Modal Bandwidth Reduction in Data-based Binaural Synthesis including Translatory Head-movements

Nara Hahn and Sascha Spors
Institute of Communications Engineering, University of Rostock, Germany
Email: {nara.hahn,sascha.spors}@uni-rostock.de

Introduction

Binaural synthesis is a sound reproduction approach aiming at the generation of ear signals for a virtual auditory scene. The ear signals are typically presented to listeners via headphones. To achieve a good perceptual performance, dynamic binaural synthesis is desirable. Here the head-movements of a listener are tracked and the ear signals are adjusted accordingly. In a recent study [1], an approach for dynamic binaural synthesis including not only rotational but also translational head-movements was proposed. The required pre-processing is to decompose a captured sound field into plane wave components, as shown in Fig. 1a. To consider a translational head-movement, each plane wave needs to be extrapolated to the translated head position by applying a phase shift. A subsequent study [2] showed that a displacement from the original head position results in decreased localization accuracy. The physical properties relevant to such degradations are not fully identified, though. In this paper, we investigate the modal spectrum of spatially translated sound fields, and show that the effective modal bandwidth is reduced.

Data-based Binaural Synthesis

We consider two-dimensional sound fields within a bounded source-free region which can be represented as circular harmonics expansion [3, Sec. 2.5.3.1],

$$S(\mathbf{x},\omega) = \sum_{m=-\infty}^{\infty} \breve{S}_m(\omega) J_m(\tfrac{\omega}{c} r) e^{im\alpha}, \tag{1}$$

where $\mathbf{x} = (r,\alpha)^T$ denotes the position in the horizontal plane, $\breve{S}_m(\omega)$ the modal expansion coefficient, $J_m(\cdot)$ the m-th order Bessel function of the first kind, and $e^{im\alpha}$ the m-th circular harmonics. The angular frequency is denoted by ω, the speed of sound by c, and the imaginary unit by i. Translatory head-movements in the horizontal plane are considered, i.e., $\mathbf{x}_\mathrm{t} = (r_\mathrm{t},\phi_\mathrm{t})^T$. It is assumed that the original sound field is captured ideally with a continuous circular microphone array, so that no spatial aliasing artifacts occur. The only practical constraint we consider is the spatial band-limitation.

The sound field can also be represented as a superposition of plane waves [3, Sec. 2.5.3.2],

$$S(\mathbf{x},\omega) = \frac{1}{2\pi} \int_0^{2\pi} \bar{S}(\phi,\omega) e^{-i\frac{\omega}{c} r\cos(\phi-\alpha)} \mathrm{d}\phi, \tag{2}$$

where $\bar{S}(\phi,\omega)$ is spectrum of the plane wave propagating in the direction of $(\cos\phi,\sin\phi)^T$. The wave vector of the respective plane wave is $\mathbf{k} = \frac{\omega}{c}(\cos\phi,\sin\phi)^T$. To compute the plane wave decomposition coefficients, we employ modal beam-forming, which is performed in the circular harmonics domain,

$$\bar{S}(\phi,\omega) = \sum_{m=-\infty}^{\infty} i^m \breve{S}_m(\omega) e^{im\phi} \tag{3}$$

$$\breve{S}_m(\omega) = \frac{i^{-m}}{2\pi} \int_0^{2\pi} \bar{S}(\phi,\omega) e^{-im\phi} \mathrm{d}\phi, \tag{4}$$

as shown in Fig. 1a.

In binaural synthesis, the plane wave spectra can be directly used for auralization. The ear signals for a given head orientation $(\cos\gamma,\sin\gamma)^T$ are computed by filtering the individual plane waves with the corresponding far-field head-related transfer functions (HRTFs),

$$P_{\mathrm{L,R}}(\gamma,\omega) = \frac{1}{2\pi} \int_0^{2\pi} \bar{S}(\phi,\omega) \bar{H}_{\mathrm{L,R}}(\phi,\gamma,\omega) \mathrm{d}\phi. \tag{5}$$

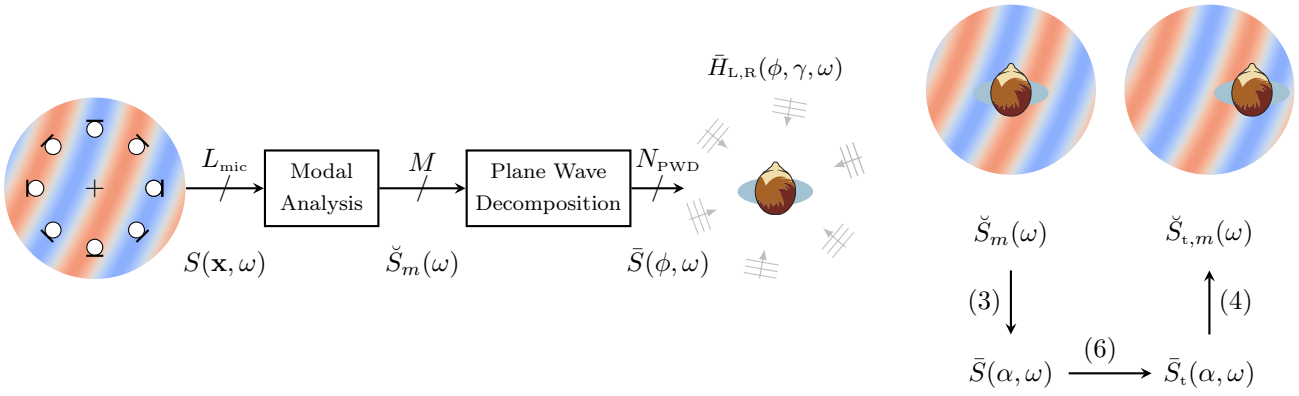
A far-field HRTF $\bar{H}_{\mathrm{L,R}}(\phi,\gamma,\omega)$ represents the linear distortion of an incident plane wave measured at a defined position in the ear canal. These can be obtained, for instance, by extrapolating a HRTF data set measured at finite distance [4]. For practical purposes, a source distance larger than 1 m suffices.

In this context, an approach for dynamic binaural synthesis including translatory head-movements was proposed in [1]. The underlying assumption is that a translational head-movement is equivalent to a spatial translation of the sound field. Instead of switching between HRTFs, the plane wave expansion coefficients are modified. For a head-movement $\mathbf{x}_\mathrm{t} = (r_\mathrm{t},\phi_\mathrm{t})^T$, a phase shift is applied to each plane wave spectrum,

$$\bar{S}(\phi,\omega) \Rightarrow \bar{S}(\phi,\omega) \times e^{-i\frac{\omega}{c} r_\mathrm{t}\cos(\phi-\phi_\mathrm{t})}, \tag{6}$$

which extrapolates the plane wave to the translated head position.

In practice, the spatial resolution of a captured sound field is limited, most prominently due to the finite number of microphones and the size of the microphone array. In [5, 6], the perceptual properties for varying spatial resolution were investigated, without considering head-movements. It was shown that a decrease in modal bandwidth of the captured sound field deteriorates the localization and also causes timbral coloration. Recently, in



(a) Data-based binaural synthesis (b) Translatory head-movements

Figure 1: Data-based binaural synthesis including translatory head-movements. (a) A sound field captured by a microphone array is decomposed into circular harmonics components, from which the plane wave spectrum is computed. (b) A translational head-movement is considered by applying phase shifts to the individual plane wave spectra. The numbers indicate the corresponding equations in the text.

[2], the influence of translatory head-movements was investigated. It was found that an additional degradation in localization accuracy occurs at translated head positions. While the amount of degradation strongly depends on the translation distance, perpendicular movements with respect to the sound propagation are more critical than parallel movements.

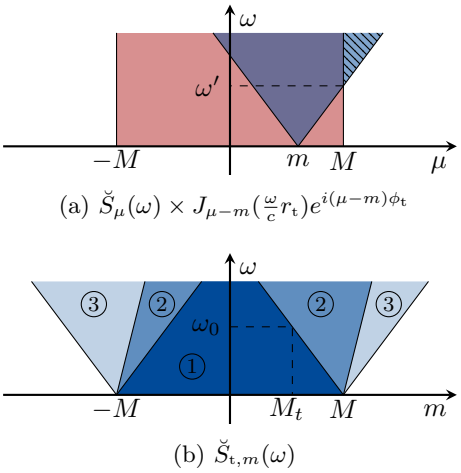
Spatial translation

In this section, the modal expansion coefficients at the translated position $\breve{S}_{\text{t},m}(\omega)$ are derived. The procedure is illustrated in Fig. 1b. The phase-shift in (6) is applied to the plane wave spectrum in (3), and the result is transformed back to the modal spectrum by using (4),

$$\begin{aligned}
&\breve{S}_{\text{t},m}(\omega) \\
&= \frac{i^{-m}}{2\pi} \int_{-\infty}^{\infty} \underbrace{\sum_{\mu=-\infty}^{\infty} i^{\mu} \breve{S}_{\mu}(\omega) e^{i\mu\phi} e^{-i\frac{\omega}{c} r_{\text{t}} \cos(\phi-\phi_{\text{t}})}}_{\bar{S}_{\text{t}}(\phi,\omega)} e^{-im\phi} \mathrm{d}\phi \\
&= \sum_{\mu=-\infty}^{\infty} \breve{S}_{\mu}(\omega) \underbrace{\frac{i^{\mu-m}}{2\pi} \int_{-\infty}^{\infty} e^{-i\frac{\omega}{c} r_{\text{t}} \cos(\phi-\phi_{\text{t}})} e^{i(\mu-m)\phi} \mathrm{d}\phi}_{J_{\mu-m}(\frac{\omega}{c} r_{\text{t}}) e^{i(\mu-m)\phi_{\text{t}}}} \\
&= \sum_{\mu=-\infty}^{\infty} \breve{S}_{\mu}(\omega) J_{\mu-m}(\tfrac{\omega}{c} r_{\text{t}}) e^{i(\mu-m)\phi_{\text{t}}}. \qquad (7)
\end{aligned}$$

This states that the translated modal expansion coefficient is given as an infinite sum of the coefficients $\breve{S}_{\mu}(\omega)$ weighted by $J_{\mu-m}(\frac{\omega}{c} r_{\text{t}}) e^{i(\mu-m)\phi_{\text{t}}}$. This can be interpreted as the cross-correlation of the two terms in the modal domain. The same result can be obtained by using the addition theorem of the Bessel functions [7].

Now we consider modal band-limitation which is applied to the captured sound field, i.e., $\breve{S}_{\mu}(\omega) = 0$ for $|\mu| > M$.



(a) $\breve{S}_{\mu}(\omega) \times J_{\mu-m}(\frac{\omega}{c} r_{\text{t}}) e^{i(\mu-m)\phi_{\text{t}}}$

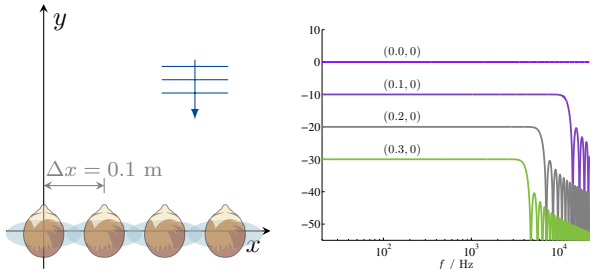
(b) $\breve{S}_{\text{t},m}(\omega)$

Figure 2: Spatial translation of a band-limited sound field. (a) The computation of the modal expansion coefficient using (8), for $0 < m < M$. (b) The modal spectrum at a translated position.

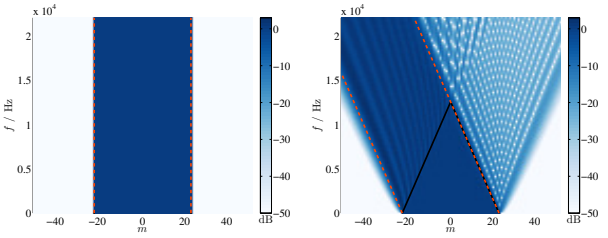
The infinite sum in (7) reduces to $2M+1$ terms,

$$\breve{S}_{\text{t},m}(\omega) = \sum_{\mu=-M}^{M} \breve{S}_{\mu}(\omega) J_{\mu-m}(\tfrac{\omega}{c} r_{\text{t}}) e^{i(\mu-m)\phi_{\text{t}}}. \qquad (8)$$

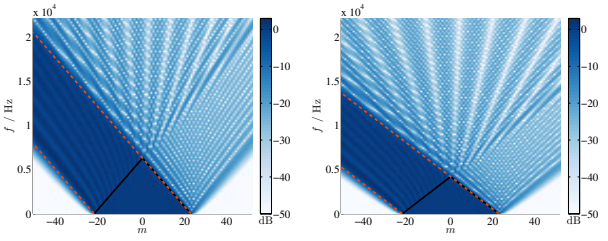
The computation of (8) is illustrated in Fig. 2a. The two spectra are overlapped with an offset of m, which are then multiplied and summed over μ. We consider the sound field of a plane wave propagating in the direction of $(\cos\phi_{\text{PW}}, \sin\phi_{\text{PW}})^T$, and thus, $\breve{S}_{\mu}(\omega) = i^{-\mu} e^{-i\mu\phi_{\text{PW}}}$. As $\breve{S}_{\mu}(\omega)$ is band-limited, it has a rectangular shape in the modal domain. The term $J_{\mu-m}\frac{\omega}{c} r_{\text{t}} e^{i(\mu-m)\phi_{\text{t}}}$, on the other hand, is not strictly band-limited, but most of the energy is contained in the region $\frac{\omega}{c} r_{\text{t}} > |m|$ due to the properties of the Bessel functions. Thus, it has a triangular shape in the m-ω domain, where the slopes of the lateral sides are $\pm\frac{\omega}{c} r_{\text{t}}$. For a given m, the resulting



(a) Head positions (b) Transfer functions

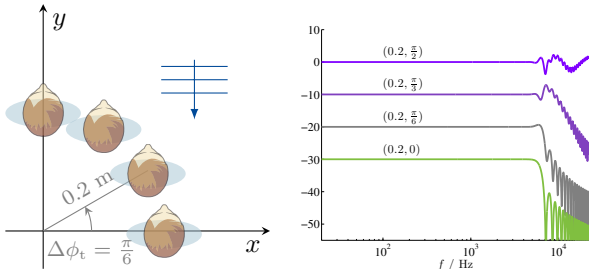


(c) $r_\mathrm{t} = 0.0$ m (d) $r_\mathrm{t} = 0.1$ m

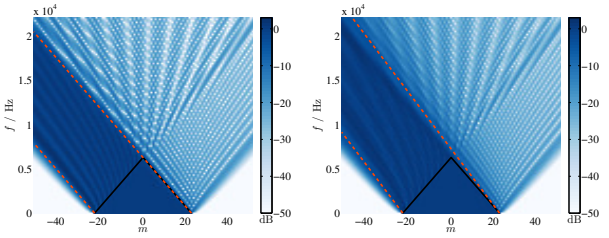


(e) $r_\mathrm{t} = 0.2$ m (f) $r_\mathrm{t} = 0.3$ m

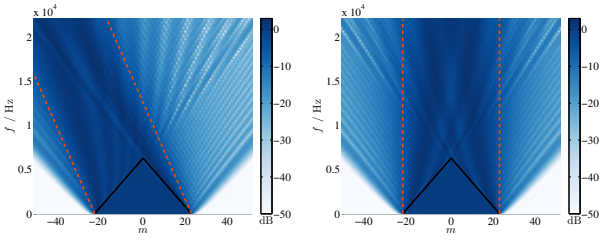
Figure 3: Spatial translation with varying distance. An incident plane wave ($\phi_\mathrm{PW} = -\frac{\pi}{2}$) is captured with a modal bandwidth of $M = 23$.



(a) Head positions (b) Transfer functions



(c) $\phi_\mathrm{t} = 0$ rad (d) $\phi_\mathrm{t} = \frac{\pi}{6}$ rad



(e) $\phi_\mathrm{t} = \frac{\pi}{3}$ rad (f) $\phi_\mathrm{t} = \frac{\pi}{2}$ rad

Figure 4: Spatial translation with varying direction. An incident plane wave ($\phi_\mathrm{PW} = -\frac{\pi}{2}$) is captured with a modal bandwidth of $M = 23$.

$\breve{S}_{\mathrm{t},m}(\omega)$ is not affected by modal limitation in the frequency band below $\omega' = \frac{c(M-m)}{r_\mathrm{t}}$. Above this frequency, there is some loss of energy in the region indicated by diagonal stripes ▨. This introduces errors in $\breve{S}_{\mathrm{t},m}(\omega)$.

The modal spectrum at the translated point is depicted in Fig. 2b. It has a trapezoidal shape and the slopes of the lateral edges are $\pm\frac{\omega}{c}r_\mathrm{t}$. Inside the spectrum, there are three regions that exhibit different properties. There is a error-free region ① which is not influenced by modal truncation. It has a triangular shape and the slopes of the lateral edges are $\pm\frac{\omega}{c}r_\mathrm{t}$. The area of this region depends on the translation distance r_t. There is a low-error region ② which is contaminated by modal truncation, but still maintains relatively high signal-to-error ratio. This region has the shape of a parallelogram and the slope of the edges is $\frac{\omega}{c}r_\mathrm{t}\sin(\phi_\mathrm{PW} - \phi_\mathrm{t})$. The shape of this region not only depends on r_t, but also on the angular difference between ϕ_PW and ϕ_t. The third region ③ contains very low energy, and suffers from strong errors.

From this observation, it can be concluded that a spatial translation results in a reduction at the modal bandwidth, in the sense that, the translated modal spectrum is accurate only up to a reduced modal order, $M_\mathrm{t} = M - \frac{\omega}{c}r_\mathrm{t}$ (See Fig. 2). The modal bandwidth linearly decreases with r_t, and the reduction rate depends on the temporal frequency. The influence of the low-error region ② on $\breve{S}_{\mathrm{t},m}(\omega)$ is not clear at this stage, and will be discussed in the following section by numerical simulations.

Evaluation

In this section, the modal spectra are numerically simulated for different head positions and their properties are discussed. We consider a broad band plane wave propagating in the negative y direction ($\phi_\mathrm{PW} = -\frac{\pi}{2}$). The modal expansion coefficient is given as $\breve{S}_m(\omega) = i^{-m}e^{-im\phi_\mathrm{PW}}$ at the origin, and $\breve{S}_{\mathrm{t},m}(\omega) = \breve{S}_m(\omega)e^{-i\frac{\omega}{c}r_\mathrm{t}\cos(\phi-\phi_\mathrm{t})}$ at the translated position. Thus, it has unit magnitude. It is assumed that the sound field is ideally captured up to a limited modal order $M = 23$.

In the first scenario, four different head positions including the origin (no movement) are considered, as shown in Fig. 3a. The movements are perpendicular to the propagation of the incident plane wave, i.e., $\phi_\mathrm{PW} - \phi_\mathrm{t} = -\frac{\pi}{2}$. Transfer functions at the respective head positions are

shown in Fig. 3b. The absolute values of the modal spectra are shown in Fig. 3c–3f. The error-free region is indicated by solid lines, and the low-error region by dashed lines. At the origin, Fig. 3c, the modal-bandwidth is ideally band-limited and the frequency response is flat. For off-center head positions, Fig. 3d–3f, the error-free region gets smaller with increasing distance, as pointed out in the previous section. The transfer functions at off-center positions exhibit a low-pass characteristic. Note that the cut-off frequency corresponds to the intersection of the peak of the error-free region appearing at $m = 0$. This is because the spectral property at a point is determined by the zeroth-order modal spectrum. Due to the low-pass characteristic, it is expected that a lateral head-movement results in timbral coloration.

In the second scenario, the distance from the origin was fixed to 0.2 m, and the polar angle was varied within the first quadrant, as shown in Fig. 4a. As r_t is unchanged, the error-free regions are not affected by the head position. The low-error region, however, depends on the polar angle of the translation ϕ_t. When ϕ_t gets closer to $\frac{\pi}{2}$, the low-error region affects the transfer functions, as can be seen in Fig. 4b. The frequency responses are flat up to a same frequency (≈ 6 kHz), but the behavior at higher frequencies varies with ϕ_t. For head-movements parallel to the propagation direction, more high-frequency energy is supplied by the low-error components. Therefore, different timbral coloration is likely to be perceived for head-movements in different directions.

For informal listening, ear signals were generated for the same scenarios as considered above ($\phi_\mathrm{PW} = -\frac{\pi}{2}, M = 23$). The incident sound field was decomposed into 360 plane waves, which were then filtered with HRTFs measured at 3 m distance [8]. Dry speech and castanets samples were used as source signals [9]. The ear signals at the original head position and each translated head position were pairwise compared by successive listening. In the case of head-movements parallel to the sound propagation $\phi_\mathrm{t} = \frac{\pi}{2}$, the difference in localization and timbre are barely perceivable for r_t up to several meters. On the contrary, head-movements perpendicular to ϕ_PW cause timbral differences that are clearly audible already for $r_\mathrm{t} \approx 0.2$ m. As r_t further increases, the ear signal gets more muffled. Localization is also deteriorated for $r_\mathrm{t} > 0.2$ m, and more than two sources with different timbre are perceived. Occasionally, in-head localization occurs. The listening examples are available for download at `http://spatialaudio.net/modal-bandwidth-reduction`.

Conclusion

Spatial translation of a band-limited sound field was discussed in the context of dynamic binaural synthesis. It was shown that the spatial bandwidth is reduced for a translated head position. Such spatial bandwidth reduction occurs as soon as the head is displaced from the origin, and there is no extended region where constant spatial bandwidth is maintained. This is accompanied with a spectral distortion which depends on the distance and direction of the translation. To avoid any perceptual degradation, it might be desirable to capture the sound field with a spatial bandwidth higher than that would be required for a fixed head position.

To investigate only the properties of spatial translation, spatial sampling was excluded from our discussion. In practice, at least two spatial sampling stages have to be taken into account. The first one occurs when a sound field is captured with a finite number of microphones. The second spatial sampling happens when the captured sound field is decomposed into a chosen number of plane waves. The technical and perceptual properties of these are still open topics.

References

[1] F. Schultz and S. Spors, "Data-based Binaural Synthesis including Rotational and Translatory Head-movements," in *52nd Conference on Sound Field Control - Engineering and Perception, Audio Engineering Society*, Sept. 2013.

[2] F. Winter, F. Schultz, and S. Spors, "Localization Properties of Data-based Binaural Synthesis including Translatory Head-Movements," in *Forum Acusticum*, Sept. 2014.

[3] A. Kuntz, *Wave Field Analysis using Virtual Circular Microphone Arrays*. Verlag Dr. Hut, 2008.

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

[5] S. Spors and H. Wierstorf, "Evaluation of Perceptual Properties of Phase-Mode Beamforming in the Context of Data-Based Binaural Synthesis," in *International Symposium on Control, Communications, and Signal Processing*, May 2012.

[6] S. Spors, H. Wierstorf, and M. Geier, "Comparison of Modal versus Delay-and-Sum Beamforming in the Context of Data-based Binaural Synthesis," in *132nd Convention of the Audio Engineering Society*, Apr. 2012.

[7] M. Abramowitz and I. A. Stegun, *Handbook of mathematical functions: with formulas, graphs, and mathematical tables*. No. 55, Courier Corporation, 1964.

[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 Convention of the Audio Engineering Society*, May 2011.

[9] EBU, "Sound Quality Assessment Material Recordings for Subjective Tests," Sept. 2008.