Auditory object localization of the variable-directivity icosahedral loudspeaker

Franz Zotter, Matthias Frank, Florian Wendt, Markus Zaunschirm

University of Music and Performing Arts, Institute of Electronic Music and Acoustics, Graz, Austria, Email: zotter@iem.at

Introduction

The strengths of the various acoustic propagation paths in a room influence what we perceive in terms of distance, direction, etc. Our variable-directivity icosahedral loudspeaker array (ICO) provides a means to control these strengths and hereby to alter emerging auditory objects.

The recent PhD Theses of Frank [1] and Stitt [2] established vector-based models for the localization of auditory events caused by coherent sound coming to the ear from multiple directions. While Frank's work considers a weighted $\boldsymbol{r}_\mathrm{E}$ model for perceived direction and width [3, 4] at central listening positions in Ambisonics (coincident arrival), the Stitt's work involves attenuation of preceded sound paths for the case of off-center positions in Ambisonics (non-coincident arrival) [5]. Applying to typical sound propagation paths of the ICO, i.e. for an enlarged timespan of non-coincident arrival times, one of our recent studies [6] applied a weighted $\boldsymbol{r}_\mathrm{E}$ model involving a $-\frac{1}{4}$ dB/ms threshold as a precedence threshold.

This contribution is an attempt to merge previous extended $\boldsymbol{r}_\mathrm{E}$ vector models into a new one. The model should be able to work with information of the coherent sound arriving from multiple paths either based on (a) delay and direction information of a simplified multipath propagation model, e.g. image source model, or (b) measured spherical microphone array impulse responses.

In an attempt of validation, we investigates whether the auditory event localization caused by the ICO can be modeled on the basis of Eigenmike measurements or a third-order image source model. The underlying perceptual data deal with static directivity in 4 directions.

Continuous $\boldsymbol{r}_\mathrm{E}$ vector model

For the desired versatile applicability, the model is assumed to operate on the basis of a directional impulse response $g(\boldsymbol{\theta}, t)$ with the continuous direction vector $\boldsymbol{\theta}$ and the continuous time t, to postpone a spatio-temporal discretization. This is mainly in favor of continuous symbolic directions and times of simple multipath models.

Accordingly, the typical definition of the $\boldsymbol{r}_\mathrm{E}$ vector that estimates direction and width for coherent sound incidence from discrete directions and strengths is modified to continuous directions and time. If precedence were ineffective, strengths would be obtained from integration over time and the $\boldsymbol{r}_\mathrm{E}$ vector model would become

$$\boldsymbol{r}_\mathrm{E} = \frac{\int_{\mathbb{S}^2} \left[\int g^2(\boldsymbol{\theta}, t)\,\mathrm{d}t\right]\,\boldsymbol{\theta}\,\mathrm{d}\boldsymbol{\theta}}{\int_{\mathbb{S}^2} \left[\int g^2(\boldsymbol{\theta}, t)\,\mathrm{d}t\right]\,\mathrm{d}\boldsymbol{\theta}}. \tag{1}$$

To involve precedence, however, the strengths need to be modified differently.

Suitable coordinates

As precedence is based on ITD, it is convenient to use a coordinate system in which ITD is an independent coordinate. With the ISO 31-11 spherical coordinate angles for the unit-length direction vector

$$\boldsymbol{\theta} = \begin{pmatrix} \sin\vartheta\,\cos\varphi \\ \sin\vartheta\,\sin\varphi \\ \cos\vartheta \end{pmatrix} = \begin{bmatrix} \sqrt{1-\mu^2}\begin{pmatrix}\cos\varphi\\ \sin\varphi\end{pmatrix} \\ \mu \end{bmatrix}$$

the most suitable orientation of the z axis is aligned with the right ear direction, x with the frontal direction, and y with the upward direction. In this orientation, the ITD (ear distance $d = 0.17m$, ITD$\approx 2r/c\,\cos\vartheta$) is accurately enough approximated by half the z component of $\boldsymbol{\theta}$

$$ITD \approx \frac{\mu}{2} \text{ in ms.}$$

For each lateral slice in μ, i.e. ITD, we calculate excitation E^xy and 2D $\boldsymbol{r}_\mathrm{E}^\mathrm{xy}$ vector from the response $g(\varphi, \mu, t)$

$$E^\mathrm{xy}(\mu, t) = \int_0^{2\pi} g^2(\varphi, \mu, t)\,\mathrm{d}\varphi, \tag{2}$$

$$\hat{\boldsymbol{r}}_\mathrm{E}^\mathrm{xy}(\mu, t) = \frac{\int_0^{2\pi} g^2(\varphi, \mu, t)\begin{pmatrix}\cos\varphi\\ \sin\varphi\end{pmatrix}\,\mathrm{d}\varphi}{\int_0^{2\pi} g^2(\varphi, \mu, t)\,\mathrm{d}\varphi}. \tag{3}$$

A laterally weighted $\boldsymbol{r}_\mathrm{E}$ vector model

According to Stitt [5], a weighting function w with the range $[0; 1]$ is involved to model precedence by attenuating the affected acoustic paths. We define an $\boldsymbol{r}_\mathrm{E}$ vector by integration over μ including the weight $w(\mu, t)$ that attenuates the excitation at corresponding μ, i.e. ITDs,

$$\boldsymbol{r}_\mathrm{E} = \frac{\int\int_{-1}^{1} w(\mu, t)\,\mathrm{E}^\mathrm{xy}(\mu, t)\begin{pmatrix}\sqrt{1-\mu^2}\,\hat{\boldsymbol{r}}_\mathrm{E}^\mathrm{xy}(\mu, t)\\ \mu\end{pmatrix}\,\mathrm{d}\mu\,\mathrm{d}t}{\int\int_{-1}^{1} w(\mu, t)\,\mathrm{E}^\mathrm{xy}(\mu, t)\,\mathrm{d}\mu\,\mathrm{d}t}. \tag{4}$$

Following Stitt [5], we define the weight to attenuate excitation that does not sufficiently lie above the precedence threshold, depending on the ratio of the excitation value for each μ to an excitation threshold E_th defined later:

$$w(\mu, t) = \min\left\{\frac{1}{8}\frac{E(\mu, t)}{E_\mathrm{th}(\mu, t)}, 1\right\}. \tag{5}$$

The precedence threshold

Interpreting Stitt's work [5] and what was used in [6], we define a precedence threshold $E_\mathrm{th}(\mu, t)$ by two-dimensional convolution of the lateral-temporal excitation $E^\mathrm{xy}(\mu, t)$ with a threshold mask $M(\mu, t)$

$$E_\mathrm{th}^\mathrm{xy}(\mu, t) = E^\mathrm{xy}(\mu, t) \star\star M(\mu, t). \tag{6}$$

The mask considers a build-up ranging from 0 to 1 of about $0 \leq t \leq 1\text{ms}$ over time and of about $|\mu| \leq \frac{1}{2}$ over laterality and a decay by a factor of $-\frac{1}{4}$dB/ms thereafter

$$M(\mu, t) = \alpha \cdot \min\left\{\frac{\max(t,0)}{1\text{ms}}, \frac{|\mu|}{2}, 10^{-\frac{1}{4}\frac{1}{20}\frac{t-1\text{ms}}{1\text{ms}}}\right\} \quad (7)$$

as shown in Fig. 1(a). As in Stitt's work, the constant α is involved to cover stationary excitation, for which a value of zero makes precedence ineffective. Moreover, α needs to attenuate for array measurements as they yield larger thresholds than simplified multipath propagation models because of their limited spatio-temporal resolution.

Comparison to experimental data

We compare the model to data from Zaunschirm's experiments [7] that describes auditory localization of either transient (sequence of irregular bursts) or stationary (pink noise with 500ms onset and release) sounds of both the virtual and the physical ICO. Fig. 2 shows the corresponding confidence ellipses of auditory localization in the room when using 4 beam directions to the front (0°), right (−110°), back (180°), and left (90°) of the room.

The setup can be recognized in Fig. 2 and has been measured in terms of impulse responses using the ICO (as in the experiment) and the Eigenmike (at the listening position), using the filters described in [7, 8] to obtain $g(\boldsymbol{\theta}, t)$ for the 4 different beamforming directions. On the other hand, the directional room impulses responses $g(\boldsymbol{\theta}, t)$ have been modeled in terms of weighted and superimposed Dirac delta functions in time and direction based on a 3rd order image source model using a frequency-independent 3rd order $\max - \boldsymbol{r}_\text{E}$ directivity pattern in the 4 different look directions.

Results obtained by shoebox model

Fig. 1(b) shows the mask and the excitation pattern due to the image-source model of the directional spatial impulse response for the 180° beam direction. Directional and temporal structure are strictly impulsive. A free scaling $\alpha = 0.4$ was used to improve readability.

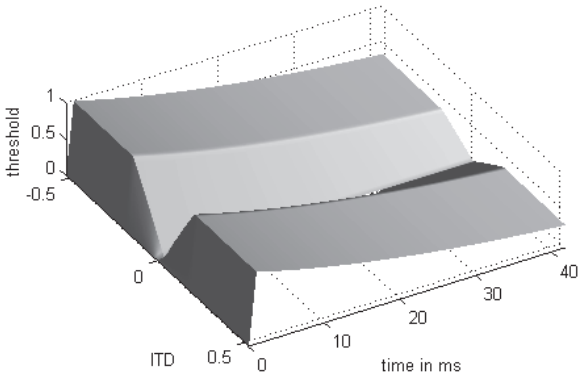
Fig. 2(a) shows results for the shoebox model with $\beta \cdot (1 - \|\boldsymbol{r}_\text{E}\|)$ used to plot distance along the angle $\angle\boldsymbol{r}_\text{E}$ on the horizon, cf. first 4 data lines in Table 1 for values, for a sequence of the directions $[0°, -110°, 180°, 90°]$. Lines with $\alpha = 0$ denote stationary sounds (free parameter $\beta = 10$), $\alpha = 1$ transient sounds ($\beta = 35$). Lateralization is over-estimated, but lateral sounds are correctly ranked in distance, which is not true for frontal sounds.

Results obtained by measurements

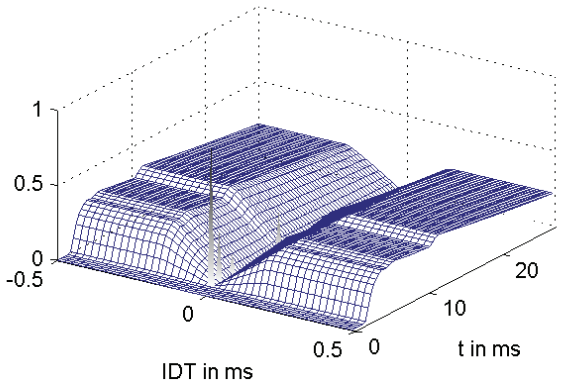
Fig. 1(c) shows the mask and the excitation pattern due to the directional spatial impulse response for the 180° beam direction measured by the Eigenmike. The direct sound is not strictly impulsive because of the band-limitation of the Eigenmike, and the directional mapping is not infinitely narrow either due to the Eigenmike's 4th-order resolution. The mask was scaled by a factor $\frac{1}{1000}$ to stand in reasonable relation to the excitation.

Table 1: $\boldsymbol{r}_\text{E}$ model of four beam directions.

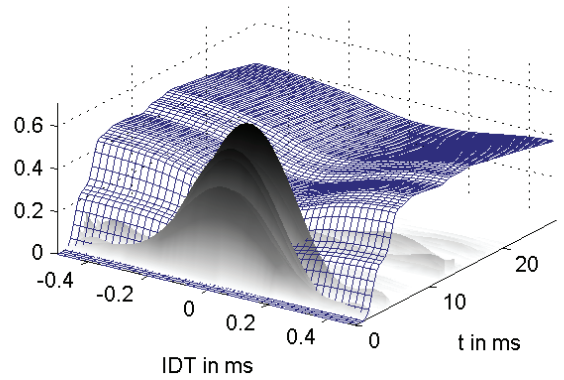
direction	0°	−110°	180°	90°
$\|\boldsymbol{r}_\text{E}\|$, $\alpha = 0$, mod.	0.67	0.76	0.57	0.70
$\angle\boldsymbol{r}_\text{E}$, $\alpha = 0$, mod.	5.4°	57°	9.4°	−63°
$\|\boldsymbol{r}_\text{E}\|$, $\alpha = 1$, mod.	0.88	0.98	0.91	0.97
$\angle\boldsymbol{r}_\text{E}$, $\alpha = 1$, mod.	3.6°	56°	6.2°	−67°
$\|\boldsymbol{r}_\text{E}\|$, $\alpha = 0$, meas.	0.50	0.51	0.53	0.52
$\angle\boldsymbol{r}_\text{E}$, $\alpha = 0$, meas.	8.7°	49°	7.4°	−38°
$\|\boldsymbol{r}_\text{E}\|$, $\alpha = \frac{1}{1000}$, meas.	0.82	0.81	0.88	0.72
$\angle\boldsymbol{r}_\text{E}$, $\alpha = \frac{1}{1000}$, meas.	4°	52°	3.9°	−31°



(a) Threshold build-up over (ITD,t) and decay over t.

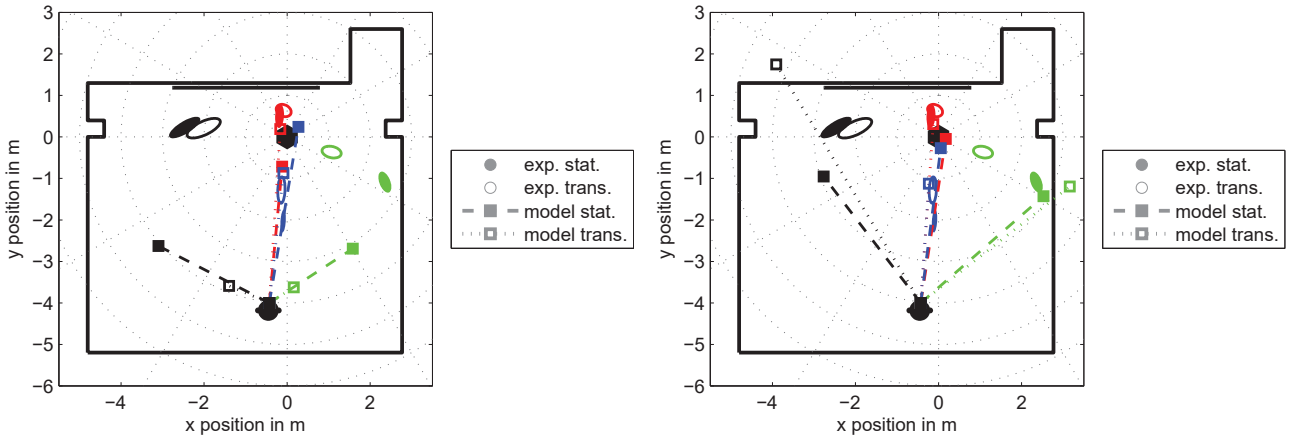


(b) Example of E and E_th based on image source model



(c) Example of E and E_th based on measurement

Figure 1: Convolution template M for the precedence threshold and its two-dimensional convolution over an exemplarily estimated binaural exitation pattern E.



(a) Model using 3rd order image source model.

(b) Model using Eigenmike32 measured responses.

Figure 2: Filled and open ellipses are 95% confidence regions of auditor localization for slowly and fast-attacking noise sound, respectively, for the directions $[0°(\text{red}), -110°(\text{green}), 180°(\text{blue}), 90°(\text{black})]$ with 0° (front) and 90° (left). The $\boldsymbol{r}_\text{E}$ model results (lines) use $\alpha = \{0, 1\}$ for the modeled room and $\alpha = \{0, 1/1000\}$ for the measured room to capture stationary (no precedence) and transient behavior. $1 - \|\boldsymbol{r}_\text{E}\|$ has been manualy scaled for every setting of α to roughly fit the auditory distances.

Fig. 1(c) and Tab. 1 show results that match experiments better in lateralization than with the simulated model. This might be because the measurement does not idealize the ICO to a frequency-independent 3rd order source. Nevertheless, angle- and distance-related ranking are not fully constistent with the experimental data.

Conclusion and outlook

We presented an extended $\boldsymbol{r}_\text{E}$ vector model involving precedence suitably to estimate auditory localization from measured and modeled spatial responses.

In this model, the typical $\boldsymbol{r}_\text{E}$ part covers the determination of elevation cues, avoiding monaural spectral cues. The precedence effect is then modeled on the basis of a binaural excitation pattern in which precedence is modeled to suppress affected lateral excitation. The rest is a classical $\boldsymbol{r}_\text{E}$ model, which, similarly as some known models, gathers the lateral information by a centroid, however without discarding information about elevation.

Despite the comparison we gave to experimental data is not entirely satisfactory, closer parameter matching would not be reasonable either. Our data material is not exhaustive enough for model calibration. We think that our model could be generic enough to calibrate the shape of the mask M and the attenuation α as soon as exhaustive-enough experimental data is available, and it might be useful to involve directivity of hearing, cf. [4].

It is probable that the model not only applies to impulse responses but to signals as well. In slowly attacking signals, the shape of M automatically yields unsharpness between early and late sound instances. The parameter α could be adaptive similar to adaptive IACC thresholds.

Our model neither separates multiple perceivable directions, nor does it spectrally segregate sources. Nevetheless, it combines the simple and powerful $\boldsymbol{r}_\text{E}$ estimator with what is known from binaural excitation patterns.

Acknowledgments

The Austrian Science Fund (FWF) supports our work by the artistic research project *Orchestrating Space by Icosahedral Loudspeaker* (OSIL), AR 328-G21.

References

[1] M. Frank, "Phantom sources using multiple loudspeakers in the horizontal plane," Ph.D. dissertation, University of Music and Performing Arts, Graz, 2013.

[2] P. Stitt, "Ambisonics and higher-order ambisonics for off-centre listeners: Evaluation of perceived and predicted image direction," Ph.D. dissertation, Queen's University, Belfast, 2015.

[3] M. Frank, "Source width of frontal phantom sources: Perception, measurement, and modeling," *Archives of Acoustics*, vol. 38, no. 3, 2013.

[4] ——, "Localization using different amplitude-panning methods in the frontal horizontal plane," in *EAA Symposium Auralization and Ambisonics, Berlin*, 2014.

[5] P. Stitt, S. Bertet, and M. van Walstijn, "Extended energy vector prediction of ambisonically reproduced image direction at off-centre listening positions," *J. Audio Eng. Soc.*, 2016.

[6] F. Zotter and M. Frank, "Investigation of auditory objects caused by directional sound sources in rooms," *Acta Physica Polonica*, vol. 128, no. 1-A, 2015.

[7] M. Zaunschrim, M. Frank, and F. Zotter, "An interactive virtual icosahedral loudspeaker array," in *Fortschritte der Akustik, DAGA, Aachen*, 2016.

[8] S. Lösler and F. Zotter, "Comprehensive radial filter design for practical higher-order ambisonic recording," in *Fortschritte der Akustik, DAGA, Nürnberg*, 2015.