Universität Rostock
Traditio et Innovatio

Frank Schultz, Sascha Spors
Institute of Communications Engineering
{frank.schultz,sascha.spors}@uni-rostock.de

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

Introduction

In the context of binaural synthesis it appears elegant to separate the free-field head-related transfer function (HRTF) from the room. 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 auralized with HRTFs. Head-rotations can be considered efficiently by rotation of the plane wave expansion coefficients or the HRTF dataset before superposition. This study introduces a novel technique to cope for translatory head-movements in dynamic binaural synthesis by manipulating the captured soundfield in the angular spectrum domain.

Data-Based Binaural Synthesis

For the proposed approach we utilize the plane-wave decomposition (PWD). The inverse PWD is defined as

$$P(\mathbf{x},\omega) \propto \int_0^{2\pi}\int_0^{\pi} \overline{P}(\phi,\theta,\omega)\, \mathrm{e}^{+\mathrm{j}\langle\mathbf{k},\mathbf{x}\rangle}\, \sin\theta \mathrm{d}\theta \mathrm{d}\phi \qquad (1)$$

which can be interpreted as a weighted superposition of plane waves coming from $\mathbf{k} = \|\omega/c\| \cdot (\cos\phi\sin\theta, \sin\phi\sin\theta, \cos\theta)$ over a notional sphere [1, 2]. A single plane wave with temporal spectrum $F(\omega)$ coming from $\mathbf{k_0} = \|\omega/c\| \cdot (\cos\phi_0\sin\theta_0, \sin\phi_0\sin\theta_0, \cos\theta_0)$ may then be defined as a spatial delta function

$$\overline{P}(\phi,\theta,\omega) \propto F(\omega)\,\delta(\cos\theta - \cos\theta_0)\,\delta(\phi-\phi_0). \qquad (2)$$

The basic concept of the presented approach is to place a virtual head into the center and to replace the unit amplitude plane waves $\mathrm{e}^{+\mathrm{j}\langle\mathbf{k},\mathbf{x}\rangle}$ in eq. (1) by far-field HRTFs. Thus the sound pressure at the left/right ear $P_{\mathrm{L,R}}(\gamma,\psi,\omega)$ for a certain head orientation γ,ψ is given by [3]

$$P_{\mathrm{L,R}}(\gamma,\psi,\omega) \propto \int_0^{2\pi}\int_0^{\pi} \overline{P}(\phi,\theta,\omega)\cdot\overline{H}_{\mathrm{L,R}}(\phi,\theta,\gamma,\psi,\omega)\, \sin\theta \mathrm{d}\theta \mathrm{d}\phi. \qquad (3)$$

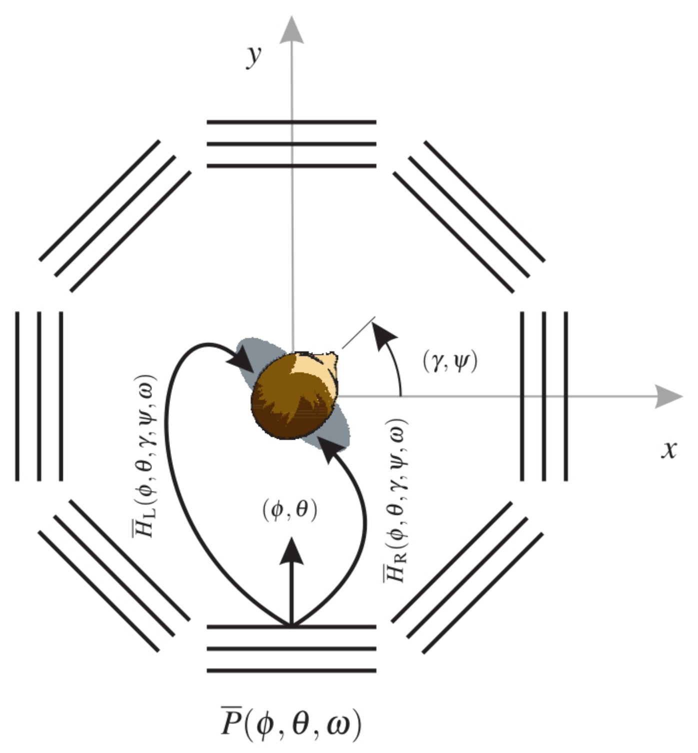

Figure : Data-based binaural synthesis using a plane wave expansion of the virtual sound field. For illustration, the filtering of the left/right HRTFs by the plane wave expansion coefficients $\overline{P}(\phi,\theta,\omega)$ is shown only for one particular direction. The z-axis points towards the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

Translatory Head-Movements

Since the plane wave decomposition coefficients $\overline{P}(\phi,\theta,\omega)$ can be directly linked to the angular wave spectrum of the soundfield the shift theorem of the Fourier transform holds. Hence, for a shift to $\mathbf{x_T}$ the equation

$$\overline{P}(\phi,\theta,\omega)\,\mathrm{e}^{+\mathrm{j}\langle\mathbf{k},\mathbf{x_T}\rangle} \bullet\!\!-\!\!\circ\; \overline{p}(\phi,\theta,t+\frac{\langle\mathbf{k},\mathbf{x_T}\rangle}{\omega}) \qquad (4)$$

is valid and it becomes evident that the spatial shift for $\overline{P}(\phi,\theta,\omega)$ results in a temporal shift in the temporal response $\overline{p}(\phi,\theta,t)$. The soundfield superposition then reads [cf. eq. (3)]

$$P_{\mathrm{L,R},\mathbf{T}}(\gamma,\psi,\omega,\mathbf{x_T}) \propto \int_0^{2\pi}\int_0^{\pi} \overbrace{\overline{P}(\phi,\theta,\omega)\,\mathrm{e}^{+\mathrm{j}\langle\mathbf{k},\mathbf{x_T}\rangle}}^{\text{Head Translation}} \cdot \overline{H}_{\mathrm{L,R}}(\phi,\theta,\gamma,\psi,\omega)\, \sin\theta \mathrm{d}\theta \mathrm{d}\phi. \qquad (5)$$

for a translation of the head to $\mathbf{x_T}$. When introducing the translational shift within the inverse discrete-time Fourier transform (DTFT)

$$\overline{p}_{\mathbf{T}}(\phi,\theta,k,\mathbf{x_T}) \propto \int_{-\pi}^{+\pi} \overbrace{\overline{P}(\phi,\theta,\Omega)\,\mathrm{e}^{+\mathrm{j}\frac{\langle\mathbf{k},\mathbf{x_T}\rangle}{\Omega}\Omega}}^{\text{Head Translation}}\,\mathrm{e}^{+\mathrm{j}\,\Omega\,k}\,\mathrm{d}\Omega \qquad (6)$$

for time-discrete PWD data with sample index $k \in \mathbb{Z}$, sampling frequency f_s and $\Omega = 2\pi\frac{f}{f_s}$ it can be deduced that the spatial shift yields to a sinc-interpolation, i.e. a fractional delay of the temporal PWD-response.

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 $\left(\frac{f_s}{c}\langle\mathbf{n_k},\mathbf{x_T}\rangle\right) \in \mathbb{Z}$. All other translational shifts $\frac{\langle\mathbf{k},\mathbf{x_T}\rangle}{\Omega} \notin \mathbb{Z}$ result in temporal fractional sample delays for which $\left(\frac{f_s}{c}\langle\mathbf{n_k},\mathbf{x_T}\rangle + \frac{1}{2}\right) \in \mathbb{Z}$ is considered to be the worst case for fractional sample delay interpolation, because the impulse energy is most widely spread over time. In practice the PWD data are temporally and spatially discretized which has directly impact on the interpolation via the resolution of $\mathbf{k}$ and Ω.

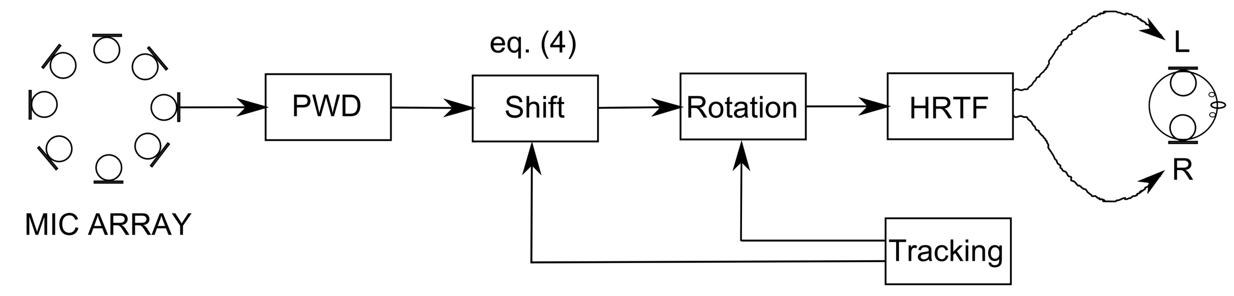

Figure : Signal-flow graph

Evaluation

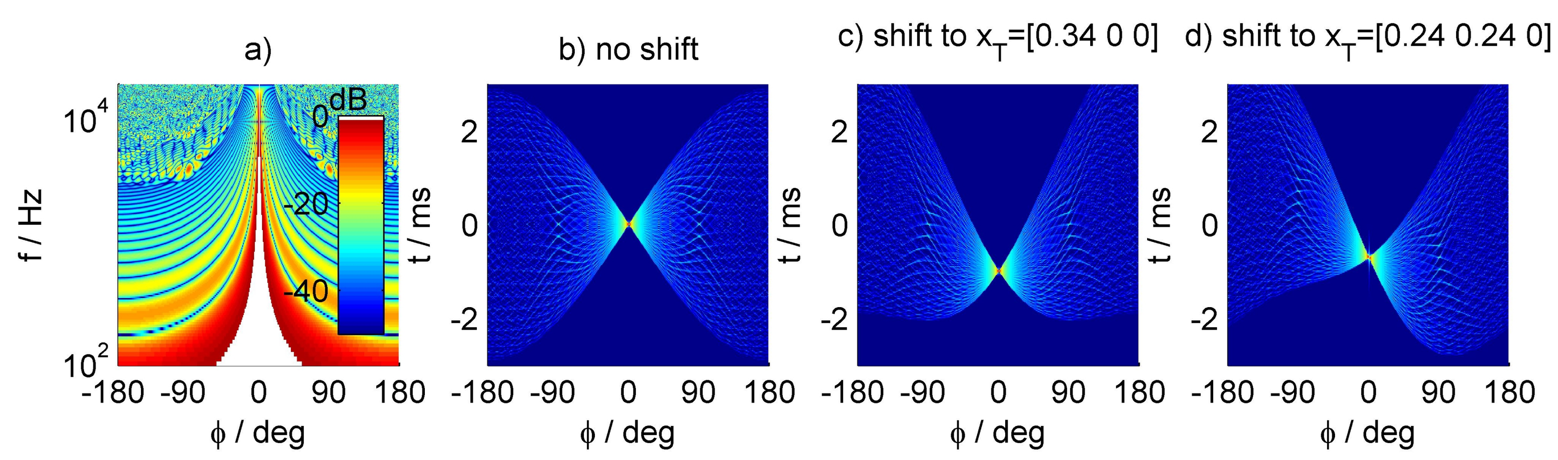

Figure : Plane wave decomposition using **delay-and-sum beamforming** of a broadband plane wave coming from $\phi_0 = 0°$, $\theta_0 = \pi/2$ captured by a **spatially discrete** sphere sampled with 770 microphones on a Lebedev grid with radius $r = 0.5$ m. a): the frequency response $|\overline{P}(\phi,\theta,\omega)|$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

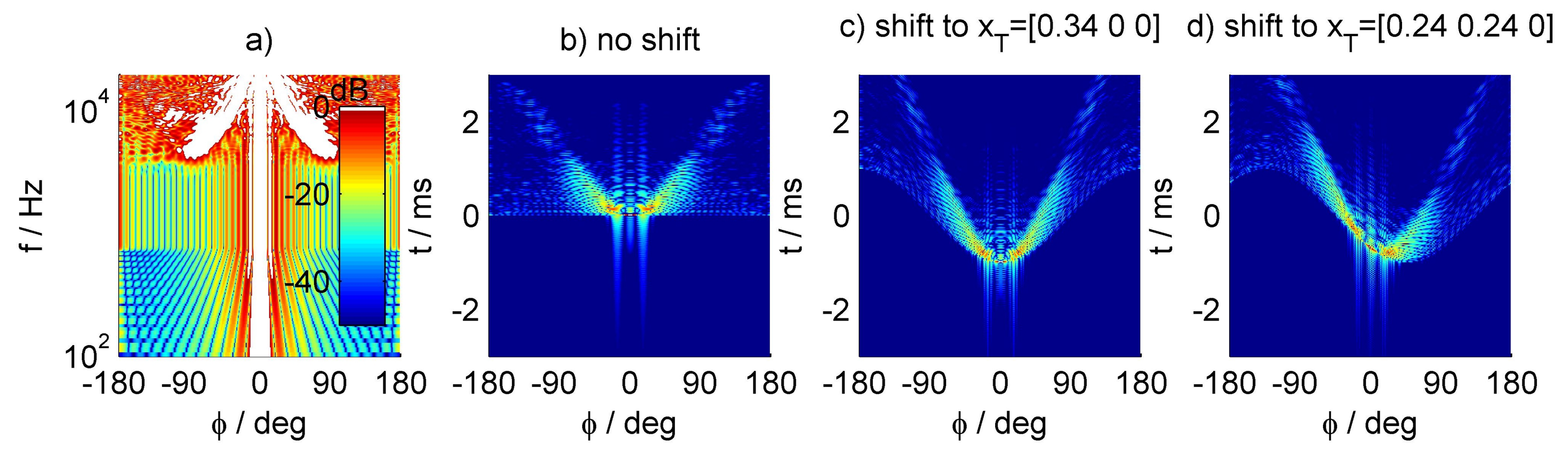

Figure : Plane wave decomposition using **modal beamforming** of a broadband plane wave coming from $\phi_0 = 0°$, $\theta_0 = \pi/2$ captured by a **spatially discrete** sphere sampled with 770 microphones on a Lebedev grid with radius $r = 0.5$ m. a): the frequency response $|\overline{P}(\phi,\theta,\omega)|$ is shown in dB with reference to $f = 500$ Hz; $\phi = 0°$. The other subfigures show the temporal responses $|\overline{p}(\phi,\theta,t)|$ in dB with reference to the peak for different **translational shifts**. b): $\mathbf{x_T} = [0,0,0]$ m, i.e. no shift. c): $\mathbf{x_T} = [0.3422,0,0]$ m, i.e. shift only on x-axis. d): $\mathbf{x_T} = [0.2420, 0.2420, 0]$ m, i.e. equally shifted on x and y. Raw data generated with [4].

Conclusion

For data-based binaural synthesis translatory head-movements of the listener can be considered explicitly with no additional measurement effort, when separating the HRTF data capture from the acquisition of the soundfield. The shift theorem of the Fourier transform is applied to the plane wave decomposition coefficients. This can be interpreted as spatially fullband translational shifts in the spatial frequency domain. For temporal and spatial discrete signals this results in fractional delay interpolation. The perceptual implications of translational shifting remain for future research.

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

[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] 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.

[4] 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.

AES 52ND CONFERENCE ON SOUND FIELD CONTROL, SEPTEMBER 2-4, 2013, GUILDFORD, UK