

PAPER • OPEN ACCESS

Cinema Sound: Characteristics and 3D Acoustic Measurements

To cite this article: L Tronchin and N Scaroni 2020 *J. Phys.: Conf. Ser.* **1655** 012079

View the [article online](#) for updates and enhancements.



IOP | ebooks™

Bringing together innovative digital publishing with leading authors from the global scientific community.

Start exploring the collection—download the first chapter of every title for free.

Cinema Sound: Characteristics and 3D Acoustic Measurements

L Tronchin¹ and N Scaroni²

¹DA-CIARM, University of Bologna, Via dell'Università 50, 47521 Cesena, Italy

²Politecnico of Milan, Via Sesto 39, 26100 Cremona, Italy

lamberto.tronchin@unibo.it

Abstract. In 2014, an SMPTE report stated that sound consistency between dubbing stages and performance venues was not respected. SMPTE technical commission studied this issue in 3 commercial cinemas and 3 dubbing stages in Hollywood using the impulse response method. This lack in sound consistency is probably imputable to the use of a wrong calibration system (RTA + Pink Noise Method) that is time-blind and does not mimic the perceptual listening of the human being. A campaign of measurements has been conducted in Rome to test the acoustic properties of a dubbing stage room and a cinema performance room. Two measurement systems were implemented: One for the extraction of ISO 3382 parameters (Omni, Binaural, Soundfield) and one for studying the spatial distribution of sound in the room (Eigenmike™).

1. Introduction

The evolution of technology in the last 20 years brought big improvements in the entertainment world, raising the level of video and audio, making quality to be affordable by the most.

In many commercial venues, however, after the advent of digital cinema, image standards upgraded a big deal, leaving sound status a step behind.

Cinema owners must face the advent of new entertainment platforms such as Netflix and Amazon prime video that keep people at home.

Improving audio quality with more immersive systems that are already available on the market can be the way.

Also, they should abandon or at least revise the old calibration methods such as RTA Analysis through equalization, being this technique lacking scientific support and if not used properly can ruin sound perception. In particular, the use of a system that can simultaneously record sound in every direction, such as Eigenmike™, can be a precise tool for visualize the source of disturbance in a performance room.

With the knowledge of how the sound is spatially distributed in the room, one can act on the elements that bother the perception thus improving the acoustic quality of the space.

2. Cinema Standards

Traditional Cinema Sound fields are generated by a group of loudspeakers organized as a base layer, so they give spatial perception only in the horizontal plane (5.1, 7.1 etc...).

Immersive Sound Systems add the third dimension of height above the listener with the implementation of an additional group of loudspeakers positioned above the base layer.

Here are listed and briefly described some of the state-of-art Immersive systems.



Content from this work may be used under the terms of the [Creative Commons Attribution 3.0 licence](https://creativecommons.org/licenses/by/3.0/). Any further distribution of this work must maintain attribution to the author(s) and the title of the work, journal citation and DOI.

In Dolby Atmos we have 5 screen loudspeakers in the horizontal plane, two rows of ceiling loudspeakers and a base layer of surround loudspeakers. In Auro-3D system we have two vertical layers of screen loudspeaker, two layers of surround loudspeakers and two layers of ceiling loudspeakers. IOSONO system is 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.

NHK 22.2 is a system whose 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” so, to playback the sound as intended, one has to use a Renderer that uses technologies such as WFS, HOA or VBAP.

As digital cinema replaced the old distribution formats (35 mm or 70 mm) the audio moved from being printed on the strip to being carried as MXF Track Files in the DCP packet.

The DCP standard allows for up to 16 channels for audio and the currently supported are the 5.1 and 7.1 audio formats and SMPTE is working towards a common distribution format for Immersive Systems. 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.

3. HOA & SPS

HOA (High Order Ambisonics) and SPS (Spatial PCM Sampling) are two methods for recording and playing back sound in the 3D space.

Ambisonics is based on the reproduction of the sound field excitation as a decomposition into Spherical harmonics so, any pressure at a location in spherical coordinates with a certain frequency k can be expressed as an infinite sum of these terms.

$$p(k, r, \vartheta, \varphi) = \sum_{n=0}^{\infty} \sum_{m=-n}^n 4\pi i^n j^n(kr) A_{mn} Y_n^m(\vartheta, \varphi) \quad (1)$$

Where j 's are the Spherical Bessel function and Y 's are the Spherical Harmonics.

With the extraction of the weighting coefficients A_{mn} we uniquely define the Soundfield.

To find these coefficients all it's needed is the implementation of a microphone that has the directivity pattern of the specific coefficient.

So, as an example, a Soundfield microphone (that records in w, x, y, z) can extract these coefficients for First Order Ambisonics format ($m=1, n=1$), by combining the directivity patterns of its four capsules.

SPS (P-Format) is instead derived from the PCM RAW recording (A-Format).

PCM signal is convolved with an SPS filtering matrix so, the capsules signals are transformed into an arbitrary number of ultra-directive virtual microphones that point in every direction and divide uniformly the sphere.

This SPS encoding matrix has dimension $M \times V$, being M the number of real microphones and V the number of virtual microphones.

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 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.

$$\sum_{m=1}^M c_{m,d} * h_m \Rightarrow p_d \quad d = 1 \dots D \quad (2)$$

$c_{m,d}$ 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 .

$$p_d = Q_d \cdot \delta \quad (3)$$

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 the 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 the frequency domain applying the Kirkeby algorithm in this way.

$$[H_k]_{M \times V} = \frac{[C_k]^*_{M \times D} \cdot [Q]_{D \times V} \cdot e^{-j\pi k}}{[C_k]^*_{M \times D} \cdot [C_k]_{M \times D} + \beta_k \cdot [I]_{M \times M}} \quad (4)$$

In our case $M=V=32$.

4. Case Study: Cinema Lux & Dubbing stage Cinema

Measurements are performed in two rooms representing the starting point (dubbing stage room) and the destination (cinema performance room) of the soundtrack of a commercial movie.

The dubbing stage room (Left) is a simple rectangular room of 5,75 mt x 7,55 mt x 3,15 mt.

A big mixing-deck, two pc screens, and two rows of 4 padded chairs in the back are the main elements present in the room.

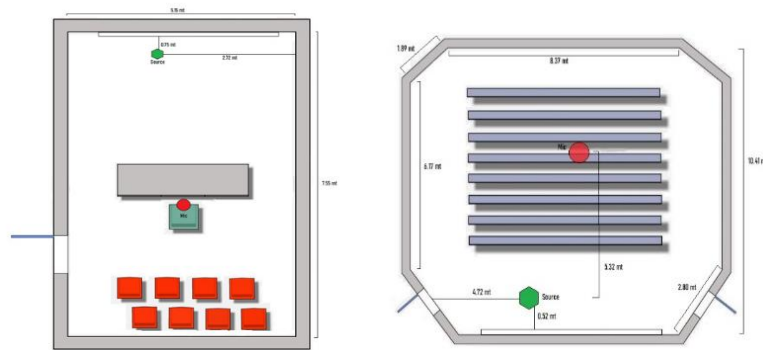


Figure 1: Dubbing Stage Room & cinema Lux performance room measurement layout

The projector is located behind the back wall and it points, through the glass, towards the screen. The audio system installed is a 5.1 JBL so, three screen channels (L, R, C), two surround channels arrays (Ls, Rs) attached on left, right and back wall and an LFE unit set asymmetrically under the screen.

The Source for the Sweep test is set asymmetrically at 2,72 mt from the right wall and the recordings are done at the re-recording mixer's location.

The second space of study is the performance cinema room n.10 located in multiplex Cinema Lux (Right). The floor is covered in wood and the walls, the ceiling and the armchairs are padded with absorbing material.

The audio system installed here is a Dolby Processor 750 with a Dolby 5.1 system with screen speakers (L, R, C), an LFE unit installed asymmetrically under the screen and a set of surround speakers at growing height.

The Source for the Sweep test is set asymmetrically at 4,72 mt from the right oblique wall and at 0.52 mt from the screen and the recordings are done in an asymmetric point at row n.5.

5. Measurements

We used two systems for the measurements:



Figure 2. First system layout

The first for the Ambisonics First Order, Binaural and Omnidirectional recordings, composed of three microphones (Neumann KU 100 Dummy Head, Sennheiser Ambeo Soundfield microphone, Behringer ECM 8000 Omnidirectional microphone) with different orientations. Zoom F8 multitrack field recorder is used to record simultaneously (on a SDHC memory unit) all the signals. Recordings are stored as a multitrack 7 channels .wav file (4 channels Ambeo, 2 channels Dummy Head, 1 channel Behringer).

With this system is recorded the ESS reproduced from an omnidirectional Source: the “*SI03AC*” Lookline dodecahedron.

The second one (Eigenmike™) for 3D recording with 32 capsules on a sphere. Signals are delivered to the audio interface through a digital CAT-6 cable (Ethernet) employing the A-net protocol.

The audio interface is linked to the PC via Firewire to Thunderbolt cable adapter. With Eigenmike two sound scenes were recorded: The ESS played by the dodecahedron and a Dolby trailer in 5.1 configuration (“<https://thedigitaltheater.com/dolbytrailers>”) reproduced by the entire electro-acoustic system as found in the rooms.



Figure 3. Eigenmike system layout

6. Results

In the following, the results are presented separately for both the implemented measurement systems.

6.1 First system

From the multitrack 7-channels recording, mono files are extracted and elaborated through Audition 3.0. These recordings are used to extract ISO 3382 parameters for performance spaces.

With “Aurora Plugin for Audition 3.0” the Impulse Responses are extracted as a deconvolution with the inverse ESS.

Now, the extraction of ISO 3382 parameters is performed with Aurora (based on Schroeder integral theory) through the command “*Acoustical Parameters*”.

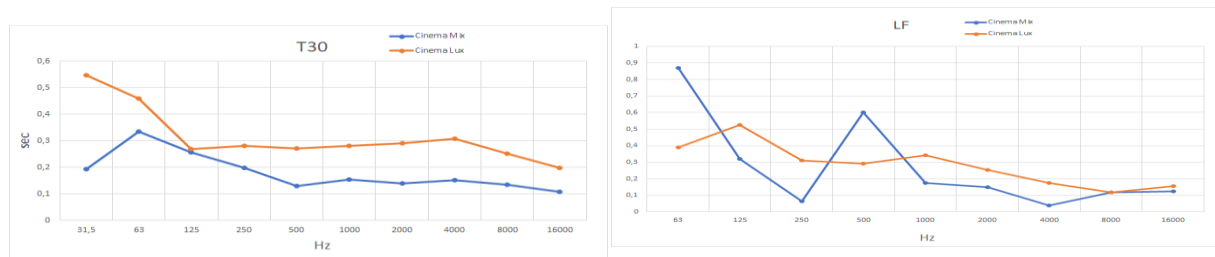


Figure 4. Reverberation T_{30} and Lateral Fraction LF

6.2 Second system

If we start from 5.1 Dolby Trailer recordings or from ESS recordings, the processing to obtain the results is different.

For both the Test Signals, the Eigenmike recordings are performed through the software “*Plogue*” that saves them into a single multitrack 32 channels wav file in PCM format.

Trailer recordings are converted into 3rd order Ambix (B-Format) while ESS recordings are converted into SPS 32 channels (P-Format).

6.3 Creation of VR videos

The ready-made 16 channels Ambix format recording must be aligned (with “*Audition CC 2019*”) to the video shot with the Ricoh Theta V camera

The aligned Ambix 16 channels audio file and the video in MP4 format recorded by the 360° camera are loaded on “*FB360 Encoder*” and two VR video outputs are generated: FB360 Matroska in TBE audio format (8 channels for Oculus VR) and Youtube video in Ambix 1st order audio format (4 channels for jump inspector on Android or VLC).

To see the perceivable differences between the original 5.1 Trailer and the recorded one, the same procedure is performed with the original audio, extracted from the Dolby Trailer video file.

6.4 Visualizing Dolby 5.1 Trailer sound on Panoramic Image

The VST used for the goal is called “*O3A Flare*” (hosted on *Audition CC 2019*) and it produces a view of an Ambix 3rd order stream shown using a rectangular screen region that can be loaded directly in the plugin window through a “load image” button.

Here are presented some snapshots at 18 seconds (Left and Right Surround radiating) of the Trailer.

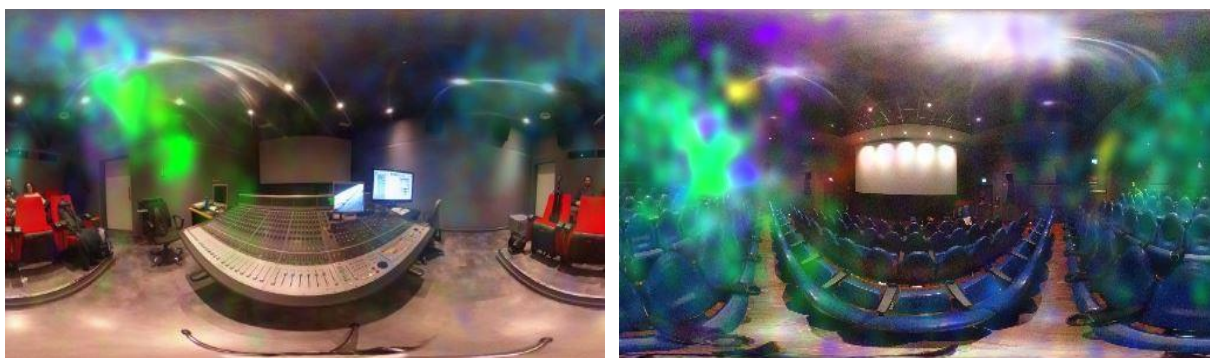


Figure 5: Trailer 5.1 Sound distribution in Dubbing Stage room (left) and Cinema lux n. 10 (right)

6.5 Sound Reflections from Impulse Response on Panoramic Image

Here we start from the ESS RAW recordings of the Eigenmike and we chose to use a MATLAB Suite. ESS 32 channels PCM RAW recording, SPS Matrix, Inverse Sweep and the panoramic image are loaded into MATLAB with a first script (“a_configureFileNamesAndDirectories”).

A second script (“b_multiChConvolve”) allows to extract the 32 IRs.

The third script (“c_generateEncodedFormat”) makes the encoding of extracted IRs from A-format to P-format, so it convolves the 32 IRs in RAW format with the 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.

Last script (“d_spatialIRsVideoMap_sps_image”) creates the video map of sound levels over the panoramic image. Here some meaningful results of this process are reported.

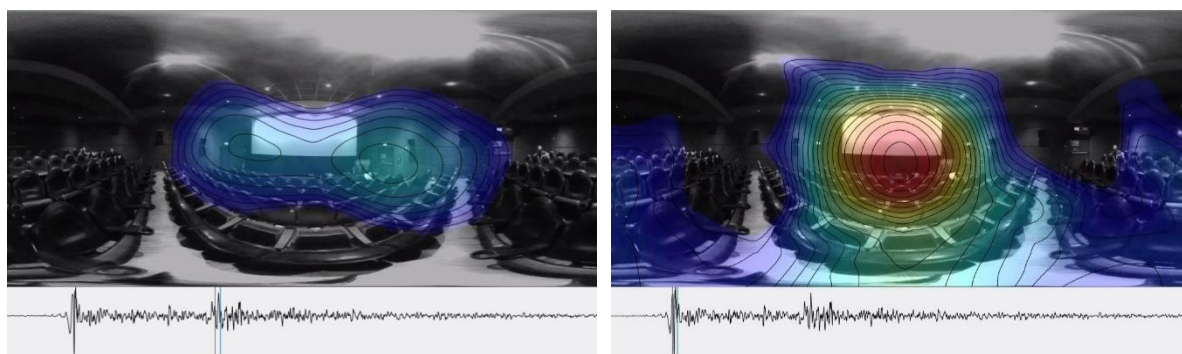


Figure 6. Strong reflection from left upper wall (Left), strong reflection from projector glass (Right), Dubbing Stage room

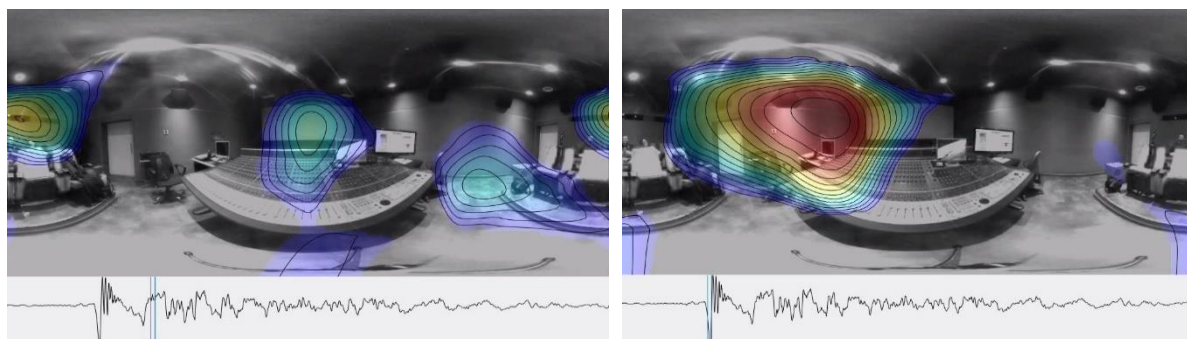


Figure 7. Strong reflection from screen (Left), medium reflection from entrance doors (Right), Cinema Lux

7. Conclusions

To conclude, the idea that the acoustic perception of a movie soundtrack should be the same for both Dubbing rooms and performance cinema rooms 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 systems through equalization should be abandoned because it is based on wrong scientific assumptions, so it is in many cases unhelpful in improving the overall quality of the in-room sound.

To enhance the acoustic of a room at a good degree, one could instead measure the spatial distribution of sound (through impulse response method or through Test material sound visualization over panoramic images) to see which elements produce excessive reflections and to act on them, for example, by putting absorbing or diffusive materials.

The implementation of array microphones (such as Eigenmike 32 capsules or similar), for the recordings in 3D space, and 3D cameras for shooting panoramic images, allow the operator to precisely identify the source of disturbances in the room under study.

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

8. References

- [1] Farina A 2000 110th AES Convention, February 2000
- [2] Lokki T and Pätynen J 2020 *Ac. Science Technology* **41(1)**
- [3] SMPTE, “TC-25CSS B-Chain Frequency and Temporal Response Analysis of Theatres and Dubbing Stages”, October 2014
- [4] Tronchin L and Farina A 1997 *JAES* **45(12)**
- [5] Postma B N J and Katz B F G 2015 *Virtual Reality* **19(3-4)**
- [6] Gerzon M 1975 *JAES* **23(7)**
- [7] Shimokura R, Tronchin L, Cocchi A and Soeta Y 2011 *Journal of Sound and Vibration* **330(14)**
- [8] Farina A and Ayalon R 2003 *AES Conf. Surr. Sound Techniques Techn. Percep*
- [9] Tronchin L 2012 *AES: Journal of the Audio Engineering Society* **60(12)**
- [10] Farina A, Binelli M, Capra A, Armelloni E, Campanini S and Amendola A 2011 International Seminar on Virtual Acoustics, Valencia (Spain), 24-25 November 2011
- [11] Kuunisen A and Lokki T 2015 *Jour. Ac. Society America* **138(5)**
- [12] Tronchin L 2013 *Acta Acustica united with Acustica* **99(1)**
- [13] Postma B N J and Katz B F G 2017 *Jour. Ac. Society America* **142(5)**
- [14] Tervo S, Pätynen J, Kuusinen A and Lokki T 2013 *AES Jour. Audio Eng. Society* **61(1-2)**
- [15] Poletti M A 2018 *JASA* **144(3)**
- [16] Tronchin L and Knight D J 2016 *Int J of Historical Archaeology* **20(1)**
- [17] Pulkki V, Pöntynen H and Santala O *AES Journ. Audio Eng. Society* **67(11)**
- [18] Kaplanis N, Bech S, Lokki T, Van Waterschoot T and Holdt Jensen S 2019 *Journ. Ac. Society America* **146(5)**
- [19] Tronchin L and Coli V L 2015 *AES: Journal of the Audio Engineering Society* **63(9)**
- [20] Hammond B and Jackson P 2018 *IEEE Intern. Conf. Ac. Speech Signal Proc*
- [21] Farina A and Tronchin L 2013 *Acta Acustica United With Acustica* **99(1)**