# Study and Implementation of 3D Sound Decoding Algorithms for Loudspeaker Arrays of Different Geometries

Pinto, F.[34], Paulo, J.[12], Tavares, G.[34]

*Abstract*—**Nowadays, the increasing development of 3D audio in several fields of multimedia, has offered a greater under- standing as to how spatial audio provides a immersive and realistic sound environment perception. Spatial audio mapping and decoding can be achieved using existing technology, such as Vector-Base Amplitude Panning (VBAP) and Ambisonics.**

**VBAP places a virtual sound source in different locations by controlling each loudspeaker's output amplitude. Ambisonics is a recording and reproduction method containing a sound field representation, which may be decoded to any loudspeaker array. Perceptual tests were carried out using headphones and a multichannel loudspeaker array in an anechoic chamber at Instituto Superior Técnico (IST) to compare the performance of the two spatialization methods, in terms of localization accuracy.**

***Index Terms*—3D sound, immersive sound, auditory percep- tion, multichannel loudspeaker array, VBAP, Ambisonics.**

## I. Introduction

3D audio has the ability to assist the listener to perceive sound as it would exist in the real world, offering the user a faithful reproduction of the spatial aspects of a recorded sound. Using already existing technology, it is now possible to position sound at any location in space and then decode it to be played through either standard stereo headphones or multiple loudspeakers arranged in a specific geometry.

Vector Base Amplitude Panning (VBAP) places a virtual sound source at different locations by adaptively controlling the output amplitude of each individual loudspeaker. This technique relies heavily on the fact that the human ear can perceive sound directions effectively based on the level differ- ence between the ear drums.

Ambisonics is another method and it takes into account the sound field's directional properties. First order Ambisonics (FOA) recordings contain a representation of the sound field, taking into account certain physical properties of the acoustic field, such as the sound pressure.

[1] ISEL-Instituto Superior de Engenharia de Lisboa, Portugal
[2] LAA - Lab. de Áudio e Acústica do ISEL, Portugal

[3] INESC-ID - Instituto de Engenharia de Sistemas e Computadores -ID Portugal
[4] IST - Instituto Superior Técnico, Portugal

Francisco Pinto - franciscomcpinto@tecnico.ulisboa.pt
Joel Paulo - joel.paulo@isel.pt
Gonçalo Tavares - goncalo.tavares@inesc-id.pt

This project is comparative study between the VBAP and Ambisonics methods by using a multichannel array of loud- speakers and a number of subjects, in order to evaluate the differences in multiple settings by conducting perceptual tests.

## II. Background

### A. Spatial Hearing

The ability of humans to localize sound sources is based on real- time analysis of two signals entering the ear canals. Binaural (or interaural) and monaural cues are important to spatial perception of sound [1], [2].

Binaural cues are derived from the differences in the signals between the two ears. Interaural Time Difference (ITD) and Interaural Level Difference (ILD) are the main acoustic cues in general [2]. ILD is also referred to as Interaural Intensity Difference (IID).

However, ILD and ITD cues could be ambiguous. Due to the symmetry of the head, sounds emitted from different directions could share the same ILD and ITD. To solve this ambiguity, two mechanisms have been proven to work: head movements and spectral filtering by the outer ears (pinnae) [2], [3].

Generally, spatial hearing is a three-dimensional phe- nomenon and therefore requires a three-dimensional coordi- nate system to describe the acoustic environment. To denote sound source directions in relation to the listener's head in an acoustic environment spherical coordinates are often used. Conventionally, spherical coordinates are denoted by the azimuth angle $\varphi$ , $-180^\circ \leqslant \varphi < 180^\circ$, the elevation angle $\delta$, $-90^\circ \leqslant \delta < 90^\circ$ and the distance $r$ of a sound source.

### B. Vector Base Amplitude Panning

Vector base amplitude panning (VBAP) [5] is a method proposed by Ville Pulkki in the late 90's, to calculate gain factors for pair-wise or triplet-wise amplitude panning [6]. This technique uses a triplet of speakers with gain weightings to pan a point source in a 3D speaker array. Three vectors are calculated for the origin of the a virtual sound source, defined by a vector ($p$) with components:

$$p_x = \cos\varphi\cos\lambda,\ \ p_y = \sin\varphi\cos\lambda,\ \ p_z = \sin\varphi \qquad (1)$$

The gain coefficients for the loudspeakers triplets is deter- mined using equation 2:

https://doi.org/10.17758/EIRAI19.F0623115

$$g_{123} = p^T L_{123}^{-1} = \begin{bmatrix} p_x & p_y & p_z \end{bmatrix} \begin{bmatrix} l_{1x} & l_{1y} & l_{1z} \\ l_{2x} & l_{2y} & l_{2z} \\ l_{3x} & l_{3y} & l_{3z} \end{bmatrix}^{-1} \quad (2)$$

where $g_{123}$ corresponds to the gains for a triplet of loudspeakers, $p^T$ is the transpose of the point source vector and $L_{123}^{-1}$ is the inverse matrix of the same loudspeaker triplet. It is also important to point out that the loudspeaker gains are normalized. $l_1$, $l_2$ and $l_3$ are unit vectors that point in the $x, y$ or $z$ direction of loudspeaker 1, 2 and 3.

*C. Ambisonics*

Ambisonics [7] represents the surrounding sound field with a specific number of directive components (audio channels) carrying information about the physical properties of sound. It was first developed by Michael Gerzon in the early 1970s. Ambisonics is based on the expansion of the surround signal into spherical harmonics. The sound field information is encoded together with the sound source itself into a given number of audio channels, independent of the speaker layout. This representation of the sound field is the *B Format*.

*1) Encoding:* The *B Format* contains essentially a truncated spherical harmonic decomposition of the sound field. The Ambisonics channels are given by the expression:

$$B^{\sigma}_{mn}(t) = S(t)\, Y^{\sigma}_{mn}(\varphi_S, \delta_S) \quad (3)$$

where $Y^{\sigma}_{mn}(\varphi_S, \delta_S)$ are the coefficients, related by the direction of the sound source $S(t)$.

*2) Decoding:* Ambisonic decoding aims to acoustically recompose (or "re-encode") the encoded Ambisonics signals $B^{\sigma}_{mn}$. Signals $S_i$ are encoded as plane waves with coefficient vectors $c_i$, each representing a single direction of sound propagation. The encoded Ambisonics components with $N$ loudspeakers can be written as:

$$B = C \cdot S \quad (4)$$

where $C = [c_1 ... c_N]$ is the "re-encoding matrix" [8]. The re-encoding matrix has $K = (M + 1)^2$ rows and as many columns as loudspeakers in the setup [10]. The loudspeaker signals needed to reconstruct a particular Ambisonic-encoded sound field can be solved by inverting the linear system in (4) [8]:

$$S = C^{-1} \cdot B \quad (5)$$

where $C^{-1}$ is the inverse of $C$. This inversion can be done using the pseudoinverse matrix of $C$, typically defined as [8], [17]:

$$D = \mathrm{pinv}(C) = C^T.(C.C^T)^{-1} \quad (6)$$

however, for asymmetrical or irregular loudspeaker setups this decoding solution is often unstable and provides low spatial resolution [12]. For this reason, other decoding strategies were developed for irregular loudspeaker layouts.

A common decoding method is the all-round Ambisonic decoding method [12] (All-RAD). It involves two stages: first, a regular virtual loudspeaker layout and its decoding matrix $D$ is considered, and second, the signals of the virtual loudspeakers are mapped to the real loudspeaker by the gain matrix from VBAP.

The HO-DirAC decoding method [16] will essentially analyse the sound scene, infer the number of sound sources and isolate their signal, and then pan it, using the VBAP method for the loudspeaker setup.

Inserting imaginary loudspeakers in the vertical direction has proven to be a useful strategy to control the signal loss at missing directions in irregular loudspeaker setups [12]. When rendering the signals to the speaker setup, the imaginary loudspeaker's signals are dismissed, since these loudspeakers do not really exist. Different objective tests were conducted to evaluate how different decoding designs behaved, in terms of power distribution through a loudspeaker setup.

## III. Development Methodology

*A. Multichannel Loudspeaker System*

The loudspeaker array system was assembled on an semi-hemisphere of a geodesic dome shaped-infrastructure. The dome was installed in the acoustic anechoic chamber of Instituto Superior Técnico (IST), with dimensions of 5.30 m × 3.80 m × 2.70 m (length × width × height). Fig. 1 shows the geodesic dome with the loudspeakers attached.

Fig. 1. The physical implementation of the multichannel loudspeaker array.

To complete the setup, the loudspeakers were attached and suspended from the wooden edges of the infrastructure. The diameter of the wooden structure is approximately 3.40 m. The array is comprised of eight loudspeakers with positions defined based on a spherical coordinate system, as presented in Table I.

TABLE I
LOUDSPEAKERS' POSITIONS IN SPHERICAL COORDINATES IN DEGREES

| Loudspeaker | Azimuth | Elevation |
|---|---|---|
| 1 | -90 | 0 |
| 2 | -28.88 | 0 |
| 3 | 28.88 | 0 |
| 4 | 90 | 0 |
| 5 | 50.7 | 65 |
| 6 | -50.7 | 65 |
| 7 | 0 | 61 |
| 8 | 0 | 74 |

The four loudspeakers of the lower horizontal plane are positioned at a height of 1.20 m, at ear level when the subjects are seated. The three loudspeakers in the middle plane are in front of the subjects at a height of 1.85 m and the remaining loudspeaker is almost directly above the subject's head, at a height of 2.20 m.

*1) Hardware:* The eight loudspeakers vary in model and/or manufacturer. The M-Audio BX5 D2 is used in loudspeakers 1, 4, 7 and 8. In the center of the setup, the M-Audio BX8 are used in loudspeakers 2 and 3. For loudspeakers 5 and 6 the Behringer Studio 50 USB are used and they work in a bi-amping method, that essentially means that they use two channels of amplification to power a single speaker.

To transmit the audio and control the signals the Behringer ADA8200 Ultragain pre-amplifier module was used in combi- nation with the MiniDSP MHC Streamer multi-channel USB interface.

The loudspeakers are connected to the interface through XLR cables and are powered by IEC cables. In the percep- tual tests described in Section III-D, the AKG K512 MKII headphones were used.

### *B. Loudspeaker Calibration*

To calibrate the loudspeaker array, two types of calibration were performed: frequency calibration and level calibration.

To perform the frequency calibration, it is necessary to measure the frequency response of each loudspeaker. For this a test tone consisting of an exponential sweep with length of $2^{17}$ samples at a sampling frequency of 48 kHz between 20 Hz and 20 kHz was generated using Matlab [22]. The test tone was then played back over each loudspeaker and recorded using an omnidirectional measurement microphone, directly aimed at the loudspeaker and positioned in the same placed in the center of the array, where the subject would sit.

The loudspeaker level calibration was performed by playing a pink noise test tone. The sound amplitude was measured using a sound level meter (Bruel & Jaer, model 2260 Investi- gator), placed right in front of each loudspeaker, in the center of the array, in the same position as in the frequency calibration with the microphone.

*1) Results:* After measured, all loudspeakers were adjusted to 60 dB SPL (sound pressure level). Loudspeaker level was adjusted in the DAW (Digital Audio Workstation), in- suring all loudspeakers were at the same level. To ensure all the loudspeakers share the same frequency response at the center of the system frequency calibration was carried out. Commonly, loudspeaker frequency response shows linear distortion, meaning the output level is not constant through the whole frequency range. To ensure a flat frequency response, frequency compensation was carried out, by increasing the output level mainly in the low frequency region.

To improve the stability and accuracy of the audio signal, regularization was also carried out to smooth the response curve, thus preventing it from amplifying high-frequency noise or other unwanted artifacts in the audio signal.

The head tracking sensor, further discussed in Section III-F was also calibrated to ensure accurate and precise tracking of the subjects' position and orientation in space. The calibration was carried out in the Audio and Acoustics Laboratory at Instituto Superior de Engenharia de Lisboa (ISEL).

### *C. Software*

Sound reproduction and spatialization is achieved using an open-source VST (Virtual Studio Technology) audio plug-in suite, SPARTA (Spatial Audio Real-time Applications) [18], a collection of flexible VST audio plug-ins for spatial audio. All plug-ins are tested using REAPER [19].

*1) VBAP Decoder:* The Panner [18] allows the direction for up to 64 channels to be independently controlled, for both inputs and outputs. The inputs correspond to the virtual sound sources' positions in the sound space and the outputs correspond to the loudspeakers' positions.

*2) Ambisonics Encoder:* The AmbiENC [18] inputs multi- ple monophonic signals and spatially encodes them into spher- ical harmonics signals at specified directions. The spherical harmonics signals describe a synthetic anechoic sound-field, where the spatial resolution of the encoding is determined by the spherical harmonic order and the higher order the greater the spatial resolution.

*3) Ambisonics Decoders:* In theory, as explained in Section II, decoding the B-format signals for irregular loudspeaker arrangements is not a trivial task. For this reason, the four decoding methods previously mentioned were tested to assess which decoder works best in this particular loudspeaker ar- rangement. The tests aimed at understanding how the power was distributed amongst the loudspeakers for a virtual sound source positioned in the same place.

When dealing with irregular layouts in the decoding process, inserting one or more imaginary loudspeakers in the vertical direction has proven to be a useful strategy to control the signal loss at missing directions [12]. Imaginary loudspeaker are phantom loudspeakers that are created through mathematical algorithms to enhance the spatialization of sound sources. When rendering the signals to the speaker setup, the imaginary loudspeaker's signals are dismissed, since these loudspeakers do not really exist.

All the previously mentioned decoding methods were tested with and without imaginary loudspeakers. The virtual sound source was placed between loudspeakers 2 and 3 at ear level in all analytical tests.

In the All-Round Ambisonics Decoder [12] (All-RAD), when the imaginary loudspeaker is added, the sound source width is reduced and therefore the spatial resolution is im- proved. This is because the All-RAD decoder is signal- independent, meaning that it would need infinite encoding and decoding orders to achieve an optimal result.

On the other hand, adding an imaginary loudspeaker to the Higher Order Directional Audio Coding (HO-DirAC) [20] method does not influence the power distribution in the loudspeakers. The HO-DirAC decoding method is signal- dependent, which means that it will instead take into account the first-order sound scene. By doing this, the HO-DirAC method is able to estimate the amount of sound sources in that sound scene, evaluate their direction in the sound space and then pan the specific sound source signal, using VBAP, for the used loudspeaker setup.

The difference in the power between the different decoders is also worth pointing out. The All-RAD decoder produces a higher level of energy than the HO-DirAC decoder, meaning that even the loudspeakers that are at a greater distance from the virtual sound source produce more energy than the same loudspeakers in the HO-DirAC decoder. This leads to the conclusion that, in terms of spatial resolution, the HO-DirAC decoder outperforms the All-RAD even with an imaginary loudspeaker.

*4) Binauralizer:* The Binauraliser plug-in [20] convolves input audio (up to 64 channels) with interpolated HRTFs, obtained from the ARI database [20] in the frequency domain.

### *D. Perceptual Tests*

The listening tests consisted in the comparison of two spatialization techniques, FOA and VBAP, to evaluate the localization performance of virtual sound sources. The tests were comprised of three different scenaries: a binaural stage, a "shooting target" stage and a moving sound source stage.

In the binaural stage, FOA and VBAP are compared over headphones by synthesising the position of the real loud- speaker setup implemented in III-A. To do so, the subjects performed a localization task that consisted in reporting from which individual virtual loudspeaker where the perceived virtual sound source was originated.

In the "shooting target" stage the subjects' task was to report the perceived position of each virtual sound source, by rotating their body and pointing with their head towards the perceived sound source localization. The subjects were asked to close their eyes during the presentation of the sound. The virtual sound sources in relation to the loudspeaker setup are shown in Fig. 2.

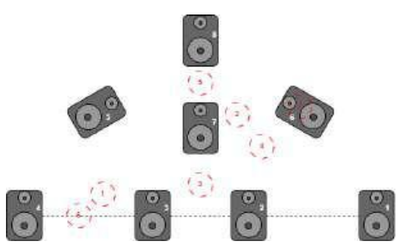

Fig. 2. Listening system and virtual sound source positions. Virtual sound source number 7 is positioned directly in loudspeaker 6.

The rotation of the subject's body and head is mapped by an head-tracking sensor, further discussed in Section III-F.

In the last stage of the perceptual tests, moving sound sources were used. The trajectory recognition was also eval- uated using an head-tracking sensor. However, in this stage, the subjects had to follow with their head and body the virtual sound source's trajectory. The trajectories chosen to evaluate the spatialization methods are represented in Fig. 3.

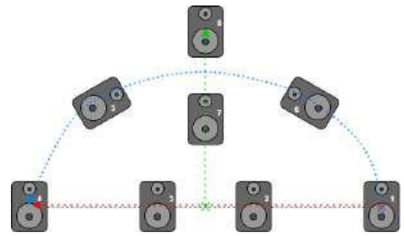

Fig. 3. Listening system and virtual sound source trajectories. There are 3 virtual sound source trajectories, each one represented in a different color (red, blue or green).

### *E. Subjects and Stimuli*

The study group consisted of 22 subjects (4 women and 18 men) aged between 19 and 58, 3 subjects had already performed perceptual tests. Three stimuli were used: bursts of pink noise, a recording of birds chirping and a saxophone recording.

### *F. Head tracking*

The monitorization of head movements is achieved using a Bluetooth head-tracking sensor, the WitMotion WT901BLECL [21].

### *G. Procedure*

In the binaural stage, prior to the beginning of the tests, subjects were asked to get familiar with the loudspeakers po- sitions around the setup. Then they put on heaphones and short bursts of pink noise (2 seconds long), were reproduced using VBAP and FOA from each individual virtual loudspeaker at random order.

In the second stage of the tests, subjects sat in a rotating chair in the center of the dome and in front of a reference mark located at head height. To calibrate the sensor, the subjects were asked to stay still, while staring at the reference mark, making this the reference position, with 0° elevation and 0° of azimuth. Following the calibration, the stimuli were presented. Each stimulus lasted five seconds and came from the differ- ent positions represented in Fig. 2. After hearing a stimulus, the subjects turned their bodies using the rotating chair and point with their nose, towards the perceived sound source location, while the sensor recorded the movement.For each method the the virtual sound source positions were presented randomly preventing the subjects to form habits that could affect the results.

Prior to the test itself, the subjects performed a training test, in which they were simply asked to turn their bodies and point with their nose to one individual loudspeaker at a time, to get familiar with the sensor and the pointing technique, preventing the results to be biased. The experiment was divided into 2 blocks, VBAP and FOA. In each block, the three stimuli were reproduced from the seven sound source positions at random order.

In the third and final stage of the perceptual tests the stimuli used was pink noise. The task was to track the moving sound source's trajectory by following the sound source's movement, by turning their bodies to the sound source, pointing with the nose and rotating the chair while the sound source moved around them.

## IV. Results

### A. Perceptual Tests

*1) Binaural results:* In FOA, the total number of misclassifications is 26 and the loudspeakers that were more confusing to the subjects were loudspeakers 3, 5 and 6. In VBAP the total number of misclassifications is 17 and the loudspeakers that produced more misclassifications were again loudspeakers 3 and 5.

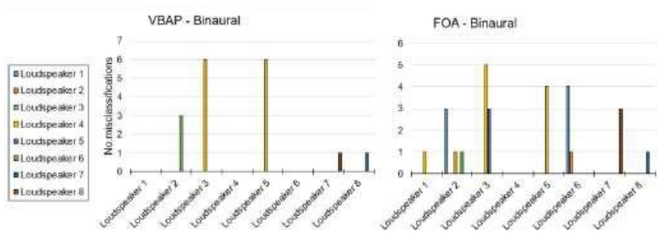


Fig. 4. The binaural results of VBAP and FOA.

Most misclassifications occurred in loudspeakers 3, 5, 6, 7 and 8, positioned in the center of the array. Loudspeakers 1 and 4, positioned on the right and left sides of the array, respectively were always correctly classified in both methods.

*2) Static sound source results:* In this stage, the subjects' task was to determine the location of seven different static sound sources in auditory space. To keep this section as concise as possible, solely two virtual sound source positions will be presented, due to the large volume of results gathered in the perceptual tests.

The evaluation metric used in this stage of the perceptual tests is the difference between the actual position of the sound source and the sound source position perceived by the listener, in both azimuth and elevation coordinates.

In Fig. 5, the individual responses are quite accurate but not precise, because despite all the results being close to the real value, the results remain scattered and therefore, not consistently close to each other. One particular case stands out, which is the results of the azimuth in FOA, where the difference in degrees between the subjects' answers and the real sound source location is low (less than 5 °) and the deviation is high.

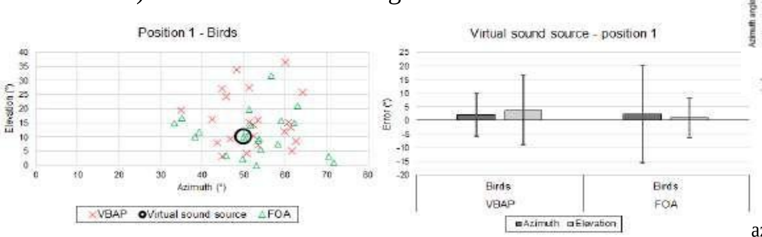


Fig. 5. Perceptual results for position 1. △ and × represent the results for FOA and VBAP, respectively and ○ corresponds to the sound source position. The bar plots display the average error and the whiskers represent the standard deviation for each case.

In Fig. 6, the individual test results are widely spread in relation to the real sound source coordinates. It is possible to observe from the subjects' individual test results, that no subject was able to perceive the sound source at ear level. This happened in both spatialization methods.

In the azimuth the subjects perceived the location more coherently, since the data points are less spread. However, the error remains high, since very few of the subjects correctly perceived the azimuth.

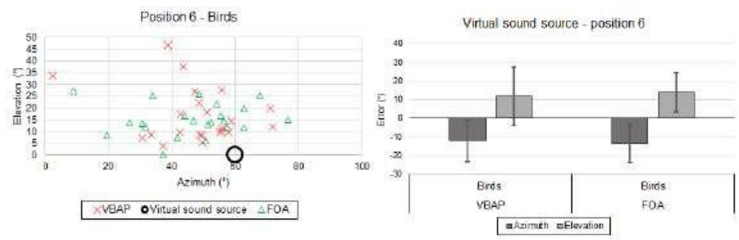


Fig. 6. Perceptual results for position 6. △ and × represent the results for FOA and VBAP, respectively and ○ corresponds to the sound source position. The bar plots display the average error and the whiskers represent the standard deviation for each case.

*3) Moving sound source results:* In this stage of the per- ceptual tests, the two spatialization techniques were compared using moving sound sources, that would travel different tra- jectories along the loudspeaker array. To better analyse and compare the results obtained from all 22 subjects, an average trajectory was calculated.

When comparing both results represented in Fig. 7, it can be seen that the azimuth of trajectory 1 in FOA, the data points are widely spread out, while in VBAP the data points are clustered together. This indicates that the subjects struggled to make sense of the path when the moving sound source was spatialized with FOA, thus forming a less perceptible pattern. With VBAP the subjects were able to perceive the trajectory with a greater sense of accuracy.

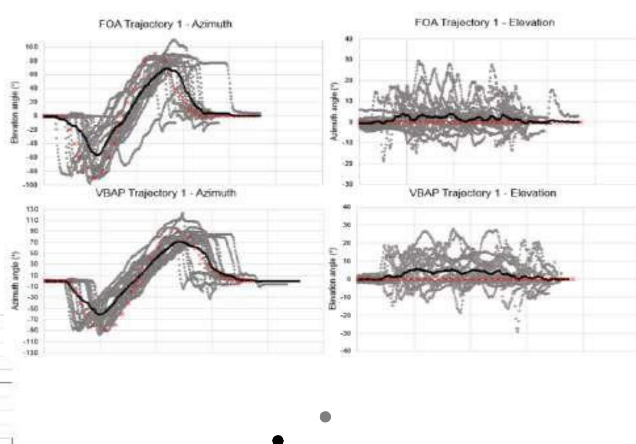


Fig. 7. Perceptual results for trajectory 1. The results are represented in terms of azimuth and elevation coordinates for both VBAP and FOA. Each plot represents three different datasets: represents the individual head-tracking sensor data from the subjects, corresponds to the average trajectory traveled by the subjects and × represents the moving sound source trajectory.

In VBAP the subjects were able to keep up with the moving sound source's trajectory, since the average trajectory almost does not lag behind the real sound source's trajectory. In FOA, the mismatch between the average trajectory and real sound source's trajectory is considerable especially when the sound source reaches 0° and 90°.

In terms of elevation, the results are similar. The reason why there still exist small differences between the average trajectory and real sound source's trajectory is due to natural unconscious head movements made while the subjects were following the moving sound source.

### B. Discussion

*1) Binaural results:* In the binaural tests, VBAP yielded better localization than FOA. Although the misclassification errors for both localization techniques often occur in the same loudspeakers, overall VBAP produced less misclassifications than FOA. In both techniques, the subjects often mistaken the loudspeakers positioned in the center of the array and not the loudspeakers located in the sides.

This indicates that the subjects' perception of the local- ization of virtual sound sources can vary depending on the position in space from where the sound source is presented. This is because of the binaural cues, the Interaural Time Differ- ence (ITD) and Interaural Level Difference (ILD), previously explained in Section II-A. The fact that misclassifications tended to occur more frequently in virtual loudspeakers closer to the center of the array results from the ambiguity of ITD and ILD, making the localization task more difficult.

*2) Static sound sources results:* In general, subjects expe- rienced more difficulty perceiving the location of the sound source in terms of elevation, as compared to azimuth, for both FOA and VBAP. This was observed in the majority of positions and stimuli, except in position 7, where the average error associated with the azimuth component was consistently higher than that associated with elevation. Position 7 is the only the virtual sound source positioned directly over loudspeaker 6, which sets it apart from the other positions where multiple loudspeakers are used to generate the virtual sound source.

As for the remaining 6 positions, 4 were rendered by 3 loudspeakers and 2 positions were generated by 2 loudspeakers in VBAP. The virtual sound source position in which the error for azimuth and elevation was lower was position 1 (rendered by 3 loudspeakers), in which VBAP slightly outperformed FOA in the azimuth, by yielding an average error around 0.5° for the azimuth, while FOA yielded 1.3°. In the elevation, FOA surpassed VBAP with an average error of 2.5°, in VBAP was 4°.

Position 6 registered the highest error in both azimuth and elevation. VBAP produced errors around of 9° in terms of elevation and 8° in terms of azimuth. FOA registered errors of approximately 10° in both azimuth and elevation. Despite the results not being strong, VBAP slightly outperformed FOA.

In this stage of the perceptual tests an intentional level of difficulty was introduced, by placing different sound sources close to each other to assess the subjects' ability to differenti- ate the various positions. This is depicted in positions 6 and 1, and positions 2 and 4. For the first pair of positions, it could be said that the perception of position 6 was often mistaken for position 1, due to the high elevation error. In the second pair, the subjects were able to differentiate the two positions. Further testing with different Ambisonics decoding methods and higher orders could be carried out using identical test procedures.

*3) Moving sound source results:* In all three trajectories in both azimuth and elevation the subjects were able to follow the general path of the sound source, but unable to accurately track its movement in terms of the azimuth or elevation coordinates. The elevation component of trajectories 2 and 3, for VBAP and FOA registered the greatest mismatch between the real sound source trajectory and the average trajectory perceived by the subjects. However, in trajectory 3 FOA became closer to the real sound source trajectory, while in trajectory 2 both techniques yielded similar results.

The delay between the subject and the moving sound source is also another important aspect. If the delay is long, it could indicate that the subject struggled to make sense as to where the sound source was located in that particular time.

Only trajectories 1 and 3 have a non-constant azimuth trajec- tory, since in trajectory 2 the sound source moves only in terms of elevation. The highest delay in trajectory 1 occurs when the sound source is rendered using FOA, and in trajectory 3 is using VBAP. Regarding elevation, trajectories 2 and 3 have a non-constant trajectory, since trajectory 1 moves only at ear level. In the two trajectories, there is no significant difference between FOA and VBAP, in terms of delay.

## V. Conclusion

In this study, VBAP and FOA techniques were compared in 3D space both through headphones and an 8 loudspeaker mul- tichannel system assembled for this effect. In the perceptual tests the virtual sound sources were either static or in motion in the acoustic space surrounding the subjects.

Despite not covering the entire sphere, the loudspeaker infrastructure provided sufficient coverage for the perceptual tests that were conducted.

The head-tracking sensor yielded results consistent with the subjects' perception of the sound source localization, whether it was static or moving.

Using headphones, VBAP showed better localization preci- sion than FOA, while in the multichannel loudspeaker array the perceptual tests using a static sound source, both tech- niques performed better in azimuth than in elevation. In the moving sound source test results, no significant differences were observed.

Overall, when comparing VBAP and FOA, the preferable choice would be VBAP considering its localization precision and ease of application.

## VI. Acknowledgements

I would like to express my gratitude to the Instituto Superior Técnico (IST) for providing access to the anechoic chamber, and to the Instituto Superior de Engenharia de Lisboa (ISEL) for the resources available at the Audio and Acoustics Labo- ratory.

## REFERENCES


[1]Grantham, D. W. (1995). Spatial hearing and related phenomena. In B. C. J. Moore (Ed.), Hearing (pp. 297–345). Academic Press. https://doi.org/10.1016/B978-012505626-7/50011-X

[2]Blauert, J., "Spatial Hearing. The Psychophysics of Human Sound Localization", MIT Press, 1997. https://doi.org/10.7551/mitpress/6391.001.0001

[3]Harris, J. D., "Localization of Sound: Theory and Applications", Ear and Hearing, 1982.

[4]Pulkki, V., "Communication Acoustics: An Introduction to Speech, Audio and Psychoacoustics", Wiley, 2015. https://doi.org/10.1002/9781119825449

[5]Pulkki, V., "Virtual Sound Source Positioning Using Vector Base Ampli- tude Panning", The Audio Engineering Society, 1997.

[6]Pulkki, V., "Spatial Sound Generation and Perception by Amplitude Panning Techniques", Helsinki University of Technology, 2001.

[7]Gerzon, M. A., "Periphony: With-height sound reproduction", Journal of the Audio Engineering Society, 1973.

[8]Daniel, J., "Further investigations of High-Order Ambisonics and Wave- field synthesis for Holophonic sound Imaging", Journal of the Audio Engineering Society, 2003.

[9]Morse, P., "Theoretical Acoustics", Princeton University Press, 1986.

[10]Daniel, J., "Spatial Sound Encoding Including Near Field Effect: Intro- ducing Distance Coding Filters and a Viable, New Ambisonic Format", Journal of the Audio Engineering Society, 2003.

[11]Marruffo, A. "A real-time encoding tool for Higher Order Ambisonics", 2014.

[12]Zotter, F., "Ambisonics: A Practical 3D Audio Theory for Recording, Studio Production, Sound Reinforcement, and Virtual Reality", 2019. https://doi.org/10.1007/978-3-030-17207-7

[13]Daniel, J., "Repr ésentation de champs acoustiques, application à la transmission et à la reproduction de scènes sonores complexes dans un contexte multimédia", 2000.

[14]Ward, D.B., "Reproduction of a plane-wave sound field using an array of Loudspeakers", IEEE Transactions on Speech and Audio Processing, 2001.

[15]Poletti, M., "Three-dimensional Surround Sound Systems Based on Spherical Harmonics", Journal of the Audio Engineering Society, 2005.

[16]Pulkki, V., "Directional Audio Coding in Spatial Sound reproduction and stereo upmixing", Journal of the Audio Engineering Society, 2006.

[17]Daniel, J., "Repr ésentation de champs acoustiques, application à la transmission et à la reproduction de scènes sonores complexes dans un contexte multimédia", 2000.

[18]McCormack, L., "SPARTA & COMPASS: Real-time implementations of linear and parametric spatial audio reproduction and processing methods", 2019.

[19] *Reaper*. [Online]. Available: https://www.reaper.fm/

[20] *ARI HRTF Database*. [Online]. Available: https://www.oeaw.ac.at/isf/das-institut/software/hrtf-database/

[21] *Head tracking sensor*. [Online]. Available: https://www.wit-motion.com/9-axis/witmotion-bluetooth-ble-5-0.html

[22] *Matlab*. [Online]. Available: https://www.mathworks.com/products/matlab.html