

# How to make Ambisonics sound good

Matthias Frank

Institute of Electronic Music and Acoustics,  
University of Music and Performing Arts, Graz, Austria.

## Summary

Ambisonics is a recording and reproduction method that is based on the representation of the sound field excitation as a decomposition into spherical or circular harmonics, respectively. This achieves physically accurate sound field reproduction restricted within a sweet spot in the center of a loudspeaker array in an anechoic room. However, experiments show that a perceptually defined sweet spot is far less restrictive, even with a small number of loudspeakers in non-anechoic listening rooms. In this case, Ambisonics is rather understood as a simple amplitude-panning method based on the psychoacoustic phenomenon of a phantom source as it is known from stereophony. Taking the current opportunity, this contribution gathers recent experimental results, brings them together with the concept of quality, and hereby discusses the effect of quality elements (e.g. reproduction room, number and equalization of loudspeakers, order weighting, and decoder design) on perceived quality features (e.g. localization, source width, and coloration). The discussion reveals that a physically accurate reproduction does not necessarily yield good perceived quality. For this reason, the contribution puts optimal quality elements of Ambisonics in perspective that ensure optimal sound.

PACS no. 43.38.Tj, 43.38.Vk, 43.38.Md, 43.66.Qp, 43.60.Fg

## 1. Introduction

Although Ambisonics has been invented more than 40 years ago [1, 2], it has hardly left research laboratories to enter a broad commercial audience. Reasons for this could either be the complexity of the mathematics involved or demonstrations that did not fully convince the audience, especially sound engineers. Members of the latter group often report that they perceived sound sources that always stuck to the closest loudspeaker or severe coloration changes when moving their head by a few centimeters. However, these people are very pleased whenever given the opportunity to enjoy a well-adjusted good-sounding Ambisonic system.

The area around the center of an Ambisonic system in which a sound field can be recreated accurately, the *physical sweet spot*, increases with the Ambisonic order and decreases with frequency. Using 1<sup>st</sup> order, the physical sweet spot for 700 Hz tightly includes the head of a single listener ( $r_{\text{sweet}} = 0.08$  m) [3]. In the last 15 years, higher-order systems have been investigated increasingly [4]. However, a physical sweet spot that provides sound to 4 listeners simultaneously ( $r_{\text{sweet}} = 0.50$  m) requires at least an order of 7 (64 loudspeakers for a full sphere) at 700 Hz.

Nevertheless, perceptual evaluations and practical experience during that time indicated that there might be a *perceptual sweet spot* that is much larger. Alternatively to the interpretation of Ambisonics as sound field synthesis technique [5], it can be seen as an amplitude-panning method that takes advantage of the perceptual effect of a phantom source [6]. It is known that level differences between the loudspeakers create inter-aural time differences at low frequencies [7], which are the dominant localization cues in the horizontal plane [8]. This contribution summarizes selected experimental results in order to render assistance in how to set up an Ambisonic system to our best current knowledge.

The paper first gives an overview about the basics of Ambisonics, including decoder strategies and order weighting. The overview introduces quality elements that have been adjusted in the reported experiments. Referring to the concept of quality from [9, 10], experimental results are presented for various quality features individually: localization, source width, coloration, and loudness. The conclusion gives practical recommendations for a good Ambisonic sound.

## 2. Ambisonics

Ambisonics [1, 2, 4, 11] records and reproduces the sound field excitation in terms of orthogonal basis functions. In the horizontal two-dimensional case, these functions are the circular harmonics, in the three-dimensional case the spherical harmonics, respectively. Their maximum order  $N$  determines the spatial resolution and the minimum number density of loudspeakers required for reproduction. In comparison to vector-base amplitude panning (VBAP) [12] that does not have such a modal representation, Ambisonics is designed to enable playback with direction-independent, smooth quality on arbitrary loudspeaker arrangements. Typically, similarly to VBAP, Ambisonics rather works based on the perceptual effect of phantom sources [6] than as physical sound field synthesis.

In this contribution, the Cartesian direction vector is defined as  $\boldsymbol{\theta} = [\cos(\varphi)\sin(\vartheta), \sin(\varphi)\sin(\vartheta), \cos(\vartheta)]^T$  that depends on the azimuth  $\varphi$  and zenith angles  $\vartheta$ . For one source at a direction  $\boldsymbol{\theta}_s$ , the Ambisonic spectrum  $\mathbf{y}_N(\boldsymbol{\theta}_s)$  is calculated by evaluating the circular/spherical harmonics at  $\boldsymbol{\theta}_s$ . This calculation is frequency-independent and assumes that all sources and loudspeakers lie at the same radius  $r$ .

The decoder derives the gains  $\mathbf{g} = \{g_1, \dots, g_L\}$  for the  $L$  loudspeakers of an arrangement from the Ambisonic spectrum  $\mathbf{y}_N(\boldsymbol{\theta}_s)$  by multiplication with the decoder matrix  $\mathbf{D}$ :

$$\mathbf{g} = \mathbf{D} \text{diag}\{\mathbf{a}_N\} \mathbf{y}_N(\boldsymbol{\theta}_s). \quad (1)$$

The matrix is derived from the circular/spherical harmonic spectra  $\mathbf{y}_N(\boldsymbol{\theta}_l)$  of each loudspeaker  $\mathbf{Y}_N = [\mathbf{y}_N(\boldsymbol{\theta}_1), \mathbf{y}_N(\boldsymbol{\theta}_2), \dots, \mathbf{y}_N(\boldsymbol{\theta}_L)]$ .

### 2.1. Decoder Strategies

Decoding Ambisonics to regular loudspeaker arrangements, e.g. full spheres or circles, is simple and the application of the traditional decoder strategies *sampling* and *mode-matching* [13] is feasible. The sampling decoder can be calculated by transposition of  $\mathbf{Y}_N$ , whereas mode matching requires its pseudo-inverse  $\mathbf{Y}_N^T(\mathbf{Y}_N\mathbf{Y}_N^T)^{-1}$ . On irregular arrangements, both strategies can cause fluctuations in localization, source width and energy. Introducing regularization [14] to mode-matching  $\mathbf{Y}_N^T(\mathbf{Y}_N\mathbf{Y}_N^T + \alpha\mathbf{I})^{-1}$  can improve its performance (this article uses  $\alpha = \frac{1}{4}\frac{L}{\pi}$ , cf. [15]).

Alternatively, the *energy-preserving* strategy reduces the set of basis functions according to the geometry of the loudspeaker arrangement and subsequently applies singular value decomposition in order to omit all singular values [16]. Other strategies employ elaborate optimization to find suitable decoder matrices, e.g. [17, 18].

In contrast, *AllRAD* is a very efficient strategy [15]. It decodes Ambisonics to an optimal virtual loudspeaker arrangement by sampling which results in a decoder that is at the same time mode-matching and energy-preserving. The signals of these virtual loudspeakers are mapped to the real loudspeakers using VBAP.

### 2.2. Order Weighting

A weighting vector  $\mathbf{a}_N$  can be applied as a spatial low-pass filter in the harmonics domain in order to control the main and side lobes that emerge from the truncation of the harmonics [4]. The attenuation of the side lobes is intended to stabilize localization at off-center listening positions. The *basic* weighting uses a vector of ones  $\mathbf{a}_N = \mathbf{1}$  and results in the narrowest possible main lobe and the strongest side lobes. In contrast, *in-phase* weighting completely suppresses the side lobes at the cost of a wider main lobe. A trade-off can be found in the *max-r<sub>E</sub>* weighting that maximizes the energy towards the panning direction.

## 3. Concept of Quality

In [9] the assessment of *quality* is defined as the "*judgment of the perceived composition of an entity with respect to its desired composition*". The desired composition typically refers to external or internal references [10]. On the one hand, quality is composed of multiple *quality features*, i.e. the particular perceived characteristics of the entity. On the other hand, *quality elements* are the parameters that can be adjusted and that may influence quality.

### 3.1. Quality Elements

In the experimental results that are discussed here, the influence of the following quality elements of Ambisonics has been evaluated:

- number of loudspeakers,
- Ambisonic order and weighting,
- loudspeaker array radius and delay compensation,
- decoder strategy,
- reverberation time of the reproduction room.

### 3.2. Quality Features

This contribution discusses a selection of particular quality features that have been evaluated in the literature about Ambisonics:

- localization,
- source width,
- coloration,
- loudness.

Although according to [19], timbral features dominate preference, evaluation mostly focused on spatial features. The author is not aware of any publication about the importance of loudness fluctuations, however they are assumed to be similarly important as timbral features.

## 4. Localization

The localization error (absolute difference between the perceived and the panning direction) decreases for higher orders at the central listening position for both horizontal panning on a loudspeaker ring (Figure 1 left, [20]) and for vertical panning on a loudspeaker hemisphere (Figure 1 right, [21]). This finding agrees with the results from other studies [22, 23, 24]. As the order is proportional to the spatial resolution, higher orders require large number of loudspeakers. Increasing the number of loudspeakers without increasing the order does not improve localization.

The localization error increases with the distance to the central listening position. Figure 2 indicates a doubling of the error for a listening position that was about 2.5m away from the center of a loudspeaker ring with an average radius of about 5 m. Independent of the listening position, max- $r_E$  weighting yields the smallest errors. Interestingly, the compensation of the different delays towards the central listening position increases the errors for all weightings tested [20]. The results in Figure 3 illustrate the effect of weighting at off-center positions: Without weighting (i.e. basic weighting), the phantom source splits into two different auditory events, one at the desired direction and one at the direction of the nearest loudspeaker (at  $90^\circ$ ). With max- $r_E$ , the side lobes that drive this loudspeaker are attenuated, so that the splitting does not occur [25].

A recent study [26, 27] investigated the influence of the loudspeaker array radius on the off-center localization errors. For two different radii, the errors at the same relative positions were similar, cf. Figure 4. The results suggest that the level differences (that are depending on the relative position) are more important than the different delays (that are depending on the absolute position). This suggestion is supported by results from [28, 29].

The investigation in [15] compared different decoder strategies on an irregular semicircular loudspeaker arrangement and employed the direction of the energy vector as localization predictor [25, 30, 31, 32]. The traditional sampling and particularly mode-matching yielded large localization errors, whereas regularization of the mode-matching decoder and the energy-preserving decoder resulted in smaller errors. The smallest errors were achieved by AllRAD.

In general, long reverberation times of the reproduction room impair localization [33] and should be avoided for any sound reproduction system that uses loudspeakers.

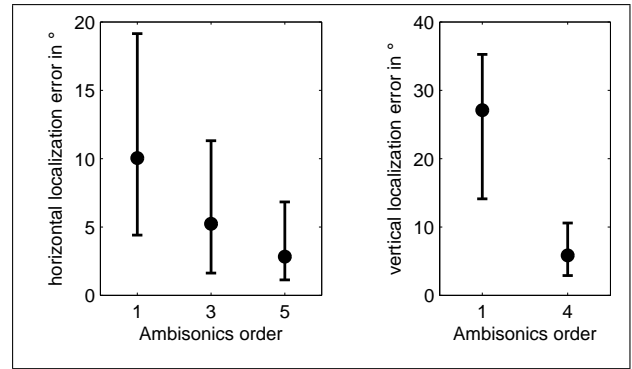


Figure 1. Medians and corresponding interquartile ranges of horizontal (left, [20]) and vertical (right, [21]) localization errors (absolute difference between perceived and panning angle) for max- $r_E$  Ambisonics with different orders.

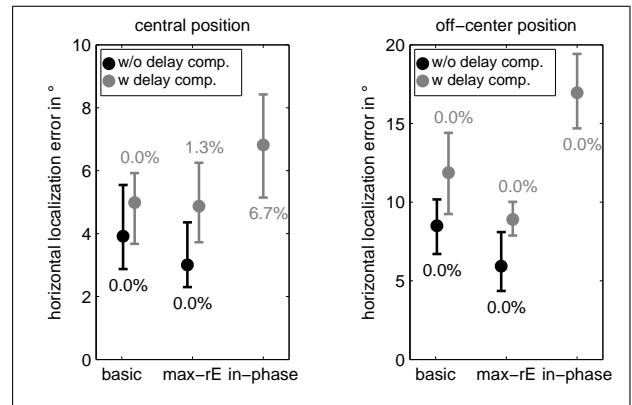


Figure 2. Medians of and corresponding 95% confidence intervals of horizontal localization errors for 5<sup>th</sup> order using different weightings with/without delay compensation to the central listening position, evaluated at the central (left) and an off-center listening position (right) [20].

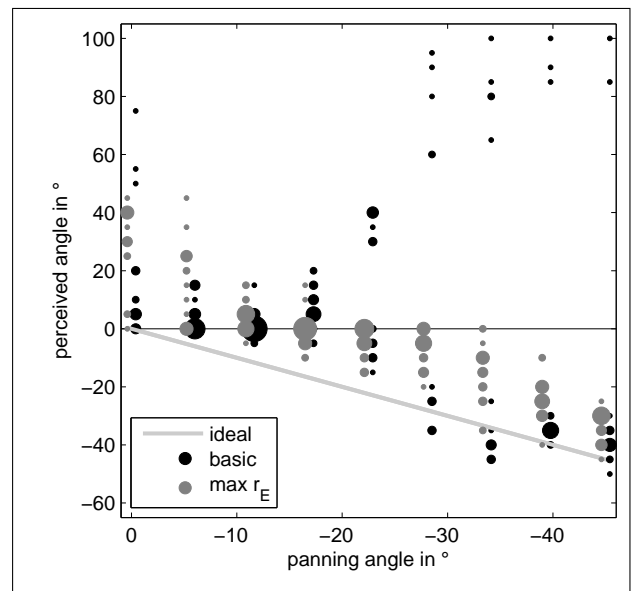


Figure 3. Histograms of perceived angles in dependence of the panning angle for different weightings at an off-center listening position [25]. The radii of the circles indicate the relative number of values within a range of  $5^\circ$ .

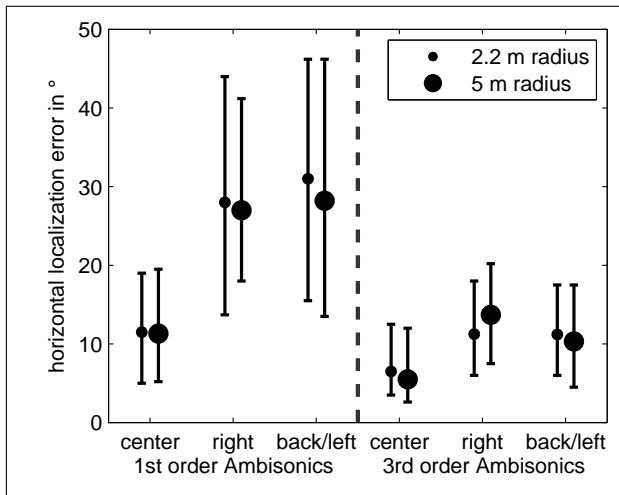


Figure 4. Medians and corresponding interquartile ranges of horizontal localization errors in dependence of listening position and Ambisonics order for 2 loudspeaker array radii [26, 27].

## 5. Source Width

Smooth movements of sources require a source width that is independent of the panning angle. Figure 5 shows for 3<sup>rd</sup> order on a regular ring of 8 loudspeakers, that the widest source was perceived for a panning direction that coincides with the direction of a loudspeaker. Panning further in between the loudspeakers (from quarter to half way between the loudspeakers) yields significantly narrower sources. This dependency disappears for higher orders and more loudspeakers [25, 34].

The dependency of the source width on different decoder strategies was investigated in [15] for an irregular semicircular loudspeaker arrangement employing the length of the energy vector as a width predictor [25, 35]. As for the localization, the mode-matching strategy yields the worst results, i.e. the strongest dependency on the panning direction. Regularization, sampling, and energy-preserving strategies result in a smoother source width distribution. The best results were achieved by AllRAD.

## 6. Coloration

The coloration changes of a moving pink noise source (clockwise and counter-clockwise rotation with 0.1° per ms) were investigated in [25]. The investigation revealed that Ambisonics with max- $r_E$  weighting produced less coloration changes than basic Ambisonics. This result was independent of the order/number of loudspeakers and the listening position. Interestingly, the amount of coloration significantly increases with the order/number of loudspeakers (only weakly for max- $r_E$  at the central position). This agrees with the results using musical stimuli in [36], where the conditions with fewer loudspeakers were mostly preferred.

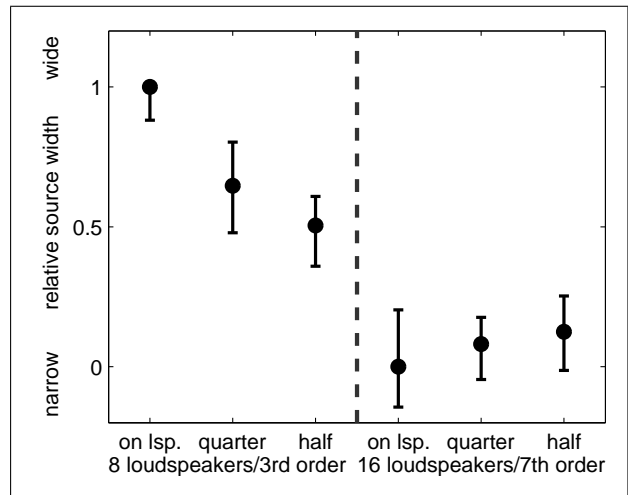


Figure 5. Perceived phantom source width using max- $r_E$  Ambisonics with different orders/numbers of loudspeakers in dependence of the panning angle relative to the loudspeakers [25].

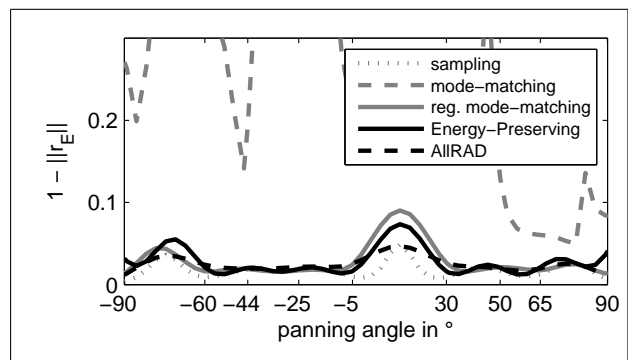


Figure 6. Predicted source width ( $1 - ||r_E||$ ) in dependence of the panning angle using different decoder strategies on an irregular loudspeaker semicircle (loudspeaker angles indicated on the x-axis) with 7<sup>th</sup> order [15].

Due to the large number of neighboring active loudspeakers, in-phase weighting was reported to cause stronger coloration than all other weightings in [20]. The same effect occurs when using a large number of loudspeakers with low orders [37].

Delay compensation of differently distant loudspeakers towards the central listening position creates a physically accurate sound field around this position. However, the accurate area is typically smaller than a human head at high frequencies. Thus, small head movements around the central listening position can yield strong phase variations that cause severe changes in coloration and localization [20, 38].

Although one might consider the anechoic listening condition ideal, it draws the listeners' attention to comb filters that emerge from the superposition of the loudspeaker signals. The audibility of these comb filters can be concealed by adding reflections or choosing a more reverberant playback environment [39].

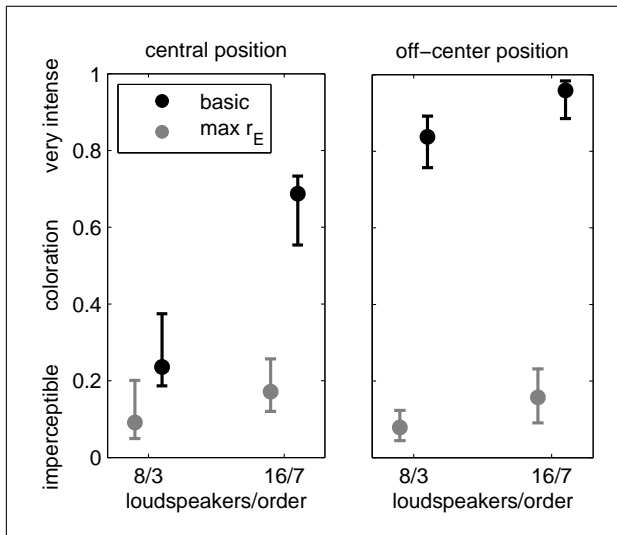


Figure 7. Coloration changes during source movement employing Ambisonics with different orders/numbers of loudspeakers and weightings at central (left) and off-center listening position (right) [25].

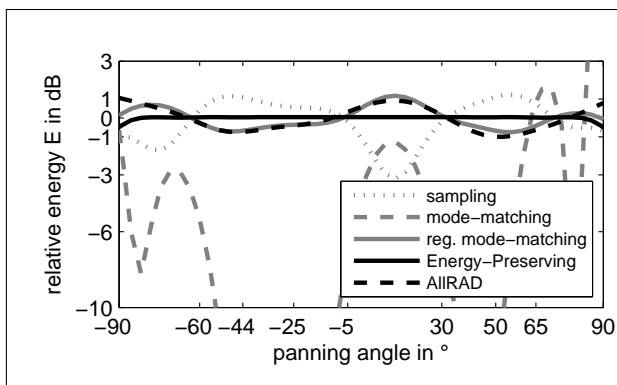


Figure 8. Relative overall energy variation in dependence of the panning angle using different decoder strategies on an irregular loudspeaker semicircle (loudspeaker angles indicated on the x-axis) with 7<sup>th</sup> order [15].

## 7. Loudness

Whenever a suitable decoder is found for a given loudspeaker arrangement, the loudness does not depend on the number of loudspeakers or the order. Employing the overall energy as a simple predictor for loudness, the study in [15] investigated loudness fluctuation for different decoder strategies. Again, the un-regularized mode-matching decoder is impractical. With regularization the fluctuation decreases to  $\pm 1$ dB. AllRAD yields similar results, whereas sampling is worse. Best results were achieved by the energy-preserving decoding strategy that lives up to its name.

## 8. Conclusion

Physically, the most accurate Ambisonics system would employ a mode-matching decoder with basic weighting and delay compensation in an anechoic room. Perceptually, this is not a good idea unless infinitely many loudspeakers and infinite order are available to bridge the apparent uncanny valley of spatial sound reproduction [40].

A large number density of loudspeakers can yield a good localization performance, especially at off-center positions. However, the order has to be chosen accordingly. Too low orders activate too many neighboring loudspeakers with the same signal resulting in coloration. This holds true for in-phase weighting that exaggerates side lobe attenuation. Too high orders can cause imbalanced timbre, source width and loudness, especially when employing traditional decoder strategies on irregular loudspeaker arrangements. In contrast, a large number of loudspeakers can yield stronger coloration. Severe coloration can also be caused by delay compensation, which indeed does not improve localization in the center. Nevertheless, a non-anechoic listening room can help to conceal coloration.

The author knows from personal experience that 4<sup>th</sup> or 5<sup>th</sup> order max- $r_E$ -weighted AllRAD-decoded Ambisonics on hemispheres with about 20...30 loudspeakers works well. Without delay compensation in studio-like rooms, such a system offers a perceptual sweet-spot of about 2/3 of the loudspeaker array's radius, i.e. 30 listeners for an array radius of 3 m.

## Acknowledgement

The author thanks Peter Stitt for providing his currently submitted article. This work was partly 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 Styria. The Austrian Research Promotion Agency (FFG) conducted the funding under the Competence Centers for Excellent Technologies (COMET, K-Project), a program of the institutions above.

## References

- [1] D. H. Cooper, T. Shiga: Discrete-matrix multichannel stereo. *Journal of the Audio Engineering Society* **20** (1972) 346–360.
- [2] M. A. Gerzon: With-height sound reproduction. *Journal of the Audio Engineering Society* **21** (1973) 2–10.
- [3] D. Ward, T. Abhayapala: Reproduction of a plane-wave sound field using an array of loudspeakers. *IEEE Transactions on Speech and Audio Processing* **9** (September 2001) 697–707.
- [4] J. Daniel: Représentation de champs acoustiques, application à la transmission et à la reproduction de scènes sonores complexes dans un contexte multimédia. Dissertation. Université Paris 6, 2001.

- [5] J. Ahrens, S. Spors: An analytical approach to sound field reproduction using circular and spherical loudspeaker distributions. *Acta Acustica united with Acustica* **94** (2008) 988–999.
- [6] K. Wendt: Das Richtungshören bei der Überlagerung zweier Schallfelder bei Intensitäts- und Laufzeitstereophonie. Dissertation. RWTH Aachen, Germany, 1963.
- [7] D. M. Leakey: Some measurements on the effects of interchannel intensity and time differences in two channel sound systems. *The Journal of the Acoustical Society of America* **31** (1959) 977–986.
- [8] F. L. Wightman, D. J. Kistler: The dominant role of low-frequency interaural time differences in sound localization. *The Journal of the Acoustical Society of America* **91** (1992) 1648–1661.
- [9] U. Jekosch: Voice and speech quality perception: Assessment and evaluation. Springer, 2005.
- [10] A. Raake, S. Egger: Quality of experience. Springer, 2014, Ch. Quality and Quality of Experience, 11–33.
- [11] D. G. Malham, A. Myatt: 3D Sound Spatialization using Ambisonic Techniques. *Computer Music Journal* **19** (1995) 58–70.
- [12] V. Pulkki: Virtual sound source positioning using vector base amplitude panning. *Journal of the Audio Engineering Society* **45** (1997) 456–466.
- [13] M. A. Poletti: A Unified Theory of Horizontal Holographic Sound Systems. *Journal of the Audio Engineering Society* **48** (2000) 1155–1182.
- [14] M. Poletti: Robust two-dimensional surround sound reproduction for nonuniform loudspeaker layouts. *Journal of the Audio Engineering Society* **55** (2007) 598–610.
- [15] F. Zotter, M. Frank, H. Pomberger: Comparison of energy-preserving and all-round ambisonic decoders. *Fortschritte der Akustik, AIA-DAGA*, Meran, 2013.
- [16] F. Zotter, H. Pomberger, M. Noisternig: Energy-Preserving Ambisonic Decoding. *Acta Acustica united with Acustica* **98** (2012) 37–47.
- [17] D. Moore, J. Wakefield: A design tool to produce optimized ambisonic decoders. *Audio Engineering Society Conference: 40th International Conference: Spatial Audio: Sense the Sound of Space*, 2010.
- [18] N. Epain, C. Jin, F. Zotter: Ambisonic decoding with constant angular spread. *EAA Joint Symposium on Auralization and Ambisonics*, Berlin, 2014. Submitted to *Acta Acustica united with Acustica*.
- [19] F. Rumsey, S. Zieliński, R. Kassier, S. Bech: On the relative importance of spatial and timbral fidelities in judgments of degraded multichannel audio quality. *The Journal of the Acoustical Society of America* **118** (2005) 968–976.
- [20] M. Frank, F. Zotter: Localization experiments using different 2D Ambisonics decoders. 25. Tonmeistertagung, Leipzig, 2008.
- [21] S. Braun, M. Frank: Localization of 3D Ambisonic Recordings and Ambisonic Virtual Sources. *International Conference on Spatial Audio*, Detmold, 2011.
- [22] E. Benjamin, A. Heller, R. Lee: Localization in Horizontal-Only Ambisonic Systems. *Audio Engineering Society Convention 121*, San Francisco, 2006.
- [23] G. Kearney, E. Bates, F. Boland, D. Furlong: A comparative study of the performance of spatialization techniques for a distributed audience in a concert hall environment. *Audio Engineering Society Conference: 31st International Conference: New Directions in High Resolution Audio*, London, 2007.
- [24] S. Bertet, J. Daniel, E. Parizet, O. Warusfel: Investigation on localisation accuracy for first and higher order ambisonics reproduced sound sources. *Acta Acustica united with Acustica* **99** (2013) 642–657.
- [25] M. Frank: Phantom sources using multiple loudspeakers in the horizontal plane. Dissertation. University of Music and Performing Arts Graz, Austria, 2013.
- [26] P. Stitt, S. Bertet, M. van Walstijn: Perceptual investigation of image placement with ambisonics for non-centred listeners. *DAFx-13*, Maynooth, 2013.
- [27] P. Stitt, S. Bertet, M. van Walstijn: Off-centre localisation performance of ambisonics and HOA for large and small loudspeaker array radii. *EAA Joint Symposium on Auralization and Ambisonics*, Berlin, 2014. Submitted to *Acta Acustica united with Acustica*.
- [28] N. Peters: Sweet [re]production: developing sound spatialization tools for musical applications with emphasis on sweet spot and off-center perception. Dissertation. McGill University, 2011.
- [29] J. Nettingsmeier, D. Dohrmann: Preliminary studies on large-scale higher-order ambisonic sound reinforcement systems. *Ambisonics Symposium*, Lexington, 2011.
- [30] M. A. Gerzon: General metatheory of auditory localisation. *Audio Engineering Society Convention 92*, Vienna, 1992.
- [31] F. Wendt, M. Frank, F. Zotter: Amplitude panning with height on 2, 3, and 4 loudspeakers. *International Conference on Spatial Audio*, Erlangen, 2014.
- [32] M. Frank: Localization using different amplitude-panning methods in the frontal horizontal plane. *Proc. of the EAA Joint Symposium on Auralization and Ambisonics*, Berlin, 2014, 42–47.
- [33] W. M. Hartmann: Localization of sound in rooms. *The Journal of the Acoustical Society of America* **74** (1983) 1380–1391.
- [34] M. Frank, G. Marentakis, A. Sontacchi: A simple technical measure for the perceived source width. *Fortschritte der Akustik, DAGA*, Düsseldorf, 2011.
- [35] M. Frank: Source width of frontal phantom sources: Perception, measurement, and modeling. *Archives of Acoustics* **38** (2013) 311–319.
- [36] G. Marentakis, F. Zotter, M. Frank: Vector-base and ambisonic amplitude panning: A comparison using pop, classical, and contemporary spatial music. *EAA Joint Symposium on Auralization and Ambisonics*, Berlin, 2014. Submitted to *Acta Acustica united with Acustica*.
- [37] A. Solvang: Spectral impairment of two-dimensional higher order ambisonics. *Journal of the Audio Engineering Society* **56** (2008) 267–279.
- [38] A. J. Tucker, W. L. Martens, G. Dickens, M. P. Hollier: Perception of reconstructed sound-fields: The dirty little secret. *Audio Engineering Society Conference: 52nd International Conference: Sound Field Control - Engineering and Perception*, Guildford, 2013.
- [39] O. Santala, H. Vertanen, J. Pekonen, J. Oksanen, V. Pulkki: Effect of listening room on audio quality in ambisonics reproduction. *Audio Engineering Society Convention 126*, Munich, 2009.
- [40] F. Rumsey: Sound field control. *Journal of the Audio Engineering Society* **61** (2013) 1046–1050.