Simple Uncertainty Prediction for Phantom Source Localization

Matthias Frank
Institute of Electronic Music and Acoustics, Email: frank@iem.at
University of Music and Performing Arts, 8010 Graz, Austria

Introduction

A horizontally arranged pair of loudspeakers can create an auditory event between the loudspeakers, a so-called phantom source [1]. Phantom sources are known to be localized with larger uncertainty compared to single sound sources [2, 3, 4]. However, only few studies differentiated between the uncertainty that is due to inter- and intra-subjective variation [5, 6]. This contribution discusses the influence of both variations and presents a reasonable explanation for inter-subjective variation that can be incorporated into a simple localization model.
The paper first introduces binaural and vector-based localization models. Their predictions of the experimental results from [4] are compared pointing out that the weighted energy vector yields best results. A suitable weighting is derived from the directivity of human hearing. The subsequent section presents experimental results that reveal the intra-subjective variation to be surprisingly small and that uncertainty is mainly due to inter-subjective variation. Finally, the localization models are modified in order to predict this uncertainty.

Localization Models

This section introduces a binaural model, as well as vector models for the prediction of phantom source localization.

Binaural Model

The employed binaural localization model after Lindemann [7, 8] is part of the Auditory Modeling Toolbox [1]. It divides the binaural input signals into 36 frequency bands, and applies half-wave rectification and low-pass filtering to model the auditory nerve. In each band, the inter-aural time-difference (ITD) is computed as the centroid of the inter-aural cross correlation function [9].
The ITD value of the phantom source is compared to the values of a single sound source from a lookup table. The best matching ITD is selected and the corresponding angle is regarded as the angle of the phantom source for the present frequency band. The median value of all bands represents the broadband prediction result. In this contribution, best results have been achieved when using 18 frequency bands covering the range 124 Hz . . . 1979 Hz. The model can be fed with head-related impulse responses (HRIR) or binaural room impulse responses. This work employs an HRIR database from the Acoustic Research Institute (ARI) with data from 104 persons [2].

[1] freely available on `amtoolbox.sourceforge.net/`

[2] freely available on `https://www.kfs.oeaw.ac.at/index.php?option=com_content&view=article&id=608:ari-hrtf-database&catid=158:resources-items&Itemid=606&lang=en`

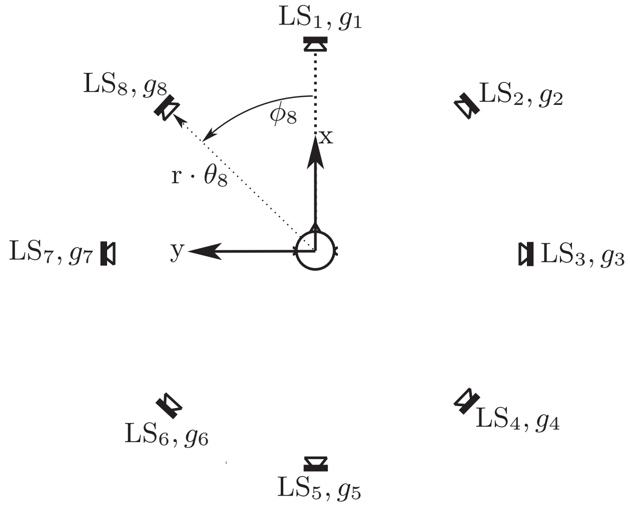

Figure 1: Coordinate system and loudspeaker setup used in the experiments from [4, 19].

Vector Models

Linear summation of the weighted loudspeaker directions $\boldsymbol{\theta}_l = [\cos(\phi_l),\ \sin(\phi_l)]^{\mathrm{T}}$, cf. Figure 1, is used to calculate the *velocity vector* [10, 11]

$$\boldsymbol{r}_{\mathrm{V}} = \frac{\sum_{l=1}^{\mathrm{L}} g_l \boldsymbol{\theta}_l}{\sum_{l=1}^{\mathrm{L}} g_l}. \tag{1}$$

The direction of this vector is assumed to correspond to the localization of low frequencies ($\leq$ 700Hz) and is identical to the direction predicted by vector-base amplitude panning (VBAP, [12]) and the tangent law [13].

Similarly, the *energy vector* [11] is defined as

$$\boldsymbol{r}_{\mathrm{E}} = \frac{\sum_{l=1}^{\mathrm{L}} g_l^2 \boldsymbol{\theta}_l}{\sum_{l=1}^{\mathrm{L}} g_l^2}. \tag{2}$$

The energetic superposition of the loudspeaker signals is expected to model the localized direction for higher frequencies or broadband signals. Just as the velocity vector, the energy vector has no restriction regarding the number of simultaneously active loudspeakers. The magnitude of the energy vector can also used to describe the perceived width of phantom sources [5, 14, 15].
Direction-dependent weighting $w(\boldsymbol{\theta}_l)$ can be applied to the energy vector [11] by weighting of the loudspeaker gains [14]. This yields the *weighted energy vector*

$$\boldsymbol{r}_{\mathrm{E}}^{\mathrm{w}} = \frac{\sum_{l=1}^{\mathrm{L}} (g_l w(\boldsymbol{\theta}_l))^2 \boldsymbol{\theta}_l}{\sum_{l=1}^{\mathrm{L}} (g_l w(\boldsymbol{\theta}_l))^2}. \tag{3}$$

It seems useful to apply the directivity of human hearing as directional weighting.

Directivity of Human Hearing

Jahn conducted experiments on how a single value for the perceived binaural loudness could be determined from the sound pressure of both ears $p_{left}(\phi)$ and $p_{right}(\phi)$ [16, 17]. He found the following frequency-independent summation formula for the directional loudness or directivity $D(\phi)$

$$D(\phi) = \left(\frac{|p_{left}(\phi)|^k + |p_{right}(\phi)|^k}{|p_{left}(0)|^k + |p_{right}(0)|^k} \right)^{\frac{1}{k}}. \quad (4)$$

Although he found the exponent to be $k = 2.1$ for levels close to the absolute hearing threshold and $k = 1.77$ for higher levels (50-70 phon), later work found that an exponent of $k = 2$ yields the best correlation to listening experiments [18]. This value is also used in this work.
Using Eq. (4), the directivity of human hearing can be computed directly from head-related transfer functions. In order to have a good estimate of the average directivity, this work again employs the ARI database. The median of this broadband directivity is used as direction-dependent weighting $w(\boldsymbol{\theta}_l)$ for $\boldsymbol{r}_\mathrm{E}^\mathrm{w}$, cf. Figure 3.

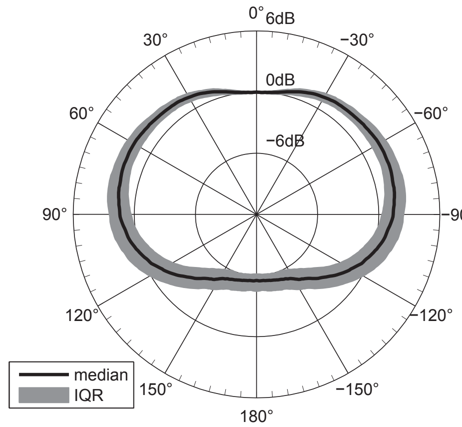

Figure 3: Median and IQR of human broadband hearing directivity, computed from 104 listeners of the ARI database.

Prediction of Localization Results from [4]

All presented models are employed to predict the median localization results from Simon [4]. The localization experiment studied the perceived direction of phantom sources created by pairwise panning with level differences between ±18dB in steps of 4.5dB on loudspeaker pairs with 45° aperture angle on a horizontal semicircle, cf. Figure 1. The prediction results in Figure 2 show that the binaural model with a lookup table for the whole 0° . . . 180° range is not capable of correctly distinguishing between front and back. Improvement is achieved by limiting the lookup table to the angular range of the active loudspeaker pair, e.g. 90° . . . 135° for the [90°, 135°] loudspeaker pair. For the frontal directions, the predictions of the vector models and the range-limited binaural model are similar. The advantage of the weighted energy vector is most obvious for lateral directions. Figure 5 compares the median prediction errors of the models. The weighted energy vector performs best ($p \leq 0.1$). The range-limited binaural model and the velocity vector are at par ($p = 0.76$), while the latter is worse than the energy vector ($p = 0.077$).

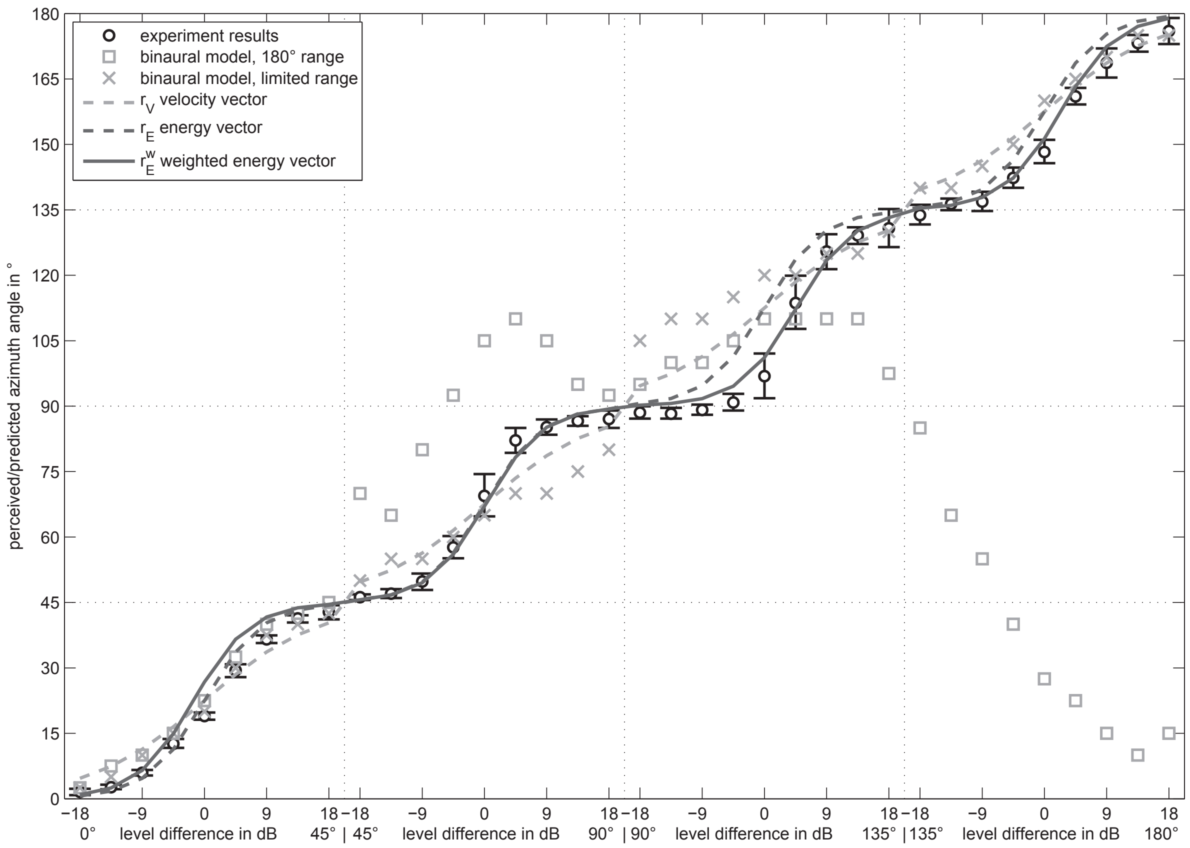

Figure 2: Localization results from [4] (medians and confidence intervals) and prediction of median results by different predictors.

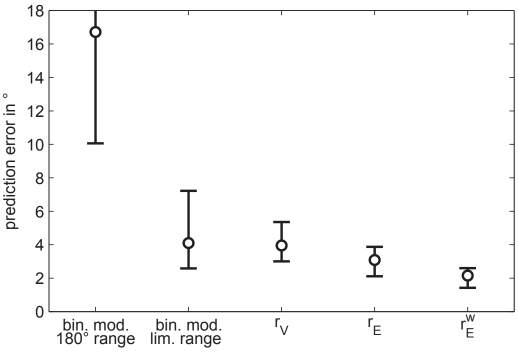

Figure 5: Medians and 95% confidence intervals of the prediction errors for all configurations from [4].

Inter- or Intra-Subjective Variation?

Is the uncertainty due to inter- or intra-subjective variation? To clarify, a detailed dataset from an own experiment [19] is analyzed. The experiment employed a regular ring of Genelec 8020 loudspeakers at a radius of r = 2.5 m at an ear height of 1.2 m in the 10.3 m × 12 m × 4.8 m large IEM CUBE [5], cf. Figure 1. The stimulus consisted of 3 pink noise bursts, each with 100 ms fade-in, 200 ms at 65 dB(A), 100 ms fade-out, and 200 ms silence before the next fade-in. The perceived direction was assessed in random order using a pointing method with a toy-gun that was captured by an infrared tracking system [20].
The selected dataset includes 2 conditions, that were evaluated 6 times by each subject: 0dB level difference between loudspeaker pairs at $[0°, 45°]$ and $[90°, 135°]$. All 14 subjects were part of a trained listening panel [21, 22, 23].
Table 1 compares the total, inter-subjective, and intra-subjective inter-quartile ranges (IQR) of the localized directions for both loudspeaker pairs. The total IQRs are computed from the $84 = 14$ (subjects) × 6 (repetitions) answers for each loudspeaker pair representing the overall variation caused by both inter-subjective and intra-subjective variation. For each loudspeaker pair, the inter-subjective IQR is calculated after replacing each subject's answers by her/his median answer for this loudspeaker pair, i.e. using a single value for each subject. Vice versa, the intra-subjective IQR is calculated after suppressing the inter-subjective variation of each subject by subtracting the subject's median answer.

Table 1: Inter-quartile ranges of the localized directions for different pairs of equally loud loudspeakers.

loudspeaker pair in °	inter-quartile range in °		
	total	inter-subjective	intra-subjective
[0, 45]	6.1	3.3	2.6
[90, 135]	17.1	16.3	8.1

For the frontal loudspeaker pair at $[0°, 45°]$, the inter-subjective variation is a bit greater than the intra-subjective variation. An analysis of variance shows a significant effect of the subjects ($p \ll 0.001$), but no effect of the repetitions ($p = 0.179$). For the lateral loudspeaker pair at $[90°, 135°]$, the inter-subjective variation is twice as big as the intra-subjective variation. In this case, nearly the whole overall variation is caused by inter-subjective variation. Again, the effect of the subjects is significant ($p \ll 0.001$), but the effect of the repetitions is not ($p = 0.755$). This finding shows that the subjective differences in the localization is dominated by differences between the subjects.

Prediction of Localization Uncertainty

Results in the literature [2, 3, 4] show the largest uncertainty for level differences close to 0dB. Combining this knowledge with the results from the above experiment, it seems reasonable to attribute the inter-subjective differences to individual directivity patterns. Imagine a listener that is more sensitive in a certain direction $\boldsymbol{\theta}_1$. If a loudspeaker at this direction contributes to a phantom source, the listener will perceive it closer to $\boldsymbol{\theta}_1$ compared to another listener that is less sensitive in $\boldsymbol{\theta}_1$.
The differences in the directivity can be incorporated in the weighted energy vector by employing the IQR of the directivity, cf. Figure 3. For two loudspeakers, this can be done by first weighting one loudspeaker with the upper quartile of the directivity and the other one with the lower quartile, and then vice versa. The resulting area between both localization curves predicts the inter-subjective variation. In the binaural model, inter-subjective variation is predicted by the standard deviation of the directions calculated for the 104 individual HRIR datasets.

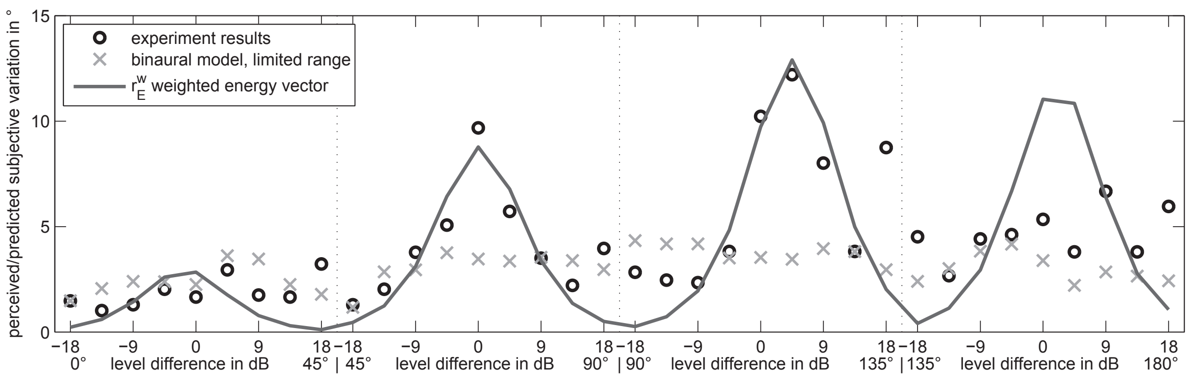

Figure 4: Confidence interval of experimental results and predicted variation by different models for all conditions from [4].

Figure 4 shows the confidence interval from [4] and the two predictions. The vector-based prediction achieves a correlation to the experimental results of 73%, whereas the binaural prediction yields only 33%.

Conclusion

Uncertainty of phantom source localization is mainly due to inter-subjective differences, especially at lateral directions. The hypothesis seems natural that the differences are caused by differences in the directivity of hearing: Incorporated in the simple weighted energy vector model, the localization uncertainty can be predicted with a 73% correlation.

Acknowledgment

This work was partly supported by the project ASD, which is funded by Austrian ministries BMVIT, BMWFJ, the Styrian Business Promotion Agency (SFG), and the departments 3 and 14 of the Styrian Government. The Austrian Research Promotion Agency (FFG) conducted the funding under the Competence Centers for Excellent Technologies (COMET, K-Project), a program of the above-mentioned institutions.

References

[1] K. Wendt, *Das Richtungshören bei der Überlagerung zweier Schallfelder bei Intensitäts- und Laufzeitstereophonie.* PhD thesis, RWTH Aachen, Germany, 1963.

[2] G. Theile and G. Plenge, “Localization of lateral phantom sources,” *J. Audio Eng. Soc*, vol. 25, no. 4, pp. 196–200, 1977.

[3] V. Pulkki, “Localization of amplitude-panned virtual sources II: Two- and three-dimensional panning,” *J. Audio Eng. Soc*, vol. 49, no. 9, pp. 753–767, 2001.

[4] L. S. R. Simon, R. Mason, and F. Rumsey, “Localization curves for a regularly-spaced octagon loudspeaker array,” in *Audio Engineering Society Convention 127*, 10 2009.

[5] M. Frank, *Phantom Sources using Multiple Loudspeakers in the Horizontal Plane.* PhD thesis, University of Music and Performing Arts Graz, Austria, 2013.

[6] F. Wendt, M. Frank, and F. Zotter, “Amplitude panning with height on 2, 3, and 4 loudspeakers,” in *International Conference on Spatial Audio*, 2014.

[7] W. Lindemann, “Extension of a binaural cross-correlation model by contralateral inhibition. I. Simulation of lateralization for stationary signals,” *The Journal of the Acoustical Society of America*, vol. 80, no. 6, pp. 1608–1622, 1986.

[8] W. Lindemann, “Extension of a binaural cross-correlation model by contralateral inhibition. II. The law of the first wave front,” *The Journal of the Acoustical Society of America*, vol. 80, no. 6, pp. 1623–1630, 1986.

[9] L. A. Jeffress, “A place theory of sound localization,” *Journal of comparative and physiological psychology*, vol. 41, no. 1, pp. 35–39, 1948.

[10] Y. Makita, “On the directional localization of sound in the stereophonic sound field,” tech. rep., EBU, Review Part A, 73, 1962.

[11] M. A. Gerzon, “General metatheory of auditory localisation,” in *Audio Engineering Society Convention 92*, 3 1992.

[12] V. Pulkki, “Virtual sound source positioning using vector base amplitude panning,” *J. Audio Eng. Soc*, vol. 45, no. 6, pp. 456–466, 1997.

[13] D. M. Leakey, “Some measurements on the effects of interchannel intensity and time differences in two channel sound systems,” *The Journal of the Acoustical Society of America*, vol. 31, no. 7, pp. 977–986, 1959.

[14] M. Frank, G. Marentakis, and A. Sontacchi, “A simple technical measure for the perceived source width,” in *Fortschritte der Akustik, DAGA*, 2011.

[15] M. Frank, “Source width of frontal phantom sources: Perception, measurement, and modeling,” *Archives of Acoustics*, vol. 38, pp. 311–319, September 2013.

[16] G. Jahn and S. Vogelsang, “Die einohrige Richtcharakteristik des menschlichen Gehörs,” *Hochfrequenztechnik und Elektroakustik*, vol. 68, 1959.

[17] G. Jahn, “Zum Unterschied zwischen einohrigem und zweiohrigem Hören,” *Hochfrequenztechnik und Elektroakustik*, vol. 72, 1963.

[18] V. P. Sivonen and W. Ellermeier, “Directional loudness in an anechoic sound field, head-related transfer functions, and binaural summation,” *The Journal of the Acoustical Society of America*, vol. 119, no. 5, pp. 2965–2980, 2006.

[19] M. Frank, “Localization using different amplitude-panning methods in the frontal horizontal plane,” in *Proc. of the EAA Joint Symposium on Auralization and Ambisonics*, (Berlin), pp. 42–47, 2014.

[20] M. Frank, L. Mohr, A. Sontacchi, and F. Zotter, “Flexible and Intuitive Pointing Method for 3-D Auditory Localization Experiments,” in *Audio Engineering Society Conference: 38th International Conference: Sound Quality Evaluation*, 6 2010.

[21] A. Sontacchi, H. Pomberger, and R. Höldrich, “Recruiting and evaluation process of an expert listening panel,” in *Fortschritte der Akustik, NAG/DAGA*, 2009.

[22] M. Frank, A. Sontacchi, and R. Höldrich, “Training and guidance tool for listening panels,” in *Fortschritte der Akustik, DAGA*, 2010.

[23] M. Frank and A. Sontacchi, “Performance review of an expert listening panel,” in *Fortschritte der Akustik, DAGA*, 2012.