E. C. 1935. The Characteristics of Sound Transmission in Rooms. *The Journal of the Acoustical Society of America,* 7**,** 76-76.

WENZEL, E. M. 1999. Effect of Increasing System Latency on Localization of Virtual Sounds. *AES 16th International Conference - Spatial Sound Reproduction.* Rovaniemi, Finland: Audio Engineering Society.

WOLF, S. K. 1929a. Reproduction in the Theatre. *Academy Technical Digest.* Hollywood, CA USA: Academy of Motion Picture Arts and Sciences.

WOLF, S. K. 1929b. Theatre Acoustics for Reproduced Sound *Academy Technical Digest.* Hollywood, CA USA: Academy of Motion Picture Arts and Sciences.

YAIRI, S., IWAYA, Y. & SUZUKI, Y. 2008. Influence of Large System Latency of Virtual Auditory Display on Behaviour of Head Movement in Sound Localization Task. *Acta Acustica united with Acustica,* 94**,** 1016-1023.

ZACHAROV, N. & KOIVUNIEMI, K. Unravelling the Perception of Spatial Sound Reproduction: Language Development, Verbal Protocol Analysis and Listener Training. Audio Engineering Society Convention 111, 2001. Audio Engineering Society.

ZAHORIK, P. 2002. Assessing Auditory Distance Perception Using Virtual Acoustics. *The Journal of the Acoustical Society of America,* 111**,** 1832-1846.

ZAJONC, R. B. 2001. Mere Exposure: A Gateway to the Subliminal. *Current Directions in Psychological Science,* 10**,** 224-228.

ZIELINSKI, S. 2016. On Some Biases Encountered in Modern Audio Quality Listening Tests (Part 2): Selected Graphical Examples and Discussion. *Journal of the Audio Engineering Society,* 64**,** 55-74.

ZIELINSKI, S., RUMSEY, F. & BECH, S. 2008. On Some Biases Encountered in Modern Audio Quality Listening Tests—A Review. *Journal of the Audio Engineering Society,* 56**,** 427-451.

Appendix A - Survey Invitation

Sender Name: Linda Gedemer
Sender Email Address: l.a.gedemer@edu.salford.ac.uk
Bcc Email Address: None
Subject: PhD Survey on Cinema Audio Calibration

University of Salford MANCHESTER

Hello {{Name}},

Thank you in advance for helping me with my PhD research survey on cinema audio calibration.

The survey should take about 15 minutes, you will be anonymous on the survey and your individual survey results will be kept confidential. As a participant in the survey, you will be entered into a drawing to win one of three sets of AKG K242HD or K272HD headphones, your choice.

Your answers and input on this survey are extremely valuable and are part of a larger research effort towards the future of cinema calibration.

In order to participate, you may either:

1. Click on this link

or

2. Copy-paste the entire following link between quote marks (NOT including the quote marks) in a web browser
" http://www.sogosurvey.com/k/QsVRPTsSsQRSTUVsP "

or

3. Click on the following URL and enter the login information provided below:
http://www.sogosurvey.com/static/surveykey.aspx
Key: QsVRPTsSsQRSTUVsP

I appreciate your time and thank you for your participation.

Best Regards,

Linda A. Gedemer

Post Graduate Researcher
Acoustics & Audio Engineering
School of Computing, Science & Engineering
University of Salford
l.a.gedemer@edu.salford.ac.uk

This email is sent on behalf of the person/organization whose name appears in the FROM field by SoGoSurvey . If you have any questions about the email, please contact the sender by replying to this email.

If you prefer not to receive future reminders about this survey, please click here.

If you prefer not to receive future surveys from the organization behind this survey, please click here.

POWERED BY SoGoSurvey

Appendix B - Survey Questions

* **Required Information**

page 1

*** 1. What is your age?** (Select one option)

- ○ 20-25
- ○ 26-32
- ○ 33-40
- ○ 41-49
- ○ 50-59
- ○ 60-70
- ○ 70+

*** 2. How many years have you worked in post production audio?** (Select one option)

- ○ 0-5
- ○ 6-10
- ○ 11-15
- ○ 16-20
- ○ 21-25
- ○ 26-30
- ○ 31-40
- ○ 40+

*** 3. How many years have you worked as a re-recording mixer?** (Select one option)

- ○ 0
- ○ 1-5
- ○ 6-10
- ○ 11-15
- ○ 16-20
- ○ 21-25
- ○ 26-30
- ○ 31-40
- ○ 40+

*** 4. How many years have you worked as a studio engineer (i.e., staff technical engineer).** (Select one option)

- ◯ 0
- ◯ 1-5
- ◯ 6-10
- ◯ 11-15
- ◯ 16-20
- ◯ 21-25
- ◯ 26-30
- ◯ 31-40
- ◯ 40+

*** 5. When was your last audiometric (hearing) test?** (Select one option)

- ◯ Haven't had one.
- ◯ Within the last year
- ◯ 1-2 years ago
- ◯ 3-5 years ago
- ◯ 6-10 years ago
- ◯ More than 10 years ago.

*** 6. Were the results of your hearing test within normal range?** (Select one option)

- ◯ Yes
- ◯ No
- ◯ Not sure
- ◯ Did not have hearing tested

*** 7. What country / geographicregion do you work in most often?**
(Select one option)

- ◯ USA
- ◯ Canada
- ◯ Mexico
- ◯ South America
- ◯ UK
- ◯ Northern Europe
- ◯ Southern Europe
- ◯ Russia
- ◯ Africa
- ◯ Asia
- ◯ Australia / New Zealand
- ◯ Middle East

*** 8. Currently, what is the approximate percentage of work you do in each category.**

Film	___
Marketing / Trailers	___
Television	___
Games / Interactive Media	___
Other	___
Total	100

*** 9. Whatis the approximate size of thedub stagethat youmost oftenwork in? (Please select only one)**

- ☐ Small (40' x 30' x 15'), (12m x 9m x 4.5m)
- ☐ Medium (50' x 40' x 20'), (15m x 12m x 6m)
- ☐ Larger than those stated
- ☐ Smaller than those stated
- ☐ Varies

page 2

*** 10. Do you typically work at a dub stage thatutilizes a calibration or roomequalization curve?** (Select one option)

◯ Yes	Go to Page No. 3
◯ No	Go to Page No. 6
◯ Not Sure	Stop, you have finished the survey

page 3

*** 11. How often do you work at a dub stage that uses an un-altered SMPTE ST202-2010 (X-Curve)?** (Select one option)

◯ Always ◯ Sometimes ◯ Never ◯ Not Sure

Always:Go to Page No. 5
Sometimes:Go to Page No. 4
Never:Go to Page No. 7
Not Sure:Continue to next question

page 4

*** 12. If you answered theprevious question "sometimes" (or "not sure"), what is the approximatepercentage (or best estimate)of time that you workat a dub-stage that uses an un-altered SMPTE X-curve?** (Select one option)

◯ 0-20%
◯ 21-40%
◯ 41-60%
◯ 61-80%
◯ 81-100%

page 5

When you have worked on a dub stage that uses the SMPTE X-Curve, howwould you rate the followingsonic qualities?

13. Sonic Quality

	Very Dull 0	1	2	3	4	Average 5	6	7	8	9	Very Bright 10
*(a) High frequency response (Select one option)	○	○	○	○	○	○	○	○	○	○	○

14. Sonic Quality

	Very Thin 0	1	2	3	4	Average 5	6	7	8	9	Very Full 10
*(a) Low frequency response (Select one option)	○	○	○	○	○	○	○	○	○	○	○

15. Sonic Quality

	Very Unclear 0	1	2	3	4	Average 5	6	7	8	9	Very Clear 10
*(a) Clarity of dialog (Select one option)	○	○	○	○	○	○	○	○	○	○	○

16. Sonic Quality

	Very Closed or Flat 0	1	2	3	4	Average 5	6	7	8	9	Very Spacious 10	No effect N/A
*(a) Spatial quality (Select one option)	○	○	○	○	○	○	○	○	○	○	○	○

17. Sonic Quality

	Poor 0	1	2	3	4	Average 5	6	7	8	9	Excellent 10
*(a) Translation of dub stage mix to commercial cinema (Select one option)	○	○	○	○	○	○	○	○	○	○	○

18. Sonic Quality

	Poor 0	1	2	3	4	Average 5	6	7	8	9	Excellent 10
*(a) Overall sonic quality of dub stage system (Select one option)	○	○	○	○	○	○	○	○	○	○	○

*** 19. Do you compensate for the SMPTE X-curve in your mix and if so, how?**

___

___

Go to Page No. 7

page 6

*** 20. Please explain why the dub stage that you typically work at does not use a calibration or equalization curve. If you are unsure of this information, please notate "not sure".**

___

___

Stop, you have finished the survey

page 7

*** 21. How often do you work at a dub stage thatuses a modified version of the SMPTE X-Curve?** (Select one option)

○ Always ○ Sometimes ○ Never ○ Not Sure

Always:Go to Page No. 9
Sometimes:Go to Page No. 8
Never:Go to Page No. 10
Not Sure:Continue to next question

page 8

*** 22. If you answered theprevious question "sometimes" (or "not sure"), what is the approximatepercentage (or best estimate)of time that you workat a dub-stage that uses a modified version of theSMPTE X-curve?** (Select one option)

○ 0-20%

○ 21-40%

○ 41-60%

○ 61-80%

○ 81-100%

page 9

*** 23. Ifthe dub stagewhere you work usesa modified SMPTE X-Curve, please explain how and why itwas modified. If you are unsure of this information, please notate "not sure".**

page 10

*** 24. How often do you work at a dub stage thatuses a custom or "house" curve? (A curve that is customized beyond a simple modification of the X-Curve)** (Select one option)

◯ Always ◯ Sometimes ◯ Never ◯ Not Sure

Always:Go to Page No. 12
Sometimes:Go to Page No. 11
Never: Stop, you have finished the survey
Not Sure:Continue to next question

page 11

*** 25. If you answered theprevious question "sometimes" (or "not sure"), what is the approximatepercentage (or best estimate)of time that you workat a dub-stage that uses a custom or "house" curve?** (Select one option)

◯ 0-20%

◯ 21-40%

◯ 41-60%

◯ 61-80%

◯ 81-100%

page 12

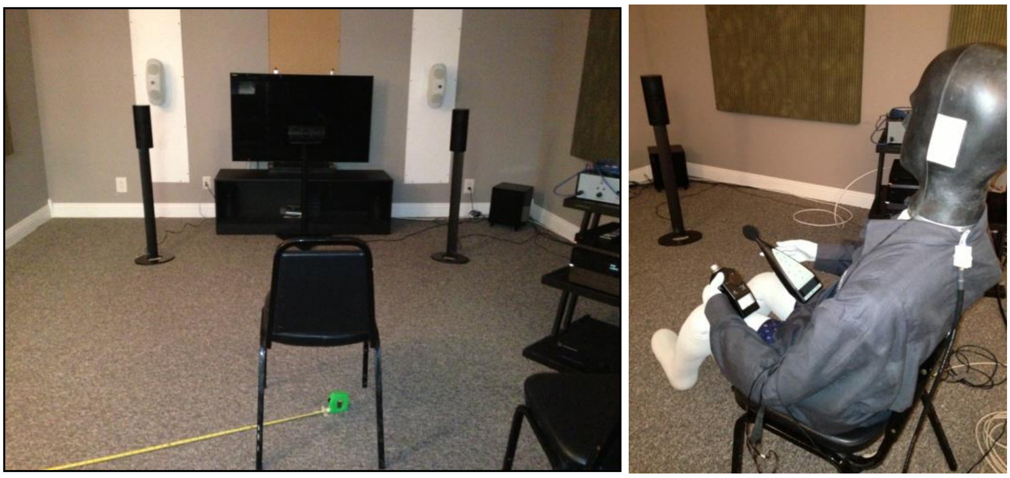
*** 26. Please describe the custom or "house" curve used and why it is used instead of a SMPTE or modified SMPTE curve. If you are unsure of this information, please notate "not sure".**

Appendix C - Example of Survey Answers

Response No : 1
1. What is your age?
33-40
2. How many years have you worked in post production audio?
16-20
3. How many years have you worked as a re-recording mixer?
11-15
4. How many years have you worked as a studio engineer (i.e., staff technical engineer).
0
5. When was your last audiometric (hearing) test?
6-10 years ago
6. Were the results of your hearing test within normal range?
Yes
7. What country / geographic region do you work in most often?
USA
8. Currently, what is the approximate percentage of work you do in each category.
Film:90
Marketing / Trailers:2
Television:1
Games / Interactive Media:5
Other:2
Total:100
9. What is the approximate size of the dub-stage that you most often work in?
Large (70' x 50' x 30')
Varies
10. Do you typically work at a dub-stage that utilizes a calibration or room equalization curve?
Yes
11. How often do you work at a dub-stage that uses an un-altered SMPTE ST202-2010 (X-Curve)?
Sometimes
12. If you answered the previous question "sometimes" (or "not sure"), what is the approximate percentage (or best estimate) of time that you work at a dub-stage that uses an un-altered SMPTE X-Curve? (Select one option)

81-100%
When you have worked on a dub-stage that uses the SMPTE X-Curve, how would you rate the following sonic qualities?
13. Sonic Quality: High frequency response
2
14. Sonic Quality: Low frequency response
4
15. Sonic Quality: Clarity of dialog
5
16. Sonic Quality: Spatial quality
2
17. Sonic Quality: Translation of dub-stage mix to commercial cinema
5
18. Sonic Quality: Overall sonic quality of dub-stage system
5
19. Do you compensate for the SMPTE X-curve in your mix and if so, how?
Yes...tend to push high and low end.
21. How often do you work at a dub-stage that uses a modified version of the SMPTE X-Curve?
Sometimes
22. If you answered the previous question "sometimes" (or "not sure"), what is the approximate percentage (or best estimate) of the time that you work at a dub-stage that uses a modified version of the SMPTE X-Curve (Select one option)
0-20%
23. If the dub-stage where you work uses a modified SMPTE X-Curve, please explain how and why it was modified. If you are unsure of this information, please notate "not sure".
Lower frequency range subwoofers and slightly extended high end. (Most of the systems are 3-way.)
24. How often do you work at a dub-stage that uses a custom or "house" curve? (A curve that is customized beyond a simple modification of the X-Curve)
Never

Appendix D - Photos of Listening Rooms at Harman

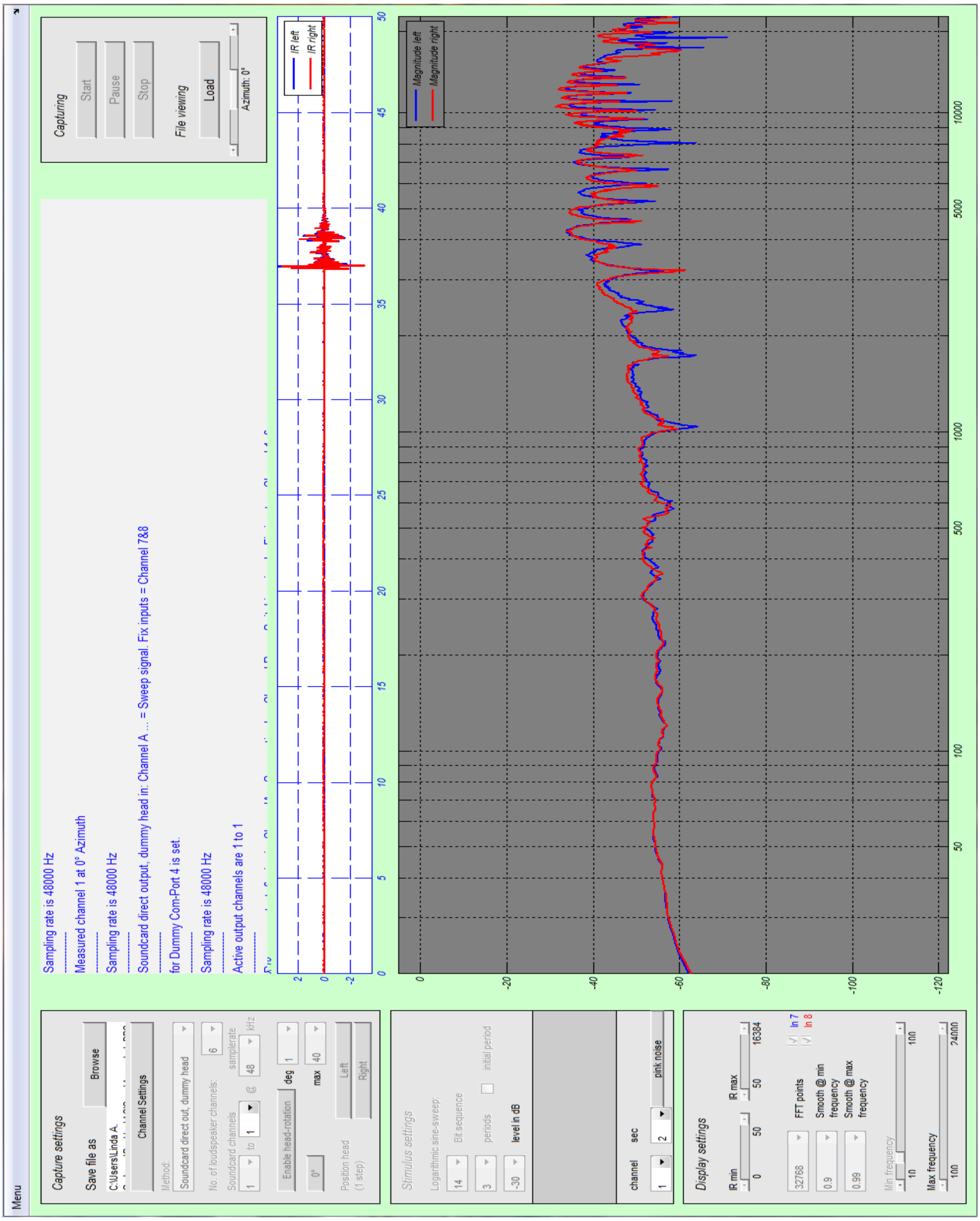


Living Room

Eargle Theatre

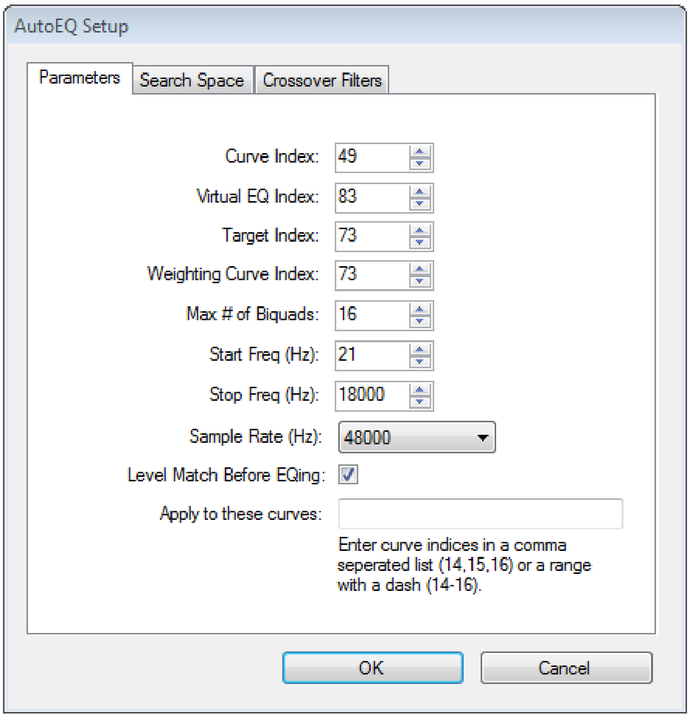
Arena Lab

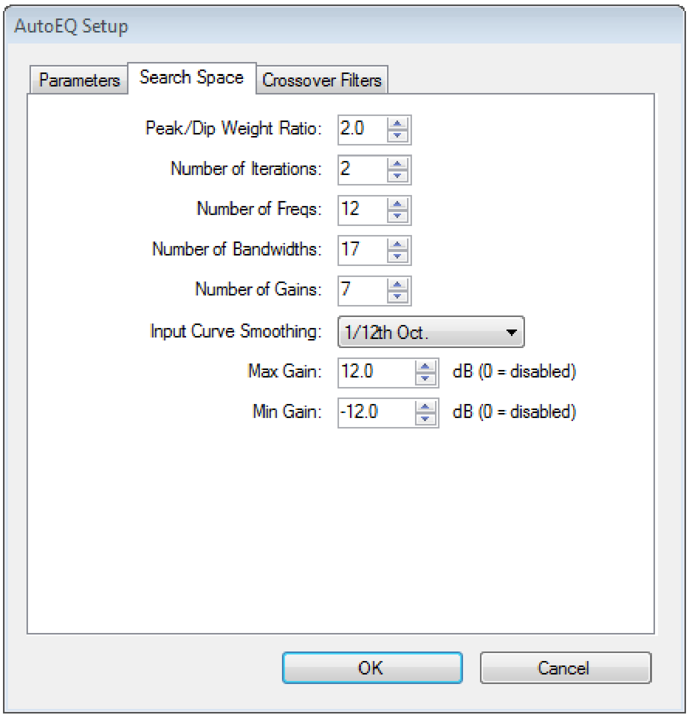
Appendix E - Matlab Binaural Room Scanning Interface



Appendix F - Harman Audio Test Software (HATS)

Auto EQ Setup Screens





Appendix G - Matlab Code for Mauchly's Test of Sphericity

```
function [Mausphercnst] = Mausphercnst(X,alpha)
%MAUSPHERCNST Mauchly's sphericity test for orthogonal contrasts.
%
%     Syntax: function [Mausphercnst] = Mausphercnst(X,alpha)
%
%     Inputs:
%          X - multivariate data matrix.
%      alpha - significance level (default = 0.05).
%     Output:
%          n - sample-size
%          p - variables
%          L - Mauchly's statistic used to test any deviation from
%              an expected sphericity
%         df - degrees of freedom
%         X2 - Chi-square statistic
%          P - probability that null Ho: is true

%  Created by A. Trujillo-Ortiz, R. Hernandez-Walls and K. Barba-Rojo
%             Facultad de Ciencias Marinas
%             Universidad Autonoma de Baja California
%             Apdo. Postal 453
%             Ensenada, Baja California
%             Mexico.
%             atrujo@uabc.mx
%
%  Copyright. April 18, 2008.
%

if nargin < 2,
   alpha = 0.05;  %(default)
end;

if nargin < 1,
   error('Requires at least one input arguments.');
end;

[n,p] = size(X);
S = cov(X);  %variance-covariance matrix
R = diag(ones(p,1))-diag(ones(p-1,1),1);
q = p-1;
A = R(1:q,:);  %nonorthonormalized contrasts matrix
B = A';
C = cgrscho(B);  %m×n column-orthonormal output matrix
L = q^q*det(C'*S*C)/(trace(C'*S*C)^q);  %Mauchly's statistic
LL = -(n-1-((2*q^2+q+2)/(6*q)))*log(L);  %approximate to Chi-square
distribution
F = q*(q+1)/2-1;  %degrees of freedom
P = 1-chi2cdf(LL,F);  %Chi-square probability-value

fprintf('----------------------------------------------------------------------------
--------\n');
disp(' Sample-size    Variables        L          df         Chi-square
P')
```

```
fprintf('------------------------------------------------------------------
--------\n');
fprintf('%8.i%13.i%16.4f%8.i%17.4f%12.4f\n',n,p,L,F,LL,P);
fprintf('------------------------------------------------------------------
--------\n');
fprintf('With a given significance level of: %.2f\n', alpha);

if P >= alpha;
   fprintf('Assumption of sphericity is tenable.\n\n');
else
   fprintf('Assumption of sphericity is not tenable.\n\n');
end

return;
```

Appendix H - ANOVA Table for BRS Experiment #1

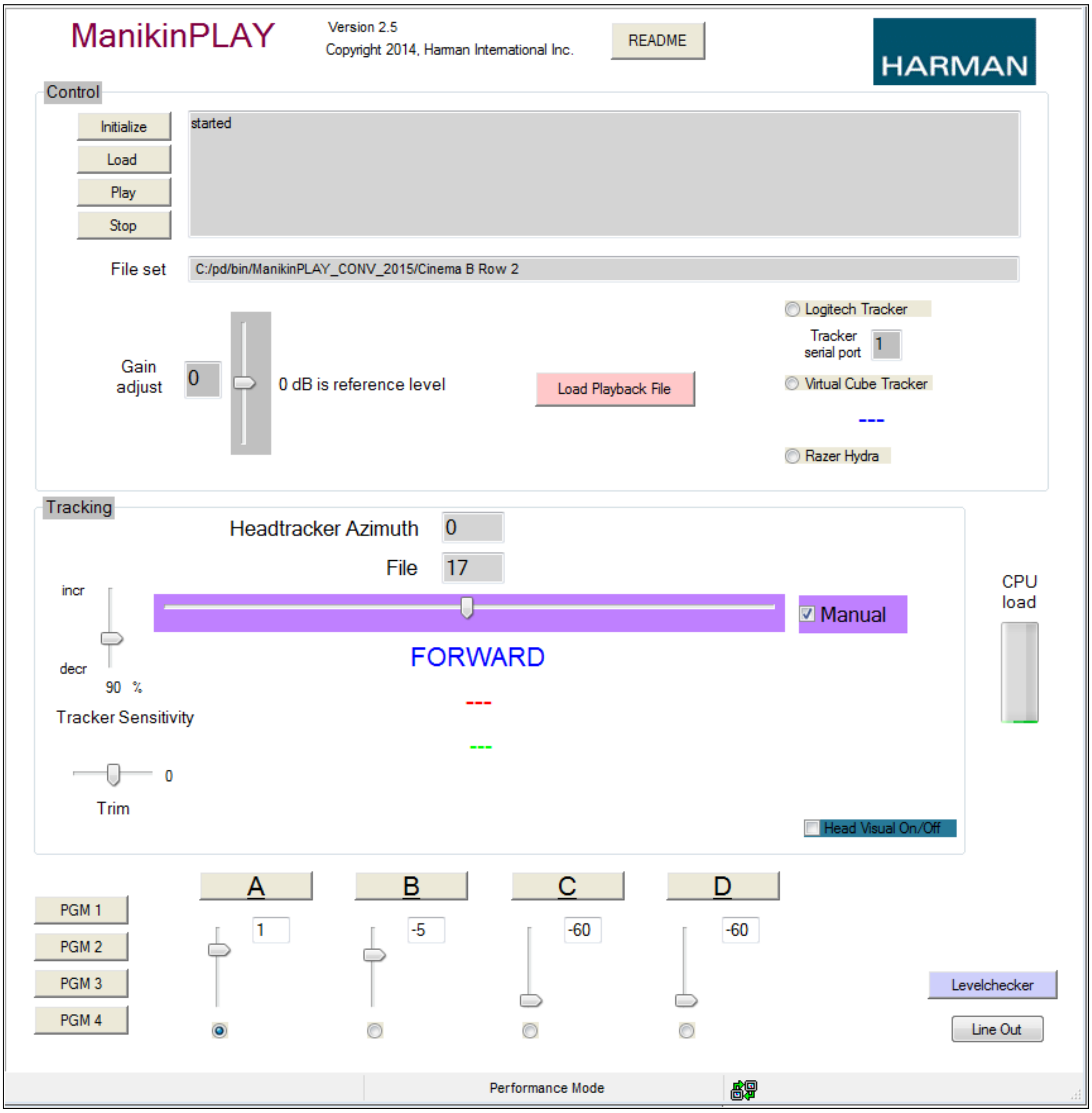
	DF	Sum of Squares	Mean Square	F-Value	P-Value	Lambda	Power
Subject	8	311.153	38.894				
Room	1	25.651	25.651	2.488	0.1534	2.488	0.275
Room * Subject	8	82.484	10.31				
Method	1	3.241	3.241	3.672	0.0916	3.672	0.383
Method * Subject	8	7.059	0.882				
Target Curve	3	3477.193	1159.064	96.781	<.0001	290.343	1
Target Curve * Subject	24	287.428	11.976				
Program	3	2.345	0.782	0.905	0.4532	2.715	0.213
Program * Subject	24	20.73	0.864				
Observation	1	0.003	0.003	0.091	0.7711	0.091	0.058
Observation * Subject	8	0.277	0.035				
Room * Method	1	0.001	0.001	0.003	0.9608	0.003	0.05
Room * Method * Subject	8	2.186	0.273				
Room * Target Curve	3	91.608	30.536	5.682	0.0044	17.045	0.914
Room * Target Curve * Subject	24	128.987	5.374				
Room * Program	3	6.826	2.275	2.791	0.0622	8.372	0.593
Room * Program * Subject	24	19.569	0.815				
Room * Observation	1	0.03	0.03	2.802	0.1327	2.802	0.304
Room * Observation * Subject	8	0.086	0.011				
Method * Target Curve	3	1.489	0.496	0.672	0.5777	2.015	0.167
Method * Target Curve * Subject	24	17.727	0.739				
Method * Program	3	0.437	0.146	0.46	0.7129	1.379	0.127
Method * Program * Subject	24	7.606	0.317				
Method * Observation	1	0.278	0.278	1.856	0.2102	1.856	0.217
Method * Observation * Subject	8	1.199	0.15				
Target Curve * Program	9	13.246	1.472	1.305	0.2496	11.744	0.585
Target Curve * Program * Subject	72	81.208	1.128				
Target Curve * Observation	3	0.477	0.159	1.737	0.1861	5.212	0.387
Target Curve * Observation * Subject	24	2.196	0.091				
Program * Observation	3	0.177	0.059	0.431	0.7326	1.293	0.122
Program * Observation * Subject	24	3.283	0.137				
Room * Method * Target Curve	3	0.284	0.095	0.09	0.9647	0.271	0.064
Room * Method * Target Curve * Subject	24	25.156	1.048				
Room * Method * Program	3	0.148	0.049	0.188	0.9038	0.563	0.08
Room * Method * Program * Subject	24	6.305	0.263				
Room * Method * Observation	1	0.394	0.394	4.752	0.0609	4.752	0.476
Room * Method * Observation * Subject	8	0.663	0.083				
Room * Target Curve * Program	9	8.841	0.982	1.318	0.2431	11.859	0.591
Room * Target Curve * Program * Subject	72	53.677	0.746				
Room * Target Curve * Observation	3	0.424	0.141	2.602	0.0754	7.805	0.559

Room * Target Curve * Observation * Subject	24	1.304	0.054				
Room * Program * Observation	3	0.088	0.029	0.322	0.8094	0.966	0.102
Room * Program * Observation * Subject	24	2.194	0.091				
Method * Target Curve * Program	9	2.534	0.282	1.09	0.3809	9.81	0.493
Method * Target Curve * Program * Subject	72	18.6	0.258				
Method * Target Curve * Observation	3	0.143	0.048	0.397	0.7563	1.191	0.116
Method * Target Curve * Observation * Subject	24	2.876	0.12				
Method * Program * Observation	3	0.145	0.048	0.622	0.6075	1.867	0.157
Method * Program * Observation * Subject	24	1.858	0.077				
Target Curve * Program * Observation	9	0.677	0.075	0.952	0.4864	8.57	0.43
Target Curve * Program * Observation * Subject	72	5.685	0.079				
Room * Method * Target Curve * Program	9	1.615	0.179	0.859	0.565	7.735	0.388
Room * Method * Target Curve * Program * Subject	72	15.035	0.209				
Room * Method * Target Curve * Observation	3	0.219	0.073	1.178	0.339	3.533	0.269
Room * Method * Target Curve * Observation * Subject	24	1.488	0.062				
Room * Method * Program * Observation	3	0.112	0.037	0.694	0.5646	2.083	0.171
Room * Method * Program * Observation * Subject	24	1.294	0.054				
Room * Target Curve * Program * Observation	9	0.405	0.045	0.548	0.8345	4.932	0.246
Room * Target Curve * Program * Observation * Subject	72	5.917	0.082				
Method * Target Curve * Program * Observation	9	0.394	0.044	0.585	0.8055	5.261	0.262
Method * Target Curve * Program * Observation * Subject	72	5.387	0.075				
Room * Method * Target Curve * Program * Observation	9	0.672	0.075	1	0.4479	9.002	0.452
Room * Method * Target Curve * Program * Observation * Subject	72	5.372	0.075				

Appendix I - ANOVA Table for BRS Experiment #2

	DF	Sum of Squares	Mean Square	F-Value	P-Value	Lambda	Power
Subject	8	148.551	18.569				
Method	1	0.019	0.019	0.008	0.9298	0.008	0.051
Method * Subject	8	18.31	2.289				
Target Curve	3	2187.589	729.196	53.954	<.0001	161.862	1
Target Curve * Subject	24	324.364	13.515				
Program	3	1.123	0.374	1.11	0.3646	3.329	0.255
Program * Subject	24	8.096	0.337				
Observation	1	0.001	0.001	0.013	0.9112	0.013	0.051
Observation * Subject	8	0.514	0.064				
Method * Target Curve	3	2.253	0.751	0.725	0.5467	2.176	0.177
Method * Target Curve * Subject	24	24.845	1.035				
Method * Program	3	0.769	0.256	1.158	0.3461	3.475	0.265
Method * Program * Subject	24	5.308	0.221				
Method * Observation	1	0.049	0.049	0.378	0.5557	0.378	0.083
Method * Observation * Subject	8	1.032	0.129				
Target Curve * Program	9	1.878	0.209	0.685	0.7201	6.164	0.307
Target Curve * Program * Subject	72	21.934	0.305				
Target Curve * Observation	3	0.146	0.049	0.374	0.7728	1.121	0.111
Target Curve * Observation * Subject	24	3.136	0.131				
Program * Observation	3	0.515	0.172	1.369	0.2762	4.106	0.309
Program * Observation * Subject	24	3.009	0.125				
Method * Target Curve * Program	9	1.148	0.128	0.436	0.9112	3.923	0.198
Method * Target Curve * Program * Subject	72	21.066	0.293				
Method * Target Curve * Observation	3	0.247	0.082	0.849	0.4808	2.547	0.202
Method * Target Curve * Observation * Subject	24	2.329	0.097				
Method * Program * Observation	3	0.49	0.163	1.389	0.2702	4.168	0.314
Method * Program * Observation * Subject	24	2.824	0.118				
Target Curve * Program * Observation	9	1.639	0.182	1.86	0.072	16.74	0.778
Target Curve * Program * Observation * Subject	72	7.048	0.098				
Method * Target Curve * Program * Observation	9	1.609	0.179	1.502	0.1636	13.522	0.662
Method * Target Curve * Program * Observation * Subject	72	8.567	0.119				

Appendix J - Pure Data Convolution Engine Graphical User Interface (GUI)



Appendix K - ANOVA Table for Cinema Experiment

	DF	Sum of Squares	Mean Square	F-Value	P-Value	Lambda	Power
Subject	13	515.237	39.634				
Cinema	2	47.229	23.615	15.639	<.0001	31.278	0.999
Cinema * Subject	26	39.26	1.51				
Listening Position	2	5.487	2.743	1.04	0.3676	2.081	0.205
Listening Position * Subject	26	68.566	2.637				
Target Curve	4	30556.961	7639.24	547.123	<.0001	2188.491	1
Target Curve * Subject	52	726.054	13.963				
Program	4	83.334	20.834	12.607	<.0001	50.428	1
Program * Subject	52	85.932	1.653				
Observation	1	0.022	0.022	0.186	0.6731	0.186	0.068
Observation * Subject	13	1.543	0.119				
Cinema * Listening Position	4	19.992	4.998	2.744	0.038	10.977	0.717
Cinema * Listening Position * Subject	52	94.704	1.821				
Cinema * Target Curve	8	87.491	10.936	7.873	<.0001	62.983	1
Cinema * Target Curve * Subject	104	144.468	1.389				
Cinema * Program	8	26.868	3.358	5.608	<.0001	44.865	1
Cinema * Program * Subject	104	62.281	0.599				
Cinema * Observation	2	0.785	0.392	2.402	0.1104	4.804	0.43
Cinema * Observation * Subject	26	4.248	0.163				
Listening Position * Target Curve	8	17.375	2.172	1.679	0.1121	13.434	0.704
Listening Position * Target Curve * Subject	104	134.507	1.293				
Listening Position * Program	8	3.742	0.468	0.83	0.5783	6.638	0.363
Listening Position * Program * Subject	104	58.619	0.564				
Listening Position * Observation	2	0.527	0.264	2.038	0.1506	4.076	0.37
Listening Position * Observation * Subject	26	3.363	0.129				
Target Curve * Program	16	514.636	32.165	34.868	<.0001	557.896	1
Target Curve * Program * Subject	208	191.871	0.922				
Target Curve * Observation	4	0.836	0.209	2.326	0.0685	9.303	0.631
Target Curve * Observation * Subject	52	4.674	0.09				
Program * Observation	4	0.055	0.014	0.094	0.984	0.375	0.068
Program * Observation * Subject	52	7.621	0.147				
Cinema * Listening Position * Target Curve	16	77.42	4.839	5.266	<.0001	84.251	1
Cinema * Listening Position * Target Curve * Subject	208	191.136	0.919				
Cinema * Listening Position * Program	16	29.986	1.874	3.954	<.0001	63.257	1
Cinema * Listening Position * Program * Subject	208	98.599	0.474				
Cinema * Listening Position * Observation	4	0.294	0.074	0.394	0.8118	1.577	0.132

Cinema * Listening Position * Observation * Subject	52	9.707	0.187				
Cinema * Target Curve * Program	32	79.078	2.471	4.739	<.0001	151.65	1
Cinema * Target Curve * Program * Subject	416	216.923	0.521				
Cinema * Target Curve * Observation	8	0.436	0.055	0.386	0.9258	3.09	0.174
Cinema * Target Curve * Observation * Subject	104	14.678	0.141				
Cinema * Program * Observation	8	1.086	0.136	1.229	0.2893	9.836	0.537
Cinema * Program * Observation * Subject	104	11.481	0.11				
Listening Position * Target Curve * Program	32	58.08	1.815	3.376	<.0001	108.031	1
Listening Position * Target Curve * Program * Subject	416	223.652	0.538				
Listening Position * Target Curve * Observation	8	1.257	0.157	0.882	0.5343	7.056	0.386
Listening Position * Target Curve * Observation * Subject	104	18.529	0.178				
Listening Position * Program * Observation	8	0.484	0.06	0.434	0.8984	3.469	0.193
Listening Position * Program * Observation * Subject	104	14.503	0.139				
Target Curve * Program * Observation	16	2.086	0.13	0.962	0.4996	15.394	0.64
Target Curve * Program * Observation * Subject	208	28.188	0.136				
Cinema * Listening Position * Target Curve * Program	64	84.744	1.324	2.674	<.0001	171.149	1
Cinema * Listening Position * Target Curve * Program * Subject	832	411.963	0.495				
Cinema * Listening Position * Target Curve * Observation	16	1.712	0.107	0.876	0.5972	14.022	0.587
Cinema * Listening Position * Target Curve * Observation * Subject	208	25.396	0.122				
Cinema * Listening Position * Program * Observation	16	1.685	0.105	0.965	0.4961	15.445	0.642
Cinema * Listening Position * Program * Observation * Subject	208	22.689	0.109				
Cinema * Target Curve * Program * Observation	32	3.869	0.121	0.963	0.5283	30.807	0.874
Cinema * Target Curve * Program * Observation * Subject	416	52.252	0.126				
Listening Position * Target Curve * Program * Observation	32	3.799	0.119	0.992	0.4825	31.748	0.887
Listening Position * Target Curve * Program * Observation * Subject	416	49.784	0.12				
Cinema * Listening Position * Target Curve * Program * Observation	64	7.828	0.122	1.053	0.3695	67.362	0.995
Cinema * Listening Position * Target Curve * Program * Observation * Subject	832	96.683	0.116				

Appendix L - Effect Size, Program wrt Target Curve

	d	*r*	r^2	*CL*
Curve A - Program				
SD/TM	0.02	0.01	0.000	50.47%
SD/Sc	0.32	0.16	0.025	58.93%
SD/Sc-T	0.32	0.16	0.026	59.05%
SD/Sc-F	0.04	0.02	0.000	51.24%
TM/Sc	0.29	0.14	0.020	57.99%
TM/Sc-T	0.29	0.14	0.020	58.10%
TM/Sc-F	0.03	0.00	0.000	50.04%
Sc/Sc-T	0.00	0.00	0.000	50.04%
Sc/Sc-F	0.26	0.13	0.017	57.32%
Sc-T/Sc-F	0.26	0.06	0.004	53.65%
Curve B - Program				
SD/TM	0.13	0.06	0.004	53.59%
SD/Sc	0.05	0.03	0.001	51.45%
SD/Sc-T	0.03	0.01	0.000	50.73%
SD/Sc-F	0.14	0.07	0.005	53.97%
TM/Sc	0.07	0.36	0.126	52.00%
TM/Sc-T	0.10	0.05	0.002	52.77%
TM/Sc-F	0.02	0.01	0.000	50.46%
Sc/Sc-T	0.03	0.01	0.000	50.71%
Sc/Sc-F	0.09	0.04	0.002	52.40%
Sc-T/Sc-F	0.11	0.06	0.003	53.16%
Curve C - Program				
SD/TM	0.09	0.05	0.002	52.65%
SD/Sc	0.08	0.04	0.002	52.38%
SD/Sc-T	0.38	0.19	0.035	60.64%
SD/Sc-F	0.63	0.30	0.091	67.30%
TM/Sc	0.17	0.09	0.007	54.83%
TM/Sc-T	0.29	0.15	0.021	58.24%
TM/Sc-F	0.56	0.27	0.073	65.41%
Sc/Sc-T	0.43	0.21	0.045	62.04%
Sc/Sc-F	0.67	0.32	0.100	68.12%
Sc-T/Sc-F	0.30	0.15	0.022	58.47%
Curve D - Program				
SD/TM	0.43	0.21	0.045	62.02%
SD/Sc	0.28	0.14	0.019	57.90%
SD/Sc-T	0.50	0.24	0.058	63.72%
SD/Sc-F	0.83	0.38	0.147	72.11%
TM/Sc	0.68	0.32	0.103	68.44%
TM/Sc-T	0.04	0.02	0.000	51.23%
TM/Sc-F	0.36	0.18	0.031	60.05%
Sc/Sc-T	0.75	0.35	0.123	70.19%
Sc/Sc-F	1.07	0.47	0.222	77.51%
Sc-T/Sc-F	0.33	0.16	0.026	59.15%
Curve E - Program				
SD/TM	0.14	0.07	0.005	53.92%
SD/Sc	0.35	0.17	0.030	59.83%
SD/Sc-T	1.26	0.53	0.285	81.38%
SD/Sc-F	1.10	0.48	0.233	78.18%
TM/Sc	0.49	0.24	0.057	63.62%
TM/Sc-T	1.40	0.57	0.328	83.82%
TM/Sc-F	1.24	0.53	0.278	81.01%
Sc/Sc-T	0.93	0.42	0.178	74.46%
Sc/Sc-F	0.75	0.35	0.124	70.28%
Sc-T/Sc-F	0.21	0.10	0.011	55.86%

Appendix M - Effect Size, Target Curve wrt Program

	d	*r*	r^2	*CL*
SD - Curve				
A/B	0.50	0.24	0.059	63.80%
A/C	-7.13	0.96	0.927	100.00%
A/D	-6.20	0.95	0.906	100.00%
A/E	-4.10	0.90	0.808	99.81%
B/C	-7.61	0.97	0.935	100.00%
B/D	-6.68	0.96	0.918	100.00%
B/E	-4.54	0.92	0.837	99.93%
C/D	0.85	0.39	0.154	72.71%
C/E	2.42	0.77	0.595	95.66%
D/E	1.61	0.63	0.392	87.20%
TM - Curve				
A/B	0.60	0.29	0.082	66.36%
A/C	-6.78	0.96	0.920	100.00%
A/D	-5.00	0.93	0.862	99.98%
A/E	-3.76	0.88	0.780	99.61%
B/C	-7.84	0.97	0.939	100.00%
B/D	-5.81	0.95	0.894	100.00%
B/E	-4.53	0.91	0.837	99.93%
C/D	1.19	0.51	0.262	80.05%
C/E	2.54	0.79	0.617	96.37%
D/E	1.23	0.52	0.274	80.74%
Sc - Curves				
A/B	0.79	0.37	0.134	71.09%
A/C	-5.76	0.94	0.892	100.00%
A/D	-5.54	0.94	0.885	100.00%
A/E	-3.86	0.89	0.789	99.68%
B/C	-6.78	0.96	0.920	100.00%
B/D	-6.66	0.96	0.917	100.00%
B/E	-4.83	0.92	0.854	99.97%
C/D	0.59	0.28	0.080	66.18%
C/E	1.96	0.70	0.490	91.73%
D/E	1.47	0.59	0.352	85.13%
ScT - Curves				
A/B	0.78	0.36	0.132	70.96%
A/C	-5.97	0.95	0.899	100.00%
A/D	-4.52	0.91	0.836	99.93%
A/E	-4.56	0.92	0.839	99.94%
B/C	-7.18	0.96	0.928	100.00%
B/D	-5.49	0.94	0.883	99.99%
B/E	-5.45	0.94	0.881	99.99%
C/D	0.95	0.43	0.183	74.85%
C/E	0.60	0.29	0.081	66.31%
D/E	-0.28	0.14	0.019	57.89%
ScF - Curves				
A/B	0.63	0.30	0.090	67.17%
A/C	-5.17	0.93	0.870	99.99%
A/D	-4.33	0.91	0.824	99.89%
A/E	-4.92	0.93	0.858	99.98%
B/C	-5.87	0.95	0.896	100.00%
B/D	-5.03	0.93	0.863	99.98%
B/E	-5.67	0.94	0.889	100.00%
C/D	0.86	0.39	0.155	72.75%
C/E	0.48	0.23	0.055	63.37%
D/E	-0.40	0.20	0.039	61.15%

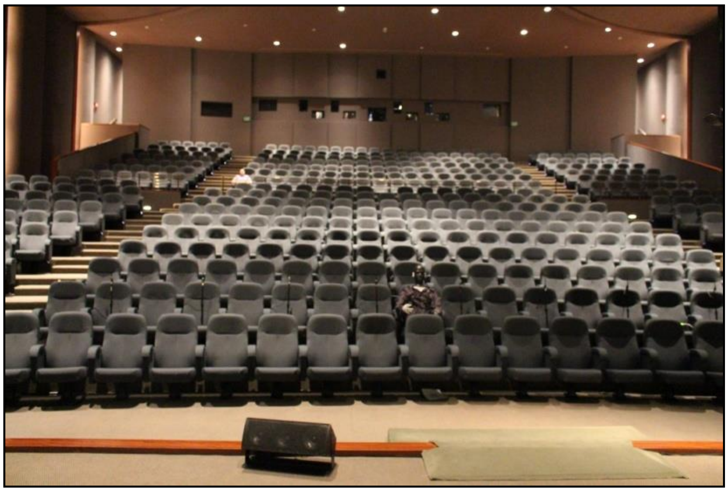
Appendix N - Cinema Photos

Cinema A - Warner Bros. Room 2

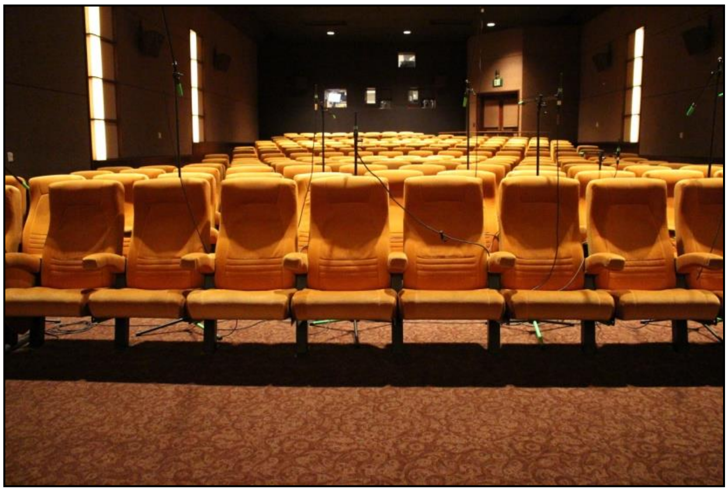
Cinema B - Warner Bros. Room 1

Cinema C - Dolby Burbank

Cinema D - Warner Bros. Room 5



Cinema E - Technicolor



Cinema F - Warner Bros. Room 12

Cinema F - Warner Bros. Steven J. Ross Theatre

Appendix O - Matlab Code for Impulse Response Analysis

```
function [rt,drr,cte,cfs,edt] = irStats(filename,varargin)
%IRSTATS Calculate RT, DRR, Cte, and EDT for impulse response file
%
%   RT = IOSR.ACOUSTICS.IRSTATS(FILENAME) returns the reverberation time
%   (to -60 dB) using a method based on ISO 3382-1:2009. The function uses
%   reverse cumulative trapezoidal integration to estimate the decay curve,
%   and a linear least-square fit to estimate the slope between 0 dB and
%   -60 dB. Estimates are taken in octave bands and the overall figure is
%   an average of the 500 Hz and 1 kHz bands.
%
%   FILENAME should be the full path to an audio file or the name of an
%   audio file on the Matlab search path. The file can be of any format
%   supported by the AUDIOREAD function, and have any number of channels;
%   estimates (and plots) will be returned for each channel.
%
%   The function returns a 1xN vector of RTs, where N is the number of
%   channels in the audio file.
%
%   The function determines the direct sound as the peak of the squared
%   impulse response.
%
%   [RT,DRR] = IOSR.ACOUSTICS.IRSTATS(FILENAME) returns the
%   direct-to-reverberant-ratio DRR for the impulse; DRR is the same size
%   as RT. This is calculated in the following way:
%
%   DRR = 10 * log10( X(T0-C:T0+C)^2 / X(T0+C+1:end)^2 )
%
%   where X is the approximated integral of the impulse, T0 is the time of
%   the direct impulse, and C=2.5ms [1].
%
%   [RT,DRR,CTE] = IOSR.ACOUSTICS.IRSTATS(FILENAME) returns the
%   early-to-late index CTE for the impulse; CTE is the same size as RT.
%   This is calculated in the following way:
%
%   CTE = 10 * log10( X(T0-C:T0+TE)^2 / X(T0+TE+1:end)^2 )
%
%   where TE is 50 ms.
%
%   [RT,DRR,CTE,CFS] = IOSR.ACOUSTICS.IRSTATS(FILENAME) returns the
%   octave-band centre frequencies CFS used in the calculation of RT.
%
%   [RT,DRR,CTE,CFS,EDT] = IOSR.ACOUSTICS.IRSTATS(FILENAME) returns the
%   early decay time EDT, which is the same size as RT. The slope of the
%   decay curve is determined from the fit between 0 and -10 dB. The decay
%   time is calculated from the slope as the time required for a 60 dB
%   decay.
%
%   ... = IOSR.ACOUSTICS.IRSTATS(...,'PARAMETER',VALUE) allows numerous
%   parameters to be specified. These parameters are:
%
%       'graph'      : {false} | true
%           Controls whether decay curves are plotted. Specifically, graphs
%           are plotted of the impulse response, decay curves, and linear
%           least-square fit for each octave band and channel of the audio
```

```
%           file. If the EDT output is specified, the EDT fit will also be
%           plotted.
%       'te'          : {0.05} | scalar
%           Specifies the early time limit (in seconds).
%       'spec'        : {'mean'} | 'full'
%           Determines the nature of RT and EDT outputs. With spec='mean'
%           (default) the reported RT and EDT are the mean of the 500 Hz
%           and 1 kHz bands. With spec='full', the function returns the
%           RT and EDT as calculated for each octave band returned in
%           CFS; RT and EDT have size [M N] where M=length(CFS).
%       'y_fit'       : {[0 60]} | two-element vector
%           Specifies the decibel range over which the decay curve should
%           be evaluated. For example, 'y_fit' may be [-5 -25] or [-5 -35]
%           corresponding to the RT20 and RT30 respectively.
%       'correction' : {0.0025} | scalar
%           Specifies the correction parameter C (in seconds) given above
%           for DRR and CTE calculations. Values of up to 10 ms have been
%           suggested in the literature.
%
%   Octave-band filters are calculated according to ANSI S1.1-1986 and IEC
%   standards. Note that the OCTDSGN function recommends centre frequencies
%   fc in the range fs/200 < fc < fs/5.
%
%   The author would like to thank Feifei Xiong for his input on the
%   correction parameter.
%
%   References
%
%   [1] Zahorik, P., 2002: 'Direct-to-reverberant energy ratio
%       sensitivity', The Journal of the Acoustical Society of America,
%       112, 2110-2117.
%
%   See also AUDIOREAD, OCTDSGN.

%   Copyright 2016 University of Surrey.

    %% validate inputs and set options

    % check dependency
    if exist('octdsgn','file')~=2
%         web('http://uk.mathworks.com/matlabcentral/fileexchange/69-
octave','-new','-browser')
        error('Please download and install the OCTAVE toolbox from:
http://uk.mathworks.com/matlabcentral/fileexchange/69-octave')
    end

    % check file exists
    assert(exist(filename,'file')==2,['iosr.acoustics.irStats: ' filename '
does not exist'])

    % set defaults
    options = struct(...
        'graph',false,...
        'te',0.05,...
        'spec','mean',...
        'y_fit',[0 -60],...
        'correction',0.0025);
```

```
    % read parameter/value inputs
    if nargin>1 % if parameters are specified
        % read the acceptable names
        optionNames = fieldnames(options);
        % count arguments
        nArgs = length(varargin);
        if round(nArgs/2)~=nArgs/2
           error('IRSTATS needs propertyName/propertyValue pairs')
        end
        % overwrite defults
        for pair = reshape(varargin,2,[]) % pair is {propName;propValue}
           IX = strcmpi(pair{1},optionNames); % find match parameter names
           if any(IX)
              % do the overwrite
              options.(optionNames(Fleischmann et al.)) = pair{2};
           else
              error('%s is not a recognized parameter name',pair{1})
           end
        end
    end

    % check options size and type
    assert(isvector(options.y_fit) && numel(options.y_fit)==2,'''y_fit''
must be a two-element vector.')
    assert(isscalar(options.correction),'''correction'' must be a scalar.')
    assert(isscalar(options.te),'''te'' must be a scalar.')
    assert(ischar(options.spec),'''spec'' must be a char array.')
    assert(islogical(options.graph) && numel(options.graph)==1,'''graph''
must be logical.')

    % check for reasonable values
    assert(all(options.y_fit<=0),'''y_fit'' values must be less than or
equal to 0.')
    assert(options.te>=0,'''te'' must be greater than or equal to 0.')
    assert(options.correction>=0,'''correction'' must be greater than or
equal to 0.')

    %% read in audio file

    % read in impulse
    [x,fs] = audioread(filename);
    assert(fs>=5000,'Sampling frequency is too low. FS must be at least
5000 Hz.')

    % set te in samples
    te = round(options.te*fs);

    % Check sanity of te
    assert(te<length(x),'The specified early time limit te is longer than
the duration of the impulse!')

    % get number of channels
    numchans = size(x,2);

    %% set up octave-band filters

    % octave-band center frequencies
    cfs = [31.25 62.5 125 250 500 1000 2000 4000 8000 16000];
```

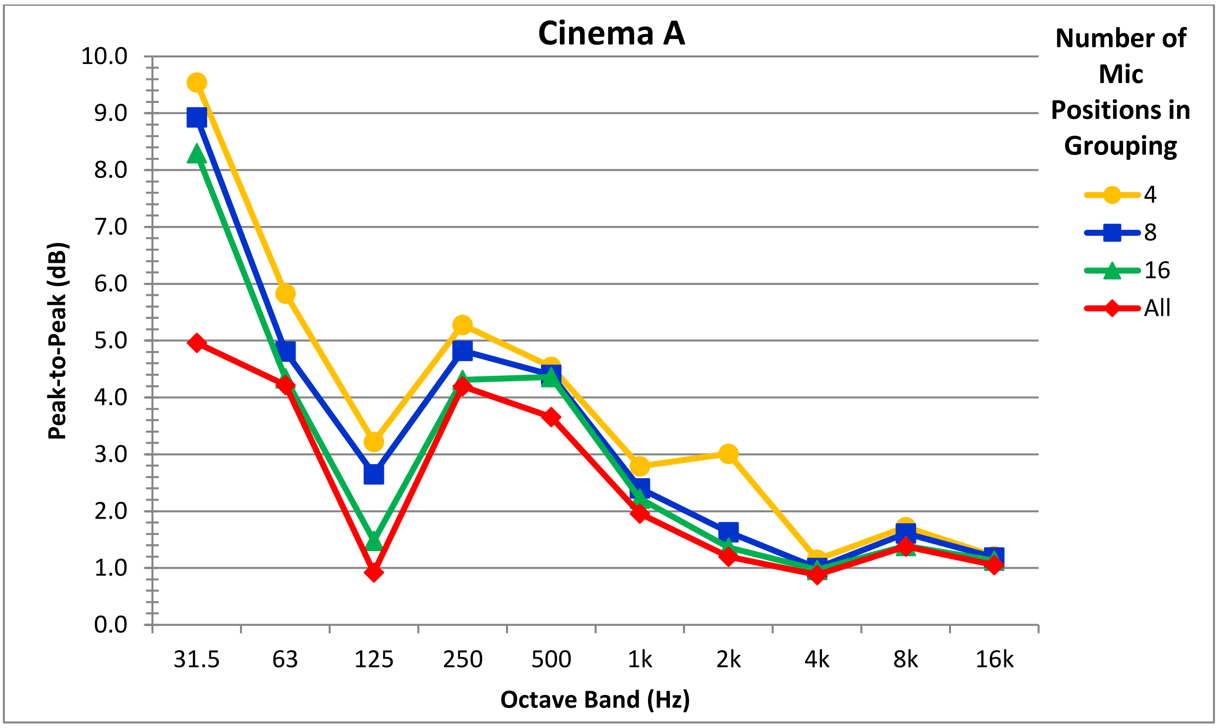
```
    % octave-band filter order
    N = 3;

    % limit centre frequencies so filter coefficients are stable
    %cfs = cfs(cfs>fs/200 & cfs<fs/3);
    %cfs = cfs(:);

    % calculate filter coefficients
    a = zeros(length(cfs),(2*N)+1);
    b = zeros(length(cfs),(2*N)+1);
    for f = 1:length(cfs)
        [b(f,:),a(f,:)] = octdsgn(cfs(f),fs,N);
    end

    %% perform calculations

    % empty matrices to fill
    z = zeros([length(cfs) size(x)]);
    rt_temp = zeros([length(cfs) numchans]);
    edt = zeros([length(cfs) numchans]);
    t0 = zeros(1,numchans);
    drr = zeros(1,numchans);
    cte = zeros(1,numchans);

    correction = round(options.correction*fs);

    % filter and integrate
    for n = 1:numchans
        % find direct impulse
        peak = find(x(:,n).^2==max(x(:,n).^2));
        if numel(peak)>1
            warning('More than one peak found. Choosing first peak. Are you
sure this is an impulse response?')
        end
        t0(n) = peak(1);
        % draw figure
        if options.graph
            scrsz = get(0,'ScreenSize');
            figpos = [((n-1)/numchans)*scrsz(3) scrsz(4) scrsz(3)/2
scrsz(4)];
            figure('Name',['Channel ' num2str(n)],'OuterPosition',figpos);
        end
        % evaluate impulse in each octave band
        for f = 1:length(cfs)
            y = filter(b(f,:),a(f,:),x(:,n)); % octave-band filter
            temp = cumtrapz(y(end:-1:1).^2); % decay curve
            z(f,:,n) = temp(end:-1:1);
            [rt_temp(f,n),E_rt,fit_rt] =
calc_decay(z(f,t0:end,n),options.y_fit,60,fs,cfs(f)); % estimate RT
            [edt(f,n),E_edt,fit_edt] = calc_decay(z(f,t0:end,n),[0,-
10],60,fs,cfs(f)); % estimate EDT
            if options.graph % plot
                % time axes for different vectors
                ty = ((0:length(y)-1)-t0(n))./fs;
                tE_rt = (0:length(E_rt)-1)./fs;
                tE_edt = (0:length(E_edt)-1)./fs;
```

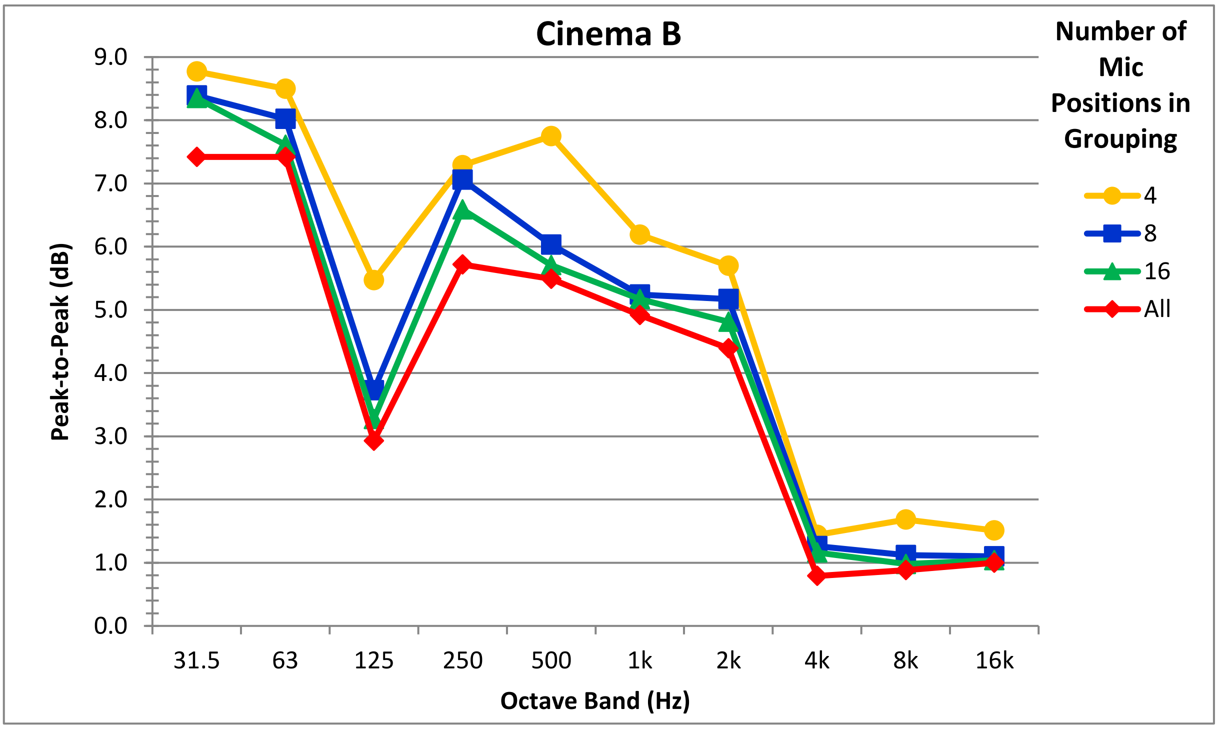
```
                % plot
                subplot(length(cfs),2,(2*f)-1)
                plot(ty,y,'k') % octave-band impulse
                if f==1
                    title({'Impulse response'; ''; [num2str(cfs(f)) ' Hz
octave band']})
                else
                    title([num2str(cfs(f)) ' Hz octave band'])
                end
                if f==length(cfs)
                    xlabel('Time [s]')
                else
                    set(gca,'xticklabel',[]);
                end
                ylabel('Amplitude')
                set(gca,'position',[1 1 1
1.05].*get(gca,'position'),'xlim',[min(ty) max(ty)]);
                subplot(length(cfs),2,2*f)
                % energy decay and linear least-square fit
                if nargout==5
                    % plot EDT fit if EDT wanted
                    plot(tE_rt,E_rt,'-k',tE_rt,fit_rt,'--
r',tE_edt,fit_edt,':b')
                else
                    plot(tE_rt,E_rt,'-k',tE_rt,fit_rt,'--r')
                end
                % title for top row
                if f==1
                    title({'Decay curve'; ''; [num2str(cfs(f)) ' Hz octave
band']})
                else
                    title([num2str(cfs(f)) ' Hz octave band'])
                end
                % x label for bottom row
                if f==length(cfs)
                    xlabel('Time [s]')
                else
                    set(gca,'xticklabel',[]);
                end
                ylabel('Energy [dB]')
                set(gca,'position',[1 1 1
1.05].*get(gca,'position'),'ylim',[-70 0],'xlim',[0 max(tE_rt)]);
                % choose legend according to EDT request
                fitstr = num2str(abs(diff(options.y_fit)));
                if nargout==5
                    legend('Energy decay curve',['Linear fit (RT' fitstr
')'],'Linear fit (EDT)','location','northeast')
                else
                    legend('Energy decay curve',['Linear fit (RT' fitstr
')'],'location','northeast')
                end
            end
        end
        % DRR
        if nargout>=2
            drr(n) = 10.*log(...
                trapz(x(max(1,t0(n)-correction):t0(n)+correction,n).^2)/...
                trapz(x(t0(n)+correction+1:end,n).^2)...
                );
```

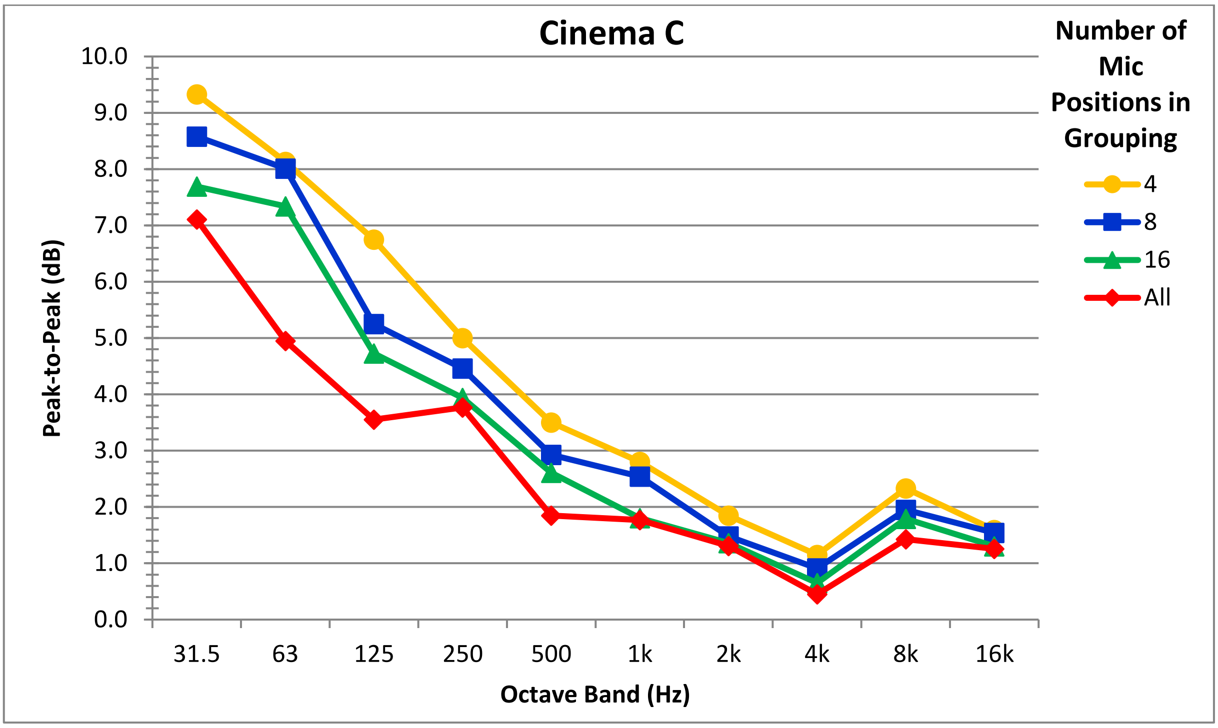
```
        end
        % Cte
        if nargout>=3
            if t0(n)+te+1>size(x,1)
                warning(['Early time limit (te) out of range in channel '
num2str(n) '. Try lowering te.'])
                cte(n) = NaN;
            else
                cte(n) = 10.*log(...
                    trapz(x(max(1,t0(n)-correction):t0(n)+te).^2)/...
                    trapz(x(t0(n)+te+1:end,n).^2)...
                    );
            end
        end
    end

    %% write output

    switch lower(options.spec)
        case 'full'
            rt = rt_temp;
        case 'mean'
            rt = mean(rt_temp(cfs==500 | cfs==1000,:)); % overall RT
            edt = mean(edt(cfs==500 | cfs==1000,:)); % overall EDT
        otherwise
            error('Unknown ''spec'': must be ''full'' or ''mean''.')
    end

end

function [t,E,fit] = calc_decay(z,y_fit,y_dec,fs,f)
% CALC_DECAY calculate decay time from decay curve
% Returns the time for a specified decay y_dec calculated
% from the fit over the range y_fit. The input is the
% integral of the impulse sample at fs Hz. The function also
% returns the energy decay curve in dB and the corresponding
% fit.

    E = 10.*log10(z); % put into dB
    E = E-max(E); % normalise to max 0
    if any(isinf(E))
        E = E(1:find(isinf(E),1,'first')-1); % remove trailing infinite
values
    end

    % find yfit x-range
    IX1 = findDB(E,max(y_fit),1,f);
    IX2 = findDB(E,min(y_fit),length(E),f);
    IX = IX1:IX2;

    % calculate fit over yfit
    diff_y = abs(diff(y_fit)); % dB range diff
    x = reshape(IX,1,length(IX));
    y = reshape(E(IX),1,length(IX));
    p = polyfit(x,y,1);
    fitLength = max(length(E),(1.1*diff_y/abs(length(E)*p(1)))*length(E));
% evaluate fit over sufficient dynamic range
    fit = polyval(p,1:fitLength); % actual fit
```

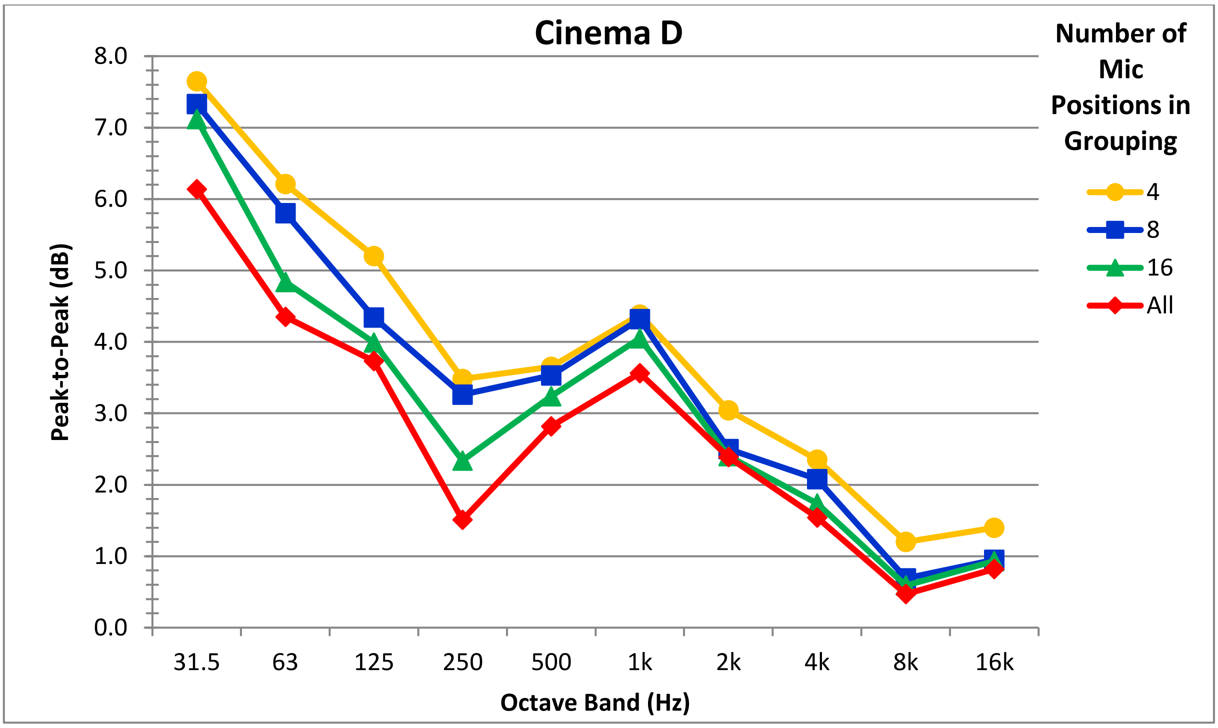
```
    fit0 = fit-max(fit); % fit anchored to 0dB
    t = (y_dec/diff_y)*findDB(fit0,-diff_y,[],f)/fs; % estimate decay time
    fit = fit(1:length(E));

end

function IX = findDB(E,dB,default,f)
% FINDDB find dB value in energy decay

    IX = find(E<=dB,1,'first');
    if isempty(IX)
        if isempty(default)
            error('Impulse response has insufficient dynamic range at %i Hz
to evaluate at %i dB.',f,dB)
        else
            warning('Impulse response has insufficient dynamic range at %i
Hz to evaluate at %i dB. Evaluating at %.1f dB instead.',f,dB,E(default))
            IX = default;
        end
    end

end
```

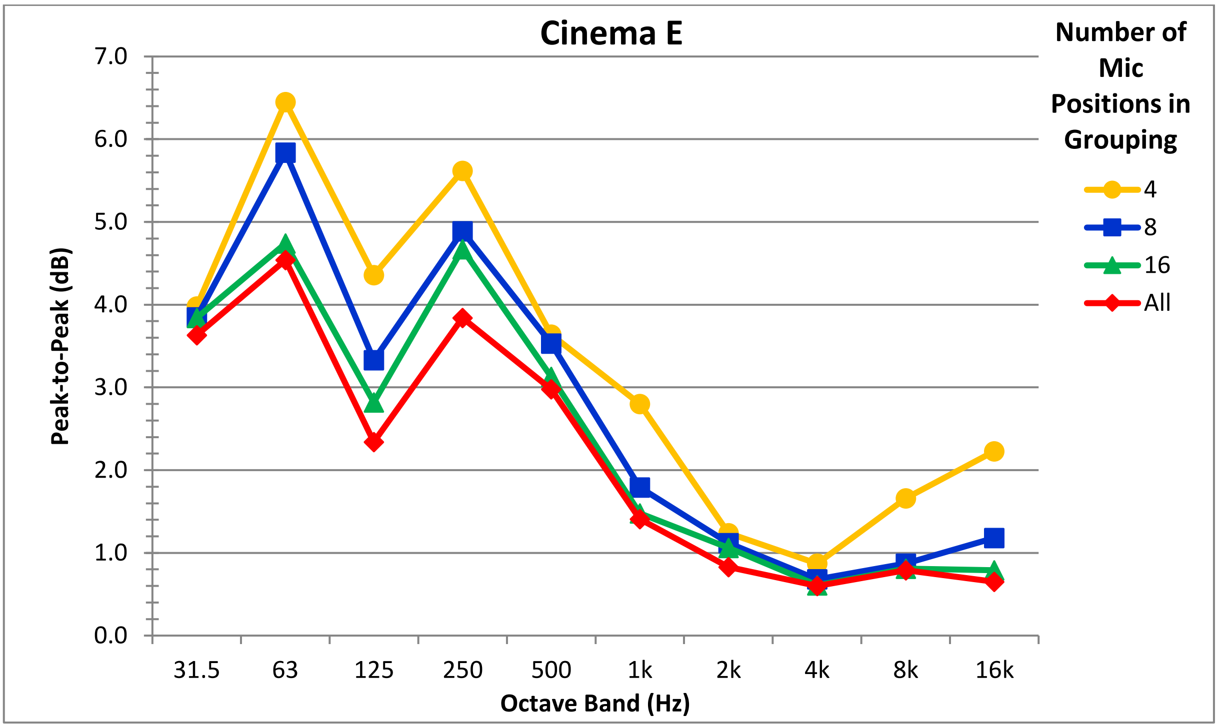
Appendix P - Peak-to-Peak Values by Position Grouping

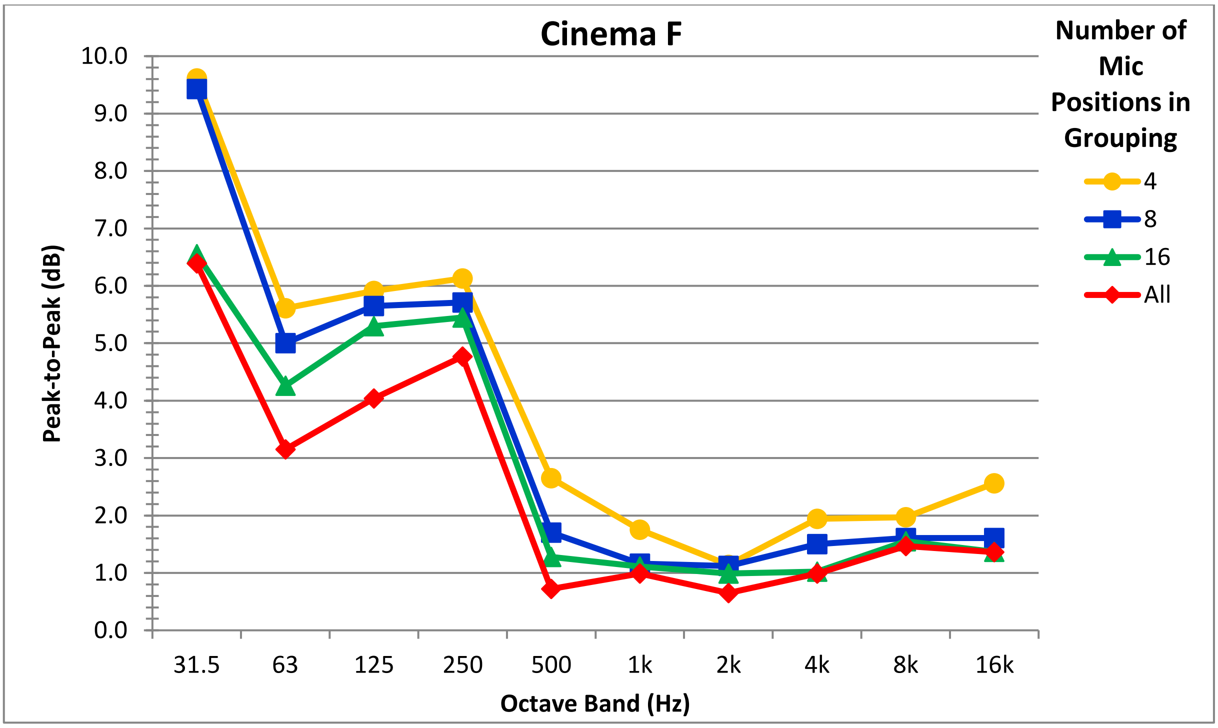


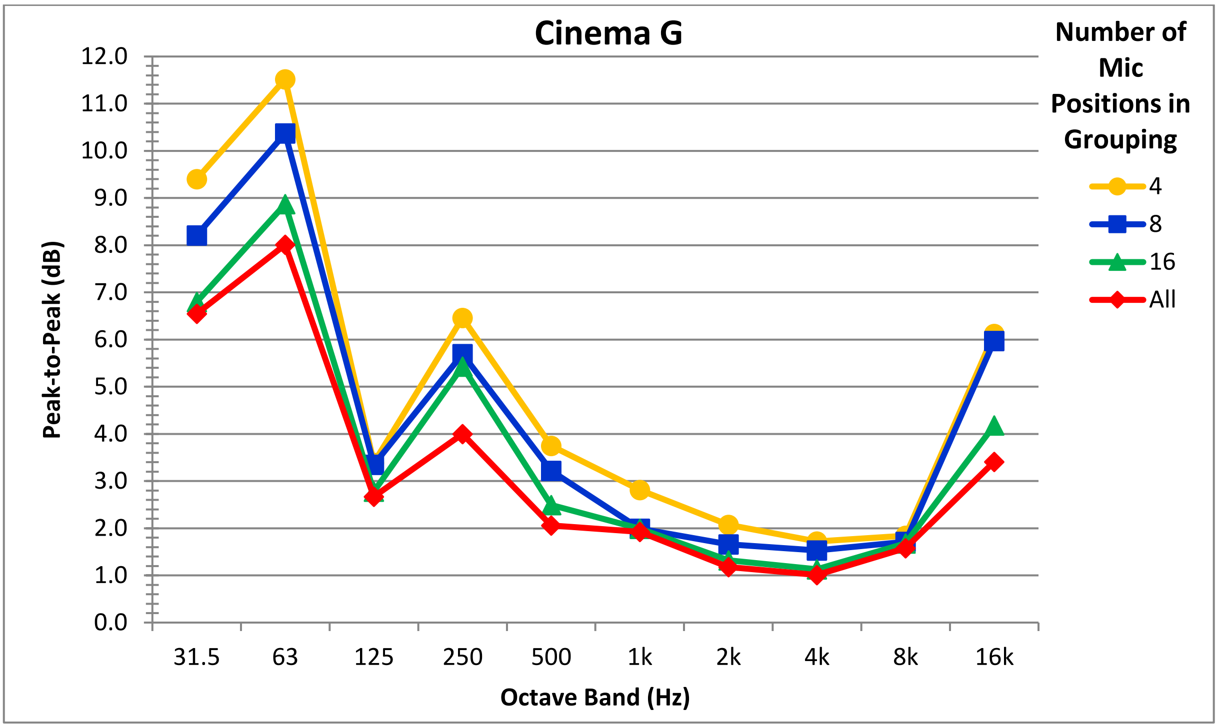



Cinema C
Number of Mic Positions in Grouping
4
8
16
All
Peak-to-Peak (dB)
10.0
9.0
8.0
7.0
6.0
5.0
4.0
3.0
2.0
1.0
0.0
31.5
63
125
250
500
1k
2k
4k
8k
16k
Octave Band (Hz)


Cinema D
Number of Mic Positions in Grouping
4
8
16
All
Peak-to-Peak (dB)
8.0
7.0
6.0
5.0
4.0
3.0
2.0
1.0
0.0
31.5
63
125
250
500
1k
2k
4k
8k
16k
Octave Band (Hz)


Cinema E
Number of Mic Positions in Grouping
4
8
16
All
Peak-to-Peak (dB)
7.0
6.0
5.0
4.0
3.0
2.0
1.0
0.0
31.5
63
125
250
500
1k
2k
4k
8k
16k
Octave Band (Hz)


Cinema F
Number of Mic Positions in Grouping
4
8
16
All
Peak-to-Peak (dB)
10.0
9.0
8.0
7.0
6.0
5.0
4.0
3.0
2.0
1.0
0.0
31.5
63
125
250
500
1k
2k
4k
8k
16k
Octave Band (Hz)



Cinema G
Number of Mic Positions in Grouping
4
8
16
All
Peak-to-Peak (dB)
12.0
11.0
10.0
9.0
8.0
7.0
6.0
5.0
4.0
3.0
2.0
1.0
0.0
31.5
63
125
250
500
1k
2k
4k
8k
16k
Octave Band (Hz)