

# Data-based Binaural Synthesis Including Rotational and Translatory Head-Movements

Frank Schultz<sup>1</sup> and Sascha Spors<sup>1</sup>

<sup>1</sup>*Institute of Communications Engineering, Universität Rostock, R.-Wagner-Str. 31 (H8), D-18119 Rostock, Germany*

Correspondence should be addressed to Frank Schultz (frank.schultz@uni-rostock.de)

## ABSTRACT

Several approaches to data-based binaural synthesis have been published that are based on the analysis of sound fields captured by spherical microphone arrays. The captured sound field is decomposed into plane waves which are then auralized with head-related transfer functions (HRTFs). So far head-rotations of the listener were considered for dynamic binaural synthesis. We propose an analytic method to consider translatory head-movements as well. A straightforward calculus in the spatial frequency domain is presented and will be evaluated.

## 1. INTRODUCTION

Head-related transfer functions (HRTFs) represent the acoustic transmission path from an acoustic source to the outer ears. HRTFs vary amongst individuals and head/body-orientation with respect to the source. The case of free-field propagation is commonly referred to as HRTFs, while transfer functions which include additional room reflections are referred to as binaural room transfer functions (BRTFs), respectively binaural room impulse responses (BRIRs). Left and right ear drum signals for a virtual acoustic scene are rendered by filtering an anechoic and monaural signal of a virtual source with the left and right ear HRTF/BRTF. This approach to sound reproduction is termed as binaural synthesis. Dynamic binaural synthesis adapts to the actual head-orientation of the listener. Typically headphones are used to reproduce the signals. This requires a sampled grid of HRTFs/BRTFs for all possible head-orientations, which is called dataset below. So far only head-rotations of the listener have been considered in the literature, since this is easily accomplished with a single BRTF dataset. Translatory head-movements would require many BRTF datasets for a (dense) grid of listener positions, which cover the potential translatory head-movements. The measurement effort for such datasets would be considerably. Binaural synthesis using HRTFs/BRTFs suffers from two limitations: (i) scenarios with many sound sources or diffuse sound fields cannot be synthesized efficiently due to the large number of required BRIRs and/or their length, and (ii) measuring

individual BRTFs (for each listener and each individual room) is technically demanding and very time-consuming. These limitations can be overcome by combining techniques from sound field analysis (SFA) and (individual) HRTF-based binaural synthesis. In SFA diverse approaches are known that decompose a captured sound field into plane waves. Spherical microphone arrays exhibit properties, which are independent from the direction of the impinging waves and are therefore preferred for the analysis of multi-path sound fields. Due to the geometry, the captured sound field may be decomposed into surface spherical harmonics which is known as modal beamforming. An alternative decomposition can be realized by so called delay-and-sum beamforming. Practical implementations are limited with respect to the number of sampling positions of the spherical microphone array and equipment self-noise. As a result, data-based binaural synthesis may suffer from inaccuracies resulting from limited spatial bandwidth and noise amplification. Combining SFA and HRTFs, the sound pressure at the left respectively right ear for an actual head-orientation is given as the superposition of the associated (far-field) HRTFs filtered by the plane wave expansion coefficients of the captured sound field. Head-rotations can be considered efficiently by rotation of the plane wave expansion coefficients or the HRTF dataset before superposition. Thus, the SFA data can be used for several individuals. For translatory head-movements an enhanced and new approach can be utilized. The plane wave expansion is first computed with respect to the cen-

ter position of the microphone array. This reference point can then be shifted by applying a phase shift to the plane wave expansion coefficients in the spatial frequency domain. If this phase shift is applied before the filtering process with the far-field HRTFs and the subsequent superposition, the ear-signals for a translatory shifted head position are yielded. Hence, translatory head-movements of the listener can be considered explicitly with no additional measurement effort. The paper introduces the mathematical and practical foundations of the novel approach for translatory head-movements in the context of data-based dynamic binaural synthesis. First the decomposition of a captured sound field into plane waves and the shift theorem for the plane wave expansion is recollected. Then the ear-signals for a shifted head will be derived by filtering and superposition of the shifted plane wave expansion coefficients with the respective HRTFs. The degree of translatory head-movements that can be considered as well as other practical aspects are discussed in the paper.

## 2. NOMENCLATURE

This section defines conventions, assumptions and notations that are used throughout this study. With the angles azimuth  $\alpha, \phi, \gamma \in [0, 2\pi]$  and colatitude  $\beta, \theta, \psi \in [0, \pi]$  the spatial position vector  $\mathbf{x} = (x, y, z) = \|\mathbf{x}\| \cdot (\cos \alpha \sin \beta, \sin \alpha \sin \beta, \cos \beta)$  with  $\|\mathbf{x}\| = r = \sqrt{x^2 + y^2 + z^2}$  and the wavespace vector  $\mathbf{k} = (k_x, k_y, k_z) = \|\mathbf{k}\| \cdot (\cos \phi \sin \theta, \sin \phi \sin \theta, \cos \theta)$  with  $\|\mathbf{k}\| = \omega/c = \sqrt{k_x^2 + k_y^2 + k_z^2}$  and the speed of sound  $c$ ,  $[c]=\text{m/s}$  is used. The head looking direction is described with the unit length vector  $\mathbf{n}_H = (\cos \gamma \sin \psi, \sin \gamma \sin \psi, \cos \psi)$ . The angular temporal frequency  $\omega = 2\pi f$ ,  $[\omega]=\text{rad/s}$  is linked to the temporal frequency  $f$ ,  $[f]=\text{Hz}$ . The temporal sampling frequency is denoted by  $f_s$ . The spherical harmonics definition

$$Y_n^m(\alpha, \beta) = (-1)^m \sqrt{\frac{2n+1}{4\pi} \cdot \frac{(n-|m|)!}{(n+|m|)!}} \times \quad (1)$$

$$P_n^{|m|}(\cos \beta) e^{+im\alpha}$$

[1, eq. (2.1.59)] with  $i^2 = -1$ ,  $n \in \mathbb{N}_0$ ,  $-n \leq m \in \mathbb{Z} \leq +n$ , the Legendre polynomials  $P_n(\cdot)$  and the associated Legendre polynomials  $P_n^{|m|}(\cdot)$  [1, ch. 2.1.2.1] is used. Note that for this definition  $Y_n^m(\alpha, \beta)^* = Y_n^{-m}(\alpha, \beta)$  is valid. The spherical Bessel function of the first kind with order  $n$  is denoted by  $j_n(\cdot)$ . Within the following calculus

the wave vector  $\mathbf{k}$  shall point into the direction of where the wave is coming from. The Fourier transform convention  $p(\mathbf{x}, t) \propto \int P(\mathbf{x}, \omega) e^{+i\omega t} d\omega$  is used for the relationship of the sound pressure  $p(\mathbf{x}, t)$  in time domain and its temporal spectrum  $P(\mathbf{x}, \omega)$ . Thus,  $e^{+i(\mathbf{k}_{PW}, \mathbf{x})} \cdot e^{+i\omega t}$  describes the unit amplitude plane wave with radiating direction  $-\mathbf{k}_{PW}$ . The scalar product notation  $\langle \mathbf{k}_{PW}, \mathbf{x} \rangle = k_{x,PW}x + k_{y,PW}y + k_{z,PW}z$  is used. Note that scaling coefficients in the transformation equations are omitted.

## 3. PLANE WAVE DECOMPOSITION

For the proposed approach we utilize the plane-wave decomposition (PWD) on spherical apertures. The inverse PWD can be defined as

$$P(\mathbf{x}, \omega) \propto \int_0^{2\pi} \int_0^\pi \bar{P}(\phi, \theta, \omega) e^{+i(\mathbf{k}, \mathbf{x})} \sin \theta d\theta d\phi \quad (2)$$

which is a weighted superposition of plane waves coming from  $\mathbf{k}$  over a notional sphere [2]. A single plane wave with the temporal spectrum  $F(\omega)$  coming from  $\mathbf{k}_0$  may then be defined as a spatial delta function

$$\bar{P}(\phi, \theta, \omega) \propto F(\omega) \delta(\cos \theta - \cos \theta_0) \delta(\phi - \phi_0) \quad (3)$$

within the PWD-domain. The modal decomposition of a soundfield captured on a spherical surface into its plane wave contributions [3] is usually referred as to modal beamforming or modal plane-wave decomposition. The spherical harmonics transform (SHT) of a soundfield  $P(\mathbf{x}, \omega)$  captured on a sphere with  $\|\mathbf{x}\| = r = \text{const.}$  is required [1, eq. (2.1.68)]

$$\hat{P}_n^m(\omega) \propto \int_{\alpha=0}^{2\pi} \int_{\beta=0}^\pi P(\mathbf{x}, \omega) Y_n^m(\alpha, \beta)^* \sin \beta d\beta d\alpha. \quad (4)$$

in order to get the modal plane-wave decomposition coefficients for the finite SHT-order  $N_{\text{SHT}}$

$$\bar{P}_{\text{MB}}(\phi, \theta, \omega) \propto \sum_{n=0}^{N_{\text{SHT}}} \sum_{m=-n}^{+n} \frac{\hat{P}_n^m(\omega)}{d_n(\omega)} Y_n^m(\phi, \theta) \quad (5)$$

with the appropriate so called radial filter  $d_n(\omega)$ . For a open sphere, i.e. a notional sphere under condition of free air impedance  $Z_0 = \rho_0 c$  and pressure sensors the radial filter is defined by  $d_n(\omega) = 4\pi i^{+n} j(\frac{\omega}{c} r)$  for the conventions that are used here. The spatially fullband plane-wave decomposition may defined by

$$\bar{P}_{\text{DSB}}(\phi, \theta, \omega) \propto \int_0^{2\pi} \int_0^\pi P(\mathbf{x}, \omega) e^{-i(\mathbf{k}, \mathbf{x})} \sin \theta d\theta d\phi. \quad (6)$$

This method is also known as delay-and-sum beamforming [4, 5]. The delay-and-sum beamformer can be considered as a high frequency and/or farfield approximation of the modal beamformer for  $N_{\text{SHT}} \rightarrow \infty$ .

#### 4. DATA-BASED BINAURAL SYNTHESIS

The basic concept of the presented approach for data-based binaural synthesis is to setup a virtual head into  $\mathbf{0} = (0, 0, 0)$  and to replace the unit amplitude plane waves  $e^{+i(\mathbf{k}, \mathbf{x})}$  in eq. (2) by far-field HRTFs of the virtual head. Thus the sound pressure at the left/right ear  $P_{\text{L,R}}(\gamma, \psi, \omega)$  for a certain head orientation  $\gamma, \psi$  is given by superposition of the respective farfield HRTFs  $\bar{H}_{\text{L,R}}(\phi, \theta, \gamma, \psi, \omega)$  filtered by the plane wave expansion coefficients  $\bar{P}(\phi, \theta, \omega)$

$$P_{\text{L,R}}(\gamma, \psi, \omega) \propto \int_0^{2\pi} \int_0^\pi \bar{P}(\phi, \theta, \omega) \times \bar{H}_{\text{L,R}}(\phi, \theta, \gamma, \psi, \omega) \sin \theta d\theta d\phi. \quad (7)$$

Figure 1 illustrates the used angles for one specific head orientation  $\mathbf{n}_\text{H}$  and one plane wave direction  $\mathbf{k}$ . In [5] the delay-and-sum beamformer as well as the modal beamformer for a continuous and spatially sampled sphere were evaluated within the framework of data-based binaural synthesis. Dynamic binaural synthesis requires the tracking of the actual head position in order to adapt to the actual required HRTF. Commonly, only head rotations will be tracked and dynamically adapted to, while ignoring translatory head-movements, mainly due to the fact that the required HRTF data is missing. The approach in eq. (7) allows to introduce translatory head-movements without changing the HRTF dataset but rather manipulating the PWD dataset  $\bar{P}(\phi, \theta, \omega)$ .

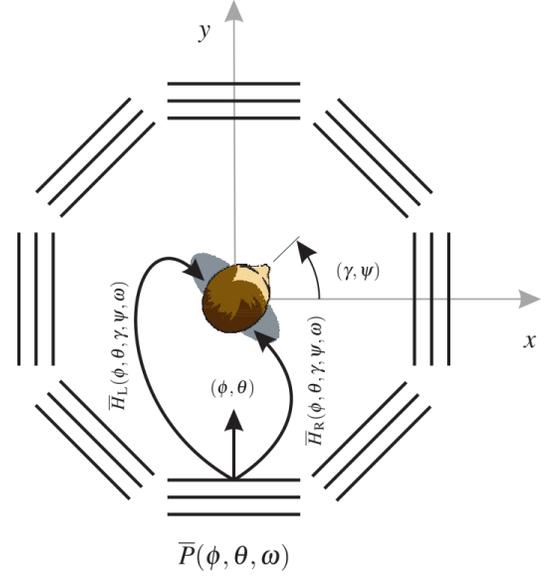
#### 5. TRANSLATORY HEAD-MOVEMENTS

A temporal inverse Fourier transform of the PWD coefficients reveals the temporal response  $\bar{p}(\phi, \theta, t)$  of the beamformer

$$\bar{P}(\phi, \theta, \omega) \bullet \text{---} \bar{p}(\phi, \theta, t). \quad (8)$$

It can be concluded that with  $F(\omega) = 1$  in eq. (3) the single and unit amplitude plane wave coming from  $\phi_0, \theta_0$  results in a temporal Dirac function, i.e. the 'perfect' beamformer

$$\begin{aligned} \bar{P}(\phi, \theta, \omega) &\propto \delta(\cos \theta - \cos \theta_0) \delta(\phi - \phi_0) \bullet \text{---} \\ \bar{p}(\phi, \theta, t) &\propto \delta(t) \delta(\cos \theta - \cos \theta_0) \delta(\phi - \phi_0). \end{aligned}$$



**Fig. 1:** Data-based binaural synthesis using a plane wave expansion of the virtual sound field. For illustration, the filtering of the left/right HRTFs by the plane wave expansion coefficients  $\bar{P}(\phi, \theta, \omega)$  is shown only for one particular direction. The  $z$ -axis points upwards.

Since the plane wave decomposition coefficients can be directly linked to the angular wave spectrum of the soundfield [6] the shifting theorem of the Fourier transform holds [7]. Hence, for a shift to  $\mathbf{x}_\text{T}$  the equation

$$\bar{P}(\phi, \theta, \omega) e^{+i(\mathbf{k}, \mathbf{x}_\text{T})} \bullet \text{---} \bar{p}(\phi, \theta, t + \frac{\langle \mathbf{k}, \mathbf{x}_\text{T} \rangle}{\omega}) \quad (9)$$

is valid and it becomes evident that the spatial shift for  $\bar{P}(\phi, \theta, \omega)$  results in a temporal shift in  $\bar{p}(\phi, \theta, t)$ . Thus, again for  $F(\omega) = 1$  a plane wave coming from  $\phi_0, \theta_0$  results in

$$\begin{aligned} \bar{P}(\phi, \theta, \omega) e^{+i(\mathbf{k}, \mathbf{x}_\text{T})} &\propto \delta(\cos \theta - \cos \theta_0) \delta(\phi - \phi_0) e^{+i(\mathbf{k}, \mathbf{x}_\text{T})} \\ \bullet \text{---} \bar{p}(\phi, \theta, t) &\propto \delta(t + \frac{\langle \mathbf{k}, \mathbf{x}_\text{T} \rangle}{\omega}) \delta(\cos \theta - \cos \theta_0) \delta(\phi - \phi_0). \end{aligned}$$

Note that due to the definition of the shift theorem the translation is exact and perfect for arbitrary translational shifts by  $\mathbf{x}_\text{T}$  as long as the PWD data contains only exact plane waves and is spatially and temporal fullband, i.e. not discretized in space and time. This may be referred to as spatially fullband translation. Thus, a translatory head-movement can be realized by manipulating

the phase of the PWD coefficients. With the reference point  $\mathbf{x} = (0, 0, 0)$  of the virtual head (i.e. also the expansion origin of the PWD data) the soundfield superposition then reads

$$P_{L,R,T}(\gamma, \psi, \omega, \mathbf{x}_T) \propto \int_0^{2\pi} \int_0^\pi \underbrace{\bar{H}_{L,R}(\phi, \theta, \gamma, \psi, \omega)}_{\text{Head Translation}} \times \bar{P}(\phi, \theta, \omega) e^{+i(\mathbf{k}, \mathbf{x}_T)} \sin \theta d\theta d\phi. \quad (10)$$

for a translation of the head to  $\mathbf{x}_T$ . It appears promising and straightforward to discuss the effects of translational shifting combined with spatial and temporal discretization within the spatial PWD domain only and not for the whole eq. (10). When considering time-discrete PWD data with sample index  $k \in \mathbb{Z}$  the temporal Fourier transform is done by the inverse DTFT [7]

$$\bar{p}(\phi, \theta, k) \propto \int_{-\pi}^{+\pi} \bar{P}(\phi, \theta, \Omega) e^{+i\Omega k} d\Omega \quad (11)$$

with angular frequency  $\Omega = 2\pi \frac{f}{f_s}$ . Introducing the translational shift within the inverse DTFT

$$\bar{p}_T(\phi, \theta, k, \mathbf{x}_T) \propto \int_{-\pi}^{+\pi} \underbrace{\bar{P}(\phi, \theta, \Omega)}_{\text{Head Translation}} e^{+i\frac{\langle \mathbf{k}, \mathbf{x}_T \rangle}{\Omega} \Omega} e^{+i\Omega k} d\Omega \quad (12)$$

it can be deduced that the spatial shift results in a sinc-interpolation, i.e. a fractional delay of the temporal PWD-response [8]. The spatial shifting itself can be realized continuously ( $\mathbf{x}_T \in \mathbb{R}$ ). In practice the PWD data are temporally and spatially discretized which has directly impact on the interpolation via the resolution of  $\mathbf{k}$  and  $\Omega$ . The fractional delay results in a perfect impulse, i.e. an integer sample delay when  $\frac{\langle \mathbf{k}, \mathbf{x}_T \rangle}{\Omega} \in \mathbb{Z}$ . With  $\mathbf{n}_k = \mathbf{k}/\|\mathbf{k}\|$  this requirement may be linked to the sampling frequency

$$\left( \frac{f_s}{c} \langle \mathbf{n}_k, \mathbf{x}_T \rangle \right) \in \mathbb{Z} \quad (13)$$

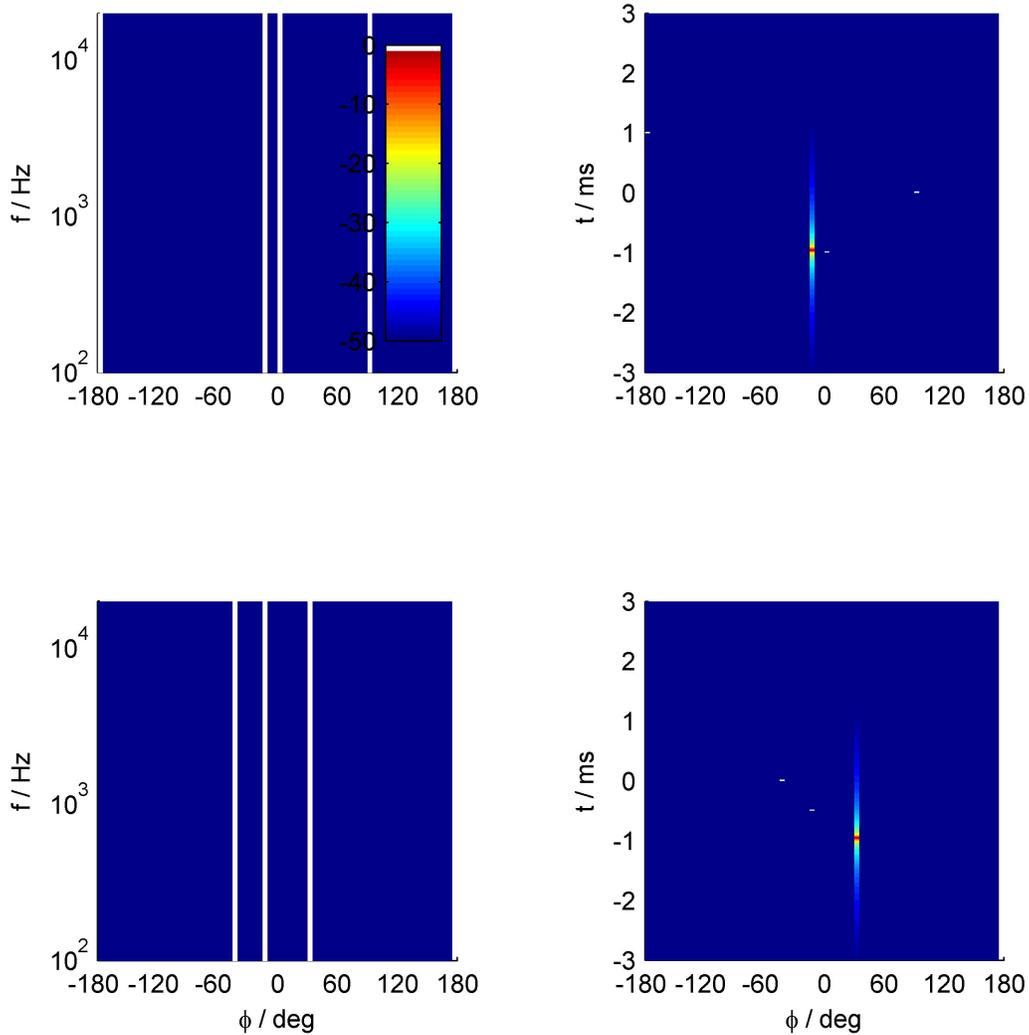
All other translational shifts  $\frac{\langle \mathbf{k}, \mathbf{x}_T \rangle}{\Omega} \notin \mathbb{Z}$  result in temporal fractional sample delays for which

$$\left( \frac{f_s}{c} \langle \mathbf{n}_k, \mathbf{x}_T \rangle + \frac{1}{2} \right) \in \mathbb{Z} \quad (14)$$

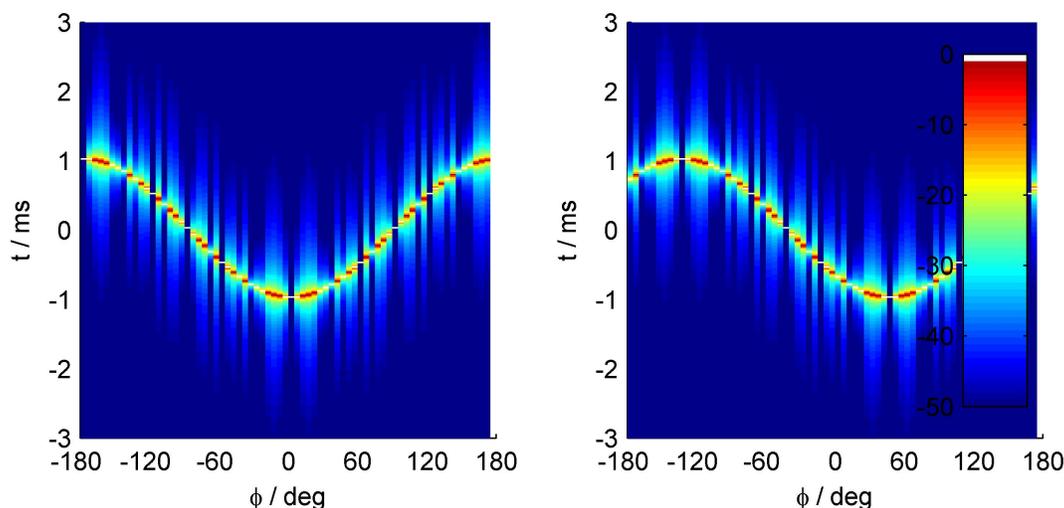
is considered to be the worst case for fractional sample delay interpolation, because the impulse energy is most widely spread over time. It is worth to note that interestingly the translational shift can be even outside the sphere from where the PWD data were acquired, since the plane wave expansion coefficients only decode the amplitude and phase, respectively the temporal characteristics of plane waves.

## 6. EVALUATION

The proposed approach was evaluated by numerical simulations. The speed of sound was set to  $c = 343$  m/s. The temporal sampling frequency was set to  $f_s = 44.1$  kHz. The investigations were limited to the horizontal plane, i.e.  $\theta = \pi/2$ ,  $\theta_0 = \pi/2$  for a meaningful graphical treatment of the results. At first the influence of translational shifts was inspected for ideal, unit amplitude and zero-phase plane waves. The temporal frequency resolution was set by an FFT-length  $N = 2^{11}$  and the angular discretization was set to  $\Delta\phi = 5^\circ$ . Four plane waves coming from  $\phi_0 = -180^\circ, -15^\circ, 0^\circ, +90^\circ$  with unit amplitude and zero-phase were setup as raw PWD data. A translational shift along  $x$ -axis was realized with  $\mathbf{x}_T \approx [0.3422, 0, 0]$  m, for which a temporal shift of exact 44 samples ( $\approx 1$  ms) was intended with the given  $c$  and  $f_s$ . The magnitude of the frequency response and the temporal response are shown in Fig. 2 on top. Note that unit amplitude and zero-phase plane waves without translational shifts result in a temporal Dirac impulse at  $t = 0$  s. The temporal responses behave differently for different impinging directions according to the translational shift. The temporal response for the wave coming from  $+90^\circ$  remains unaffected by the shift, since the wave vector  $\mathbf{k}$  is orthogonal to the shift vector  $\mathbf{x}_T$  resulting in  $\langle \mathbf{k}, \mathbf{x}_T \rangle = 0$ . The temporal response for the wave coming from  $-180^\circ$  results in a delay by exact 44 samples ( $\approx 1$  ms), since the wave fronts arrive later at the shifted position  $\mathbf{x}_T$ . Vice versa the temporal response for the wave coming from  $+0^\circ$  yields to a negative delay, since the wave fronts arrive earlier compared to the reference position  $\mathbf{x} = (0, 0, 0)$ . Note that those shifts intentionally result in an integer sample delay for the chosen temporal resolution. Thus, ideal temporal Dirac impulses are realized. The temporal response for the wave coming from  $-15^\circ$  results in a negative temporal fractional delay of almost exact 42.5 samples. Thus the sinc-interpolation for the fractional delay of half an sample can be clearly seen. Fig. 2 bottom demonstrates a different scenario with ideal plane waves. A transla-



**Fig. 2:** Ideal plane wave decompositions with **translational shifts**. Top: unit amplitude and zero-phase plane waves coming from  $\phi_0 = -180^\circ, -15^\circ, 0^\circ, +90^\circ$ ,  $\theta_0 = \pi/2$  with shift to  $\mathbf{x}_T = [0.3422, 0, 0]$  m, Bottom: unit amplitude and zero-phase plane waves coming from  $\phi_0 = -45^\circ, -15^\circ, +30^\circ$ ,  $\theta_0 = \pi/2$  with shift to  $\mathbf{x}_T = [0.2420, 0.2420, 0]$  m. Left: the frequency responses  $|\bar{P}(\phi, \theta, \omega)|$  are shown in dB with reference to unit amplitude. Right: the temporal responses  $|\bar{P}(\phi, \theta, t)|$  of the shifted PWD data are shown in dB with reference to unit amplitude. All subfigures use the same color scaling and the colorbar from top, left. Note the subtle white dots which represent the ideal temporal Dirac impulses right, top at:  $-180^\circ|1\text{ms}, 0^\circ|-1\text{ms}, +90^\circ|0\text{ms}$  and right, bottom at:  $-45^\circ|0\text{ms}, -15^\circ|0.5\text{ms}$ . PWD resolution:  $5^\circ$ ,  $f_s = 44.1\text{kHz}$ ,  $\Delta f = 2^{11}/f_s$ .



**Fig. 3:** Temporal responses  $|\bar{p}(\phi, \theta, t)|$  of ideal plane wave decompositions with **translational shifts**. Left: unit amplitude and zero-phase plane waves coming from all angles  $\phi_0$ ,  $\theta_0 = \pi/2$  with shift to  $\mathbf{x}_T = [0.3422, 0, 0]$  m, Right: unit amplitude and zero-phase plane waves coming from all angles  $\phi_0$ ,  $\theta_0 = \pi/2$  with shift to  $\mathbf{x}_T = [0.2420, 0.2420, 0]$  m. The temporal responses  $|\bar{p}(\phi, \theta, t)|$  of the shifted PWD data are shown in dB with reference to unit amplitude. Subfigures use the same color scaling and the colorbar from right. PWD resolution:  $5^\circ$ ,  $f_s = 44.1$  kHz,  $\Delta f = 2^{11}/f_s$ .

tional shift to  $\mathbf{x}_T = [0.2420, 0.2420, 0]$  m was chosen for three unit amplitude and zero-phase plane waves coming from  $\phi_0 = -45^\circ, -15^\circ, +30^\circ$ . The vector magnitude  $\|\mathbf{x}_T\| = 0.3422$  m again corresponds to a shift by 44 samples ( $\approx 1$  ms). The temporal response for the wave coming from  $\phi_0 = -45^\circ$  remains unaltered, i.e. a Dirac impulse at  $t = 0$  s again due to  $\langle \mathbf{k}, \mathbf{x}_T \rangle = 0$ . The temporal response for the wave coming from  $\phi_0 = -15^\circ$  is a perfect Dirac impulse due to the negative integer delay of 22 samples ( $\approx 0.5$  ms). The temporal response for the wave coming from  $+30^\circ$  results again in a negative temporal fractional delay of almost exact 42.5 samples, where the sinc-interpolation for the fractional delay of half an sample can be clearly seen again. Thus, it can be concluded, that the spatial shift of the PWD data leads to a sinc-interpolation of the temporal response, which heavily depends on the plane wave direction, the actual shift vector  $\mathbf{x}_T$  and the sampling frequency  $f_s$ . This is also illustrated in Fig. 3. Temporal responses are shown for the same translational shifts however this time with unit amplitude and zero-phase plane waves impinging from *all* horizon-

tal directions. The delay varies between  $\pm 1$  ms and the sinc-interpolation differs between angles as expected.

A further simulation series was performed with model-based plane wave expansion coefficients from delay-and-sum and modal beamformers for a single plane wave coming from  $\phi_0 = 0^\circ$  similar to [5]. The sphere radius was set to  $r = 0.5$  m. The temporal frequency resolution was set by an FFT-length  $N = 2^{12}$  and the angular discretization was set to  $\Delta\phi = 1^\circ$ . Fig. 4 and 6 show PWD data from a spatially continuous sphere using delay-and-sum respectively modal beamforming with  $N_{\text{SHT}} = 23$ . Fig. 5 and 7 on the other hand illustrate PWD data from a spatially discrete sphere sampled with 770 microphones on a Lebedev grid for the two beamformer versions. Each top, left subfigure represents the magnitude response of the beamformer. Each top, right subfigure shows the temporal response without any translational shift, i.e.  $\mathbf{x}_T = [0, 0, 0]$  m. Each bottom, left subfigure shows the temporal response for the shift to  $\mathbf{x}_T = [0.3422, 0, 0]$  m, and each bottom, right subfigure shows the shift to  $\mathbf{x}_T = [0.2420, 0.2420, 0]$  m. Note that in

practice the translatory head-movements are expected to be smaller for resting listener bodies. The shift amounts here were chosen for straightforward and convenient exemplification and to make them comparable to Fig. 2 and 3. The temporal responses of delay-and-sum and modal beamformers strongly differ as was already shown in [5]. All bottom, left subfigures (most apparent in Fig. 6) show the shift towards  $x$ -axis and reveal the earlier arriving wave front for  $-\pi/2 < \phi < +\pi/2$  resulting in negative delays, while for remaining angles positive delay times indicate later arriving wave fronts. This leads to an axisymmetric temporal response around  $\phi_0 = 0^\circ$ . For a translational shift to  $\mathbf{x}_T = [0.2420, 0.2420, 0]$  m artifacts from fractional delay interpolation can be observed at the main impulse. Note that fractional delay interpolation will occur at all angles (cf. Fig. 3). The graphical resolution for very low amplitudes however forbids the visibility. The delay-and-sum beamformer in general spreads more energy over time, however it may be deduced from the graphics that it is more robust against translational shifts when considering the main impulse energy, see Fig. 4, 5 bottom, left/right vs. 6, 7.

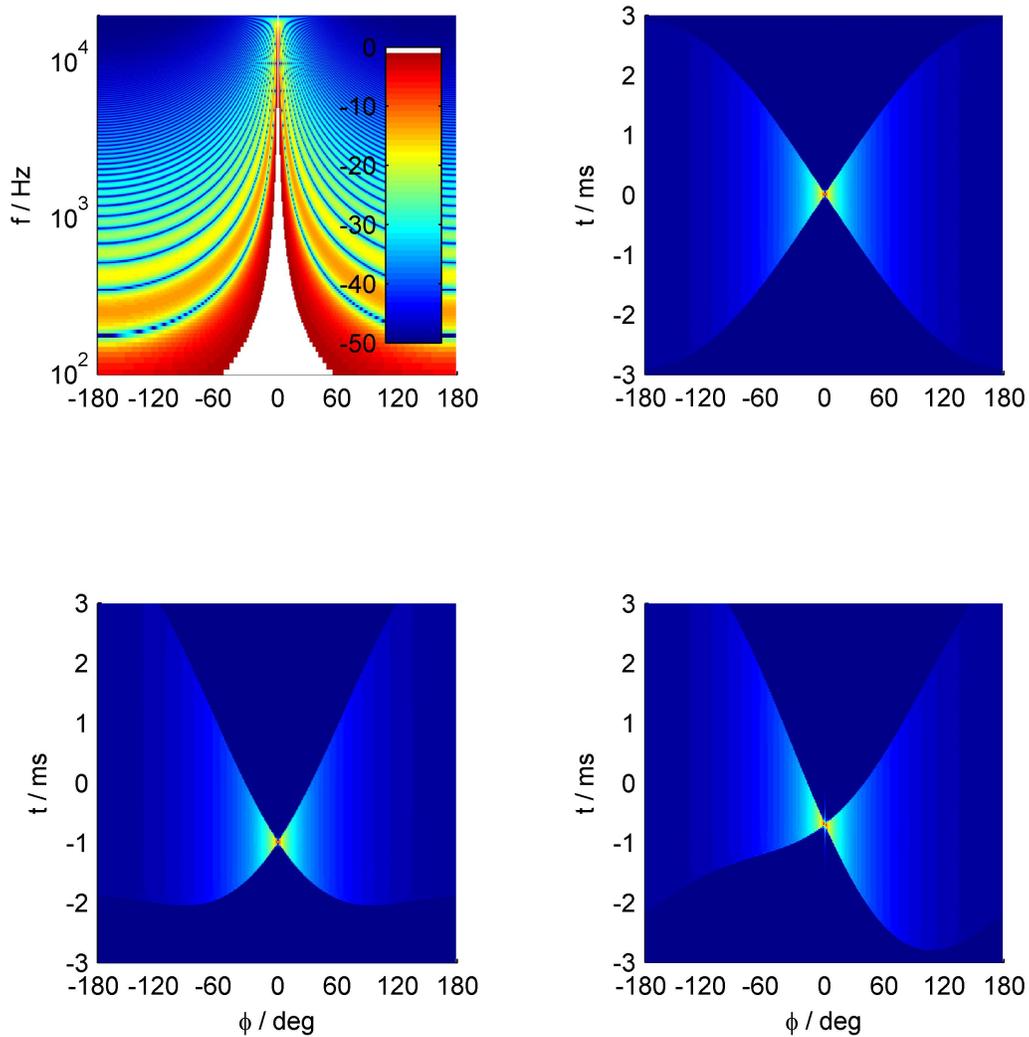
## 7. SUMMARY AND CONCLUSION

In the context of binaural synthesis it appears elegant to separate the HRTF data capture from the acquisition of the soundfield, which has to be auralized. It is then not only much easier to individualize the binaural synthesis for listeners by using individualized HRTF datasets but also provides more degrees of freedom within the acoustic signal processing chain. This study used this benefit to realize translatory head-movements for dynamic binaural synthesis by manipulating the soundfield data in the angular spectrum domain. More explicitly the shifting theorem of the Fourier transform was applied to the plane wave decomposition coefficients which employs spatially fullband translational shifts. The spatial (continuous) shift results in a temporal shift, which involves fractional sample delays for time-discrete signals due to temporal and spatial sampling of the soundfield. All known resampling and fractional sample interpolation techniques may be utilized to improve the resolution. Furthermore time-discrete signals should have sufficient length to map the desired spatial shifts into the time domain and to avoid cyclic shifts within the temporal DFT-data. For the realization of negative delays a pre-delay is required. Dynamic binaural synthesis including translational head-movements would require realtime tracking of the head-movements and an interpolating adaptation

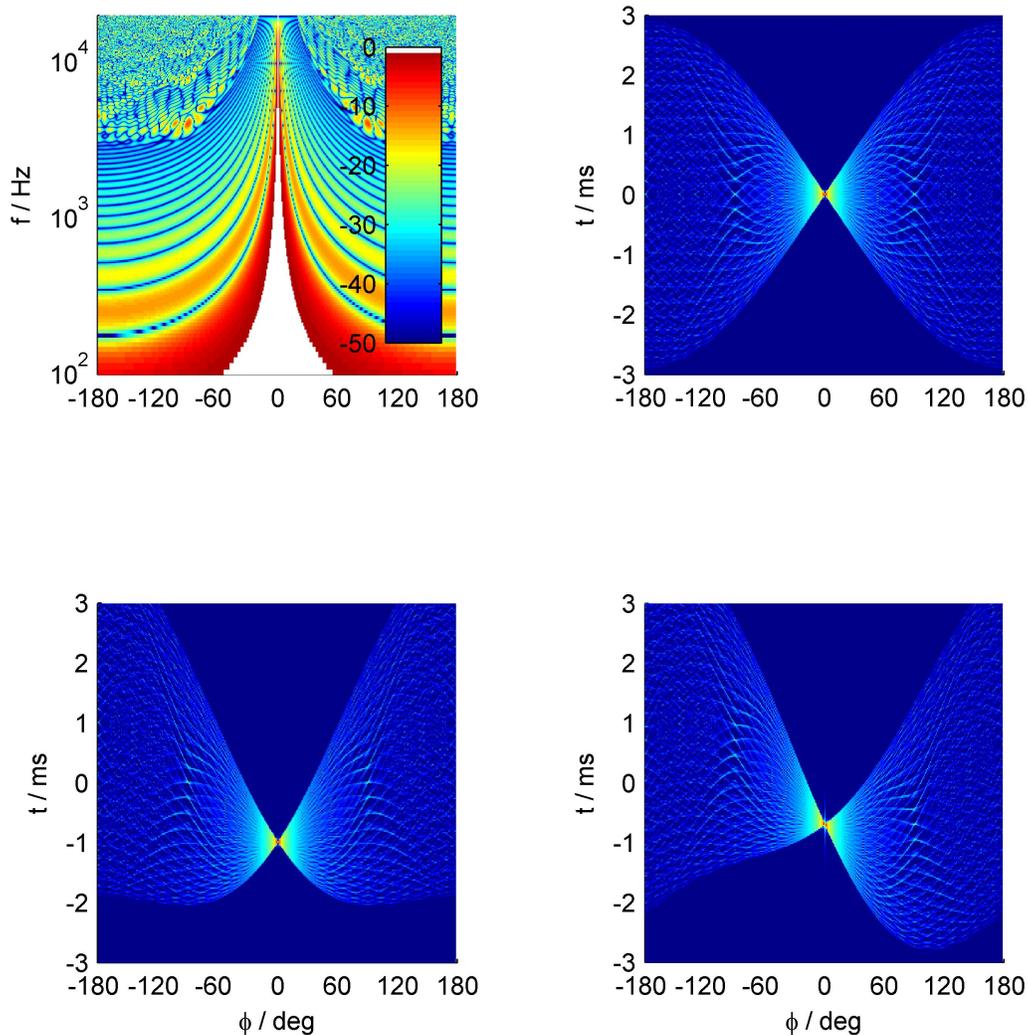
of the transition as it will be usually handled for head-rotations. The perceptual implications of translational shifting remain for future research. Binaural listening examples are available under <http://spatialaudio.net/translatory-head-movements/>.

## 8. REFERENCES

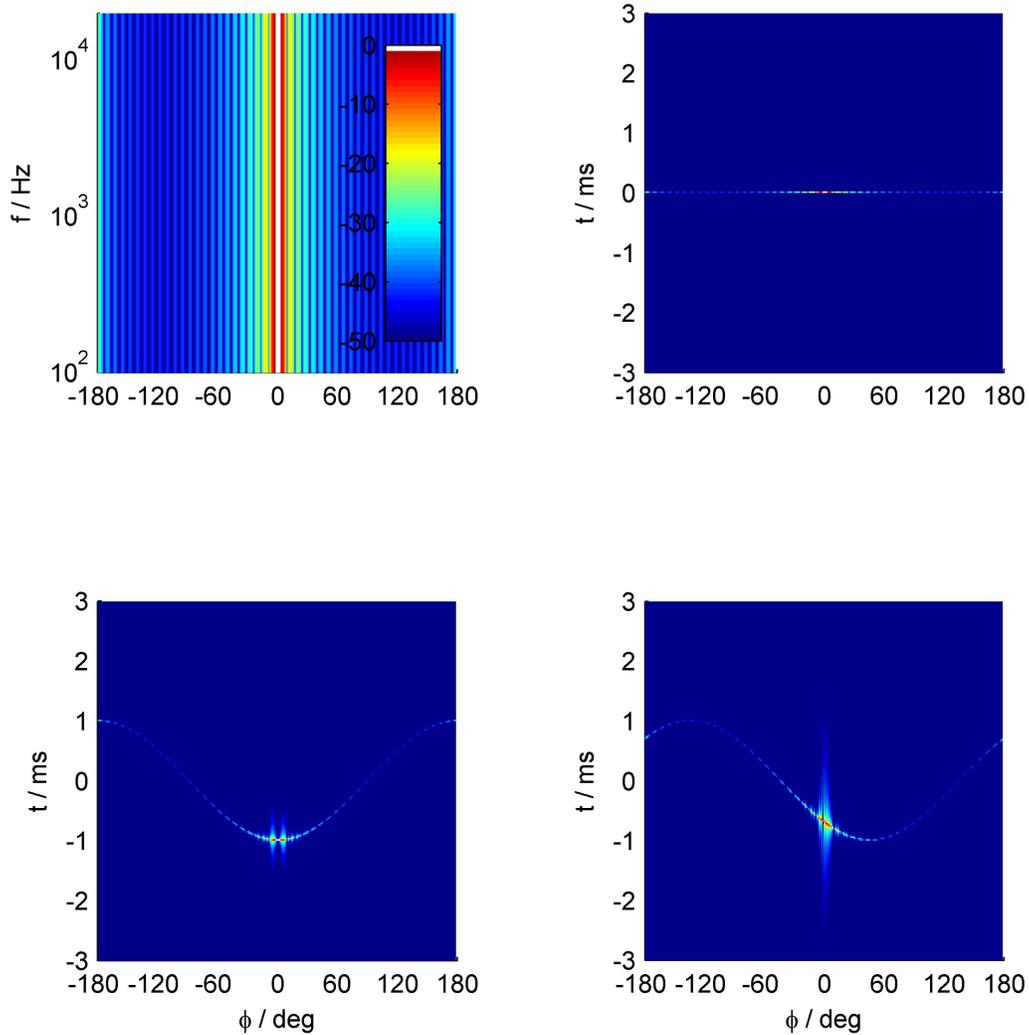
- [1] N. A. Gumerov and R. Duraiswami, *Fast multipole methods for the Helmholtz equation in three dimensions*. Oxford: Elsevier Science, 2004.
- [2] R. Duraiswami, Z. Li, D. N. Zotkin, E. Grassi, and N. A. Gumerov, "Plane-wave decomposition analysis for spherical microphone arrays," in *Proc. of IEEE Workshop on Applications of Signal Processing to Audio and Acoustics, Oct. 16-19, 2005, New Paltz, NY, USA, 2005*.
- [3] M. Park and B. Rafaely, "Sound-field analysis by plane-wave decomposition using spherical microphone array," *J ACOUST SOC AM*, vol. 118, no. 5, pp. 3094–3103, 2005.
- [4] B. Rafaely, "Phase-mode versus delay-and-sum spherical microphone array processing," *IEEE Signal Process. Letters*, vol. 12, no. 10, pp. 713–716, 2005.
- [5] S. Spors, H. Wierstorf, and M. Geier, "Comparison of modal versus delay-and-sum beamforming in the context of data-based binaural synthesis," in *Proc. of the 132nd AES Conv., Budapest*, no. #8669, 2012.
- [6] A. Kuntz, *Wave Field Analysis Using Virtual Circular Microphone Arrays*. Munich: Verlag Dr. Hut, 2009.
- [7] B. Girod, R. Rabenstein, and A. Stenger, *Signals and Systems*. Chichester: Wiley, 2001.
- [8] T. Laakso, V. Valimaki, M. Karjalainen, and U. Laine, "Splitting the unit delay," *IEEE Signal Processing Magazine*, vol. 13, no. 1, pp. 30–60, 1996.
- [9] B. Bernschütz, C. Pörschmann, S. Spors, and S. Weinzierl, "Sofia sound field analysis toolbox," in *Proc. of the ICSA International Conference on Spatial Audio, Detmold, Germany, November 2011, 2011*.



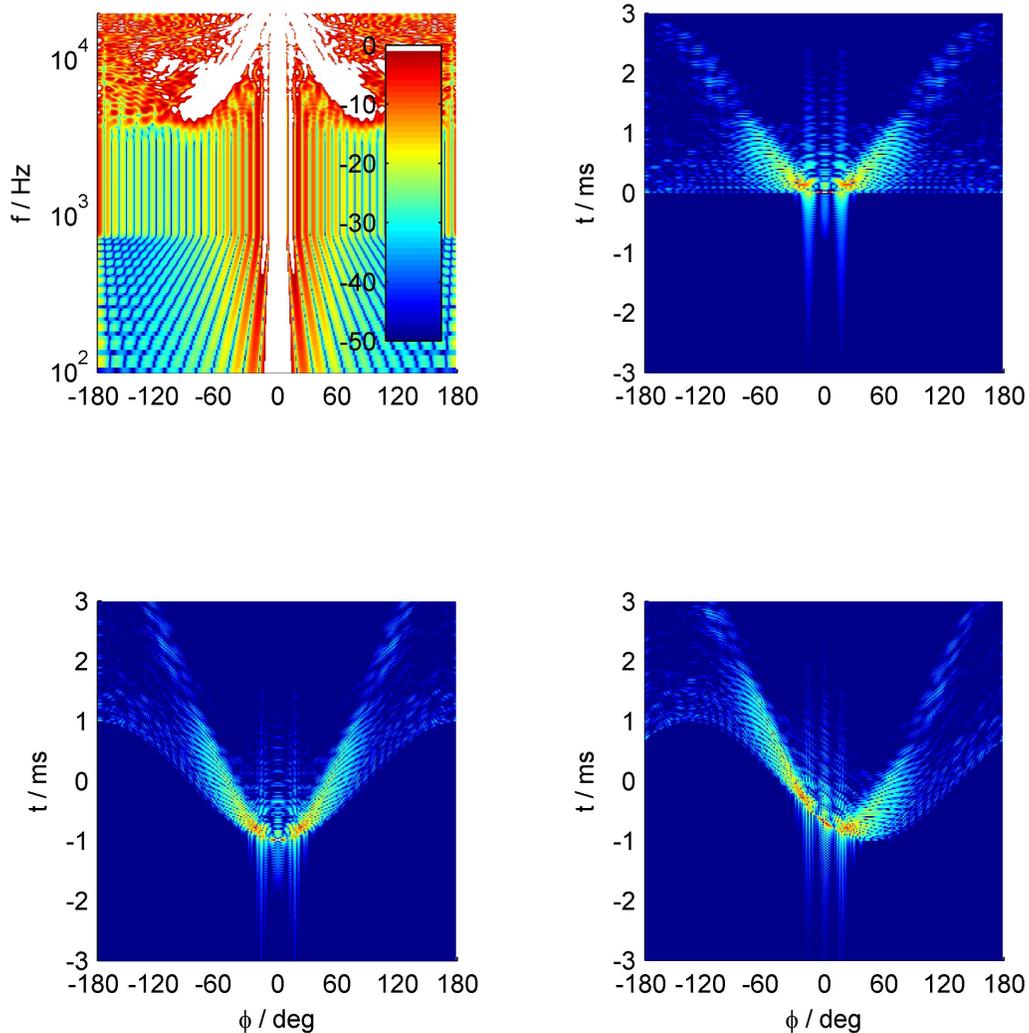
**Fig. 4:** Plane wave decomposition using **delay-and-sum beamforming** of a broadband plane wave coming from  $\phi_0 = 0^\circ$ ,  $\theta_0 = \pi/2$  captured **spatially continuous** on a spherical aperture with radius  $r = 0.5$  m. Left, top: the frequency response  $|\bar{P}(\phi, \theta, \omega)|$  is shown in dB with reference to  $f = 500$  Hz;  $\phi = 0^\circ$ . The other subfigures show the temporal responses  $|p(\phi, \theta, t)|$  in dB with reference to the peak for different **translational shifts**. Right, top:  $\mathbf{x}_T = [0, 0, 0]$  m, i.e. no shift. Left, bottom:  $\mathbf{x}_T = [0.3422, 0, 0]$  m, i.e. shift only on  $x$ -axis. Right, bottom:  $\mathbf{x}_T = [0.2420, 0.2420, 0]$  m, i.e. equally shifted on  $x$  and  $y$ . All subfigures use the same color scaling and the colorbar from top, left. Top corresponds to [5, Fig. 3 b), d)]. PWD resolution:  $1^\circ$ ,  $f_s = 44.1$  kHz,  $\Delta f = 2^{12}/f_s$ .



**Fig. 5:** Plane wave decomposition using **delay-and-sum beamforming** of a broadband plane wave coming from  $\phi_0 = 0^\circ$ ,  $\theta_0 = \pi/2$  captured by a **spatially discrete** sphere sampled with 770 microphones on a Lebedev grid with radius  $r = 0.5$  m. Left, top: the frequency response  $|\bar{P}(\phi, \theta, \omega)|$  is shown in dB with reference to  $f = 500$  Hz;  $\phi = 0^\circ$ . The other subfigures show the temporal responses  $|p(\phi, \theta, t)|$  in dB with reference to the peak for different **translational shifts**. Right, top:  $\mathbf{x}_T = [0, 0, 0]$  m, i.e. no shift. Left, bottom:  $\mathbf{x}_T = [0.3422, 0, 0]$  m, i.e. shift only on  $x$ -axis. Right, bottom:  $\mathbf{x}_T = [0.2420, 0.2420, 0]$  m, i.e. equally shifted on  $x$  and  $y$ . All subfigures use the same color scaling and the colorbar from top, left. Top corresponds to [5, Fig. 4 b), d)]. PWD resolution:  $1^\circ$ ,  $f_s = 44.1$  kHz,  $\Delta f = 2^{12}/f_s$ .



**Fig. 6:** Plane wave decomposition using **modal beamforming** of a broadband plane wave coming from  $\phi_0 = 0^\circ$ ,  $\theta_0 = \pi/2$  captured **spatially continuous** on a spherical aperture with radius  $r = 0.5$  m. Left, top: the frequency response  $|\bar{P}(\phi, \theta, \omega)|$  is shown in dB with reference to  $f = 500$  Hz;  $\phi = 0^\circ$ . The other subfigures show the temporal responses  $|p(\phi, \theta, t)|$  in dB with reference to the peak for different **translational shifts**. Right, top:  $\mathbf{x}_T = [0, 0, 0]$  m, i.e. no shift. Left, bottom:  $\mathbf{x}_T = [0.3422, 0, 0]$  m, i.e. shift only on  $x$ -axis. Right, bottom:  $\mathbf{x}_T = [0.2420, 0.2420, 0]$  m, i.e. equally shifted on  $x$  and  $y$ . All subfigures use the same color scaling and the colorbar from top, left. Top corresponds to [5, Fig. 3 a), c)]. PWD resolution:  $1^\circ$ ,  $f_s = 44.1$  kHz,  $\Delta f = 2^{12}/f_s$ ,  $N_{\text{SHT}} = 23$ .



**Fig. 7:** Plane wave decomposition using **modal beamforming** of a broadband plane wave coming from  $\phi_0 = 0^\circ$ ,  $\theta_0 = \pi/2$  captured by a **spatially discrete** sphere sampled with 770 microphones on a Lebedev grid with radius  $r = 0.5\text{m}$ . Left, top: the frequency response  $|\bar{P}(\phi, \theta, \omega)|$  is shown in dB with reference to  $f = 500\text{Hz}$ ;  $\phi = 0^\circ$ . The other subfigures show the temporal responses  $|p(\phi, \theta, t)|$  in dB with reference to the peak for different **translational phase shifts**. Right, top:  $\mathbf{x}_T = [0, 0, 0]\text{m}$ , i.e. no shift. Left, bottom:  $\mathbf{x}_T = [0.3422, 0, 0]\text{m}$ , i.e. shift only on  $x$ -axis. Right, bottom:  $\mathbf{x}_T = [0.2420, 0.2420, 0]\text{m}$ , i.e. equally shifted on  $x$  and  $y$ . All subfigures use the same color scaling and the colorbar from top, left. Top corresponds to [5, Fig. 4 a), c)]. PWD resolution:  $1^\circ$ ,  $f_s = 44.1\text{kHz}$ ,  $\Delta f = 2^{12}/f_s$ ,  $N_{\text{SHT}} = 23$ . The SOFiA Sound Field Analysis Toolbox [9] was used to generate the raw PWD data.