Audio Engineering Society

Conference Paper 22

Presented at the International Acoustics & Sound Reinforcement Conference
2024 January 23-26, Le Mans, France

Function group approach to immersive audio system design for stage-based applications

Javier Frutos-Bonilla[1]

[1]d&b audiotechnik

Correspondence should be addressed to Javier Frutos-Bonilla (`javier.frutos-bonilla@dbaudio.com`)

ABSTRACT

The latest generation of sound reinforcement systems should not only aim to achieve excellent sound amplification but also to recreate a plausible spatial and room-acoustical impression. This paper describes an object-based sound reproduction approach based on assigning function groups to the loudspeakers planned in the design phase. Finally, it will be shown how these concepts are implemented in an actual theatre-style installation for 3780 people.

1 Introduction

The development of the first delta stereophony systems [1] back in 70´s and 80´s can be considered the first modern approach (DSP based) to spatial audio for sound reinforcement applications, beyond stereo or LCR design. For the last decade, the use of different spatial audio solutions for a wide range of applications has increased noticeably. Nowadays, it is a serious (if not the first) deliberation when designing a new sound system for most of the renowned theatres worldwide.

This paper is structured as follows: section 2 gives an insight into the d&b audiotechnik immersive audio platform; section 3 underlines the psychoacoustic model in which the algorithms are based; sections 4, 5, and 6 explain the system design approach, the concept of function groups and the alignment between them; and finally, a theatre-style installation is presented as a case study.

2 d&b Soundscape immersive platform

The DS100 processor is the platform underneath d&b Soundscape. It is able to handle 64 Dante inputs and outputs, in a 64 x 64 audio matrix with level and delay adjustments at all cross points. It includes input and output processing including gain, EQ, delay or polarity switches. To pre-configure the processor, the user needs to use the d&b ArrayCalc simulation software for planning and can control the system using d&b R1 Remote control software.

The d&b En-Scene software module provides dynamic object positioning within the acoustic space. En-Scene algorithm is a form of vector-based panning between all the available loudspeakers. Vector-based panning could best be described by a source or signal from the stage that is distributed to all relevant loudspeakers with a unique set of gains, delays and, if required, filters. With this approach, a loudspeaker system is converted into an acoustic environment in which sound objects can be placed.

The En-Space software allows the user to mimic a different acoustic environment (on top of the existing one) for the entire audience using an in-line approach. The algorithm is based on multiple pre-captured impulse responses, which have been acquired in a variety of different well-designed and renowned rooms. Thus, the acoustic imprint, gathered at specific points in these rooms, is adapted to the actual design of the loudspeaker system which is used for reproduction in the actual case.

R1 software allows an object to be positioned on a graphic representation. However, most parameters within the DS100 Signal Engine can be controlled via OSC messages. It is then possible to play a show just using show control systems, DAWs or tracking devices. VST, and AAX plugins supporting the DS100 OSC protocol are also provided.

3 Psychoacoustic model

The psychoacoustic model behind the implemented panning algorithm will set the design rules of the loudspeaker system. The En-Scene algorithm considers the mix of the listeners psycho-acoustical perception as well as acoustic effects between the sources to calculate the transfer functions for the relevant matrix cross-points [5]. Maintaining the rules of the precedence effect (often referred as "law of the first wavefront") between the first impinging loudspeaker and its neighbours ensures accurate spatial localization (yellow area of Figure 1).

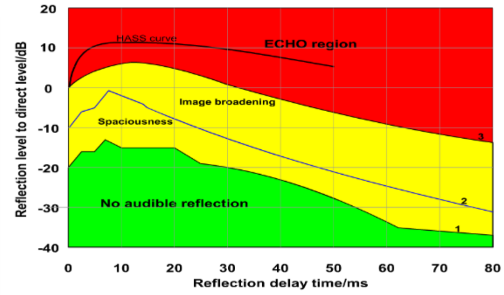

Figure 1. Adaptation of single reflection perception from [3].

The object energy distribution between adjacent loudspeakers can be controlled using the spread parameter. It influences the perceived width of an object and can be adjusted between 0 and 1. A value close to 1 will distribute the energy more homogeneously between all sources creating a diffuse object perception (but increasing the risk of breaking the precedence effect for certain listening positions and certain loudspeaker setups). Opposite to this, a value closer to 0 will focus the energy on fewer sources sharpening the image perception and creating a directness effect for the object. As a drawback of this configuration, the object reproduction may rely on a single source, on its coverage and SPL capabilities in order to cover the complete audience. There is no one-correct-value for all cases and the spread is also used as part of the artistic mix.

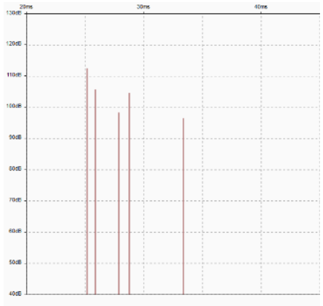

Figure 2. Exemplary echogram for an object with an spread of 0.5, placed 1m. behind the stage lip and a listener at coordinates [x=2m, y=8m] from the stage. First incoming loudspeaker determines the direction while the neighbouring ones slightly broaden the spatial image.

En-Scene supports all kinds of horizontal loudspeaker distribution. Linear arrangement as well as convex or concave designs are possible. The algorithm considers both, the position, and the orientation of the loudspeakers towards the sound object. The individual levels and delays of each source depend on the horizontal angle between the object and the loudspeaker axis, its distance, and the object properties [5].

4 System design: coverage and venue segmentation

The required SPL and target uniformity of the given application, the architecture of the venue, and the position where the loudspeakers can be installed will dictate whether a point source, loudspeaker cluster, or line array should be used for each source position.

The number of positions needed to recreate a realistic object imaging effect depends on the type of loudspeakers chosen, the distance to the loudspeakers, and the accuracy of the localization required. The localization resolution of the human binaural hearing system depends on the incidence angle [2], having the highest precision in front of us. However, when we have the visual reference of the source generating the sound, a certain offset between the visual and acoustical stimulus is tolerated by our brain. This effect is known as Ventriloquism and simplifying, it states that visual information dominates when the audio-visual spatial information mismatch (to a certain degree) perceiving both stimuli at the direction of the visual cue. In [4], the authors review the different results published on multiple papers based on the stimulus, setup, and type of test and they performed their own one. We also considered up to 10 degrees mismatch as a good value, while up to 20 degrees could be acceptable for certain non-localization critical applications. ArrayCalc software can be used not only to simulate the direct SPL of the loudspeaker included in the design but also to predict the perceived direction of the object position.

Vertical coverage represents the first challenge when doing immersive system designs for stage-based applications. Independently of the number of positions, there will be multiple source positions above the stage. Although it can be possible to cover from first row to the last seat with one single source (especially line arrays or clusters), it is advisable to use frontfills to cover the first rows to lower down the spatial image and keep it on the stage.

In terms of horizontal loudspeaker coverage, the worst-case scenario would be when an object is so configured that one single source should cover most of the audience. In this case, loudspeakers covering close audiences need to have larger horizontal dispersion compared to conventional system design. The resulting overlap between loudspeakers is also needed to recreate the sensation of a moving object since the listeners should be covered by all the loudspeakers activated by such movement. Apart from carefully choosing the right dispersion for each loudspeaker position, turning the loudspeakers at the outer edges of the system towards the center will improve the cross-coverage and the localization of objects.

As a result of this, the design of sound reinforcement systems entails the segmentation of the venue into different zones that need to be addressed individually and aligned together to play as a complete system. This fact is a clear distinction against studio or cinema applications in which there is a unique zone that is covered (ideally) by all loudspeakers.

Apart from the source selection decision, the most common factors that influence the segmentation of the venue into different areas are visual constraints like the height of the lowest edge of the loudspeaker above the stage; weight limitations of the rigging points (forcing the use of smaller and shorter lines arrays together with delay lines); and finally the restrictions imposed by the architects in terms of loudspeaker placement.

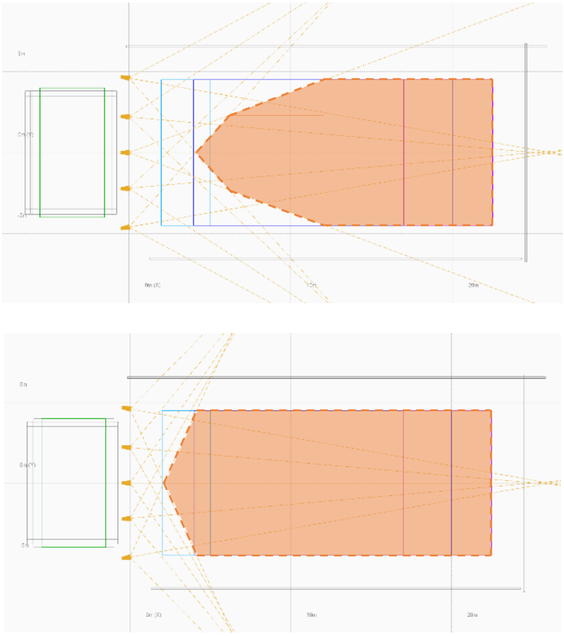

Figure 3. Overlap area of 5 point sources for 75° horizontal dispersion (upper picture) and 110° (bottom picture).

5 Function groups

Once the venue has been properly segmented into different areas based on loudspeaker characteristics and external restrictions, the loudspeakers assigned to each area will be processed so that they work together to create the spatial image of the sound object. This process is known as function group (FG) assignment, and it is the core of the En-Scene software module.

There are 11 different function groups shown in Table 1 [5]. Although the goal of each function group is the same, representing the audio object in its given position, the processing to achieve this goal varies significantly depending on the function assigned. For example, the processing of the function group “Frontfill” is completely different from the function group “Outfill” due to the fact that the first one needs to cover an extremely wide stage opening angle against the narrow one for the seats under its coverage.

An exemplary immersive system design with 3 function groups is graphically represented in Figures 4 and 5. Notice that the narrow stage opening at the outfill area in blue makes the use of a single source position sufficient. The minor spill from the red area would help to widen the spatial image.

	Function group	Short description
1	Sub array	Mono downmix central subwoofer
2	Main	Object positioning
3	Frontfill	Object positioning first rows
4	Surround	Lateral and rear object positioning
5	Subs group	Low frequency object positioning
6	Outfill	Lateral extension to Main function group
7	Delay line	Depth extension with object positioning
8	Mono Out	Mono downmix for auxiliary purposes
9	Ceiling	Overhead loudspeakers
10	Outfill embedded	Similar to standard Outill but sharing level with main
11	Delay embedded	Similar to standard Delay but sharing level with main

Table 1. Function groups available and their functional purpose.

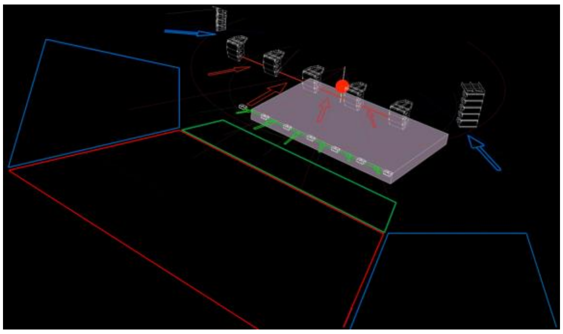

Figure 4. The different areas covered by the Main FG (red), Frontfill FG (green) and Outfill FG (blue).

In the case of Figure 5, the length of the arrays is not long enough to cover the last rows, so a delay line needs to be included. Notice that due to the limitation of the sources' horizontal coverage, the delay area is again divided into 3 narrower zones and addressed individually.

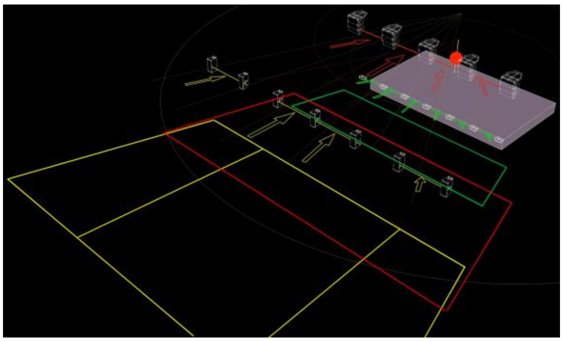

Figure 5. The areas covered by the Main FG (red), Frontfill FG (green) and Delay FG (yellow).

The different relationships between function groups are also stably defined in their implementation, establishing a hierarchal relationship between them. We can define 4 kinds of relations:

1. Combination: when two or more function groups combine themselves to play together sharing the object energy distribution (Main and Surround). The loudspeakers assigned need to cover the same area so that the movement and placement are smooth enough.

2. Parallel: when two or more function groups are representing at the same time the same object position for two different areas with minor overlap between them (Main and Frontfill).

3. Dependency: when the object energy played back by the primary function group influences the level of the secondary (Main and Outfill / Delay Embedded).

4. Independency: when the function groups work by themselves independently of each other (Main and Aux).

Except for the "dependency" relation, the En-Scene algorithm ensures a constant unity gain for each object, independently of the number of sources contained in each function group.

6 Function group alignment

The last important step in the design phase is the alignment of the different function groups. For a stereo reinforcement scenario, the alignment reference is always a static loudspeaker; the delay under balcony is aligned to left or right, out-fill left to stereo left and so on. However, for immersive audio, the references are the multiple objects that not only are at different positions but also can be moved. It means the alignment delay changes depending on the relative position between the object, the loudspeakers in Main FG, and the loudspeakers in the Delay or Outfill FG. As a result of this, the alignment delay can only be handled dynamically by the audio processor. The wider the venue is, the higher the variation between alignment delays.

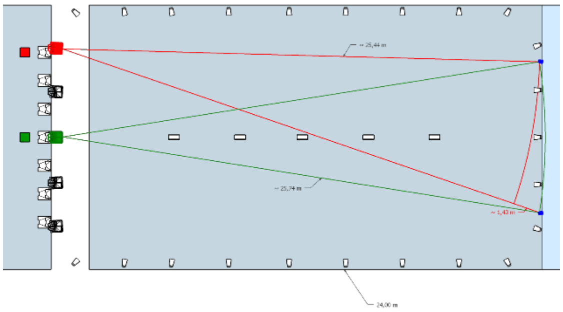

Figure 6. Path difference between green and red objects separate by just 4.5m to the delay loudspeakers is already ~1.5m.

7 Case study: Church Impact Centre Chrétien.

The Church Impact Centre Chrétien is a novel installation that hosts modern Christian services in which the spoken word from the pastor is supported by music and audio-video-lights effects in the venue.

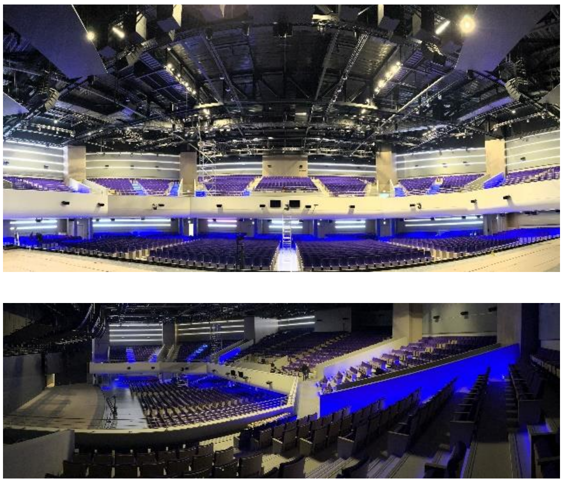

Figure 7. Front and side pictures of the venue.

The project started as a stereo design, focusing the attention on video and light. After visiting an immersive audio demo in a smaller venue, the pastor asked for the same experience in a bigger room. The audio system had to provide excellent speech and audio quality and natural localization of talkers on the stage. It should help to increase the connection of the pastor with the prayers and allow the possibility to use theatrical and room emulation effects to enhance the experience of the parishioners.

The venue is built as a half oval with two levels and has a capacity of 3780 seats. The lower level measures approximately 60m. wide and 30m. long with the last 10m. covered by the upper balcony. The upper level measures 60m. wide and extends up to a depth of 40m. with the balcony starting at 20m. from the stage lip (see Figure 7).

Mounting positions for the main system were limited to the position of the mother grid, front-fills should be installed inside the stage lip and the position of surrounds and delays could be freely chosen. The lowest height of the main arrays had to be 9.5m. above the stage due to visual shadowing onto the LED screen.

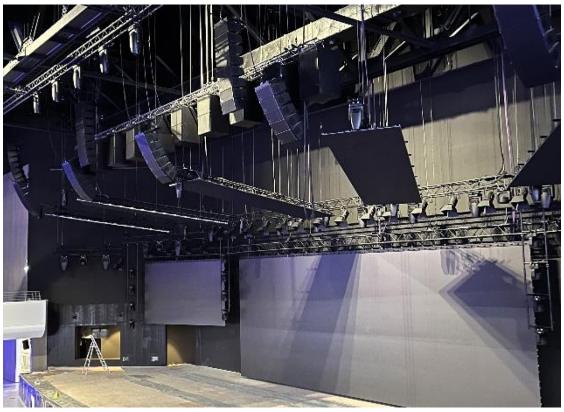

Figure 8. Picture of the main system flown above the LED screens.

The system design was designed by EAS team in France back in 2020 using ArrayCalc V10. Although initially it was intended to use 5 line arrays covering the lower and upper levels (most straightforward approach), different constraints forced us to split the main coverage into two groups: the first one with a higher density with 5 arrays for the stalls in the lower level, and a second group with just two main arrays for the upper balcony. It was decided to segment the venue into 3 areas with its own sub-divisions: the stalls and the under-balcony in the lower level, and the upper balcony. The detailed segmentation and loudspeaker selection can be found in Table 2 and Figure 9.

Segmented area	Loudspeakers	Amount
Stalls main stalls	5 arrays V-series	6 cabinets per array
Stalls front fills	44S Flush-mount	14
Stalls surround	Y10P	10
Under-balcony delay front	Y10P	20
Under-balcony delay surround	Y10P	18
Upper balcony main	2 arrays V-series	8 cabinets per array
Upper balcony outfill	2 arrays V-series	7 cabinets per array
Upper balcony surround	Y10P	16
Subwoofers	V-Sub as flown subarray	10 double positions

Table 2. Venue segmentation and loudspeaker selection.

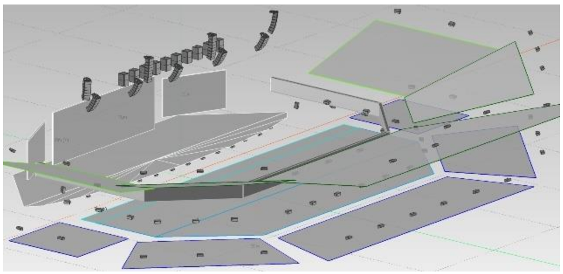

Figure 9. Colour coded venue model. In light blue, the main area, dark blue the under-balcony and in green the upper balcony.

The system requires 90 DS100 outputs, therefore two devices had to be cascaded in order to increase the number of outputs. Table 3 shows the DS100 output and function group distribution. Input routing into the two devices from the console output is straightforward with Dante routing. Metadata and control synchronization are achieved directly using R1. For third party OSC control software that cannot send messages to two recipients, a new piece of software was used for the first time: En-bridge. Between other functionalities, it can work as an OSC router receiving packages on one port and forwarding them to multiple ones.

Segmented area	DS100 outputs	Function group
Stalls main stalls	5 - DS100_A	Main_A
Stalls front fills	14 - DS100_A	4x Frontfill
Stalls surround	10 - DS100_B	Surround_A
Under-balcony delay front	20 - DS100_A	Delay
Under-balcony delay surround	18 - DS100_A	Delay
Upper balcony main	2 - DS100_B	Main_B
Upper balcony outfill	2 - DS100_B	Outfill
Upper balcony surround	16 - DS100_B	Surround_B
Subwoofers	1 - DS100_A	Sub-array

Table 3. DS100 output and function group assignment

Like in any novel installation, there were some interesting learnings during the commissioning of the system:

1. The division of the main system into two groups of sources worked better than expected. Although the localization accuracy in the upper balcony is not so precise as in the stalls, the longer distance to the stage helped to relativize the problem.

2. Initially all frontfill loudspeakers were included in a single FG. When the object moved to the side of the stage, there was no coverage on the other side. Increasing the spread of the object improved the energy distribution but made it sound too diffuse. The solution was to divide the frontfill loudspeakers into 4 Frontfill FG according to the seating block. Since each FG keeps a unity gain, each seating block always has at least one frontfill loudspeaker activated. Its cross coverage was reduced from 40m (stage width) to 8m (block width). The big distance from the stage to the first seat (4m) also helps with the uniformity of the coverage.

3. Although initially the object positioning was intended to be restricted to the stage, once the customer discovered the possibilities of object-based audio, they started to move objects outside the stage. The FG of the frontal under-balcony delay loudspeakers

was then changed from delay embedded to delay so that they not only work for objects on the stage but also, act as delay line for objects placed on both sides of the venue.

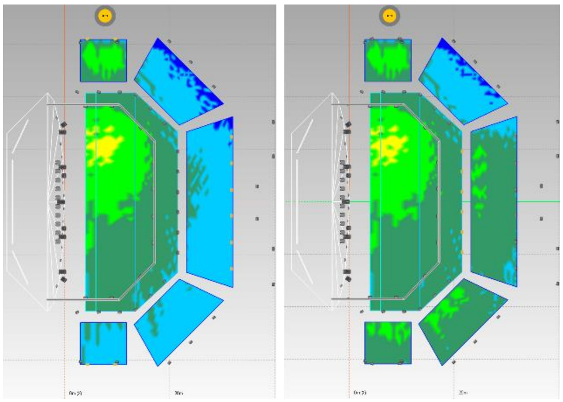

Figure 10. SPL at 2KHz +/- 1 octave for the given object position (yellow bubble). On the left the initial configuration, on the right with the frontal under-balcony delay loudspeakers assigned to the delay FG. Each colour change represents 6dB.

4. The communication with the light and video departments took place in time so that each department had time to adjust its systems.

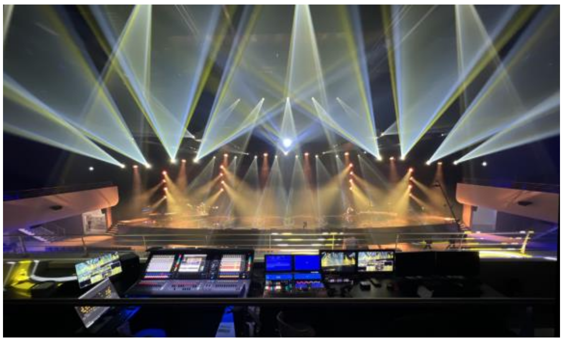

Figure 11. Light design helps to conceal the arrays.

5. The FG alignment works almost out of the box when the object delay is in use. We just had to make minor manual fine-tuning by hearing (a couple of milliseconds here and there).

6. The Y10P for the under-balcony delay surround and under-balcony delay front were oversized. A smaller loudspeaker like 8S would have done the job without any problem. The fact that the 8S has a conical 100°-degree coverage would have helped the listeners close to the walls.

7. Due to budget restrictions, there were no transition loudspeakers between the main and the surround systems. Although the customer thought that they would not need it, we have learned that it is a good practice to include them.

8 Conclusions

The function group approach to immersive audio system design for stage-based applications has been proven to be fully functional even for very demanding venues like the Church Impact Centre Chrétien. Being in control of the time parameter not only to reproduce the object but also to align the different function groups is critical for this kind of application.

Figure 12. Venue during one of the services.

References

[1] W. Ahnert et al., “The complex simulation of acoustical sound fields by the Delta Stereophony System”. *81th Convention of the Audio Engineering Society* (1986)

[2] J. Blauert, *Spatial Hearing: the Psychophysics of Human Sound Localization*. The MIT Press, 2 ed. (1997)

[3] F. Toole, *Sound Reproduction: The Acoustics and Psychoacoustics of Loudspeakers and Rooms*, Routledge, 3 ed. (2017).

[4] H. Stenzel and P. J. B Jackson, “Perceptual thresholds of audio-visual spatial coherence for a variety of audio-visual objects” *AES International Conference on Audio for Virtual and Augmented Reality* (2018)

[5] d&b audiotechnik TI 501, “d&b Soundscape System design and operation,” version 1.10 (2023), www.dbsoundscape.com