Panning with Height on 2, 3, and 4 Loudspeakers

Florian Wendt, Matthias Frank, Franz Zotter
University of Music and Performing Arts Graz, Austria
Institute of Electronic Music and Acoustics, Email: {frank, zotter}@iem.at

Abstract

A phantom source is typically created by two loudspeakers playing the same signal. In the horizontal plane, phantom source localization can be explained by binaural localization cues, i.e. the inter-aural time and level differences. Three-dimensional sound systems based on amplitude panning such as Vector-Base Amplitude Panning (VBAP) and Ambisonics assume the existence of vertical phantom sources.

Obviously, vertical phantom sources cannot be explained by inter-aural differences, since cues that decode elevation are spectral modifications caused by pinna, head, and torso. Indeed studies exist which imply that localization of vertical phantom sources for multichannel loudspeaker arrays is possible, even though the localization is intersubjective and therefore phantom sources are localized individually. The ICSA 2011 was the starting point for a discussion that raised the question if and how such localization works.

Since then, studies were presented that confirm the existence of amplitude-panned phantom sources created by two loudspeakers placed in the median plane. This article examines the localization of amplitude-panned phantom sources created by horizontal and vertical loudspeaker pairs, as well as loudspeaker triangles and rectangles. The results are compared to existing vector models yielding a conclusive proposal for three-dimensional amplitude panning.

1. Introduction

Surround playback was restricted to the horizontal plane for decades. Stereophonic panning methods make use of level and/or time delay differences between the loudspeakers to control the position of auditory objects between the loudspeakers, so-called phantom sources [1]. In recent years, the introduction of 3D displays in movie theaters enabled the utilization of all three dimensions and allow a new form of entertainment. On the one hand, the development of suitable marketable sound systems providing 3D sound environments is still in the early stage, and there is a variety of manifestations, be it in the form of Ambisonics [2], Vector-Base Amplitude Panning (VBAP) [3], Dolby Atmos [4], AURO 3D [5], and 22.2 [6], that differs in theory and number of loudspeakers. On the other hand, all of these systems assume that vertical phantom sources exist and can be panned between directions of the loudspeakers.

Localization of single sound sources in the horizontal plane is well studied, and inter-aural level and time differences (ILD and ITD) are proved to be the relevant localization cues [7]. This holds true for the localization of horizontal phantom sources, because the superposition of the head-related transfer functions (HRTFs) for the loudspeaker pair yields inter-aural differences resembling those of a single sound source [8].

For localization in vertical planes, and especially in the median plane, such cues are weak or absent. As auditory perception still lets us localize a single elevated sound source, other cues, predominantly spectral properties of the HRTFs [9], play an important role complementing the inter-aural differences.

Largely the same ILD and ITD values as those of a single source on the horizon are caused by a vertical pair of loudspeakers lying at the same interaural angle, symmetrical to the horizontal plane. ILD and ITD of both loudspeakers are then invariant to amplitude and time-delay panning. Moreover, spectral cues of their superimposed HRTFs need not resemble the HRTF of the single source in between. Still a few studies exist which imply that localization and panning of vertical phantom sources is possible [10, 11, 12, 13, 14]. To examine the existence of vertical phantom sources created by the various 3D audio systems in general, we describe the problem as simple and consistent as possible. The listening experiment presented here to investigate horizontal and vertical phantom sources uses only basic arrangements of 2, 3, and 4 loudspeakers spanning a horizontal and a vertical dimension in front of the listener.

This paper is arranged as follows: It briefly discusses to what extent known two-channel panning methods apply to vertical loudspeaker arrays. The subsequent section on the listening experiment describes method, conditions, and results. In order to examine the controllability of the phantom source localization, subjective variation is discussed. Section four presents simple localization predictors that are based on the loudspeaker gains and positions and compares them to the experimental results.

2. Panning with Height

In horizontal stereophonic setups, panning of phantom sources is achieved by the use of inter-channel time and level differences (ICTD, ICLD) [1]. Studies report that amplitude panning using ICLDs is also applicable on vertical loudspeaker arrays in a similar way [10, 11, 12, 13, 14]. However, none of the studies mentions the use of ICTD as control parameter.

To investigate the impact of ICLDs and ICTDs on the vertical phantom source and to determine appropriate control parameter, a preliminary test was performed using a vertical loudspeaker pair at elevation angles of ±20° in the median plane [15]. The results show that the ICTD randomly changes the vertical position of the phantom source and is thus no suitable control parameter. In contrast, ICLDs yield a monotonic localization curve, even if additional ICTDs are applied to the loudspeakers. The present study therefore focuses on the application of ICLDs (amplitude panning) only.

3. Experiment

The listening experiment evaluates the horizontal and vertical localization of amplitude-panned phantom sources using arrays of 2, 3, and 4 loudspeakers.

3.1. Setup & Method

Figure 1 shows a sketch of the entire loudspeaker setup to evaluate the various loudspeaker arrays under test. The 14 Genelec 8020A loudspeakers were placed on a ring with a diameter of $1.7\,\mathrm{m}$ centred around an additional center loudspeaker at $0°$ elevation and azimuth. From the subject's head, all loudspeakers were $2.5\,\mathrm{m}$ away resulting in an aperture angle of $40°$. The exact position of the loudspeakers are shown in Table 1.

The whole loudspeaker setup was covered by an acoustically transparent screen. The experiment was performed in the IEM CUBE, a $10.3\,\mathrm{m} \times 12\,\mathrm{m} \times 4.8\,\mathrm{m}$ large room with a mean reverberation time of $470\,\mathrm{ms}$ that fulfils the recommendation for surround reproduction in ITU-R BS.1116-1 [16]. The central listening position lies within the effective critical distance.

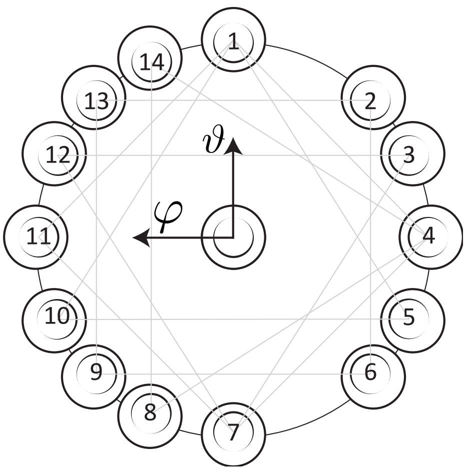

Figure 1: Sketch of the loudspeaker setup with implied loudspeaker arrays.

Loudspeaker	φ / °	ϑ / °
1	0.0	20.0
2	−14.1	14.1
3	−17.3	10.0
4	−20.0	0.0
5	−17.3	−10.0
6	−14.1	−14.1
7	0.0	−20.0
8	10.0	17.3
9	14.1	−14.1
10	17.3	−10.0
11	20.0	0.0
12	17.3	10.0
13	14.1	14.1
14	10.0	17.3

Table 1: Azimuth and elevation angle φ, ϑ of the loudspeakers.

During the listening experiment, the subjects were requested to keep their heads immobile and to face the $0°$ direction, which corresponds to the position of the center loudspeaker and to the angular origin of the coordinate system. The perceived directions were recorded by pointing with a motion tracked toy-gun [17]. The corresponding azimuth and elevation angles φ and ϑ were stored when the subjects pulled the trigger of the toy-gun.

3.2. Conditions

The evaluated loudspeaker arrays were pairs, equilateral triangles, and squares in different rotations, cf. Table 2.

Abbreviation	Geometry	Active Loudspeakers
H	–	4, 11
V	|	1, 7
T1	△	1, 5, 10
T2	▽	3, 7, 12
T3	▷	4, 8, 14
S1	◇	2, 6, 9, 13
S2	□	1, 4, 7, 11

Table 2: Overview of the examined loudspeaker arrays.

Each loudspeaker array shown in Table 2 was evaluated for an ICLD of $0\,\mathrm{dB}$, i.e. all 2, 3, or 4 loudspeakers were playing the stimulus with the same level. Additionally to these conditions, different ICLD values were tested in order to evaluate amplitude panning on the different arrays. The loudspeaker pairs H and V employed ICLDs of $\Delta g = \{\pm 3, \pm 6\}\,\mathrm{dB}$.
Vertical panning on array T1 evaluated ICLDs of $\Delta g = \{+11.5, -\infty\}\,\mathrm{dB}$ between the upper loudspeaker (1) and the lower loudspeaker pair (5, 10). These ICLD values result from VBAP for panning a phantom source to elevation angles of $\pm 10°$. In the latter case, only the lower loudspeaker pair was active. The same ICLDs were used on array T2 between the lower loudspeaker (7) and the upper loudspeaker pair (3, 12). Array T3 was evaluated for horizontal panning employing again the same ICLD values between the right loudspeaker (4) and the left loudspeaker pair (8, 14).
On array S1 horizontal panning was achieved by a level of $+6\,\mathrm{dB}$ for the left (11) or right loudspeaker (4). Correspondingly, a level of $+6\,\mathrm{dB}$ for the upper (1) and lower (7) loudspeaker was employed for vertical panning. Vertical panning on array S2 used a level of $+6\,\mathrm{dB}$ for the upper (13, 2) or the lower loudspeaker pair (9, 6), whereas horizontal panning used $+6\,\mathrm{dB}$ for the left (9, 13) or the right (6, 2) loudspeaker pair.

The different conditions were normalized energetically to yield sufficiently uniform loudness. Each condition was evaluated twice, resulting in 52 conditions for each subject. The order of the conditions was generated randomly for each subject. The stimulus consisted of 3 pink noise bursts, each one with a $100\,\mathrm{ms}$ fade-in and $100\,\mathrm{ms}$ fade-out and $200\,\mathrm{ms}$ of silence in between.
There were 15 subjects participating the experiment. All of them are experienced listeners with normal hearing.

4. Results

For a neat comparison of the both perceived azimuth and elevation angles for the tested array, a two-dimensional plot presents each mean value within the elliptical border of its 95% confidence area.

In a first step the normally distributed spherical coordinates φ and ϑ are re-interpreted in terms of Cartesian coordinates x and y with less than < 0.5° distortion as they stay < 20°. The bivariate normal distribution of the direction $\boldsymbol{r} = (x\ y)^T$ is then defined as:

$$\mathcal{N}(\boldsymbol{r}) = \frac{1}{2\pi\sqrt{det(\boldsymbol{C})}} e^{\left\{-\frac{1}{2}(\boldsymbol{r}-\boldsymbol{\mu})^T \boldsymbol{C}^{-1}(\boldsymbol{r}-\boldsymbol{\mu})\right\}}. \quad (1)$$

The vector $\boldsymbol{\mu} = (\mu_x\ \mu_y)^T$ describes the expectation vector of the direction $\boldsymbol{r}$ and the matrix $\boldsymbol{C}$ is the covariance matrix of x and y. The matrix $\boldsymbol{C}$ describes the two-dimensional distribution around $\boldsymbol{\mu}$. The roots of its eigenvalues multiplied by the corresponding eigenvectors describe the semi-axes of an ellipse, the two-dimensional area of standard deviation. By scaling the ellipse with a value depending on the number of data points for each condition (two answers from each 15 subjects), we obtain the 95% confidence area for $\boldsymbol{\mu}$.

4.1. Loudspeaker Pairs

The mean values and the corresponding confidence areas obtained for loudspeaker pairs H and V are presented in **Figure 2**. For array H, an analysis of variance (ANOVA) reveals the ICLD to be a significant factor for the azimuth angle ($p_\varphi \ll 0.001$), but not for the elevation ($p_\vartheta = 0.200$). In contrast, the ICLD is a significant factor for the elevation ($p_\vartheta \ll 0.001$) and weakly for the azimuth angle ($p_\varphi = 0.070$) on array V. Furthermore, all neighboring ICLD conditions of array V yield significantly different elevation angles ($p_\vartheta \leq 0.027$). For array V the subjects are a highly significant factor ($p_\vartheta \ll 0.001$), whereas for array H they are not ($p_\varphi = 0.301$). For both arrays, the repetition is at least a weakly significant factor (array H: $p_\varphi \ll 0.001$, array V: $p_\vartheta = 0.080$).

For further investigation of these factors, intrasubjective and intersubjective standard deviations (STD) are examined. The intrasubjective STD is calculated from the two repetitions of each subject, individually. The intersubjective STD is calculated from the 15 individual medians. For an ICLD of $0\,\mathrm{dB}$ on array V, the intrasubjective STD of most subjects is at most half of the intersubjective STD. Only for one subject it is larger. Similar results are obtained for all other conditions.

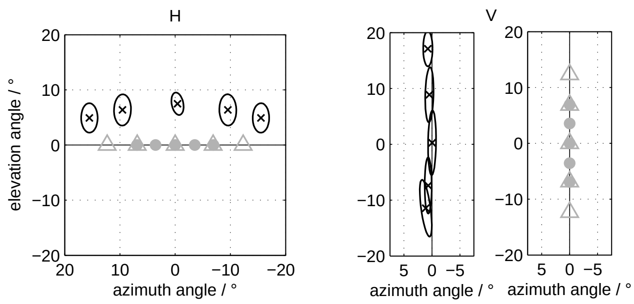

Figure 2: Mean (×) and 95% confidence areas (⬯) of localization and energy vector (△) and velocity vector (●) for array H and V.

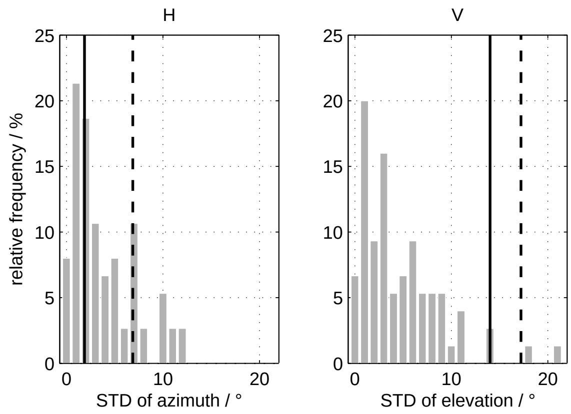

Figure 3: Histogram of the intrasubjective standard deviation (■) compared to the median of the intersubjective standard deviation (–) and the total standard deviation of all conditions and answers (**- -**) tested with array H and V.

The distribution of the intrasubjective STD is similar for the azimuth angle on array H and the elevation angle on array V, cf. **Figure 3**. However, the total STD of the azimuth angles is only half as for the elevation angles. This is because the median of the intersubjective standard deviation of array V is almost eight times larger than for array H. Thus, the larger STD of the elevation angles is dominated by the intersubjective STD, i.e. the differences between the subjects. This finding agrees with the results in [14, 10].

4.2. Loudspeaker Triangles

The 95% confidence areas do no overlap for both triangles T1 and T2, cf. **Figure 4**. Obviously, the ICLD is a highly significant factor for the elevation angle ($p_\vartheta \ll 0.001$). Interestingly, for the azimuth angle, it is also a significant factor in case of array T1 ($p_\varphi = 0.008$), but not in case of T2 ($p_\varphi = 0.306$).

Horizontal panning was applied on array T3, cf. **Figure 5**. As expected, the ICLD is a highly significant factor for the azimuth angle ($p_\varphi \ll 0.001$), whereas it is not for the elevation ($p_\vartheta = 0.184$). Despite the horizontal panning, the size of the 95% confidence areas of T3 is comparable to that of T1 and T2. Summarizing the results of all triangles, the number of active loudspeakers weakly affects the size of the confidence interval for the azimuth angle ($0.053 \leq p_\varphi \leq 0.081$) but not for the elevation angle ($p_\vartheta \geq 0.505$). Similar to array V, the subjects are a highly significant factor ($p_\vartheta \ll 0.001$ for T1 and T2, $p_\varphi = 0.008$ for T3) for all triangles.

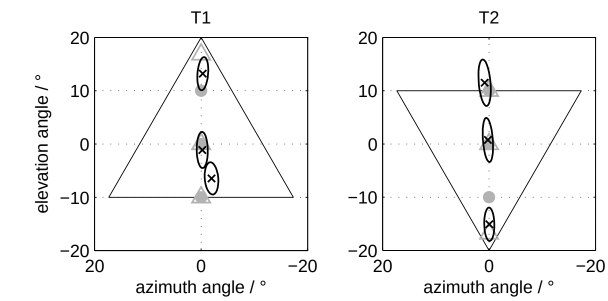

Figure 4: Mean (×) and 95% confidence areas (⬯) of localization and energy vector (△) and velocity vector (●) for array T1 and T2.

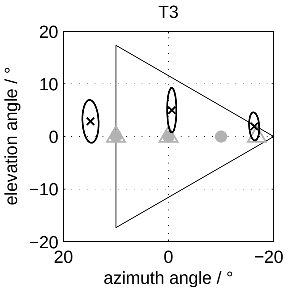

Figure 5: Mean (×) and 95% confidence areas (O) of localization and energy vector (△) and velocity vector (•) for array T3.

4.3. Loudspeaker Squares

The loudspeaker squares S1 and S2 were employed for both horizontal and vertical panning, cf. **Figure 6**.
As expected for horizontal panning on array S1, the ICLD is a highly significant factor for the azimuth ($p_\varphi \ll 0.001$) but not for the elevation angle ($p_\vartheta = 0.469$). Correspondingly, for vertical panning, the ICLD is a highly significant factor for the elevation ($p_\vartheta \ll 0.001$) and not significant for the azimuth angle ($p_\varphi = 0.843$).
Similarly to S1, horizontal panning on array S2 yields highly significant differences for the azimuth ($p_\varphi \ll 0.001$) and not for the elevation angle ($p_\vartheta = 0.612$). However, the ICLD of vertical panning is a (weakly) significant factor for both the elevation ($p_\vartheta \ll 0.001$) and the azimuth angle ($p_\varphi = 0.063$). For vertical panning on both square arrays, the subjects are a highly significant factor for the elevation angle ($p_\vartheta \ll 0.001$). In contrast, the azimuth angle of horizontal panning is only affected by the subjects on array S2 ($p_\varphi = 0.015$) and not on array S1 ($p_\varphi = 0.647$).

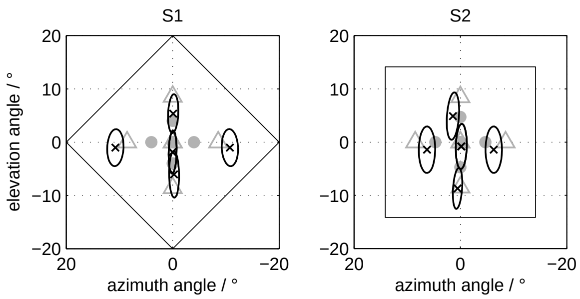

Figure 6: Mean (×) and 95% confidence areas (O) of localization and energy vector (△) and velocity vector (•) for array Q1 and Q2.

5. Localization Predictions

Localization models can be used to predict the localization of phantom sources. They are typically based on dummy head measurements or HRTFs and evaluate binaural differences for horizontal localization [18] or monaural spectral cues for vertical localization [9]. This section discusses simple vector-based predictors and employs them for the prediction of both azimuth and elevation angles of the phantom sources. For these predictors, the directions $\boldsymbol{\theta_l}$ are defined as vectors of unit length that depend on the azimuth φ and the elevation angle ϑ

$$\boldsymbol{\theta_l}(\varphi, \vartheta) = \begin{pmatrix} \cos(\varphi)\ \cos(\vartheta) \\ \sin(\varphi)\ \cos(\vartheta) \\ \sin(\vartheta) \end{pmatrix}. \quad (2)$$

5.1. Velocity Vector

The velocity vector is based on the summation of the L loudspeaker directions $\boldsymbol{\theta_l}$ weighted with g_l. It is defined as

$$\boldsymbol{r}_\mathrm{V} = \frac{\sum_{l=1}^{L} g_l \boldsymbol{\theta_l}}{\sum_{l=1}^{L} g_l}. \quad (3)$$

The direction of this vector is assumed to correspond to the localization of low frequencies (≤ 700Hz) [19].

5.2. Energy Vector

Following the idea of the velocity vector, the energy vector $\boldsymbol{r}_\mathrm{E}$ [19] is defined as

$$\boldsymbol{r}_\mathrm{E} = \frac{\sum_{l=1}^{L} g_l^2 \boldsymbol{\theta_l}}{\sum_{l=1}^{L} g_l^2}. \quad (4)$$

This model assumes an energetic superposition of the loudspeaker signals and is expected to model the localization direction of the higher frequencies or broadband signals.

Without limitation, both vector models can be calculated for arbitrary numbers of simultaneously active loudspeakers.

5.3. Prediction of the Experimental Results

Along with the experimental results, **Figures 2** and **4** to **6** show the different predictions. **Figures 7** summarizes the deviations of the predictions from the mean experimental results for each array. For both predictors, the deviations of the elevation angles are mostly larger than those of the azimuth angles. Interestingly, the largest deviations of the elevation angles can be found for the horizontal loudspeaker pair (array H). This elevation effect can be attributed to spectral characteristics of the horizontal phantom source and is investigated in [20].

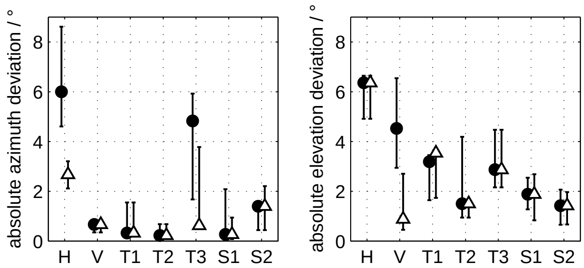

Figure 7: Median and 95% confidence interval of the absolute azimuth and elevation deviation of the velocity vector (•) and energy vector (△) from the mean experimental results for each array.

For all tested arrays, the direction of the energy vector matches the experimental results better than or at least equally good as the one of the velocity vector. Therefore a panning method assuming energetic superposition of the loudspeaker signals seems to be the most suitable technique.
With the good predictability of the experimental data by the energy vector, we finally have solid evidence for what has often only been assumed in the design of three-dimensional Ambisonic [2] amplitude panning and decoding systems.

6. Conclusion

This contribution investigated the horizontal and vertical localization of phantom sources generated by basic loudspeaker arrays composed of 2, 3, and 4 loudspeakers in a listening experiment. Along with the investigation of the phantom source location for all loudspeakers of an array playing equally loud, horizontal and vertical amplitude panning was examined. Besides horizontal panning, the experiment proved that vertical amplitude panning is possible. However, the standard deviation of the elevation angles is much larger than the one of the azimuth angles. Agreeing with the findings in [10, 14], an analysis of the results revealed the intrasubjective standard deviation, i.e. the difference between the subjects, to be the predominant factor for the large standard deviation of the elevation angles. In contrast, the intersubjective standard deviation, i.e. the difference between repetitions by the same subject, is similar for both azimuth and elevation angles.

The standard deviation was only weakly influenced by the number of loudspeakers. However, it is not assumed that this holds true for the perceived source width [21]. The investigation of phantom source width and colouration for three-dimensional amplitude panning is still an ongoing task.

The experimental results were compared to the velocity and energy vector as localization predictors. The latter yielded better predictions and is suitable as a simple predictor that solely depends on the loudspeaker positions and gains. Therefore, it seems reasonable for three-dimensional amplitude panning to employ methods that assume energetic superposition of the loudspeaker signals.

7. Acknowledgements

The authors thank all subjects for their participation in the listening experiment. 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 abovementioned institutions.

References

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

[2] M. A. Gerzon, "With-height sound reproduction," *Journal of the Audio Engineering Society*, vol. 21, pp. 2–10, 1973.

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

[4] Dolby Laboratories, "Dolby Atmos," `http://www.dolby.com/us/en/professional/technology/cinema/dolby-atmos.html`.

[5] Auro Technologies, "Auro 3D," `http://www.auro-technologies.com/system/concept`.

[6] K. Hamasaki, K. Hiyama, and R. Okumura, "The 22.2 multichannel sound system and its application," in *Audio Engineering Society Convention 118*, 2005.

[7] J. Blauert, *Spatial Hearing*. MIT Press, 1983.

[8] 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.

[9] E. H. A. Langendijk and A. W. Bronkhorst, "Contribution of spectral cues to human sound localization," *The Journal of the Acoustical Society of America*, vol. 112, no. 4, pp. 1583–1596, 2002.

[10] F. Wendt, M. Frank, and F. Zotter, "Application of localization models for vertical phantom sources," in *Fortschritte der Akustik, AIA-DAGA*, (Meran, Italy), 2013.

[11] T. Kimura and H. Ando, "Listening test for three-dimensional audio system based on multiple vertical panning," in *Acoustics Hong Kong*, 2012.

[12] S. Braun and M. Frank, "Localization of 3D Ambisonic Recordings and Ambisonic Virtual Sources," in *International Conference on Spatial Audio*, (Detmold, Germany), 2011.

[13] P. Bremen, M. M. van Wanrooij, and A. J. van Opstal, "Pinna cues determine orienting response modes to synchronous sounds in elevation," *The Journal of Neuroscience*, vol. 30, no. 1, pp. 194–204, 2010.

[14] 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.

[15] F. Wendt, "Untersuchung von Phantomschallquellen vertikaler Lautsprecheranordnungen," Master's thesis, University of Music and Performing Arts, Graz, Austria, 2013.

[16] ITU, "ITU-R BS.1116-1 methods for the subjective assessment of small impairments in audio systems including multichannel sound systems," tech. rep., ITU, 1994-1997.

[17] 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*, 2010.

[18] W. Lindemann, "Extension of a binaural cross-correlation model by contralateral inhibition. I. Simulation of lateralization for stationary signals," *J. Acoust. Soc. Am.*, vol. 80, no. 6, pp. 1608–1622, 1986.

[19] M. A. Gerzon, "General metatheory of auditory localisation," in *Audio Engineering Society Convention 92*, vol. 3, 1992.

[20] M. Frank, "Elevation of horizontal phantom sources," in *Fortschritte der Akustik, DAGA*, (Oldenburg, Germany), 2014.

[21] M. Frank, "Source width of frontal phantom sources: Perception, measurement, and modeling," *Archives of Acoustics*, vol. 38, no. 3, pp. 311–319, 2013.