# Dynamic Measurement of Binaural Room Impulse Responses Using an Optical Tracking System

Nara Hahn, Wiebke Hahne, and Sascha Spors

*Institute of Communications, University of Rostock, Germany, Email: nara.hahn@uni-rostock.de*

**Abstract**

The binaural room impulse responses for different head orientations are captured by using a dynamic measurement technique. The system is excited by a so called periodic perfect sequence which exhibits the self-orthogonal property. A head-and-torso simulator is mounted on a turntable and is continuously rotated. During the measurement, the head orientation is also recorded with an optical tracking system. The BRIRs for selected azimuth angles are then computed by using a time-varying system identification method based on spatial interpolation. The performance of the proposed approach is evaluated by comparing the BRIRs obtained from a dynamic measurement and a conventional static measurement.

## Introduction

Binaural room impulse responses (BRIRs) characterize the acoustic transmission from a sound source to the outer ears of a listener. BRIRs can be used for an authentic auralization of a loudspeaker array system that is installed in a room. For each loudspeaker, BRIRs are measured for different head orientations, so that rotational head movements can be included in binaural synthesis. Such a binaural re-synthesis has been often employed in listening experiments for the comparison of sound reproduction systems in different rooms, comparison of different spatial audio techniques, and for the perceptual evaluation in different listening positions, to name only a few [**?**, 1, 2, 3]. Another application can be found in music production, where the ear signals of a mixing studio are simulated, e.g. Nx 3D Audio plug-in by WAVES. This allows the sound engineer to mix using headphone in a virtual mixing room.

Since the temporal and spectral structure of BRIRs strongly depends on the listener's anthropometry, it is often advised to use individual BRIRs for higher timbral and spatial fidelity [4, 5, 6]. In a typical measurement technique, linearity and time-invariance of the system is assumed. It is thus important to immobilize the listener's head during the measurement. To obtain a pair (left and right) of BRIRs, an excitation signal (e.g. sine sweep, maximum length sequence) is played back and the responses at the outer ears are captured. This is repeated for every possible combinations of loudspeakers and head-orientations. Therefore, the total duration scales with the dimensionality (azimuth-only or azimuth-and-elevation) of the measurement, the spatial resolution of the head orientations, and the number of loudspeakers.

To avoid such a tedious and time-consuming procedure, a dynamic measurement technique can be considered. In recent years, various approaches have been proposed for a time-efficient measurement of a large numger of impulse responses, e.g. head-related impulse responses (HRIRs) and spatial room impulse responses [7, 8, 9, 10, 11, 12]. Typically, the loudspeaker or the receiver is continuously moved and an excitation signal is played back during the entire measurement period. The movement has to be either controlled accurately based on a pre-defined trajectory, or captured with a tracking method. The impulse responses are then estimated from the captured signal by using a time-varying system identification method.

In the present work, real room BRIRs are measured by using the approach introduced by the authors in [11]. The room is excited by a periodic perfect sequence and the signal captured at each outer ears are considered as a spatio-temporal sampling of the sound field. The original sound field is reconstructed by means of spatial interpolation. The BRIRs are then obtained by deconvolving the sound field with the excitation signal. The performance of the dynamic measurement are evaluated by comparing them with a conventional static measurement.

## System Identification

For a head orientation $\phi$, the acoustic signals (sound pressure) at the left and right ears are represented as

$$p_{\mathrm{L,R}}(\phi, n) = \sum_{k=0}^{N_h-1} \psi(n-k) h_{\mathrm{L,R}}(\phi, k), \tag{1}$$

where $\psi(n)$ denotes the excitation signal emitted by a sound source at a fixed position, $h_{\mathrm{L,R}}(\phi, k)$ the filter coefficients of the corresponding BRIR, and $p_{\mathrm{L,R}}(\phi, n)$ the resulting ear signals. It is assumed that the length of $h_{\mathrm{L,R}}(\phi, k)$ is always shorter than $N_h$. Since the derivations are identical for both ears, the subscripts 'L' and 'R' are omitted in the remainder.

For a dynamic measurement, the system is excited by a perfect sequence of period $N$ which exhibits the self-orthogonality,

$$\sum_{n=0}^{N-1} \psi(n+m)\psi(n) = \sigma_\psi^2 \cdot \delta(m \bmod N), \tag{2}$$

where $\sigma_\psi^2 = \sum_{n=0}^{N-1} |\psi(n)|^2$ denotes the energy within one period and $\delta(n)$ the unit impulse function. Without loss of generality, $\sigma_\psi^2 = 1$ is assumed. To avoid temporal aliasing, the excitation period has to longer than $N_h$. In a noiseless case, the impulse response $h(\phi, k)$ is given as the length-$N$ circular cross-correlation of $p(\phi, n)$ and $\psi(n)$,

$$\hat{h}(\phi, k) = \sum_{n=0}^{N-1} p(\phi, n+k)\psi(n), \tag{3}$$

which can be proven by substituting (1) into (2).

If the head orientation varies over time, i.e. $\phi(n)$, the captured signal $s(n)$ reads

$$s(n) = p\left(\phi(n), n\right) = p\left(\phi(n), n \bmod N\right), \tag{4}$$

where the periodicity of $p(\phi(n), n)$ is exploited in the second equality. This states that $s(n)$ constitutes a spatio-temporal sampling of $p(\phi, n)$ along the curve $(\phi(n), n)$. The $n$-th sample of the sound field is sampled at $\phi(n + \mu N)$, $\mu \in \mathbb{Z}$. If the total length of the captured signal is denoted by $L$, the effective number of spatial sampling points is $M_{\text{eff}} = \frac{L}{N}$.

Since the individual samples are captured in different angular positions, (2) cannot be directly used for the computation of the BRIR. Instead, the sound field has to be reconstructed for the desired position,

$$\hat{p}(\phi, n \bmod N) = \mathcal{L}\left(\phi | s(n+\mu N),\ \mu \in \mathbb{Z}\right) \tag{5}$$

where $\hat{(\cdot)}$ denotes the estimate of the argument and $\mathcal{L}(\cdot)$ a spatial interpolation. In order to obtain $\hat{p}(\phi, \nu)$, for instance, a decimated sequence of $s(n)$ is used:

$$s_\nu: \quad s(\nu) \quad s(N+\nu) \quad s(2N+\nu) \quad \ldots$$

Once the sound field $\hat{p}(\phi, \nu), \nu = 0, \ldots, N-1$ is estimated, the BRIR $\hat{h}(\phi, k)$ is computed by using (3). Since this is an interpolation problem, the performance depends on the distribution of the spatial sampling points and the interpolation order [13].

Considering that the ear microphone moves on a circle, the spatial bandwidth of the sound field in the circular harmonics domain has to be taken into account. The sound field consists of an incoming sound field (represented as an interior circular harmonics expansion) and a sound field scattered by the head (represented as an exterior circular harmonics expansion). For both expansions, the expansion coefficients exhibits an approximate spatial bandwidth of $\frac{2\pi f}{c} R$ [14], where $f$ denotes the temporal frequency, $R$ the radius of the circle, and $c$ the speed of sound.

For a uniform rotation with angular speed $\Omega$ (°/s), the sound field is sampled at equi-angular positions on the circle (if $M_{\text{eff}} \in \mathbb{Z}$). The number of spatial sampling points has to be sufficient, so that the reconstructed sound field does not suffer from spatial aliasing. Based on the approximated spatial bandwidth mentioned above, an anti-aliasing condition was derived in [11, Eq. (14)],

$$\Omega < \frac{c}{RN} \times \frac{180}{\pi}. \tag{6}$$

If this condition is not fulfilled, using a higher-order interpolation does not improve the performance of the dynamic measurement [13].

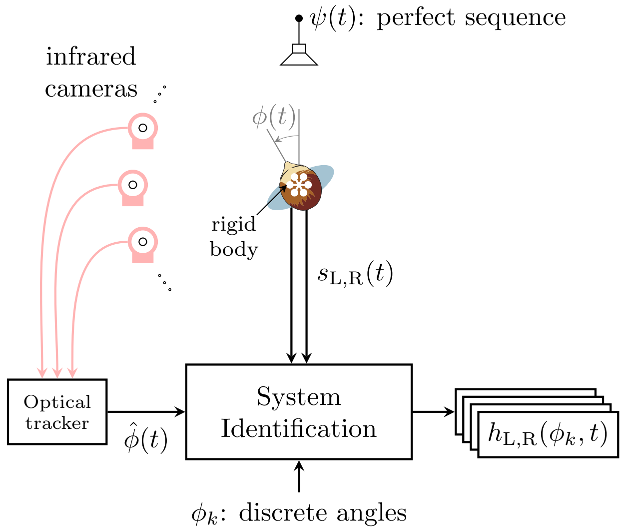



**Figure 1:** Continuous measurement of BRIRs. A perfect sequence $\psi(n)$ is played back by the loudspeaker on the $y$-axis. The responses at the outer ears $s_{\mathrm{L,R}}(t)$ are captured by a dummy head which rotates on a turntable. The head-orientation is tracked by an optical tracking system. The BRIRs are computed for selected angles $\phi_k$.

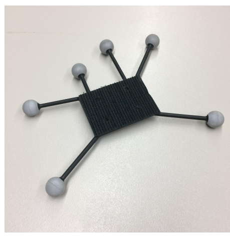


**Figure 2:** Six retro-reflective markers forming a rigid body. The position and rotation are tracked by 12 cameras.

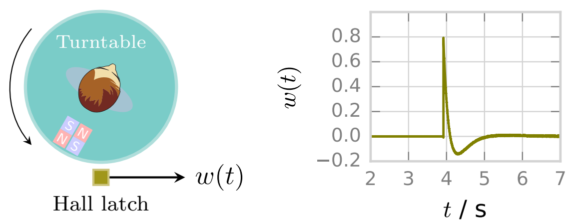



**Figure 3:** Bipolar Hall-effect sensor. Two small-sized permanent magnets are attached on the turntable with opposite polarities. The Hall latch generates a step signal when the magnetic field changes from positive to negative. The signal is recorded by the audio interface together with the microphone signals. The DC components of the step signal is removed by the mic preamp.

## Measurement

The BRIRs were measured in a rectanguler room ($W \times L \times H = 5.8 \times 5.0 \times 3.0$ m$^3$) at the Institute of Communications Engineering, University of Rostock. The room was treated with absorptive materials which gives the reverberation time of 0.22 s.

The measurement set-up is depicted in Figure 1. A head-and-torso simulator (G.R.A.S. type 45BA, Large ears type KB0065 & 0066, condenser microphones type 40AO) was mounted on a turntable (Varisphear [15]) at the center of the room. The microphone signals were recorded with a pre-amp (Lake People C360) and an audio interface (RME Fireface UC). A full-range 2-way loudspeaker (Neumann KH120A) was placed in front of the dummy head ($\phi = 0$) at a distance of 2.2 m. The loudspeaker was driven by a periodic perfect sequence of length $N = 88200$ at sampling rate of $f_\mathrm{s} = 44.1$ kHz [16]. A rigid body, consisting of six infra-reflective markers (Fig. 2), was attached on top of the head. Twelve cameras surrounding the dummy head capture the rigid body with a frame rate of $f_\mathrm{opt} = 120$ Hz. The position and orientation were computed by a dedicated software (Motive). In this software, the smoothing parameter for tracking was set to 5 (0: no smoothing, 100: maximum smoothing).

In order to compensate the latency of the optical tracking system, a bipolar Hall latch (Unisonic U18) was mounted on the static part of the turntable. Two small-sized permanent magnets were attached on the moving part of the turntable with opposite polarities [17, Ch. 3]. The midpoint of the two magnets, where the magnetic field changes abruptly, was placed at a predefined azimuth angle $\phi_\mathrm{ref}$. The Hall latch is switched on when the turntable reaches $\phi_\mathrm{ref}$ (Fig. 3). The sensor output was recorded by the audio interface synchronously with the ear signals. Once the dynamic measurement is completed, the audio signal $s(n)$, the Hall latch signal $w(n)$,

| $t_0$ | $t_1$ | $\dots$ | $t_n$ | $\dots$ |
|---|---|---|---|---|
| $s(0)$ | $s(1)$ | $\dots$ | $s(n)$ | $\dots$ |
| $w(0)$ | $w(1)$ | $\dots$ | $w(n)$ | $\dots$ |

and the tracking data $\phi(l)$,

| $\tau_0$ | $\tau_1$ | $\dots$ | $\tau_l$ | $\dots$ |
|---|---|---|---|---|
| $\phi(0)$ | $\phi(1)$ | $\dots$ | $\phi(l)$ | $\dots$ |

are obtained. Note that the sampling rate of $\phi(l)$ is different from other signals, i.e. $\tau_l = l/f_\mathrm{opt}$ and $t_n = n/f_\mathrm{s}$. The time $\tau_\mathrm{ref}$ and $t_\mathrm{ref}$ corresponding to the reference angle $\phi_\mathrm{ref}$ are computed from $\phi(l)$ and $w(n)$, respectively. By applying a shift of

$$\tau_l \leftarrow \tau_l - \tau_\mathrm{ref}$$
$$t_n \leftarrow t_n - t_\mathrm{ref},$$

the two time axes are aligned, i.e. $\tau_0 = t_0$.

Three different angular speeds ($\Omega = 0.5, 4.0, 8.0°/\mathrm{s}$) were considered. The anti-aliasing condition according to (6) is $\Omega < 2.2°/\mathrm{s}$. The dummy head starts at $\phi = -100°$ and stops at $\phi = 100°$. The acceleration ($\dot{\Omega}$) and jerk ($\ddot{\Omega}$) were set to $30°/\mathrm{s}^2$ and $10000°/\mathrm{s}^3$, respectively. As shown in the table below, the higher the angular speed, the smaller is the effective number sampling points. The signal-to-noise ratio (SNR) decreases for faster movement due to the motor noise.[1]

| $\Omega$ (°/s) | $M_\mathrm{eff}$ | SNR (dB) |
|---|---|---|
| 0.5 | 202 | 30.0 |
| 4.0 | 50 | 17.1 |
| 8.0 | 25 | 13.1 |

The BRIRs are computed for 181 head orientations,

$$\varphi = -90, -89, \dots, 90°.$$

For the reconstruction of the original sound field, linear interpolation was used in all cases. The BRIRs for the same angles were measured by using the conventional static technique. A perfect sequence of the same period ($N = 88200$) was used and the $\hat{h}(\varphi, n)$ was computed by using (3).

## Evaluation

The accuracy of the dynamic measurement is evaluated in terms of normalized system distance $\mathcal{D}(\varphi)$,

$$\mathcal{D}(\varphi) = \frac{\sum_{k=0}^{N_0-1} |\hat{h}(\varphi, k) - h(\varphi, k)|^2}{\sum_{k=0}^{N_0-1} |h(\varphi, n)|^2}, \tag{7}$$

where the static measurement is considered as the reference $h(\varphi, k)$. Only the first part ($N_0 = 22050$ samples) of the BRIRs was used for the evaluation, as the later part goes far below the noise floor in both static and dynamic measurements.

The system distance for different $\Omega$ is compared in Fig. 4. As expected, the higher the angular speed, the larger is the system distance. This is attributed to the increased SNR and the reduced spatial sampling points. The angular speed of $\Omega = 4, 8$ does not satisfy the anti-aliasing condition (6), and thus the sound field cannot be reconstructed from the captured signal. This results in spatial aliasing as shown in Fig. 5. The time-of-arrival of the wavefront exhibits discontinuities. On the contrary, for $\Omega = 0.5$ which satisfies the anti-aliasing condition, the spatio-temporal structure of BRIRs is successfully recovered.

The BRIRs achieving the best and worst system distance are shown in Fig. 6. The FIR coefficients in the range of $[0, 250]$ ms are shown in decibel for dynamic measurements together with the coefficient errors (red). In the best case 6b, the amplitude of the coefficient error remains 20 dB below the BRIR in the first 30 ms, and gradually decreases. In the worst case 6a, the coefficient error has almost the

[1] For the computation of SNR, it was assumed that the motor sound is dominant compared to other sources of noise. The energy of the signal was obtained from the static measurement and averaged. The energy of the noise was obtained by moving the motor without the excitation signal.

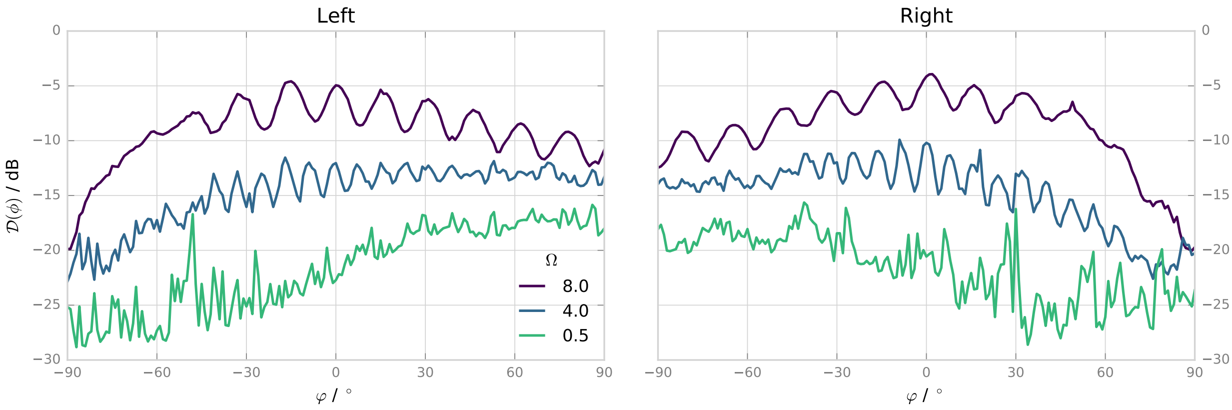



**Figure 4:** System distance for varying microphone speeds ($\Omega = 0.5, 4.0, 8.0$ °/s). The original sound field is reconstructed from the captured signals, each of which constitutes a spatio-temporal sampling of the sound field. The Lagrange polynomial of 3rd order was used for the spatial interpolation.

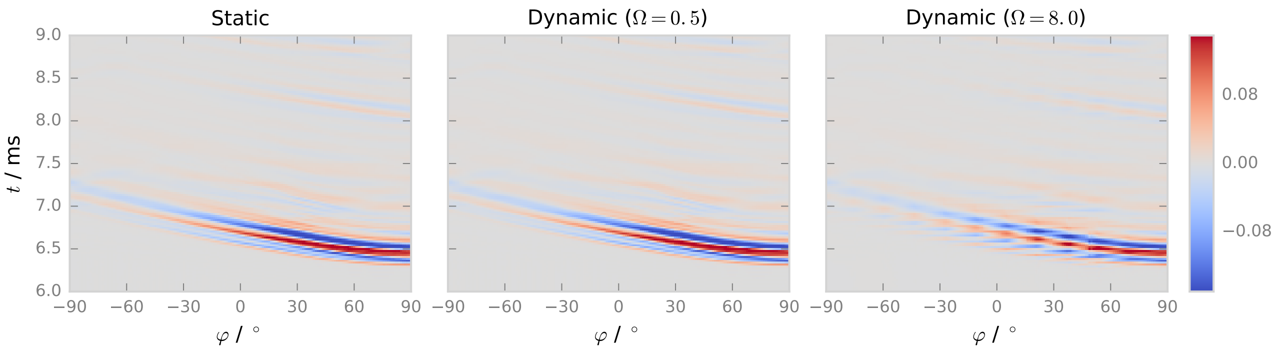



**Figure 5:** The early part ($t \in [6, 9]$ ms) of the static and dynamic measurements (right ear). Each vertical slice corresponds to the BRIR for the respective angle. The time-of-arrival of the direct sound (high-amplitude impulse) varies with the orientation of the head. The BRIRs for $\Omega = 8$ (right) are not able to track the change and suffer from discontinuities. This is attributed to the spatial aliasing which occurs if the angular speed does not satisfy the anti-aliasing condition, or equivalently, if there are not enough spatial sampling points (low $M_{\mathrm{eff}}$).

same envelope of the BRIR. The BRIRs obtained from dynamic and static measurement are directly compared in the lower figures. The corresponding ranges are indicated by horizontal bars ⊢⊣. Note the different scale of the axes in the lower plots. Even in the worst case, the temporal structure of the early part (left) is reasonably identified, but the later part (right) suffers from a low frequency error. The best case BRIR, on the other hand, exhibits a very good accuracy.

For an informal listening, ear signals were generated with the static and dynamic BRIR measurements. Dry recordings of speech [18] and castanets [19] signals were filtered with the BRIRs (available for download at `http://www.spatialaudio.net/dynamic_brir_measurement`). In all cases, the azimuth localization was not impaired even for the least accurate BRIRs. For $\Omega = 0.5$, the ear signals were indistinguishable from those generated with the static measurement. The ear signals for $\Omega = 4, 8$ suffers from clearly audible artifacts, which make it not suitable for an auralization. The artifacts are more sever for the speech than the castanets.

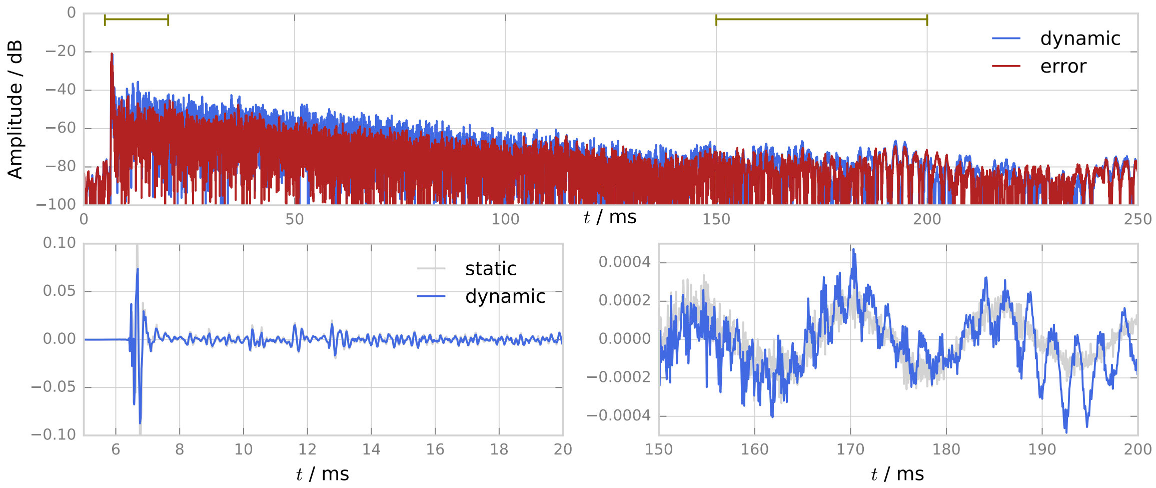
# Conclusion

The time-varying identification method based on spatial interpolation was used for the dynamic measurement of BRIRs in a real room. An optical tracking system was used to capture the rotation of the head-and-torso simulator which was rotated on a turntable. The BRIRs measured at $\Omega = 0.5°$/s was able to achieve an averaged system distance of $-20.9$ dB with the measurement period of 7 min. For $\Omega = 4, 8°$/s, where the anti-aliasing condition was not fulfilled, the BRIRs are not usable for auralization. This is attributed to the poor SNR and spatial aliasing, but it is not clear which has more impact than the other.

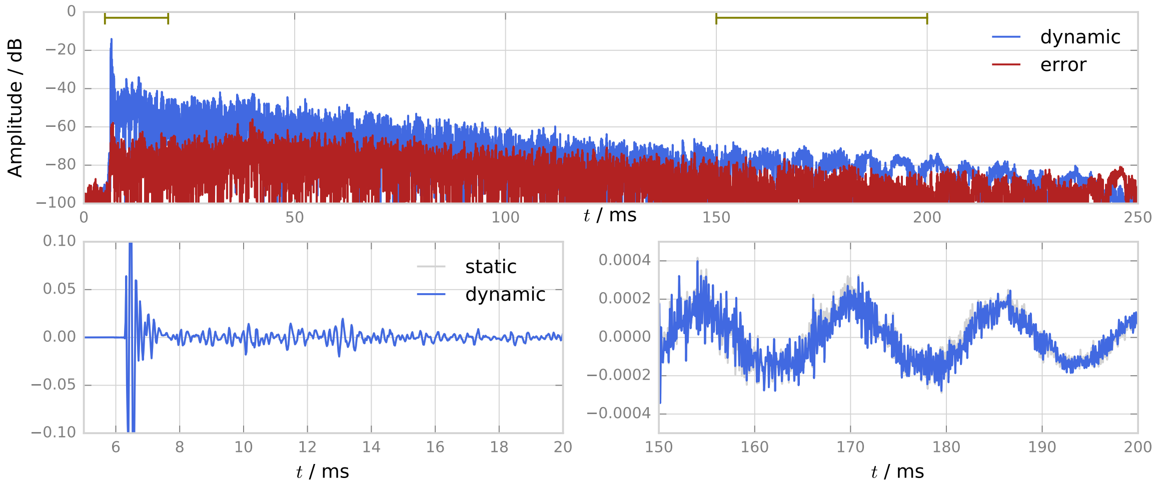
The presented method can be applied for the measurement of individual HRIRs/BRIRs, where the listener can freely moves the head during the measurement. To be able to incorporate more general head movements, the method has to be extended for more complex trajectories.

## Acknowledgement


This research was supported by a grant of the *Deutsche Forschungsgemeinschaft (DFG)* SP 1295/7-1.




**(a)** Right ear, $\phi = 2°$, $\Omega = 8°/\text{s}$, $\mathcal{D}(\phi) = -3.9$ dB (worst case)




**(b)** Left ear, $\phi = -87°$, $\Omega = 0.5°/\text{s}$, $\mathcal{D}(\phi) = -28.8$ dB (best case)

**Figure 6:** BRIRs from the dynamic measurement that exhibit the (a) worst and the (b) best system distance.

# References


[1] H. Wierstorf, A. Raake, and S. Spors, *Binaural Assessment of Multichannel Reproduction.* Berlin Heidelberg: Springer, 2013, pp. 255–278.

[2] H. Wierstorf, C. Hohnerlein, S. Spors, and A. Raake, "Coloration in wave field synthesis," in *Proc. 55th Audio Engineering Society (AES) International Conference*, Helsinki, Finland, August 2014, pp. Paper 5–3.

[3] A. Lindau, T. Hohn, and S. Weinzierl, "Binaural resynthesis for comparative studies of acoustical environments," in *Proc. of the 122nd Audio Engineering Society (AES) Convention.* Vienna, Austria: Audio Engineering Society, may 2007.

[4] E. M. Wenzel, M. Arruda, D. J. Kistler, and F. L. Wightman, "Localization using nonindividualized head-related transfer functions," *The Journal of the Acoustical Society of America (JASA)*, vol. 94, no. 1, pp. 111–123, 1993.

[5] H. Møller, M. F. Sørensen, C. B. Jensen, and D. Hammershøi, "Binaural technique: Do we need individual recordings?" *Journal of the Audio Engineering Society (JAES)*, vol. 44, no. 6, pp. 451–469, 1996.

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

[7] E. M. Hulsebos, "Auralization using Wave Field Synthesis," Ph.D. dissertation, Delft University of Technology, Delft, The Netherlands, 2004.

[8] T. Ajdler, L. Sbaiz, and M. Vetterli, "Dynamic Measurement of Room Impulse Responses Using a Moving Microphone," *The Journal of the Acoustical Society of America (JASA)*, vol. 122, no. 3, pp. 1636–1645, 2007.

[9] C. Antweiler and G. Enzner, "Perfect Sequence LMS for Rapid Acquisition of Continuous-azimuth Head Related Impulse Responses," in *Proc. of IEEE*

*Workshop on Applications of Signal Processing to Audio and Acoustics (WASPAA)*, New Paltz, NY, USA, Oct. 2009, pp. 281–284.

[10] N. Hahn and S. Spors, "Identification of Dynamic Acoustic Systems by Orthogonal Expansion of Time-variant Impulse Responses," in *Proc. of the 6th International Symposium on Communications, Control and Signal Processing (ISCCSP)*, Athens, Greek, May 2014.

[11] ——, "Continuous Measurement of Impulse Responses on a Circle Using a Uniformly Moving Microphone," in *Proc. of the European Signal Processing Conference (EUSIPCO)*, Nice, France, Aug. 2015.

[12] J. He, R. Ranjan, and W.-S. Gan, "Fast continuous HRTF acquisition with unconstrained movements of human subjects," in *Acoustics, Speech and Signal Processing (ICASSP), 2016 IEEE International Conference on.* IEEE, 2016, pp. 321–325.

[13] N. Hahn and S. Spors, "Spatial Aliasing in Continuous Measurement of Spatial Room Impulse Responses," in *Proc. of the 43nd German Annual Conference on Acoustics (DAGA)*, Kiel, Germany, Mar. 2017.

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

[15] B. Bernschütz, C. Pörschmann, S. Spors, and S. Weinzierl, "Entwurf und Aufbau eines variablen sphärischen Mikrofonarrays fÃ$\frac{1}{4}$r Forschungsanwendungen in Raumakustik und virtual Audio," in *Proc. of the 36th German Annual Conference on Acoustics (DAGA)*, Berlin, Germany, Mar. 2010.

[16] N. Aoshima, "Computer-generated Pulse Signal Applied for Sound Measurement," *The Journal of the Acoustical Society of America (JASA)*, vol. 69, no. 5, pp. 1484–1488, 1981.

[17] E. Ramsden, *Hall-Effect Sensors: Theory and Application.* Newnes, 2011.

[18] E. B. U. (EBU), "Sound quality assessment material recordings for subjective tests," Compact Disc, October 2008.

[19] M. Frank, "Castanets dry," Audio, 2009. [Online]. Available: http://ambisonics.iem.at/Members/frank/sounds/castanets-dry