The Influence of the Floor Reflection on the Perception of Sound Elevation

Florian Wendt, Robert Höldrich, Matthias Frank
University of Music and Performing Arts Graz, Austria
Institute of Electronic Music and Acoustics, Email: {wendt, hoeldrich, frank}@iem.at

Introduction

Auditory perception of sound sources includes two principal localization subsystems. Horizontal localization cues comprise inter-aural time and level differences extracted mainly from the direct sound [1]. Vertical localization cues on the other hand are not as well understood. One cue that we clearly do use, however, is the pattern of peaks and valleys in the spectrum of a broadband sound mainly caused by the characteristics of the pinna [2].
In real-life situations the direct sound of an emitting source is typically followed by a strong floor reflection. The physical elevation of the source determines the delay time and level of this reflection providing additional information about the elevation of the source.
This contribution focuses on the influence of the floor reflection on the perceived height of a sound source. We sketch a listening experiment inspired by real-life listening situations. Gained results are compared to an existing localization model finding evidence that, additional to spectral cues, temporal information is evaluated for assessing the elevation of a sound source.

Experimental Setup

Starting point for this work is a study presented by Guski in [3], focusing on the influence of reflections from floor, ceiling, and side walls on the localization. The study could show that the addition of a floor reflector in an anechoic environment reduces the vertical localization error of a speech signal significantly. To determine what relevant information of the floor reflection is evaluated by the auditory system, we performed a listening experiment that varied both time delay and level of the reflection.
The experimental layout was similar to [3] and consisted of six vertically arranged Genelec 8020A loudspeakers set up in our anechoic laboratory. Figure 1 shows a sketch of the setup. Loudspeakers LS1. . . LS5 were equally distributed between $\varphi = 20° \ldots -20°$ in the median plane and covered by an acoustically transparent screen, cf. Figure 2. The floor reflections were simulated as image sources using LS6. The small directional mismatches between LS6 and the actual image source was assumed to be negligible. All loudspeakers were level- and delay-compensated to the central listening position.

Table 1: Time delay and level decrease ($\Delta T[ms]/\Delta L[dB]$) of floor reflection (LS6) for reference and test conditions.

i	-10	-5	0	5	10
A_i	1.1/2.6	1.6/3.1	2.1/3.6	2.6/4.0	3.0/4.4
B_i	2.1/3.6	2.6/4.0	3.0/4.4	3.5/4.8	3.9/5.1
C_i	3.0/4.4	3.5/4.8	3.9/5.1	4.3/5.4	4.7/5.7

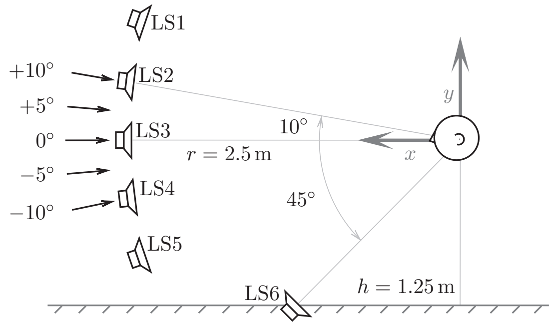

Figure 1: Sketch of the loudspeaker setup. Black arrows exemplary indicate simulated shifts of the emitting source whose specular reflections were tested together with direct sound played back by LS3 yielding conditions $B_{10,5,\ldots,-10}$.

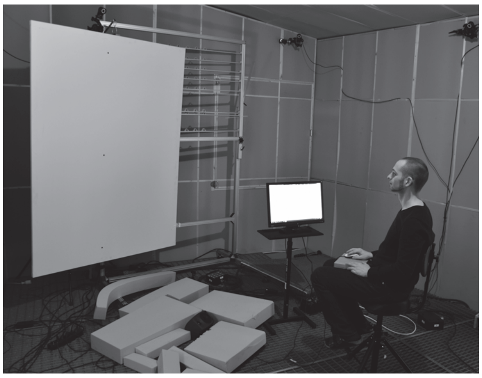

Figure 2: Experimental setup in the anechoic laboratory.

Tested conditions were based on three references, denoted as A_0, B_0, and C_0 (see Table 1), each consisting of direct sound emitted from one of the loudspeakers LS4, LS3, or LS2 ($\varphi = [-10°, 0°, 10°]$), and the corresponding floor reflection played back by LS6. Time delay and level of the reflection corresponded to the respective image source calculated with a frequency-independent absorption coefficient of $\alpha = 0.15$.
For each reference, four additional test conditions were created by varying level and time delay of the floor reflection. The variations represented floor reflections from sources shifted in elevation by $\pm[5°, 10°]$ with respect to the reference condition. This shift is indicated by the index i of each condition, cf. Table 1. For example, condition C_{10} is composed of direct sound emitted by LS2 and a floor reflection evoked by an emitting source at $\varphi = 20°$. Each of the sets A, B, and C was completed by an upper and lower anchor at ±10° relative to their reference. Time delay and level of their floor reflections corresponded to the respective image source. This yields 3 sets × $(1 + 4 + 2)$ conditions.

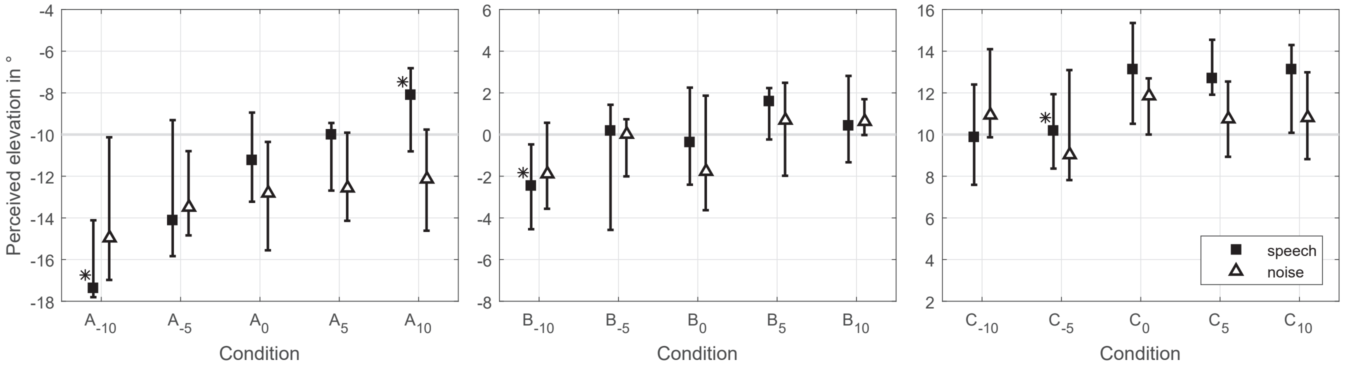

Figure 3: Median and corresponding 95% confidence interval of responses for the three sets. Conditions that are significantly different from the respective reference are marked with an asterisk.

Tested sounds were female speech (taken from CD *B&O 101*, 1992) and a 1-second-long pink noise burst (onset and release time of 10 ms) followed by a pause of the same length. Samples were played back in loop at comfortable level of 70 dB(A) and could be repeated at will.

In a preliminary experiment, listeners were asked to indicate the perceived elevation of each condition presented in random order using a pointing device [4]. Apart from the anchors, there were no significant differences between the conditions in each set. However, informal listening by the authors indicated an effect of the varying floor reflection within each set, when comparing only the seven conditions of the set.
Therefore, the final experiment was conducted as a MUSHRA-like procedure [5] for each set and sound as separate trials. The listeners' task was to indicate the perceived height of each randomly ordered sample of the set with a continuous slider on a relative scale. Eight listeners $S_{1...8}$ (all male; age $29-56$ years) participated in this experiment. All of them were staff of the IEM and experienced listeners with normal hearing. Except for S_5, listeners performed two runs.

Experimental Results

The lower anchor was always perceived as the least elevated and the upper anchor as the most elevated condition within each set, except for one out of 90 response sets (3 test sets × 2 signals × 15 runs). This allows us to map the relatively-scaled responses to elevation angles by scaling them with the actual angles of the anchors for each response set. Figure 3 shows medians and corresponding 95% confidence intervals of responses for reference and test conditions of test sets A, B, and C.

By comparing test conditions with the respective reference conditions, we see an impact of the floor reflection on the perceived elevation. Interestingly, listeners were more sensitive for speech then for noise indicating, that in addition to the spectrum, temporal cues are used for assessing elevation.

The medians reveal that the largest shift was achieved for set A (direct sound at $-10°$), showing a monotone dependency for both speech and noise. Similar dependency for elevation of direct sound was found in [3], showing that a loudspeaker at $-12°$ benefits most (compared to 0° and 12°) when a floor reflector is added.

The minimum separation between two sources that can be reliably detected is indicated by the minimum audible angle (MAA). For horizontal separation, the threshold is known to be in the range of 1°, whereas for vertical localization it inter-subjectively ranges from 2° to 5° [6]. An analysis of variance (ANOVA) of set A revealed two test conditions to be (weakly) significantly different from the reference ($p_{0/-10} \ll 0.01$, $p_{0/10} = 0.10$). The shift of respective medians to the median of the reference condition corresponds to one MAA threshold. Similarly, one different test condition was found for set B ($p_{0/-10} = 0.05$) and one for set C ($p_{0/-5} = 0.10$) with shifts in the range of one MAA threshold. For noise no significance was observed.

For an evaluation of a combined data set for each sound, summarizing all sets A, B, and C, each set was normalized to the median of its reference condition. Responses of the combined and normalized data are shown in Figure 4. Medians of both signals yield monotone curves. However, only for speech an ANOVA revealed two significant steps ($p_{-10/-5} = 0.05$, $p_{-5/0} = 0.05$).

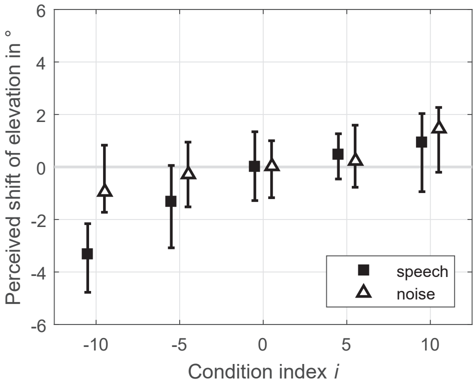

Figure 4: Median and corresponding 95% confidence interval of the combined data set for reference and test conditions.

The individual thresholds for MAA found in [6] indicate subjective differences of vertical localization. Further hints thereof are reported in studies on the perception of phantom sources created by loudspeakers with vertical aperture angle [7, 8]. To obtain more insights, individual medians of the combined data set are shown in Figure 5.

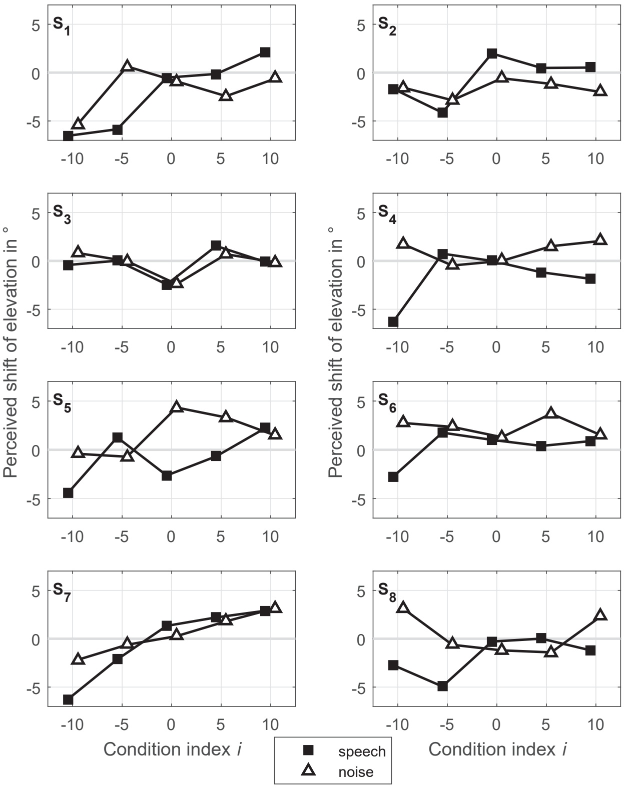

Figure 5: Individual median values of the combined data set for reference and test conditions.

Interestingly, medians of the speech signal show a jump in the perception of elevation between two neighboring conditions for each listener. The median increase of this jump is 5.0° (min. 4.0°, max. 7.0°) and occurs inter-subjectively between two different conditions. Although for noise, some listeners show similar increases between two neighboring conditions the overall jump size is significantly smaller (median 2.7°).

Modeling the Perceived Height

The question is whether the responses can be explained by models of sagittal-plane localization. These functional models are based on spectral shape cues assuming that listeners create an internal template set from their specific head-related transfer functions (HRTFs) as a result of a monaural learning process. Obviously, when using HRTFs, only the spectral information of the direct sound is considered.

The incoming sound, simulated by a weighted superposition of HRTFs for direct sound and floor reflection, is compared with a database of HRTFs in order to predict the perceived elevation.

The model proposed by Langendijk and Bronkhorst [2] demonstrated good performance, e.g. [9, 10], and is used for predicting our results.

Expecting relatively small changes in elevation, an HRTF set with high angular sampling in the range of the MAA (at least 5°) is required. Moreover, it needs to sample the angle of the floor reflection. The freely accessible HRTF measurement database of the Austrian Research Institute[1] complies with the first requirement, however the low end of the measurement range is limited. Thus, floor reflections were simulated at $\varphi = -30°$.

The model's predictions clearly distinguished the elevation of anchors and reference conditions. However, it could not predict any significant differences between the test conditions and references. Therefore we can conclude that a pure spectral cue is not sufficient to explain the results of our experiment.

Conclusion

We investigated the influence of level and delay of a simulated floor reflection on the perception of height. Two different signals were tested to determine relevant information for the auditory system. Broadband noise did not show any significant influence of the floor reflection and obtained results agree with predictions of a vertical localization model. For speech signals, on the other hand, the floor reflection contributes to the perception of height. Indications are found implying that, in addition to the spectrum, temporal cues extracted from the fine structure of a signal are evaluated. Further research is needed to get more insights to what extend temporal cues are used and what role the fine structure plays.

Acknowledgments

Our research was partly funded by the Austrian Science Fund (FWF) project nr. AR 328-G21, Orchestrating Space by Icosahedral Loudspeaker (OSIL). We thank all listeners taking part in the experiment.

References

[1] J. Blauert, *Spatial hearing - the psychophysics of human sound source localization.* The MIT Press, 1983.

[2] 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, p. 1583, oct 2002.

[3] R. Guski, “Auditory localization: effects of reflecting surfaces.,” *Perception*, vol. 19, no. 6, pp. 819–830, 1990.

[4] M. Frank, L. Mohr, A. Sontacchi, and F. Zotter, “Flexible and intuitive pointing method for 3D auditory localization experiments,” in *Proceedings of the 38th AES International Conference, Piteå, Sweden*, 2010.

---

[1] available at `www.kfs.oeaw.ac.at/index.php?lang=eng`

[5] ITU-R Recommendation, "Method for the Subjective Assessment of Intermediate Sound Quality (MUSHRA)," 2001.

[6] D. R. Perrott and K. Saberi, "Minimum audible angle thresholds for sources varying in both elevation and azimuth.," *The Journal of the Acoustical Society of America*, vol. 87, no. 4, pp. 1728–1731, 1990.

[7] V. Pulkki, "Localization of Amplitude-Panned Virtual Sources II: Two- and Three-Dimensional Panning," *Journal of the Audio Engineering Society*, vol. 49, no. 9, pp. 753–767, 2001.

[8] F. Wendt, M. Frank, and F. Zotter, "Panning with height on 2, 3, and 4 loudspeakers," in *2nd International Conference on Spatial Audio (ICSA)*, 2014.

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

[10] M. Frank, "Elevation of Horizontal Phantom Sources Vertical Localization Model," in *Fortschritte der Akusitk*, 2014.