Spatial impression and directional resolution in the reproduction of reverberation

Matthias Frank and Franz Zotter
Institute of Electronic Music and Acoustics
University of Music and Performing Arts, 8010 Graz, Austria, Email: frank@iem.at

Introduction

Diffuse sound fields are typically characterized by sound that arrives at the listener randomly, but equally distributed over space and time. Thus, one could think that low spatial resolution is sufficient for the representation of late reverberation. In recording technology, it is practical knowledge that the impression of spaciousness requires decorrelated signals. Decorrelation is typically achieved by large distances between the microphones rather than coincident microphone arrays. Higher correlation during playback is assumed to reduce the spatial impression by reducing the perceived mapping of spatial depth. This contribution investigates the dependency of the spatial impression on the directional resolution for different listening positions inside an Ambisonic playback system when playing back a reverberant sound scene.

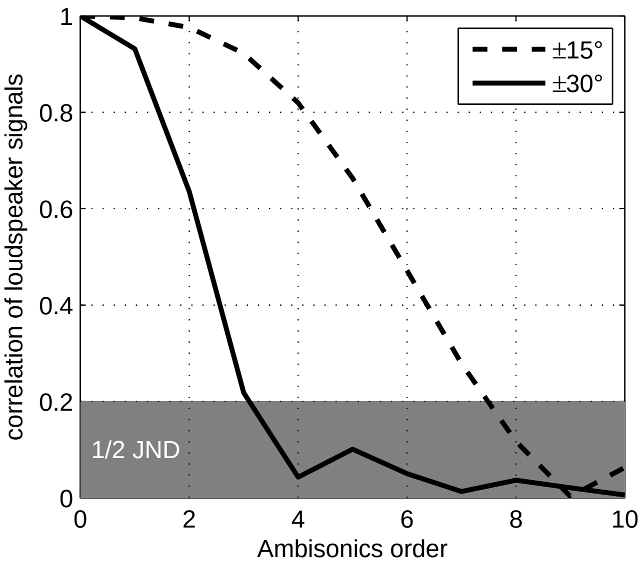

Figure 1: Correlation of 2 independent noise signals at $\pm 15°$ and $\pm 30°$ after Ambisonic encoding/decoding in dependence of Ambisonics order.

Ambisonics [1] permits adjustable spatial resolution by truncation of the spherical harmonic representation at a certain Ambisonic order. Figure 1 shows the correlation of two independent noise sources encoded into Ambisonics and then decoded at the encoding angles $\pm 15°$ and $\pm 30°$, respectively. As the crosstalk of each source pair decreases with the order, also the correlation of the loudspeaker signals decreases. To estimate the required order, Figure 1 shows the just-noticeable difference (JND) for the inter-aural cross correlation coefficient (IACC) close to zero [2]. Signal correlation in the range $[0; 1]$ typically induces IACCs within the range $[0.2; 0.85]$ in studios [3]. Simplistically, we may assume a direct mapping from signal correlation to IACC, which gives a rough idea about how Ambisonic order determines IACC. Preservation of decorrelation is roughly achieved by the orders above 3 for sources at $\pm 30°$, and above 7 for $\pm 15°$.

In this contribution, we employ the spatial decomposition method [4] to create higher-order Ambisonics reverberation from first-order room impulse responses. The higher-order impulse responses are compared to the first-order original in listening experiments that use the impulses responses for a 3D convolution reverb.

The first listening experiment determines the perceptual sweet spot size in dependence of the Ambisonic order and the second experiment evaluates the perceived mapping of spatial depth at two listening positions.

Spatial Decomposition Method (SDM)

The spatial decomposition method (SDM [4]) can be employed to increase the spatial resolution in measured directional impulse responses. This is basically done by estimation of the direction of every single sample in the impulse response and reassignment of each sample at the estimated direction. For the investigation in this paper, the directional reassignment is done using different Ambisonic orders. The increase of spatial resolution is assumed to improve both the localization of the direct sound and the diffuseness of the diffuse sound.

SDM uses the intensity vector to estimate the direction of sound. The intensity is computed from the scalar sound pressure p (W channel) and the velocity vector $\boldsymbol{v}$:

$$\boldsymbol{I} = p\boldsymbol{v}. \tag{1}$$

Practically, the velocity vector is not measured explicitly, but rather by the pressure gradients in the directions of the x-, y-, and z-axis (X, Y, Z channels).

Due to the non-infinitesimally small distance between the four cardioid capsules of typical 1^{st}-order microphone arrays, the application of a short moving average filter is reasonable for the directional estimation from the intensity vector. We choose a filter length of 10 samples ($226\,\mu$s at 44.1 kHz). For the same reason, the analysis is limited to a frequency range below 4 kHz, i.e. the spatial aliasing frequency.

Figure 2 shows the directional analysis of the first 100 ms of a first-order directional impulse response taken from the openair lib[1]. This response was measured in St. Andrew's Church Lyddington, UK ($2600\,\text{m}^3$ volume, 11.5 m source-receiver distance) with a Soundfield SPS422B microphone.

[1] http://www.openairlib.net

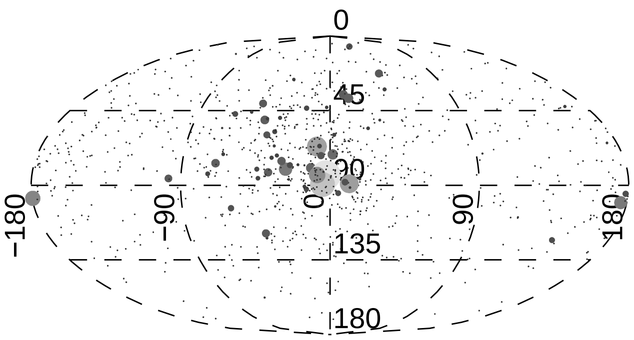

Figure 2: Spatial distribution of the first 100 ms of the employed first-order directional impulse response analyzed by SDM. Brightness and size of the circles indicate the level.

The direct sound from the front is clearly visible, as well as strong early reflections from front and back, and equally distributed weak directions from the diffuse reverb.
The estimated directions are employed to encode the omni-directional sound pressure samples with arbitrarily high directional resolution using Ambisonics. Resulting impulse responses exhibit increased spectral brightness [4] as an artifact when compared to the original. According to the impression when listening, this brightness increase mainly modifies the diffuse reverberation tail and thus cannot be equalized by time-invariant filtering. Figure 3 presents a clear analysis of this behavior in terms of an unnatural increase of the reverberation time at high frequencies, especially when using high encoding orders. The behavior originates in spectral effects of amplitude modulation caused by the strong directional fluctuations in the diffuse tail. In this regard, the encoding order can be interpreted as the modulation depth.
In order to equalize the reverberation times of the SDM-enhanced impulse responses to the those of the original in third-octave resolution, we multiplied the impulse responses with decaying exponentials, similar as in [5].

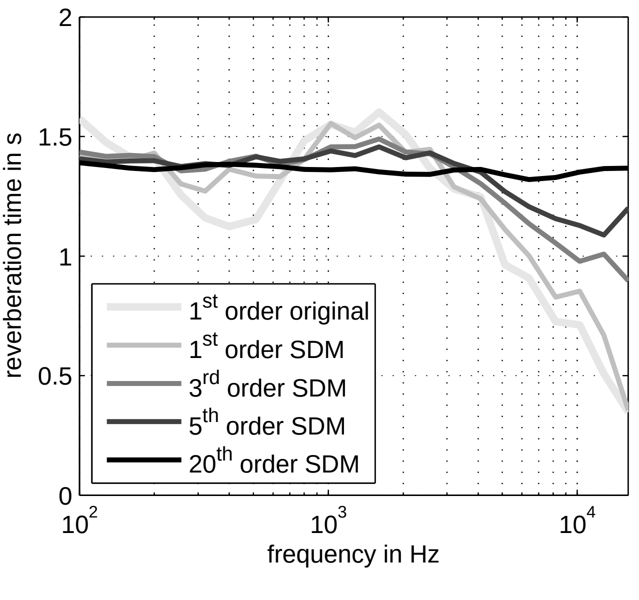

Figure 3: Frequency-dependent reverberation time calculated from original and SDM-enhanced impulse responses.

Experimental Setup and Conditions

The experiments we present were done at the IEM CUBE, a 10.3 m × 12 m × 4.8 m studio with a reverberation time of 750 ms. A hemisphere of 24 Tannoy System 1200 loudspeakers were used for playback, cf. [6] for exact angular loudspeaker angles. The lowest loudspeaker ring at ear height is shown in Figure 4 along with the listening positions of the second experiment. Ambisonic [7] decoding used AllRAD [6] including appropriate max-r_{E} weighting [8] in dependence of the maximum playback order, which was shown to achieve best results in terms of localization and coloration at off-center listening positions in previous studies [9, 10]. Moreover accordingly, delay differences due to the different loudspeaker distances to the center were not compensated.
The audio scene was composed of a direct sound arriving from the front and enveloping reverberation. The first sentence of EBU's male speech reference recording [11] has been convolved with different versions of the spatial impulse responses based on the measurements from St. Andrew's Church, considering 4 test conditions:

- original 1^{st}-order,
- SDM-enhanced 1^{st}-order,
- SDM-enhanced 3^{rd}-order,
- SDM-enhanced 5^{th}-order.

Convolution and Ambisonic decoding of the test signal was done using the *mcfx* and *ambiX* plugin-suites[2].
Both listening experiments used the same 15 listeners as participants. All of them were part of the trained expert listening panel [12, 13, 14].

Experiment 1: Sweet Spot Size

The size of the perceptual sweet spot can be determined by localization experiments at various listening positions, e.g. using established pointing methods [15], and subsequent selection of listening positions with localization errors below a certain plausible threshold, e.g. 10°. By contrast, this contribution employs a more comprehensive approach that is not restricted to localization.
We defined the sweet spot as the listening area within which the perception of the sound scene is plausible, i.e. the direct sound is localized from the front and the reverberation is perceived enveloping. To determine the corresponding sweet spot geometry, listeners were asked to approach each of the horizontal loudspeakers until either (a) the direct sound is localized outside the loudspeaker triplet L-C-R, (b) localization collapses into a single loudspeaker, or (c) the reverberation stops being enveloping. While approaching each loudspeaker, they walked on a straight line from the center of the arrangement towards the corresponding loudspeaker and were facing the front loudspeaker C. The observed distance limit was filled in a paper form and interpreted as the sweet spot radius for the respective direction of motion. The experiment was part (1/3) of a more exhaustive study including two more sound scenes and took an average duration of 56 minutes.

[2]freely available on `http://www.matthiaskronlachner.com`

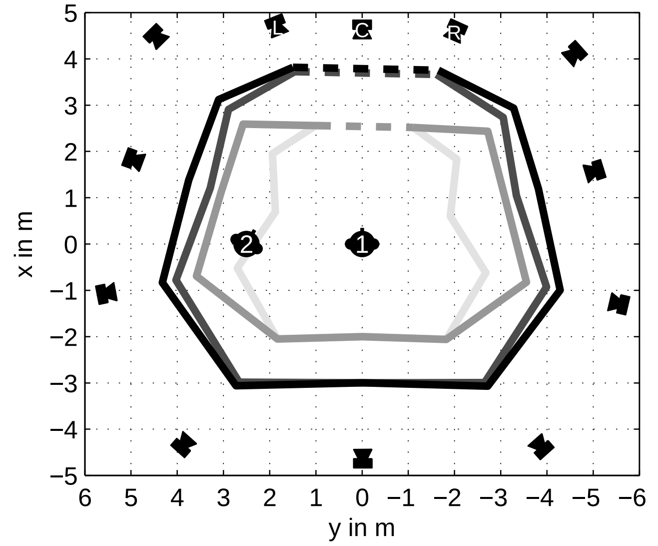

Figure 4: Median sweet spot radius for each loudspeaker direction except C. Gray levels indicate conditions: light gray - 1^{st}-order original, gray - 1^{st}-order SDM, dark gray - 3^{rd}-order SDM, black - 5^{th}-order SDM.

Figure 4 shows the resulting median sweet spot radius for each loudspeaker direction. Note that there are no results for the front loudspeaker C, as the sweet spot radius for its direction would equal the loudspeaker radius. Obviously the sweet spot size increases with the order. Interestingly, the lateral sweet spot radius using the SDM-enhanced 1^{st}-order impulse responses is larger than for the original impulse responses, despite the same Ambisonic order. This is because the SDM version sharpens each sample to the max-r_{E} [8] optimum of the 1^{st}-order.

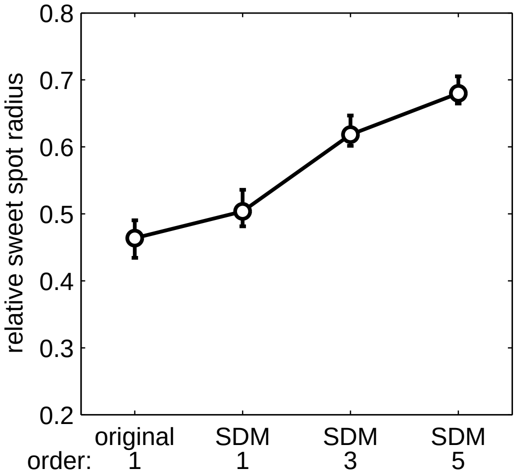

Figure 5: Median relative sweet spot radius and corresponding 95% confidence interval, summarizing all loudspeaker directions except C.

The same results can be found in Figure 5, where the median sweet spot radius is shown in relation to the loudspeaker radii. All conditions yield significantly different median and mean values ($p < 0.001$). The sweet spot radius is more than 2/3 of the loudspeaker radius for 5^{th} order, while is stays clearly below 1/2 for 1^{st} order.

Experiment 2: Mapping of Spatial Depth

In addition to the determination of the perceptual sweet spot geometry, the effect of SDM has been evaluated with respect to spaciousness, focusing on the mapping of spatial depth. The evaluation has been done at two listening positions: (1) in the center of the loudspeaker arrangement and (2) 2.5 m left from the center, cf. Figure 4. The second listening position has been chosen as to be just outside the sweet spot of the original 1^{st}-order playback.

Each listener evaluated all 4 versions of the reverberant speech signal simultaneously in random order using a MUSHRA-like [16] graphical interface with scales for the perceived mapping of spatial depth ranging from "very small" to "very large". The evaluation was performed once per listener and listening position resulting in 2 trials with 4 stimuli each. The average duration of the whole experiment was 11 minutes but also included another audio scene that is not analyzed here.

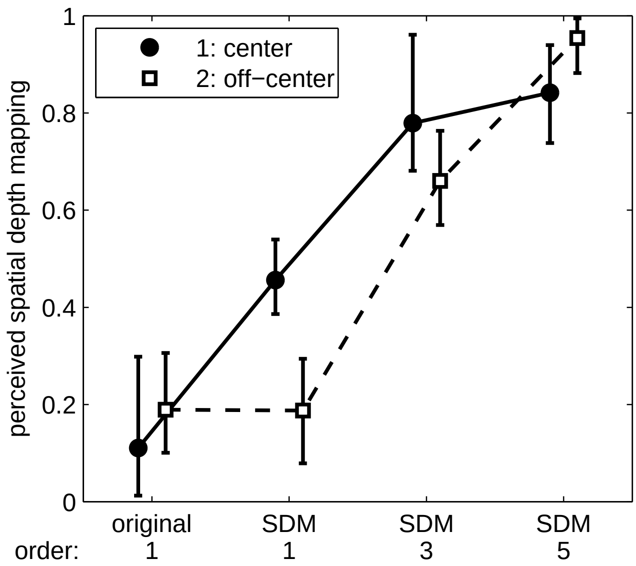

Figure 6: Median and corresponding 95% confidence intervals for the perceived mapping of spatial depth for both listening positions.

As expected, the results in Figure 6 show that the perceived mapping of spatial depth increases with the Ambisonics order. In particular, all differences between median/mean values are significant ($p \leq 0.003$) at the central listening position, except for the comparison of 3^{rd} and 5^{th} order (median $p = 0.74$, mean $p = 0.73$). At the central listening position, increasing the order above 3 seems to be perceptually ineffective.

In contrast, the differences between all SDM conditions are significant at the off-center position ($p < 0.001$), whereas the difference between original and SDM-enhanced 1^{st} order is not (median $p = 0.72$, mean $p = 0.87$). The results show that the effect of an increased Ambisonics order is obviously relevant at off-center listening positions. For 1^{st}-order playback, the improved mapping of spatial depth by SDM gets ineffective at the off-center listening position, which already lies slightly outside the sweet spot.

Conclusion

This contribution investigated the dependency of the spatial impression of reverberant sound on the directional resolution for different listening positions in an Ambisonic playback system. We employed the Spatial Deomposition Method (SDM) to create first-, third-, and fifth-order Ambisonic impulse responses from measured first-order directional room impulse responses. Strong directional fluctuations in the diffuse reverberation tail yield an increase of the reverberation time at high frequencies and orders, and we were able to compensate for this effect by exponential windowing in third-octave resolution.
In the first listening experiment, we determined the geometry of a perceptual sweet spot for the convolution reverb employing the original first-order impulse response and its SDM-enhanced versions. As expected, the sweet spot size increases with the Ambisonic order. What is more, also the SDM-enhanced first-order playback exhibits a larger sweet spot than the original impulse response. For fifth order, the perceptual sweet spot has a size of more than 2/3 of the loudspeaker radius, which agrees with the experimental results of Stitt [17] and is much larger than the sweet spot for physically accurate reproduction [18].
In the second experiment, we investigated the perceived mapping of spatial depth when using SDM. At the central listening position, SDM increases the perceived spatial depth for playback orders up to 3, while off-center this increase is only perceived for orders greater than 1. Apparently off-center, the sweet spot size of first-order playback limits the depth-increasing effect of SDM.
Our results clearly suggest to use higher-order playback: Not only does it provide well-focused direct sound, but also it achieves a large sweet spot with reverberant sound fields convincing in terms of envelopment and depth.

Acknowledgments

The authors thank all listeners for their participation. This work was partly supported by the project ASD, which is funded by Austrian ministries BMVIT, BMWFW, 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. Partial support was also granted by the project OSIL (AR 328-G21), which is funded by the Austrian Science Fund (FWF).

References

[1] M. A. Gerzon, “Ambisonics. Part two: Studio techniques,” *Studio Sound*, vol. 17, pp. 24–26, 1975.

[2] I. Pollack and W. J. Trittipoe, “Binaural listening and interaural noise cross correlation,” *The Journal of the Acoustical Society of America*, vol. 31, no. 9, pp. 1250–1252, 1959.

[3] F. Zotter and M. Frank, “Efficient phantom source widening,” *Archives of Acoustics*, vol. 38, no. 1, pp. 27–37, 2013.

[4] S. Tervo, J. Pätynen, A. Kuusinen, and T. Lokki, “Spatial decomposition method for room impulse responses,” *J. Audio Eng. Soc*, vol. 61, no. 1/2, pp. 17–28, 2013.

[5] C. Pörschmann, S. Schmitter, and A. Jaritz, “Predictive auralization of room modifications by the adaptation of measured room impulse responses,” *Fortschritte der Akustik – AIA-DAGA*, pp. 1653–1656, 2013.

[6] F. Zotter and M. Frank, “All-round Ambisonic panning and decoding,” *Journal of the Audio Engineering Society*, vol. 60, no. 10, pp. 807–820, 2012.

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

[8] J. Daniel, *Représentation de champs acoustiques, application à la transmission et à la reproduction de scénes sonores complexes dans un contexte multimédia.* PhD thesis, Université Paris 6, 2001.

[9] M. Frank and F. Zotter, “Localization experiments using different 2D Ambisonics decoders,” in *25th Tonmeistertagung*, (Leipzig), 2008.

[10] M. Frank, *Phantom Sources using Multiple Loudspeakers in the Horizontal Plane.* PhD thesis, University of Music and Performing Arts Graz, Austria, 2013.

[11] EBU, “EBU SQAM CD: Sound Quality Assessment Material recordings for subjective tests,” 2008.

[12] A. Sontacchi, H. Pomberger, and R. Höldrich, “Recruiting and evaluation process of an expert listening panel,” in *Fortschritte der Akustik, NAG/DAGA*, (Rotterdam), 2009.

[13] M. Frank, A. Sontacchi, and R. Höldrich, “Training and guidance tool for listening panels,” in *Fortschritte der Akustik, DAGA*, (Berlin), 2010.

[14] M. Frank and A. Sontacchi, “Performance review of an expert listening panel,” in *Fortschritte der Akustik, DAGA*, (Darmstadt), 2012.

[15] 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*, 6 2010.

[16] ITU, “ITU-R BS.1534-1: Method for the subjective assessment of intermediate sound quality (mushra),” 2006.

[17] P. Stitt, S. Bertet, and M. van Walstijn, “Off-centre localisation performance of ambisonics and HOA for large and small loudspeaker array radii,” *Acta Acustica united with Acustica*, vol. 100, no. 5, 2014.

[18] D. Ward and T. Abhayapala, “Reproduction of a plane-wave sound field using an array of loudspeakers,” *IEEE Transactions on Speech and Audio Processing*, vol. 9, pp. 697–707, September 2001.