ECHO Project

ESMA-3D & PCMA-3D

Prof. Hyunkook Lee

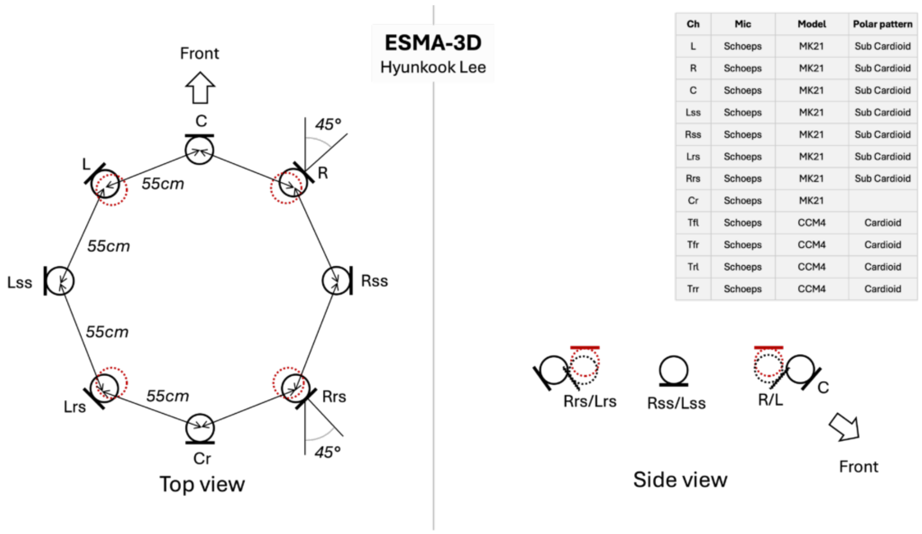
Applied Psychoacoustics Lab (APL)
University of Huddersfield

1. ESMA-3D

1.1. Background

The Equal Segment Microphone Array 3D (ESMA-3D) is optimised for 360° sound field recordings, providing accurate and continuous imaging across all loudspeakers during reproduction, as well as during sound field rotation for head-tracked VR applications. The original ESMA concept was proposed by Michael Williams in 1991 [1]. I optimised the configuration based on my MARRS ICTD/ICLD-based localisation model and augmented it with height channels for 3D recording in 2016 [2].

The ESMA design concept requires that: (i) the subtended angle between microphones in each stereophonic segment matches the base angle between the corresponding loudspeakers during reproduction, and (ii) the stereophonic recording angle (SRA) also matches the base angle. For example, in an octagonal layout, the microphone subtended angle, loudspeaker base angle, and SRA are all 45°, whereas in a quadraphonic layout, they are 90°. Despite this, the octagonal ESMA still provides good imaging across the L, C, and R channels in a Dolby Atmos 7.1.4 loudspeaker setup. The 55 cm microphone spacing ensures sufficient decorrelation at low and mid frequencies, contributing to a convincing spatial impression.



Ch	Mic	Model	Polar pattern
L	Schoeps	MK21	Sub Cardioid
R	Schoeps	MK21	Sub Cardioid
C	Schoeps	MK21	Sub Cardioid
Lss	Schoeps	MK21	Sub Cardioid
Rss	Schoeps	MK21	Sub Cardioid
Lrs	Schoeps	MK21	Sub Cardioid
Rrs	Schoeps	MK21	Sub Cardioid
Cr	Schoeps	MK21	
Tfl	Schoeps	CCM4	Cardioid
Tfr	Schoeps	CCM4	Cardioid
Trl	Schoeps	CCM4	Cardioid
Trr	Schoeps	CCM4	Cardioid

Figure 1. Configuration of the ESMA-3D.

The original octagonal ESMA employs cardioid microphones for the main layer. In the ECHO project, however, sub-cardioids (Schoeps MK21) were used instead to enhance low-end response. ESMA-3D extends the original ESMA by adding four cardioid microphones (Schoeps CCM4) for the height layer, arranged in a vertically coincident configuration. These height microphones face directly upwards to ensure at least 10 dB of channel separation—sufficient to prevent vertical image shift and undesirable colouration [3]. This arrangement also provides flexibility in adjusting the level of the height channels during mixing.

1.2. Balancing and Processing

My aim with ESMA-3D recordings is to achieve linear 360° imaging and a balanced spatial impression with a neutral tonal balance. This approach typically requires minimal balancing and processing, as the array design is based on an experimentally validated psychoacoustic model for localisation and spatial impression. As a result, the spatial characteristics are predictable, and consistent results can be achieved with careful array placement across a variety of recording situations.

At the same time, the array provides sufficient channel separation, allowing flexible adjustment of microphone levels without compromising imaging accuracy. For example, the supercardioid or cardioid height microphones, which face directly upwards, capture minimal interchannel crosstalk (i.e., direct sound in the height signals). This means that the height channel levels can be boosted, if needed, without causing upward image shift or undesirable colouration.

Balancing: Each microphone in the array was directly mapped to its corresponding loudspeaker without any panning or level adjustment. The only exception was the centre rear (Cr) microphone, which was routed to both the rear left and rear right loudspeakers to create a phantom rear centre image. (As mentioned above, the ideal loudspeaker setup for the octagonal ESMA-3D is eight loudspeakers spaced at 45° intervals in the horizontal layer.)

The initial balance used unity gain for all channels. Since all microphones employed Schoeps capsules with identical sensitivity, manual level matching was unnecessary. Depending on the room acoustics and the positions of sound sources, I occasionally adjust the levels of the rear and height microphones, but most of the time they remain at unity.

In the ECHO session, however, the ESMA-3D array was positioned slightly farther forward than the other arrays due to spatial constraints. As a result, the string sections in the Circular layout—located toward the rear—were slightly more distant from the ESMA-3D than from some of the other arrays. To compensate, I applied a 3 dB boost to the rear signals (Lrs, Rrs, and Cr). In the Traditional layout, all array signals were maintained at unity level.

Equalisation: In both the ArrayOnly and Presentation mixes, I applied EQ only to the rear and height channels. For the rear channels in the mixes for the Circular layout, I used FabFilter Pro-Q3 to apply a 2 dB boost at 1 kHz (Q = 2.0) and a 2 dB cut at 4 kHz (Q = 4.0). This was intended to reduce potential back-to-front confusion observed during the mixing process—likely due to limitations in phantom imaging between the rear speakers.

Blauert's directional bands theory [4] suggests that 4 kHz is associated with frontal perception, while 1 kHz is associated with rear perception, regardless of loudspeaker position. I believe this relates to the differences in head-related transfer functions (HRTFs) at 0° and 180° azimuths—where 0° tends to have relatively more energy around 4 kHz, and 180° around 1 kHz.

I found that applying this EQ helped slightly reduce back-to-front localisation confusion without significantly affecting the tonal character of the string sections. For the Traditional layout, this EQ was not necessary.

Figure 2. EQ applied for the rear channels in the ESMA-3D array mixes for the Circular layout.

A high-pass filter at 100 Hz (−12 dB/oct) and a 3 dB boost at 10 kHz (Q = 2.0) were applied to all the height channels. The high-pass filter was used to ensure clarity. Firstly, low-end energy is already sufficiently captured by the main layer microphones. Furthermore, due to the pitch–height effect, low frequencies reproduced from height loudspeakers are localised at ear height. This means that coherent low-frequency signals, such as those from percussion and contrabass, originating from both layers could potentially become overpowering.

The boost at 10 kHz was intended to add a sense of “air” to the height channels. While Schoeps capsules have a completely flat frequency response, which is ideal for achieving a neutral tonal balance in the main layer, I found that a slight boost around 10 kHz gave the height channels a more open and spacious quality.

Figure 3. EQ applied for the height channels in all of the ESMA-3D array mixes.

2. PCMA-3D

2.1. Background

PCMA-3D, which stands for Perspective Control Microphone Array 3D, aims to provide a natural and realistic representation of the sound field, ensuring accurate localisation and a balanced spatial impression.

I originally proposed the PCMA concept in 2011 for effective direct-to-reverberation ratio control in the 5.1 format, utilising two coincidentally arranged directional microphones at each pick-up point of the microphone array. In 2014, I adapted this concept for 3D recording, based on experimental findings suggesting that vertical microphone spacing or vertical decorrelation has little effect on overall spatial impression [5]. PCMA-3D, therefore, employs a horizontally spaced, vertically coincident microphone arrangement. The vertical pair is angled so that there is at least 7-10 dB of channel separation as in ESMA-3D. The supercardioid microphones (Schoeps CCM41) facing directly upwards captures the height information, providing maximal separation from the main channels. The sufficient channel separations across all the channels not only ensure stable imaging and natural tonality, but also allows flexible level balancing without affecting the imaging.

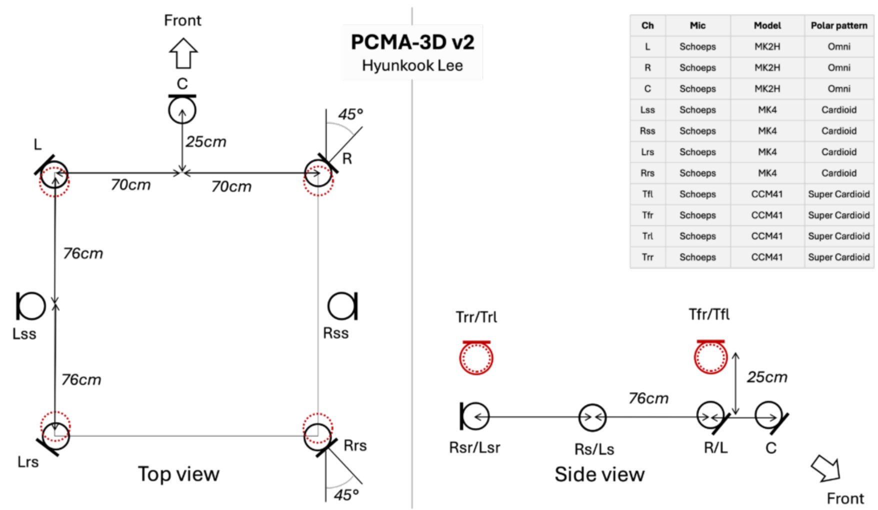
For the ECHO project, PCMA-3D v2 was used instead, featuring omnidirectional microphones (Schoeps MK2H) for the main layer (L, C, and R) to extend low-frequency response for the orchestra. This necessitates the height layer to be spaced from the main layer. PCMA-3D v2 places the height layer 25 cm above the main layer, providing sufficient channel separation through interchannel decorrelation above 1 kHz [6].

2.2. Balancing and Processing

As with the ESMA-3D, minimal balancing and processing was applied.

Balancing: Each microphone signal was directly routed to the corresponding loudspeaker without panning. The levels were kept the unity level, except for the centre channel being reduced by 3 dB. This is because the original configuration of PCMA-3D uses cardioids, but using omni microphones for L/C/R in v2 make the interchannel level differences reduced. By lowering the centre channel by 3 dB, narrowing of the L/C/R imaging was avoided.

Equalisation: The equalisations applied to the PCMA-3D mixes was identical to those for the ESMA-3D described in Section 1.2.



Ch	Mic	Model	Polar pattern
L	Schoeps	MK2H	Omni
R	Schoeps	MK2H	Omni
C	Schoeps	MK2H	Omni
Lss	Schoeps	MK4	Cardioid
Rss	Schoeps	MK4	Cardioid
Lrs	Schoeps	MK4	Cardioid
Rrs	Schoeps	MK4	Cardioid
Tfl	Schoeps	CCM41	Super Cardioid
Tfr	Schoeps	CCM41	Super Cardioid
Trl	Schoeps	CCM41	Super Cardioid
Trr	Schoeps	CCM41	Super Cardioid

Figure 4. Configuration of the PCMA-3D v2.