INTERAURAL CUES CARTOGRAPHY: LOCALIZATION CUES REPARTITION FOR THREE SPATIALIZATION METHODS

Eric Méaux and Sylvain Marchand,

L3i, University of La Rochelle, France
eric.meaux1@univ-lr.fr, sylvain.marchand@univ-lr.fr

ABSTRACT

The Synthetic Transaural Audio Rendering (STAR) method, first introduced at DAFx-06 then enhanced at DAFx-19, is a perceptive approach for sound spatialization aiming at reproducing the acoustic cues at the ears of the listener, using loudspeakers. To validate the method, several comparisons with state-of-the-art spatialization methods (VBAP and HOA) were conducted. Previously, quality comparisons with human subjects have been made, providing meaningful subjective results in real conditions. In this article an objective comparison is proposed, using acoustic cues error maps. The cartography enables us to study the spatialization effect in a 2D space, for a listening position within an audience, and thus not necessarily located at the center. Two approaches are conducted: the first simulates the binaural signals for a virtual KEMAR manikin, in ideal conditions and with a fine resolution; the second records these binaural signals using a real KEMAR manikin, providing real data with reverberation, though with a coarser resolution. In both cases the acoustic cues were derived from the binaural signals (either simulated or measured), and compared to the reference value taken at the center of the octophonic loudspeakers configuration. The obtained error maps display comforting results, our STAR method producing the smallest error for both simulated and experimental conditions.

1. INTRODUCTION

The main problematics of sound spatialization research projects are to find solutions able to convince a listener that a sound is coming from a very specific location within a room space. This research topic is a well-documented one, and efficient methods have already been proposed in the literature such as Vector Base Amplitude Panning (VBAP) proposed by Pulkki [1], Ambisonics proposed by Gerzon [2] and generalized to higher orders (Higher-Order Ambisonics or HOA) by Daniel [3], and the Synthetic Transaural Audio Rendering (STAR) method [4] we proposed. If these methods are based on different premisses, they all aim at recreating a believable 3D sound to be interpreted by our brain.

Indeed, the human brain uses perceptive cues for localizing sources [5], which are mainly the Interaural Level Difference (ILD) and the Interaural Time Difference (ITD). Experimental tests have been conducted in 2015 and 2019 to confront the different methods in real conditions [4], and differences were exacerbated (such as the continuity of HOA and STAR); but the three methods are very similar. This article also focuses on comparing these methods, but in another way. Unlike the previous tests, this paper introduces an objective comparison based on perceptive cues obtained using simulation or recorded via a KEMAR manikin, instead of human subjects. This paper will focus only on the ILD error, since the ITD measures are ambiguous by nature (obtained by a phase difference known only modulo 2π). Furthermore, the comparison is computed on cartography error, which enables the simulation under real diffusion conditions: in a place where an audience can be spatially dispersed.

The remainder of this article is organized as follows: Section 2 describes the methodology used to obtain the cartography, Section 3 shows the ILD error maps obtained, and Section 4 presents some perspectives before concluding the article.

2. METHODOLOGY

In this paper we will stay in the horizontal plane and focus on the azimuth of the sound source.

2.1. Acoustic Cues

Human listeners use acoustic cues [6] for sound source localization, binaural cues such as the Interaural Level Differences (ILDs) and Interaural Time Differences (ITDs) [5] being the most important for the estimation of the azimuth. They measure the level or time differences between the two ears when a sound wave travels from some position in space (with a certain azimuth, elevation, and distance). For ILD, the greater the difference between the two ears, the more lateralized the sound will be perceived. However, in practice ITDs are measured through Interaural Phase Differences (IPDs), which are angular differences determined in radians only up to a modulo 2π factor. Thus the computation of the ITDs has to deal with this ambiguity. For this reason, in the present study we will focus only on the ILDs, which can be computed in a non-ambiguous way from the binaural signals using Equation 1:

$$\mathrm{ILD}(f) = 20 \log_{10}(|S_r(f)/S_l(f)|) \tag{1}$$

where S_l and S_r are the spectra of the signals corresponding to the left and right ears, respectively.

The main contribution of the present work is to verify, by simulations and measures, if the ILDs produced by several spatialization methods (including ours [4]) are close the one expected (using the KEMAR with large pinnae of the CIPIC database [7] as a reference).

2.2. Spatialization Methods

Different spatialization methods exist to date, and in this paper we consider three of them working with loudspeakers (and not headphones), since we are interested in concerts with large audiences.

Two of these methods are from the state of the art, namely Vector Base Amplitude Panning (VBAP) [1] and Higher-Order Ambisonics (HOA) [2, 3]. The third method we want to study is Synthetic Transaural Audio Rendering (STAR), a method introduced by Mouba [8] and improved since then [4].

These methods follow different approaches. VBAP performs a geometrical interpolation in order to reconstruct the sound wave. HOA aims at reconstructing the sound field at the center of the loudspeaker array, where the listener has to be. STAR attempts to recreate the acoustic cues at the ears of the listener.

Since we focus on the azimuth and stay in the horizontal plane, the methods are in fact in 2D. The consequence for VBAP is that only 2 loudspeakers are used for one source. And the consequence for HOA is that the spherical harmonics reduce to classic Fourier harmonics (the polynomial parts of the spherical harmonics being for the elevation, which will be 0 in this study). We used a direct implementation for HOA rendering, without optimization.

These methods, contrary to the ones based on headphones, involve multiple loudspeakers and the resulting configuration, plus the position of the listener, may strongly impact sound reproduction and perception. The goal of the present study is to characterize to which extent the acoustic cues are respected. For that purpose, an ILD cartography will be done for each method, and different listener positions, for comparison. More precisely, Section 3 will show the difference between measured and reference ILDs. These reference ILDs will be obtained using Equation 1 on binaural signals.

2.3. Binaural Rendering

In order to calculate the ILD it is essential to know the sound signal which arrives at each ear. Binaural rendering is a technique which allows this signal to be produced in order to send it to headphones, and which thus also makes it possible to calculate the acoustic cues. Several techniques have been proposed to render a binaural signal from a multi-channel one, especially in the case of HOA. However, for equity sake, we want to use the same rendering technique for every spatialization method. Thus, we chose the simplest one, considering the acoustic paths between each loudspeaker and the left and right ears (see Figure 1). These paths are in theory the Head-Related Impulse Responses (HRIRs) for a source placed at the loudspeaker position. Thus, for an octophonic configuration consisting of $N = 8$ loudspeakers, we have:

$$s_{l,r} = \sum_{n=1}^{N} \mathrm{HRIR}_{l,r}(\theta_n) * s_n \qquad (2)$$

where $s_{l,r}$ represents the left or right binaural signal, $\mathrm{HRIR}_{l,r}(\theta)$ is the left or right HRIR for a given azimuth θ, s_n is the signal played by loudspeaker number n, and $*$ denotes the convolution. Moreover, for loudspeakers regularly placed on a circle, we have $\theta_n = (n-1) \cdot 2\pi/N$ (radians). In practice for the HRIRs we use the CIPIC database [7] without interpolation.

3. ILD CARTOGRAPHY

In this section we compare the ILD error for different spatialization methods and listener positions. The ILD error is the difference between the ILDs under consideration (coming either from simulations or measurements) and the reference ILD, coming from the KEMAR manikin with large pinnae of the CIPIC database [7] supposed at the center. In theory, the smaller the ILD error, the more accurate the method. For a given configuration (source, loudspeakers, listener), we consider the mean of all ILD errors over time and frequency (using the absolute values).

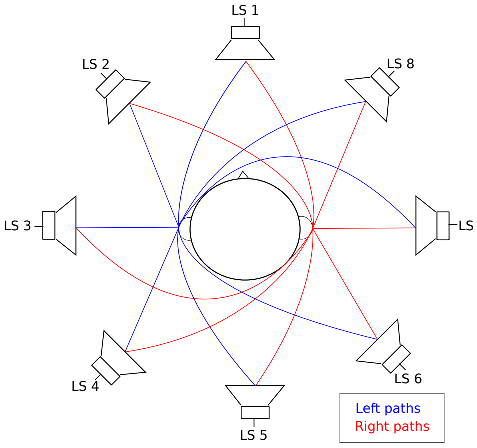

Figure 1: *Acoustic paths for the octophonic configuration.*

For both simulation and measurements, we use an octophonic setup for the loudspeakers, a white noise of duration 6s located at 30° for the sound source, and a KEMAR manikin with large pinnae (either virtual or real) to emulate the listener. The 30° azimuth was chosen because then the source lies in between two loudspeakers, but not in the exact middle.

The reference ILD is calculated with Equation 1 from the ideal binaural signals obtained by simple convolution of the source signal (white noise) with the pair of HRIRs from the CIPIC database for the desired azimuth (30°).

For this computation of the ILDs, Equation 1 is fed with a (pair of) Short-Time Fourier Transform(s) with a Hann window, 50% overlap, and size 2048 points, for (binaural) sounds recorded at a sampling rate of 44100Hz.

3.1. Simulation Process

In the case of simulation, the spatialization methods are run *in silico*, and the output of the (virtual) loudspeakers are rendered into binaural signals using Equation 2. Then the ILDs are calculated in turn using Equation 1.

To consider several positions of the listener, in the simulations we use a 100×100 grid covering a surface of 1 square meter and try each position on this grid for the (virtual) KEMAR manikin, always looking forward (towards azimuth 0). The sound source is located at its azimuth (here 30°), then for each grid position the relative angle is used for each spatialization method (see Figure 2).

3.2. Simulation Results

The produced maps show the error for all real positions, the X and Y axes represent the reference position (0 0 being the center of the speakers).

Figures 3, 4, and 5 show the mean ILD error cartography for the spatialization methods under consideration, with the 100×100

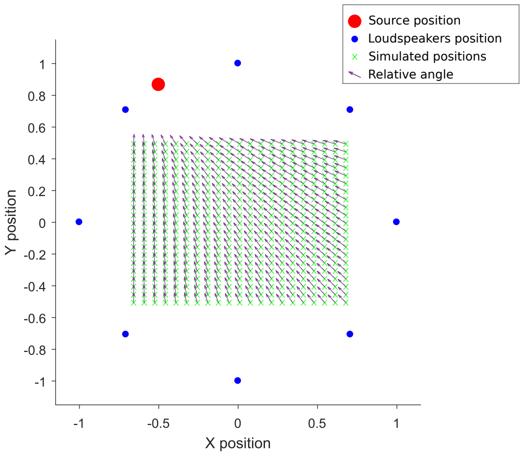

Figure 2: *The mesh used for the simulation.*

grid for the listener position, a regular octophonic loudspeaker display, and a sound source at azimuth 30°.

On these maps, we first see that HOA is rather chaotic (thus very dependent of the listener position), while VBAP appears to be the smoothest, STAR being intermediate. We also see a diagonal effect (dark line oriented roughly between the sound source and center), which is normal since the displacement on that lane affects only the distance, and not the relative azimuth which impacts the ILD. Last but not least, for all methods it appears that the center exhibits only small errors (approaching 0, plotted in dark).

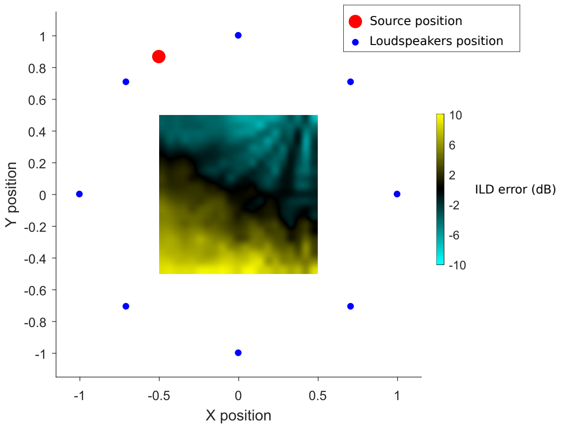

Figure 3: *ILD cartography simulated using HOA spatialization method, octophonic display, and source at* 30°.

Figure 6 shows the evolution of the three methods for the $[10; 160]^\circ$ range (by increments of 10°), and confirms the previous observations. VBAP appears to have more brutal changes than the other methods as the azimuth varies (which has already been observed in our DAFx-19 tests [4]).

3.3. Experiments

3.3.1. Experimental Process

The measurements have been performed at the *Studio de Création et de Recherche en Informatique et Musiques Expérimentales* (SCRIME), University of Bordeaux, France. This studio is used by musicians and has quite good acoustics, even if it is not physically controlled. There, 18 Genelec 8030 loudspeakers are mounted on three loudspeakers rings. The studio has a surface of 40 square meters. One wall has three windows, two have a wooden door, and some acoustic panels are disposed against them. The floor is covered with a thin carpet. For the source signal a white noise of 6s length, sampled at 44.1kHz was spatialized at 30° with the three different spatialization methods using an octophonic loudspeaker ring of diameter 2.6m at the same horizontal level than the KEMAR with large pinnae manikin. The manikin was moved in the studio on markers placed beforehand to produce a 5×5 grid covering a surface of 4 square meters (see Figure 7). That mimics the position of each listener in some audience around the center. Figure 8 shows this audience of KEMAR manikins, the center displayed in bold. The KEMAR manikin was mounted on an office chair support, which provided simple rotation and displacement.

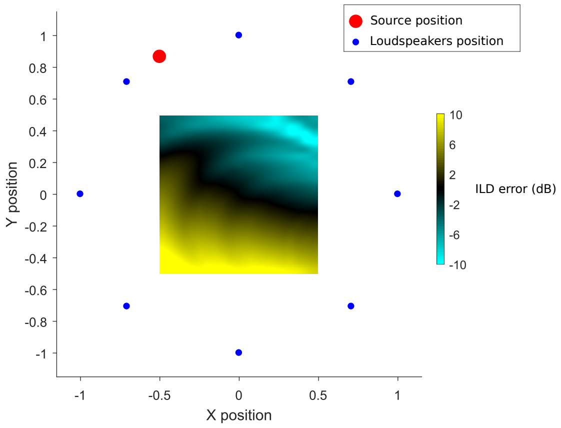

Figure 4: *ILD cartography simulated using STAR spatialization method, octophonic display, and source at* 30°.

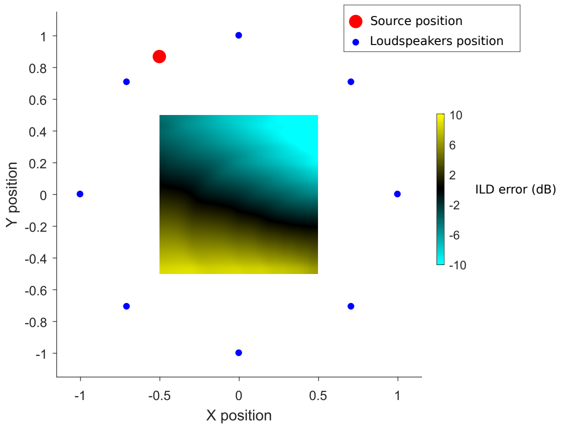

Figure 5: *ILD cartography simulated using VBAP spatialization method, octophonic display, and source at* 30°.

3.3.2. Experimental Results

Although the experimental cartography tends to confirm the observations done on the simulation, reverberation seems to have an effect comparable to the spatialization methods themselves. Moreover the mesh being only of dimensions 5×5 for the experiments to be tractable, the resolution is much lower than for the simulation. Since the graphics on Figures 9, 10 and 11 are interpolated, we provide the data tables corresponding to the measurements (Tables 1, 2 and 3).

Y\X	-1	-0.5	0	0.5	1
-1	4.2	-3.1	0.5	-7.6	-3.1
-0.5	-6.2	-1.8	-1.5	-12.2	-2.9
0	-3.9	-0.5	**0.8**	-9.6	-4.4
0.5	-0.7	-6.6	0.4	-6.6	-2.3
1	-0.5	-0.6	-1.9	-5.1	-3.7

Table 1: *ILD error (in dB) using HOA spatialization method, with source at azimuth* 30°*, and a* 5×5 *grid for the position of the listener (central position in bold).*

Y\X	-1	-0.5	0	0.5	1
-1	-1.1	0.3	6.2	-0.3	6.6
-0.5	-4.0	5.9	0.5	1.8	5.2
0	-4.1	2.4	**-0.3**	8.3	3.1
0.5	6.4	0.1	7.4	11.4	2.9
1	3.9	0.5	-0.5	9.4	4.5

Table 2: *ILD error (in dB) using STAR spatialization method, with source at azimuth* 30°*, and a* 5×5 *grid for the position of the listener (central position in bold).*

Table 4 resumes the different values for the central position. It appears clearly that STAR has better ILD reconstruction (lower error) than the two other methods, for both simulated and experimental situations.

While VBAP and HOA should reconstruct the sound wave or field at the center of the display, they do not take into account the presence of the head of the listener. STAR does. We have also investigated ILDs only for lower frequencies, and tried different HOA decoders, but the observations above were still valid.

Y\X	-1	-0.5	0	0.5	1
-1	4.0	-6.0	-0.5	-1.9	-5.6
-0.5	4.2	-3.2	0.6	-6.9	-3.1
0	-6.0	-0.0	**-1.8**	-12.3	-2.9
0.5	-4.1	-0.5	0.4	-9.2	-4.3
1	-1	-6.5	0.3	-6.6	-2.3

Table 3: *ILD error (in dB) using VBAP spatialization method, with source at azimuth* 30°*, and a* 5×5 *grid for the position of the listener (central position in bold).*

	HOA	STAR	VBAP
Simulated	0.7	**0.0**	-2.4
Experimental	0.8	**-0.3**	-1.8

Table 4: *ILD error at the center for simulated and experimental situations, with different spatialization methods, and a source at azimuth* 30°.

4. CONCLUSION

In this paper, we introduced the perceptive cues cartography (here limited to the ILD). This cartography represents an objective measure for a perceptive evaluation of sound spatialization methods. Most of the time, only the central spot is considered to characterize these methods. Having a map is particularly interesting since it mimics listeners distributed in a single room. In this article, two approaches have been conducted: a simulation first, then an experimentation. The two cartographic results agree, even if it remains difficult to fully compare them due to the different resolutions (for the experiments to be tractable in practice). On the experimental approach, the result are very noisy. We can make the assumption that, under real conditions, room reverberation takes precedence over the spatialization methods themselves. Whether on simulated or real tests, it appears than the STAR method reproduces more faithfully the expected acoustic cues (notably at the center, where the error has been minimized at the design of the reference system). Another observation is about the disturbance of the STAR method between the chaotic aspect of HOA and the smooth aspect of VBAP visible on the maps. These two results paves the way for further developing the STAR spatialization method, and extend it to distance and elevation, in order to generate a complete 3D sound system with a perceptive approach.

5. ACKNOWLEDGMENTS

The authors would like to thank the SCRIME *(Studio de Création et de Recherche en Informatique et Musiques Expérimentales)*, University of Bordeaux, France, which supplied the studio and materials for experiments.

6. REFERENCES

[1] Ville Pulkki, “Virtual Sound Source Positioning using Vector Base Amplitude Panning,” *Journal of the Acoustical Society of America*, vol. 45, no. 6, pp. 456–466, 1997.

[2] Michael A. Gerzon, “Periphony: With-height sound reproduction,” *Journal of the Audio Engineering Society*, vol. 21, no. 1, pp. 2–10, 1973.

[3] Jérôme Daniel, *"Représentation de champs acoustiques, application à la transmission et à la reproduction de scènes sonores complexes dans un contexte multimédia"*, Ph.D. thesis, Université Paris 6, 2001.

[4] Eric Méaux and Sylvain Marchand, “Synthetic Transaural Audio Rendering (STAR): a Perceptive Approach for Sound Spatialization,” in *Proceedings DAFx*, Birmingham, United Kingdom, September 2019.

[5] John W. Strutt (Lord Rayleigh), “On Our Perception of Sound Direction,” *Philosophical Magazine*, vol. 13, no. 6, pp. 214–302, 1907.

[6] Jens Blauert, *"Spatial Hearing"*, MIT Press, Cambridge, Massachusetts, revised edition, 1997, Translation by J. S. Allen.

[7] V. Ralph Algazi, Richard O. Duda, Dennis M. Thompson, and Carlos Avendano, “The CIPIC HRTF Database,” in *Proc. of the IEEE Workshop on Applications of Signal Processing to Audio and Acoustics (WASPAA)*, New Paltz, New York, 2001, pp. 99–102.

[8] Joan Mouba and Sylvain Marchand, “A Source Localization/Separation/Respatialization System Based on Unsupervised Classification of Interaural Cues,” in *Proceedings DAFx*, Montreal, Quebec, Canada, September 2006, pp. 233–238.

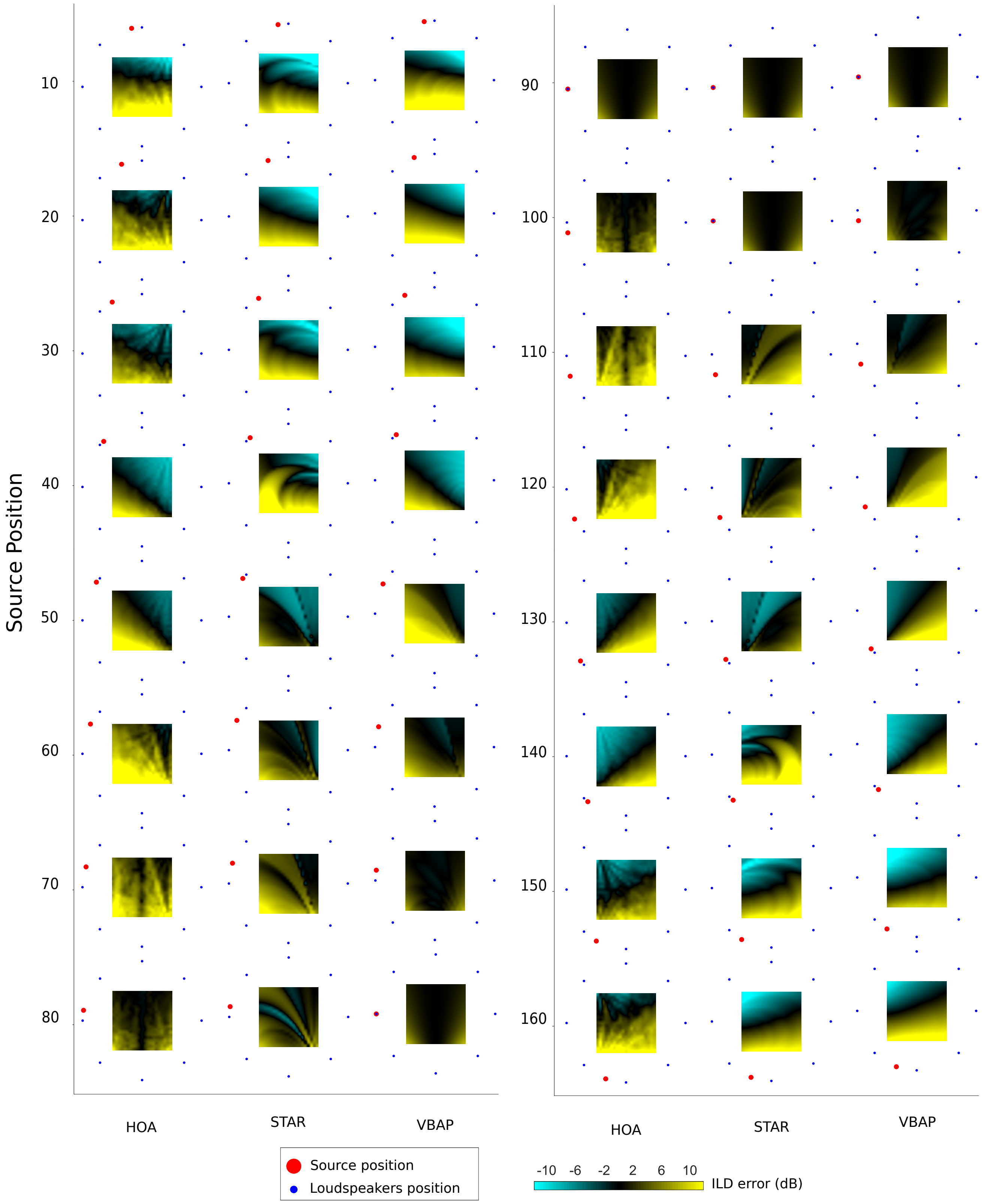

Figure 6: *ILD error cartography for all spatialization methods and different source positions (azimuth from* 10° *to* 160° *by steps of* 10°*).*

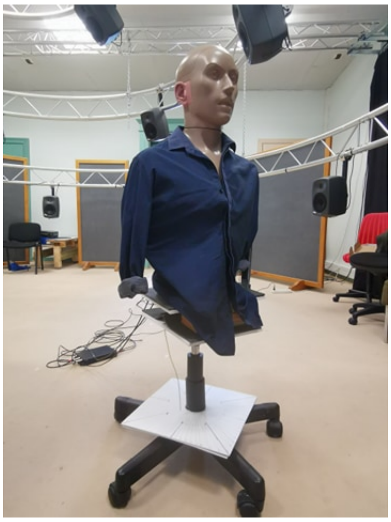

Figure 7: *Experimental recording setup with the KEMAR manikin. The manikin was moved into the space intended for listeners in order to produce the cartography*

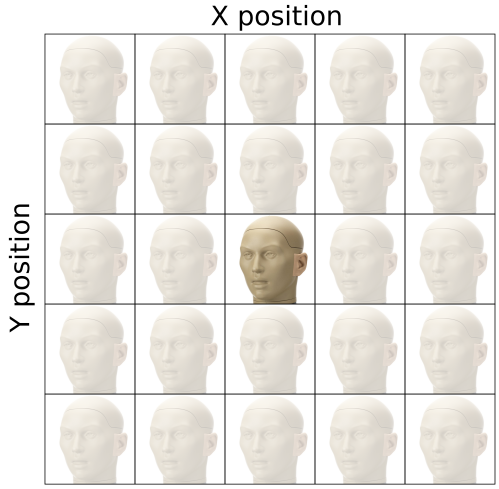

Figure 8: *Audience of 25 KEMAR manikins, the central position being in bold.*

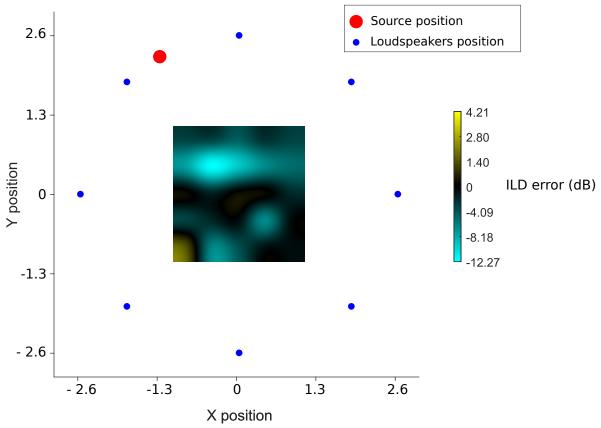

Figure 9: *ILD error cartography measured using HOA spatialization method, octophonic display, and source at* 30°.

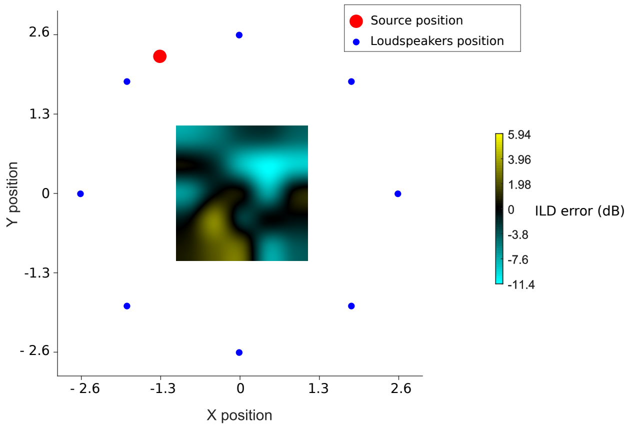

Figure 10: *ILD error cartography measured using STAR spatialization method, octophonic display, and source at* 30°.

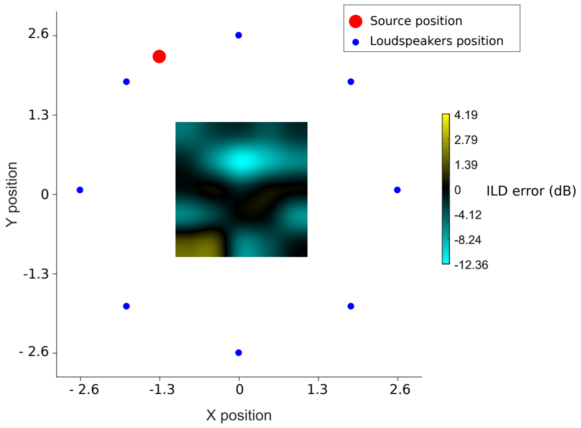

Figure 11: *ILD error cartography measured using VBAP spatialization method, octophonic display, and source at* 30°.