Comparison of Continuous Measurement Techniques for Spatial Room Impulse Responses

Nara Hahn and Sascha Spors

Institute of Communications Engineering

University of Rostock, Germany

E-mail: {nara.hahn,sascha.spors}@uni-rostock.de

Abstract—A large number of spatial room impulse responses can be measured efficiently by using a moving microphone in combination with a time-varying system identification method. The microphone moves on a predefined trajectory and captures the response of the acoustic system which is periodically excited. The instantaneous impulse responses are computed from the captured signal by taking the time-variance explicitly into account. In this paper, three different continuous measurement techniques are investigated and compared in a unified framework. It is shown that impulse response estimation constitutes a spatial interpolation process, where each method corresponds to a specific interpolation filter. In numerical simulations the performance of theses approaches are evaluated in terms of system distance and spatial bandwidth.

I. Introduction

Multichannel sound reproduction systems with a high number of loudspeakers have recently gained considerable attention [1]. For the auralization of real environments, like cathedrals and concert halls, room impulse responses are measured in multiple positions in order to capture the spatio-temporal structure of the reverberant sound field. These spatial room impulse responses are often decomposed into plane waves [2], [3], and the plane wave components are reproduced by loudspeakers using a suitable multichannel reproduction technique, such as Wave Field Synthesis [4]. The achievable spatial resolution of the reproduced sound field scales with the number of impulse responses. Considering the dimensionality of a sound field within the audio frequency band (< 20 kHz), the theoretically required number of impulse responses is in the order of 10^5 even for a small bounded region (< 1 m) [5]. Measuring such a large number of impulse responses requires a lot of time and effort, and the measurement may suffer from time-variance of the system due to the change of room temperature [6] and the time-varying characteristics of electro-acoustics devices. Such changes are not easily predicted nor can be conveniently compensated.

The measurement time can be reduced considerably by employing a continuous measurement technique. Typically, the acoustic space is excited by a periodic signal and the response is captured by a continuously moving microphone. The instantaneous impulse responses are then computed from the captured signal. Since the dynamics of the system can be predicted, the time-variance can be explicitly considered. Several approaches have been proposed for this purpose [7]–[10]. Ajdler, for instance, proposed an analytic solution derived

in the spatio-temporal frequency domain by exploiting the projection-slice theorem [8]. The excitation signal is a sum of multiple sinusoids with appropriately spaced frequencies. Antweiler, on the other hand, employed a normalized least mean square (NLMS) algorithm. The system is excited by a so-called periodic perfect sequence, and the instantaneous impulse responses are computed sample-by-sample [9]. The computational complexity of this approach was reduced by Carini [11].

In [10], the authors proposed a novel system identification method, where each impulse response is represented as an orthogonal expansion. The expansion coefficients are computed from the captured signal by applying a spatial interpolation filter. The order and complexity of the interpolation filter can be chosen freely depending on the required technical and/or perceptual accuracy. The method is used for spatial room impulse response measurement [12] and binaural room impulse response (BRIR) measurement [13].

In this paper, the methods [8] and [9] are reviewed in the framework we proposed in [10]. It is shown that each method corresponds to a specific type of spatial interpolation filter. Along with the theoretical analysis, numerical simulations are presented where the technical performance is evaluated.

II. Continuous Measurement Techniques

We consider the measurement of impulse responses on a circular contour as depicted in Fig. 1(a). A discrete finite impulse response (FIR) model is assumed. For a given source signal $\psi(n)$, the sound field at (r_0, ϕ) is

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

where $h(\phi, l)$ denotes the l-th coefficient of the impulse response at angular position ϕ, and N the length of the impulse response. For convenience, r_0 is omitted.

We assume that the system is excited by a periodic perfect sequence $\psi(n) = \psi(n + \gamma N), \gamma \in \mathbb{Z}$, where the period N is longer than any impulse response of the system. The circular (cyclic) autocorrelation $\varphi_{\psi\psi}(n)$ of a perfect sequence yields

$$\varphi_{\psi\psi}(n) = \frac{1}{N}\sum_{n'=0}^{N-1} \psi(n')\psi(n'+n) = E \cdot \delta(n) \tag{2}$$

for $n = 0, \ldots, N-1$ where $\delta(\cdot)$ denotes the unit impulse function and E the energy within a period. Without loss of generality it is assumed that $E = 1$. Due to the autocorrelation of the perfect sequence, the impulse response is obtained by a length-N circular convolution of the captured signal and the time-reversed input signal $\psi(-n)$,

$$h(\phi, l) = \sum_{\eta=0}^{N-1} \left\{ \underbrace{\sum_{\nu=0}^{N-1} h(\phi, \nu) \psi(\eta - \nu)}_{p(\phi, \eta)} \right\} \psi(\eta - l) \quad (3)$$

for $l = 0, \ldots, N-1$.

For a given radius, the spatial-temporal impulse responses are represented as circular harmonics expansion [14],

$$H(\phi, \omega_k) = \sum_{m=-\infty}^{\infty} \mathring{H}_m(\omega_k) e^{im\phi} \quad (4)$$

$$\mathring{H}_m(\omega_k) = \frac{1}{2\pi} \int_0^{2\pi} H(\phi, \omega_k) e^{-im\phi} \mathrm{d}\phi \quad (5)$$

where $H(\phi, \omega_k)$ denotes the discrete Fourier transform (DFT) of $h(\phi, n)$, $\omega_k = \frac{2\pi}{L} k$ the k-th frequency bin, and $\mathring{H}_m(\omega_k)$ the m-th circular harmonics expansion coefficient.

Assume that the microphone moves at a constant angular speed Ω in rad/sample (Ωf_s in rad/s with f_s denoting the sampling frequency). The microphone captures the sound field at the discrete positions $\phi_n = \Omega n$ and thus the captured signal reads

$$s(n) = p(\Omega \cdot n, n), \quad n = 0, \ldots, L-1 \quad (6)$$

where the total length of the captured signal $L = \frac{2\pi}{\Omega}$ is assumed to be an integer ($\phi_n = \frac{2\pi}{L} n$). Note that only one sample is available for each angular position, and thus (3) cannot be used straightforwardly for system identification.

In the subsequent, three different approaches are introduced and their relation is revealed.

A. Interpolation of Orthogonal Expansion Coefficients

In our work [10], (3) is interpreted as an orthogonal expansion

$$h(\phi_n, l) = \sum_{\eta=0}^{N-1} a_\eta(\phi_n) \psi(\eta - l) \quad (7)$$

where the basis functions are time-shifted and time-reversed sequences $\psi(\eta - l)$ and $a_\eta(\phi_n) = p(\phi_n, \eta)$ the corresponding expansion coefficients. The orthogonality of the basis functions can be deduced from (2).

By plugging (7) into (1), it is easily shown that [12, Eq. (9–11)]

$$s(n) = a_{\bar{n}}(\phi_n) \quad (8)$$

where $\bar{n} = n \bmod N$. This states that the captured signal corresponds to a spatial sampling of the expansion coefficients. As illustrated in Fig. 1(b), each coefficient is sampled at $M = \frac{L}{N}$ equiangular positions, and thus $a_\eta(\phi_n)$ is decimated by a factor of N.

The missing expansion coefficients can be recovered if the spatial bandwidth of $a_\eta(\phi_n)$, and equivalently the spatial
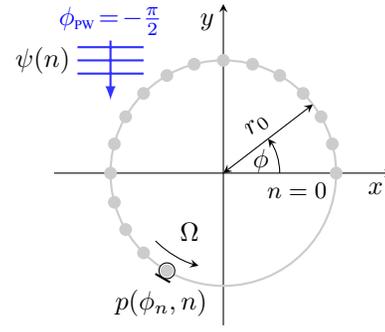
(a) configuration
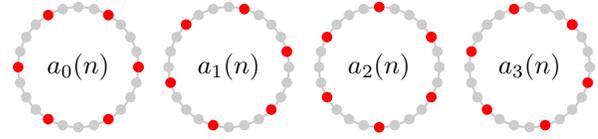
(b) decimated $a_\eta(n)$

Fig. 1: (a) Continuous measurement on a circle ($N = 4, L = 24$). The gray circles indicate the positions where the signal is captured. (b) The relation of $s(n)$ and $a_\eta(n)$. The red circles indicate the positions where the respective coefficients are captured. In this example, the effective number of spatial sampling positions is $\frac{L}{N} = 6$.

bandwidth of $h(\phi_n, l)$, is smaller than M [3]. For a bounded circular region, the circular harmonics coefficients exhibit a low-pass characteristic and the bandwidth is approximated by $\frac{2\omega_{\max}}{c} r_0 = \frac{2\pi f_s}{c} r_0$ with c denoting the speed of sound [5]. The anti-aliasing condition $\frac{2\pi f_s}{c} r_0 < M$ leads to the maximum allowable angular speed of the microphone [12, Eq. (14)]

$$\Omega < \frac{c}{r_0 N f_s}. \quad (9)$$

Once the latter is fulfilled, $a_\eta(\phi_n)$ can be obtained by applying an interpolation filter and finally $h(\phi_n, l)$ is computed by using (7). The order of the spatial interpolation filter is determined by application specific requirements. In [12], we observed that perceptually acceptable results can be achieved even with linear or cubic spline interpolation. Though, the perceptual evaluation is out of the scope of this paper, and only the technical performance is considered.

B. Adaptive Filtering

The NLMS algorithm is often used in combination with a periodic perfect sequence for system identification. The filter coefficients are updated sample-by-sample [15],

$$\hat{h}(\phi_n, l) = \hat{h}(\phi_{n-1}, l) + \Delta \frac{\varepsilon(n)}{E} \psi(n - l), \quad (10)$$

where Δ denotes the step size of the adaptation. The error $\varepsilon(n)$ is the difference between the captured signal $s(n)$ and

its estimate $\hat{s}(n)$,

$$\varepsilon(n) = s(n) - \underbrace{\sum_{l=0}^{N-1} \hat{h}(\phi_{n-1}, l)\psi(n-l)}_{\hat{s}(n)}. \qquad (11)$$

The step size Δ is set to unity which was shown to be the optimal step size for fast convergence under noiseless conditions which is also assumed here [16]. Again, $E = 1$ is assumed.

Note that $\psi(n-l)$ in (10) and (11) is the $\bar{n}$-th orthogonal basis introduced in (7). As pointed out in [9] and [17], each orthogonal component is updated once in a period. When a coefficient is updated it remains unchanged during the following $N-1$ samples. By using (8) and the fact that

$$\hat{s}(n) = \hat{a}_{\bar{n}}(\phi_n) = a_{\bar{n}}(\phi_{n-N}), \qquad (12)$$

(10) can be reformulated as

$$\hat{h}(\phi_n, l) = \hat{h}(\phi_{n-1}, l) + \left\{ a_{\bar{n}}(\phi_n) - a_{\bar{n}}(\phi_{n-N}) \right\} \psi(\bar{n} - l). \qquad (13)$$

The old value of the $\bar{n}$-th coefficient is replaced with $a_{\bar{n}}(\phi_n)$ which is equal to $s(n)$. The NLMS algorithm thus interpolates the decimated expansion coefficients with a rectangular window

$$g_{\text{NLMS}}(\phi_n) = \begin{cases} 1, & \frac{2\pi}{M} \le \phi_n \le 0 \\ 0, & \text{elsewhere.} \end{cases} \qquad (14)$$

In [10] and [18], it was pointed out that NLMS with $\Delta = 1$ is equivalent to the length-N circular convolution of $s(n)$ and the time reversed perfect sequence $\psi(-n)$

$$\hat{h}(\phi, l) = \sum_{\nu=0}^{N-1} s(l-\nu)\psi(-\nu). \qquad (15)$$

This indicates that this method implicitly assumes piece-wise time-invariance of the system.

Note that $g_{\text{NLMS}}(\phi_n)$ in (14) is not centered at the origin but shifted by $-\frac{\pi}{M}$. This is attributed to the fact that NLMS is an on-line algorithm and takes only the values up to the n-th sample. As a result, the tracking of the system is lagged and $\hat{h}(\phi_n, n)$ suffers from an angular shift. If on-line processing is not a requirement, the accuracy can be improved by shifting $g_{\text{NLMS}}(\phi_n)$ by $\frac{\pi}{M}$, or equivalently by modifying (10) as

$$\hat{h}(\phi_n, l) = \hat{h}(\phi_{n-1}, l) + \varepsilon(n + \tfrac{N}{2})\psi(n + \tfrac{N}{2} - l). \qquad (16)$$

In the remainder of this paper, only this modified NLMS is considered.
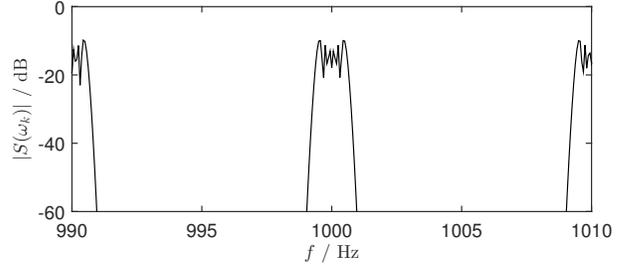
C. Projection-Slice Theorem

The approach introduced in [8] was formulated in the continuous-time and continuous-space domain. In the following, its discrete-time and -space variant is derived in order to enable a direct comparison with the other continuous measurement techniques.
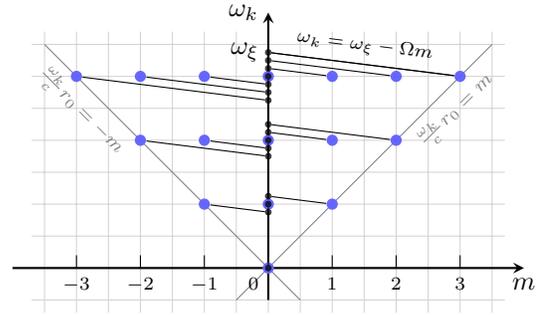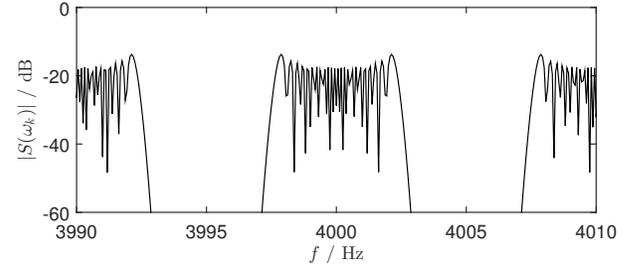

(a) mapping of $\mathring{P}_m(\omega_\xi)$ and $S(\omega_k)$
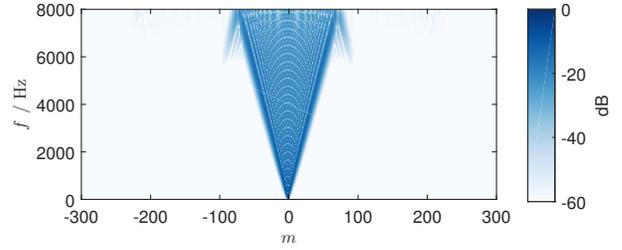
(b) $\mathring{P}_m(\omega_k)$

(c) $S(\omega_k)$, $f_{\text{center}} = 1$ kHz

(d) $S(\omega_k)$, $f_{\text{center}} = 4$ kHz

Fig. 2: Projection-slice approach. (a) The blue and black circles indicate $\mathring{P}_m(\omega_k)$ and $S(\omega_k)$, respectively. The bandpass components in (c) and (d) correspond to horizontal slices of (b) at f_{center}.

The sound field on the circular contour can be represented as a double inverse DFT of the spatio-temporal spectrum $\mathring{P}_m(\omega_k)$ (refer to (4) and (5)),

$$p(\phi_n, n) = \frac{1}{L} \sum_{k=0}^{L-1} P(\phi_n, \omega_k) e^{i\omega_k n}$$

$$= \frac{1}{L} \sum_{k=0}^{L-1} \left\{ \sum_{m=0}^{L-1} \mathring{P}_m(\omega_\xi) e^{im\phi_n} \right\} e^{i\omega_k n}. \qquad (17)$$

Due to the N-periodicity of $p(\phi_n, n)$, the temporal frequency spectra $P(\phi_n, \omega_k)$ and $\mathring{P}_m(\omega_k)$ are discrete and have nonzero values only for $k = \gamma M, \gamma \in \mathbb{Z}$.

Since $s(n)$ is a slice of the two-dimensional function $p(\phi_n, n)$ along $\phi_n = \Omega n$, the temporal frequency spectrum of the captured signal is

$$S(\omega_k) = \sum_{n=0}^{L-1} p(\Omega n, n) e^{-i\omega_k n} \qquad (18)$$

where $\omega_k = \frac{2\pi}{L} k, k = 0, \ldots, L-1$. If we plug (17) into (18)

$$S(\omega_k) = \sum_{n=0}^{L-1} \left\{ \frac{1}{L} \sum_{\xi=0}^{L-1} \left(\sum_{m=0}^{L-1} \mathring{P}_m(\omega_\xi) e^{i\Omega m n} \right) e^{i\omega_\xi n} \right\} e^{-i\omega_k n}$$

$$= \frac{1}{L} \sum_{\xi=0}^{L-1} \sum_{m=0}^{L-1} \mathring{P}_m(\omega_\xi) \left(\sum_{n=0}^{L-1} e^{i(\Omega m + \frac{2\pi}{L}\xi - \frac{2\pi}{L}k)n} \right). \qquad (19)$$

Since $N = \frac{L}{M}$ and $\Omega = \frac{2\pi}{L}$, the bracketed term in (19) is an impulse train with period L

$$\sum_{n=0}^{L-1} e^{i\frac{2\pi}{L}(m+\xi-k)n} = L \sum_{\gamma \in \mathbb{Z}} \delta(m + \xi - k + \gamma L), \qquad (20)$$

and therefore,

$$S(\omega_k) = \sum_{m=0}^{L-1} \mathring{P}_m(\omega_{k-m}) = \sum_{\xi=0}^{L-1} \mathring{P}_{k-\xi}(\omega_\xi). \qquad (21)$$

This constitutes a projection of $\mathring{P}_m(\omega_k)$ along $k = m + \xi$ in the m-ω_k plane. The result is known as the *projection-slice theorem* and states that the Fourier transform of a slice of a two-dimensional function is equivalent to a projection of the two-dimensional Fourier transform. To be able to obtain $\mathring{P}_m(\omega_k)$ from $S(\omega_k)$, there has to be only one term on the right hand side of (21) so that it is a one-to-one mapping. This requires $\mathring{P}_m(\omega_k)$ to be spatially band-limited ($|m| \leq \frac{M}{2}$) and Ω to satisfy condition (9). As depicted in Fig. 2(a), the frequency bins of $S(\omega_k)$ indicated by black circles • are mapped back to $\mathring{P}_m(\omega_k)$ indicated by blue circles ●. In terms of vector/matrix computations, the $L \times 1$ vector composed of $S(\omega_k)$ is converted to an $N \times M$ matrix. An example is shown in Fig 2(b), 2(c) and 2(d). As $\mathring{P}_m(\omega_k)$ exhibits a low-pass characteristic along the m-axis, $S(\omega_k)$ consists of bandpass components with varying bandwidth.

In order to obtain the impulse responses, the sound field $p(\phi, n)$ or equivalently $a_\eta(\phi)$ is computed from the band-limited circular harmonics coefficients $\mathring{P}_m(\omega_k)$ by using (4). This corresponds to an ideal low-pass filter in the spatial domain, and thus the decimated orthogonal expansion coefficients are interpolated by the periodic sinc function

$$g_{\text{PS}}(\phi_n) = \frac{1}{M} \frac{\sin\left(M \frac{\phi_n}{2}\right)}{\sin\left(\frac{\phi_n}{2}\right)}. \qquad (22)$$

In practice, a successful implementation of the projection-slice approach is not trivial. If $\frac{L}{N}$ is not exactly an integer, (21) is no longer a one-to-one mapping due to the leakage

(a) $h(\phi, t), \phi = \frac{\pi}{2}$

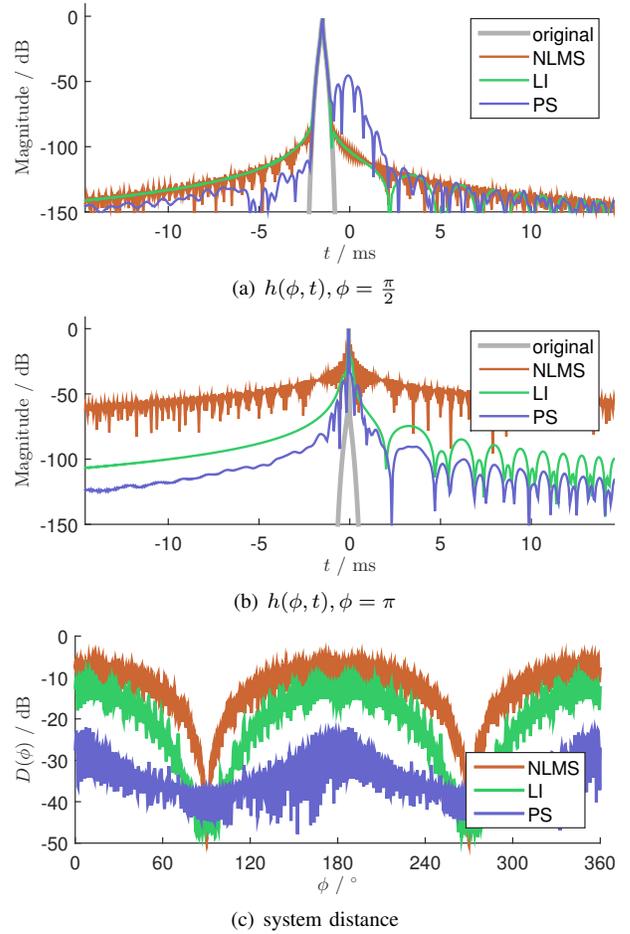

(b) $h(\phi, t), \phi = \pi$

(c) system distance

Fig. 3: Comparison of the measured impulse responses. (a) and (b) show the magnitude of the impulse responses in dB. (c) shows the system distance (23) of the estimated impulse responses for different angles.

effect. Moreover, if the assumption of spatial band-limitation does not hold, the result is strongly degraded.

III. EVALUATION

The performance of the continuous measurement techniques is evaluated in numerical simulations. A Dirac-shaped plane wave propagating in the direction $\phi_{\text{PW}} = -\frac{\pi}{2}$ is assumed as incident sound field for $f_s = 16$ kHz, $r_0 = 0.5$ m, $\Omega \times f_s = \frac{\pi}{8}$ rad/s, $N = 1600$ (0.1 s). The spatial room impulse response $h(\phi, t) = \delta\left(t - \frac{r_0}{c} \cos(\phi - \phi_{\text{PW}})\right)$ was simulated by using fractional delay filters (Lagrange filter of order 23) [19]. Three different approaches were implemented:

- linear interpolation (LI) of $a_\eta(\phi_n)$ (Sec. II-A)
- modified NLMS algorithm (Sec. II-B) and
- the projection-slice (PS) approach (Sec. II-C).

The original and estimated impulse responses are compared in Fig. 3(a) and 3(b) for $\phi_n = \frac{\pi}{2}$ and $\phi_n = \pi$, respectively. High accuracy is achieved where a high-order spatial interpolation filter is used. Clearly, the performance is strongly influenced by the time variability of the system. For $\phi_n = \pi$

(a) NLMS

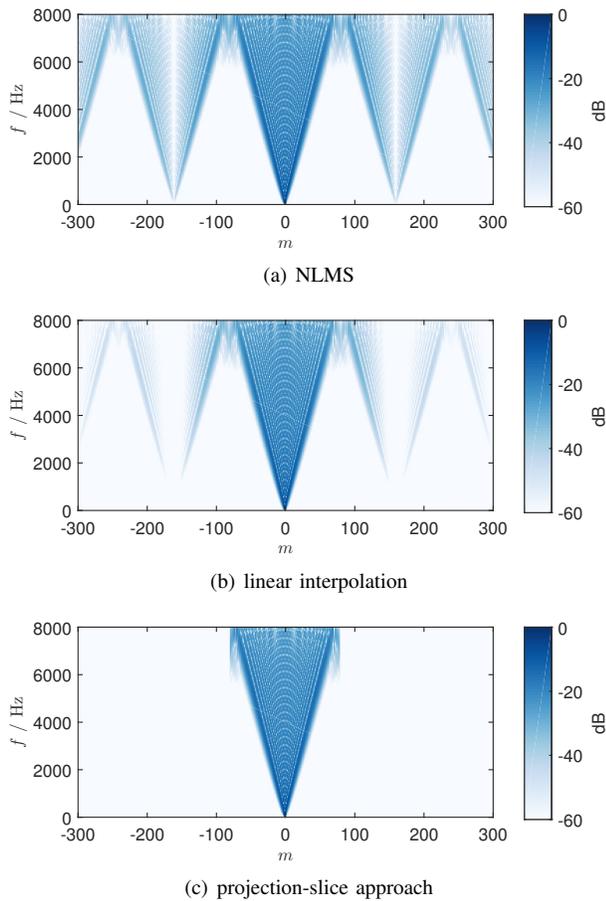(b) linear interpolation

(c) projection-slice approach

Fig. 4: Circular harmonics spectra of continuously measured impulse responses.

where the time-of-arrival of the plane wave varies rapidly, the performance is degraded irrespective of the employed method. This can be seen more clearly in Fig. 3(c) where the normalized system distance is compared for different angles

$$D(\phi_n) = \left(\frac{\sum_{l=0}^{N-1} |h(\phi_n, l) - \hat{h}(\phi_n, l)|^2}{\sum_{l=0}^{N-1} |h(\phi_n, l)|^2} \right)^{\frac{1}{2}}. \quad (23)$$

Close to $\phi_n = \frac{\pi}{2}$ and $\phi_n = \frac{3\pi}{2}$, two low-order interpolators (NLMS and linear interpolation) exhibit comparable or even better performance as the projection-slice approach. It is likely that the system is almost time-invariant in this range, thus high-order interpolation has little advantage.

Since the spatial room impulse responses are intended for sound field analysis with high spatial resolution, the spatial frequency spectrum is also examined. The circular harmonics expansion coefficients are shown in Fig. 4. Compare them with Fig. 2(b). NLMS and linear interpolation both suffer from spatial aliasing, whereas the projection-slice approach suppresses the spectral repetitions.

IV. CONCLUSION

Currently available continuous measurement techniques are formulated and compared in a unified framework. It was shown that these methods can be considered as spatial interpolation filters: NLMS corresponds to rectangular window whereas the projection-slice approach corresponds to the periodic sinc function. Numerical simulations indicate that the projection-slice approach exhibits favorable properties in terms of system distance and suppression of spatial aliasing.

REFERENCES

[1] S. Spors, H. Wierstorf, A. Raake, F. Melchior, M. Frank, and F. Zotter, "Spatial Sound with Loudspeakers and its Perception: A Review of the Current State," *IEEE Proc.*, vol. 101, no. 9, pp. 1920–1938, September 2013.

[2] B. Rafaely, "Plane-wave Decomposition of the Sound Field on a Sphere by Spherical Convolution," *J. Acoust. Soc. of Am.*, vol. 116, no. 4, pp. 2149–2157, 2004.

[3] A. Kuntz, *Wave Field Analysis Using Virtual Circular Microphone Arrays*. München: Verlag Dr. Hut, 2009.

[4] A. J. Berkhout, D. de Vries, and P. Vogel, "Acoustic control by wave field synthesis," *J. Acoust. Soc. of America*, vol. 93, no. 5, pp. 2764–2778, 1993.

[5] R. Kennedy, P. Sadeghi, T. Abhayapala, and H. Jones, "Intrinsic Limits of Dimensionality and Richness in Random Multipath Fields," *IEEE T. Signal Proces.*, vol. 55, no. 6, pp. 2542–2556, 2007.

[6] B. Bernschütz, C. Pörschmann, S. Spors, and S. Weinzierl, "Zeitin-varianzen durch Temperaturveränderung bei sequentiellen sphärischen Mikrofonarrays im Plane Wave Decomposition Verfahren," in *Proc. of the 37th German Annual Conference on Acoustics (DAGA)*, Düsseldorf, Germany, Mar. 2011.

[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," *J. Acoust. Soc. of America*, vol. 122, no. 3, pp. 1636–1645, 2007.

[9] C. Antweiler, "Multi-Channel System Identification with Perfect Sequences." John Wiley & Sons, 2008, pp. 171–198.

[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)*. IEEE, May 2014, pp. 161–164.

[11] A. Carini, "Efficient NLMS and RLS Algorithms for Perfect Periodic Sequences," in *International Conference on Acoustics Speech and Signal Processing (ICASSP)*. IEEE, 2010, pp. 3746–3749.

[12] N. Hahn and S. Spors, "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.

[13] ——, "Measurement of Time-Variant Binaural Room Impulse Responses for Data-Based Synthesis of Dynamic Auditory Scenes," in *Proc. of the 40th German Annual Conference on Acoustics (DAGA)*, Oldenburg, Germany, Mar. 2014.

[14] E. Williams, *Fourier Acoustics: Sound Radiation and Nearfield Acoustical Holography*. Academic Press, 1999.

[15] S. Haykin, *Adaptive Filter Theory*. Prentice Hall, 2002.

[16] C. Antweiler, A. Telle, and P. Vary, "NLMS-type System Identification of MISO Systems with Shifted Perfect Sequences," *Proc. International Workshop on Acoustic Signal Enhancement (IWAENC)*, 2008.

[17] N. Hahn and S. Spors, "Analysis of Time-Varying System Identification using Normalized Least Mean Square (NLMS) in the Context of Data-Based Binaural Synthesis," in *Proc. of the 42nd German Annual Conference on Acoustics (DAGA)*, Aachen, Germany, Mar. 2016.

[18] C. Antweiler, S. Kuehl, B. Sauert, and P. Vary, "System Identification with Perfect Sequence Excitation-Efficient NLMS vs. Inverse Cyclic Convolution," in *Proc. of the 11th ITG Conference on Speech Communication*. VDE, 2014, pp. 1–4.

[19] T. I. Laakso, V. Valimaki, M. Karjalainen, and U. K. Laine, "Splitting the Unit Delay," *Signal Processing Magazine, IEEE*, vol. 13, no. 1, pp. 30–60, 1996.