Perceptual Evaluation of Spatial Resolution in Directivity Patterns

Matthias Frank[1] and Manuel Brandner[1]

[1] *Institute of Electronic Music and Acoustics, University of Music and Performing Arts, Graz, Austria, Email: frank@iem.at*

Introduction

Plausible and authentic auralization of sound sources in rooms benefits from the incorporation of source and receiver directivity to control the direct-to-reverberant energy ratio, and thus the perceived distance [1, 2]. While receiver directivity is typically well represented by measured head-related transfer functions, source directivity requires more measurement effort, especially for musical instruments. However, there exist surround microphone arrays of 64 microphones to capture source directivity in a resolution up to 7th order spherical harmonics [3]. Even if the directivity pattern is not of high spatial resolution, high orders are sometimes necessary to compensate for imprecise centering [4, 5].

Clearly, the computational effort for auralization with high orders is high and can make real-time processing very challenging, in particular when using multiple sources simultaneously. Moreover, typical compact spherical loudspeaker arrays to play back directional sources are limited to 3rd order [6]. As a consequence, it is desirable to reduce the spatial resolution of directivity patterns. Still, little is known about the perceptual impact of such reduction and what the best strategy for reduction is, i.e. which parameters of the directivity pattern are most important to preserve.

This contribution tries to take a step forward in answering these questions. As a simplified model, we assume that every arbitrary directivity pattern can be composed of multiple directional components and a diffuse component. The directional components are characterized by parameters, such as direction, beam width, and side lobe suppression. The resulting phase of the combined directional components is assumed to be perceptually irrelevant when playing the source in a reverberant environment. The diffuse component distributes decorrelated sound equally into all directions and its level is adjusted to complement the directional components in a way that the overall sound power is preserved in comparison to the high order directivity pattern.

In a listening experiment, we investigate the perceptual impact of order reduction for both a generic directional beam and a diffuse source. The experiment uses speech and noise as stimuli and is performed at two listening positions in a simulated room. The experimental results are finally related to technical measures to investigate which parameters of the source directivity are most important to preserve. This helps to develop an efficient reduction strategy.

Setup and Conditions

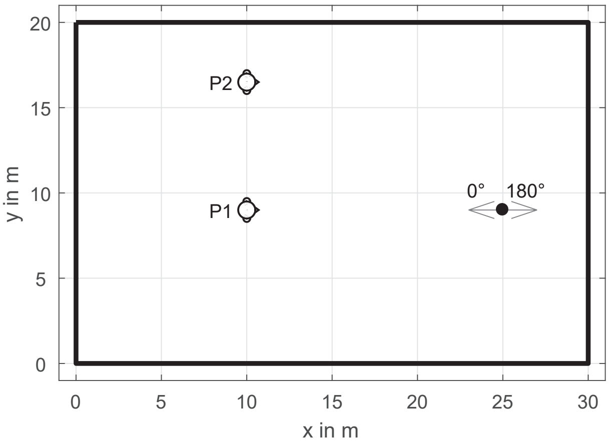


Figure 1: Position of listener and source in the horizontal cross section of the simulated room.

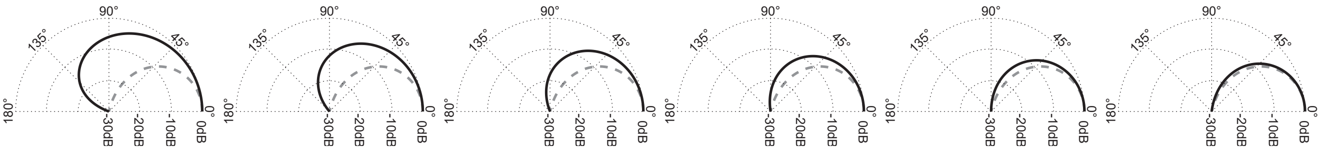
The simulated room had a size of 30 m × 20 m × 10 m, cf. Figure 1. The reverberation time for frequencies between 200 Hz and 2 kHz was 1.9 s and it doubled/halved for frequencies below 100 Hz and above 4 kHz, respectively. That resulted in a critical distance of 3.2 m for mid frequencies. The simulation employed a 7th-order image-source model (236 reflections) from the IEM `RoomEncoder` VST plug-in[1]. Headphone playback employed 7th-order head-tracked [7] binaural Ambisonics [8] using the IEM `BinauralDecoder`.

The source was positioned 4 m above the floor and the two listening positions were at a height of 2 m to recreate typical concert conditions. Listening position P1 was positioned exactly in the frontal direction of the source to provoke changes in the direct-to-reverberant energy ratio mainly, whereas P2 had an angle of 30° to the source to study the influence of an increased beam width at low orders.

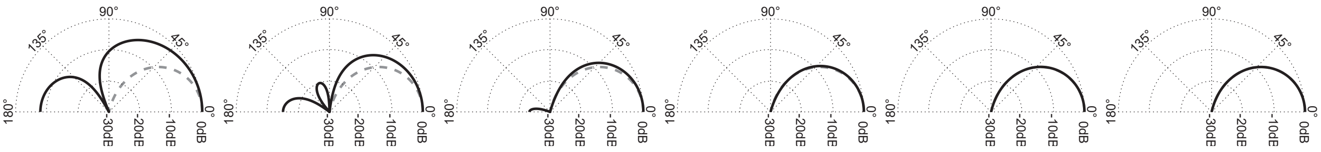
The reference for the directional beam was a 7th-order inphase [9] design facing exactly at P1 (0°) and away from it (180°), respectively. The limitation to 7th order was done due to practical reasons, as the largest available microphone array to measure simultaneously has 64 microphones and the VST plug-in is also limited to 64 channels. The second orientation resulted in no direct sound from the source at P1, as the inphase design has a null at the back, cf. gray dashed directivity pattern in Figure 2. The distance between the source and P1 was chosen to be the effective critical distance of the reference beam. Note that the orientation angle of the source in the plug-in is reversed in comparison to this contribution.

---

[1] freely available at `plugins.iem.at`



(a) inphase order 1 (b) inphase order 2 (c) inphase order 3 (d) inphase order 4 (e) inphase order 5 (f) inphase order 6



(g) truncated order 1 (h) truncated order 2 (i) tunrcated order 3 (j) truncated order 4 (k) truncated order 5 (l) truncated order 6

Figure 2: Directivity patterns of beams in the experiment; gray dashed line indicates 7^{th}-order inphase beam as reference.

Two different strategies were employed to reduce the order of the directional beam: (I) appropriate inphase design for the reduced order, resulting in beams with strongly increasing width of the main lobes while preserving nulls at the back and (II) simple truncation that better preserves the width of the main lobe at the cost of increasingly strong side lobes at the back of the pattern, cf. Figure 2.

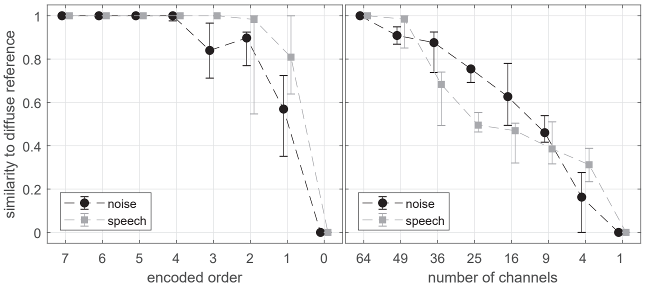
The diffuse source employed a 64×64 feedback delay network (FDN) [10, 11] to generate multiple decorrelated signals out of a single input using the IEM `FDNReverb` plug-in. The plug-in was set to a reverberation time of 1 s (0.3 s above 2 kHz) and a fade-in time of 0.1 s to ensure maximally decorrelated output signals while keeping the reverberation by the FDN clearly below the reverberation time of the room. There were two strategies to spatially distribute the 64 output signals on the diffuse source: (A) encoding the 64 signals at 64 position equally distributed on a sphere and varying the order from 0 to 7 (reference) and (B) selecting a set of output channels ranging from 1, 4, 9 to 64 (reference) and apply them as spherical harmonic channels directly.

The experiment employed two different sounds: (α) continuous pink noise for maximum sensitivity to coloration and loudness and (β) male English speech [12] that facilitates better spatial perception and familiarity.

Overall, there were $20 = 2$ (sounds) $\times$ 2 (reduction strategy for directional beam) $\times$ 2 (orientation of the directional beam) $\times$ 2 (listening positions for directional beam) + 2 (diffuse strategy) trials with 8 (0^{th} to 7^{th} order or 1, 4, 9 to 64 channels) stimuli each in multi-stimulus comparison. The listeners task was to compare the similarity of the 8 stimuli to the corresponding reference on a continuous scale from *very different* to *identical*. Each of the 10 experienced listeners (average age 32 years) spent about 33 min on the entire experiment.

Results

The results for the diffuse source revealed that independent of the strategy, the similarity to the reference increases with the encoding order and number of FDN output channels, cf. Figure 3. The encoding strategy requires orders of 4 and 1 to generate perceptually indistinguishable results (Wilcoxon signed rank test with Bonferroni-Holm correction) to the reference for noise and speech, respectively. Reducing the number of FDN output channels has a stronger perceptual effect, as this strategy would require 64 and 49 channels (equivalent to orders of 7 and 6).



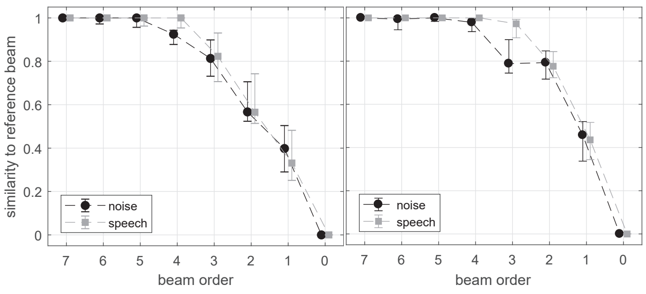
(a) 64 encoded FDN channels (b) number of FDN channels

Figure 3: Medians and 95% confidence intervals of perceived similarity for diffuse source.

The results indicate that the number of FDN output channels is crucial for the diffuseness of the source while the spatial resolution for distributing the 64 channels on the surface of the source is not important for decorrelation as long as the spherical harmonics order exceeds 0. However, as the results for noise show, coloration-less reproduction may require higher orders.

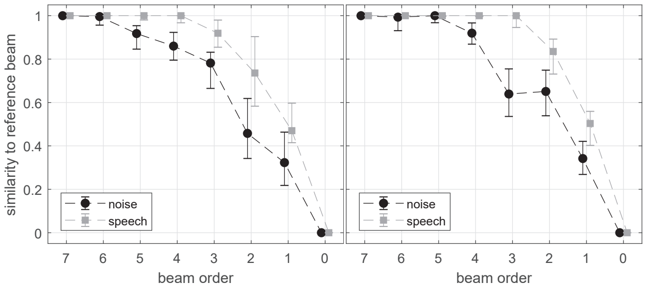
For the directional beams, similarity increases with the order, cf. Figures 4 and 5. For each sound, reduction strategy, and source orientation, the minimum required order to be indistinguishable from the reference is shown in Table 1. As for the diffuse source, higher resolution is required for the reproduction of noise. The truncation strategy seems to be more efficient, especially for noise at listening position P2, where the beam is not facing directly at or away from the listener. This finding gives a hint that preserving the width of the main lobes is more important than the exact reproduction of the nulls.

In order to derivate rough thresholds for perceptual differences, i.e. similar to just noticeable differences (JNDs), the next section calculates some technical measures and relates them to the experimental results for the directional beams.



(a) inphase beams (b) truncated beams

Figure 4: Medians and 95% confidence intervals of perceived similarity summarizing 0° and 180° beam directions at listening position P1.



(a) inphase beams (b) truncated beams

Figure 5: Medians and 95% confidence intervals of perceived similarity summarizing 0° and 180° beam directions at listening position P2.

Table 1: Minimum required order to be indistinguishable from reference at 5% level with Bonferroni-Holm correction.

	list. pos. P1		list. pos. P2	
beam	noise	speech	noise	speech
inphase 0°	5	4	7	4
inphase 180°	5	3	6	3
truncated 0°	4	4	5	3
truncated 180°	5	3	5	3

Technical Measures

The first kind of technical measures is independent of the room and solely depends on the beam itself. The measures are:

- side lobe: level of the strongest side lobe in dB,
- width: aperture angle of the cap exceeding -6 dB relative to the maximum in °,
- F/B-R_{40}: front-to-back ratio in dB, with lower dynamic limitation at -40 dB relative to the maximum,
- F/B-R_{25}: front-to-back ratio in dB, with lower dynamic limitation at -25 dB relative to the maximum.

Table 2 presents the values for the reference beam and the two reduction strategies at different orders. Values for orders that resulted in indistinguishable experiment results for speech and noise are printed in italics and bold, respectively.

The minimum required order for the truncation strategy when facing the listener at P1 was found to be 5 and 3, respectively, cf. Table 1. Thus for noise and speech, a side lobe attenuation of 49.1 dB and 23.4 dB is enough. This finding indicates that a more precise preservation of nulls in the beam pattern are not necessary. The F/B-R_{40} reflects this limitation of the side lobe attenuation and reveals that a difference of 1.1 dB is perceptually irrelevant for noise. The difference when further reducing the order to 4 (inphase) and 3 (truncation) results in a clear increase of at least 4.4 dB. Similarly for speech, differences in F/B-R_{25} of 0.8 dB are tolerable, while the next lower order results in a difference of at least 2.5 dB. Interestingly, for both noise and speech, there is a clear point of discontinuity in the F/B-R_{40} and F/B-R_{25} differences at the minimum required order.

The differences in beam width are visible for the 0° direction at P2. For noise, a difference of 3° was not tolerated, whereas for speech 23° or 32% were tolerated.

Table 2: Side lobes, beam width, and front-to-back energy ratios of the tested beams. Values that resulted in indistinguishable results for speech and noise are printed in italics and bold, respectively.

beam	side lobe in dB	width in °	F/B-R_{40} in dB	F/B-R_{25} in dB
inphase 7	$-\infty$	71	34.8	19.8
inphase 6	$-\infty$	*77*	**35.0**	*20.1*
inphase 5	$-\infty$	*84*	**34.0**	*20.5*
inphase 4	$-\infty$	*94*	30.4	*20.9*
inphase 3	$-\infty$	108	24.7	*20.6*
inphase 2	$-\infty$	131	18.1	17.3
inphase 1	$-\infty$	180	10.8	10.8
0	0	360	0	0
truncated 6	**−70.7**	**71**	**34.8**	*19.8*
truncated 5	**−49.1**	**72**	**34.8**	*19.8*
truncated 4	*−34.2*	*74*	**33.7**	*19.9*
truncated 3	*−23.4*	*81*	24.9	*20.1*
truncated 2	-15.1	99	16.1	15.9
truncated 1	-8.0	147	9.5	9.5

Table 3: Direct-to-reverberant energy ratio of the tested beams in dB for both listening positions. Values that resulted in indistinguishable results for speech are printed in italics.

	list. pos. P1		list. pos. P2	
beam	180°	0°	180°	0°
inphase 7	$-\infty$	-0.2	$-\infty$	-7.3
inphase 6	$-\infty$	*−0.7*	$-\infty$	*−7.5*
inphase 5	$-\infty$	*−1.3*	$-\infty$	*−7.9*
inphase 4	$-\infty$	*−2.0*	$-\infty$	*−8.4*
inphase 3	$-\infty$	-2.8	*−80.3*	-9.0
inphase 2	$-\infty$	-3.9	-56.2	-9.8
inphase 1	-50.1	-5.3	-32.8	-10.8
0	-9.1	-9.1	-13.8	-13.8
truncated 6	*−71.6*	*−0.2*	*−81.3*	*−7.3*
truncated 5	*−49.4*	*−0.2*	*−62.4*	*−7.3*
truncated 4	*−34.1*	*−0.4*	*−59.0*	*−7.2*
truncated 3	*−23.4*	-0.9	*−40.6*	*−7.5*
truncated 2	-16.1	-2.3	-25.5	-8.5
truncated 1	-10.7	-4.5	-16.5	-10.1

The second kind of technical measure is the direct-to-reverberant energy ratio and it depends on the combination of the beam pattern, its orientation, and the room. For the 180° orientation at P1, the resulting values are obviously similar to the level of the strongest side lobe in Table 2. The same tendency can be found at P1. The results indicate that the direct-to-reverberant energy ratio has a lower perceptual limit around -23 dB for speech. For reference values around 0 dB, i.e. for the 0° orientation at P1, differences exceeding 1.8 dB are perceivable. This difference agrees with the JNDs found in literature [13]. For the same orientation at P2, where the beam is not facing the listener, sensitivity increases and so the difference to the reference must not exceed 1.1 dB.

Conclusion

This contribution investigated the perceptual effect of reducing the spatial resolution of 7th-order diffuse and directional sources in a virtual environment. In general, listeners were more sensitive when listening to pink noise in comparison to speech. For the diffuse source, a high number of decorrelated FDN outputs was crucial, whereas a 1st-order spatial resolution was sufficient for speech.

The minimum required order of directional beams to be indistinguishable from the 7th-order inphase reference was around 3 for speech and 5 for noise. The reduction of the order by simple truncation yielded better results than the inphase design of the same order, indicating that the exact preservation of nulls is perceptually less relevant than the approximation of the beam width. Differences in beam width of 23° or 32% were tolerable for speech, while differences of 3° or 4% were perceivable for noise. Differences in the front-to-back ratio were imperceivable below about 1 dB, as for the direct-to-reverberant ratio for a reference around 0 dB. For a reference at $-\infty$ dB, a beam with a ratio of -23 dB was not perceived differently.

References

[1] A. Kolarik, S. Cirstea, and S. Pardhan, "Discrimination of virtual auditory distance using level and direct-to-reverberant ratio cues," *The Journal of the Acoustical Society of America*, vol. 134, no. 5, pp. 3395–3398, 2013.

[2] F. Wendt, F. Zotter, M. Frank, and R. Höldrich, "Auditory distance control using a variable-directivity loudspeaker," *MDPI Applied Science*, vol. 7, no. 7, 2017.

[3] F. Hohl, "Kugelmikrofonarray zur Abstrahlungsvermessung von Musikinstrumenten," Master's thesis, TU Graz, 2009.

[4] D. Deboy, "Tangential intensity algorithm for acoustic centering," in *Fortschritte der Akustik, DAGA*, Düsseldorf, 2011.

[5] I. B. Hagai, M. Pollow, M. Vorländer, and B. Rafaely, "Acoustic centering of sources measured by surrounding spherical microphone arrays," *Journal of the Acoustical Society of America (JASA)*, vol. 130, no. 4, 2011.

[6] F. Zotter, M. Zaunschirm, M. Frank, and M. Kronlachner, "A beamformer to play with wall reflections: The icosahedral loudspeaker," *Computer Music Journal*, vol. 41, no. 3, 2017.

[7] M. Romanov, P. Berghold, M. Frank, D. Rudrich, M. Zaunschirm, and F. Zotter, "Implementation and evaluation of a low-cost headtracker for binaural synthesis," in *Audio Engineering Society Convention 142*, May 2017.

[8] C. Schörkhuber, M. Zaunschirm, and R. Höldrich, "Binaural rendering of ambisonic signals via magnitude least squares," in *Fortschritte der Akustik - DAGA*, Munich, March 2018.

[9] J. Daniel, "Représentation des champs acoustiques, application à la transmission et à la reproduction de scènes sonores complexes dans un contexte multimédia," Ph.D. dissertation, Université Paris 6, 2001.

[10] J. Stautner and M. Puckette, "Designing multichannel reverberators," *Computer Music Journal*, vol. 6, no. 1, pp. 52–65, 1982.

[11] J.-M. Jot and A. Chaigne, "Digital delay networks for designing artificial reverberators," in *90th AES Conv., prepr. 3030*, Paris, February 1991.

[12] EBU. (2008) EBU SQAM CD: Sound Quality Assessment Material recordings for subjective tests. [Online]. Available: https://tech.ebu.ch/publications/sqamcd

[13] E. Larsen, N. Iyer, C. R. Lansing, and A. S. Feng, "On the minimum audible difference in direct-to-reverberant energy ratio," *The Journal of the Acoustical Society of America*, vol. 124, no. 1, pp. 450–461, 2008.