Audio Engineering Society

Convention Paper 8601

Presented at the 132nd Convention
2012 April 26–29 Budapest, Hungary

Microphone Array Design for localisation with elevation cues

Michael WILLIAMS[1]

[1] 'Sounds of Scotland', Le Perreux sur Marne, France
mike@soundsscot.com
www.mmad.info

ABSTRACT

Analysis of the HRTF characteristics with respect to both azimut and elevation localisation cues would seem to suggest that, whilst inter-aural time difference and inter-aural level difference information give strong azimut localisation cues, we only have spectral variations in the vertical plane to generate localisation cues with respect to elevation. Of course positioning of a 2nd layer of loudspeakers above the horizontal reference plane will already introduce listener spectral differences relative to the horizontal plane reproduction, but the microphone array that feeds this 2nd layer must not generate information that will be in conflict with the localisation cues already generated in the horizontal plane of the 1st layer or main array. In the event of Time Difference and Level Difference information being generated between microphones in both layers, localisation characteristics must be considered as projected onto the main horizontal plane of localisation information.

1. THE PSYCHOACOUSTICS

Many of the microphone array systems that are being proposed for sound recording in the upper hemisphere of the surround sound field, assume that the same parameters that generate localisation information in the horizontal reference plane, will also produce elevation information. This is unfortunately far from being the case. However very little actual psychoacoustic information has been published with respect to the detection of localisation in elevation when using various loudspeaker configurations. On the other hand analysis of the HRTF database characteristics with respect to both azimut and elevation localisation cues shows that, whilst inter-aural time difference and inter-aural level difference information give strong azimut localisation cues, we only have a set of complex spectral variations in the vertical plane that can generate localisation cues with respect to elevation, which when 'interpreted' by the listener are inherently very weak localisation cues. Figures 1 & 2 show the spherical ISO Interaural Level Difference (ILD) and ISO Interaural Time Difference (ITD) curves measured over a 1.5kHz to 10kHz energy band. These curves were published in 2009 by P.Guillon[1] in his PhD thesis at the University of Maine in France.

The Orange Labs HRTF database that was used to generate these diagrams is not available to the general public but reference to the CIPIC database[2] of HRTF curves will confirm these measurements.

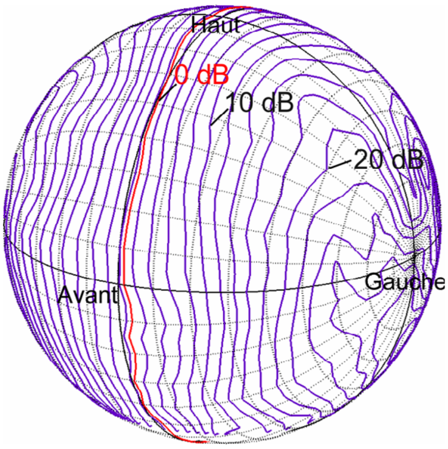

FIGURE 1 – VARIATION OF ILD ON A SPHERE, BASED ON HRTF (FROM THE ORANGE LABS PRIVATE DATABASE) FOR ENERGY WITHIN THE 1.5KHZ TO 10KHZ ENERGY BAND.

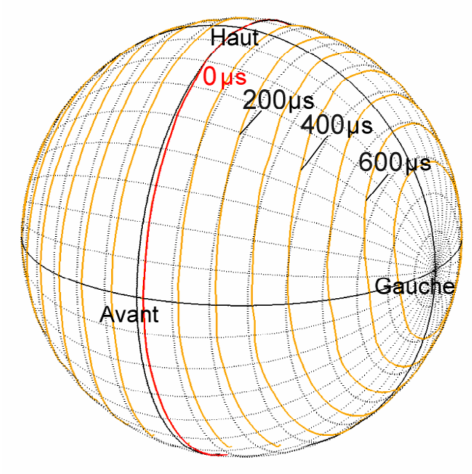

FIGURE 2 – VARIATION OF ITD ON A SPHERE, BASED ON HRTF (FROM THE ORANGE LABS PRIVATE DATABASE) FOR ENERGY WITHIN THE 1.5KHZ TO 10KHZ ENERGY BAND.

If we examine in more detail the Interaural Level Differences for the whole of the audible frequency band then we can see from the examples in Figures 3 & 4 that for a given azimut of 30° and 4 elevation angles, the situation is much more complex.

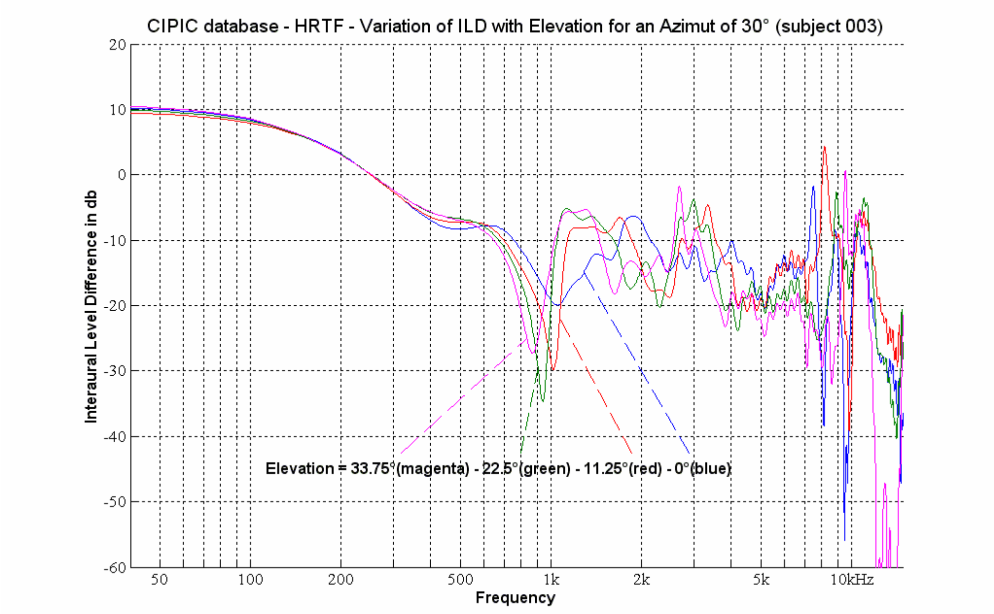

FIGURE 3 – ILD VARIATIONS FOR SUBJECT 003 (CIPIC DATABASE)[2]
FOR AN AZIMUT OF 30° AND VARIOUS ELEVATION ANGLES

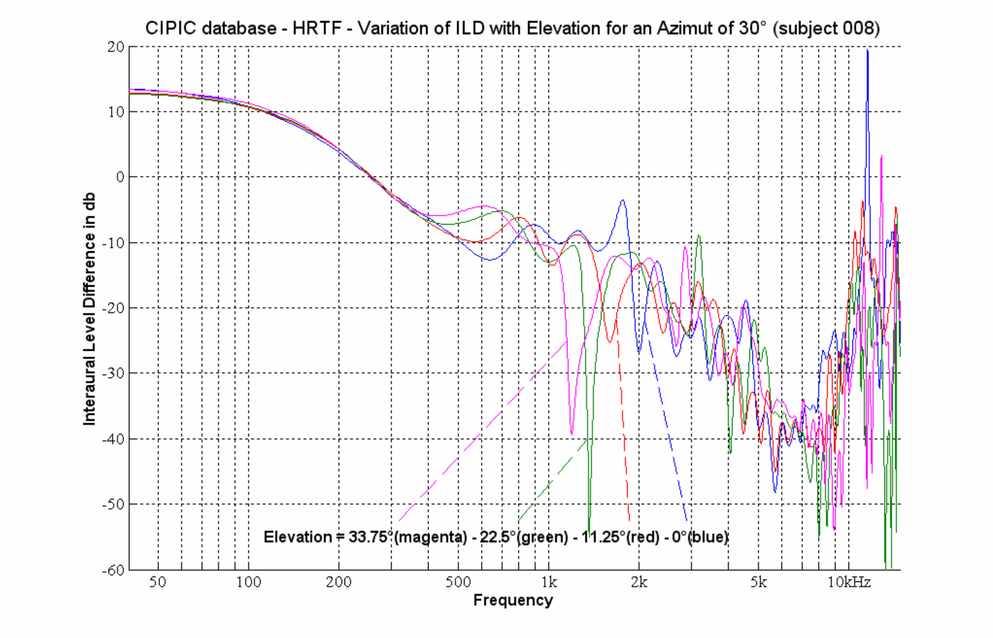

FIGURE 4 – ILD VARIATIONS FOR SUBJECT 008 (CIPIC DATABASE))[2]
FOR AN AZIMUT OF 30° AND VARIOUS ELEVATION ANGLES

Whilst the overall energy level may vary but little from one elevation to another, the 'devil is in the detail' - there are complex variations generating both negative and positive ILD variations due to the individual morphology of the pinna of each specific person. This explains why, in Binaural Technology, there is so much attention paid to matching the specific pinna characteristics for each individual listener, but it also shows why these variations are totally impossible to recreate in standard loudspeaker reproduction.

2. LOUDSPEAKER REPRODUCTION

Of course it is well known that when listening to loudspeaker stereophony we need to about double the level and time difference to obtain the equivalent localisation effect with respect to binaural localisation cues. However the weak and complex spectral cues needed for elevation localisation are, to all intents and purposes, impossible to generate with standard loudspeaker stereophony. Even artificial head recording or similar techniques still do not overcome the loss of elevation information when passing from binaural listening to loudspeaker reproduction. There is therefore no reason why we should expect to be able to directly recuperate the effect of elevation in a normal multichannel recording when using a single plane of microphones as an array. In fact we can see clearly that there is NO general or progressive variation of ILD or ITD with respect to elevation that can be used to create upper hemisphere sound localisation for any form of surround sound array.

2.1. Generating spectral differences with elevation

Positioning of a second set of loudspeakers some distance above the horizontal reference plane will of course naturally introduce spectral differences – the listener is by definition placed in the horizontal plane of loudspeaker reproduction, and the second set of loudspeakers being placed at a certain elevation angle, will be perceived by the listener as having a different spectral balance, and this spectral differences will of course be specific to the listener.

We can consider this second set of loudspeakers as an additional plane or layer of reproduction. This in no way implies that there is a continuous and realistic reproduction of elevation characteristics in the upper sphere of the surround sound field, it is just a second layer of reproduction. The crucial question that arises is 'what is the interaction between the two layers?'

- MINIMAL interaction if the array is so designed
- CONSIDERABLE interaction if a high level of the same sound field is picked up by both layers.

2.1.1 MINIMAL Interaction

After some laboratory experimental testing, this type of array structure was used on a trial recording in January 2012 during a series of recordings in London with the London Sinfonietta, in collaboration with Neu Records (Barcelona).

The main layer was based on a 7 channel Hypocardioid MAGIC array[3] in the horizontal reference plane, whilst the elevation plane was covered either with 4 figure of eights or 4 supercardioids spaced at 55cm and orientated vertically i.e. at 90° to the horizontal reference plane. This array configuration is shown in Figures 5, 6, 7 & 8..

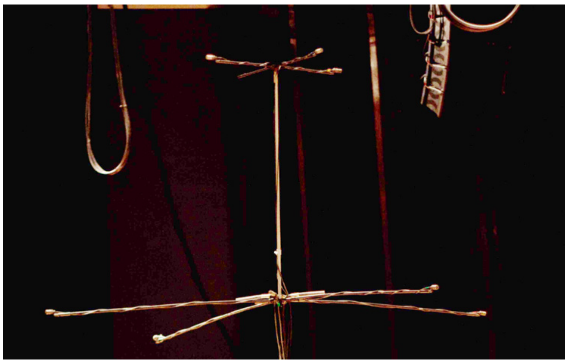

FIGURE 5 – MAGIC ARRAY + 2ND LAYER ARRAY FOR ELEVATION

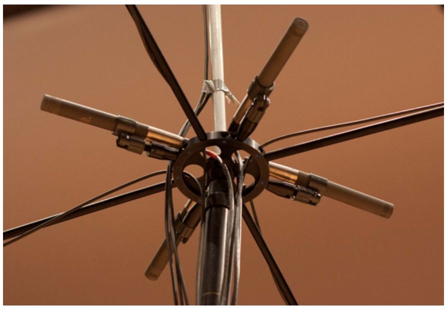

FIGURE 6 – THE 'WILLIAMS HYPOCARDIOID CROSS' - CENTRE OF THE HUB OF THE MAGIC ARRAY -

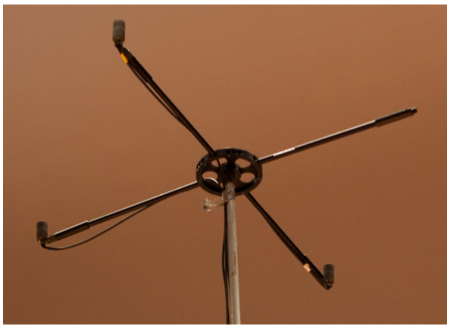

FIGURE 7 – 2ND LAYER ARRAY (3 MICROPHONES) FIGURE OF EIGHT ELEVATION ARRAY DIRECTIVITY AXIS IS AT 90° TO BODY OF MICROPHONE

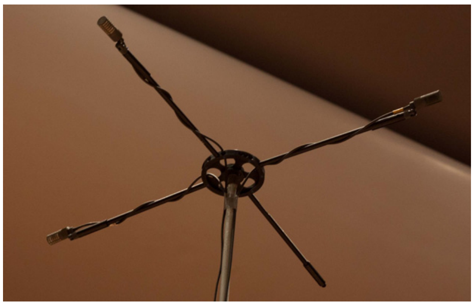

FIGURE 8 – 2ND LAYER (3 MICROPHONES) SUPERCARDIOID ELEVATION ARRAY

The details of the 7 channel Hypocardioid MAGIC array have been described in AES Preprints 7057[3] and 7480[4]. Basically it is made up of a 4 channel Hypocardioid 'Williams' cross (90°/35cm) (as in Figure 6) as described in 1991 in Preprint 3157(5), plus 3 satellite microphones positioned at 0°, 90° and 270° at a distance of 1m50 from the centre (as in Figure 5).

To this is added the 2nd Layer array either with Figure of Eight microphones (as in Figure 7) or with Supercardioid microphones (as in Figure 8). In this particular recording session the position of the musicians meant that the second solution with the Supercardioid microphones for the 2nd layer, produced better results as the 'zero' in the directivity pattern corresponded almost exactly with the musical sound sources. There was therefore a very much better separation between the elevation array which picked up almost exclusively the upper reverberation field, whereas the MAGIC array picked up the direct sound sources.

2.1.2. CONSIDERABLE Interaction

We have now introduced a further complication in that we have multiple interactions between the different elements that make up the total microphone array system. In general we can say that the microphone array that feeds the elevation set of loudspeakers (the 2nd layer) must not generate information that will be in conflict with the localisation cues already generated in the loudspeakers making up the horizontal plane information (1st layer). And cross catchment generation of localisation information between the two microphone array planes must also be either reduced to negligible levels, or designed not to conflict especially with the localisation cues generated by the 1st layer array.

The first and simplest criteria to consider in design, is that any coverage information, developed by each of the planes of array catchment or pickup, must be in the same direction and the same angular range (that is of course when both arrays have catchment of direct sound source signals). For instance if a specific direct sound source generates, via the 1st layer microphone array, localisation information in the horizontal reference plane of loudspeakers, then the 2nd layer microphone array covering the elevation must produce localisation in the same direction when reproduced by the elevation loudspeakers. In other words the coverage angles and orientation must be as near the same as possible.

The second problem of cross array catchment is much more difficult to solve in that any time difference and level difference information generated between a microphone in the 1st layer and adjacent microphones in the 2nd layer, must be considered as projected onto the horizontal plane of localisation information. The elemental segment for this phenomena situated in three dimensional space must be considered as the contribution of three adjacent microphones in a triangular configuration. The optimum configuration is by necessity an isosceles triangle, with the base of the triangle being in the horizontal plane (1st layer) and generated by two adjacent microphones in the main horizontal array plane. This also applies to the disposition of the reproduction loudspeakers.

The worst possible configuration for the two loudspeaker configurations is when they are mounted on the same support gantry and therefore are mounted one above the other. If each loudspeaker is mounted on its own independent gantry then the 2nd layer loudspeakers can be positioned so that they form the all important isosceles triangle with respect to the 1st layer of loudspeakers, and also mirroring the microphone array construction.

3. FURTHER RESEARCH

In the case of considerable interaction between each layer of reproduction, further research needs to be carried out on the individual contributions from each specific element or microphone pair. A period of about two weeks has been set aside at the University of Gothenburg[6] in August 2012, in order to carry out a systematic analysis of all the different dimensions of interference. In particular the effect of possible superimposition of images, and the role of the precedence effect in the rejection of unwanted localisation information.

4. ACKNOWLEDGEMENTS

I would like to thank Hugo Romano and Santi Barguño from Neu Records (www.neurecords.com) in Barcelona for permission to record the session of contemporary music by Ramon Humet, performed by the London Sinfonietta in the Watford Colosseum in London in January 2012. Neu Records will be publishing their own recording of this session.

I also would like to thank Rozenn Nicol (Orange Labs, France) for the valuable information published in her AES Monograph 'Binaural Technology', and for her help in guiding me to the some additional references that were needed for this presentation.

Thanks also to Mikaël Kenikssi for permission to reproduce the photos in Figures 5, 6, 7 & 8 of the array used during the London Sinfonietta recordings.

5. REFERENCES

[1] Guillon, P. « Individualisation des indices spectraux pour la synthèse binaurale: recherche et exploitation des similarités inter-individuelles pour l'adaptation ou la reconstruction de HRTF », PhD Thesis, Université du Maine, Le Mans, France, 2009

[2] CIPIC database which can be downloaded at : http://interface.cipic.ucdavis.edu/sound/hrtf.html

[3] 2007 - 122nd AES Convention in Vienna - Preprint 7057

[4] 2008 – 124th AES Convention in Amsterdam - Preprint 7480

[5] 1991 – 91st AES Convention in New York - Preprint 3157

[6] Within the framework of the project entitled 'The Organ as Memory Bank' at the GOArt Gothenburg Organ Art Centre, University of Gothenburg, Sweden.