TOWARDS THE DEVELOPMENT OF INTELLIGENT MICROPHONE ARRAY DESIGNER

Hyunkook Lee

Applied Psychoacoustics Laboratory (APL)
University of Huddersfield
h.lee@hud.ac.uk

ABSTRACT

This paper introduces a novel stereophonic localisation prediction algorithm and the concept of new interactive software that can assist recording engineers in configuring a microphone array to achieve desired stereophonic localisation characteristics in acoustic recording.

1. EXISTING TOOLS

The so-called 'Williams curves' by Williams [1] present various options for combining the subtended angle and spacing between two directional microphones in different ICTD/ICLD ratio to achieve specific stereophonic recording angles (SRAs). The curves were derived from the polynomial interpolations of ICTD and ICLD values for a full image shift obtained by Simonsen [2] (i.e., 1.12ms or 15dB for 30° shift for the ±30° loudspeaker setup). In case of a near-coincident configuration, the ICTD and ICLD vary depending on the distance between sound source and microphone array, and so does the SRA of the array. However, it is not reported in Williams's papers what source-array distance his curves were derived for. Furthermore, the SRA calculation assumes that the height of the array is the same as that of the source, which is not usual in practical most classical recording situations.

Wittek's 'Image Assistant'[1] application calculates the SRA and localisation curve of a user-defined microphone configuration with the source-array distance taken into consideration. The tool predicts the degree of phantom image shift based on a psychoacoustic model proposed by Wittek and Theile [3]. The model suggests the image shift factors of 13%/0.1ms and 7.5%/dB and a linear trade-off between the two within the 75% shift region. The amount of total image shift within this region is simply the addition of the image shifts resulting from individual ICTD and ICLD, based on [4]. Although the tool provides many useful interactive features, the height of the array is assumed to be at the source height as in Williams's model.

Sengpiel's web application[2] and its tablet version by Neumann[3] also use a similar approach of adding individual ICTD and ICLD image shifts to determine the total image shift. This tool relies on the polynomial interpolation of individual ICTD and ICLD values for 25%, 50%, 75% and 100% image shifts obtained by Sengpiel. However, since the interpolated ICTD and ICLD have a non-linear relationship with the image shift, the simple summation of individual ICTD and ICLD image shifts to derive the total image shift does not seem to be logical. Sengpiel's individual ICTD value for a full image shift (1.5ms) is considerably larger than that by the other authors (1ms or 1.12ms), and therefore the SRA calculation results by this tool tend to differ largely from those by other tools. The tool uses a fixed source-array distance, which is unknown, for the SRA calculation. The height of the array is also not taken into account.

2. NEW PSYCHOACOUSTIC MODEL

In contrast to the aforementioned models and tools, the present author's array design model uses linear image shift factors that are adaptively applied to two separate shift regions of 0 to 66.7% (13.3%/0.1ms, 7.8%/dB) and 66.7° to 100° (6.7%/0.1ms, 3.9%/dB) in the ±30° loudspeaker setup. The ICTD and ICLD values required for a full image shift are 1 and 17dB. This is based on the findings of Lee and Rumsey [5]'s ICTD and ICLD panning experiments using musical sources. These factors can be traded linearly to find the combinations of ICTD and ICLD with various ratios.

Another novel aspect of the model is that it scales the shift factors depending on the loudspeaker base angle as opposed to both Williams's and Wittek and Theile's models. Theile's 'constant relative shift' theory [3] suggests that the individual ICTD and ICLD values required for a certain proportion of image shift are independent of the loudspeaker base angle. Williams [1] also claims that the SRA remains constant regardless of the loudspeaker base angle. However, the author's informal listening experiments observed that the ICTD or ICLD values for a full image shift obtained for the ±30° base angle are not large enough to achieve a full shift for ±45° base angle. From this, a new hypothesis was established as follows. As the base angle between two loudspeakers increases, the individual ICTD and ICLD required for a certain degree of image shift also increase in proportion to the increase of interaural time difference (ITD) and interaural level difference (ILD) produced by one of the loudspeakers. For example, the ITD for the 45° azimuth angle obtained from a KEMAR dummy head is about 1.5 times larger than that for a source at the 30° position. Similarly, the ILD for the 45° is about 1.3 times larger than that for the 30°. Hence, for the ±45°

[1] H. Wittek, www.hauptmikrofon.de, 2016.
[2] E. Sengpiel, www.sengpielaudio.com/HejiaE.htm, 2016.
[3] Neumann, https://itunes.apple.com/us/app/recording-tools/id576702914?mt=8, 2016.

loudspeaker setup, the individual ICTD and ICLD required for a full image shift are those for the ±30° setup multiplied by the scale factors of 1.5 and 1.3, respectively, resulting in 1.5ms and 22dB. It is assumed that the region-adaptive linear trade-off relationship mentioned above is still valid for the ±45° setup, and thus the shift factors become 8.8%/0.1ms and 6%/dB up to the 66.7% (30°) region, and 4.4%/0.1ms and 3%/dB for the 66.7% to 100% (45°) region. The proposed linear trade-off relationship with the base-angle-dependent scaling is described in Figure 1, and also defined in the equations below.

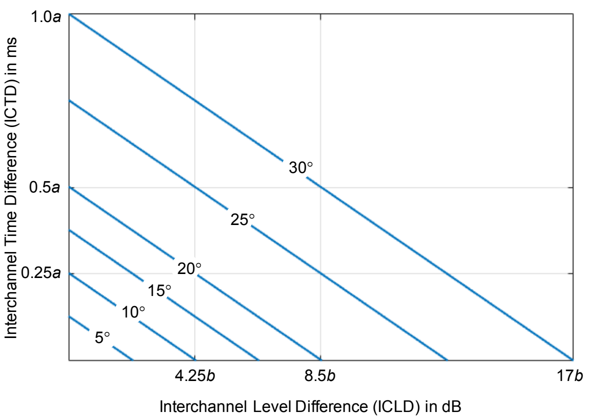

Figure 1. Proposed ICTD-ICLD trade-off functions. The angular values denote the predicted image positions φ.

$$\varphi = \left(ICTD + \frac{a}{17b} ICLD\right)\frac{4\theta}{3a}$$

$$, if\ ICTD \le -\frac{a}{17b} ICLD + \frac{a}{2}\ \&\ ICLD \le 17b\left(\frac{a}{2} - ICTD\right)$$

$$\varphi = \left(ICTD + \frac{a}{17b} ICLD + \frac{a}{2}\right)\frac{2\theta}{3a}$$

$$, otherwise$$

where φ = predicted image angle, θ = half the loudspeaker base angle, a = ITD(θ)/ITD(30°), b = ILD(θ)/ILD(30°).

3. NEW MICROPHONE ARRAY DESIGN TOOL

The new microphone array design tool by the Applied Psychoacoustics Lab, available for free download on the APL website[4], exploits the new psychoacoustic model described in Section 2 in order to produce more accurate image shift prediction results for different loudspeaker base angles. It aims to overcome the limitations of existing tools discussed above and also to provide a more interactive workflow using an object-based graphical user interface (GUI). The current version is available as tablet applications for the two-channel stereo only, but the tool will be extended for 2D and 3D multichannel formats in the future. It is also planned that the tool is made as a plugin that can simulate a microphone array recording for multitrack sources in virtual acoustics. The tool incorporates almost all available features of the existing tools. The novel features in the current version of the APL's microphone array design tool are as follows.

1) Multiple sources can be located anywhere on a virtual stage on the GUI. The predicted horizontal image positions for a user-defined or preset microphone configuration are plotted between two loudspeakers shown on the GUI, which change as the user moves the sources around. The vertical distance of the sources from the floor can also be adjusted.
2) The ICTD-ICLD relationship can be present for each source by selection. The localisation curve for a user-defined source-array distance can be also displayed.
3) The height of the microphone array and the up/down orientation angle as well as the spacing and angle between microphones and the horizontal position of the array are included in the control parameter set.
4) The tool recommends microphone configurations from 100% coincident (XY) to 100% spaced (AB) pair and lets the user select the ratio using a slider. The configurations are computed so that a user-selected physical span of a virtual ensemble to be perceived in a specific stereo width between the loudspeakers. The desired stereo width is defined by widening or narrowing a symmetrical region between the two loudspeakers displayed on the GUI. The physical span of the ensemble on the virtual stage is also visually controlled on the GUI. Additionally, as the XY/AB ratio is varied, the predicted position of each source on the display also changes accordingly.

4. REFERENCES

[1] M. Williams, "Unified theory of microphone systems for stereophonic sound recording," presented at the *82nd Convention of the Audio Engineering Society* (1987), convention paper 2466.

[2] M. Williams, "Migration of 5.0 multichannel microphone array design to higher order MMAD (6.0, 7.0&8.0) with or without the inter-format compatibility criteria," presented at the *124th Convention of the Audio Engineering Society* (2008), convention paper 7480.

[3] H. Wittek and G. Theile, "The Recording Angle - Based on Localization Curves," presented at the *112th Convention of the Audio Engineering Society* (2002 May), convention paper 5568.

[4] G. Theile, "On the performance of two- channel and multi-channel stereophony," presented at the *88th Convention of the Audio Engineering Society* (1990), convention paper 2932.

[5] H. Lee and F. Rumsey, "Level and time panning of phantom images for musical sources," *J. Audio Eng. Soc.*, vol.61 (12), pp. 753-767 (2013 Dec.).

[4] APL, Intelligent microphone array designer, https://www.hud.ac.uk/research/researchcentres/mtprg/projects/apl/resources/, 2016.