Application of localization models for vertical phantom sources

Florian Wendt[1], Matthias Frank[2], Franz Zotter[2]

[1] *Graz University of Technology, 8010 Graz, Austria, Email: florian.wendt@student.tugraz.at*

[2] *Institute of Electronic Music and Acoustics, 8010 Graz, Austria, Email: {frank, zotter}@iem.at*

Introduction

Localization of real sound sources in the horizontal plane is well studied, and the use of inter-aural level and time differences (ILD and ITD) as localization cues is proved. For localization in vertical planes and especially in the median plane, such cues are lacking or completely absent. Nevertheless, the localization of real sound sources is possible and other cues such as spectral properties play an important role in contrast to inter-aural differences.
Previous studies showed that spectral properties are mainly caused by the pinna, which causes spectral coloration above 4kHz [5]. Based on this, localization models for median plane were developed and tested.
The principle of all models is the same: The spectral coloration of an incoming sound with known spectrum is compared with a database of head-related transfer functions (HRTFs) in order to find the most similar ones. The directions thereof estimate the localization. Listening experiments verify that this localization mechanism is not as accurate as horizontal localization [3].

A single phantom source is typically created by two loudspeakers playing the same signal. For the horizontal plane, well-known studies explain the phantom source localization in terms of ILD and ITD [10]. This is because the superposition of the HRTFs for the left and right loudspeaker yields inter-aural differences resembling those of a real source [10, 12].

Phantom sources created by a vertical pair of loudspeakers arranged in the median plane (see Fig. 1) largely deliver the correct, but indifferent, ILD and ITD cues. However, the superimposed HRTFs of the loudspeaker pair does not yield a coloration that is sufficiently similar to one of a real source. Still a few studies exist which imply that localization is possible. One of them shows that vertical phantom source localization is inter-subjective and therefore perceived individually [11]. Another study reports that phantom source localization in the vertical plane is almost as good as in the horizontal plane [7].
This article applies vertical localization models to examine vertical phantom source localization. In addition a localization test is realized to evaluate the examined localization models.

Localization models

All models can be divided into two stages. In the first stage the peripheral processing is done, simulating the transmission of sound from the ear canal to the inner hair cells for both ears. In the second stage, the template-based comparison with the HTRF database is done. The best matching HRTFs estimate the localization.

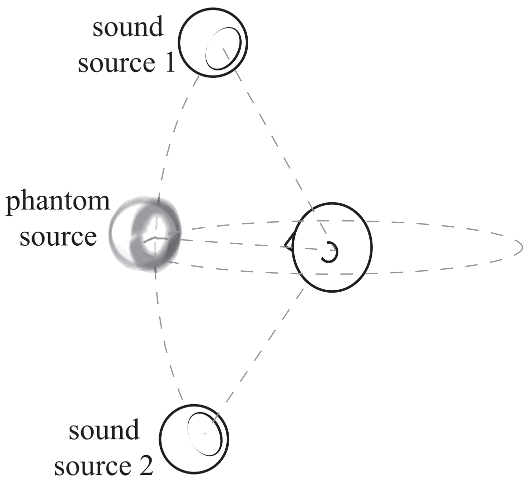

Figure 1: Does this vertical phantom source exist?

Iida/Blanco-Martin. Iida [5] analyzed the peaks and notches of the HRTFs and reported that artificial HRTFs that use the first and second notch (between 4kHz and 10kHz) and the first peak (between 7kHz and 9kHz) provide almost the same localization accuracy as the measured HRTFs. Based on this research, Blanco-Martin [2] introduced two models for median plane localization. The retrieval of notches in realistic binaural signals is not reliable in practice, thus the results can be inaccurate. To improve the results an alternative model is presented that is based on least square error minimization to the HRTFs of a database.

Langendijk/Bronkhorst. The third model examined is proposed by Langendijk and Bronkhorst [8]. It uses a differentiator which was inspired by Zakarauskas and Cynader [13]. The similarity of two HRTF spectra is examined by either their cross-correlation or their Euclidean distance.
The existing vertical localization models based on spectral properties have been verified with real sources [1,2,8].

The difference of the models lies in the decision stage. To ensure comparability the peripheral processing for all models is the same.

Model evaluation

All the presented analysis results use the freely accessible HRTF measurement database of the ARI (Austrian Research Institute, 2012). It contains measured HRTFs of 66 people. The elevation angle in the median plane is restricted to the interval from $\phi = -30°$ to $80°$ with a regular sampling of 5°, except for a gap between 70° and 80°.

We obtain the composite HRTF of a vertical loudspeaker pair playing the same signal at the elevation ±20° by summing the respective HRTFs from the database. The localization models were applied to this composite HRTF to observe whether there is a stable vertical phantom source. Moreover, the models are applied to the real sources at 0° elevation for comparison.

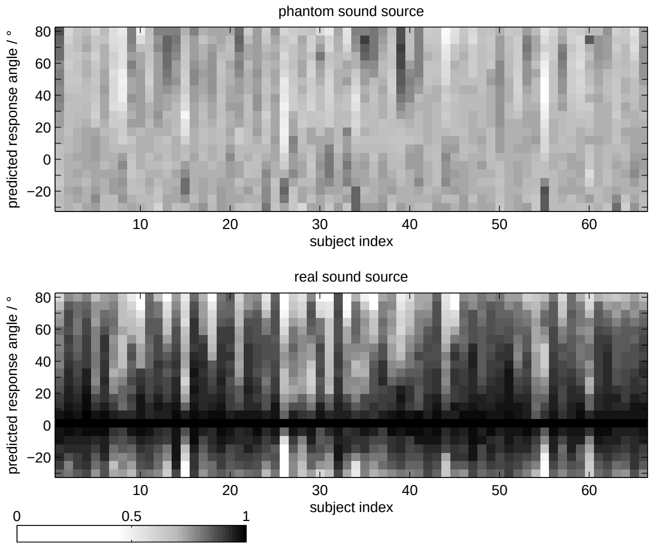

Figure 2: Localization model by Blanco-Martin [2] applied to a phantom source generated by vertical loudspeakers at ±20° and to a real source at 0° elevation.

Figure 2 shows the result for the least-squares model (Iida, Blanco-Martin). For the different HRTF datasets (subject index), this model neither yields a consistent nor a accurate vertical phantom source localization. As a reference, the localization of real sources works well and indicates a certain natural inaccuracy.
The second model introduced by Iida/Blanco-Martin (peak/notch-matching) provides similar results for phantom source localization, and therefore results are omitted.

Figure 3 shows the localization prediction using the model proposed by Langendijk/Bronkhorst. In contrast to the model above, there seems to be a phantom source at a pronounced location for most HRTF datasets, however with a great fluctuation in the range between ±20°.

Scalar elevation estimator

The model predictions have been compared to existing listening test results. Figure 3 shows that the model proposed by Langendijk/Bronhorst at least confirms Pulkki's statement in [11], according to which vertical phantom source localization is possible but highly subjective. To achieve better matching results, this model is modified in order to get one single response angle Φ for each subject. Lindemann [9] suggests to use the centroid of the inter-aural cross correlation function to indicate the horizontal location. Since the centroid did not deliver

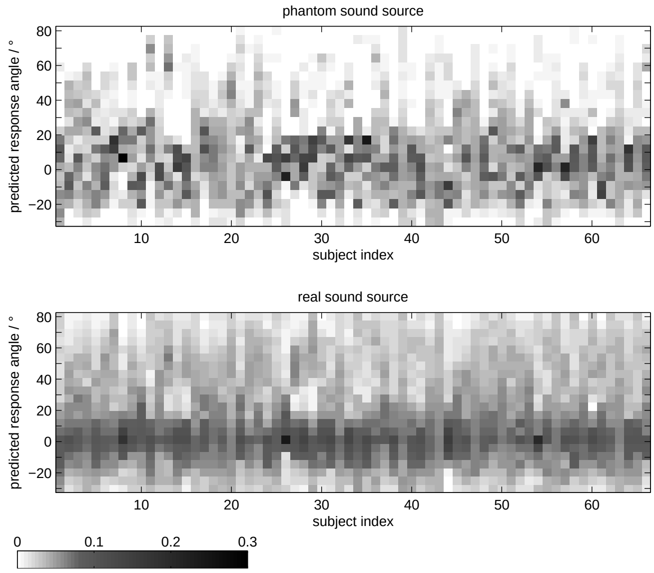

Figure 3: Localization model by Langedijk [8] applied to a phantom source generated by vertical loudspeakers at ±20° and to a real source at 0° elevation.

sufficient results we enhance Lindemann's suggestion by applying the power of n to the probability function p of the Langendijk/Bronkhorst model:

$$\Phi = \frac{\sum_{\phi=-30°}^{80°} \phi p^n}{\sum_{\phi=-30°}^{80°} p^n}. \tag{1}$$

The choice of the exponent n is done by feeding HRTFs with known elevation into the model. Good results were achieved for $n = 8$. Figure 4 shows the predicted localization for real sound sources at ±20° and a phantom source created by both of them playing the same signal.

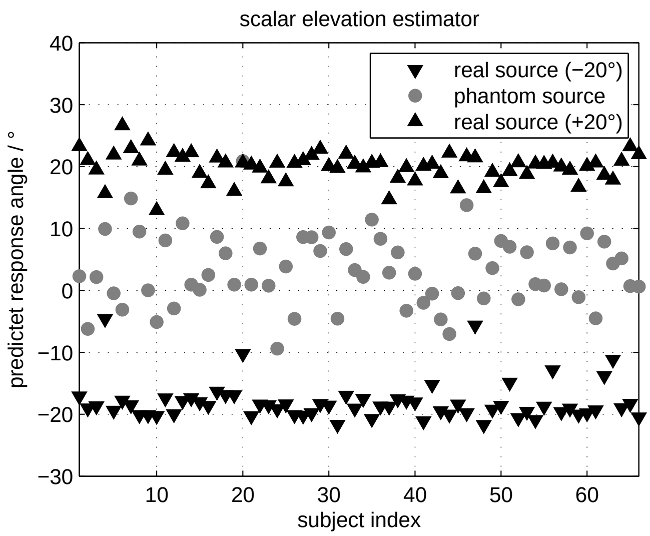

Figure 4: Predicted response angle Φ using the scalar elevation estimator applied to real sources at ±20° and to a phantom source generated by vertical loudspeakers at ±20°.

For all subjects, the estimated elevation of the phantom source lies between the estimated perceptual elevation angles of two single loudspeakers. In this case, it appears that phantom sources can be panned between the real sources.

Listening test

To confirm the existence of a panning law and to evaluate the enhanced model proposed by Langendijk [8], a listening test is performed. In the test, the subject was asked to evaluate the perceived elevation Φ of an amplitude panned phantom source generated by two vertically stacked loudspeakers at $\pm 20°$. The test setup is shown in figure 1. The loudspeakers (GENELEC 8020A) were placed at a 2.5m distance from the subject's head in the IEM CUBE, a 11m x 11m x 5m room with an average reverberation time $RT_{60} = 470$ms. Although with reference to ITU-R BS.1116-1 [6] the room is large, it is still within the recommended reverberation time limits. In addition, participants were seated within the effective critical distance of the setup. The loudspeaker setup was covered by an acoustically transparent screen. During the listening test, the subject was requested to face the 0° direction and to keep his or her head immobile. The perceived direction was recorded by pointing with a motion tracked toy-gun [4]. The position and orientation of the pointing device and the subjects head was captured. The resulting direction was stored when the subject pulled the trigger of the toy-gun. The listening test was carried out with 15 subjects. All of them are experienced listeners with normal hearing.
The test signal were pink noise bursts (3 x 300ms) at 65dB(A). The different amplitude panned phantom sources were created with 7 different inter-channel level differences (ICLD) and two repetitions for each stimulus, except of four repetitions for an ICLD of 0dB that are shown in figure 6. In total each subject had to evaluate 16 directions of sound images. The order of the stimulus playback was generated randomly, and the subjects could start the playback when they were ready, and they could repeat it at will.

Results

Localization curves are computed with the data generated by the listening test and the localization model. Figure 5 shows the comparison of the obtained curves.

Both curves are monotonically increasing over the ICLD. The localization curve obtained by the listening test exhibits a vertical offset for single loudspeakers and a saturation in both directions. Furthermore it shows a big spread in the range of small ICLDs.

The localization curve obtained from the enhanced localization model has just a slight offset, and in comparison to the listening test curve it has an almost linear progression and shows no saturation.
Even though the localization curves are not congruent, the scalar elevation estimator delivers a monotonic localization curve.

In addition the data obtained by the listening test was analyzed by ANOVA. Except for the neighboring conditions (-6dB, -3dB) and (6dB, ∞dB) all differences are at

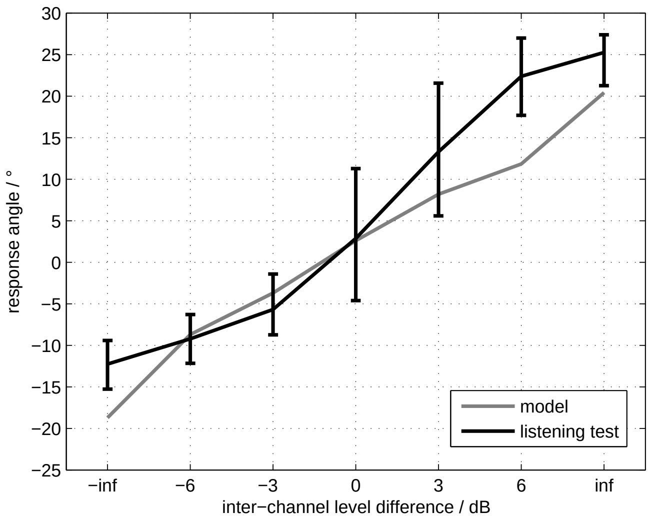

Figure 5: Comparison of localization curves extracted form listening test (medians and corresponding 95% confidence intervals) and localization model (medians) by Langendijk [8].

least weakly significant ($p \leq 0.085$).
Furthermore the intrasubjective and intersubjective standard deviation was computed. The intrasubjective standard deviation is computed from the four repetitions of each subject, individually. The intersubjective standard deviation is computed from the 15 individual medians. For 75% of the subjects the intrasubjective standard deviation is half of the intersubjective standard deviation or even less. Solely for subject 9 it is larger. This confirms Pulkki's statement [11] that elevation localization is intersubjective; see also figure 6.

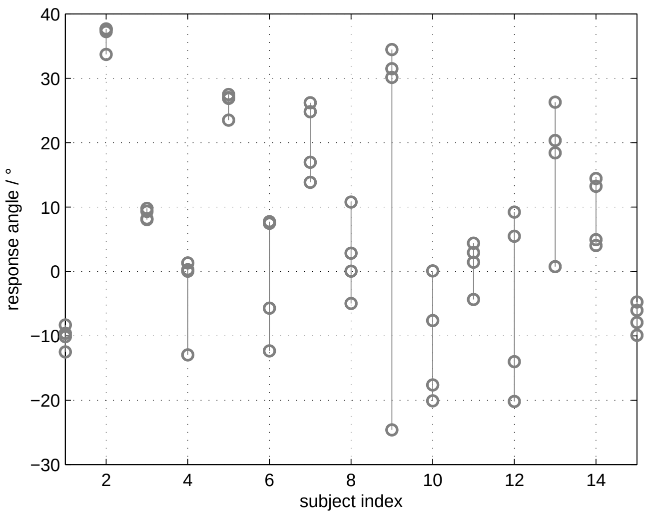

Figure 6: 4 vertical localization responses from each subject to the phantom source elevation generated by vertical loudspeakers at $\pm 20°$ with ICLD of 0dB

Conclusion

The model predictions have been compared to listening test results.
The model proposed by Iida/Blanco-Martin has been tested with other loudspeaker setups than $\pm 20°$ and dif-

ferent bands of the gammatone filterbank but it does not work for phantom source localization.
The model proposed by Langendijk/Bronkhorst seems to confirm Pulkki's statement in [11], according to which vertical phantom source localization is possible but highly subjective. A scalar elevation estimator was introduced, and a listening test was performed which partly confirms Kimura's results [7]. The estimator is able to predict vertical panned phantom source elevation although further enhancements and fine tuning is necessary to ensure a stable prediction.

Acknowledgments

We thank all subjects for their participation in the listening test.

References

[1] Robert Baumgartner. Modeling sagittal-plane sound localization with the application to subband-encoded head-related transfer functions. Master's thesis, Universität für Musik und darstellende Kunst Graz, 2012.

[2] Elena Blanco-Martin et al. Objective measurement of sound event localization in horizontal and median planes. *J. Audio Eng. Soc*, 59(3):124–136, 2011.

[3] Jens Blauert. *Spatial hearing - The psychophysics of human sound localization*. The MIT Press, 1999.

[4] Matthias Frank et al. Flexible and intuitive pointing method for 3-d auditory localization experiments. In *Audio Engineering Society Conference: 38th International Conference: Sound Quality Evaluation*, 6 2010.

[5] K. Iida. Estimation of sound source elevation by extracting the vertical localization cues from binaural signals. In *Proceedings of Meetings on Acoustics*, volume 4, page 050002, 2008.

[6] ITU. ITU-R BS.1116-1 methods for the subjective assessment of small impairments in audio systems including multichannel sound systems. Technical report, ITU, 1994-1997.

[7] Toshiyuki Kimura and Hiroshi Ando. Listening test for three-dimensional audio system based on multiple vertical panning. In *Acoustics Hong Kong*, 2012.

[8] Erno H. A. Langendijk and Adelbert W. Bronkhorst. Contribution of spectral cues to human sound localization. *The Journal of the Acoustical Society of America*, 112(4):1583–1596, 2002.

[9] W. Lindemann. Extension of a binaural cross-correlation model by contralateral inhibition. i. simulation of lateralization for stationary signals. *J. Acoust. Soc. Am.*, 80:1608, 1986.

[10] V. Pulkki and M. Karjalainen. Localization of Amplitude-Panned Virtual Sources I: Stereophonic Panning. *J. Audio Eng. Soc*, 49(9):739–752, 2001.

[11] V. Pulkki and M. Karjalainen. Localization of Amplitude-Panned Virtual Sources II: Two- and Three-Dimensional Panning. *J. Audio Eng. Soc*, 49(9):753–767, 2001.

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

[13] Pierre Zakarauskas and Max S. Cynader. A computational theory of spectral cue localization. *The Journal of the Acoustical Society of America*, 94(3):1323–1331, 1993.