

Investigation of Auditory Objects Caused by Directional Sound Sources in Rooms

F. ZOTTER* AND M. FRANK

Institute of Electronic Music and Acoustics, University of Music and Performing Arts Graz,
Inffeldgasse 10/3, 8010 Graz, Austria

The auditory impression of sound sources is strongly influenced by the room, which, e.g., determines the apparent source width. What is more, typical sources are not omnidirectional, which also makes their orientation a strong influence. This influence, however, has only been investigated a little, although it can even change the perceived location of the source. To provide more insight, we performed extensive listening experiments inside our anechoic laboratory that is equipped with a 24-channel loudspeaker playback to simulate both the directional source and the room. The directional source is described by two frequency-independent 3rd order directivity designs in 36 different orientations, and the room is simulated by the two-dimensional 1st and 2nd order image source method. Results of the experiment indicate that, in most cases, the auditory location can be determined by the loudest unmasked acoustic reflection path. This allows to explain the primary direction perceived with an astonishingly simple model including precedence effects.

DOI: [10.12693/APhysPolA.128.A-5](https://doi.org/10.12693/APhysPolA.128.A-5)

PACS: 43.55.Hy, 43.66.Qp, 43.30.Zk

1. Introduction

In a room, what we hear of a source arrives at our ears on various acoustic propagation paths (direct sound, early reflections, diffuse sound). The signal carried by each of these paths is individual in strength, direction of radiation, direction of arrival, arrival time, and coloration. The direction-dependent sound radiated by an instrument, speech, or any other acoustic source will therefore affect what is being heard or received [1]. A natural modification occurs whenever a sound source changes its orientation, cf. [2]. Some articles describe the perceived direction of a sound source in a room. Most notably the paper [3] gives a threshold above which propagation paths preceded by others change the perceived direction. A description of how the perceived direction is affected in such cases appears to be less known, experimentally. Our particular application of sound playback with sources of adjustable directivity requires investigation of this question in greater depth, to enable reliable control of the auditory objects caused.

Compact spherical loudspeaker arrays are typically used as neutral source approximating omnidirectional sound radiation in room acoustic measurements, e.g. [4–7]. More recently, spherical arrays with independently controlled transducers were used as sources featuring adjustable directivity [8–12]. Spatial computer music applied such sources to produce focused beams of sound, e.g. [13]. When used in a room, a source emitting sound beams of adjustable direction achieves perceptual effects that can be utilized musically. In particular, these auditory objects can be controlled to be perceived at other

locations than the compact spherical loudspeaker array itself. This is known to be the case whenever the level of a reflected sound exceeds the echo threshold of the preceding sound. The knowledge of the localization itself is still rough and can often only be guessed from what is known [2, 3, 14–16]. A more specific but still coarse description of the perceived direction was given in [13, 17]. Below, the first extensive investigation through a listening experiment is described, whose greater detail brings more insight but also raises further questions.

Although our laboratories employ an existing icosahedral loudspeaker array for concerts in large halls, one of our general concerns with the usefulness of experiments was reproducibility. It therefore seemed neither useful to employ this specific icosahedral loudspeaker array in the experiment, with its unique acoustical properties and customized directivity control algorithm, nor did it seem to be useful to perform experiments in a unique room acoustical setting that might be hard to re-build.

2. Experiment

The localization experiment described here considers a directional source with a frequency-independent third-order response pattern with its horizontal aiming varied in 10° steps from 0° to 350° .

The auralization of the room is simplistic, which was chosen above all reasoning in order to provide easy reproducibility of the experiment. The setting only employs horizontal directions to enable auralization by a horizontal loudspeaker ring in an anechoic chamber. The room is shoebox-shaped with ideally rigid walls, and its model is restricted to direct sound, 1st, and 2nd order reflections, which makes it easy to model using the image source method [18].

*corresponding author; e-mail: zotter@iem.at

2.1. Setup

A ring of 24 Genelec 8020 loudspeakers (15° inter-loudspeaker spacing) with a radius 1.5 m was placed in the anechoic chamber at ear height of the subject (1.25 m) sitting in the center of the arrangement.

Based on impulse responses calculated in MATLAB, auralization of the room model in all its conditions was done in the open source software pure data[†] on a standard PC with RME MADI audio interface and DirectOut D/A converters.

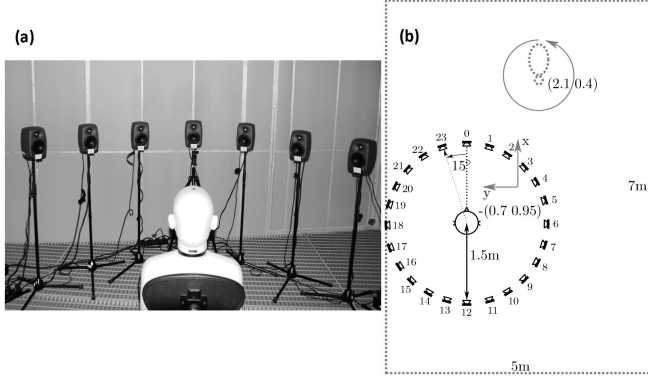


Fig. 1. (a) Experimental setup in anechoic chamber (dummy head not used for experiment but documentation thereafter); (b) auralization scenario of a shoebox room with a directional source.

The geometry of the auralized room and positions of the source and the listener are shown in Fig. 1. Positions and room dimensions were chosen as to provide a reasonably large angular and temporal separation in the sound propagation paths arriving at the listener. The Table shows the 0th, 1st, and 2nd order propagation paths of the setting, their length r_i , their delay with regard to the direct sound $\tau_i - \tau_0$, their radiation angle $\varphi_{S,i}$ at the source, their ideal incidence angle $\varphi_{R,i}$, and the matching loudspeaker direction $\varphi_{l[i]}$ and index $l[i]$ employed to auralize the propagation path in the experiment. During the experiment, auralization was switched between different conditions and fed by a test signal in real time. The test sequence for every subject was an individual random permutation of the entire set of conditions (including repeated conditions).

2.2. Conditions

The source directivity was varied in 36 orientations φ_{ori} from 0° to 350° in 10° steps. Figure 2 shows two different directivity pattern designs used for the auralized source.

Both of the 2 directivity patterns of the order $N = 3$ source were obtained by weighted superposition of Legendre polynomials

[†]Freely available on <http://puredata.info/downloads>.

TABLE

Distances, time delays, radiation and receiving angles, and simulated receiving angles using loudspeakers (horizontal lines distinguish between sound incidence of 0th, 1st, and 2nd order).

r_i	$\tau_i - \tau_0$	$\varphi_{S,i}$	$\varphi_{R,i}$	$\varphi_{l[i]}$	$l[i]$
3.1 m	0.0 ms	154°	-26°	-30°	2
5.3 m	6.3 ms	122°	58°	60°	20
5.8 m	7.8 ms	14°	-14°	-15°	1
6.2 m	9.1 ms	-117°	-63°	-60°	4
8.5 m	15.8 ms	171°	-171°	-165°	11
7.2 m	11.8 ms	38°	38°	45°	21
7.9 m	13.9 ms	-45°	-45°	-45°	3
9.1 m	17.5 ms	-108°	72°	75°	19
9.5 m	18.7 ms	152°	152°	150°	14
10.1 m	20.3 ms	-147°	-147°	-150°	10
11.3 m	23.9 ms	7°	-173°	180°	12
11.7 m	25.0 ms	104°	-76°	-75°	5
16.9 m	40.1 ms	175°	-5°	0°	0

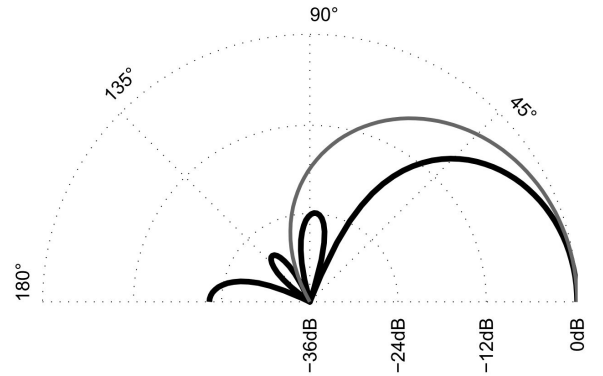


Fig. 2. Two third-order polar pattern designs, in-phase (gray) and max-rE (black), were employed to simulate the auralized source under step-wise varying orientation.

$$g_S(\varphi) = \sum_{n=0}^N (2n+1) a_n P_n(\cos \varphi), \quad (1)$$

using normalized weights $a_n = \hat{a}_n / \left[\sum_{n=0}^N (2n+1) \hat{a}_n \right]$. The two directivity pattern designs \hat{a}_n are described in [19, 20],

$$\hat{a}_n^{\text{inphase}} = \frac{N!(N+1)!}{(N+n+1)!(N-n)!}, \quad (2)$$

$$\hat{a}_n^{\text{maxre}} = P_n \left[\cos \left(\frac{137.9^\circ}{N+1.51} \right) \right]. \quad (3)$$

The impulse response of the l^{th} loudspeaker index and n^{th} time index uses the time and angle sampling intervals $T = 1/(44.1 \text{ kHz})$ and $A = 360^\circ/24$. With the Kronecker delta δ_{ij} and an integer rounding operation $\text{int}\{\}$, the impulse response yields, depending on g_S , φ_{ori} , and paths in Table,

$$h_l[n] = \sum_i \frac{g_S(\varphi_{S,i} - \varphi_{\text{ori}})}{r_i} \delta_{n, \text{int}(\tau_i/T)} \delta_{l, \text{int}(\varphi_{R,i}/A)}. \quad (4)$$

Each impulse response set was normalized to its RMS value $\sqrt{\sum_{l,n} h_l^2[n]}$ to provide constant loudness.

The *image source method* was used in *2 orders*, up to 1st order reflections, and up to 2nd order, for comparison of the simplistic auralization method.

All conditions were presented *2 times*, i.e., including one *repetition*, to enable checking the reliability of each subject.

The sound fed into auralization consisted of a sequence of three random pink noise bursts with 200 ms of silence in between. The envelope for each burst was deterministic and consisted of a 10 ms linearly rising onset, a 500 ms constant sustain, and a 10 ms linearly falling decay, cf. Fig. 3.

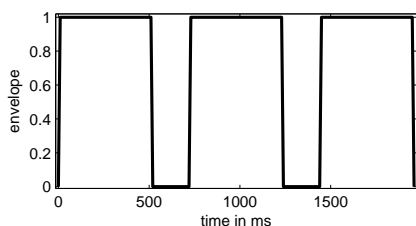


Fig. 3. Stimulus envelope.

11 subjects (staff of the lab and one audio engineering master student, min./avg./max. age: 26/35/52 years) took part in the experiment. We expect both reliable and representative results even if the number of subjects is small[‡], because all of them gained professional listening experience in 3D audio. Subjects did not report any hearing problems, and for 6 of them (min./avg./max. age 28/34/52 years), audibility threshold measurements were available that confirmed normal hearing (≤ 15 dB HL; among other screening tests for expert listening panels [23, 24]).

At the beginning of the experiment, each subject was given a short accommodation phase to familiarize to the variety of the stimuli. After this, the subject was requested to face loudspeaker 0 (0°) during listening; there was no head fixation. The subject was asked to name the number label on the loudspeaker, see Fig. 1, or half steps in between, that best indicates the direction at which the sound was perceived. In cases where there was a primary and a secondary localization percept, the subject should first name the primary localization followed by the secondary one. To record the subject's responses, monitor the experiment, and to switch to the next trial, the conductor of the experiment was sitting outside the

loudspeaker ring in the corner of the anechoic chamber, without disturbing the synthesized sound. Subjects could ask the conductor for repetition of the present trial in case they were uncertain about their response.

Each subject had to localize 36 orientations \times 2 patterns \times 2 image source orders \times 2 runs = 288 trials in total, which were all presented in a randomly permuted sequence that was unique for every subject. It took the average subject 39 min to finish the experiment.

2.3. Results

On average, in 66% of the responses (min. 56%, max. 80%), repetitions were localized idenetically in terms of their primary direction. In 3% (min. 1%, max. 6%) of the responses, only the repeated localization of the secondary direction stayed the same, and in 6% (min. 0%, max. 11%), exactly the same direction pair was named with interchanged primary and secondary direction. In the remaining 24% (min. 10%, max. 34%) of all responses, the mean directional deviation was 12° (min. 8° , max. 19°).

The correlation of each subject's primary direction to the median primary direction of the entire experiment was 86% on average (min. 70%, max. 96%).

Together with the experimental results shown in terms of histograms in Figs. 4 and 5, the time delay of each propagation path is shown above the diagram and the orientation-dependent level of the propagation path auralized by a loudspeaker level is represented by the gray shading in the background. It is obvious that for all source orientations, image source models, and directivity pattern designs, the responses of the subjects coincide at large with active propagation paths.

It is noticeable that of all the simultaneously present propagation paths, only distinguished ones will yield primary or secondary localization. For instance at the orientation 180° , it is clearly visible that the direct sound dominates localization, and other reflection paths are nearly absent in the localization histogram. The transitions between different paths is moreover remarkably determined either by precedence, e.g. for right and back wall reflections at the -150° orientation, or localization comparable to two-channel stereophony as, e.g., for the -40° orientation in Figs. 4 and 5.

In the results for the 1st order image source model, the correlation of median primary directions using max-rE and in-phase directivity patterns is 0.80. Regarding the localization histograms in Fig. 4, differences between the directivity patterns narrow down to the localization of side lobes of max-rE at the -40° orientation in Fig. 4 and a permutation of secondary and primary directions around the 90° source orientation. This holds for the 2nd order image source model in Fig. 5, where the correlation between max-rE and in-phase results is 0.86. Comparing the histograms of both image source models Figs. 4 and 5, the 2nd order model yields an increased number of secondary directions, which occasionally match primary directions of

[‡]According to literature [21, 22], experiments with native listeners tend to deliver similar results as with expert listeners, only that a number of required expert listeners can stay around ≥ 10 or use half the number of subjects.

the 1st order model, and a smoother transition between front/back/left directions due to additional reflections. Despite the correlation between the median primary

directions of the 1st and 2nd order models is reduced (max-rE: 0.54, in-phase: 0.71), their results include more or less the same tendencies.

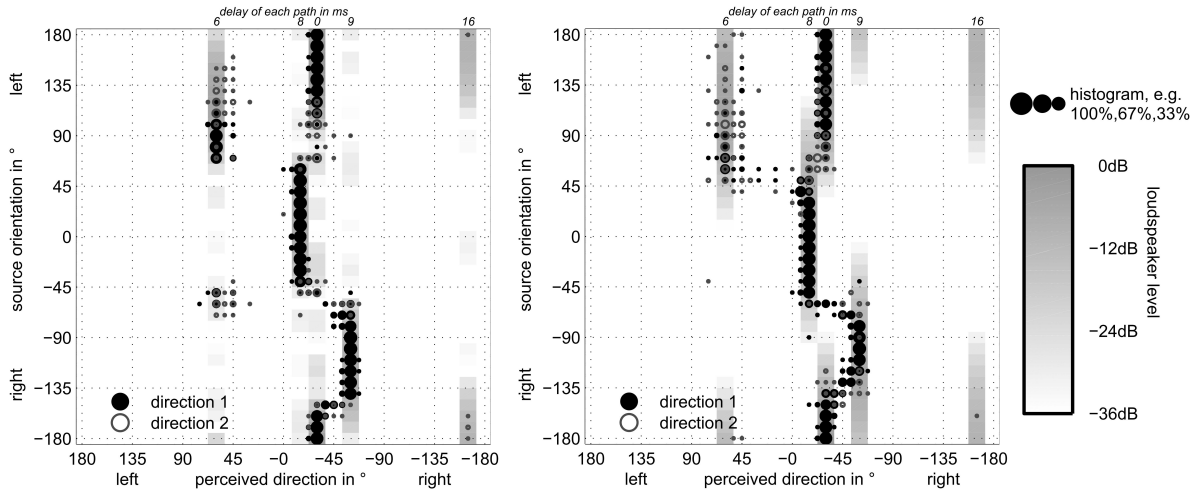


Fig. 4. Listening experiment results for 1st order image source model; radii represent the number of answers for each direction; level of each propagation path is represented by the gray shading in the background.

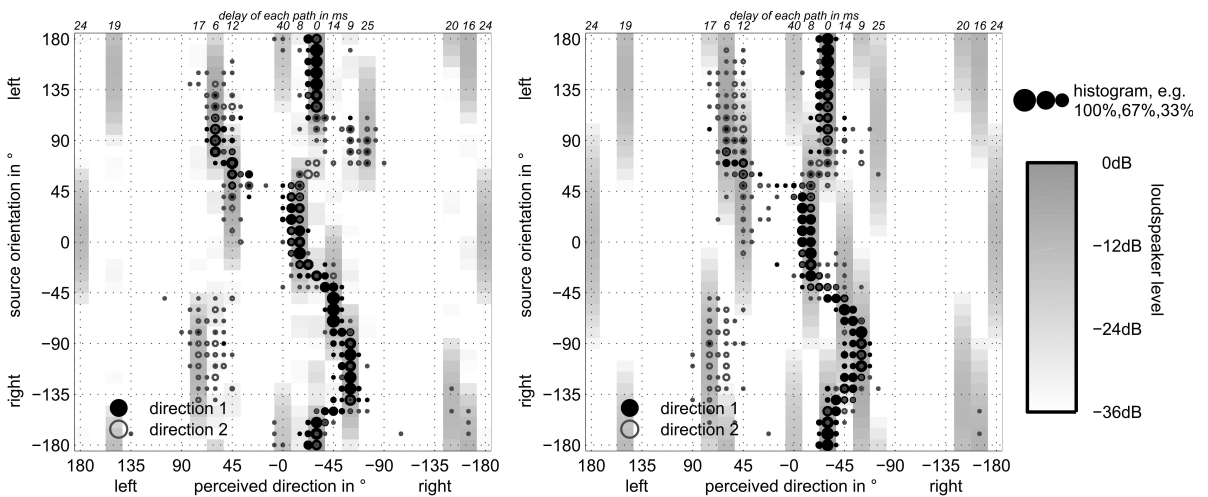


Fig. 5. Listening experiment results for 2nd order image source model; radii represent the number of answers for each direction; level of each propagation path is represented by the gray shading in the background.

Aside from the results above and in addition to what was asked, the majority of the subjects reported that they had difficulties to decide in cases sharing the same perceived direction but appearing to have different source widths. In some cases with twofold localization, it was difficult to sort primary and secondary direction. Three subjects reported that secondary directions sometimes coincide with the decay of the stimulus, and that localization might be threefold in seldom cases.

3. Modeling the primary localization

The above results raise questions of prediction and interpretation. To allow predictions, a simple model of the

auditory localization is desirable. A model that predicts all of the results appears to be difficult to establish, in particular when it comes to distinguishing between conditions with and without secondary localization. As reflection paths are differently delayed in time and arrive from quite different directions, a full model would need to consider delay time and arrival direction differences to predict when splitting and merging of the directional impression happens. To maintain simplicity of the investigation, only the primary localization is modeled here.

An energy interpretation was successful in the work of Robinson et al. on localization of diffuse reflections [15]

and in [25] to model primary auditory events caused by multiple loudspeakers. Consequently, a vector approach was also used in [17] to model the orientation-dependent localization of directional sound sources in a room, which we largely adopt here.

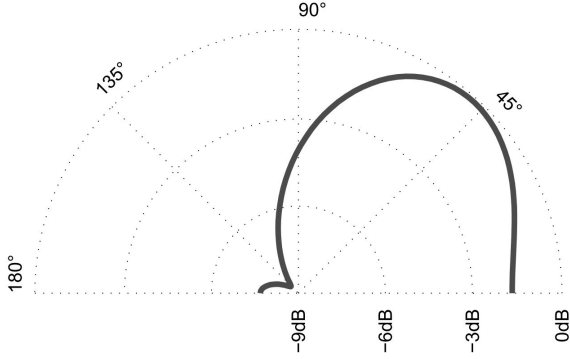


Fig. 6. Frequency-independent model of the ear directivity according to [17].

The vector model estimates the localization as weighted sum of direction-of-arrival vectors taking into account ear-directivity and precedence. The weight factor w_i^2 for each sound path takes into account the source directivity g_S from Eq. (1) evaluated at the corresponding angle of radiation $\varphi_{R,i}$ minus the adjustable source orientation φ_{ori} , which is then all divided by the path length r_i , cf. Table. Moreover, it includes a broad-band directivity of the human ear g_{ear} evaluated at the corresponding auralization direction $\varphi_{l[i]}$

$$w_i = \frac{g(\varphi_{S,i} - \varphi_{ori}) g_{ear}(\varphi_{l[i]})}{r_i}. \quad (5)$$

The broad-band human ear directivity g_{ear} can be found in [25–27], and in [17] a suitable shape was expressed by dB values expanded in terms of a cosine series (see Fig. 6)

$$20 \lg g_{ear}(\varphi) = \begin{pmatrix} -4.4 \\ 4.5 \\ -0.3 \\ -1.5 \end{pmatrix}^T \begin{pmatrix} 1 \\ \cos \varphi \\ \cos 2\varphi \\ \cos 3\varphi \end{pmatrix} \text{ in dB}. \quad (6)$$

In the 2D setup, weighted and averaged vectors are simply denoted as complex scalars $e^{i\varphi_{l[i]}}$, and the phase of the accumulated result yields the final estimator of the primary auditory localization direction:

$$\hat{\varphi}_S = \angle \left\{ \sum_i w_i^2 e^{i\varphi_{l[i]}} \right\}. \quad (7)$$

The arrival time $\tau_i = r_i/c$ must be considered by summing only over paths i that lie above the precedence threshold

$$\forall i : w_i > w_{thr}(\tau_i). \quad (8)$$

The precedence threshold given in [3, 15] decays by minus $\frac{1}{4}$ dB/ms. In this paper we use the threshold raised by a factor of two to predict our experiment well:

$$w_{thr}(\tau) = 2 \max_{\forall i: \tau_i < \tau} \left\{ w_i 10^{\frac{(\tau_i - \tau)/\text{ms}}{4 \cdot 20}} \right\}. \quad (9)$$

In the given setting, propagation paths outside $\pm 60^\circ$ are entirely masked and therefore not shown in Figs. 7 and 8. The average absolute deviation of modeled directions from the medians of the perceived primary directions is 3.5° . The amount of deviation is mainly caused by the fact that the model sometimes predicts the secondary direction instead of the primary one for source orientations around -50° and 90° . The median absolute deviation is less sensitive to such special cases and amounts to only 0.3° .

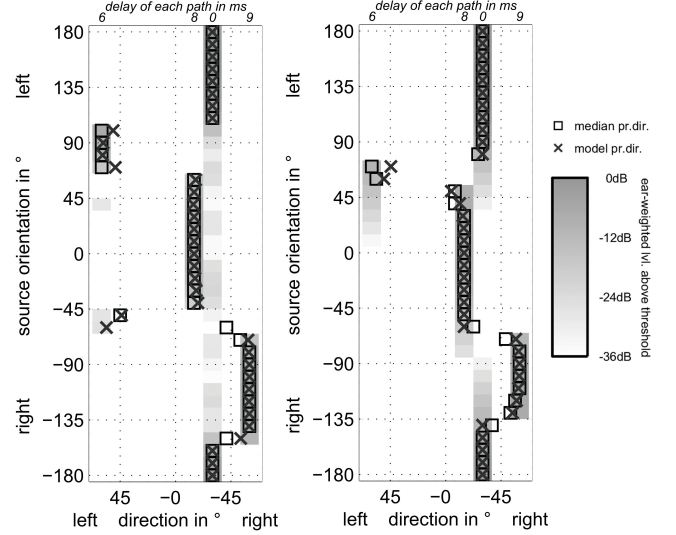


Fig. 7. Medians of perceived primary directions (squares) and modeled directions (crosses) for 1st order image source model; level of each propagation path after weighting with ear directivity and precedence masking is represented by the gray shading in the background.

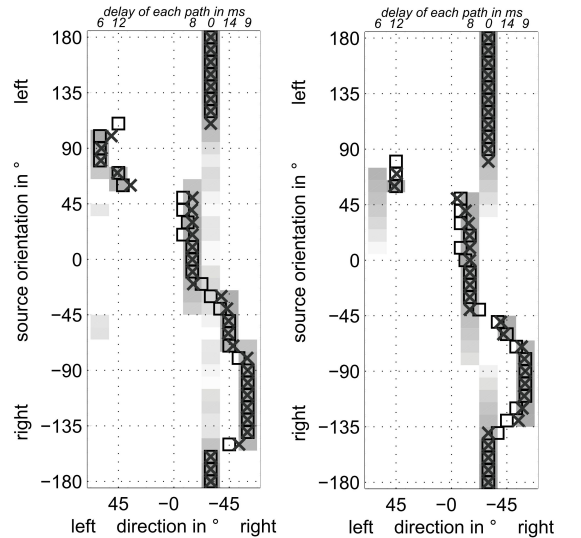


Fig. 8. Medians of perceived primary directions (squares) and modeled directions (crosses) for 2nd order image source model; level of each propagation path after weighting with ear directivity and precedence masking is represented by the gray shading in the background.

4. Conclusion and outlook

We could show results of a listening experiment indicating where auditory objects of a variable-directivity source are perceived in a shoebox-shaped room. In most cases, the primary auditory direction could be successfully modeled. There, localization is dominated by the acoustic reflection path that remains after applying to the ear-directivity-weighted time sequence of incoming reflections an exponentially-decaying mask that models precedence. The resulting model could be valuable in computer music and sound engineering practice to predict achievable spatial effects of not only compact spherical loudspeaker arrays but also in surrounding loudspeaker systems when listeners are seated off-center.

Other effects that are visible in the data were not modeled yet. In these special cases, either a onefold auditory localization is determined by stereophonic localization, or splitting occurs into twofold (or threefold) auditory directions. Stereophonic localization requires to be modeled using time-intensity trading curves such as those in [28, 29]. Less is known under which delay and directional offset the localization splits. Predicting from listener statements about different localization at the end of the sound, the envelope of the sound [30] may play a significant role. Envelope may determine the sequence or strength, in which onefold, twofold, or a thinkable threefold localization is perceived.

Acknowledgments

This work partly was 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] B.-I. Dalenbäck, M. Kleiner, P. Svensson, *J. Audio Eng. Soc* **41**, 905 (1993).
- [2] L.M. Ronsse, L.M. Wang, *Acta Acustica u. Acustica* **98**, 768 (2012).
- [3] B. Rakerd, W.M. Hartmann, J. Hsu, *J. Acoust. Soc. Am.* **107**, 1061 (2000).
- [4] V. Tarnow, *Technical Review* **4**, 23 (1974).
- [5] T.W. Leishman, S. Rollins, H.M. Smith, *J. Acoust. Soc. Am.* **120**, 1411 (2006).
- [6] A.M. Pasqual, *J. Sound Vib.* **333**, 4930 (2014).
- [7] Ch. Quested, A. Moorhouse, B. Piper, B. Hu, *Appl. Acoust.* **85**, 161 (2014).
- [8] J.L. Butler, S.L. Ehrlich, *J. Acoust. Soc. Am.* **61**, 1427 (1977).
- [9] O. Warusfel, P. Derogis, R. Caussé, in *Papers of AES 103rd Conv.*, New York 1997.
- [10] F. Zotter, R. Höldrich, in *Proc. 19th International Congress on Acoustics (ICA 2007)*, Instituto de Acústica, Madrid 2007.
- [11] M. Pollow, G.K. Behler, *Acta Acustica u. Acustica* **95**, 1082 (2009).
- [12] A.M. Pasqual, P. Herzog, J.R. de França Arruda, *J. Acoust. Soc. Am.* **128**, 3478 (2010).
- [13] G.K. Sharma, F. Zotter, M. Frank, in *Joint Conference ICMC/SMC 2014*, Eds. A. Georgaki, G. Kouroupetroglou, Athens 2014, p. 830.
- [14] A. Schmeder, in *Proc. of 1st Ambisonics Symposium*, IEM/KUG, Graz 2009.
- [15] P.W. Robinson, A. Walther, C. Faller, J. Braasch, *J. Acoust. Soc. Am.* **134**, 2755 (2013).
- [16] R. Mehra, L. Antani, S. Kim, D. Manocha, *IEEE T. Vis. Comput. Gr.* **20**, 495 (2014).
- [17] F. Zotter, M. Frank, A. Fuchs, D. Rudrich, in *Proc. of forum acusticum*, Kraków 2014.
- [18] J.B. Allen, D.A. Berkley, *J. Acoust. Soc. Am.* **65**, 943 (1979).
- [19] J. Daniel, Ph.D. Thesis, Université Paris 6 2001, p. 184.
- [20] F. Zotter, M. Frank, *J. Audio Eng. Soc* **60**, 807 (2012), Eq. (10).
- [21] ITU-R-BS.1116-1, 1997.
- [22] S. Bech, N. Zacharov, *Perceptual Audio Evaluation*, Wiley, 2006.
- [23] H. Pomberger, A. Sontacchi, R. Höldrich, in *NAG/DAGA Int. Conf. on Acoustics*, Ed. M.M. Boone, Rotterdam 2009.
- [24] M. Frank, A. Sontacchi, in *Fortschritte der Akustik — DAGA*, Ed. H. Hanselka, Darmstadt 2012.
- [25] M. Frank, Ph.D. Thesis, University of Music and Performing Arts Graz, 2013.
- [26] V.P. Sivonen, W. Ellermeier, *J. Acoust. Soc. Am.* **119**, 2965 (2006).
- [27] G. Jahn, *Hochfrequenztechnik und Elektroakustik* **72**, 23 (1963).
- [28] R.M. Aarts, in *Papers of 94th AES Conv.*, Berlin 1993.
- [29] H. Lee, F. Rumsey, *J. Audio Eng. Soc* **61**, 978 (2013).
- [30] B. Rakerd, W.M. Hartmann, *J. Acoust. Soc. Am.* **80**, 1695 (1986).