EVALUATION OF PERCEPTUAL PROPERTIES OF PHASE-MODE BEAMFORMING IN THE CONTEXT OF DATA-BASED BINAURAL SYNTHESIS

Sascha Spors and Hagen Wierstorf

Quality and Usability Lab, Deutsche Telekom Laboratories, Technische Universität Berlin, Germany
{Sascha.Spors, Hagen.Wierstorf}@telekom.de

ABSTRACT

Several approaches to data-based binaural synthesis have been published that process the sound field captured by a spherical microphone array. The captured sound field is typically decomposed into plane waves which are then auralized using head-related transfer functions. The decomposition into plane waves is often based upon modal beamforming techniques which represent the sound field with respect to surface spherical harmonics. The achievable spatial bandwidth is limited due to practical considerations. This paper investigates the perceptual implications of these limitations in the context of data-based binaural synthesis.

Index Terms— binaural synthesis, modal beamforming, binaural room transfer functions, perception

1. INTRODUCTION

Head-related transfer functions (HRTFs) capture the acoustic transmission path from an acoustic source to the outer ear. HRTFs vary to some degree amongst individuals. Typically the case of free-field propagation is referred to as HRTFs, while similar transfer functions captured in a room are referred to as binaural room transfer functions (BRTFs). Both are used for instance for the synthesis of virtual sources in virtual auditory environments, by filtering the (dry) signal of a virtual source with the respective left and right HRTF/BRTFs. This approach to sound reproduction is termed as binaural reproduction or synthesis. Typically headphones are used to reproduce the left and right ear signals.

Binaural synthesis using HRTF/BRTFs is limited to the auralization of a finite number of individual virtual sources. For good results, head-tracking and dynamic exchange of the HRTFs is mandatory. Diffuse sound fields, for instance cafeteria noise, cannot be represented by transfer functions. Hence, a head-tracked auralization of diffuse sound fields cannot be achieved straightforwardly by binaural synthesis. The perceived quality is furthermore increased by using individualized HRTF/BRTFs. While the measurement effort for individualized HRTFs is already quite high, similar measurements for BRTFs have to be repeated for all acoustic environments to be auralized. It would be desirable to separate the room effect from the BRTFs, so that the room effect can be added to individualized HRTFs. These limitations of BRTF-based binaural synthesis can be overcome by combining techniques from sound field analysis with HRTF-based binaural synthesis.

In this paper we follow the concept presented in [1, 2, 3]. Here the sound field is captured by a spherical microphone array and decomposed into plane waves. The plane waves are then filtered by the respective (far-field) HRTFs. This approach can be used to either synthesize diffuse sound fields or to compute BRTFs. For the former, the actual sound field has to be captured in real-time, while for the latter impulse responses measured from the source to the microphone array are suitable. By using individual HRTFs one can derive individualized BRTFs. We focus here on the computation of BRTFs.

Practical implementations are limited with respect to the number of sampling positions of the spherical microphone array. As a result, data-based binaural synthesis suffers from inaccuracies resulting from the limited spatial bandwidth that can be captured. Recently studies have been published with respect to the perceptual impact of these limitations. For instance in [3] a HRTF dataset has been represented in terms of spherical harmonics. The interaural time differences (ITDs) and interaural level differences (ILDs) have been investigated for a varying spatial bandwidth. In [2] data-based binaural synthesis of a sound field represented by spherical harmonics has been considered. Here the interaural cross correlation (ICC) has been investigated. This paper extends the results from previous studies by considering spectral cues, as well as the influence of a varying head orientation. Furthermore, a model of human perception is used for the studies.

2. DATA-BASED BINAURAL SYNTHESIS

A propagating sound field $P(\mathbf{x},\omega)$ can be represented in terms of a superposition of plane waves with spectra $\bar{P}(\phi,\theta,\omega)$ impinging from all directions with incidence angles ϕ,θ as [4]

$$P(\mathbf{x},\omega) = \frac{1}{4\pi} \int_0^{2\pi}\int_0^{\pi} \bar{P}(\phi,\theta,\omega) e^{-i\mathbf{kx}} \sin\theta \, d\theta d\phi \,, \tag{1}$$

where $\mathbf{x} = (x,y,z)$ denotes a position in space, $\mathbf{k}$ the wave vector of a plane wave with $\mathbf{k} = \frac{\omega}{c}(\cos\phi\sin\theta, \sin\phi\sin\theta, \cos\theta)$, ϕ its azimuth and θ its colatitude. The horizontal plane is defined by $\theta = \pi/2$.

Far-field or plane wave HRTFs $\bar{H}_{\text{L,R}}(\phi,\theta,\gamma,\delta,\omega)$ represent the acoustic transmission path from a plane wave with incidence angle ϕ,θ to the left and right ear under the head orientation γ,δ. In the first approximation and under free-field conditions $\bar{H}_{\text{L,R}}$ depend only on the difference between the two angle pairs ϕ,θ and γ,δ. Far-field HRTFs can be extrapolated from HRTFs measured at finite distance [5], but for most applications HRTFs with a distance greater than one meter to the source are appropriate.

The sound pressure at the left/right ear $P_{\text{L,R}}(\gamma,\delta,\omega)$ for a given head-orientation γ,δ can be computed by superposition of the respective far-field HRTFs $\bar{H}_{\text{L,R}}(\phi,\theta,\gamma,\delta,\omega)$ filtered by the plane wave expansion coefficients $\bar{P}(\phi,\theta,\omega)$

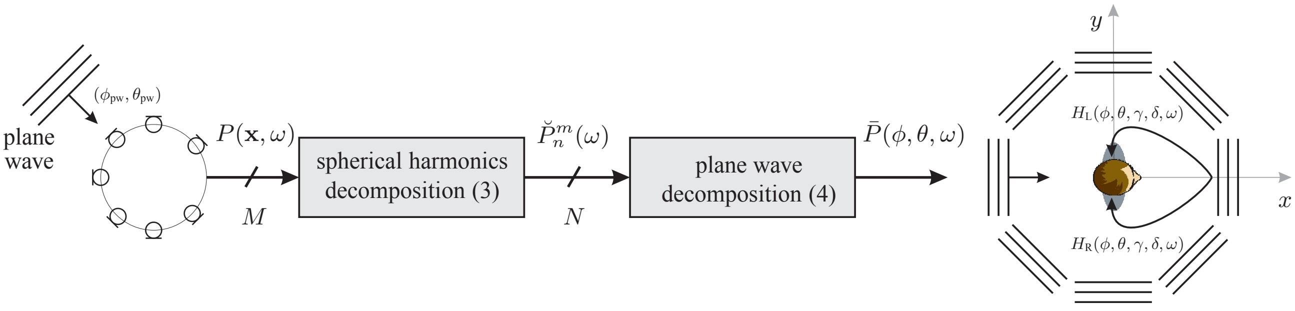
$(\phi_{\text{pw}}, \theta_{\text{pw}})$
plane wave
$P(\mathbf{x}, \omega)$
M
spherical harmonics decomposition (3)
$\breve{P}_n^m(\omega)$
N
plane wave decomposition (4)
$\bar{P}(\phi, \theta, \omega)$
y
x
$H_{\text{L}}(\phi, \theta, \gamma, \delta, \omega)$
$H_{\text{R}}(\phi, \theta, \gamma, \delta, \omega)$

Fig. 1. Block diagram of data-based binaural synthesis using a plane wave representation of the captured sound field.

$$P_{\text{L,R}}(\gamma, \delta, \omega) = \int_0^{2\pi}\int_0^{\pi} \bar{P}(\phi, \theta, \omega) \bar{H}_{\text{L,R}}(\phi, \theta, \gamma, \delta, \omega) \sin\theta \; d\theta d\phi \; . \tag{2}$$

The above concept to data-based binaural synthesis using an expansion of the captured sound field with respect to plane waves has a number of benefits: head-tracked dynamic binaural synthesis is straightforward to achieve, numerous techniques are known to perform a plane wave decomposition with microphone arrays, it provides variable spatial resolution and horizontal-only synthesis can be formulated by quite simple means.

Spherical microphone arrays exhibit properties which are independent from the incidence direction of sound and are therefore preferred for the analysis of sound fields. The next section briefly discusses the basic theory of spherical microphone arrays and the related techniques to perform a plane wave decomposition.

3. SPHERICAL MICROPHONE ARRAYS

Due to the underlying geometry it is natural to represent the sound field captured on the surface of a sphere with respect to surface spherical harmonics. Various techniques have been published for acoustically transparent or rigid spheres equipped with pressure or cardioid microphones [6, 1]. For clarity of presentation we discuss the open sphere case only.

The spherical harmonics expansion coefficients $\breve{P}_n^m(\omega)$ can be computed from the sound pressure $P(\mathbf{x}, \omega)$ captured on an acoustically transparent sphere with radius R as [6]

$$\breve{P}_n^m(\omega) = \frac{1}{j_n(\frac{\omega}{c}R)} \int_0^{2\pi}\int_0^{\pi} P(\mathbf{x}, \omega) Y_n^{-m}(\beta, \alpha) \sin\beta \; d\beta d\alpha \; , \tag{3}$$

where the $Y_n^m(\cdot)$ denotes the n-th order surface spherical harmonic of m-th degree, $j_n(\cdot)$ the n-th order spherical Bessel function [4] and $\mathbf{x} = R(\cos\alpha \sin\beta, \sin\alpha \sin\beta, \cos\beta)$. The expansion of a sound field in terms of plane waves (1) can be linked to the spherical harmonics expansion coefficients as

$$\bar{P}(\phi, \theta, \omega) = \sum_{n=0}^{\infty} \sum_{m=-n}^{n} i^n \breve{P}_n^m(\omega) Y_n^m(\theta, \phi) \; . \tag{4}$$

Equation (3) together with (4) forms the basis to calculate the plane wave expansion $\bar{P}(\phi, \theta, \omega)$ of a sound field captured by a spherical microphone array. Figure 1 illustrates the principle of data-based binaural synthesis as introduced so far.

In practice it is only possible to measure the pressure on a limited number of positions on the sphere. Therefore a spatially discretized version of (3) is used. It can be shown that spatial sampling limits the order $n \leq N$ up to which the spherical harmonics coefficients $\breve{P}_n^m(\omega)$ can be computed without prominent spatial sampling artifacts. However, the sound field of a plane wave is not strictly bandlimited in its spherical harmonics representation. As a consequence of (1), any natural sound field is not strictly bandlimited. A reasonable approximation of a plane wave expressed in spherical harmonics can be achieved within a region of radius R by choosing $N > \lceil \frac{e\pi}{c} Rf \rceil$, where $\lceil \cdot \rceil$ denotes the ceiling function [7]. Regarding the typical audio bandwidth of 20 kHz and the size of a typical human head with $R = 0.09$ m one would need to compute the expansion coefficients up to order $N \geq 45$.

The acoustic pressure signals captured by microphones are subject to equipment noise. This leads to further limitations with respect to the low frequency behavior of a spherical microphone array. In order to derive the expansion coefficients at low frequencies differential mechanisms are exploited implicitly in (3). This leads to an amplification of the equipment noise for high orders n. In order to maintain a reasonable white-noise gain (WNG) the order is typically limited for low frequencies in practical applications .

As a consequence of the considerations given so far, practical microphone arrays are not capable to capture the required order $N \geq 45$ over the full audio frequency range. This holds especially for the lower frequencies. Since we are aiming at binaural synthesis for human listeners the question arises which order is required to cope for the capabilities of the human ear. In order to investigate on this, we consider the effect of order-limitation without considering spatial sampling and equipment noise.

The plane wave decomposition of an incident plane wave with unit amplitude and incidence angle $(\phi_{\text{pw}}, \theta_{\text{pw}})$ impinging on an ideal spatially continuous microphone array without sensor noise reads [6]

$$\bar{P}_{\text{ideal}}(\phi, \theta, \omega) = \sum_{n=0}^{N} \frac{2n+1}{4\pi} P_n(\cos\Theta) \; , \tag{5}$$

where $P_n(\cdot)$ denotes the n-th Legendre polynomial and Θ the angle between the incidence angle of the plane wave $(\phi_{\text{pw}}, \theta_{\text{pw}})$ and the look direction of the plane wave decomposition (ϕ, θ). It is evident from (5) that the ideal array response is frequency independent. It can furthermore be concluded from [6] that the spatial selectivity of the plane wave decomposition decreases with decreasing order N.

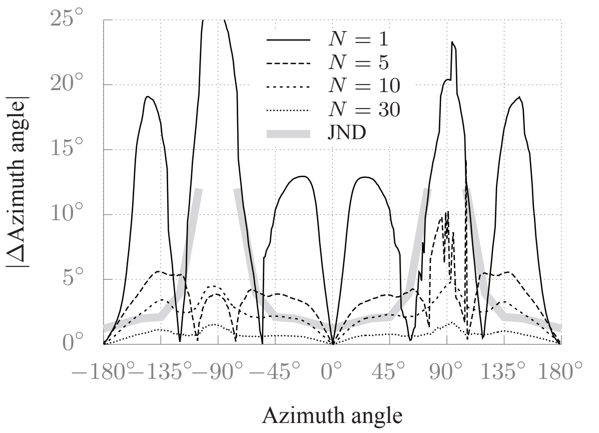

Fig. 2. Deviation of the perceived direction of the computed BRTFs from the original HRTFs for different orders N.

4. EVALUATION

4.1. Setup

We aim at investigating the influence of a limited spatial resolution on psychoacoustic relevant properties of data-based binaural synthesis. For this purpose the signal processing chain depicted in Fig. 1 was implemented in MATLAB. Due to the lack of full sphere HRTF datasets with high resolution, we used a horizontal plane HRTF database with one degree resolution [8]. Elevated plane waves are not considered as incident sound field, neither in the plane wave decomposition. Hence, the chosen setup constitutes a 2.5-dimensional scenario.

A unit amplitude plane wave with $\phi_{\text{pw}} = 0°$, $\theta_{\text{pw}} = 90°$ is chosen as incident sound field. In order to exclude effects emerging from spatial sampling, the plane wave decomposition $\bar{P}(\phi, \theta, \omega)$ was computed analytically using (5). Instead of synthetic far-field HRTFs, measured ones at 3 m distance are used. We computed BRTF datasets for $\gamma = -180° \ldots 180°$, $\delta = 90°$ and different orders N in one degree steps in order to investigate the effect of head-movements. Listening examples are available for informal listening [1].

The human auditory system has a remarkable performance in estimating the localization of a sound source even in the presence of diffuse background noise or in reverberant situations. This ability is achieved by exploring different characteristics of the sound field present at the two ears. Beside non acoustical cues like vision, or the change of the sound field with head movements, spectral cues and interaural differences between the ears are used by the auditory system. Therefore it is important that the computed BRTFs preserve these characteristics and deviate only in an inaudible range. The next section deals with the interaural differences, spectral cues are regarded in Section 4.3.

4.2. Directional perception

The two most important features for the estimation of the direction of arrival of a sound in the horizontal plane are ITDs, and ILDs. To investigate the influence of the order N on the perceived direction for a sound convolved with the computed BRTFs, a binaural model after [9] is applied. The binaural model simulates the behavior of the two ears by applying a gammatone filterbank to the input signals of the left and right ear. In the frequency bands between 200 Hz and 1300 Hz the interaural phase difference (IPD) and the ILD are calculated. The IPD is ambiguous for frequencies greater than 700 Hz. In addition the ILD has its maximum around 60° and is ambiguous for higher angles. On the other hand the sign of the ILD can be easily used to overcome the ambiguity of the IPD and to calculate the real ITD for a given stimulus. At the same time this mechanism accounts for the dominance of the perceived direction by the ITD for the considered frequency range [10].

In order to obtain an estimation of the azimuth for a calculated ITD, a lookup table is required that maps calculated ITDs on corresponding azimuth angles. Such a table was created from the 3 m HRTF data set described in the Section 4.1 for azimuth angles $\gamma = -90° \ldots 90°$ in one degree steps. The azimuth angles behind the listener can not be differentiated from the frontal ones by using only ITD or ILD values. Hence the angles coming from the back of the dummy head were handled as if they were coming from the corresponding angles in the frontal hemisphere. For the prediction of the perceived direction for the computed BRTFs, each set of BRTF signals was convolved with the same white noise signal of 1 s length. The resulting ear signals were then fed into the binaural model together with the lookup table to generate a prediction of the perceived azimuth angle. This was conducted for various orders N and the available azimuth angles (see section 4.1). The same procedure was applied to the HRTF data set to obtain the desired perceived azimuth direction. The deviation of the azimuth $\Delta\gamma$ of the computed BRTFs is characterized by the absolute difference between the estimated azimuth of the BRTF and the HRTF.

The result is presented in Fig. 2. As can be seen, $\Delta\gamma$ depends on the direction of incidence which is more pronounced for lower orders N. The performance of direction discrimination of the human listener depends also on the direction of incidence of a sound source. A common measure of this performance is the just notable difference (JND) between two angles [11]. The JND is shown as the grey line in Fig. 2. The JND for angles in the back of the listener is normally larger than in the front, but as the binaural model was not able to differentiate between front and back we used the same JND for the back as in the frontal hemisphere. In [11], Mills was not able to measure the JND for an incident angle of 90°, because it was greater than 40° which was the maximum deviation their apparatus was possible of.

As can be seen, the deviation $\Delta\gamma$ is almost imperceptible for orders of $N \geq 10$, for an order of $N = 5$ small deviations occur, but for an order of $N = 1$ the performance degraded strongly for certain angles. In this case the directivity pattern of the beamformer (plane wave decomposition) is clearly visible in the results. The data is summarized in Table 1 as mean and max values over the azimuth angle.

	Azimuth deviation	
N	mean	max
0	42°	83°
1	10°	28°
3	5°	15°
5	3°	14°
10	2°	5°
30	0.7°	2°
50	0.4°	1°

Table 1. Mean and maximum deviation of the estimated azimuth angle for the computed BRTFs with respect to the order N.

[1] http://audio.qu.tu-berlin.de/?p=778

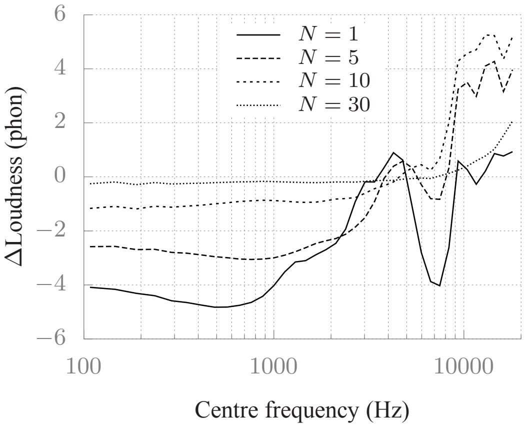

Fig. 3. Deviation of the magnitude of computed BRTFs from the measured HRTFs for different orders N.

4.3. Distance perception and coloration

In addition to the perceived direction of a source, its localization is determined by its perceived distance, which depends on the amplitude of the source and on the naturalness of the given BRTFs. Changes of the spectrum or the interaural differences of the BRTF can lead to in-head localization of the auditory event [12]. Whereby Hartmann [12] found a stronger dependency on the spectrum than on the interaural parameters. A quantitatively explanation of in-head localization is not possible at the moment and there exist results which show no strong dependency on the spectrum [13]. Another perceptual dimension which is influenced by spectral changes is coloration. The coloration of a stimuli in comparison to a baseline condition can be audible, if changes in the spectrum exceeds 1 dB.

To investigate the influence of spectral changes on the order N, the deviation in loudness of a 1 s noise signal convolved with the left ear computed BRTFs in comparison with the measured HRTFs was computed. This was performed for all azimuth angles γ. To account for the spectral resolution of the auditory system the spectrum was calculated in auditory filters ranging from 100 Hz to 20000 Hz and applying a loudness compression to the power of 0.54 to the sound pressure [14]. The result is presented in Fig. 3. In contrast to the perceived azimuth angle the deviations of the spectral content of the computed BRTFs will be audible already at quite high orders.

This is confirmed by an informal listening by the authors. Distance and coloration of the perceived source can already be distinguished from that of the HRTF at an order of $N \leq 10$. In-head localization is slightly approaching for orders of $N \leq 5$.

5. CONCLUSIONS

This paper presents a detailed analysis of the perceptual properties of data-based binaural synthesis. The influence of a limited spatial bandwidth, as occurring in practical realizations of sound field analysis, has been investigated independently of spatial sampling artifacts using a model of human auditory perception. This paper extends the results from previous studies [2, 3] by considering spectral cues, as well as the influence of a varying head orientation. The results allow for a number of conclusions. The perceived direction can be preserved until very low orders of around $N = 5$ which was confirmed by informal listening by the authors and the results presented in [3]. But the informal listening confirmed also the fact that the spectral changes are critical and will be perceived already at higher orders. The spectral deviations resulted in a coloration of the stimuli as well as in a smaller perceived distance of the source and an in-head localization for orders $N \leq 10$.

6. REFERENCES

[1] R. Duraiswami, D.N. Zotkin, Z. Li, E. Grassi, N.A. Gumerov, and L.S. Davis, "System for capturing of high order spatial audio using spherical microphone array and binaural head-tracked playback over headphones with head related transfer function cue," in *119th AES Convention*, New York, USA, October 2005, Audio Engineering Society (AES).

[2] B. Rafaely and A. Avni, "Interaural cross correlation in a sound field represented by spherical harmonics," *Journal of the Acoustic Society of America*, vol. 127, no. 2, pp. 823–828, February 2010.

[3] A. Avni and B. Rafaely, "Sound localization in a sound field represented by spherical harmonics," in *International Symposium on Ambisonics and Spherical Acoustics*, Paris, France, May 2010.

[4] N.A. Gumerov and R. Duraiswami, *Fast Multipole Methods for the Helmholtz Equation in three Dimensions*, Elsevier, 2004.

[5] S. Spors and J. Ahrens, "Generation of far-field head-related transfer functions using sound field synthesis," in *German Annual Conference on Acoustics (DAGA)*, March 2011.

[6] B. Rafaely, "Phase-mode versus delay-and-sum spherical microphone array processing," *IEEE Signal Processing Letters*, vol. 12, no. 10, pp. 713–716, 2005.

[7] R.A. Kennedy, P. Sadeghi, T.D. Abhayapala, and H.M. Jones, "Intrinsic limits of dimensionality and richness in random multipath fields," *IEEE Transactions on Signal Processing*, vol. 55, no. 6, pp. 2542–2556, 2007.

[8] H. Wierstorf, M. Geier, A. Raake, and S. Spors, "A free database of head related impulse response measurements in the horizontal plane with multiple distances," in *130th AES Convention*. Audio Engineering Society (AES), May 2011.

[9] Mathias Dietz, Stephan D Ewert, and Volker Hohmann, "Auditory model based direction estimation of concurrent speakers from binaural signals," *Speech Communication*, vol. 53, no. 5, pp. 592–605, May 2011.

[10] F L Wightman and D J Kistler, "The dominant role of low-frequency interaural time differences in sound localization.," *The Journal of the Acoustical Society of America*, vol. 91, no. 3, pp. 1648–61, Mar 1992.

[11] A W Mills, "On the minimum audible angle," *The Journal of the Acoustical Society of America*, vol. 30, no. 4, pp. 237–246, May 1958.

[12] William M Hartmann and Andrew Wittenberg, "On the externalization of sound images.," *The Journal of the Acoustical Society of America*, vol. 99, no. 6, pp. 3678–88, Jun 1996.

[13] A Kulkarni and H S Colburn, "Role of spectral detail in sound-source localization.," *Nature*, vol. 396, no. 6713, pp. 747–9, 1998.

[14] Michael Epstein and Jeremy Marozeau, *Loudness and intensity coding*, pp. 45–69, Oxford University Press, 2010.