AuroMax immersive audio

How the three-layer approach leads to a larger sweetspot

DATE 25/07/2019

AUTHORS Max Röhrbein | Solution Architect

Jan Langhammer | Systems Engineer

José Ramón Menzinger | Digital Audio Signal Processing Specialist

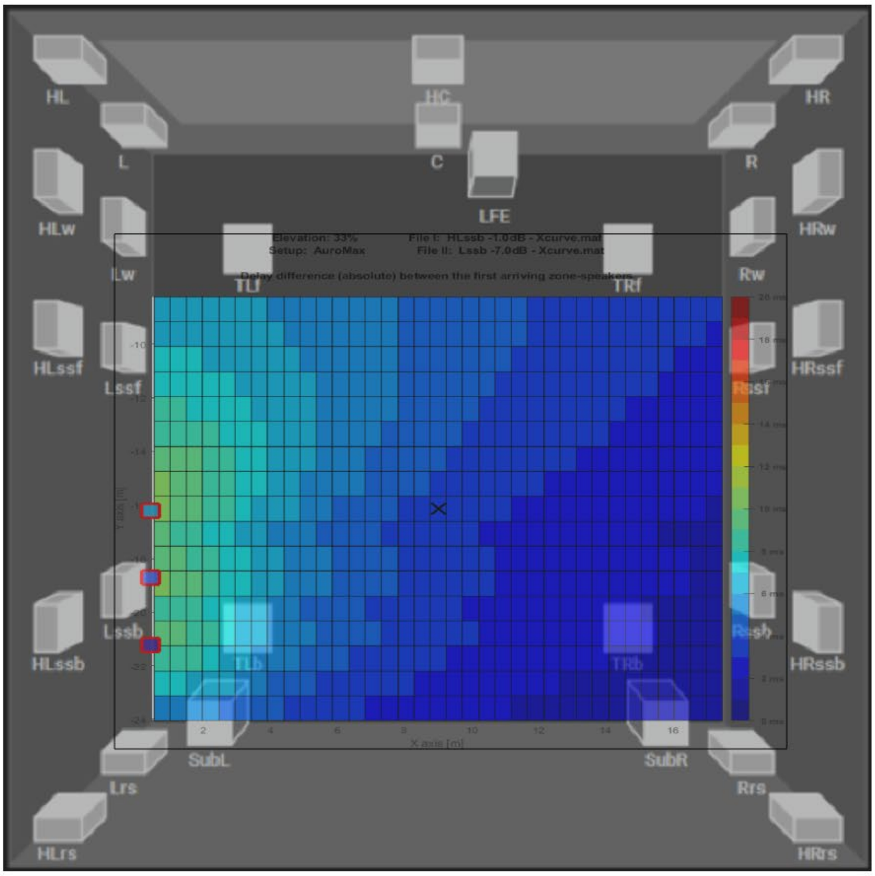


Table of content

Introduction

With immersive audio in cinema based on the open standard SMPTE ST2098-2 becoming reality, Cinionic and Barco deliver their own rendering solution to exhibitors in the cinema market: the AuroMax.

The rendering of the APX processor that drives the AuroMax systems, is based on existing Auro11.1 loudspeaker setups and advances to object-based audio as defined in the SMPTE standard which has become the standard for the premium cinema experience today. The nature of an open standard gives exhibitors the choice of different solutions on the market to playback SMPTE ST2098-2 audio content. One differentiator of existing systems is the arrangement of loudspeakers in a movie theatre. Barco's APX processor is driving loudspeaker setups containing three vertical layers (base, height and top) and three optional versions of horizontal resolution for audio object rendering are available.

Object-based audio claims to introduce an enlarged sweetspot increasing the intended sound experience for a bigger part of the audience in a state-of-the-art movie theatre using immersive audio. This white paper looks at the two most common approaches of loudspeaker arrangements on the market; vertical two-layer and three-layer approaches and their influence on the claim of an enlarged sweetspot but more specifically how the direction in space of individual audio-objects is perceived (localised) in these two different approaches.

Sweetspot definition

The sweetspot is the area where the audience perceives the soundtrack of a movie in the best representation of the intended mix from the content creators. It often refers to the places in the audience area, where correct localization and smooth loudness perception of all sound elements in the room is achieved. I.e. it is about the direction and homogeneous level distribution of a sound in a movie theatre.

Channel-based systems

In channel-based systems the sweetspot is generally small, since applied amplitude panning laws are only valid for defined loudspeaker positions around one listener position (seat) in the audience area. Due to the size of movie theatres the sweet spot becomes generally bigger because of larger distances from speakers to the relevant audience area centre.

To further enlarge the sweetspot in a 5.1 system, speaker arrays are used for the left and right surround channels along the entire side and rear walls. A major drawback of this design is that sound objects can only be drawn into the room left or right without further precise positioning in depth. The 7.1 system was introduced adding another horizontal speaker channel to move sounds from side to rear. Again 7.1 systems are driven in speaker arrays to enlarge the sweetspot in larger cinema theatres.

Object-based systems

Object-based (immersive) systems in cinemas use individual speaker zones or even single speakers along the surrounding theatre walls to increase the resolution of surround channels. With these additional zones it is now possible to create more detailed movements around the audience. In addition, with the introduction of height and top layers not only the horizontal plane can be used but also vertical positions and movements of sound objects are new measures to "immerse" the audience in these systems.

APX system

The loudspeaker layouts driven by Barco's APX processor were designed to add "just enough" zones improving the movements of sound objects in the horizontal surrounds and introduces an intermediate height layer (three-layer approach) between the common main layer and top/ceiling layer (two-layer approach). Its intention is to not only increase the horizontal but also the vertical resolution for object movements, which is considered to be critical for a true immersive experience in all three dimensions in space.

Speaker arrays

The APX system design maintains the idea of using speaker arrays in each zone. The initial and commonly accepted idea to enlarge the sweetspot when using broader zones is kept as well as having zones with better coverage for a bigger area across the audience. Not only the directivity of these zones is broader, but also the resulting sound pressure levels are higher, which results in lower speaker requirements to meet cinema sound level specifications (reference of 85dBC at the preferred listening position -PLP-) for example. Picture 1) shows a total SPL mapping of one speaker vs. an array made of three speakers representing an audio-object at the back-left side wall as simulated in a calibrated 18m wide screen movie theatre.

Picture 1: Total SPL coverage @2.5 kHz of an individual speaker (left) vs. an APX speaker array (right)

The resulting maps show that the coverage and level distribution at 2.5kHz is more homogeneous when using arrays (right) as compared to a single speaker (left). One speaker potentially introduces "dead" zones (turquoise colour) on the audience area for this frequency band as the level is significantly lower outside its directivity pattern. Using the same type of speaker but driven as an array these "dead" zones do not occur and the overall level difference in the main beam of the speaker is less drastic.

Vertical resolution

With the introduction of an intermediate height layer between the main layer and the top/ceiling layer APX systems achieve higher resolution in the vertical domain, which in return leads to better localization of elevated sound objects and smooth sound movements from surrounds into the ceiling and vice versa. By closing the gap immanent in the two-layer approach the vertical error is minimized and enlarges the sweet spot in this domain, as well.

Since the spatial resolution of human hearing is lower in the vertical as in the horizontal plane, only adding one speaker layer results in a full filed vertical coverage.

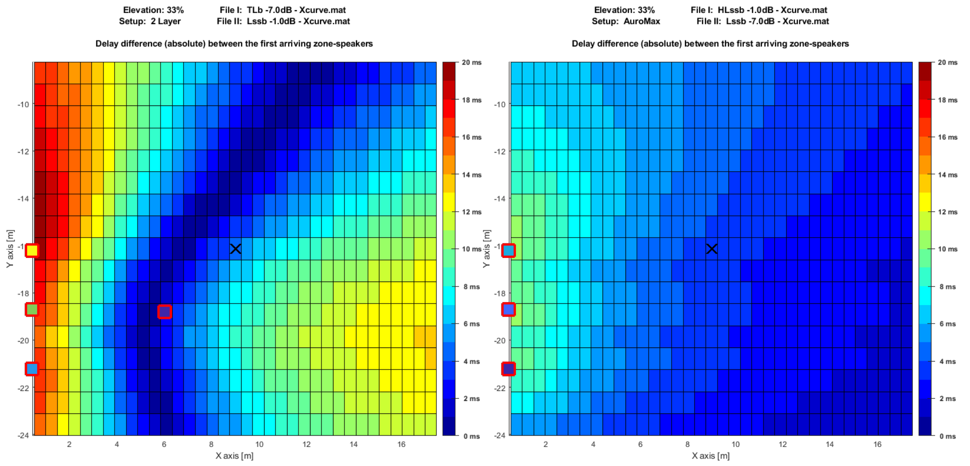
The psycho-acoustic precedence effect plays a major role of how sound object positions are perceived by the listener. The spatial location of a sound emitted by two sources is dominated by the location of the first-arriving source (the first wave front).

In immersive cinema sound systems an elevated object (sound) will be represented by a combination of speakers (source) with different levels. Psycho-acoustic research (Helmut Haas, Einfluss eines Einfachechos auf die Hörsamkeit von Sprache, in: Acustica, 1/1951, 49-58, hier: 52 f.) shows that at different arrival times between two speakers below 10ms the level differences of these speakers will be the determining factor for the perceived object position.

In a two-layer approach all elevated objects will be presented as a combination of speakers with different levels from the main layer and top layer. In a three-layer approach it will be a combination of speakers with different levels in the main and height layer between 0% and 50% object height and a combination of speakers with different levels in the height and top layer between 50% and 100% object height.

Using acoustic simulation *(EASE & Matlab™)* of a calibrated cinema speaker system (85dBC @ PLP and X-Curve tuning) in a standard movie theatre with 18m screen width it is possible to show time arrival and level differences of various speaker combinations for elevated object positions across the entire audience area.

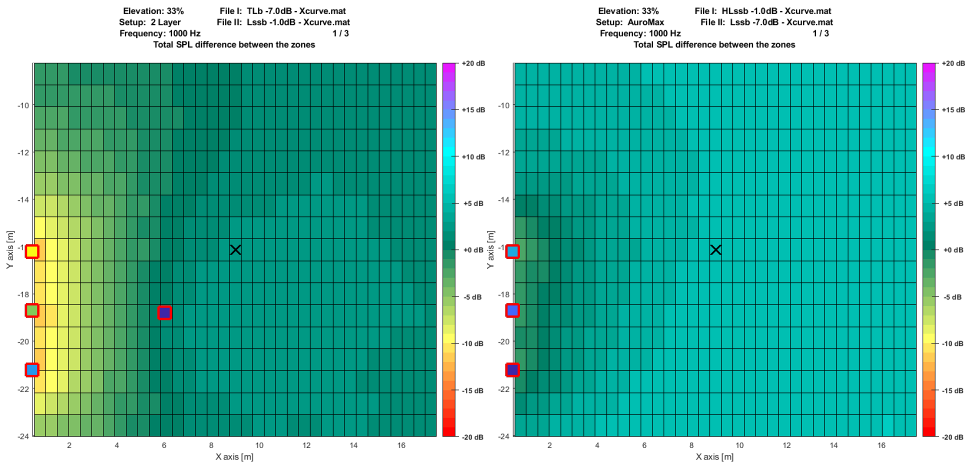
Picture 2) shows the time arrival differences for a speaker combination representing an elevated object of 33% height on the left back side wall; the room in a two-layer approach (left) vs a three-layer approach (right). Different coloration represents different time arrival values (in milliseconds).



Picture 2: Time arrival differences of an elevated object at 33% for both systems (left: two-layer, right: three-layer)

The map on the left, the two-layer approach shows time differences below 10ms in the central listening area with higher differences towards the edges. The three-layer approach in the map on the right shows arrival time differences below 10ms across the entire room. As mentioned above at time differences below 10ms research suggests to look at level differences of the speaker combinations to give a better indication on a perceived object position.

Picture 3 shows the level difference for an elevated object at the same position as in Picture 2) for a two-layer (left) and three-layer (right) approach.

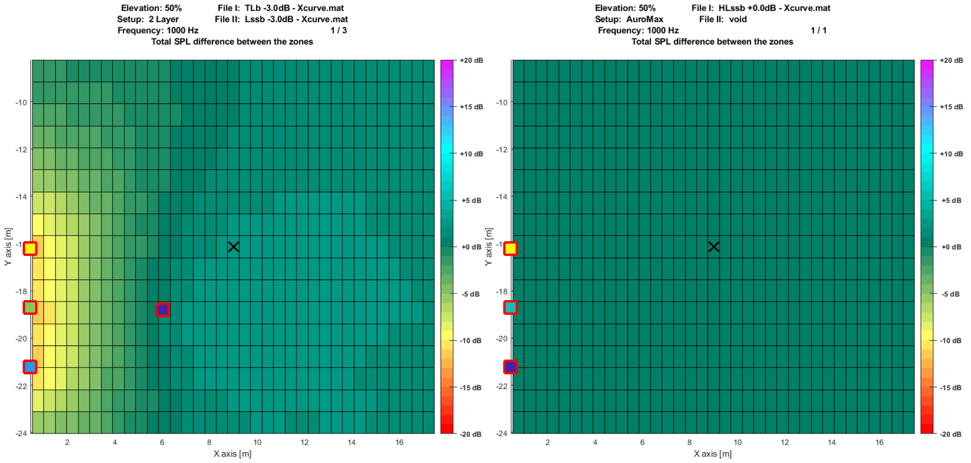


Picture 3: Level difference of an elevated object at 33% for both systems (left: two-layer, right: three-layer)

To understand this graph, a positive dB value on the color scale (towards blue and magenta) shows higher levels for the top layer (left) or height layer speaker (right). Negative values (toward yellow and red) show higher levels for the main layer speaker.

The right picture (three-layer approach) shows already a more homogeneous mapping with lower level differences across the entire audience area than the two-layer approach with level differences of up to 10dB SPL towards the left edge.

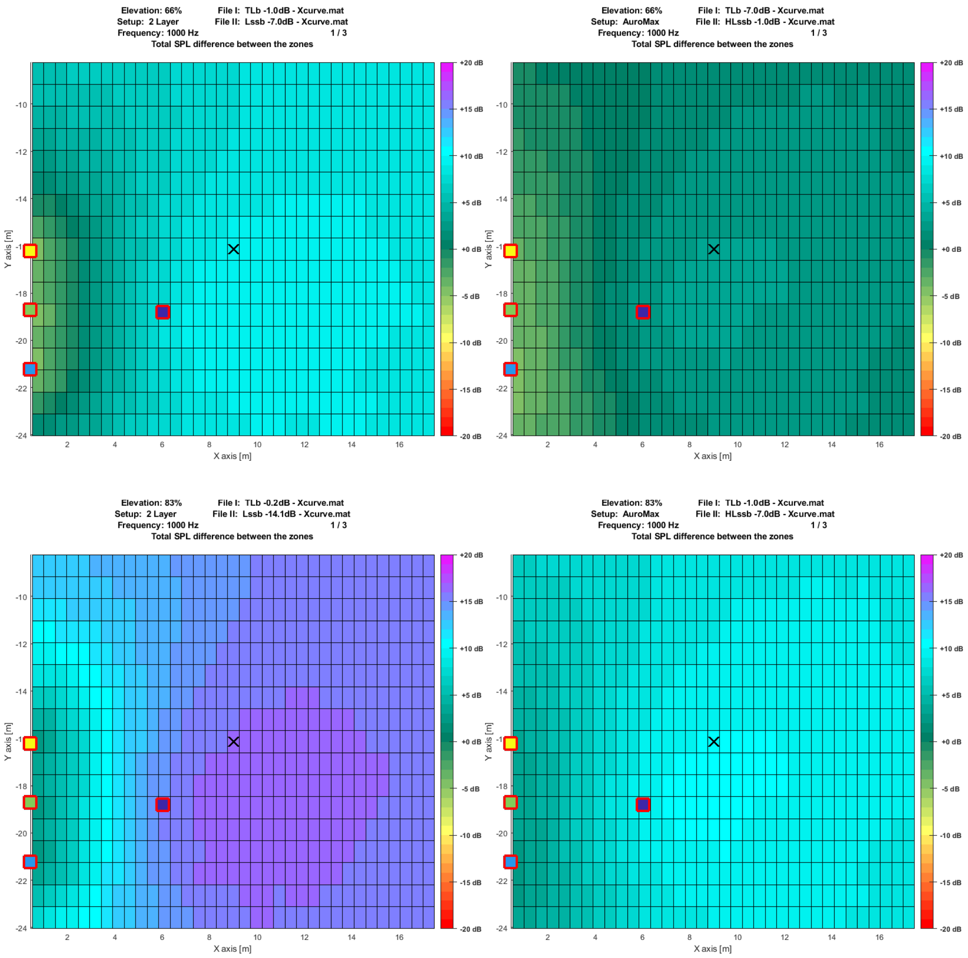
Moving the sound object upwards, Picture 4) shows the level difference for an elevated object at 50% height for both systems.



Picture 4: Level difference of an elevated object at 50% for both systems (left: two-layer, right: three-layer)

At 50% object height the two-layer system represents a virtual/phantom object as a combination of speakers from the main and top layer, whereas the three-layer system has physical speakers available leading to no level differences at all and thus a more stable perception of that position.

Continue moving up the sound object, Picture 5) shows the level difference for an elevated object at 66% and 83% for both systems.



Picture 5: Level difference of an elevated objects at 66% (top) and 83% (bottom) for both systems (left: two-layer, right: three-layer)

All graphs in Picture 5 highlight the fact that a two-layer approach always show higher level differences that are more varied across the audience area than three-layer systems. This results in different perception of object positions depending on the listener's seat in the audience. The top left picture shows level differences of around 10dB for more than the entire right half of the audience area. In this case more than half of the audience will perceive an audio-object rather from the top layer (100%) than from the intended height of 63%.

It can be assumed that a smaller level difference of two speaker zones will lead to a better representation of phantom sources between two layers, whereas at big level differences the object position will most likely be perceived from the louder layer and lead to a wrong localization of vertically elevated sound objects (due to the short arrival time difference explained in Picture 2). Psycho-acoustic research amounts a level difference of around 10dB between two speakers before one speaker (layer) will determine the perceived object position as opposed to phantom image.

Looking at the simulation results shown in this report it can be concluded that above 50% object height a two-layer approach will lead to wrong localization of elevated sources with a tendency to pronounced top layer. These differences of system approaches converge when further increasing the object height. Above 80% object height the level difference is larger than 10dB SPL for three-layer systems, as well. But the distribution of level differences is more uniform in the three-layer approach.

Conclusion

Acoustic simulations of a calibrated standard movie theatre (18m screen width) and standard signal processing show the difference of SPL coverage of speaker arrays vs. single speaker in the horizontal plane. A graphical representation was introduced to show potential localization of elevated audio objects as represented through the open standard SMPTE ST2098-2 (Immersive Audio Bitstream – IAB).

Speaker arrays in the horizontal plane

- SPL mappings show an even coverage of speaker arrays over a large audience area.
    - ➢ No "dead" zones on the audience area.
- Smoother SPL distribution of speaker arrays for close-by and far-away seats
- Lower speaker requirements to meet cinema SPL specifications

Elevated source representation in the three-layer approach

- Shows minimum delay differences (<10ms) between two speakers to represent an elevated sound object across the entire audience area.
    - ➢ Only level differences of speaker combinations will determine the localization of audio objects.
- Level differences between two speakers are below 10dB for all elevated sources below 80% height across the entire audience area.
    - ➢ Correct localization can be assumed for at least 80% of the object elevation range.
- Shows minimum deviation of level differences for elevated sources across the entire audience area
    - ➢ Homogeneous sound field for vertical phantom sources across the entire audience area
    - ➢ Above 80% elevation the localization "error" is the same for all listeners