POLITECNICO DI MILANO

Dipartimento di Elettronica, Informazione e Bioingegneria.

CINEMA SOUND: CHARACTERISTICS AND 3D ACOUSTIC MEASUREMENTS

Relatore: Prof. Lamberto Tronchin

Tesi di laurea di:

Nicola Scaroni Matr. 899457

Anno accademico 2019/2020

RINGRAZIAMENTI

Ringrazio il prof. Lamberto Tronchin per avermi concesso di utilizzare materiale per misure acustiche estremamente costoso e delicato e per il supporto professionale che ha fornito durante la campagna di misure a Roma.

Ringrazio il G. le Simone Corelli per aver condiviso con noi i suoi contatti nel mondo della cinematografia e le sue conoscenze da Re-Recording Mixer.

I miei ringraziamenti vanno inoltre ai tecnici responsabili delle sale analizzate per averci concesso di accedere a spazi solitamenti interdetti a persone non appartenenti al settore.

Essenziale per la tesi anche l'interesse e immensa disponibilità del prof A. Farina le cui profonde e chiare conoscenze nel mondo dell'audio ci hanno permesso di ottenere, a partire dalle misure, risultati acusticamente significativi.

I miei ringraziamenti vanno inoltre al Phd Daniel Pinardi, il cui supporto è stato importante, soprattutto per l'ultima parte della tesi.

SINTESI

Il Cinema sta perdendo clienti e questo anche perché oramai, con l’evoluzione della tecnologia, è possibile godere dei benefici dell’intrattenimento audio-video restando comodamente a casa propria.
Netflix e Amazon Prime stanno lentamente rimpiazzando le vecchie case di produzione cinematografica.
Gli esperti nel settore cinematografico ritengono necessario trovare un modo per fornire al cliente che entra nel Cinema un’esperienza migliore di quella domestica e questo si traduce soprattutto in un miglioramento della qualità sonora e dell’immersività.
Per ottenere questo si punta ad avere consistenza sonora nel percorso che va dalla Sala Mix al Cinema.
Dopo un’introduzione storica sul suono nel Cinema e sui suoi formati di distribuzione verranno analizzati i metodi di misura che possono essere utilizzati per verificare la consistenza sonora sopra citata.

INTRODUZIONE

Come aveva predetto Moore nel 1965, l'evoluzione della tecnologia ha assunto un andamento esponenziale negli ultimi decenni e questo fenomeno ha ovviamente interessato anche il reparto audio, consentendo una più ambia fruizione di materiale economico ma di buona qualità.

Questo ovviamente vale sia per la possibilità di poter ascoltare nelle proprie case brani musicali con una qualità sempre maggiore sia di poter godere di una soddisfacente esperienza di intrattenimento combinato audio-video che di solito sarebbe stato possibile ottenere solo andando al cinema e pagando il biglietto.

Le nuove piattaforme di streaming infatti (per esempio Netflix o Amazon Prime) forniscono un materiale di intrattenimento di buon livello sia per il video che per l'audio (Sourround 5.1 e Dolby Atmos sono infatti già implementati in molti titoli).

Un utente che possiede un sistema di riproduzione surround 5.1 ad esempio potrà godere dell'immersività sonora garantita da questo formato audio senza particolari sforzi economici e stando comodamente a casa propria.

Fortunatamente l'idea di 'andare al cinema' non ha ancora perso definitivamente la sua attrattiva essendo un'esperienza ancora differente, condivisa con altri e ancora legata alle vecchie case di produzione le quali si rivolgono come prima cosa ai cinema per le Premiere dei loro titoli.

I grandi giganti dell'intrattenimento streaming però, oltre ad aumentare la produzione di serie tv, si stanno concentrando anche sui film spesso 'rubando' il personale che prima lavorava nelle classiche case di produzione (fonici di mix, attori, rumoristi etc…).

Per non perdere la sua efficacia, l'ambiente del cinema deve impegnarsi, nonostante la crescita tecnologica, per offrire all'utente un'esperienza che non potrà trovare altrove.

In particolare, l'obiettivo ideale è quello di garantire una riproduzione in sala identica a quella nella sala Dubbing (Cinema Mix).

Già in passato sono stati fatti sforzi in questa direzione cercando degli standard che consentissero una riproduzione fedele nel percorso del suono che va dalla Sala Mix fino alla destinazione finale del Cinema e delle multisale.
Peccato che le operazioni di standardizzazione oggi utilizzate sono spesso basate su un sistema che manca di un supporto scientifico valido.
Uno dei sistemi di calibrazione delle sale cinema che viene ancora implementato si basa sull'analisi RTA tramite pink noise.
A detta di THX e molti altri questo sistema consentirebbe di 'equalizzare' la stanza in base alla forma che risulta dall'analisi in frequenza della registrazione del pink noise a 2/3 della sala, appiattendola laddove fossero presenti picchi o valli andando ad aggiustare i guadagni prima che il suono venga riprodotto dal sistema elettroacustico.
Questo sistema di misura potrebbe giustificare l'equalizzazione per ottenere una buona risposta flat per una postazione specifica ma si dimentica degli altri posti presenti in sala nell'area di ascolto.
Essendo la sala cinema un ambiente vasto, la risposta può cambiare tremendamente, a causa dei modi, non appena ci muoviamo per misurare una nuova postazione e quindi già a partire da questo fatto si può intravedere una mancanza di coerenza scientifica.
Come vedremo non è possibile 'equalizzare' una stanza e dunque se vogliamo migliorare la qualità dell'intrattenimento questa non è sicuramente la strada da seguire.
Per capire le differenze acustiche tra sala mix e sala cinema abbiamo condotto una campagna di misure a Roma studiando in loco entrambi gli ambienti.
I sistemi di misura impiegati sono stati due: uno formato da un set di tre microfoni (Soundfield, Omnidirezionale e Dummy Head Binaurale) e uno invece focalizzato sull'implementazione di un array sferico a 32 capsule chiamato Eigenmike prodotto da MH acoustics.
Con il primo sistema abbiamo calcolato i parametri ISO 3382 (per sale da performance) e con l'altro abbiamo analizzato la distribuzione spaziale del suono nella stanza.
Come potremo constatare, anche grazie al supporto della realtà virtuale, le differenze tra i due ambienti acustici sono chiare da un punto di vista percettivo.

CHAPTER 1: CINEMA SOUND THROUGH HISTORY

Figure 1: Edison's Phonograph

The story begins with one of the original American pioneers of film: Thomas A. Edison who did’ t invent motion pictures but the phonograph in 1877, a device capable of recording and playing back sound etched onto a wax cylinder.

When he met Eadweard Muybridge, the inventor of Zoopraxiscope, a moving picture device, he had the idea to use his phonograph to match sound into picture.

The problem was the little power given by the phonograph which was incapable to reach large audiences.

So, at first everyone thought the idea of cinema was meant for individual exhibition.

At the end of 1877 Edison and his lab assistant Dickson came up with the Kinetoscope and later in 1894 they put this last in combination with the phonograph creating the first Kinetophone (see fig.3) which was roughly able to synchronize the audio.

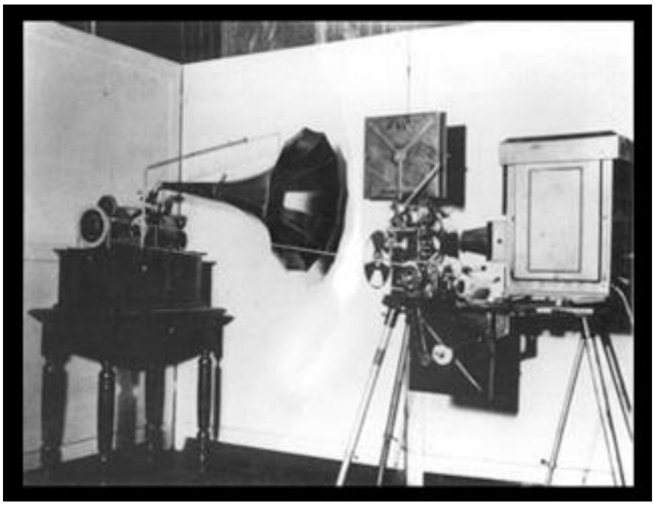

Figure 2: Kinetophone

Kinetoscope alone, without synchronized sound, took a chance in 1894 when Andrew Holland opened the first Kinetoscope parlor in New York City.

In 1900 at the Paris World Fair three new photograph synced devices were exhibited: the Phonorama, the Cronophone and the Phono-Cinema-Theatre.

The technological barriers were still a huge issue so even these brand-new devices had problems in keeping the sync between picture and sound and could playback only 5 minutes of sound.

The perfect sound-picture matching had to wait so the Cinema industry focused more on silent movies.

In 1915 Nickelodeon launched the Movie Palace in which large live orchestras were employed to play music and add some effect to the happenings on the screen.

But the implementation of musicians for live orchestra was too expensive for most of the theatres so the smallest ones decided to hire a single piano player to accompany the pictures.

One turning point happened in 1919 when three German inventors – Joseph Engl, Joseph Massole and Hans Vogt – patented the Tri-Ergon process that converted audio waves into electricity which drove a light.

This light was photographed on the film strip negative and the density was the strength of the signal.

Figure 3: A tri-ergon record

Figure 4: Variable density recording

So, instead of recording audio as a separate file this invention opened the perspective on imprinting the audio right onto the film strip itself and this solved the sync and length issues with sound on disk but not amplification.

In 1906 Dr. Lee de Forest, a giant in radio broadcasting, patented the Audition Tube, a device that could take a small signal and amplify it and in 1919 was applied to audio. In 1922 De Forest founded the De Forest Phonofilm company in the American East Coast where it supplied the entertainment for 34 theatres.
De Forest offered his knowledge to Paramount and Universal but they both preferred to keep their silent movies considering the sound as frivolous.
In the meantime, Western Electric and Bell Telephone Labs had been developing the Vitaphone, a sound on disk process that used a series of 33 inches discs but neither this took the interest of Hollywood.

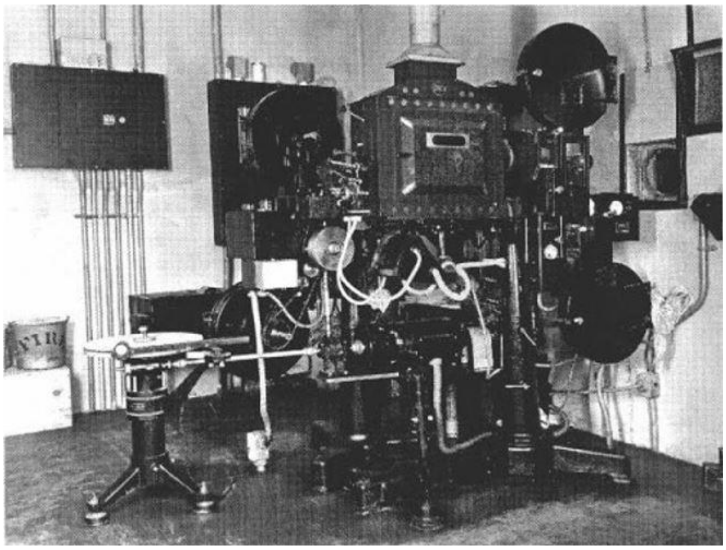

Figure 5: Vitaphone old picture

That is except for one minor studio, Warner Bros Pictures, which in 1926 established the Vitaphone corporation and they launched a premiere of 3 million dollars on August 6 of the same year in the Warner Theatre at Broadway.
The exhibition was a big success, so Warner took the show on the road hitting various city in America and touring Europe.

Despite the success the industry insiders weren’t sure about sound in film future, but the moguls began to protect themself and signed a “Big Five” agreement where the studios agreed to all adopt a single sound system.

In 1927 Warner was launched the film that cement the sound in cinema: “the Jazz Singer”.

Figure 6: Photo of “Jazz Singer” premiere

Convinced by ‘Jazz Singer’ earnings of 3.5 million dollars, a minor studio called Fox Film Corporation decided to acquire the tri-ergon technology to release three or four Newsreels per week.

From the end of 1927 everyone in film industry though that sound was not a simple fashion, but it was there to stay and in the following years it was able to attract and do big business.

In 1929 three quarters of all films made in Hollywood were released with pre-recorded sound.

It must be said that when the Great Depression started in 1929 Hollywood was able to survive thanks to audio implementation which pushed the audiences to come in droves to see the talkies.

The period that goes from 1927 to 1950s or early 60s is considered the ‘Golden Era’ (or the ‘Studio Age’) of Hollywood and this lasted till 1948 when the first Supreme Court Case United States vs Paramount took place.

This ended the studios control over their own theatre that was considered by the Court a practice on an illegal vertical monopoly.
After that a big problem came across the film industry: Television.
The launch on the market of this new device resulted in a fall of cinema attractive because people preferred to catch their shows on television than make it out to the theatres.
Cinema Industry had to find some countermeasures to give the audience an experience they couldn’t get at home, so this became the era of Widescreen Aspect Radio and huge projections alongside with the creation of Immersive feel of multitrack sound.
Till now the sound in cinema was in mono format and the first multichannel recordings were made with Fantasound in 1940’s Fantasia from Walt Disney.
But it was too expensive.
So, they came up with Cinerama (fig. 8) in 1952 that used three film strips to create a 146 degrees field of view and a total of 7 audio channels recorded magnetically onto the film strips.
Five loudspeakers behind the screen and two placed in the rear for surround sound.

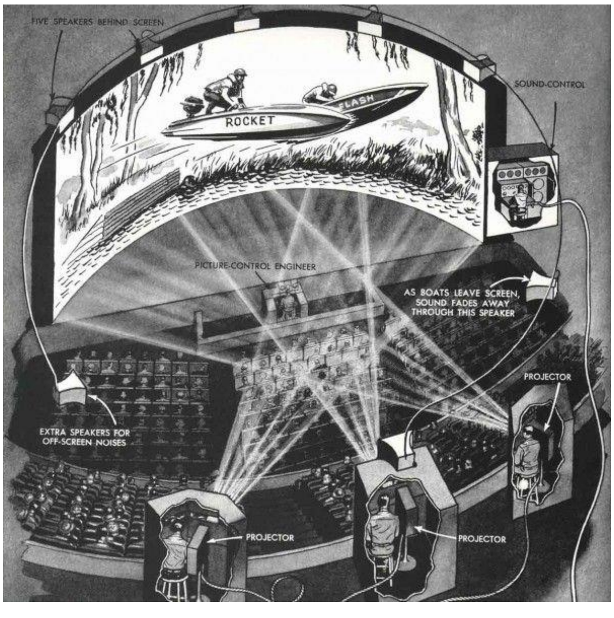

Figure 7: Cinerama

The 70mm could carry 6 channels but problems so the standard became the mono 35 mm strip that could carry only the mono sound.

In 1938 the Academy essentially standardized a frequency response - a sort of worst-case scenario which all sound mixing stages would be calibrated to so that sound editors knew what their mix would sound like even in the least capable theaters.

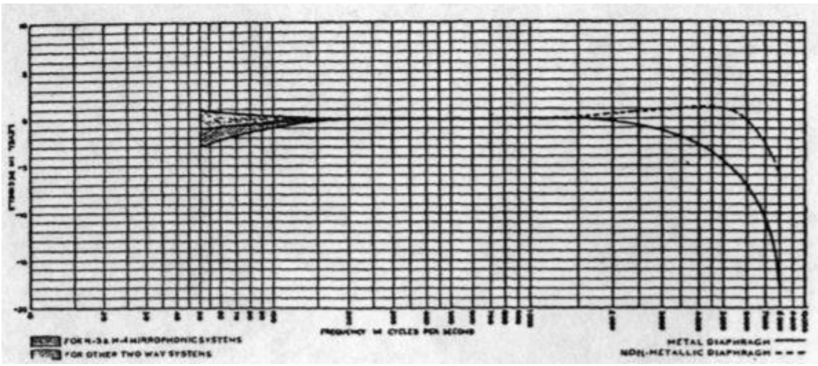

Figure 8: Academy Curve (1937)

Theaters with good sound setups would have to handicap their setups to match this Academy Curve.

To our modern sensibilities the Academy curve killed audio fidelity, but it ended up doing one very important thing: it masked the high range hiss that was so prevalent in the analog recorders of the day.

As the music industry became more sophisticated in the 60s, recording artists turned increasingly to multi-track recordings - this high-end hiss became a serious problem. If noise was bad on one channel, mixing 16 channels only amplified the problem.

One engineer by the name of Ray Dolby came up with a solution.

By splitting up the input signal into frequency bands and applying compression before recording the sound onto a tape he could record a much better signal to noise ratio on the recording medium - for playback, the Dolby would reverse the compression and the result was dramatically reduced noise.

Figure 9: Ray Dolby

In 1971, Stanley Kubrick's ‘*Clockwork Orange*’ would be the first film to use this Dolby noise reduction on all magnetic generations up to the print master though the final release print was an optical mono track.

Then one year later Dolby released Dolby Stereo with 4 channels encoded onto the two optical strips that run along the film (see fig. 11 and notice also the presence of digital audio in the middle of the strip).
These two channels were known as Left Total (Lt) and Right Total (Rt).
The center channel and surround were derived from these two channels but 3 dB down.
THX starts here making the quality assurance system so we had THX certified theatres.
In 1986 Dolby releases Dolby SR.
Coming into the 90s, in 1992, Dolby released Dolby Digital with the film Batman Returns.
Dolby Digital uses a 5.1 surround sound format using their AC-3 compression algorithm.

The digital data was printed in between the sprocket holes and the Dolby analog tracks were kept as a backup or for theaters that didn't have a digital reader.

Figure 10: Frame of the movie Star Wars

Then two years later came DTS (Digital Theatre System) that used a CD-ROM for audio playback which was synchronized to a time code embedded on the film strip.
And in the same year we see SDDS (Sony Dynamic Digital Sound) that had printed digital data on both edges of the 35 mm print supporting 7.1 surround, the first format to exceed Cinerama in terms of audio channels.
These were the last format that were developed in analog.
In the next chapter I will describe how the audio information is carried in the new movie digital format: the DCP.

CHAPTER 2: DCP

In digital cinema language, the Digital Cinema Package, or DCP, is the name that is given to the ensemble of files that are sent to an exhibition theatre.
It is simply a "box" of files that may or may not contain a complete motion picture.
On the other hand, a digital motion picture is a structured set of files that is called a Composition.

DCP SECTION

A DCP file can carry one or more Compositions (Composition Package), or eventually only a partial Composition (Asset Package).
A Packing List accompanies each DCP and it identifies the DCP's files assets.

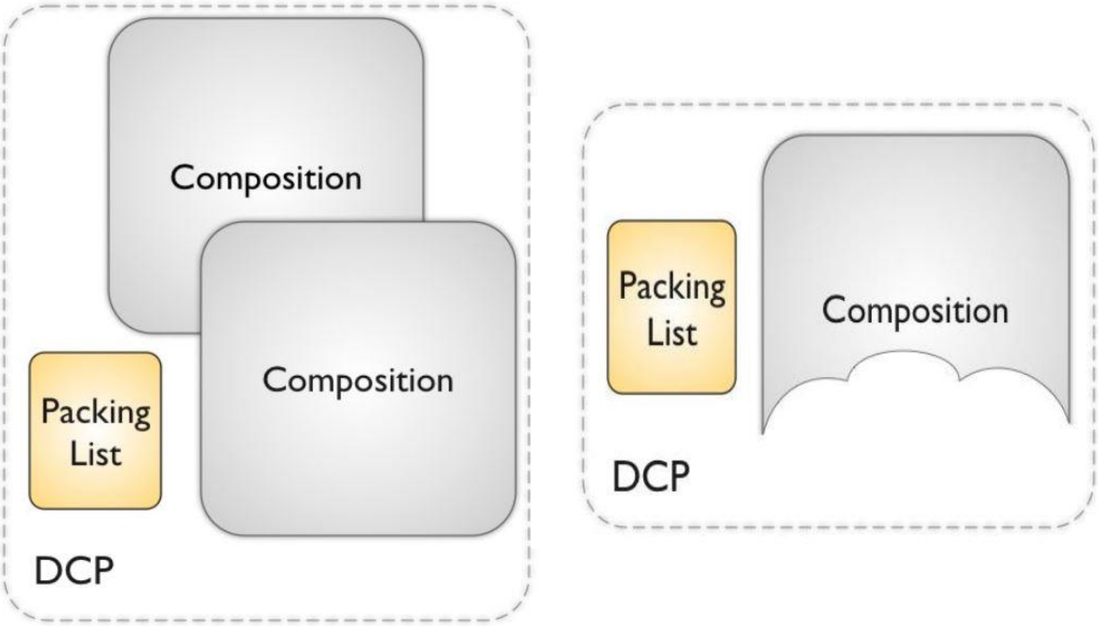

Figure 1: DCP asset

COMPOSITION SECTION

It is the final version of a work product, or title.

The Composition contains multiple files including a playlist and at least two Track Files.

Each Track File contain only one type of Essence, such as picture, sound or subtitles.

The way the files are played back in the cinema theatre is written in a playlist called the Composition Playlist or CPL.

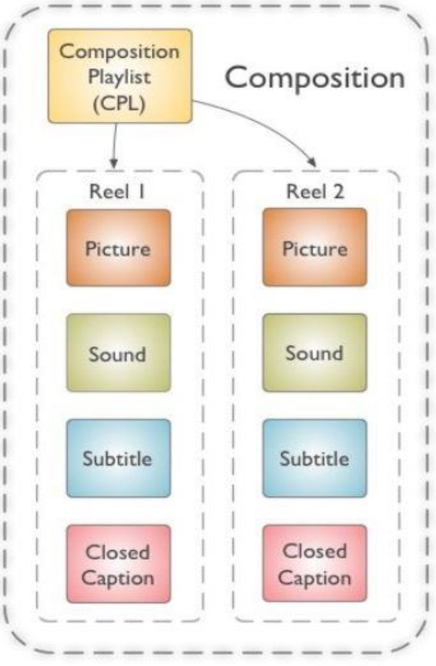

Figure 2: Composition section organization

By rule, a Composition must be composed at least of three files: A Picture Track File, a Composition playlist (CPL) and the Sound Track File.

Each Track File can be divided into multiple files consisting into pieces of essence called Reels (term coming from previous analog movie format).

The pro of the introduction of the Reels is that movies can be physically shipped easier.

On the left is shown a graphical example consisting of a Composition Playlist and four types of essence Track Files which are organized temporally as two Reels.

As we know multiple versions of a title exist.
Cinema industry need different versions to accommodate 3D, censorship cuts, subtitles or additional language soundtracks.
For each of these versions a different Composition must be created.
This can be seen as inefficient but actually allows for file assets to be shared among versions.
For example, if we want to distribute a movie in different country, we need different Compositions representing different versions of the movie.
What changes here is simply the language (English or French) but we have the same Picture Track Files.

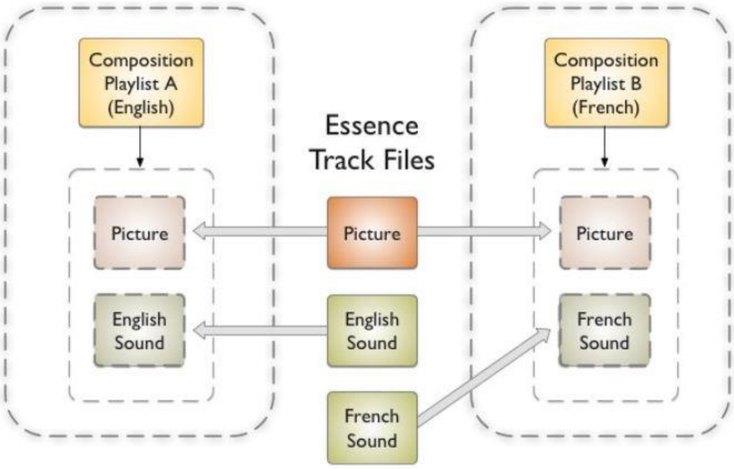

Figure 3: Interchangeability of essence files

The Composition architecture was proposed in 2001 and was a complete novelty with respect to any prior media format used in distribution.
The motion picture business has the need to distribute their content in customized versions considering different languages, sound formats, captions, aspect ratio etc.…
This is mainly driven by the fact that many cinemas have old equipment or issues so they cannot run the same Composition.
The Composition architecture was conceived to address the need to distribute multiple version of a movie to a cinema by providing a mechanism for sharing essence files among versions.

The ideas behind DCP and Composition as described in the upper lines can be found with further details in "*SMPTE ST 429-2 Operational Constraints"* which is the top-level standard for SMPTE DCP (The evolution of Interop DCP).

TITLE VERSION

In the previous discussion it was explained how multiple versions of a Composition may be created while sharing essence carried in select Track File and in the DCP section how a single Composition Package can carry multiple Compositions.
In other words, a Composition package can be used to carry multiple versions of a title (Fig 4).

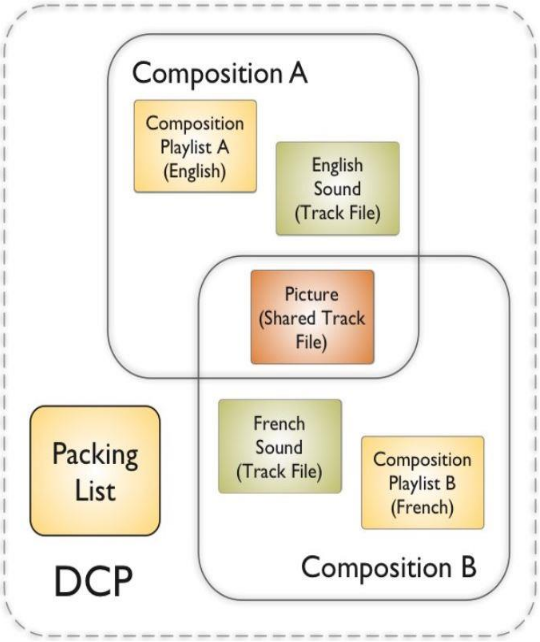

Figure 4: DCP carrying more compositions

It is often preferable to distribute title versions as separate DCP's (Fig. 5).

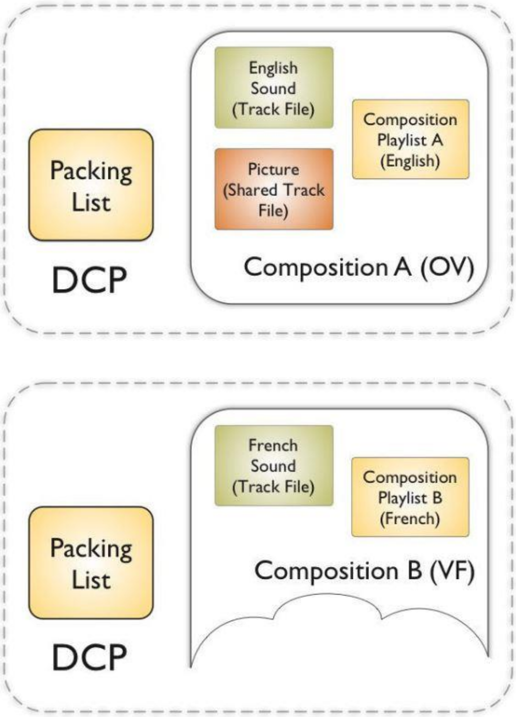

Figure 5: Multiple DCP for different compositions

In this case a parent Composition will carry the complete version of a title and one or more child Compositions versions are generated and they share some of the Track Files of the parent.

So once all the CPLs and essence Track Files are present in a common data storage, a hard disk in this case, the playback system is ready to operate.

The reproduction of a DCP child can't happen without a mechanism able to manage the distribution of parent and child Composition.

So, parent is given the tag 'Original version' or 'OV' package and is carried in its own DCP.

The child is given the tag 'Version File' or 'VF' package.

When OV package and VF package are on the same cinema server all the files needed to play the various Composition versions are present and ready to be played.

COMPOSITION PLAYLIST (CPL)

As explained before the CPL defines and coordinates the playback of all the Track Files contained in the Composition.
So, for each version of a title (different aspect ratio, sound mix, language…) there is a unique CPL.
On the contrary, the Track Files associated with each CPL do not have to be unique.
For a complete description of the SMPTE CPL refer to "*SMPTE S429—7 Composition Playlist*".

TRACK FILES

The independence of essence types in the Composition provides a good extensibility that allows new types of essence to be introduced in the future without breaking the architecture of the Composition (as happened for stereoscopic 3D).
Track Files are wrapped with a Material Exchange Format in this way:

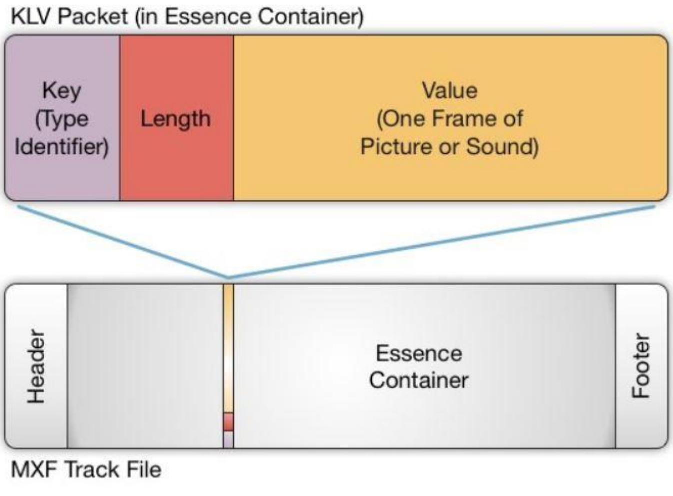

Figure 6: Track file architecture

The same happens for Subtitles and Captions files which are defined in XML and then wrapped in MXF.

ENCRYPTION

The composition is often protected with encryption (Advanced Encryption Algorithm) applied only on the Track Files.
To decrypt and be able to play the content of a Composition the theatres that purchase the movie are given a KDM of 128-bit (Key Delivery Message) file to unlock it.
For more details please refer to *Cinepedia.com.*

TODAY FORMATS

In production nowadays exist two types of DCP: Interop DCP and SMPTE DCP. And they are not interoperable.
The Interop DCP, in practice from 2004, was meant to be the preparation for the rollout of digital cinema.
However, the upgraded version of it arrived five years later and it was too late in some sense because many Venues prepared themselves as if the older version was the standard.
So, until now Interop DCP is still the main distribution format, but in the future the tests and improvement will be dedicated only to the SMPTE DCP.
Their main differences are in subtitles and audio track files organization.
Because of that many DLP projectors in America couldn't reproduce SMPTE DCP subtitle track files.
Another difference is a reliance on audio channel routing in the server, which also is not supported in many older systems.

2.1 CINEMA SOUND IN DCP: BASICS

Currently, the DCP standard allows for up to 16 channels for audio and the currently supported are the 5.1, SDDS 7.1 (five screen channels) and classic 7.1 with four surround channels.

Channel in Package	5.1	7.1	Description
1	L	L	Left
2	R	R	Right
3	C	C	Center
4	LFE	LFE	Low Frequency Effects
5	Ls	Lss	Left Surround / Left Side Surround
6	Rs	Rss	Right Surround / Right Side Surround
7	HI	HI	Hearing Impaired
8	VI-N	VI-N	Visually Impaired Narrative
9	not used	not used	Reserved for SDDS 7.1
10	not used	not used	Reserved for SDDS 7.1
11	not used	Lrs	Left Rear Surround
12	not used	Rrs	Right Rear Surround
13	Motion Data	Motion Data	Motion Seats
14	Sync	Sync	FSK Sync Signal for Immersive Sound
15	Sign Language	Sign Language	Sign Language Video
16	not used	not used	

Figure 1: ISDCF channel order

The remaining audio channels can be used to carry other data like:

- ‘HI’ +20 db for Hearing Impaired
- ‘VI’ Narrative Channel for Visually Impaired
- D-Box motion control
- Sync signal

Audio channels are hard wired to an ISDCF (Inter Society Digital Cinema Forum) agreement (Please refer to *“http://isdcf.com/papers/ISDCF-Doc4-Interop-audio - channel reccomendations.pdf “).*

There is a one to one correlation so channel 1 in the MXF Track File is routed to channel 1 of the server so to the theatre’s left speaker and so on.

Digital cinema sound is unique from every commercial distribution format because it is not compressed.

Audio is delivered to the cinema in full 24 bits/sample and 48000 samples/s.

What evolved is a more generalized concept of Cinema Sound that sees the building of sound as a forming sound bed coming from the main loudspeaker configuration with the addition of rendered sounds between the speakers.

These rendered sounds with arbitrary positioning can be produced in the 3D space of the cinema if we place another set of loudspeakers on the ceiling to add a vertical dimension.

This is the basis for what is called Immersive Sound Experience and SMPTE is working towards a common distribution format for immersive systems like Dolby Atmos, DTS or Barco/Auro (Auro 3D).

In the next chapter I’ll present some of the ideas that are behind the concept of Immersive Audio.

CHAPTER 3: IMMERSIVE SOUND

We divide the sound playouts in: Traditional Cinema Sound field and Immersive Sound field.

The Traditional Sound fields can be for example:

- Monaural (C)
- Stereo (L, R)
- 3.0 (L, C, R)
- 70mm (L, Lc, C, Rc, R)
- Surround (L, C, R, S)
- 5.1 (L, C, R, Ls, Rs, LFE)
- 7.1SDS (SDDS) (L, Lc, C, Rc, R, Ls, Rs, LFE)
- 7.1DS (L, C, R, Lss, Rss, Lrs, Rrs, LFE)

The layer of loudspeaker that produce traditional cinema Sound fields is called “Base Layer”.

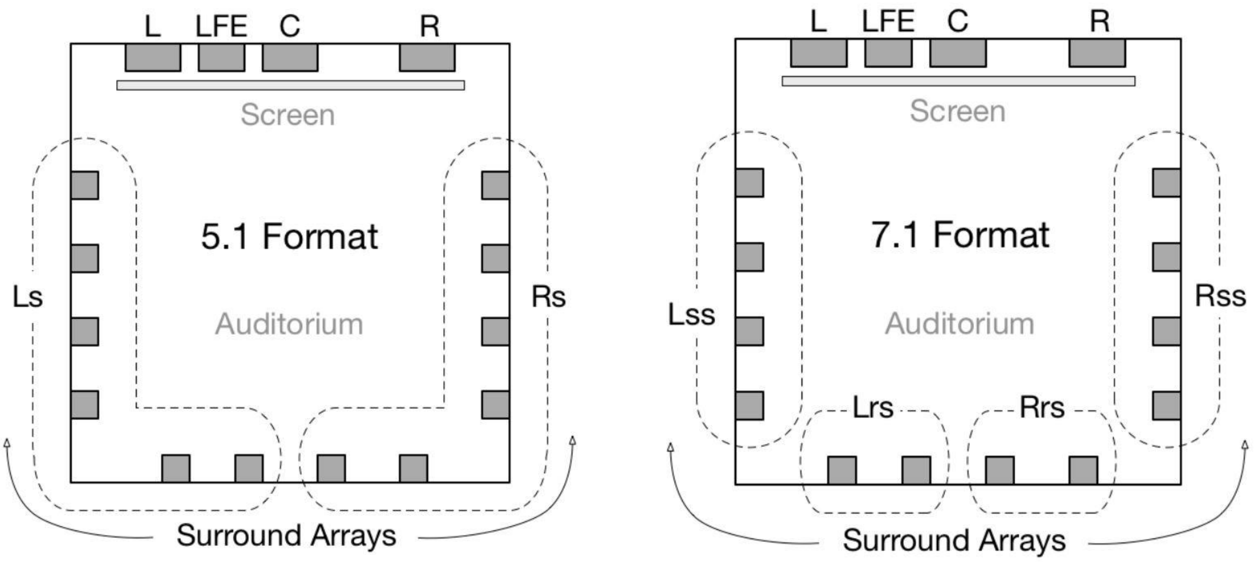

Figure 1: 5.1 & 7.1 architectures

The Immersive Sound instead adds the third dimension of height sounds above the listener with the implementation of ceiling loudspeakers.

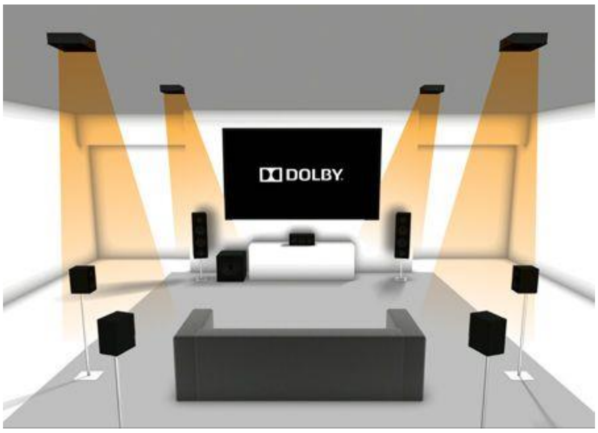

Figure 2: 7.1.4 Dolby system

In *ST 2098-5* there are two additional loudspeaker layers defined for Immersive Sound:

- Height Layer: Loudspeaker Layer placed on the walls, above the Base Layer
- Top Layer: Loudspeaker Layer placed on the ceiling over the audience

Some of the state-of-art loudspeaker architecture are shown in the following images. Please note the base layers in ‘grey’ and height and top layer in ‘blue’.

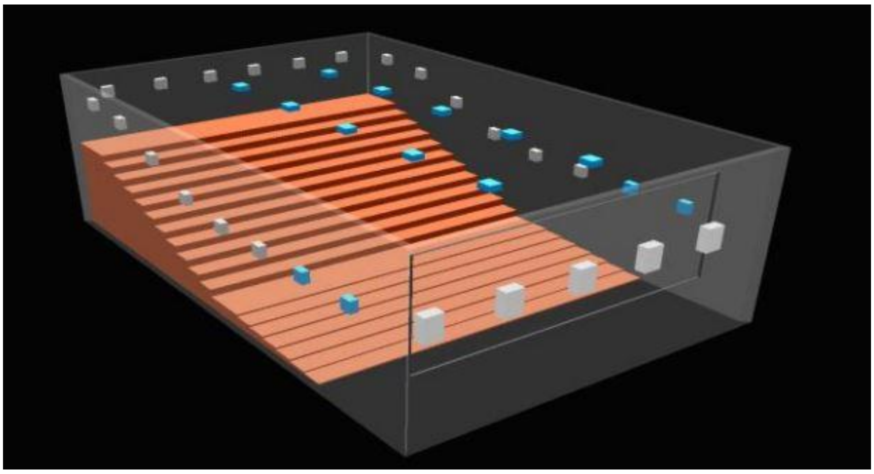

Figure 3: Dolby Atmos loudspeakers disposition

In Dolby Atmos we have 5 screen loudspeakers in the horizonthal plane, two rows of ceiling loudspeakers and a base layer of surround loudspeakers

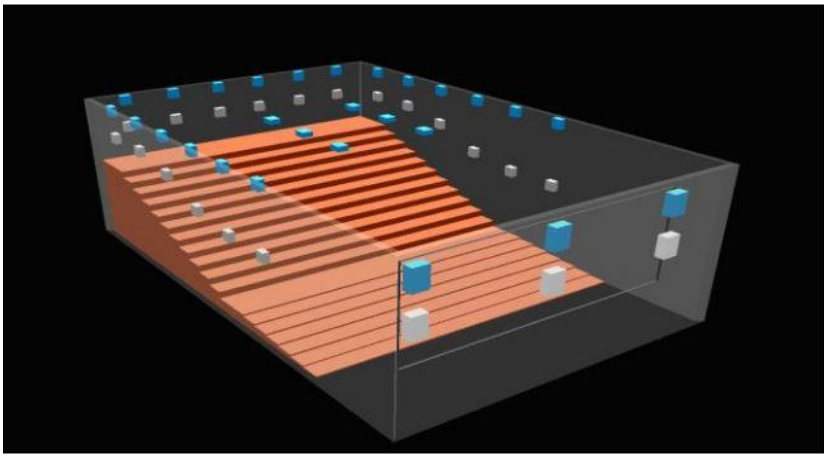

Figure 4: Auro-3D loudspeaker layout

In Auro-3D system we have two vertical layers of screen loudspeaker, two layers of surround loudspeakers and two layers of ceiling loudspeakers.
Notice the absence of audio units in the empty space from screen to audience.

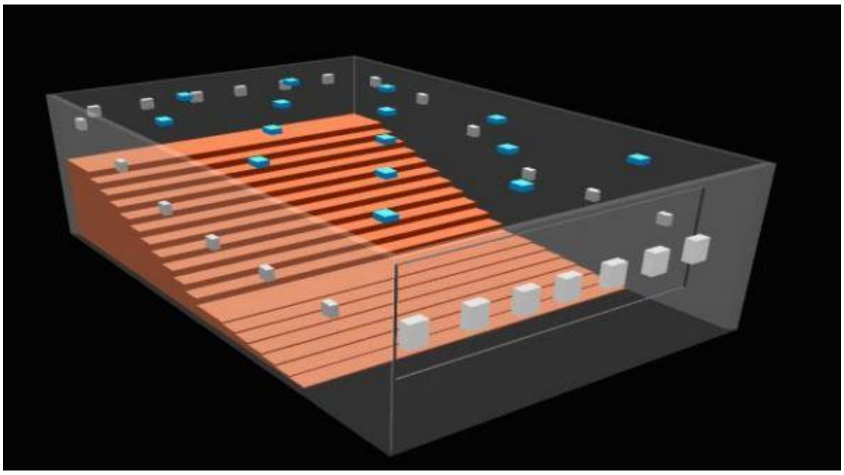

Figure 5: IOSONO loudspeakers layout

IOSONO system if composed instead by a base layer of 7 screen loudspeakers, a base layer of surround loudspeakers and some loudspeakers that cover uniformly the ceiling area.

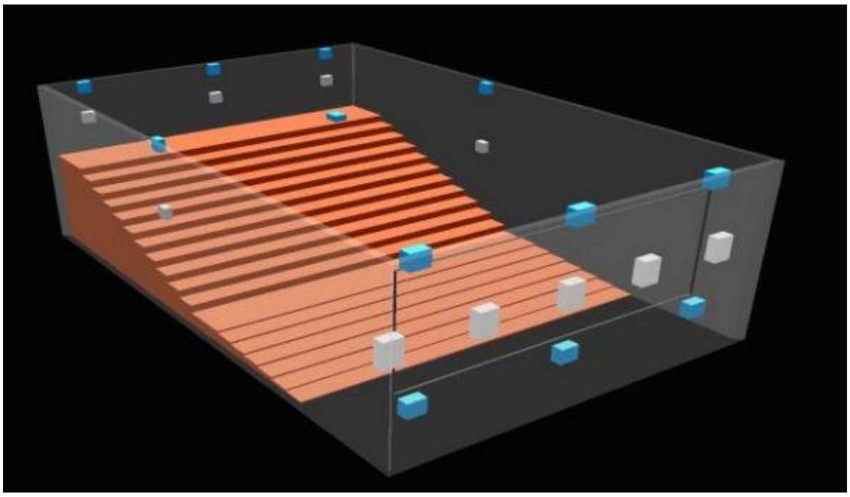

Figure 6: NHK 22.2 loudspeakers layout

This is NHK 22.2 system and its peculiarity are the three layers (base layer plus two height layers that add the vertical dimension) of screen loudspeakers.

The key idea of the Immersive audio is that, in addition to the fixed channels from the loudspeakers, we have the possibility to utilize “objects”.

The definition of “Audio Object” From SMPTE ST 2098-1:

- A segment of audio essence with associated metadata describing positional and other properties which may vary in time

- A set of audio samples and associated metadata intended for reproduction according to the position in space and other properties as indicated by the metadata.
  The position may or may not be associate with a single Loudspeaker.

In other words, an audio object is audio of any duration coupled to a metadata that describes how this is reproduced within a Sound field.
The metadata describes the position, spread, motion characteristics and other information.

According to this info the Object can move in the Sound field, being reproduced from a single speaker or virtually from any point in the 3D space.

The audio associated metadata is read by a Renderer that uses technologies such as WFS (Wave Field Synthesis), HOA (Higher Order Ambisonics), or VBAP (Vector Based Amplitude Panning), to play back the sound as intended.

Note:

By 2018 the rendering systems that were available were able to accept only one audio distribution format and were designed for one type of sound system.

This is an issue for the Cinema Industry because one need to mix and distribute specifically for the type of sound system that will play the audio, so multiple mix and multiple DCPs must be created to suit the different audio architectures.

And this is not sustainable model.

This fact scares the theatres owners that are hesitant to make an investment in the immersive sound world and this goes against the experience of the cinema goer.

The goal of SMPTE is to standardize a process to obtain single mix - single inventory - single distribution.

At the time of writing SMPTE is working also on this issue.

CHAPTER 4: MEASUREMENT METHODS

4.1 RTA METHOD

In the first part of the thesis I made a walkthrough of the history of sound in Cinema describing how it was treated in the past and how it is treated now.

In this chapter I'm going to focus more on the acoustical part of Cinema Environment describing in detail how the measurements were made in the past and how they should be done today.

I said 'should' because, as happened with Interop DCP, sometimes the scientifically and more precise way hasn't yet convinced many theatre owners who keep using the old methods that can't improve the quality of entertainment.

Figure 1: RTA Pc interface

The recommended method in "X-Curve" document suggests the use of RTA analysis to equalize the room under study.

As said in (Electro-acoustic measurements on cinema B chains in Australia) the RTA system well served the cinema industry for more than thirty years and now it's becoming obsolete.

Its assumption is: “a steady state measurement of the frequency response in the far field of a cinema when adjusted to a specific characteristic (X-curve) (red curve in Fig. 14) will ensure a flat frequency response in the near field”.

RTA is a time blind measurement so it:

- Cannot provide the complex response of the device under test which should include both magnitude and phase response.
  Phase information is important to check driver polarities and crossovers
- Cannot show the way the total system (Loudspeaker + room) frequency response evolves over time at a specific location.
- Cannot isolate direct sound and later sound so cannot show how the reproduced sound would be perceived by our hearing system which is dependent on the direct sound.

Figure 2: example of RTA measurement asset

This is the “AcoustX D2” measurement asset.

It is composed of a suitable laptop pc with the custom WinRTA software (Smaart), a set of 4 calibrated microphones, telescopic stands, a multiplexer box and long cable, a de-multiplexer box and an outboard A to D converter which connects through the USB ports of the laptop pc.

The microphones record the sound reproduced by the loudspeaker and display in real time the frequency response in 1/3 octave bands (Fig. 14).

The four microphones are disposed in non-symmetric array configuration with the main microphone on the center line at 2/3 back (Fig. 15).

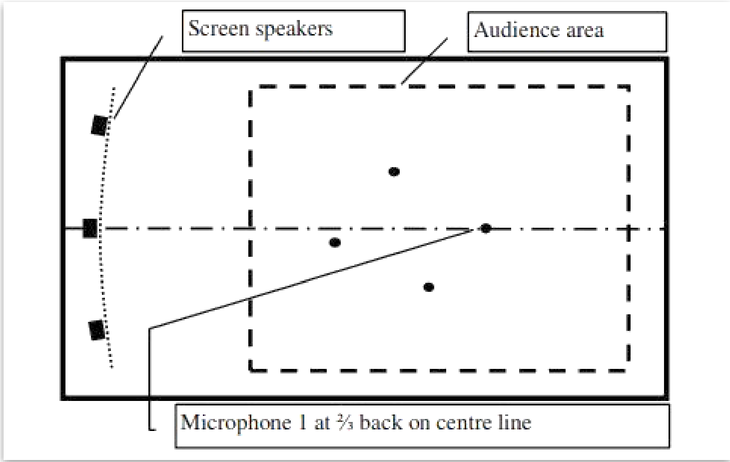

Figure 3: RTA mics layout (*SMPTE 202:2010*)

The calibration procedure starts with the Real Time frequency Analysis of recorded screen loudspeakers playing a test signal and is followed with the adjustments on the EQ and SPL levels to fit into the X-curve.

The Test signal that is being reproduced by the loudspeakers is Pink noise.

Pink noise or Flicker noise is a specific type of noise that is structured so to compensate human sensibility to different frequencies.

Its power spectral distribution follows the formula $S_{NN}(f) \propto \frac{1}{f}$

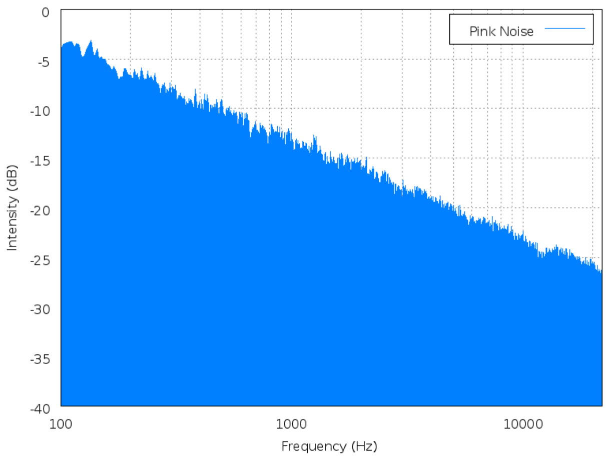

Figure 4: Pink noise Power Spectrum

4.1.1 Critics to RTA Measurement method

To understand how the RTA measurement system spread in the cinema business here is reported the statement from THX (*See Appendix A*)**:** *"Calibrated frequency response to match the X-curve, typically accomplished by loudspeaker design and 1/3 octave room EQ, which helps to standardize timbre perception".*

This is, as stated by Mc Carthy (chair of SMPTE), wrong.

The measurements techniques that has been used pink noise and real time analysis were stopped by music industry in 1968 because you're measuring modes of a room and not the loudspeaker.

In RTA if we measure the signal in one location, from the frequency analysis we distinguish peaks and valley and what everyone was doing in cinema industry was trying to fill in those peaks and valley through equalization.

Through equalization they were able to fit the frequency response into the X-Curve so they could respect the standard.

It must be said that the equalization step happens before the signal reaches the reproduction system so the myth that one can "equalize" a room is wrong because the coloration added by the environment (last filtering process before the sound arrives at the ears) is still present.

A consequence of this kind of approach was that a technician had to calibrate the movie theatres frequently because something drifted in the frequency response (changing the microphone position of some cm results in a completely different frequency response). Nothing drifted, the measurement technique was wrong.

Another wrong element present in the X-Curve standard (that suggests the use of RTA method) is that the equalization takes place by studying a frequency response that is an average of five microphones measurements.

The average has sense if we take it in the same measurement location (so to increase for example the S/N ratio) but has no purpose if it is made on five recording positions.

It has no psychoacoustical value, since the ears of human beings are not distributed in these five positions.

Moreover, we can have five bad frequency response, but their average is flat so one is brought to assume that the quality target is reached when it’s not.
All these scientific misunderstanding lead to the statement from SMPTE technicians that sometimes the equalizer was useless, and many skilled listeners appreciated more the soundtrack without the equalization, judging it to be cleaner and more enjoyable.
In the SMPTE report of 2014 the technicians made the measurements of the calibrated (as specified in SMPTE S202) rooms with FFT analysis of the impulse responses and discovered the sorry state of cinema sound.
There was no match whatsoever between the frequency responses of the various measured cinemas and dubbing stages.
The chain that brings the soundtrack from the re-recording mixer ears to the spectator in the performance theatre is broken and this is mainly due to the implementation over the years of the wrong calibration method.
So it’s impossible to equalize a room and the only way we could fix this chain is to act directly on the acoustics of the room because as Brian Mc Carty stated “*Quality loudspeakers playing in a reasonably controlled room should sound very good and should not require an equalization*”.
Now that we discussed why older measurements methods lack of scientific proofs, in the next chapter we’ll address the other measurement method based on the Impulse Response of the room.

4.2 IMPULSE RESPONSE METHOD

Another measurement technique that allows a more complete characterization of the acoustics of a room is the impulse response measurement.

As theory states, any system that is linear and time-invariant is completely characterized by its impulse response (IR).

In other words: for any input, the output can be calculated by knowing both the input signal and the impulse response.

The impulse response function of a system is the output that we obtain we excite the room with an ultra-short input signal called impulse.

So, IR is the response of the system in question when an impulse is reproduced.

The important feature of the impulse response function is that it contains all the frequencies, so it defines the response of a linear time-invariant system for all the frequencies.

With the assumption that our system (acoustics of a room) is linear (or at least a good approximation considering that some distortion is always present in the electro-acoustic system), the IRs of the loudspeakers/room system, measured with a microphone and a give source, contain all the necessary information that we need to study the frequency response AND the temporal behavior of the B-Chain and room acoustics.

The Laplace transform of the impulse response is the Transfer Function and it is defined as the ratio of the output signal to the input signal.

The Frequency Response instead is a subset of the Laplace Transform.

The Laplace Transform of a system's output in the frequency domain can be obtained by multiplying the Transfer Function by the Laplace Transform of the input signal in the complex frequency plane.

The complex frequency plane gives the complete description of the frequency domain of a system.

An example of the system described above plus some additional noise at the output in the following figure.

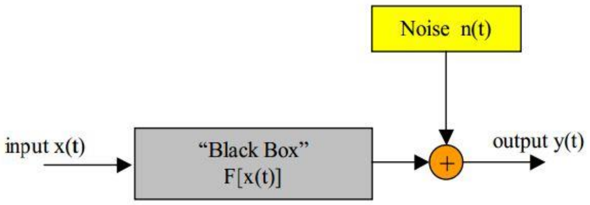

Figure 1: Generic I/O system

To determine the output directly in the time domain we require a convolution of the input with the impulse response.

In the following figures is shown an example of the extracted impulse response and its frequency analysis recorded in Teatro of Soresina, Italy.

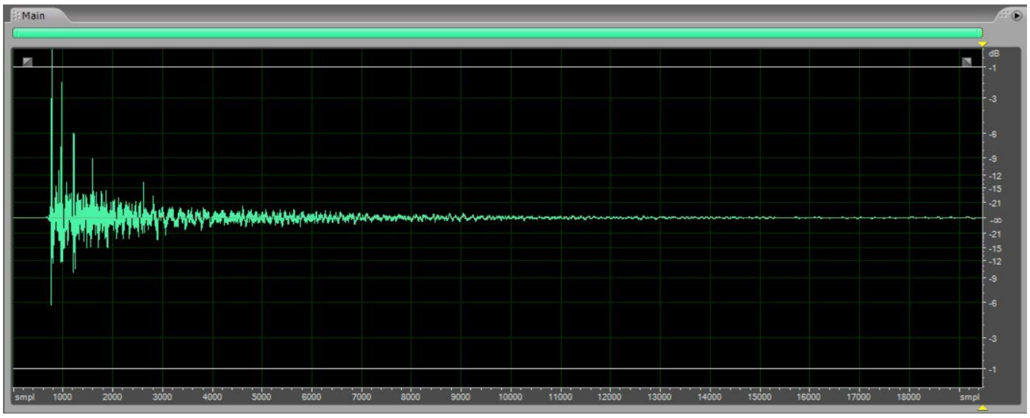

Figure 2: Impulse response in Teatro Sociale Soresina

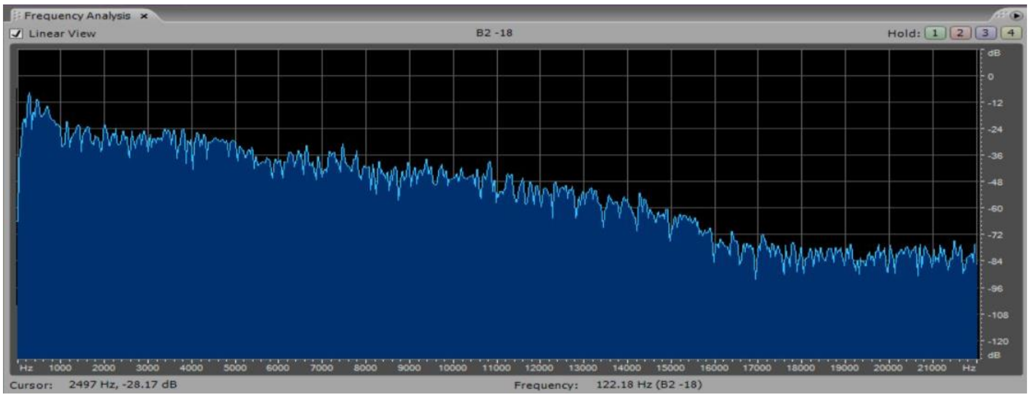

Figure 3: Frequency Analysis

In the image at the bottom of the page it is represented the division of the three temporal parts of the impulse response.
The isolated peaks represent reflections of the sound that is generated from the Source, bounces on the walls and reach the microphone.
The number of reflections is theoretically infinite so to make it easier to study the acousticians divided the impulse response’s representation in time into three parts.
The “Direct Sound” is the first peak that appears cause the sound starting from the Source reaches the microphone directly.
After the Direct Sound we can see some separated peaks.
These are called “Early Reflection” because they are the signal captured after the sound bounced a little number of times against the walls and reached the microphone.
The “early reflections" portion of the impulse response of a reverberant environment is often taken to be the first 100ms or so after the direct sound.
It is important to study the Early Reflections because they give us the information of how the room should be treated in order to get rid of them.
Last part is the “Late Reverberation” and the reflections present in this zone have happened many times against the walls so they are packed together and can’t be studied separately.

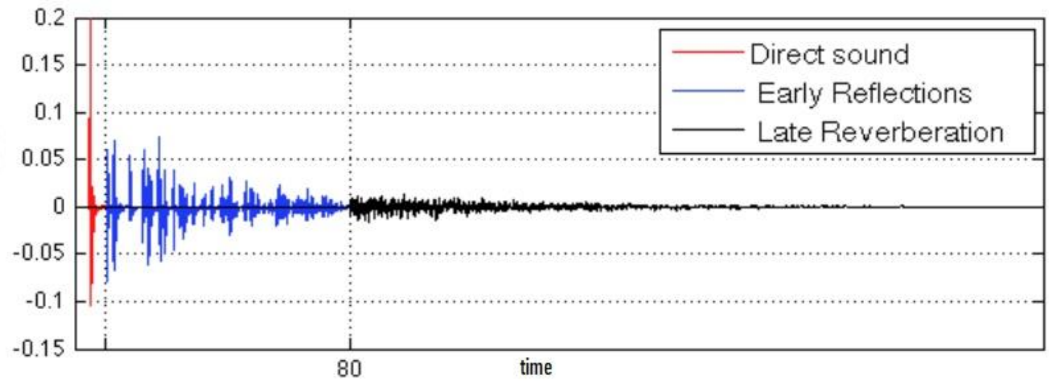

Figure 4: Temporal division of impulse response

4.2.1 TEST SIGNAL

There are a bunch of Test Signals (Pink Noise, MLS, ESS…) that are being used in the Acoustic Measurements and each of them have pros and cons.

One of the pros that some Test Signals are chosen for is the ability to distinguish the harmonic distortion artifacts (coming from loudspeakers design) from the linear response (that describes the room).

In a reverberant space excited by a loudspeaker we always find some harmonic distortion generated by the electro-mechanical transducers.

After the sound is radiated into air it travels through successive linear propagation processes like echoes, reflections or reverberation.

It's important to say that harmonic distortion can cause time aliasing artifacts: at various positions of the deconvolved impulse response some scaled copies of the impulse response appear.

These can be called "distortion products".

If we use as Test Signal a sine sweep with a linearly varying frequency these spurious peaks are not very evident as in the case of MLS and their effect can be seen in the appearance of some noise in the deconvolved h(t).

If we instead build a sine sweep with instantaneous frequency that varies exponentially with time the spurious components can be clearly seen again with their typical impulsive sound.

To visualize this phenomenon, we can look at the following Audition snapshots of the recorded sweep signal (in frequency domain) with an omnidirectional microphone in Cinema Lux, Rome.

These barely visible shadows of the sine sweep are the harmonic distortions.

It must be said that the source we used (Look Line Dodecahedron) exhibits poor harmonic distortion being a professional loudspeaker used for acoustic measurements, so this explains the scarce visibility of these artifacts.

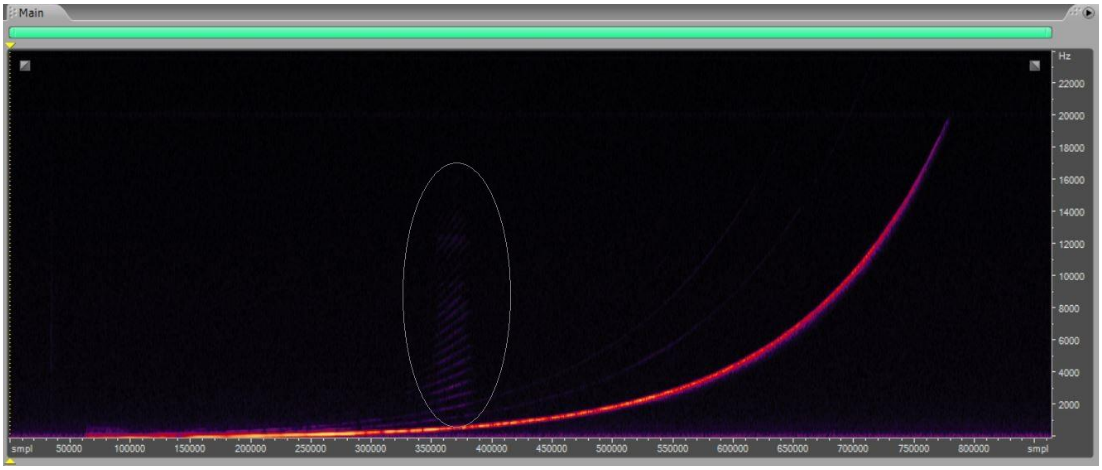

Figure 5: Cinema Lux recorded Sweep at mic position

With the use of a linear deconvolution (*) we obtain our h(t) as:

$$h(t) = y(t) * f(t) \qquad (1)$$

where f(t) is the inverse filter generated by the Pc and is simply the time-reversed version of the input Test signal x(t) that is reproduced by the loudspeaker.

With the use of linear deconvolution all the distortion products are pushed to the left of the linear response and if Exponential Sine Sweep is chosen as the Test Signal these components pack in “distortion peaks” much earlier than the linear response because their frequencies are much higher than the linear signal (see fig. 6).

This is a big advantage because even if the Loudspeaker works in non-linear region, we can still measure the system’s linear response and characterize the room under study.

Both the Linear and Exponential Sine Sweeps can do so but the second is often chosen also because of its better S/N ratio al low frequencies.

So, the harmonic distortions are packed and well separated with each other (look at the peaks in the circle) so they can be studied separately one by one as shown in the following figure.

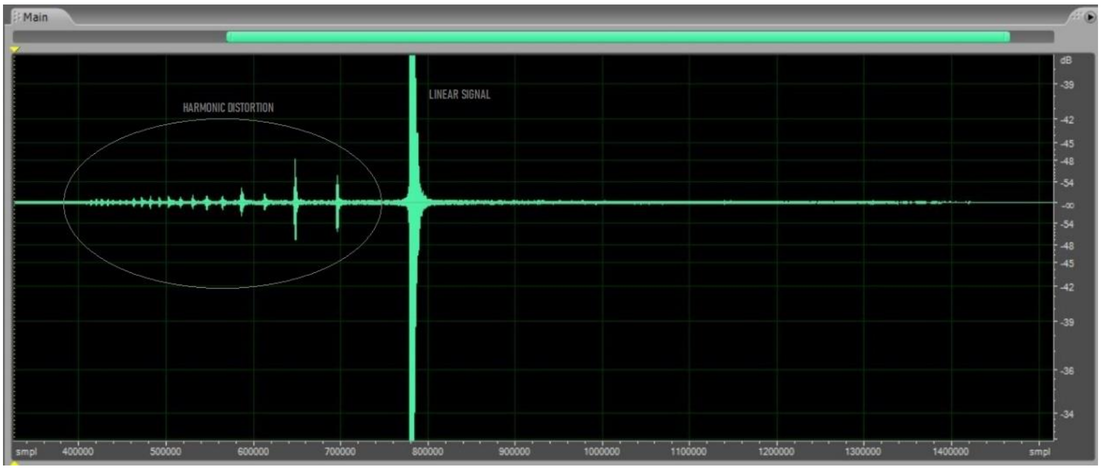

Figure 6: Cinema Lux zoomed Impulse Reponse (Behringer)

Now we will go through some theory behind the design of ESS Test Signal (our x(t)).

We start from the general form of a varying frequency sine sweep

$$\mathrm{x(t)} = \sin\bigl(f(t)\bigr)$$

with f(t) = ωt.

If the frequency varies exponentially the general form of x(t) will be the following:

$$x(t) = \sin\left\lfloor K \cdot (e^{\,t/L} - 1)\right\rfloor \qquad (2)$$

To obtain the values of the unknowns K and L we impose the conditions:

$$\frac{d\left\lfloor K \cdot (e^{\,t/L} - 1)\right\rfloor}{dt} = \omega_1 \qquad (3)$$

$$\frac{d\left\lfloor K \cdot (e^{\,t/L} - 1)\right\rfloor}{dt} = \omega_2 \qquad (4)$$

Where (2) is solved for t = 0 and (3) for t = T and yields to

$$K = \frac{T \cdot \omega_1}{\ln\left(\frac{\omega_2}{\omega_1}\right)} \; ; \qquad L = \frac{T}{\ln\left(\frac{\omega_2}{\omega_1}\right)} \qquad (5)$$

So, the required equation for the log sweep is:

$$x(t) = \sin\left(\frac{\omega_1 \cdot T}{\ln\left(\frac{\omega_2}{\omega_1}\right)} \cdot \left(e^{\frac{t}{T} \cdot \ln\left(\frac{\omega_2}{\omega_1}\right)} - 1 \right)\right) \qquad (6)$$

S/R ratio is very good because a lot of energy was diluted over a very long time and then packed into a short response.

4.2.2 HOW TO GENERATE ESS

The above all qualities and capabilities of the EES brought us to choose this last as the Test Signal.

The Test Signal was generated on my Pc with Adobe Audition 3.0 through Aurora plugin, a software developed in C and implemented as XFS plugin in Audition by prof. A. Farina.

With the series of software's plugin commands "generate ‣ Aurora ‣ sine sweep" the following window appears.

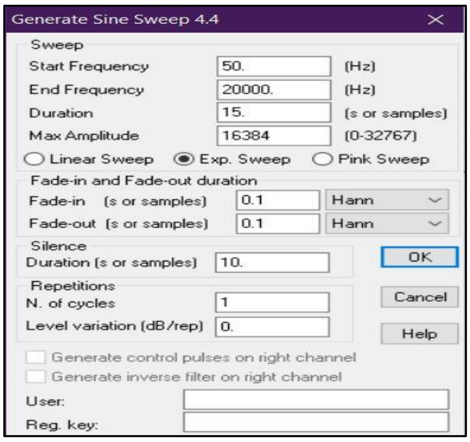

Figure 7: ESS generation window

After giving the "Ok" the signal in Fig. 8 is generated and displayed.

The inverse response of this signal (f(t) in formula 1) is stored and through the command "ctrl + v" is copied and displayed. The inverse filter's plot in time is shown in the following snapshot of the software's interface (Fig. 9).

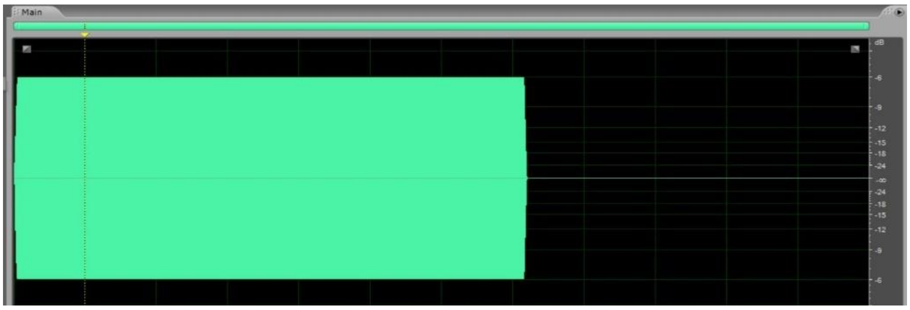

Figure 8: ESS signal in time

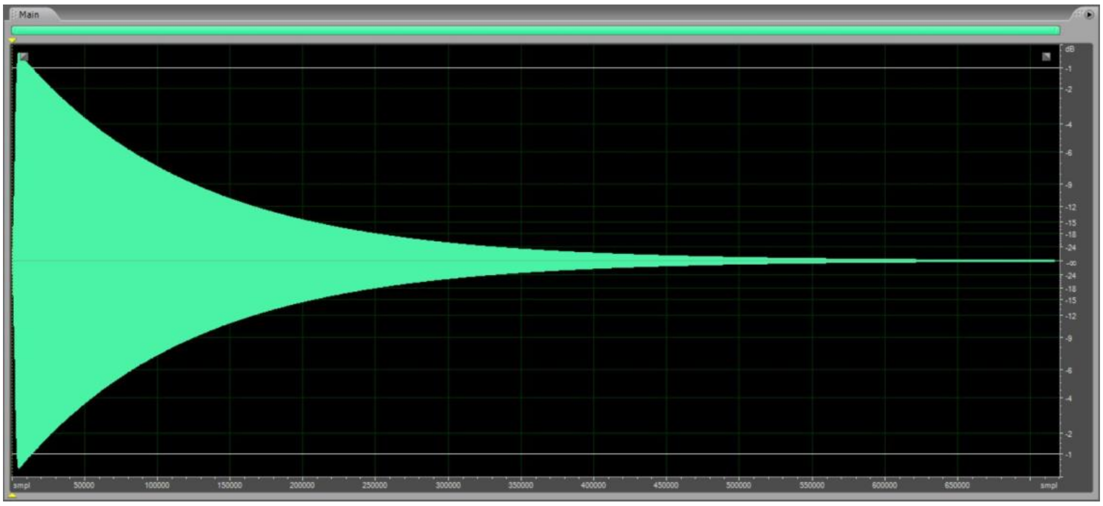

Figure 9: Inverse Sweep

For the sake of clarity this inverse sweep is the inverse filter that we'll use in the convolution with the recorded signal at the microphone (see following figure) to obtain the Impulse Response.

Figure 10: Recorded Sweep at Cinema Lux (Behringer)

4.2.3 EXTRACTION OF THE IMPULSE RESPONSE

First method

In this part it is explained how Aurora plugin implemented in Audition 3.0 computes the convolution of the single recorded signal with the inverse filter to obtain the impulse response of the room.

We used this method for the extraction of impulse response starting from the recordings took with the first implemented system (Dummy Head, Sound field and omnidirectional microphones).

So, having the generated inverse filter in memory and selecting the signal recorded at the mic with the following command order ‘effects ‣ aurora ‣ convolve with clipboard’ the following window is generated.

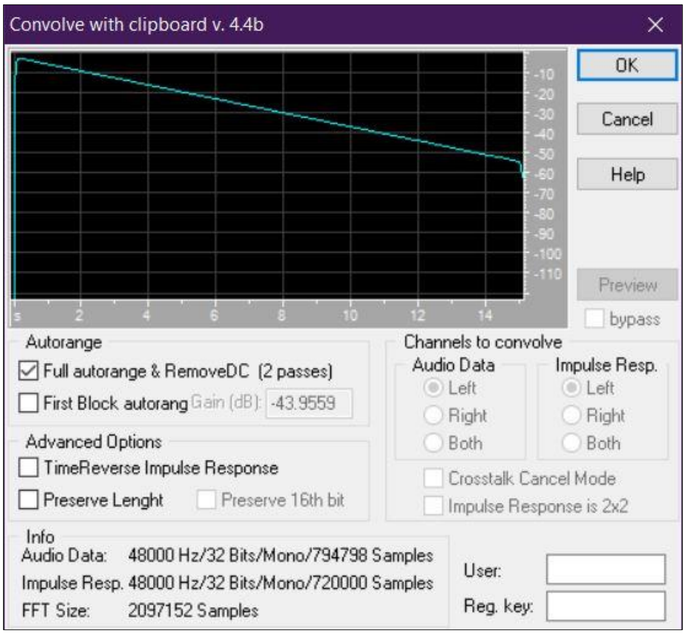

Figure 11: Convolve with Clipboard interface

After the ‘OK’ the impulse response is extracted and saved into the Audition 3.0 clipboard, ready to be saved and processed.

Second method

Due to the large number of signals recorded with the second measurement system (Eigenmike) we wanted to be able to perform the deconvolution of multiple files in one step.

This can be done through the implementation of Aurora plugins modules for "*Audacity 2.3.3*", a software that allows more freedom in the treatment of multichannel wav files. By selecting the 32 recorded sweep signals and the inverse filter in Aurora Convolver interface (that can be found in Audacity Tools) and selecting 'One for All' option, the deconvolution is performed for all the channels contained in the wav file.

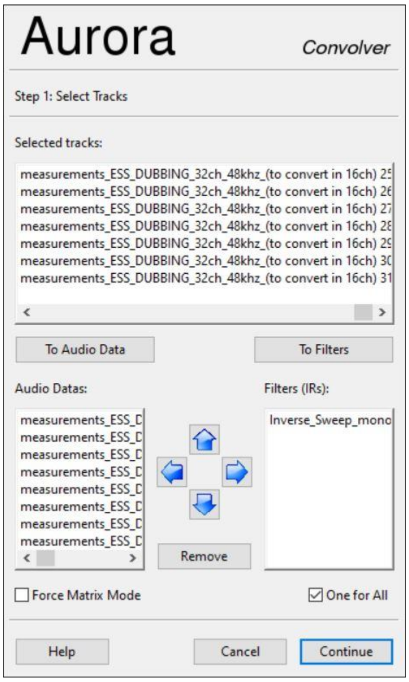

Figure 12: Aurora Convolver on Audacity interface

After pressing the "Continue" button the process starts, and the impulse responses are stored and can be saved into a new multichannel wav file.

Another way to deconvolve multiple signals is to use a MATLAB script that does the operation.

Look at Appendix F to roughly see how this operation is implemented.

CHAPTER 5: MEASUREMENT LOCATIONS

5.1 DUBBING STAGE (CINEMA MIX ROOM)

In the last week of January, I met Simone Corelli, one of the biggest Re-Recording Mixer in Italy, at “Augustus Color” dubbing stage in Rome and he explained me his fascinating job.
The first day of measurements took place in the Mixing Room n.3 of the facility.
These are the photos of the room.

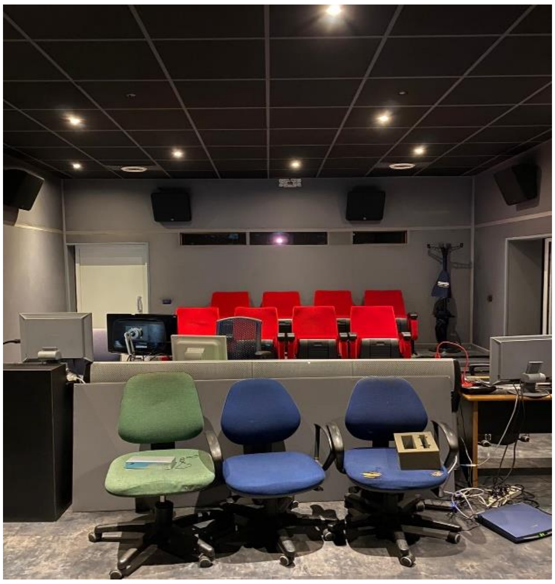

Figure 1: Back

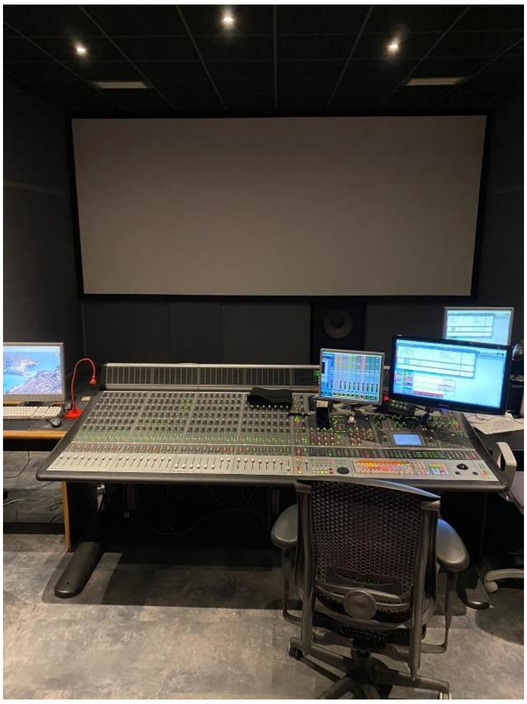

Figure 2: From mixing stage

The Re-Recording Mixer sits in the chair of fig. 10 so this should be the “sweet spot” of the room and that’s why we put our measurement systems in this location.
As we entered the room, we immediately notice the quasi-absence of any reverberation and this characteristic is probably due to the implementation of many absorbing

materials layers inside the walls, the ceiling and the two rows of chairs behind the main position of Mixing.

The installed audio system is a JBL system in 5.1 configuration, so three center channels (L, R, C) behind the perforated screen, two surround channels (Ls, Rs) attached on the walls and pointing downward and an LFE unit under the screen.

The next image shows the planimetry of the room and displays the arrangement of the measurements.

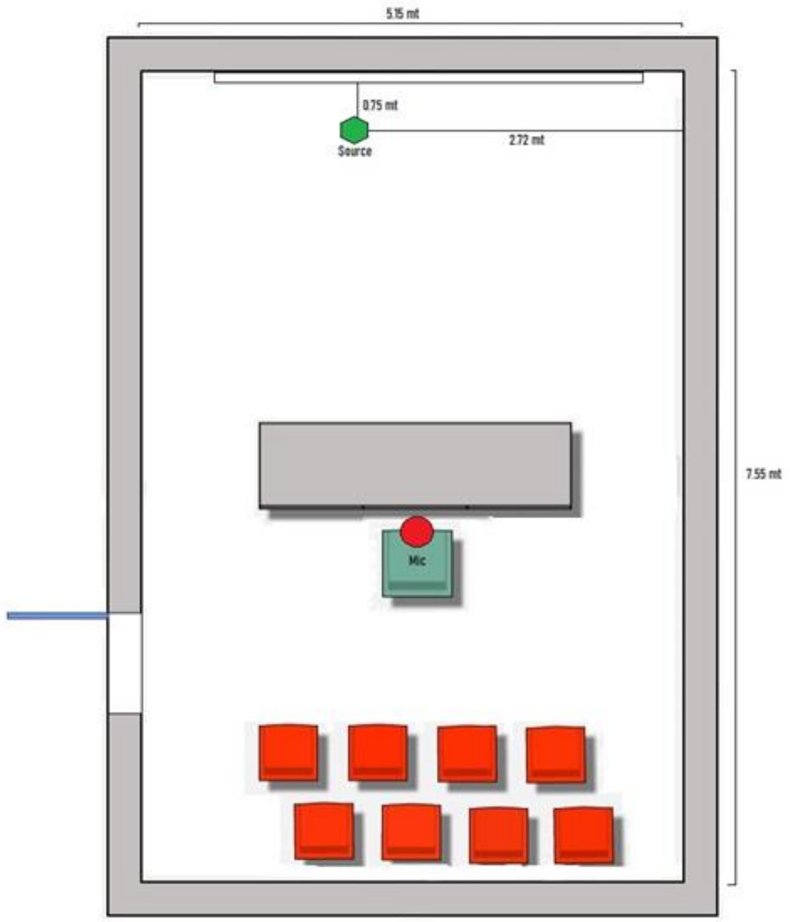

Figure 3: Cinema Mix room measurement layout

Where the Green hexagon represents the asymmetric positioning of the Source (in our case the dodecahedron) while the red circle represents the measurement point.

5.2 CINEMA LUX (ROOM 10)

After the measurements in the Dubbing Stage room I succeed in organizing another set of measurements in an exhibition theatre in Rome city center, the Cinema Lux. The possibility to measure also a performance Theatre was essential for the meaning of my thesis so I am grateful for the interest manifested by the owner of the venue. This facility is a multiplex cinema and we chose room N.10 for the measurements. Here is shown a picture of the inside

Figure 4: Inside of Room 10 Cinema Lux, Rome

Most of the walls are covered with absorbing material but the volume in cubic meters is larger than in the Dubbing Stage so we immediately perceived a longer Reverberation Time.

The geometry of this room is more complex than the simple rectangular one of the Mixing Room.

Its four angles are cut with oblique walls with different lengths if we consider the screen wall or the back wall.

Another elements of complexity come from the arrangement of the increasing rows' elevation as we head towards the back wall and the presence of a structure (see fig 12 the structure above the higher speaker) that could add some strong reflections being opposed to the screen loudspeakers (80% of the sound field comes from the central loudspeaker).

The following image shows the layout of the measurements.

The audio system installed here is a Dolby Processor 750 with a Dolby 5.1 system.

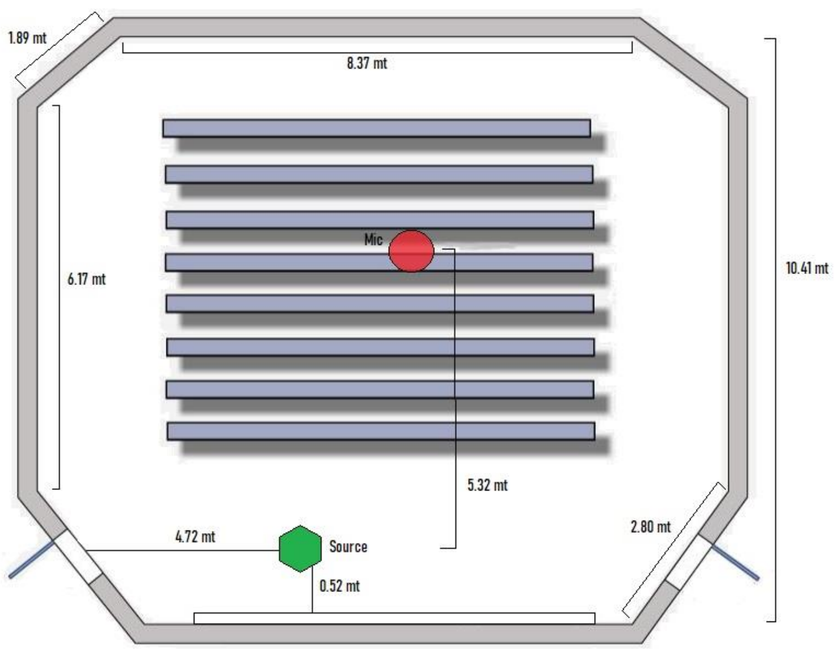

Figure 5: Cinema Lux measurement layout

The speakers are organized as following: 3 screen speakers (L, R, C), an LFE unit installed under the screen and a set of surround speakers at growing height.

We used two measurement systems: Sound field, Dummy and Behringer microphones for the first one and Eigenmike 32 capsules for the second.

So, in the following part of the thesis I analyzed separately the layout of the two systems as well as the results obtained after the processing.

CHAPTER 6: FIRST MEASUREMENT SYSTEM

The first implemented measurement method is for the Ambisonics First Order, Binaural and Omnidirectional recordings and is composed of three different microphones with different orientations.

We have a Neumann KU 100 Dummy Head, a Sennheiser Ambeo Soundfield microphone and a Behringer ECM 8000 Omnidirectional microphone.

The allocation of the listed microphones can be scoped in the following picture shot in the Mixing Room.

Figure 1: Layout of first measurement system

The entire system is located on a metal tripod where the Dummy head is screwed, and the Omnidirectional Microphone is vertically clamped (so the mic's capsule in the xy plane) with a pincer.

The Dummy Head and the Soundfield microphone located above the tripod are both pointing towards the screen.

The goal of implementing such system was to extract all the acoustical parameters described in UNI EN ISO 3382-1-2009 titled "*Acoustics - Measurement of room acoustic parameters - Part 1: Performance spaces".*

(The parameters formulas along with a brief description of them are listed in the appendix B).

As Cinema theatres can be considered a performance space, we decided to use this norm to characterize the acoustics behavior of the two rooms under study.

From here it follows some specifications on each implemented device in this kind of system.

6.1.1 LOOKLINE DODECAHEDRON

Figure 2: Look Line Dodecahedron

As Source we used the dodecahedron "*S103AC*" by Look Line which can deliver an almost flat spectrum in the 3D space.

The reason of using such a loudspeaker in the measurements is that being the speakers homogeneously distributed over the sphere it approximates the sound field generated by an omnidirectional source.

6.1.2 ZOOM F8

Figure 3: Zoom F8

The Zoom F8 is a multitrack (max 8 tracks) field recorder and this means that all the microphones can be plugged in and can record simultaneously.
The order of Microphone /Zoom channels we chose is the following:

- CH 1: W Omnidirectional-Ambeo
- CH 2: Y Ambeo
- CH 3: X Ambeo
- CH 4: Z Ambeo
- CH 5: Left Dummy
- CH 6: Right Dummy
- CH 7 Omnidirectional Behringer

Through the “rec” and “stop” buttons of the device we can register the seven Sound fields listed above and store them onto a removable SDHC unit.
All these recorded Sound fields are packed into one multitrack .wav file that can be opened through Audition or like be decomposed into mono tracks and processed easily.
The main advantage that came along with the use of this portable device was the possibility to control its interface (so to start and stop the recording) in remote, with the use of the Zoom F8 OS application installed on the iPad and linked via Bluetooth to the main unit.
Zoom F8 is also capable to provide the phantom power to feed all the microphones.

6.1.3 NEUMANN KU 100

Figure 4: Neumann KU 100 Dummy Head

The KU 100 Dummy Head is a binaural stereo microphone thought to be the substitute of the human head having two omnidirectional microphone capsules built into the ears.
If its recorded signal is listened through good quality headphones it gives us the impression of being in the scene of the acoustical event.
For its qualities this product can be used for:

- Feature Productions.
- Recording of concerts and live broadcasts in the area of classical music, jazz, pop music and entertainment shows.
- Stereo recordings with relatively simple means in acoustically very complex environments (i.e. churches).

The materials of which it is composed (which approximate human head absorbing coefficient) allow the recording of the sound field to be physiologically accurate.
The KU 100 is linked to the field recorder with balanced and unbalanced, transformer less output 3-pin XLR male connector.
The additional AC20 adapter cable that we used to link the system to the Zoom F8 can be seen in fig 14.

The yellow cable of the AC 20 adapter cable is for the left channel, the red cable for the right channel.
AC 20 male is plugged at the bottom of the Dummy Head and the red and yellow heads are plugged into the Zoom F8 in channel 5 (Left) and 6 (Right) respectively.

Figure 5: AC20 Cable

Figure 6: Sight from under

Figure 7: Sight of the inside

In this way the Zoom F8 was able to record these two mono signals coming from each ear and store them in its memory.
The next step is to create the stereo audio file that is the binaural recording, and this can be easily done with Audition CC 2019 through the creation of a new stereo file in which to paste the L/R mono.

6.1.4 BEHRINGER ECM 8000

Figure 8: Behringer ECM 8000

The Behringer ECM 8000 is a precise electret condenser microphone with a small diameter capsule and it is widely used in Acoustic Measurements because of its good quality for the price.

It is characterized by a well-balanced omnidirectional pattern and a super linear frequency response with a gentle boost in the mid-high frequencies.

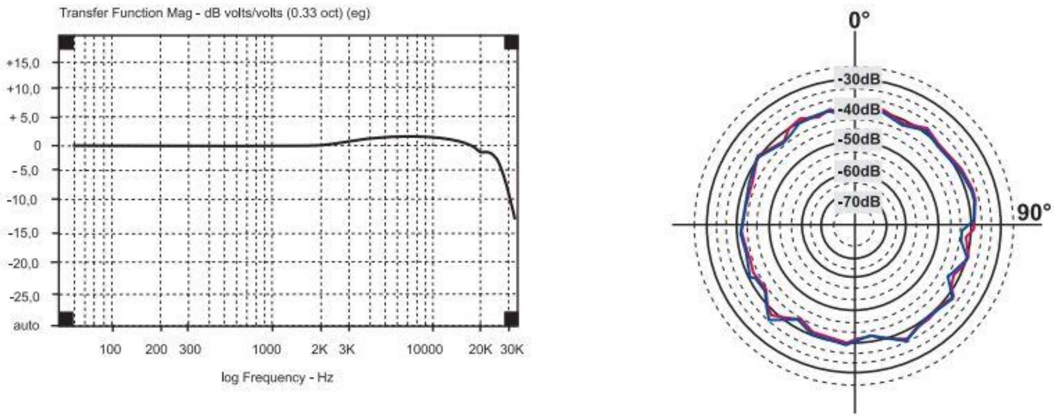

Figure 9: Frequency response and Polar Pattern of ECM 8000

It is linked to Zoom F8 channel 7 with a simple XLR cable

This microphone is the one used for the extraction of Acoustical Energetic Parameters of ISO 3382.

6.1.5 SENNHEISER AMBEO

Figure 10: Sennheiser Ambeo Soundfield microphone

The Sennheiser Ambeo Soundfield is a microphone composed of four closely spaced sub cardioid microphone capsules arranged in a tetrahedron and the first microphone of this type was invented by Michael Gerzon.
This is the typical Ambisonics First Order configuration of capsules (*See Appendix for a brief explanation of Ambisonics*).
The standard audio format produced by this kind of microphone is the following signal vector:

$$B_S = (W_S,\ Y_S,\ X_S,\ Z_S)$$

- W – a pressure signal corresponding to the output from an omnidirectional microphone
- X – the front-to-back directional information, a forward-pointing velocity or "figure-of-eight" microphone
- Y – the left-to-right directional information, a leftward-pointing "figure-of-eight" microphone
- Z – the up-to-down directional information, an upward-pointing "figure-of-eight" microphone

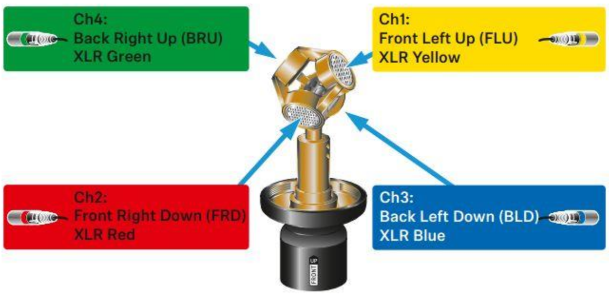

Figure 11: Ambeo capsules layout

The signals are fed into Zoom F8 with a specific cable (see fig.) that allows the output to be split into four mono tracks.

Figure 12: Ambeo channels in ZoomF8

We routed these signals in Zoom F8 starting from channel one to channel four.

It must be said that the four recorded signals together are in the so-called A-format which is a RAW format and it must be converted.

To convert to the final B-format we used “*Bidule*” coupled to the Sennheiser plugin as can be seen in the following snapshot:

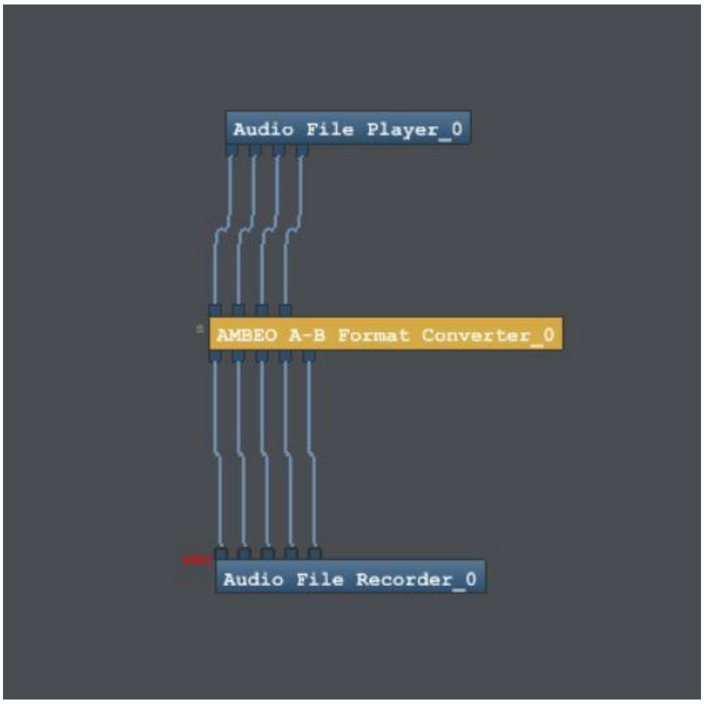

Figure 13: Bidule patch for A-B format conversion

This plugin performs the following conversion:

- C1 + C2 + C3 + C4 = W
- C1 + C2 - C3 - C4 = X
- C1 - C2 + C3 - C4 = Y
- C1 - C2 - C3 + C4 = Z

where the C terms represent the raw signals recorded by the capsules.
After the conversion the signals are used for the extraction of Spatial Parameters in ISO 3382 and precisely channels W and Y are used in the calculation of some Spatial Parameters after being merged into a stereo wav file (*see Appendix of ISO 3382*).

In the next chapter are reported and commented the results of the measurements made with the first recording system.

6.2 EXTRACTION OF ISO PARAMETERS 3382

As stated previously we used the first measurement system for the extraction of these acoustical parameters.

Formulas and description can be found in the Appendix B.

Whether we need to calculate energetic or spatial parameters we select either omnidirectional (mono) or binaural (LR) and Soundfield (WY) impulse responses to extract them.

For both energetic parameters and spatial parameters, the extraction is very simple and is performed again by the plugin Aurora implemented with Audition 3.0.

By selecting the impulse response and clicking on "Acoustical Parameters" effect the following window is displayed.

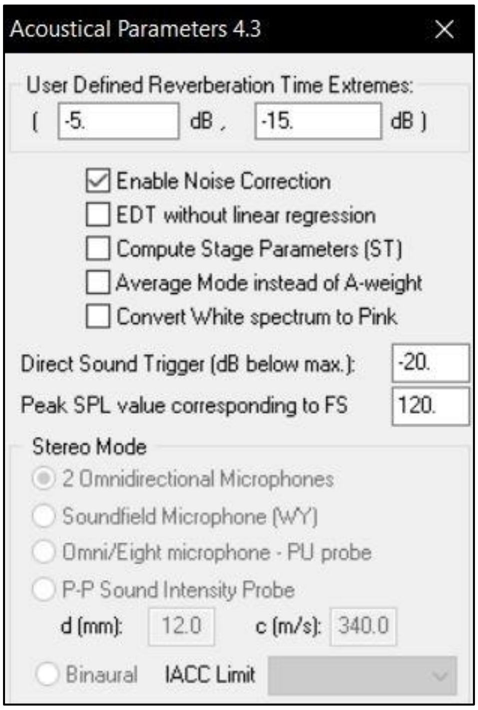

Figure 14: Acoustical parameters options

Note that the plugin allows us to define the Reverberation Time extremes and to choose some more extraction options.

We can notice also the presence of a stereo mode extraction that gets unlocked if we do the calculus of spatial parameters starting from a stereo file (binaural for IACC and fig-8 + omni for LF…).

After the “Ok” the following window is displayed and the results (in 10 1/3 octave bands) are copied to the Clipboard.

Then we can paste them as table in Excel for the processing of the graphs.

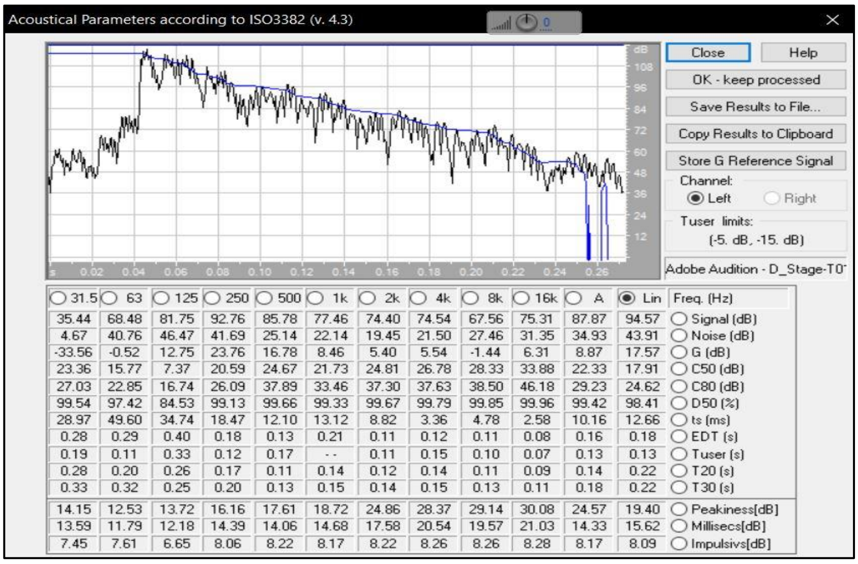

Figure 7: Parameter extraction result window

In this window is plotted the energetic impulse response and the decay.

The software is based on the Schroder integral theory: it calculates the reverberation time by simply recording an impulse response, then the decay can be obtained from a backward integration.

The software loads the impulse response from the Adobe Audition, filters the signal with the appropriate octave band filter, calculates the Schroder integral and then calculates all the parameters (for each octave band).

6.3 IMPULSE RESPONSE ANALYSIS

Here are reported the two zoomed impulse responses from omnidirectional measurement to have a glimpse on the general differences between the rooms.

It is immediately noticeable the different reflections density in the time evolution.

In Cinema Lux the reflections are more packed than in the Cinema Mix room and that’s mainly due to the different reverberation times (0.2s for the Dubbing Stage and 0.35s for the Cinema Lux).

With longer reverberation time the sound can bounce more on the walls explaining the higher density of the reflections.

The Cinema Mix room has a smaller volume so here the reflections are fewer and more separated.

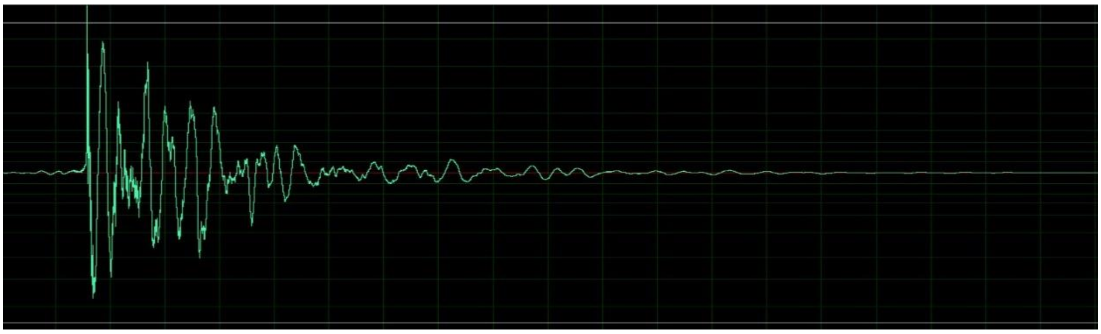

Figure 16: Cinema Mixing room IR

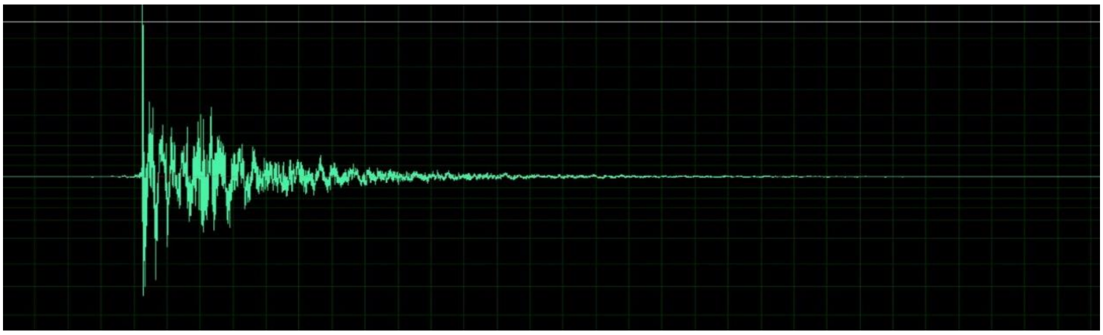

Figure 17: Cinema Lux IR

I reported also the Frequency Analysis plots of the two Impulse Responses.

In Cinema Lux the Frequency analysis shows an almost flat behavior across all the frequencies except for a gentle roll off at the high frequencies from 12 kHz probably due to air absorption.

In the Dubbing stage the frequency analysis shows instead a higher boost in the low frequencies and a general roll off towards high frequencies.

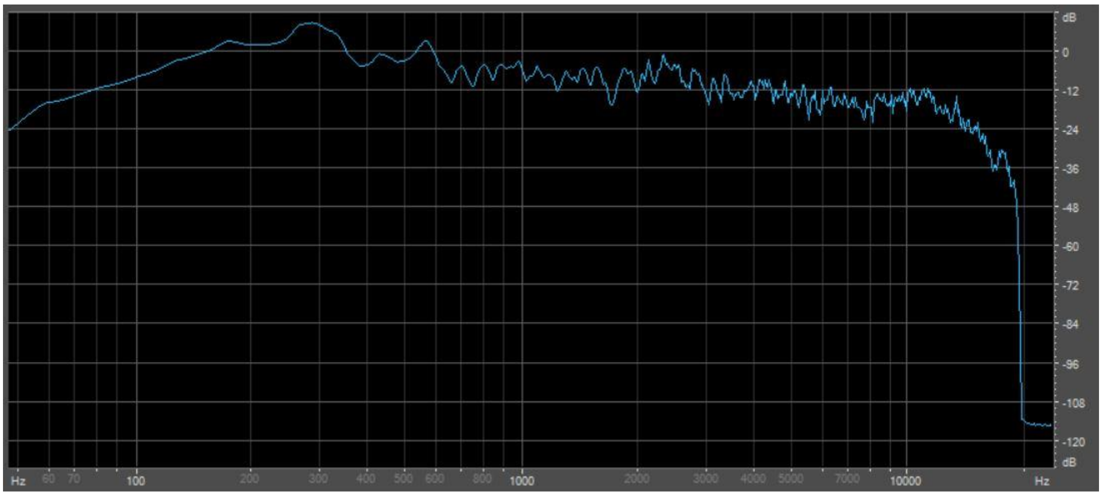

Figure 18: Cinema Mixing room frequency analysis

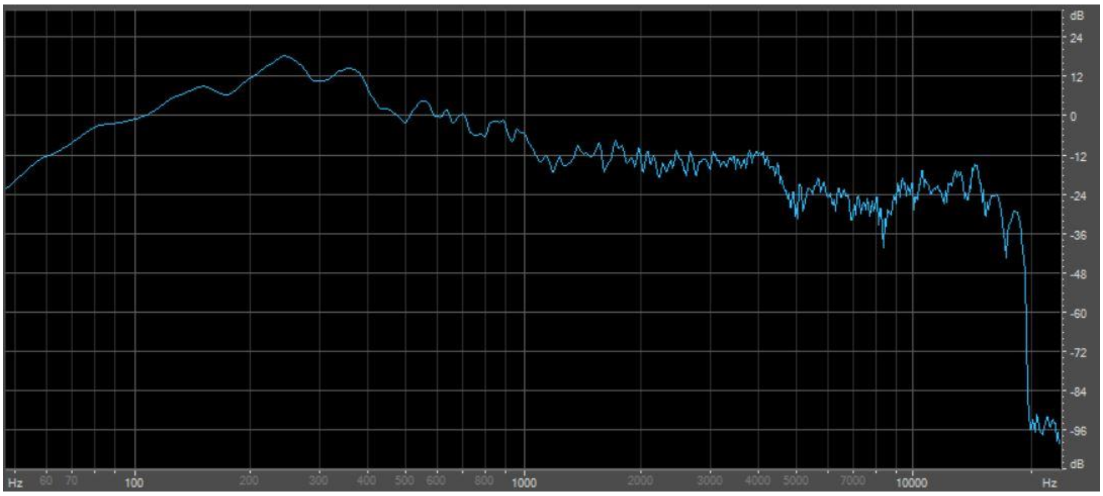

Figure 19: Cinema Lux frequency analysis

6.2.1 Energetic Parameters

The classical Reverberation Time (T60), is obtained by calculating the time interval needed for the sound level to decrease 60 dB.
From the analysis performed on “IQ-Reverb”, a VST plugin implemented in FL Studio we extracted the T60 values: 0.35s for Cinema Lux and 0.2s for Dubbing stage.
Extrapolations are necessary and usually the **T30** or the T20 (corresponding to a 30 dB decay or 20 dB, respectively) is measured, between -5 dB and -35 dB (or -25 dB), and then multiplied by 2 (or 3) in order to make it equivalent to the T60.
These approximations to T60 are performed mainly because in most of the cases the background noise appears before -60 dB and his presence can ruin the measurements.
In the following graphs are represented the values in 10 octave bands of the parameters for both the Dubbing stage room and the Cinema Room.

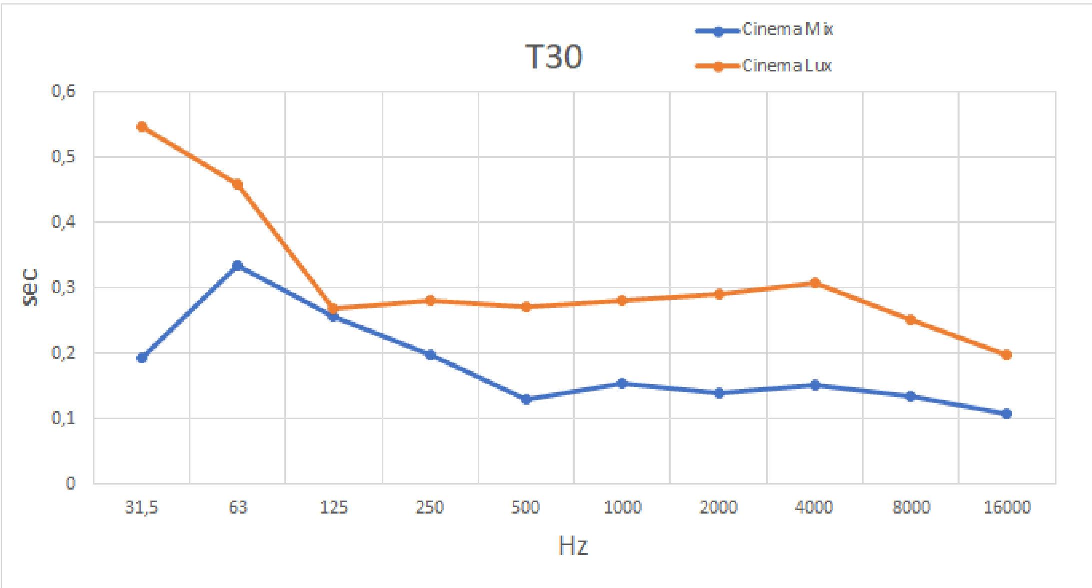

Figure 20: T30 plot

The first acoustic perception of a higher reverberation time in the Cinema room was correct.

The Orange curve representing the values of T30 of Cinema Lux is always above the Dubbing stage blue curve except for the 125 Hz octave where they match.
As expected for both the rooms we notice a general higher T30 in the low frequencies and a gentle roll off towards mid-high frequencies.

EDT (Early Decay Time) is also part of the T60 approximations being the time that the energy takes to go from -5 dB to -15 dB and multiplied by 6.
This parameter is correlated to the early decay of reflections (early reverberation).
The Ideal value for EDT stands in the range: $0.75 \cdot T_{mid} < EDT < 0.9 \cdot T_{mid}$
where T_{mid} is the average value of T30.
T_{mid} is the avg of T30: for Dubbing stage is 0.222s and for Cinema Lux is 0.288s.

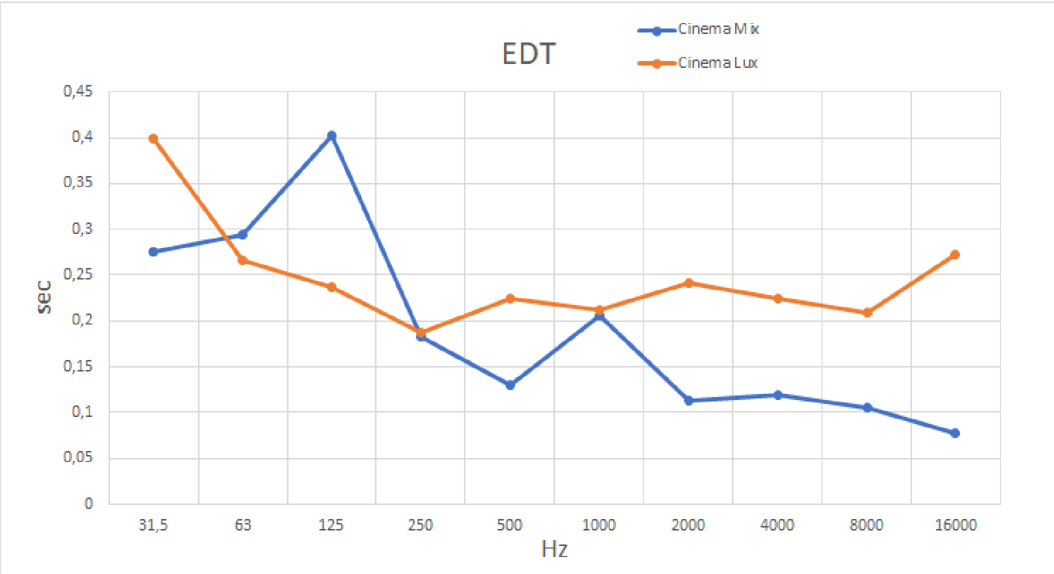

Figure 21: EDT plot

From EDT analysis we immediately notice that for 125 Hz octave band the Dubbing stage room shows a higher value than Cinema room and this is probably due to the interaction of a mode in this octave range that creates a resonance so that the sound at that frequency is slower to decay.

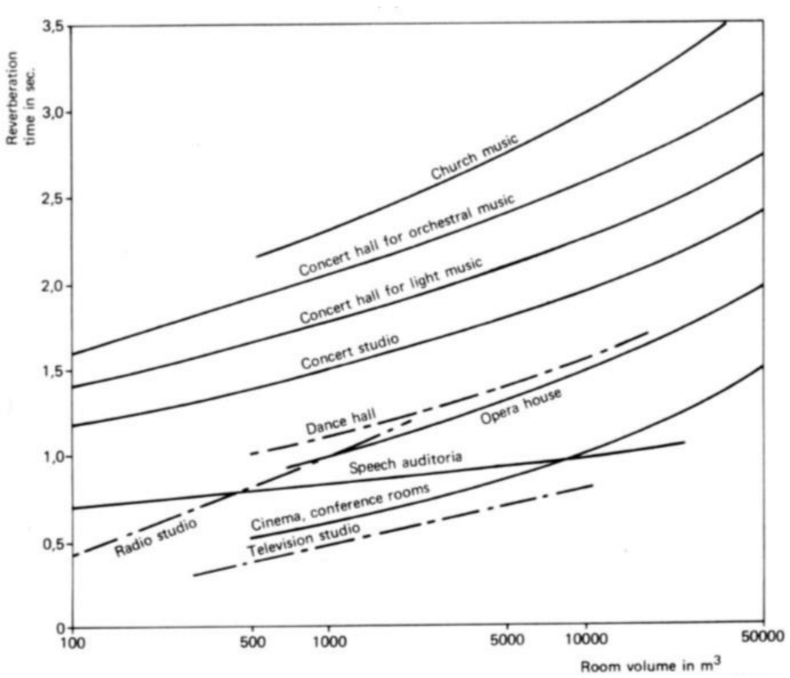

Figure 22: Ideal RT values for different performance spaces

In this graph are reported the Ideal values of Reverberation Time for different Performance Spaces.

Cinema and Conference rooms request a value that goes from 0,5s to 1,5s and in our measurements in cinema Lux we found values standing below this range.

The Dubbing stage Room instead can be considered a Television studio with Reverberation Time values ranging from 0,25s to 0,75s and the value we found falls just at the beginning of this interval.

It must be said that generally the two rooms under study were quite “dead” and maybe some reverberation could help to improve the quality of the perception.

The next Parameter we analyzed is **Clarity index**.

This Parameter express a balance between Early and Late arriving energy, that is useful to measure the Clarity as perceived by human ears.

The value of t_e (pedix value after C) is 50ms or 80ms, depending on the destination of the room, speech or music listening respectively.

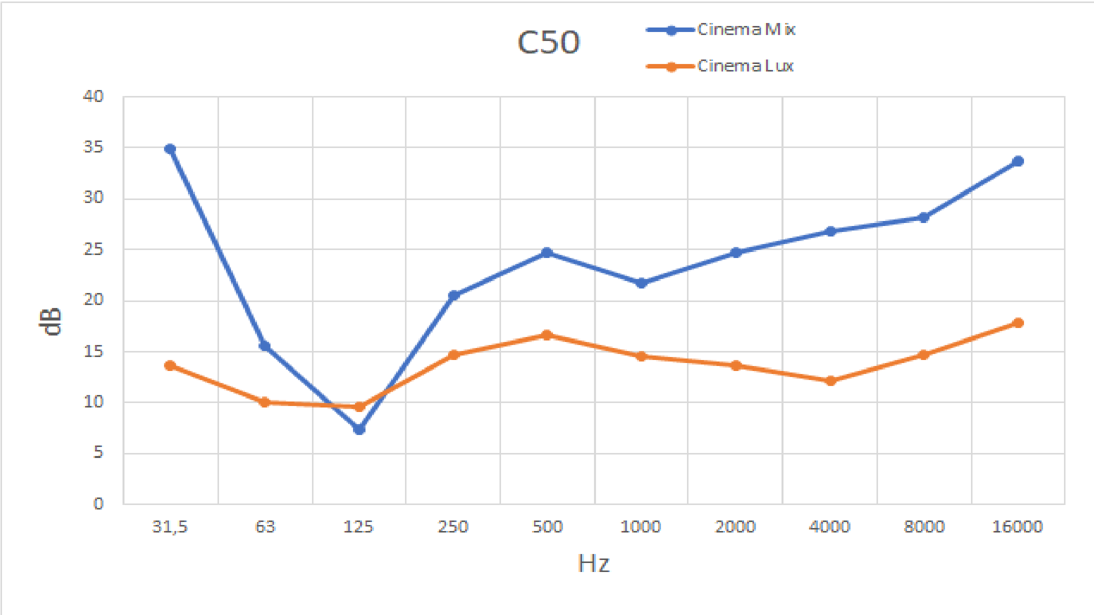

Figure 22: C50 plot

A high value of this parameter indicates a good speech/music intelligibility.
For both C50 and C80 we found a better clarity in the Dubbing stage (this result is expected being Cinema Lux a slightly more reverberant space).

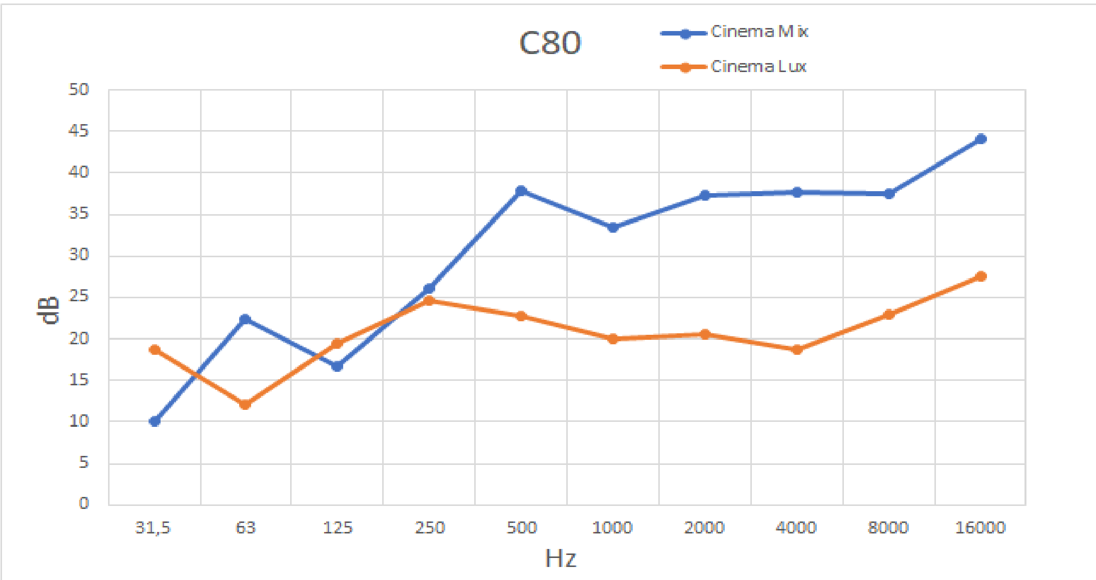

Figure 23: C80 plot

As shown in both the plots, Clarity grows towards high frequencies.

In C50 analysis of Dubbing Stage it must be noted a drop in the 125Hz octave so for these frequencies the intelligibility is bad.

The next parameter, generally less used than Clarity, is **Definition**.

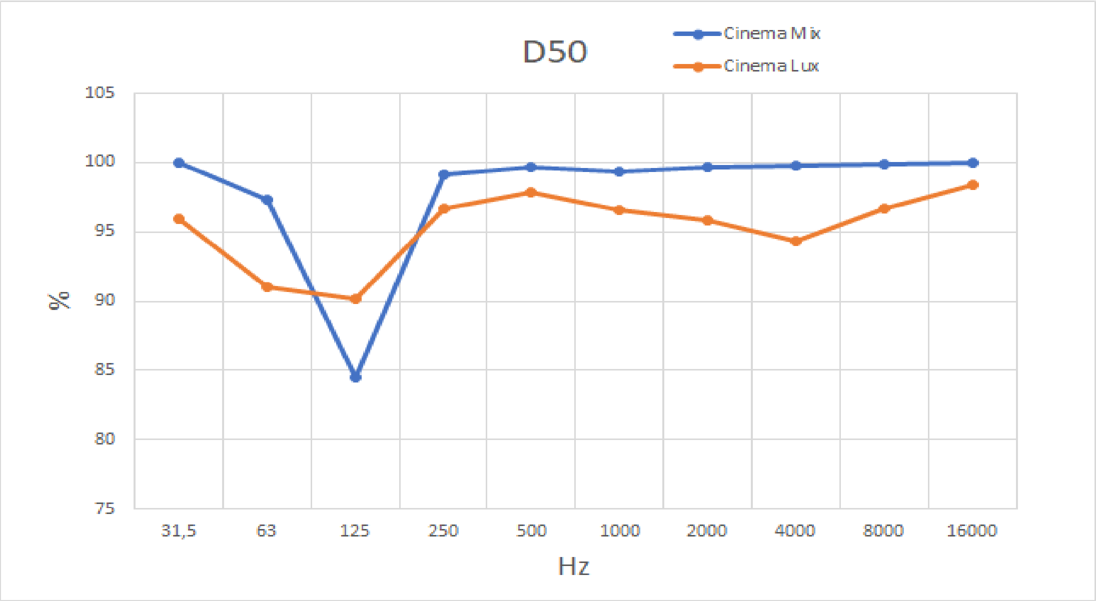

Figure 24: D50 plot

It is expressed as a ratio between the first 50 ms energy and the total energy.

For both the rooms we have a good definition (approaching unit) starting from the 250 Hz octave.

For the Dubbing stage we have a drop in the 125 Hz octave.

6.2.2 Spatial Parameters

To measure the difference in signals received by two ears of a person we use a Spatial Parameter called **Interaural Cross Correlation**.

IACC values range from -1 to +1.

A value of -1 means the signals are identical, but completely out of phase, a +1 means they are identical and 0 means they have no correlation at all.

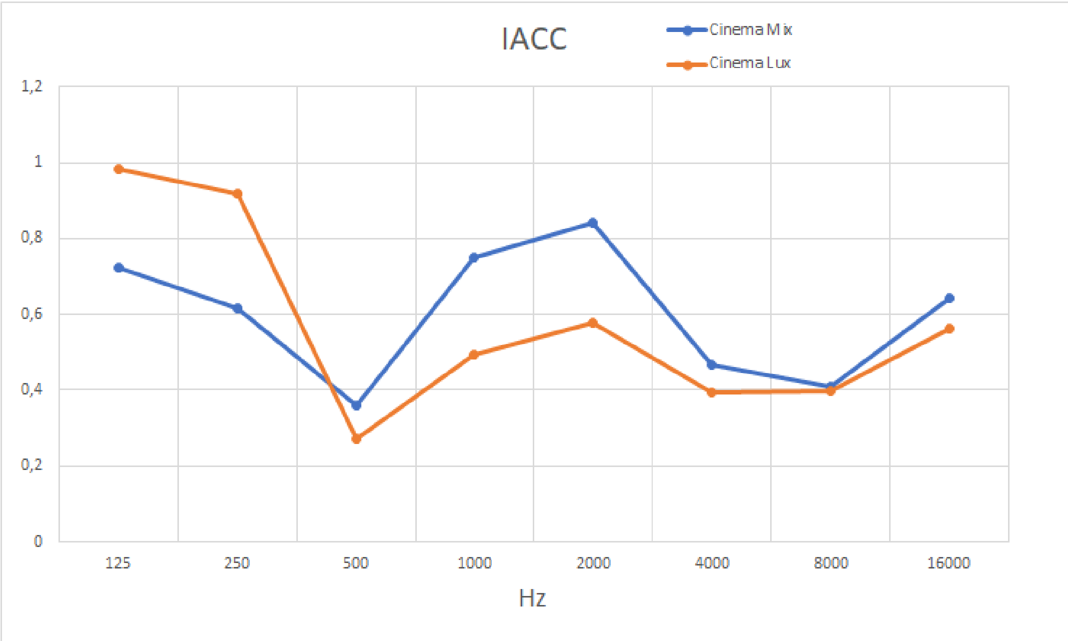

Figure 25: IACC plot

This parameter is extracted from the binaural recording (stereo) of the Neumann Dummy Head.

IACC exploits a cross-correlation operation between the two signals recorded at the Dummy ears and it reveals the *spatial degree* of the information.

For both Dubbing Stage and Cinema Lux we have, in the low frequencies, high values of this parameter.

This comes from the omnidirectional nature of the radiation at low frequencies and in fact, as we move towards higher frequencies, we have approximately decreasing correlation values.

Next spatial parameter we calculated is called **Lateral Fraction**.
LF is obtained from the Sennheiser Ambeo's omnidirectional (W channel) and figure of eight (Y channel) recordings.
This parameter calculates the ratio between the lateral sound energy and the total energy giving a measure of the *Apparent Source Width.*

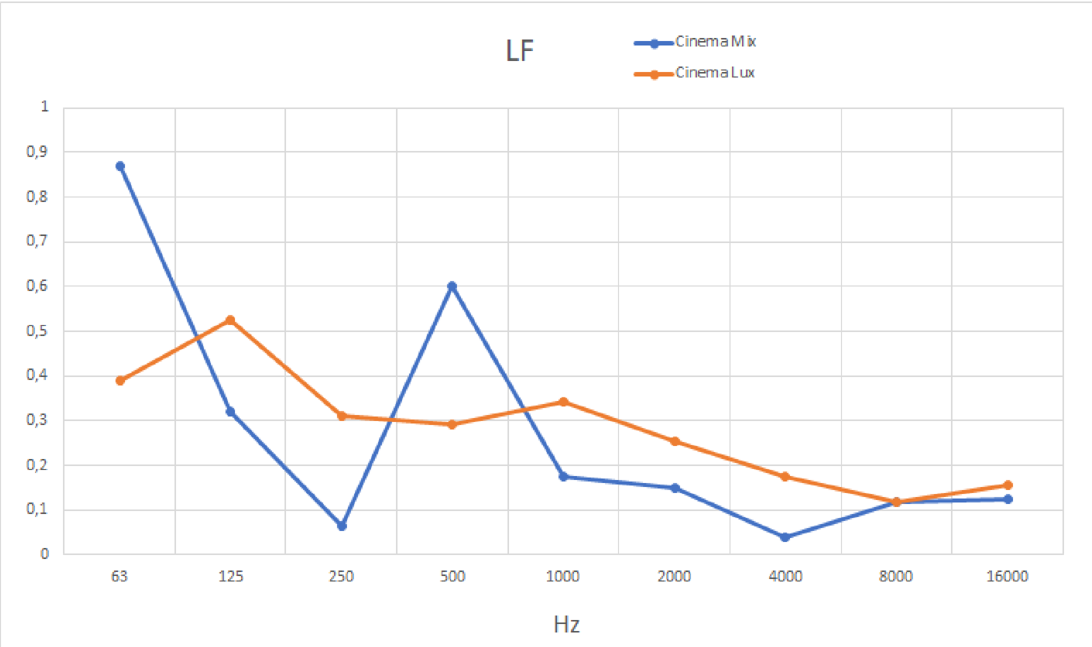

Figure 26: LF plot

In the Cinema Mix room, we can see two peaks in the graph of LF meaning that in those octaves (63 Hz and 500 Hz) the energy that comes from the lateral reflections is high.
In Cinema Lux room instead, we have a good balance between lateral energy and the total energy.

CHAPTER 7: SECOND MEASUREMENT SYSTEM

7.1.1 EIGENMIKE SYSTEM SETUP

Figure 1: Eigenmike probe

The second recording system we used is a spherical 32 Sennheiser capsules microphone array called Eigenmike™, produced by Mh Acoustics.

This system allows the recording of the Sound Field in the entire solid angle and for this reason it characterizes completely the spatial properties of the room.

The microphones, pre-amplifiers and A/D converters are packed inside the sphere with a radius of 42 mm.

The signals are delivered to the audio interface through a digital CAT-6 cable (Ethernet) employing the A-net protocol.

The audio interface is an EMIB Firewire interface: being based on the TCAT DICE II chip it works with any OS.

It is furnished with two analogue headphones outputs, one 8-channel ADAT digital output and the word clock ports for syncing with external hardware.

Figure 2: EMIB interface (front & back)

The audio interface is linked to the PC via Firewire to Thunderbolt cable adapter.

Figure 3: Eigenmike system setup

With the implementation of the linking modules available on the software "*Bidule*" by Plogue (Fig. 4) we were able to record (recording window in Fig. 5) and see real time all the 32 signals captured at the capsules.
By stopping the recording all the info is packed into a 32 channels wav file on the PC and is ready to be processed.

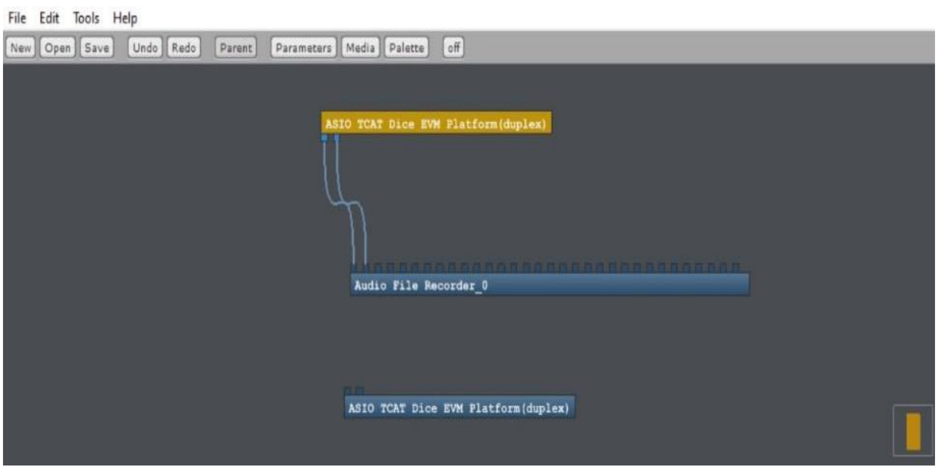

Figure 4: Plogue Bidule patch for Eigenmike recording

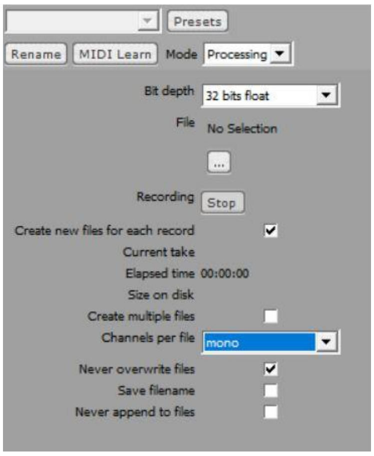

Figure 5: Recording interface

It must be said that the transducers (capsules) have a quite good frequency response from 30 Hz to 13 kHz, with a gentle roll off at higher frequencies.
With Eigenmike we record simultaneously multiple signals with 32 ultra directive virtual microphones, pointing in the same directions of the capsules.
This 32-channels recording is in RAW format also called A-format.
To process the signal in a proper way we need to transform it to P-format (SPS Signal) or B-format (HOA Signal).
For a brief explanation of what a SPS Signal is please refer to the Appendix D.

7.1.2 RECORDED SIGNALS

With Eigenmike we recorded two sound scenes: the ESS coming from the dodecahedron (recorded also with the previously described system) and a Dolby trailer in 5.1 configuration (creation of test signal is explained in next paragraph) reproduced by the entire electro-acoustic system as found in the rooms.
From the recording of the 5.1 Dolby Trailer we built:
a VR environment on the Oculus Quest to reproduce the feeling of immersive sound in the two rooms and a color plot of Sound distribution over a panoramic image.

From the ESS recordings we used instead a MATLAB script that allows to see the reflections of the rooms plotted on the panoramic image.

7.1.3 TRAILER DCP TEST SIGNAL

The Trailer was downloaded from the website "*https://thedigitaltheater.com/dolby-trailers"* and it was chosen as it matched the architecture of the reproduction systems of both the rooms under study (5.1).

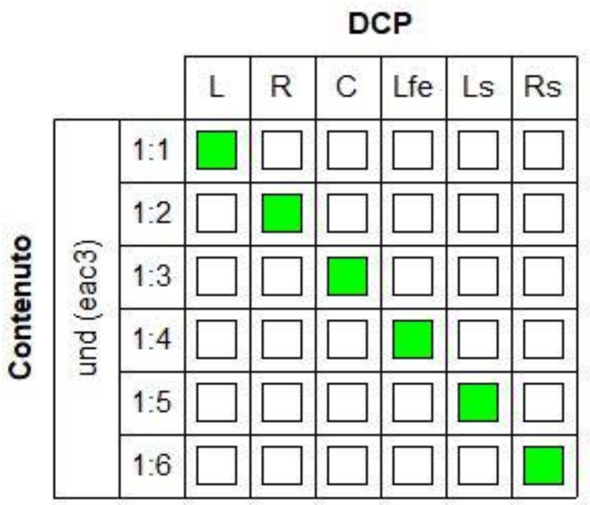

Figure 6: DCP-o-Matic audio format interface

This fact can be seen in the snapshot from "*DCP-o-Matic*", a user friendly software that I used to create the DCP file for the reproduction in Cinema Lux room (Dubbing stage doesn't need to create ad hoc DCP because the audio system is directly linked to the PC).

By simply importing the trailer in MP4 format in the file section and following the instructions the DCP file is created and its internal structure is the one described in DCP chapter.

If multichannel format audio is loaded it is possible to see also the time history of the dB levels of the different loudspeaker units that comprise the reproduction system.

As we can see in the following graph of the RMS values it is clear that this Trailer was meant to test the capabilities of the audio system under study (each unit is playing at different timing and different intensities) so it was perfect for our needs.

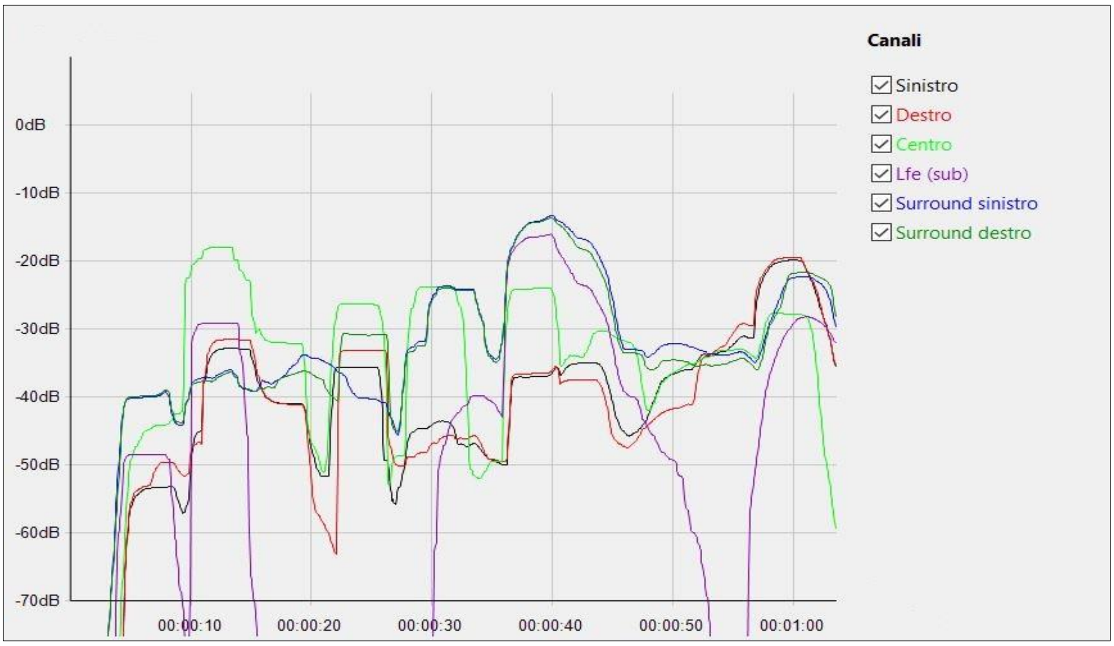

Figure 7: Loudspeakers levels in time

The sound in the DCP trailer was compressed in E-AC-3.

Dolby Digital Plus, also known as Enhanced AC-3 (and commonly abbreviated as DD+ or E-AC-3, or EC-3) is a digital audio compression scheme developed by Dolby Labs for transport and storage of multi-channel digital audio.

7.1.4 CAMERA RICOH THETA V

Figure 8: Ricoh Theta V panoramic camera

We shot a 360º video and a panoramic image of the two rooms at the position of the Eigenmike recording.

This was done mainly to see, in the virtual environment representing the inside of the room, from where the sound/reflection comes.

In other words, we want to localize a sound and match the direction of arrival of it as coming from a specific area of the room.

This virtual environment can be captured with the "*Ricoh Theta V 360*" camera which is capable of recording equirectangular panoramic video, either monoscopic (2:1 format) or stereoscopic top-bottom (1:1 format) as these are the required video format for the reproduction on a VR device.

In the next page it is shown an example of panoramic image, shot at cinema Lux.

Figure 9: Cinema Lux panoramic image

7.1.5 OCULUS QUEST & GOOGLE CARDBOARD

Figure 10: Oculus Quest & Google Cardboard

The playback of the panoramic video to obtain the virtual reality experience can happen by means of a device by Oculus called Quest.
Oculus Quest is a stand-alone system with a Qualcomm Snapdragon 835 processor.
In the inside we find two OLED displays each covered with its lens and the resolution for each eye is 1600x1440 pixel.
The movement tracking happens by means of four cameras installed on the front of the device (inside-out tracking).

Also Google cardboard implementation through “*Jump Inspector*” app installed on an Android device, can reproduce such videos.

Now that we listed all the devices involved for our objectives lets see how the RAW signals recorded with Eigenmike are processed.

The processing chains, whether we start from ESS or Trailer 5.1 recordings, are different and in the next paragraphs they are explained separately.

7.2 VR FOR 5.1 TRAILER

The 32 RAW Signals recorded and packed into a single multitrack wav file are opened with Audition CC 2019 to adjust the gain and the sampling rate (recording happened in 44100 Hz 32 bits and we needed instead 48000 Hz 24 bits).

The correct wav file is then saved and ready to be processed by another software called EigenStudio that comes along with Eigenmike.

This is the interface.

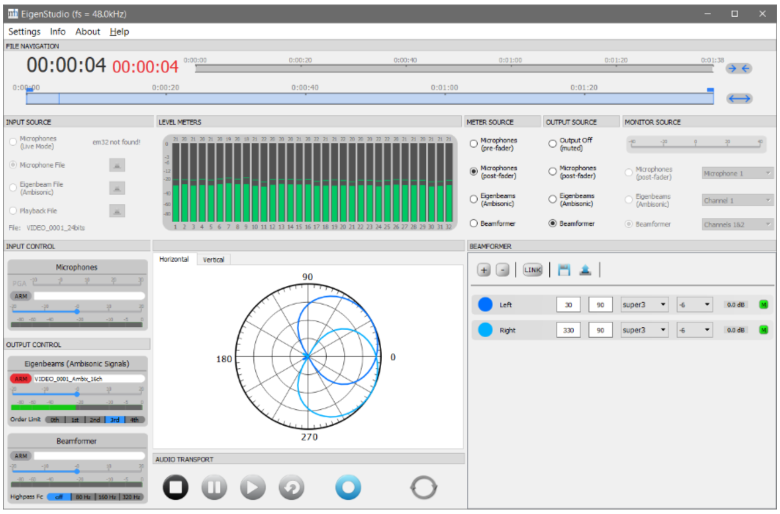

Figure 11: Eigenstudio Interface

It is capable of transform the input A-format recording into P-format and in our case we chose the settings to perform the conversion from the 32 RAW signals to the Ambisonics 3rd order format.

After the processing a new wav file with 16 channels in Ambisonics format is saved into a "pool" folder.

Now we need to edit the file using Audition CC 2019 in order to temporally line up it to the recorded video.

To do that both the Ambix 16 channels file and the audio from the Ricoh Theta V are opened in multitrack mode and the difference in samples between the two first big peaks that appear in the two types of recording is measured (1310710 samples in our case).

By cutting these number of samples from the beginning of the Ambix 16 channel file we obtain the alignment.

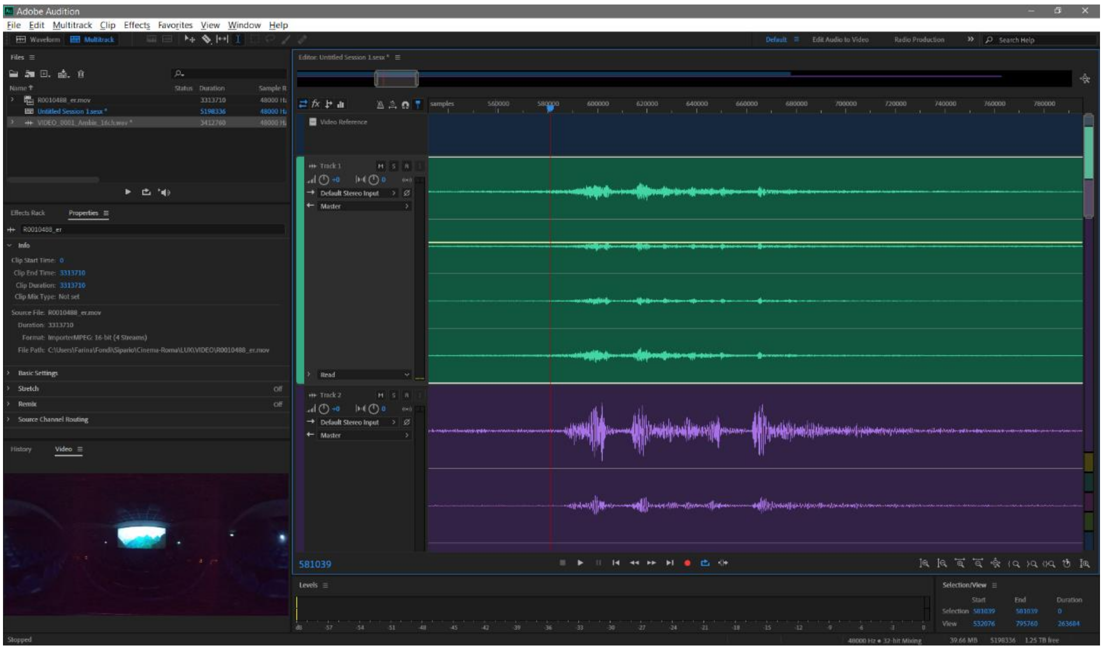

Figure 12: Tracks alignement Audition

The tails are also cut to obtain the same number of samples (3313710).

Now we save this Audition session into a new file containing the Ambix file aligned to the video.

The next step is to open another software called FB360 Encoder by Facebook and load the aligned Ambix 16 channels audio file and the video.

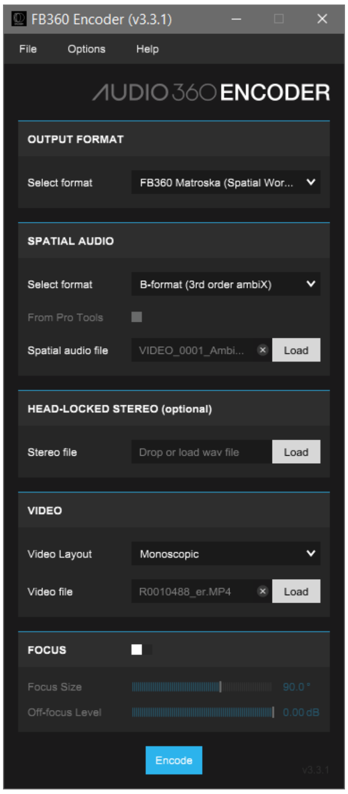

Figure 13: Audio360 Encoder

We select the input spatial audio as B-format (3rd order Ambix) and the output as FB360 Matroska which is the optimal format for the side loading through USB connection on the Oculus Quest.

This software operates the Encoding to this new format and save the file that is ready to be played back on the VR device.

The audio format of this video is TBE which is basically 2nd-order Ambisonics, where one of the Ambisonics 2nd-order channels has been removed.

To appreciate the perceptual differences between the rooms, the same procedure described above is performed for both the recording in Cinema Mixing room and Cinema Lux.

Unfortunately, Eigenmike recording of the Trailer in Cinema Mix room was corrupted at 26s (a defective cable was used) so we couldn’t process it any further.

In addiction, we decided to export such video files also in .MP4 format by selecting "Youtube Video" as Output format, to allow the playback on the PC, through VLC. The audio embedded with this type of format is FOA (first order ambisonics).

To see the perceivable correspondence between the original 5.1 Trailer and the recorded one, the same procedure for the creation of the VR video is performed with the original audio, extracted from the Dolby Trailer video file.
So, we now upload a different audio file: the Ambisonics first order transformation (4 channels) of the original 5.1 (6 channels) Dolby trailer audio track.
The conversion from 5.1 to FOA is performed again with Plogue "Bidule" throught the VST plugin Waves "B360 Ambisonics Encoder".
The following Figure shows the software interface while the conversion is happening.

Figure 8: 5.1 to FOA conversion interface

To clarify, in order to perceive the differences, two MP4 videos in Ambix first order for each room under study are created: One plays the audio from the original 5.1 Trailer and one plays the Eigenmike recording of the Trailer being reproduced by the speakers and coloured by the room.

7.3 TRAILER 5.1 SOUND ON PANORAMIC IMAGE

We start from the Eigenmike RAW recordings of the Dolby Trailer 5.1 reproduced by the electro-acoustic system that are packed into a wav 32 channel file.
The conversion from 32 RAW signals to 16 Ambisonics (third order) channels happens as described in the previous paragraph by means of Eigenstudio processing.
This transformation in 16 channels Ambix is needed as the VST used for the graphical visualization in color of the sound over the panoramic image needs this type of format and it's not able to process a higher number of channels.
This VST is contained in "*O3A Core Plugin v2.1.7"* which is a group of plugins created by Ripple Sound that provide a set of essential tools including panning, rotation, visualization and basic decoding for working with third order Ambisonics streams.
The VST we used, is the visualization tool called O3A Flare and it produces a view of an O3A stream that is shown using a rectangular screen region that can be loaded directly in the plugin window through a "load image" button.
This rectangular region shows an equal-area cylindrical projection of the directional components of the sound field interpreted over a sphere.
These directions are painted using color coding, ranging from red for low frequencies up to magenta for high frequencies (up to about 20kHz).
So after having loaded our 16 channels Ambix stream of audio on Audition CC 2019 we can simply recall the plugin from the effects curtain.
After having loaded our panoramic image we can play the multichannel audio and look at the evolving distribution of sound on the image.
Two control knobs can be adjusted to regulate the visualization: one is used to regulate the brightness of the loaded image and one is Raider Cutoff knob.
The Raider Cutoff tracks the overall signal level using a low-pass filter so for lower values we have a slower response to changes in the overall level of the signal.
It must be said that for the visualization of the Trailer 5.1 sound recorded in Cinema Mix room we used also the O3A Look plugin which allows to change where the "front" is in the 3D image.

The panoramic image shot in Cinema Mix room was in fact wrong and we had to flip it with the software "*Pano2VR*" by 180 degrees.
The same flipping operation had to be done also on sound by means of O3A Look's knobs: Azimuth and Elevation.
In the following pictures it is plot the distribution of sound being reproduced by the audio systems installed in the two room.
These are the snapshot took from "O3A Flare" Plugin interface of both the rooms at different point in the time evolution of the Trailer.
To see the differences, I chose to take a snapshot at the same timing for both the rooms. These are the plots of sound distribution at 10 sec.
At this point the majority of the sound is being radiated by the centre loudspeaker behind the screen and by the Low Frequency Effect unit (Sub) placed just under the screen.
The loacation of screen loudspeakers is almost the same for both the rooms.
To see the evolution of loudspeakers unit SPL levels, please refer to the previous section "TRAILER DCP TEST SIGNAL".

Figure 14: Cinema Mix sound distribution at 10 sec

Figure 15: Cinema Lux sound distribution at 10 sec

The first thing to say is that sound levels were higher in the Cinema Mix and this can be scoped in the higher colour tone around the screen of this room with respect to the Cinema Lux.

Some reflections are present in Cinema Mix as coming from the mixing deck and in Cinema Lux as coming from the seating area. The next snapshots are at 18 sec of the Trailer time evolution.

Figure 16: Cinema Mix sound distribution at 18 sec

Figure 17: Cinema Lux sound distribution at 18 sec

At this point the sound is majorly radiated by the left and right surround arrays with a difference in RMS SPL values of around 3 db.

As we can see in Cinema Lux the sound is more diffused than in Cinema Mix room.

7.4 SOUND REFLECTIONS FROM IMPULSE RESPONSE ON PANORAMIC IMAGE

Here we start instead from the ESS RAW recordings of the Eigenmike and we chose to use a MATLAB Suite (containing four scripts) kindly furnished by A. Farina and D. Pinardi who helped us in understanding its functions.

Please refer to Appendix F for more details on MATLAB Script.

With this method there is no need to convert the 32 channels format into 16 channels Ambix format so we don't lose spatial information and the identification of the source of the reflection is more precise.

One script allows, starting from the ESS RAW recordings, to extract the 32 IRs by means of a deconvolution process and to save them into a new multichannel wav file.

Another script makes the encoding of our IRs from A-format to P-format, so it convolves the 32 IRs in RAW format with a SPS 32x32 filtering matrix.

In this way the 32 capsules signals are transformed into 32 ultra-directive virtual microphone that point in all directions and divide uniformly the sphere.

For further explanation on how this conversion happens, look at the appendix E on the generation of the virtual microphones.

The last script we used is for the creation of an MP4 video, that shows the levels of sound reflections over the panoramic image.

No graphical algorithm is required, as a standard graphic library is employed for obtaining the colour map, based on the 32 "instantaneous" values of the sound pressure level captured by the 32 virtual microphones.

Many parameters can be changed here: we can cut our IRs, choose to visualize the panoramic image in grey scale, visualize the legend, adjust the bufferSize and the stepSize...

In particular, the bufferSize parameter controls the analysis window length so by reducing it we gain more resolution.

The stepSize parameter is instead the number of advancements sample whose reduction is used to smooth the colour map.

In the following figures we show some of the reflections of both the rooms under study.

CINEMA MIX ROOM

The first peak we find in the impulse response of cinema Mix room is the one depicted in the following image.

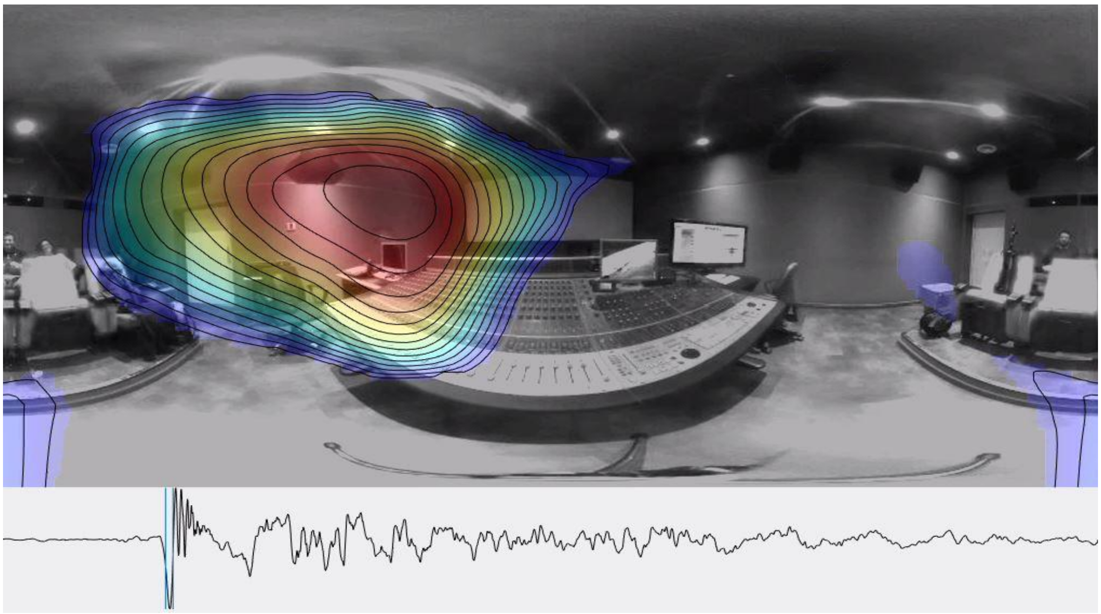

Figure 18: Strong reflection from the side

There is a strong reflection coming from the left wall upper area.
This could come from the plastic case of the surround loudspeaker or more likely from a wrong disposition of absorbing material behind the wall.
This is an issue because we find this reflection multiple times in the time evolution of the impulse response.

In Fig. 19 at the following page we record three reflections: one coming from the front screen, one from the floor and a stronger one from the glass that link the projection room to the dubbing stage.

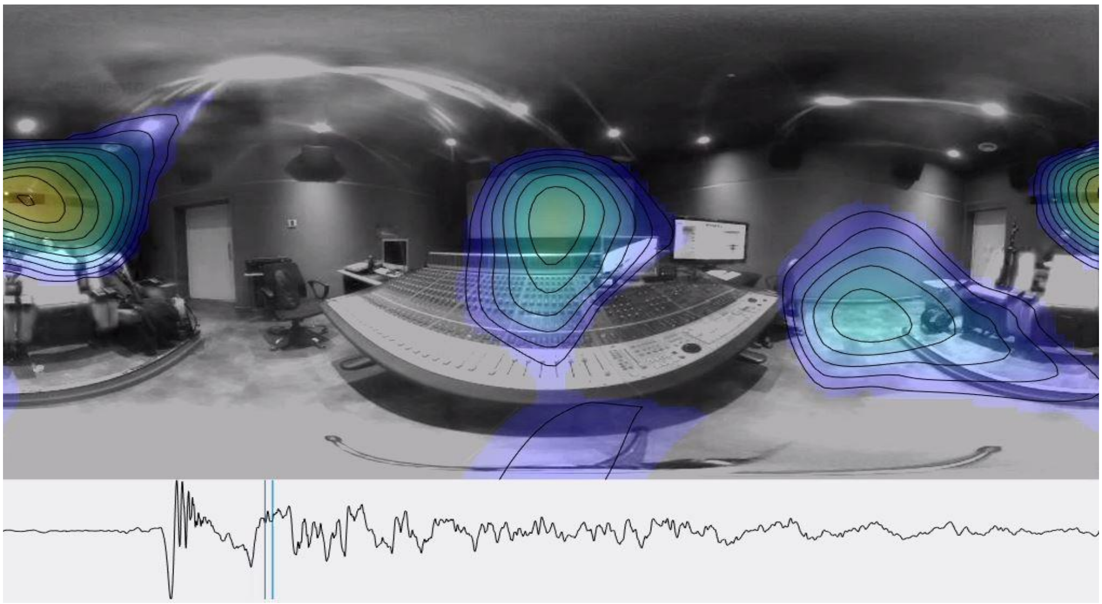

Figure 19: Reflection from projector glass, floor and screen

In Fig. 20 the reflection comes from the mixing table but it’s not too strong.

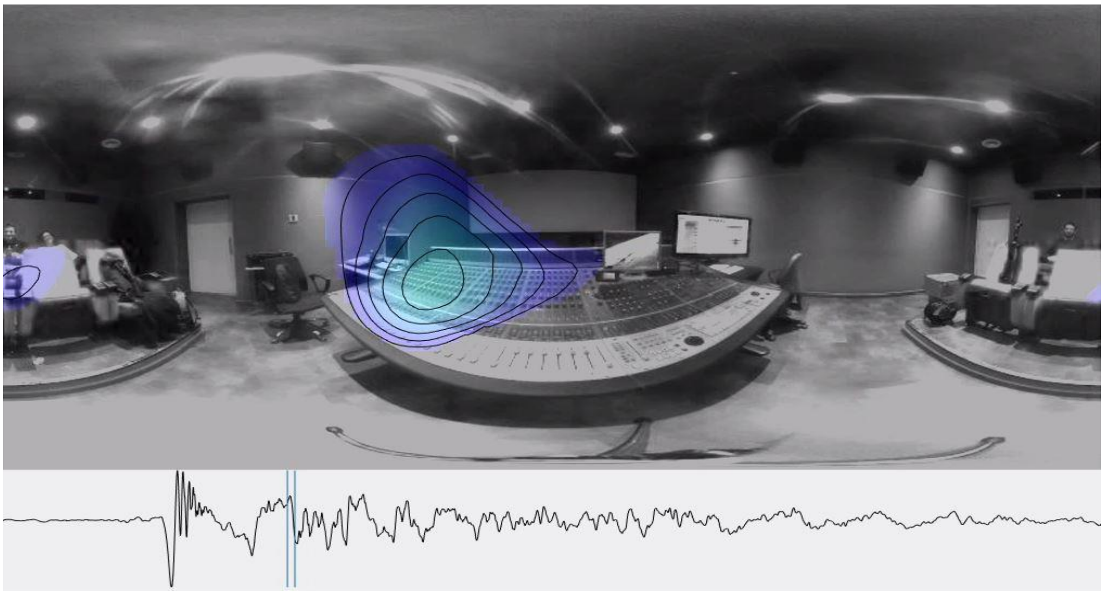

Figure 20: Reflection from the mixer

CINEMA LUX

The first peak here represents a reflection coming from screen area and from the seat area.

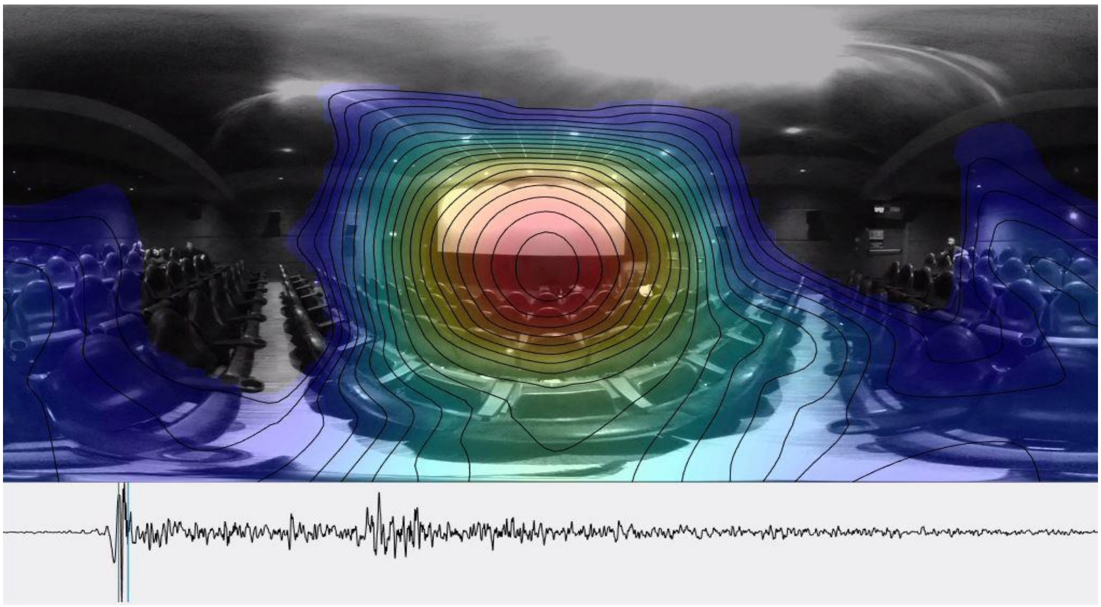

Figure 21: Reflection from screen

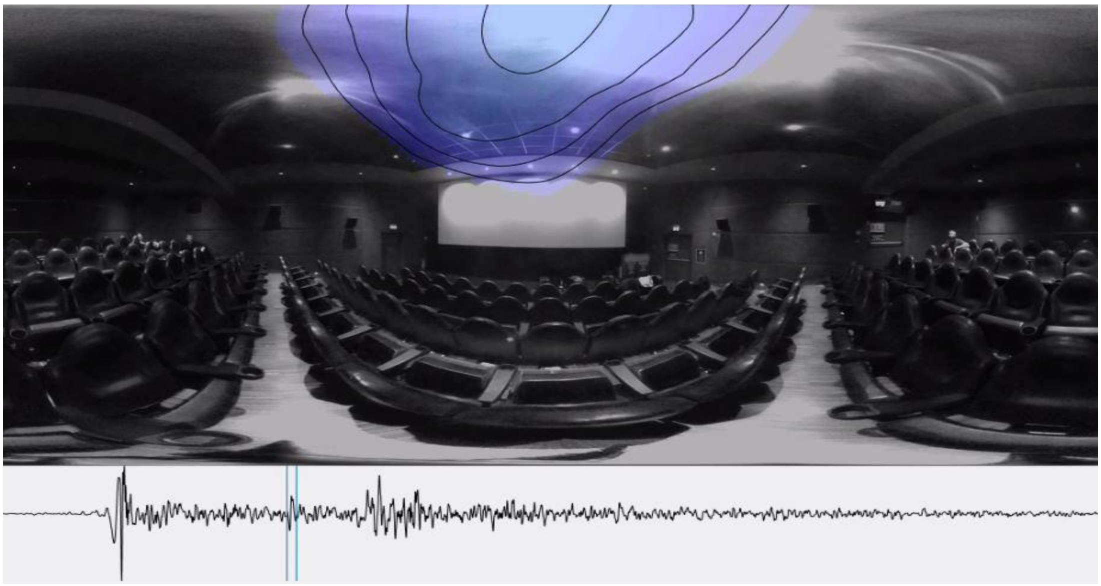

Figure 22: Reflection from the ceiling

In Fig. 22 we observe a weak reflection coming from the ceiling: the absorbing material used here does his job being this peak in the early part of the impulse response evolution in time (more risk of strong reflections).

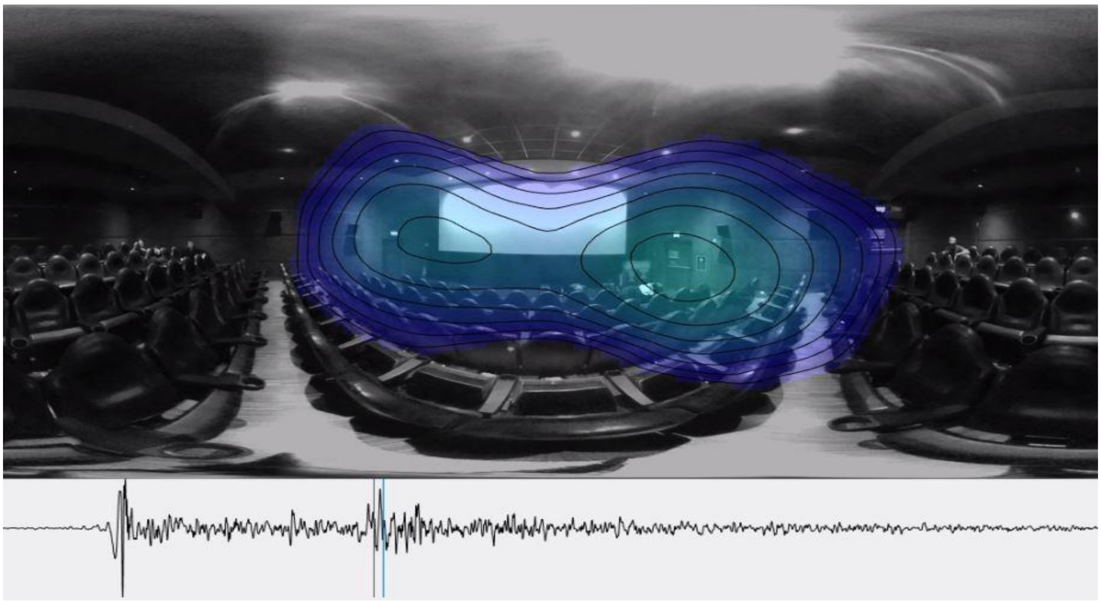

Figure 23: Reflection from right door

As we move forward in time, we notice a group of reflections and the highest peak here is shown in Fig. 23 as due to sound bouncing from the right door.
One idea to solve this reflection could be to put some absorbing material on the door.

In Cinema Mix room we have a higher risk of reflections due to the smaller volume (sound reaches reflective elements with more intensity) and to the presence of more reflective elements (glass of the projector screen, mixing table…).
In Cinema Lux the acoustic is well managed except from some reflections from the entrance doors.

CONCLUSIONS

To conclude, the idea that the perception of the soundtrack of a movie can be the same in both Dubbing room and in cinema room must be discarded.

Each room has a different acoustic fingerprint and colours the sound in its own peculiar way depending on the geometry, the furniture and the disposition of absorbing materials.

RTA measurement method for the calibration of sound system through equalization must be abandoned because it is based on wrong scientific assumptions, so it is unhelpful in improving the overall quality of the sound.

X-Curve also should be revised considering that nowadays speakers are capable to reproduce with good quality the entire human perceivable spectrum.

A good idea could be instead to measure the spatial distribution of sound to see which elements produce excessive reflections and act on them by putting absorbing or diffusive materials.

The implementation of array microphones (such as Eigenmike 32 capsules) for the recordings in 3D space allows the operator to identify precisely the source of disturbances in the room under study.

If the technician acoustically treats the room in a proper way, considering the positioning of the loudspeakers and other issues, there should be no need to equalize the signal (an operation that can ruin the perception in many cases) and the sound system should behave in an optimal way without the need to frequently re-calibrate the room.

Bibliography

[1] Rumsey, Francis, “Cinema Sound Reproduction”, JAES Volume 63 Issue 7/8 pp. 653-658; July 2015.

[2] SMPTE Professional Development Academy, “InterOP vs SMPTE DCP”, 7 August 2015.

[3] SMPTE Professional Development Academy, “What is Immersive Audio and why is it so cool?”, 9 August 2015.

[4] Murphy, David J, “Electro-acoustic measurements on cinema B chains in Australia”, http://www.aes.org/e-lib/browse.cfm?elib=17607.

[5] Tomlinson Holman, “Sound for Film and Television third edition (2010)”, Focal Press.

[6] A. Farina – “Simultaneous measurement of impulse response and distortion with a swept-sine technique”, 110th AES Convention, February 2000.

[7] A. Farina, R. Ayalon – “Recording concert hall acoustics for posterity” - 24th AES Conference on Multichannel Audio, Banff, Canada, 26-28 June 2003.

[8] SMPTE, “TC-25CSS B-Chain Frequency and Temporal Response Analysis of Theatres and Dubbing Stages”, October 2014

[9] SMPTE, “X-Curve: Dubbing Stages (Mixing Rooms, Screening Rooms and Indoor Theatres- B-Chain Electroacoustic Response)”, SMPTE ST 202:2010.

[10] A. Farina, S. Campanini, “A new Audacity feature: room objective acoustical parameters calculation module”

[11] ISO 3382:1997, “Acoustics - Measurement of the reverberation time of rooms with reference to other acoustical parameters”. International Organization for Standardization, Geneva, August 2001.

[12] UNI EN ISO 3382-2:2008

[13] S. Smyth, M. Smyth, S. Cheung, L. Kramer, “A Virtual Acoustic Film Dubbing Stage”, Smyth Research LLC, Camarillo, USA.

[14] Michael Gerzon, "Recording Concert Hall Acoustics for Posterity", JAES Vol. 23, Number 7 p. 569 (1975).

[15] A Farina. R Ayalon - “Recording Concert Hall Acoustics for posterity”. AES 24-th international Conference on Multichannel Audio. Banff Canada 26-28 June (2003).

[16] D. Malham. “3-D sound for virtual reality systems using Ambisonic techniques” (1993).

[17] A. Farina, A. Capra, L. Conti, P. Martignon, F. M. Fazi, Measuring spatial impulse responses in concert halls and opera houses employing a spherical microphone array, 19th International Congress on Acoustics (ICA), Madrid, 2007.

[18] A. Farina, M. Binelli, A. Capra, E. Armelloni, S. Campanini, A. Amendola – “Recording, Simulation and Reproduction of Spatial Soundfields by Spatial PCM Sampling (SPS)” - International Seminar on Virtual Acoustics, Valencia (Spain), 24-25 November 2011

[19] Angelo Farina, Andrea Capra, Lorenzo Chiesi, Leonardo Scopece - “A Spherical Microphone Array for Synthesizing Virtual Directive Microphones In Live Broadcasting And In Post Production” - 40th AES Conference "Spatial Audio - Sense the Sound of Space", Tokyo, Japan, 8-10 October 2010

[20] Frank, Zotter, Sontacchi, “Producing 3D Audio in Ambisonics”, Institute of Eletronic Music and Acoustics, University of Music and Performing Arts Graz, 8010 Graz, Austria

[21] A. Farina, A. Amendola, A. Capra, C. Varani, “Spatial analysis of room impulse responses captured with a 32-capsules microphone array”

[22] A. Farina, A. Amendola, L. Chiesi, A. Capra, S. Campanini, “Spatial PCM Sampling: a new method for sound recording and playback”

[23] Aurora43 manual, “www.aurora-plugins”.

[24] ST 2098-5:2018 - SMPTE Standard - D-Cinema Immersive Audio Channels and Sound field Groups

APPENDIX A: MAIN ASSOCIATIONS & IMPORTANT TERMS

Here are listed the main organizations that operate in Cinema industry some of which are referred to in the thesis:

DCI: Digital Cinema Initiatives is a consortium of the 6 major Hollywood studios: Disney, Fox, Paramount, Sony, Universal Studios, and Warner Bros.
Formed in 2002, DCI issued version 1.0 of its Digital Cinema System Specification (DCSS) in July 2005.

SMPTE: Society of Motion Picture and Television Engineers is the standards body where the majority of digital cinema standards work takes place.
Standards group activity is managed online and available to SMPTE standards committee members at the SMPTE website.
The SMPTE standards effort for digital cinema was initiated in January 2000 and continues to this day.

ISDCF: Intersociety Digital Cinema Forum meets about once a month to discuss technical and deployment issues for Theatrical Digital Cinema deployment.
They discuss DCP (Digital Cinema Package), KDM (Key Distribution Message), FLM (Facility List Management), TDL (Trusted Device List),
Formatting of distribution hard disc drives, DCI (Digital Cinema Initiatives), Upgrade scheduling, 3D luminance, subtitles, captions, closed captions, SMPTE specifications for digital cinema/audio and a large number of other TLA and FLA’s (three letter acronyms and four-letter acronyms).

NATO: National Association of Theatre Owners is a United States-based trade organization whose members are the owners of movie theatres.
Most of the worldwide major theatre chains' operators are members, as are many independent theatre operators
Collectively, they account for the operation of over 32000 motion picture screens in all 50 U.S. states and 81 other countries.

ASC: American Society of Cinematographers is a professional association of major directors of photography.

Here are defined some of the terms used in the thesis.

DUBBING STAGE OR CINEMA MIX ROOM

Applies to the motion picture industry.
When a film, as a visual, is finally edited the way the director wants it, and all the sound effects, dialog and music have been separately mixed (balanced) and finalized (“built”), they are all brought together to be merged at “the dub”, which generally occurs at a “dubbing stage.”
A dubbing stage is usually a studio facility that looks like a combination movie theatre and recording studio engineering room (projectors, screen, theatre seats, mixing board, etc.).
This is where the final decisions are executed with regard to the sonic elements of the film, and it occurs in a theatre-like environment to give the decision makers the best possible representation of how all the elements (audio and visual) are really working together in the context of how the film will be seen and heard.

A-CHAIN & B-CHAIN

The sound system of a movie theatre is divided into two parts: The A-chain and the B-chain.

You can think of this like a great big hi-fi system where the A-chain is the CD-player, radio tuner or vinyl turntable. The B-chain is the amplifier and loudspeakers.

In hi-fi, we wouldn't bother thinking about these as separate entities, but in cinema things are much larger in scale and there are conceptual benefits.

The A-chain in cinema consists of the sound recording on the film print, which will be available in Dolby analog, and a selection among Dolby Digital, DTS and SDDS digital formats.

Also, the equipment that retrieves the audio from the print and processes it so that is ready for amplification is part of the A-chain.

This is not all internal to the projector - audio systems can be retrofitted to older projectors, and the electronics will be in physically separate racks.

The B-chain consists of multi-channel volume control, equalizers, amplifiers, loudspeakers *and* the acoustics of the theatre itself.

The B-chain includes also the projection screen.

The reason for this is that the left, centre and right loudspeaker systems are *behind* the screen, therefore the screen has to be acoustically transparent. Naturally, the degree of transparency will vary with frequency, so the screen has to be included in the overall acoustic design.

The A-chain is something that can readily be changed (an improved projector could be installed.

If a manufacturer such as Dolby came up with a new and improved digital sound system, this could be retrofitted to the projector.

The B-chain is much more of a fixture - changing the acoustics of the auditorium would be a major task.

There is a high degree of standardization in both the A-chain and B-chain in movie theatres around the world.

This is so that a film soundtrack can be mixed with confidence that it will sound pretty much the same wherever the film is shown.
Artistic decisions are made in the dubbing theatre, the rest of the process all the way through to the ears of cinema goers is purely technical.

X-CURVE

X curve is a standardized roll of that evolved in the cinema business since the beginning of audio in the 1930.
At the beginning we had what was called the '*Academy curve*' and it was a roll off on both the high and low frequencies and was designed to act as a brute force noise reduction because at the extremes of the spectrum there was a lot of optical noise and in this way, it protected the loudspeakers.

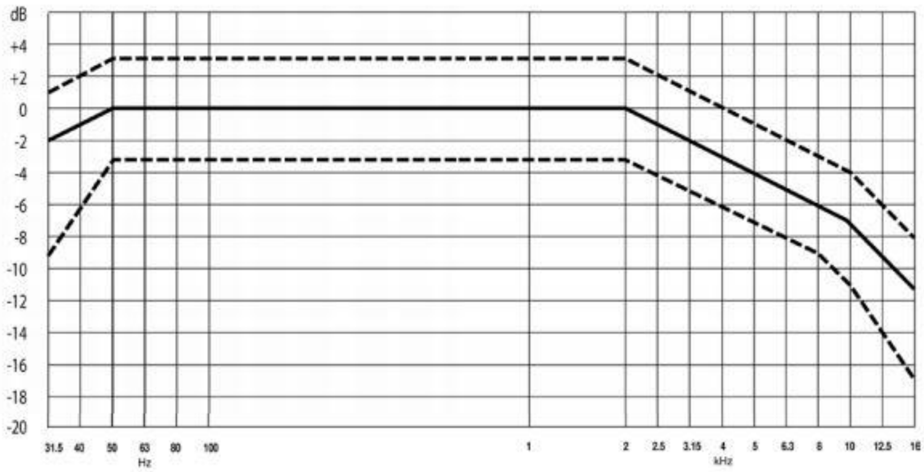

Figure 1: X-curve from SMPTE ST 202:2010

First Loudspeakers weren't able in fact to reproduce the full audio bandwidth, so they risked breaking without these imposed roll offs.
As we moved into Dolby Stereo and into the more modern soundtracks the curve was extended to the low frequencies but kept the high frequencies roll off.
The name that was given to this new standard curve was X-curve where X stands for "extended".

We have to consider that nowadays speakers are able to reproduce with good quality the full Spectral content and many professionals ask themselves what the real advantage is to keep the X-curve as a standard.

THX

THX is an American company founded in 1983 by George Lucas and it is headquartered in San Francisco, California.
It develops the “THX” high fidelity audio/visual reproduction standards for movie theatres, screening rooms, home theatres, computer speakers, gaming consoles and videogames.

APPENDIX B: ISO 3382 PARAMETERS

This ISO specifies methods for the measurements of reverberation time and other room acoustical parameters in performance spaces.
It describes the measurements procedure, the apparatus needed, the coverage required, and the method of evaluating the data and presenting the report.
It is intended for the application of modern digital measuring techniques and for the evaluation of room acoustical parameters derived from impulse responses.
The parameters are all based on measurements of the Impulse Response and are divided into two categories: Energetic Parameters and Spatial Parameters.

Energetic Parameters (from Omnidirectional measurements)

REVERBERATION TIME (T60)

The classical Reverberation Time (T60), is obtained by calculating the time interval needed for the sound level to decrease 60 dB.
Extrapolations are necessary and usually the T30 or the T20 (corresponding to a 30 dB decay or 20 dB, respectively) is measured, between -5 dB and -35 dB (or -25 dB), and then multiplied by 2 (or 3) in order to make it equivalent to the T60.

EARLY DECAY TIME (EDT)

The Early decay time is correlated to the early decay of reflections (early reverberation).

It is calculated as the time it takes for the sound to decay from 0 to -10 dB and then it is multiplied by 6, for it to be comparable with reverberation time (RT60).

CENTER TIME (T_S)

Is the time of the centre of gravity of the squared impulse response and it can be measured in seconds.
It avoids the division of the impulse response into early and late periods.

$$T_S = \frac{\int_0^{\infty} \tau \cdot p^2(\tau)\, d\tau}{\int_0^{\infty} p^2(\tau)\, d\tau} \quad [\mathrm{ms}]$$

CLARITY INDEX C

$$C_{t_e} = 10 \log\left(\frac{\int_0^{t_e} p^2(t)\, dt}{\int_{t_e}^{0} p^2(t)\, dt}\right) \quad [dB]$$

This is a logarithmic ratio between a fraction and the entire (or the remaining) IR energy and expresses a balance between early and late arriving energy which is useful to measure the clarity as perceived by human ears.
The t_e parameter changes if the room is built for human speech (t_e = 50 ms) or for musical purposes (t_e = 80 ms).
So, for example, Clarity index for music is C_{80} and it’s expressed in the following formula.

$$C_{80} = 10 \log\left(\frac{\int_0^{80ms} p^2(t)\, dt}{\int_{80ms}^{0} p^2(t)\, dt}\right) \quad [dB]$$

DEFINITION INDEX (D)

$$D_{50} = \frac{\int_0^{50ms} p^2(t)\ dt}{\int_0^{\infty} p^2(t)\ dt} \quad [\%]$$

The Definition index D is less used than Clarity index and, in the balance, it includes the entire energy from 0 to ∞.

D_{50} is defined as the percentage of the sound energy in the first 50ms after the arrival of direct sound with respect to the total sound energy.

SOUND STRENGTH (G)

$$G = 10 \log \left(\frac{\int_0^{\infty} p^2(t)\, dt}{\int_0^{\infty} {p_{10}}^2(t)\, dt} \right) \quad [dB]$$

It is a logarithmic ratio between the energy of the measured IR and a reference one.

It gives a measure of *how much* the environment increases (or decreases) the perceived loudness of a sound.

Where $p_{10}(t)$ is the instantaneous sound pressure of the impulse response measured at a distance of 10 m in a free field.

Spatial Parameters (from Binaural and WY measurements)

They give a listener surround capability measure of the room and a more complex recording equipment is needed.
Spatial parameters give a measure of the sound source virtual width or the enveloping effect.

INTERAURAL X-CORRELATION FUNCTION & COEFFICIENT (IACC & IACF)

Human spatial perception is due to the biological stereo human audio system and precisely to the difference between the signals that arrive to the two ears.
If there is no difference between left and right sounds, we are not able to locate a sound source in a scene.
With a binaural microphone (Dummy Head) it is possible to record exactly what arrives at two ears, and a cross-correlation operation between these two signals will reveal the spatial degree of the information: this is the definition of the Interaural Cross-Correlation Function (IACF), that can be expressed in the following mathematical form:

$$IACF_{\tau1,\tau2}\ (\tau) = \frac{\int_{\tau1}^{\tau2} p_L(t) \cdot p_R(\tau + t)\, d\tau}{\sqrt{\int_{\tau1}^{\tau2} p_L^2(\tau)\, dt \cdot \int_{\tau1}^{\tau2} p_R^2(t)\, dt}}$$

Generally, the extremes of integration are from 0 to 80ms:

$$IACF\ (\tau) = \frac{\int_{0}^{80ms} p_L(t) \cdot p_R(\tau + t)\, d\tau}{\sqrt{\int_{0}^{80ms} p_L^2(\tau)\, dt \cdot \int_{0}^{80ms} p_R^2(t)\, dt}}$$

IACF is a binaural measure of the difference in the sounds arriving at a listener's ears, produced by a source on stage.

Then Interaural Cross Correlation Coefficients are good indicators of the subjective quality *spatial impression* in a room and can be extracted with this formula:

$$IACC = \max(|IACF(\tau)|) \quad \text{for -1 ms} < \tau < + 1 \text{ ms}$$

LATERAL FRACTION (LF)

$$LF = \frac{\int_{5ms}^{80ms} p_8^2(\tau)\, \mathrm{d}\tau}{\int_0^{80ms} p_{omni}^2(\tau)\, \mathrm{d}\tau} \cong 1 - IACC$$

These measures need spatial information and this can be obtained with a recording system composed by an omnidirectional and a figure-of-eight pattern (with the null axis facing the source) microphones to differentiate lateral reflections from reflections that arrive to the listener from all directions:

In other words, it expresses the ratio between the lateral sound energy and the total energy that comes to the listener

It can be a measure of the *Apparent Source Width*.

Ideally it should be in the range of 0.2 - 0.25.

If $p_8(\tau)$ is the pressure coming from the sides, we are measuring the ratio of it with the total power.

It can be used also an approximation of LF called Lateral Fraction Cosine LFC:

$$LFC = \frac{\int_{5ms}^{80ms} |p_L(t) \cdot \mathrm{p(t)}|\, \mathrm{dt}}{\int_0^{80ms} p^2(t)\, \mathrm{d}t}$$

LATE LATERAL SOUND ENERGY (LG)

Surround effect can be measured with the Late Lateral Sound Energy:

$$L_{80,\infty}^{G} = 10\ log \frac{\int_{80ms}^{\infty} P_L^2(t)\, dt}{\int_0^{\infty} p_{10}^2(t)\, dt}$$

This parameter is related to the perceived listener envelopment or spaciousness in the auditorium.

APPENDIX C: PILLS OF AMBISONICS

Ambisonics is a 3D recording and playback method based on the reproduction of the sound field excitation as a decomposition into spherical harmonics.
This is the general formula that describes the main Ambisonics concept:

$$p(k,r,\vartheta,\varphi) = \sum_{n=0}^{\infty} \cdot \sum_{m=-n}^{n} 4\pi i^n j^n(kr)\, A_{mn} Y_n^m(\vartheta,\varphi)$$

Any pressure at location in spherical harmonics with a certain frequency k can be expressed as an infinite sum of these terms.
All Its needed to know to define the sound field are the coefficients A_{mn} and to obtain them we extract the spherical harmonics and spherical Bessel functions that are defined.
In other terms: to find these coefficients all we must do is to use a microphone that has the directivity pattern of that coefficient.

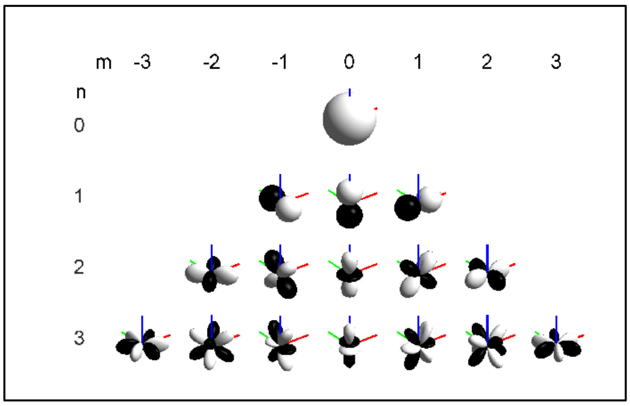

Figure 1: Spherical harmonic modes plotted in beam patterns.

So, for example if we need to extract the A_{00} coefficient we simply need an omnidirectional microphone (and this is 0 order Ambisonics).

For the first order Ambisonics we need at least 4 capsules in tetrahedral configuration and that’s the case for our Ambeo microphone.

If I want instead a microphone capable of recording second order Ambisonics I need a configuration of capsules that can create 9 patterns so the minimum number of capsules is 9.

APPENDIX D: PILLS OF SPATIAL SAMPLING

SPS stands for Spatial PCM Sampling and is considered as an alternative to Ambisonic recording.
In this Appendix it is presented the definition of the SPS concept and are shown the main differences with Ambisonics.
The SPS recording is made with a bunch of signals coming from coincident directive microphones pointing in the entire 3D space.
The fact that all the microphones are coincident means the SPS signals do not contain time differences between the channels and the only thing that differs is the amplitude which depends on the position of the Source.
And this is just like Ambisonics because both the techniques encode the spatial information based only on amplitude and not the phase.
However, the main difference of SPS is the absence of reverse polarity signals because the employed microphones here are cardioid of various orders and do not exhibit any rear lobe.
For up to 20 channels we can see the geometries of the capsule distribution choosing between regular polyhedron (tetrahedron, dodecahedron, icosahedron).
From 20 channels and above it is necessary to use “not-exactly-uniform” geometries such as the truncated icosahedron that describes the Eigenmike capsules distribution.

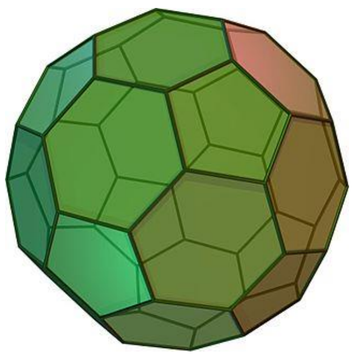

Figure 1: Truncated Icosahedron

For the Eigenmike there is a standard absolute orientation and a standard channel ordering.
PCM sampling is the representation of an analog signal by means of pulses with the form of delta functions.

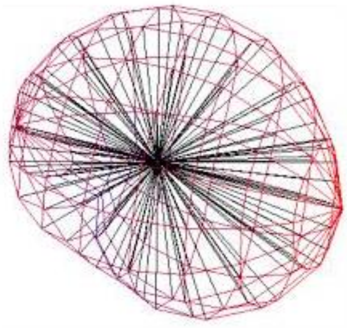

Figure 2: Spatial delta functions

Ambisonics is linked instead to the Fourier representation of this PCM signal.
In 3D space, PCM approximates the sound coming from every direction with spherical distribution of spatial pulses.
It can be said that both SPS and Ambisonics are intermediate formats capable of decomposing a complete three-dimensional sound field in discrete components.
These are just Dirac's Delta pulses for Spatial PCM Sampling, and instead are complex oscillating functions over the surface of a sphere for Ambisonics.
One difference between the two recording formats can be that "spatial equalization", which means boosting the gain in some directions and reducing the gain in other directions, is trivial with SPS signals and very tricky with Ambisonics.
Moreover, for SPS signals with large channel counts (32 and above) there is no need to "pan" across the channels, you can send each sound source to just one channel, with the exception of the case when you want to "spread" spatially the sound over more directions.

APPENDIX E: EIGENMIKE VIRTUAL MICS

It must be said that the signal that had been recorded with Eigenmike were in the so-called A-Format which is a RAW format so it must be converted.
One method for the conversion to P-format (SPS) happens by means of a convolution of the RAW signals with a matrix of FIR SPS filters.
This matrix is obtained by the creation of an arbitrary number of virtual microphones characterized with a super directive directivity pattern.

A virtual microphone signal y can be obtained as the filtered sum of M real microphone signal x, starting from a spatial sampling of the sound field performed employing an array of M microphones at different locations and aiming.

$$y_v(t) = \sum_{m=1}^{M} x_m(t) * h_{m,v}(t)$$

To obtain the filtering coefficients we impose that the measured polar pattern deviates minimally from the ideal one.
For this purpose, the Eigenmike is subject to a large number of anechoic impulse response measurements from many directions covering the whole surface.
For any direction D at any frequency the virtual microphone should provide a nominal target gain p_d and this is expressed in the following formula.

$$\sum_{m=1}^{M} c_{m,d} * h_m \Rightarrow p_d \qquad\qquad d = 1 \dots D$$

$c_{m,d}$ is the impulse response for microphone m and direction d.

p_d is obtained applying a direction-dependent gain Q_d to a delayed unit-amplitude Dirac's delta function δ.

$$p_d = Q_d \cdot \delta$$

And Q_d is defined as the directivity factor of a virtual microphone in spherical coordinates.
We know Q_d= $[0.5 + 0.5 \cdot \cos(\varphi)]^4$ for a fourth order cardioid.
φ is known from Heavyside formula starting from known azimuth and elevation of each virtual microphone.
The operation of the extraction of the filtering coefficients is then performed in frequency domain applying the Kirkeby algorithm in this way.

Using Kirkeby we extract our $[H_k]_{MxV}$.

$$[H_k]_{MxV} = \frac{[C_k]^*{}_{MxD} \cdot [Q]_{DxV} \cdot e^{-j\pi k}}{[C_k]^*{}_{MxD} \cdot [C_k]_{MxD} + \beta_k \cdot [I]_{MxM}}$$

In our case M = V = 32.
We convolve this matrix with our RAW recordings to obtain the SPS Signal that can be processed easily with MATLAB.
In the end we sample with SPS our RAW recording and we obtain a new SPS format (P-Format).

APPENDIX F: MATLAB SCRIPT

Here are reported the key parts of the scripts we used for the visualization of SPL levels of sound (Impulse Response) over the panoramic image.

This part performs the multichannel de-convolution and outputs the resulting 32 impulse responses.

```
%% DECONVOLUTION
% Recorded sweep are convolved by inverse sweep giving IRs
inverseSweepLength          = length( inverseSweep );
recordedSweepLength         = length( recordedSweep );
irLength                    = recordedSweepLength + inverseSweepLength - 1;
irs                         = zeros( irLength, numberChannels ); % pre-allocate matrix for IRs

% Initialize string to show completed convolution percentage
fprintf( '\nDeconvolving audio: ' )
percentageString            = '0%%';
percentageStringLenght      = length( percentageString );
fprintf( '%s', percentageString )
oldCompletedPercentage      = -1;

for channelIndex = 1 : numberChannels

    % Updating string to show completed convolution percentage
    CompletedPercentage     = round( channelIndex / numberChannels * 100 );
    if oldCompletedPercentage ~= CompletedPercentage
        for i = 1:percentageStringLenght
            fprintf( '\b' )
        end
        percentageString    = sprintf( '%d%%', CompletedPercentage );
        percentageStringLenght  = length( percentageString );
        fprintf( '%s', percentageString )
    end
    oldCompletedPercentage  = CompletedPercentage;

    % perform deconvolution
    irs( :, channelIndex )  = fd_conv( recordedSweep( :, channelIndex ), inverseSweep );

end

fprintf( '\n' )
```

This part performs the decoding from the A-format to P-format (RAW to SPS). The operation of convolution with SPS matrix is performed at line 12 through the command “*oa_multichannel_conv*”.

```
1.  %% PROCESSING
2.  limitLength                 = 200000; % depending on RAM of computer, might be necessary to decrease
3.  fileToProcessLength         = length( fileToProcess );
4.
5.  if  fileToProcessLength    <= limitLength
6.      % Reshaping matrix to use convolve library function: matrix_conv
7.      fileToProcessMatrix     = reshape( fileToProcess', 1, arraySize, fileToProcessLength );
8.      encodedFormat           = matrix_conv( fileToProcessMatrix, encodingMatrix );   % convolving with SPS matrix
9.  else
10.     % Reshaping matrix to use convolve library function: oa_multichannel_conv
11.     fileToProcessMatrix     = reshape( fileToProcess', arraySize, fileToProcessLength, 1 );
12.     encodedFormat           = oa_multichannel_conv( fileToProcessMatrix, encodingMatrix, precision );% convolving with SPS matrix
13. end
14. encodedFormat               = squeeze( encodedFormat )'; % eliminate exceeding matrix dimension
15.
16. figure
17. subplot( 3, 1, 1 )
18. plot( fileToProcess( :, 1 ) )
19. xlim( [ 0 size( fileToProcess, 1 ) ] )
20. xlabel( 'Samples' )
21. grid on
22. title( 'Original audio - first mic' )
23. subplot( 3, 1, 2 )
24. plot( encodedFormat( :, 1 ) )
25. xlim( [ 0 size( encodedFormat, 1 ) ] )
26. xlabel( 'Samples' )
27. grid on
28. title( 'Encoded audio - first mic' )
29.
30. for in = 1 : arraySize
31.     for out = 1 : virtualMicNumber
32.         [ ~, delayInOut ]   = max( abs( encodingMatrix( in, out, : ) ) );
33.         delay( in, out )    = delayInOut;
34.     end
35. end
36. delay                       = round( mean( delay, 'all' ) );
37. encodedFormat               = encodedFormat( delay+1 : delay + length( fileToProcess ), : );
38.
39. subplot( 3, 1, 3 )
40. plot( encodedFormat( :, 1 ) )
41. xlim( [ 0 size( encodedFormat, 1 ) ] )
42. xlabel( 'Samples' )
43. grid on
```

The last script we used is for the creation of the MP4 video showing the time evolution of SPL levels of sound distributed over the panoramic image.

I reported the key parts of this script.

Here we can choose the number of samples for the BufferSize and the StepSize, which are crucial for a meaningful visualization of the reflections.

```
% Video analysis
timeSignalEnable        = 1;
% plot time signal under video maps. Disable for VR
bufferSize              = 16;
% [samples]
stepSize                = 2;
% [samples]
frameRate               = 30;
 % Default 30
```

Here the data is processed.

The SPL values are interpolated to cover the entire panoramic image and this operation is performed for each frame.

```
for blockIndex = 1 : numbersFrame

    %%%%%%%%%%% PERCENTAGE PROCESSING %%%%%%%%%%%
    completedPercentage      = round( blockIndex/numbersFrame*100 );
    if oldCompletedPercentage ~= completedPercentage
        for i = 1 : varStrLen
            fprintf( '\b' )
        end
        varStr                = sprintf( '%d%%', completedPercentage );
        varStrLen             = length( varStr );
        fprintf( '%s', varStr )
    end
    oldCompletedPercentage   = completedPercentage;

    % perform FFT on a chunk and compute RMS with complex sum
    frfChunk         = fft( spsIrs( ( blockIndex-1 ) * stepSize + 1 : ...
                           ( blockIndex-
1 ) * stepSize + bufferSize, : ), nfft, 1 ) / fftNormFactor;
    frf              = 2 * frfChunk( 1 : nfft/2 + 1, : ) .* freqWindowing;

    complexSum       = real( frf ).^2 + imag( frf ).^2;
    rmsChunk         = ( sum( complexSum, 1 ) )' ./ 2;
    splChunk         = ( 10 * log10( rmsChunk ) + fullScaleLevel )';

```

```
    if virtualMicNumber == 122 || ( virtualMicNumber == 32 && sps32type == 1 )
        [ splChunk ] = extendSPLdir( splChunk ); % replicate SPL value on poles
    end

    % interpolation
    [ ZI ] = splInterpolation( splChunk, virtualMicNumber, extX, extY, extXI, extYI, surfpoint, sps32type );

    splMatrix( :, :, blockIndex )   = ZI;

end
```

The last part of the script maps (for each video frame) the SPL level-dependent-colours over the image.

```
for blockIndex = 1 : numbersFrame

    %%%%%%%%%%% PERCENTAGE PROCESSING %%%%%%%%%%%
    completedPercentage     = round( blockIndex/numbersFrame*100 );
    if oldCompletedPercentage ~= completedPercentage
        for i = 1 : varStrLen
            fprintf( '\b' )
        end
        varStr              = sprintf( '%d%%', completedPercentage );
        varStrLen           = length( varStr );
        fprintf( '%s', varStr )
    end
    oldCompletedPercentage  = completedPercentage;

    ZI          = splMatrix( :, :, blockIndex );                        % extract current value to plot

    % subplot colormap
    plotBG      = subplot( 'Position', [ 0 0.2 1 0.8 ] );
    videoColorMap( background, backgroundWidth, backgroundHeight, XI, YI, ZI, lowerLimit, upperLimit, plotBG , ...
                   v, alpha, contourEnable, useFixedSplPlotBounds, fixedLowerLimit, fixedUpperLimit, valueTickEnable, ...
                   cmap, tickSize, tickBold, colorbarEnable );

    % subplot time signal
    plotTime    = subplot( 'Position', [ 0 0 1 0.2 ] );
    timeSignalPlot( spsMean, verticalLimit, blockIndex, stepSize, bufferSize );

    writeVideo( videoFile, getframe( videoWindow ) );

end
```