Comparison of Higher Order Ambisonics And Wave Field Synthesis With Respect to Spatial Discretization Artifacts in Time Domain

Jens Ahrens, Hagen Wierstorf, and Sascha Spors

Quality and Usability Lab, Deutsche Telekom Laboratories, Technische Universität Berlin, Ernst-Reuter-Platz 7, 10587 Berlin, Germany

Correspondence should be addressed to Jens Ahrens (`jens.ahrens@telekom.de`)

ABSTRACT

We present a time domain analysis and comparison of spatial discretization artifacts in near-field compensated higher order Ambisonics and wave field synthesis. Simulations of both methods on the same circular loudspeaker array are investigated and the results are interpreted in terms of fundamental psycho-acoustical properties of the human auditory system, most notably the precedence effect. It can be shown that both methods exhibit fundamental different properties regarding the synthesized first arriving wave fronts as well as additional correlated wave fronts (echoes). The properties of both types of wave fronts are a consequence of the combination of the spatial bandwidth of the loudspeaker driving function and the fact that a finite number of spatially discrete loudspeakers are employed.

1. INTRODUCTION AND MOTIVATION

Wave field synthesis (WFS) and near-field compensated higher order Ambisonics (HOA) constitute the two best known representatives of analytical methods for sound field synthesis. Both methods have been derived from very different directions and have not been formulated in a common framework until recently [1].

HOA was developed from rather intuitive yet physical considerations [2]. The theory was later extended [3] and recently a solid physical interpretation in terms of the *single-layer potential* solution was found [4] which retroactively justifies the approach. The loudspeaker driving signals in order to reproduce a given desired sound field are obtained via a straightforward solution of the reproduction equation.

WFS on the other hand directly implements fundamental physical principles like the Rayleigh integrals or the Kirchhoff-Helmholtz integral [5, 6]. While HOA is restricted to spherical and circular loudspeaker arrays, WFS may be employed with planar, linear, and arbitrary convex loudspeaker contours.

It has been shown that the solutions obtained by both HOA and WFS are not essentially different from a physical perspective when continuous spherical and circular secondary source (i.e. loudspeaker) distributions are considered. The only notable difference is the fact that WFS constitutes a high-frequency approximation of HOA [7]. The psycho-acoustical consequences of this circumstance are not clear but are assumed to be marginal.

What essentially distinguishes the two approaches is the way they are implemented: The loudspeaker driving function in WFS typically exhibits infinite spatial bandwidth (or *infinite order*), the driving function in HOA is spatially bandlimited to a given order [1]. To put it in one sentence, WFS constitutes a high-frequency approximation of infinite order HOA. For convenience, we will not use the spatial bandwidth in order to refer to the different types of synthesis as it is done in [8], but will use the terms WFS and HOA as representatives for infinite spatial bandwidth and finite spatial bandwidth synthesis respectively.

It has recently been shown by the authors in [1, 8] that the spatial bandwidth of the loudspeaker driving signal has fundamental influence on the properties of the synthesized sound field when spatially discrete loudspeaker setups are considered. More explicitly, the spatial bandwidth of the driving function has a fundamental influence on the properties of the desired components of the synthesized sound field as well as on the structure and energy distribution of spatial discretization artifacts. However, analyses of the consequences on human perception are only partly available, e.g. in [9]. Note that we exclusively consider the physical synthesis of sound fields in this paper. At this stage, we do not take into account any psychoacoustic optimizations such as performed e.g. in [10, 11].

The analyses available in the literature such as [1, 8] focus on monochromatic scenarios and therefore on spectral characteristics of spatial discretization artifacts. But it has recently been shown in [12] that the temporal characteristics of spatial discretization artifacts in sound field synthesis can have essential influence on the perceived quality. In the temporal domain such artifacts can occur as correlated signals arriving before (pre-echoes) or after (echoes) the desired virtual source signal. So far, pre-echoes have only be observed in the synthesis of focused virtual sound sources in WFS [13]. Time domain simulations of WFS have also been presented in the classical WFS literature such as [14] and HOA simulations have been presented in [3]. However, detailed analysis and comparison have not been performed.

The critical property of the human auditory system to mention at this point is the *precedence effect* which is a fundamental mechanism in spatial hearing [15, 16]. The precedence effect describes the phenomenon that the direction of a perceived sound is not altered by echoes of this sound which may arrive from different directions in a time window of 1–40 ms after the leading wave front. Also, the echoes are not perceived as such but as a room impression, so that in the time window of 1–40 ms fusion to one auditory percept occurs. In the case of virtual sources the possibility hence exists that the spatial discretization artifacts have no influence on the perceived direction of the auditory event and are not perceived as echoes. This means also that pre-echoes are more critical than echoes, because they arrive before the desired wave front and can influence the perceived direction due to the precedence effect. On the other hand, the precedence effect only occurs if the relative level of the repetition occurring after the leading wave front is not higher than 10–15 dB. So if the amplitude of the wave front from the virtual source is much higher than the amplitudes of the pre-echoes, the pre-echoes will be audible as an additional auditory event.

2. FRAMEWORK

For the simulations analyzed in Sec. 3 a loudspeaker array is assumed using parameters which correspond to those of the loudspeaker array installed at the Usability Laboratory at Deutsche Telekom Laboratories. It is composed of 56 equiangularly spaced loudspeakers on a circle with a nominal radius of 1.495 m. For simplicity, we assume omnidirectional loudspeakers and free-field conditions in the simulation.

This paper does not introduce new aspects of the theory. A review of the theory is therefore waived in favor of a compact analysis of the properties of the synthesized sound field (references are mentioned below).

The scenario investigated in this paper is so-called *2.5-dimensional* synthesis. The term 2.5-dimensional refers to the fact that loudspeakers with a three-dimensional spatio-temporal transfer function are used and synthesis in a plane, i.e. in two dimensions, is targeted. This is the scenario which is mostly implemented in practice, e.g. [17, 18].

2.5-dimensional sound field synthesis exhibits a number of restrictions compared to three-dimensional sound field synthesis. Most notably, only sound fields propagating in the target plane, typically the horizontal plane, can be synthesized. Furthermore, the amplitude decay of synthesized sound fields typically deviates from the desired one. Both restrictions do not play a fundamental role in the presented investigation. It can be shown that spatial discretization artifacts have similar fundamental properties in three and 2.5 dimensions. The reader is referred to [19] for a detailed treatment of the theory of 2.5-dimensional HOA and to [5, 6] for treatments of 2.5-dimensional WFS. A brief description of the preparation of the simulations follows below.

The WFS solution for the individual loudspeaker driving signals is directly available in time domain apart from the well known pre-filter which has to be designed in temporal frequency domain. The HOA solution is exclusively available in temporal frequency domain. Infinite impulse response representations of the HOA driving signals in time domain are available [20] which are not convenient to be employed in the presented simulations. We therefore performed a sampling of the analytical temporal frequency domain representation and obtained the (finite-length) time domain representation via an inverse numerical Fourier transform [21].

As mentioned above, we assume the loudspeakers to be omnidirectional. They do thus not alter the input signals. If the response of the system at a specific position is desired, the individual loudspeaker signals have to be delayed and attenuated according to the distance of the position under consideration to the individual loudspeakers. Finally, the delayed and attenuated loudspeaker signals have to be added.

In the remainder of the paper the synthesis of a virtual plane wave sound field which propagates inside the horizontal plane is considered.

3. RESULTS

Fig. 1 shows still images of the spatio-temporal impulse response of the loudspeaker system under consideration when driven in order to reproduce a virtual plane wave with propagation direction $\theta_{\text{pw}} = -\frac{\pi}{2}$ (i.e. downwards in the plots) for different time instances. A cross-section through the horizontal plane is shownn. Fig. 1(a), 1(c), and 1(e) show HOA, Fig. 1(b), 1(d), and 1(f) show WFS.

Fig. 2 shows impulse responses of the loudspeaker system for a specific listening position for HOA (Fig. 2(a) and 2(c)) and WFS (Fig. 2(b) and 2(d)). Fig. 2(c) and 2(d) show the impulse responses from Fig. 2(a) and 2(b) respectively but lowpass and highpass filtered with cutoff frequencies f_{cutoff} as indicated. In all figures the absolute value of the sound pressure is shown in dB. The time t is chosen such that the virtual plane wave front passes the center of the loudspeaker array at $t = 0$ ms.

As described in Sec. 3.1, the major findings which can be deduced from time domain simulations are the properties of the first arriving wave fronts and the occurrence of additional and correlated wave fronts (echoes) which are a consequence of the chosen spatial bandwidth of the driving function in combination with the fact that a finite number of spatially discrete loudspeakers is employed. As stated earlier this leads to spatial discretization artifacts above the *spatial aliasing frequency* f_{al}. In WFS reproduction, f_{al} is approximately constant over the entire listener area. For the present loudspeaker array, it lies between $f_{\text{al}} = 1400$ Hz and $f_{\text{al}} = 2500$ Hz depending on the listening position.

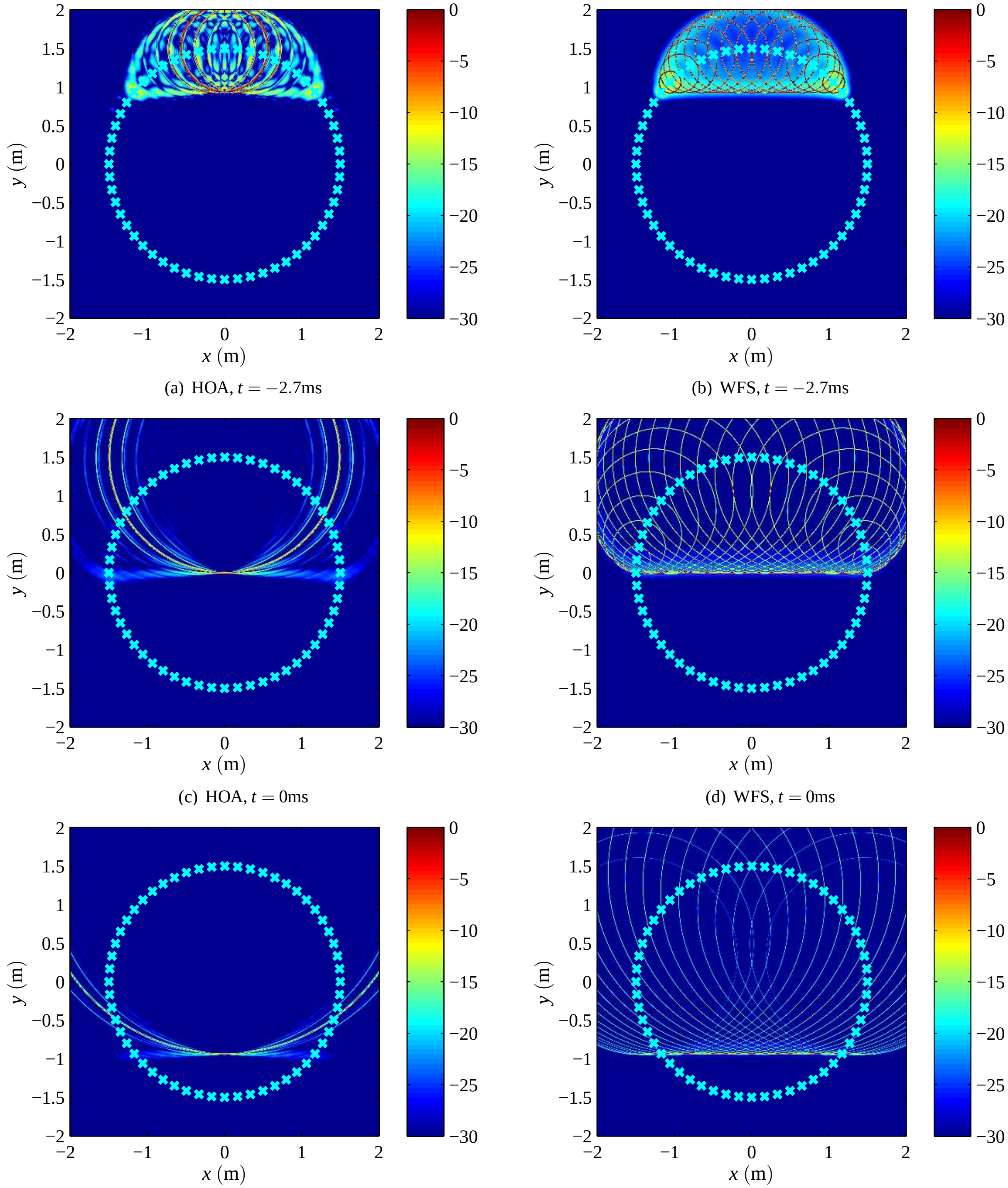
This situation is more complicated for HOA reproduction. Here, spatial aliasing in the strict sense does not occur but a reconstruction error which is also referred to as spatial aliasing [19]. In HOA it is such that an almost artifact-free region evolves around the center of the secondary source distribution which gets smaller with frequency. For frequencies below 1400 Hz, this artifact-free region fills the entire receiver area and reaches the size of a human head at approximately 10 kHz for the present loudspeaker array [1].

Sec. 3.1 summarizes the observations deduced from the illustrations in Fig. 1 and Fig. 2. Sec. 3.2 interprets the observations in terms of perception.

3.1. First wave front and echoes

WFS exhibits a pronounced first wave front at all listening positions. Above the spatial aliasing frequency this first wave front is slightly distorted but keeps its planar shape. Spatial aliasing artifacts in the form of high-frequency echoes ($> f_{\text{al}}$) follow the first wave front for all listening positions. As pointed out in [5], WFS can be seen as *wave front synthesis*. The broadband first wave front is followed by a dense sequence of echoes of approximately similar amplitude for $0 \text{ ms} < t < 0.2$ ms (refer to Fig. 1(d)). This dense sequence is followed by a slightly sparser sequence of high-frequency echoes for $0.2 \text{ ms} < t < 6$ ms with decreasing amplitude. The time interval between successive echoes in the sparser part of the impulse response is in the order of some hundred μs. These high-frequency echoes arrive from various directions and are rather homogeneously distributed over the entire receiver area. It can be shown that each of the active loudspeakers produces one of these echoes. Consequently, larger loudspeaker setups lead to longer impulse responses and a larger loudspeaker spacing leads to longer intervals between the echoes.

In HOA the plane wave front is accurately synthesized around the central listening position (refer to Fig. 1(c)). At other listening positions, especially at positions lateral to the center, the synthesized sound field consists of a number of echoes which impinge at different times and from different directions on the listener. Comparison of Fig. 1(c) with monochromatic simulations from [8] reveals that the first wave front arriving carries the temporal low frequency content. This is also confirmed by the impulse response of the loudspeaker system driven with HOA, as depicted in Fig. 2(c). The thick red curve represents energy below $f_{\text{cutoff}} = 2200$ Hz, the thin blue curve represents energy above f_{cutoff}. The virtual plane wave is accurately synthesized at these low temporal frequencies whereby it exhibits a slightly concave shape containing some distortion for positions lateral to the center. After



(a) HOA, $t = -2.7$ms

(b) WFS, $t = -2.7$ms

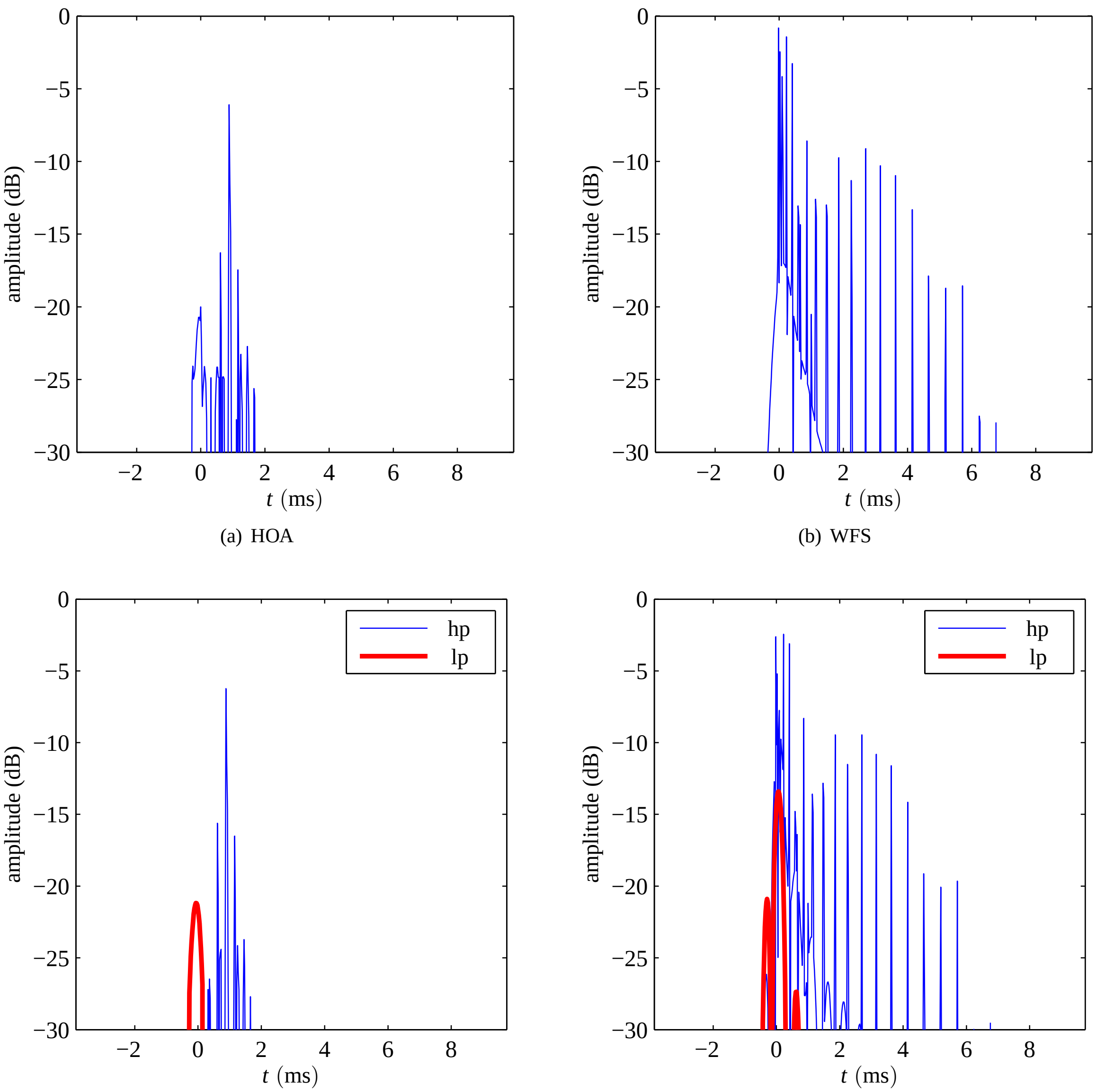
(c) HOA, $t = 0$ms

(d) WFS, $t = 0$ms

(e) HOA, $t = 2.7$ms

(f) WFS, $t = 2.7$ms

Fig. 1: Impulse responses of the loudspeaker system in the horizontal plane when driven in order to reproduce a virtual plane wave with propagation direction $\theta_{\text{pw}} = -\frac{\pi}{2}$ (downwards in the plot). The absolute value of the time domain sound pressure is shown in dB for different instances of time. The left column shows HOA, the right column shows WFS. The marks indicate the positions of the loudspeakers.



(a) HOA (b) WFS

(c) HOA, $f_{\text{cutoff}} = 2200\text{Hz}$ (d) WFS, $f_{\text{cutoff}} = 2000\text{Hz}$

Fig. 2: Impulse responses of the loudspeaker system at position $x = 1$ m, $y = 0$ m when driven in order to reproduce a virtual plane wave with propagation direction $\theta_{\text{pw}} = -\frac{\pi}{2}$. Fig. 2(c) and 2(d) show the impulse responses from Fig. 2(a) and 2(b) but highpass ('hp') and lowpass ('lp') filtered with a cutoff frequency of f_{cutoff}. The absolute value of the sound pressure is shown in dB. The plane wave passes the center of the array at $t = 0$ ms with amplitude 0 dB.

the first wave front, a number of echoes arrive successively from a direction which approximately coincides which the direction of that loudspeaker at which the virtual plane wave first "touches" the loudspeaker contour. Comparison of Fig. 1(c) with monochromatic simulations from [8] reveals that these echoes contain temporal high frequencies. Again, this is confirmed by Fig. 2(c). Note that the loudest echo is almost at 15 dB above the first wave front (Fig. 2(a)). The distance in time between the adjacent wave fronts is significantly lower than 1 ms for the loudspeaker system under consideration. A wider loudspeaker spacing leads to a larger distance between the wave fronts.

It is evident from comparing Fig. 1(c) and 1(f) (and Fig. 2(a) and 2(b)) that the impulse response of the system is significantly shorter for HOA than for WFS for a given listener position. While no considerable energy is present at all positions for $y > 0$ m in HOA for $t = 2.7$ ms in Fig. 1(c) the discretization artifacts in WFS are still obvious (Fig. 1(d)).

For completeness we mention here one property of HOA which is not evident in Fig. 1 or 2 but has to be deduced from the monochromatic simulations in [8]. It is the fact that the energy distribution over the entire listening area is very inhomogeneous for frequencies above the spatial aliasing frequency. At certain locations dependent on the considered frequency, the synthesized sound field exhibits a significantly lower amplitude than desired.

3.2. Perception

The accurate synthesis of the first wave front in WFS leads to very good auditory localization for non-focused sources over the entire listening area [22, 23, 24]. This is in accordance with the conclusion that can be drawn from Sec. 3.1 based on the precedence effect. The high-frequency echoes due to spatial discretization are not perceivable as echoes nor do they change the perceived direction of the virtual plane wave. Recall that the echoes arrive in a time window smaller than 6 ms, are lower in amplitude, and contain fewer spectral components than the first wave front. Informal listening[1] confirms absence of perceivable echoes, but the echoes do add some sense of spaciousness. This is another well known phenomenon of the precedence effect and enables us to perceive rooms. Due to the unnatural pattern of echoes and the corresponding comb filter spectrum also slight coloration is perceivable.

For HOA we have a temporal separation between the wave front for low and high frequencies, therefore it exists no spectral overlap between the first wave front and the later echoes. This leads to a weaker precedence effect [26]. Also the high-frequency echoes are 15 dB greater in amplitude than the first wave-front. This suggests that the high temporal frequency content of the virtual source is localized in direction of the loudspeakers producing these echoes (see above) in contrast to the low frequency content which impinges from the desired direction. Informal listening shows that high and low temporal frequency content are indeed localized at different directions for listening positions lateral to the center. The virtual source is thus split into two sources. One source emits exclusively the high temporal frequency content, the other source emits the low temporal frequency content.

In general, HOA provides a less homogeneous perception than WFS when the entire listening area is considered. On the other hand, at the center of the loudspeaker array HOA is expected to cause less coloration than WFS due to the absence of any echoes at this location in HOA.

4. CONCLUSIONS

We have presented a time domain analysis of artifacts occurring in 2.5-dimensional near-field compensated higher order Ambisonics (HOA) and wave field synthesis (WFS). Based on numerical simulations it was shown that the wave fronts in WFS are properly synthesized and are followed by discretization artifacts which resemble echoes from various directions. These echoes contain energy exclusively above the spatial aliasing frequency. Based on existing literature on the precedence effect, it was concluded that these echoes are inaudible, but that they add some sense of spaciousness and timbral coloration. The interpretation from [5] that WFS is actually *wave front synthesis* is thus confirmed. This fact is also supported by the close relation of the WFS theory to Huygens' principle.

The first wave front in HOA on the other hand is split into a low-frequency first wave front and high-frequency echoes which arrive from different directions for some listening positions. Since there is no spectral overlap between the first wave front and the echoes we expect that the virtual source is split into two sources at different positions. One of the perceived sources emits the low-frequency content, the other source emits the high-frequency content.

Informal listening confirms above described conclusions[1]. A formal perceptual experiment in order to confirm the presented analysis is in preparation.

[1] Audio examples including all binaural cues can be downloaded from [25].

5. REFERENCES

[1] S. Spors and J. Ahrens. A comparison of wave field synthesis and higher-order Ambisonics with respect to physical properties and spatial sampling. In *125th Convention of the AES*, San Francisco, CA, Oct. 2–5 2008.

[2] M. A. Gerzon. With-height sound reproduction. *JAES*, 21:2–10, 1973.

[3] J. Daniel. Spatial sound encoding including near field effect: Introducing distance coding filters and a viable, new Ambisonic format. In *23rd International Conference of the AES*, Copenhagen, Denmark, May 23–25 2003.

[4] F. Fazi, P. Nelson, and R. Potthast. Analogies and differences between 3 methods for sound field reproduction. In *Ambisonics Symposium*, Graz, Austria, June 25–27 2009.

[5] A. J. Berkhout, D. de Vries, and P. Vogel. Acoustic control by wave field synthesis. *JASA*, 93(5):2764–2778, May 1993.

[6] S. Spors, R. Rabenstein, and J. Ahrens. The theory of wave field synthesis revisited. In *124th Convention of the AES*, Amsterdam, The Netherlands, May 17–20 2008.

[7] J. Ahrens and S. Spors. On the secondary source type mismatch in wave field synthesis employing circular distributions of loudspeakers. In *127th Convention of the AES*, New York, NY, Oct. 9–12 2009.

[8] J. Ahrens and S. Spors. Alterations of the temporal spectrum in high-resolution sound field reproduction of varying spatial bandwidths. In *126th Convention of the AES*, Munich, Germany, May. 7–10 2009.

[9] H. Wittek. Perceptual differences between wavefield synthesis and stereophony. PhD thesis, University of Surrey, 2007.

[10] M. A. Gerzon. Practical periphony: The reproduction of full-sphere sound. In *65th Convention of the AES*, London, UK, Feb. 25–28 1980.

[11] H. Wittek, F. Rumsey, and G. Theile. Perceptual enhancement of wavefield synthesis by stereophonic means. *JAES*, 55(9):723–751, Sep. 2007.

[12] M. Geier, H. Wierstorf, J. Ahrens, I. Wechsung, A. Raake, and S. Spors. Perceptual evaluation of focused sources in wave field synthesis. In *128th Convention of the AES*, London, UK, May 22–25 2010.

[13] S. Spors, H. Wierstorf, M. Geier, and J. Ahrens. Physical and perceptual properties of focused sources in wave field synthesis. In *127th Convention of the AES*, New York, NY, Oct. 9–12 2009.

[14] P. Vogel. Application of wave field synthesis in room acoustics. PhD thesis, Delft University of Technology, 1993.

[15] H. Wallach, E. B. Newman, and M. R. Rosenzweig. The precedence effect in sound localization. *American Journal of Psychology*, 57:315–336, 1949.

[16] W. Haas. The influence of a single echo on the audibility of speech. *Acustica*, 1:49–58, 1951.

[17] J. Daniel, R. Nicol, and S. Moreau. Further investigations of high order Ambisonics and wavefield synthesis for holophonic sound imaging. In *114th Convention, March 22-25*, Amsterdam, The Netherlands, 2003.

[18] D. de Vries. *Wave Field Synthesis*. AES Monograph. AES, New York, 2009.

[19] J. Ahrens and S. Spors. An analytical approach to sound field reproduction using circular and spherical loudspeaker distributions. *Acta Acustica utd. with Acustica*, 94(6):988–999, Nov./Dec. 2008.

[20] H. Pomberger. Angular and radial directivity control for spherical loudspeaker arrays. M. Sc. thesis, IEM Graz, 2008.

[21] B. Girod, R. Rabenstein, and A. Stenger. *Signals and Systems*. J.Wiley & Sons, New York, 2001.

[22] E. W. Start. Direct sound enhancement by wave field synthesis. PhD thesis, Delft University of Technology, 1997.

[23] W. de Brujin. Application of wave field synthesis in videoconferencing. PhD thesis, Delft University of Technology, 2004.

[24] J. Sanson, E. Corteel, and O. Warusfel. Objective and subjective analysis of localization accuracy in wave field synthesis. In *124th Convention of the AES*, Amsterdam, The Netherlands, May 2008.

[25] Quality and Usability Lab. Spatial Audio Research - Audio research at Quality and Usability Lab. http://audio.qu.tu-berlin.de/?p=175.

[26] R. Y. Litovsky, H. S. Colburn, W. A. Yost, and S. J. Guzman. The precedence effect. *JASA*, 106(4):1633–1654, 1999.