Spatial Aliasing in Continuous Measurement of Spatial Room Impulse Responses

Nara Hahn and Sascha Spors

Institute of Communications Engineering, University of Rostock, Germany

Email: `nara.hahn@uni-rostock.de`

Introduction

In sound field analysis and spatial sound reproduction, impulse responses are measured at multiple positions in order to capture the spatio-temporal structure of a sound field. This constitutes a spatial sampling [1, 2]. To achieve a high spatial resolution, a large number of measurements have to be performed, which requires a lot of time and effort.

Recently, continuous measurement techniques have gained attention [3, 4, 5, 6, 7]. In a continuous measurement, either the loudspeaker or the microphone moves on a predefined path, while the system is excited by an excitation signal. The instantaneous impulse responses are computed from the captured signal using a time-varying system identification method. Compared to conventional static measurement methods, a large number of impulse responses can be measured in a short period of time. Continuous techniques have been used for the measurement of spatial room impulse responses [3, 7, 8], head-related impulse responses [4, 9], and binaural room impulse responses [10, 6].

The discrete-time signal captured by the moving microphone constitutes a spatio-temporal sampling of the sound field [6]. To avoid spatial aliasing, the movement of the microphone has to be controlled carefully by considering the spatial bandwidth of the sound field [8]. In this paper, the impact of the microphone speed on spatial aliasing and the accuracy of the impulse responses is investigated.

The scope of this paper is restricted to the measurement of impulse responses on a circle of radius r_0, as illustrated in Fig. 1(a). The microphone moves at a constant angular speed Ω. It is further assumed that the sound field consists of a plane wave propagating under free-field conditions.
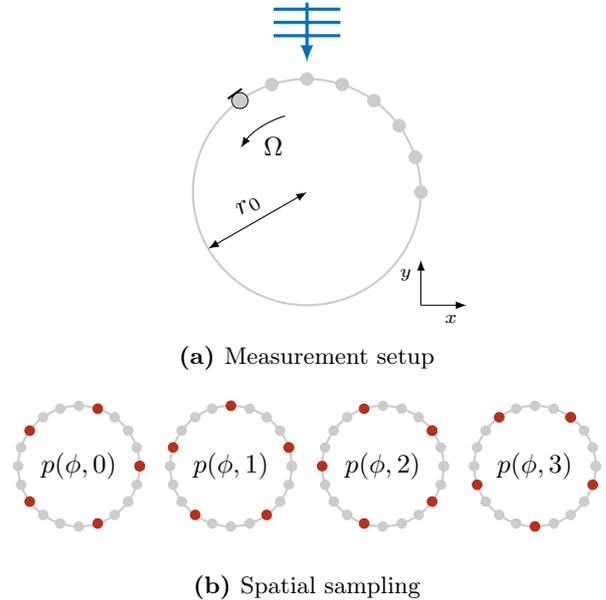
Perfect Sequence Excitation

In a continuous measurement, the acoustic system is typically excited by a periodic signal,

$$\psi(n) = \psi(n + N), \tag{1}$$

that exhibits a self-orthogonal property,

$$\sum_{m=0}^{N-1} \psi(m)\psi(m+n) = \sigma_s^2 \, \mathrm{III}_N(n), \tag{2}$$

where σ_ψ^2 denotes the energy of the signal within a period and III_N the impulse train with period N. Without

(a) Measurement setup

(b) Spatial sampling

Figure 1: Spatial sampling of the sound field in a continuous measurement. (a) The impulse responses of a Dirac shaped plane ☰ ($\phi_{\mathrm{pw}} = 270\,°$) wave are measured on a circle of radius r_0. The microphone ⟳ moves at a constant angular speed Ω. (b) The excitation signal has a period $N = 4$ and the total number of sampling points ● is $L = 20$. The effective number of sampling points ● is $\frac{L}{N} = 5$.

loss of generality, $\sigma_\psi^2 = 1$ is assumed in the remainder. The period of the excitation signal N has to be longer than the length of the impulse responses, so that the system is fully excited within a period, and also that the impulse response is not truncated or aliased in the time domain. A discrete-time signal satisfying (2) is referred to as a (periodic) perfect sequence [11]. Maximum length sequences (MLSs) and perfect sweeps [12] are well-known perfect sequences.

If the plane wave is driven by a perfect sequence, the sound field can be represented by a finite impulse response (FIR) model,

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

where ϕ denotes the polar angle of the receiver position, and $h(\phi, n)$ the impulse response. Note that the sound field is periodic in the time domain, $p(\phi, n) = p(\phi, n+N)$.

By exploiting (2), it can be shown that the impulse response is the circular cross-correlation of the sound field

and the excitation signal,

$$h(\phi, n) = \sum_{m=0}^{N-1} p(\phi, m)\psi(m + n). \qquad (4)$$

Due to the N-periodicity of $p(\phi, n)$ and $\psi(n)$, (4) also holds if m is replaced with $m + \mu N$ for an arbitrary integer $\mu \in \mathbb{Z}$.

Spatial Sampling

Although the movement of the microphone is continuous, the sound field can be captured only at a finite number of positions on the trajectory. The captured signal $s(n)$ thus constitutes a slice of the sound field in the (ϕ, n)-plane [4],

$$s(n) = p(\phi_{\mathrm{mic}}(n), n), \quad n = 0, \ldots, L - 1, \qquad (5)$$

where $\phi_{\mathrm{mic}}(n) = \Omega \times n$ denotes the polar angle of the time-varying microphone position. The total length of the signal is denoted by $L \equiv \frac{360}{\Omega} \times f_s$ with f_s denoting the sampling frequency. The number of sampling positions is thus proportional to f_s, and inversely proportional to Ω.

As illustrated in Fig. 1(b), the captured signal $s(n)$ can be decomposed into N sequences, where the ν-th sequence s_ν corresponds to a uniform sampling of the sound field at time $n = \nu + \mu N$,

$$\begin{aligned} s_\nu(\mu) &= s(\nu + \mu N) \\ &= p(\phi_{\mathrm{mic}}(\nu + \mu N), \nu + \mu N) \\ &= p(\phi_{\mathrm{mic}}(\nu + \mu N), \nu), \end{aligned} \qquad (6)$$

for $\mu \in \mathbb{Z}$. In the third equality, the periodicity of $p(\phi, n)$ is exploited. For each ν, the number of sampling points is $\frac{L}{N}$, and the distribution of the sampling points is angularly shifted by $\frac{2\pi\nu}{L}$ on the circle.

Spatial Interpolation

It was proposed in [6] to interpolate the time-domain sound field from the sampled values in (6). Once the sound field $p(\phi, n)$ is estimated, the impulse response at the corresponding position is computed by (4). This approach is quite flexible since the interpolation method can be chosen by considering the dynamics of the system, required technical/perceptual accuracy, and the available computational power. Linear and cubic spline interpolations were used in [6, 10] for the measurement of binaural room impulse responses, whereas higher-order interpolation was used in [8] for spatial room impulse responses.

Moreover, it was shown in [13], that currently available methods can be regarded as implicit spatial interpolations. The normalized least mean square (NLMS) algorithm, for instance, is equivalent to the nearest neighbour interpolation [14]. The method proposed in [4] corresponds to a sinc interpolation, which is ideal if the anti-aliasing condition is fulfilled.
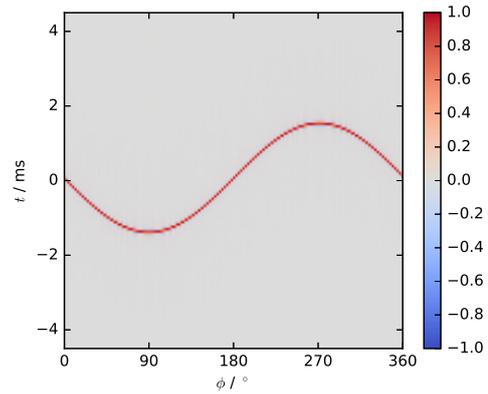
Figure 2: Impulse responses of a Dirac-shaped plane wave ($\phi_{\mathrm{pw}} = 270\,°$) on a circle of radius $r_0 = 0.5$ m. See (13).

Anti-aliasing Condition

To determine the required number of sampling points, the spatial bandwidth of the sound field has to be taken into account. For a given angular frequency $\omega = 2\pi f$, the sound field on a circle can be represented by a circular harmonics expansion [2],

$$P(\phi, \omega) = \sum_{m=-\infty}^{\infty} \mathring{P}_m(\omega) e^{im\phi}, \qquad (7)$$

where $\mathring{P}_m(\omega)$ denotes the m-th expansion coefficient, and c the speed of sound. The expansion coefficient for a plane wave $e^{-i\frac{\omega}{c} r_0 \cos(\phi - \phi_{\mathrm{pw}})}$ reads

$$\mathring{P}_m(\omega) = i^{-m} J_m(\tfrac{\omega}{c} r_0) e^{-im\phi_{\mathrm{pw}}}, \qquad (8)$$

where $J_m(\frac{\omega}{c} r_0)$ denotes the Bessel function of the first kind of order m. Although $\mathring{P}_m(\omega)$ is not band-limited in the circular harmonics domain, its magnitude decays exponentially for large m [15, Eq. (9.2.1)]. The spatial bandwidth is often approximated by [2, Sec. 4.2]

$$M_0 = \lceil \tfrac{2\pi f}{c} r_0 \rceil, \qquad (9)$$

where $\lceil \cdot \rceil$ denotes the ceiling function. Under this approximation, the number of sampling points must satisfy

$$\frac{L}{N} \geq 2M_0 = 2\lceil \tfrac{\pi f_s}{c} r_0 \rceil, \qquad (10)$$

which leads to the anti-aliasing condition for the angular speed [8, Eq. (14)],

$$\Omega \leq \Omega_0 \equiv \frac{c}{r_0 N}. \qquad (11)$$

A more pessimistic condition can be derived by approximating the spatial bandwidth as

$$M_\eta = M_0 + \eta, \quad \eta \in \mathbb{Z}. \qquad (12)$$

It is worth noting that the maximum allowable Ω can be also derived considering the Doppler shifts of the individual frequencies in $\psi(n)$. The requirement for avoiding the overlap of the Doppler shifted frequencies leads to almost the same condition as (11) [4, Eq. (29)].

(a) $\Omega = 45\ °/s$ **(b)** $\Omega = 24\ °/s$ **(c)** $\Omega = 15\ °/s$
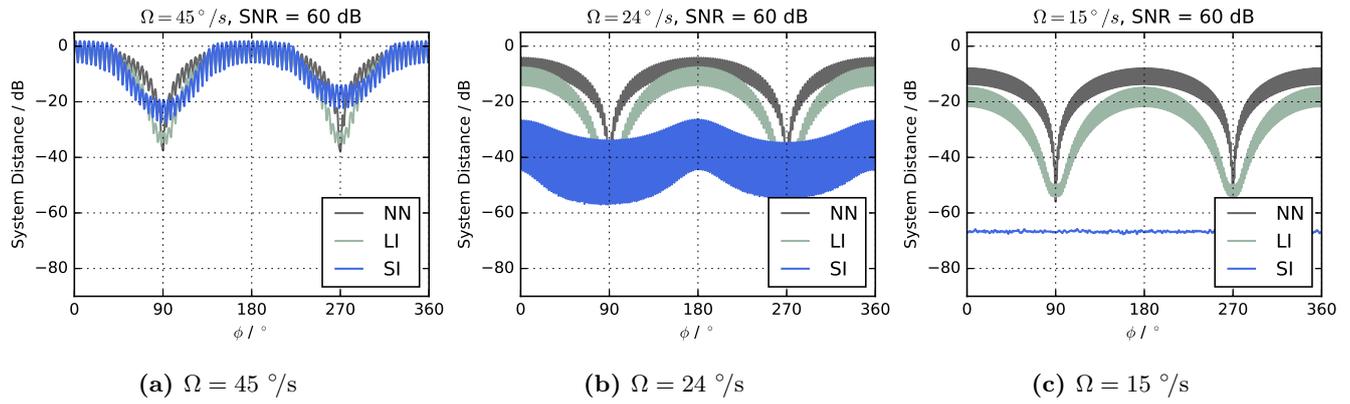
Figure 3: System distances for different angular speeds Ω. Three different interpolation methods are employed (NN: nearest neighbour, LI: linear interpolation, SI: sinc interpolation). The anti-aliasing angular speed according to (11) is $\Omega_0 \approx 24.57\ °/s$. Thus, (a) constitutes an undersampling, (b) a critical sampling, and (c) an oversampling. The SNR at the microphone is 60 dB.

(a) SNR $= 40$ dB   **(b)** SNR $= 60$ dB   **(c)** SNR $= 80$ dB
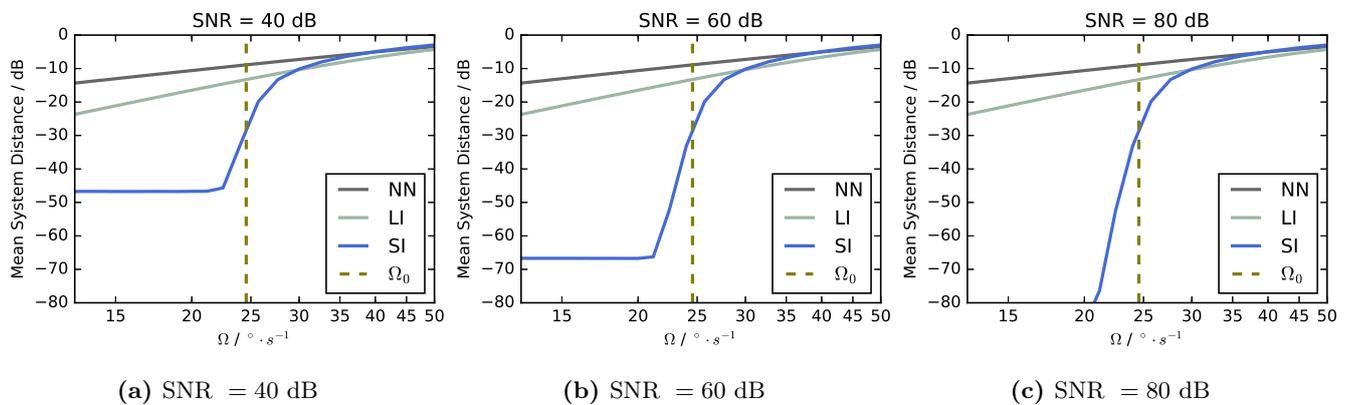
Figure 4: System distances averaged over ϕ for different angular speeds. Three different methods (NN: nearest neighbour, LI: linear interpolation, SI: sinc interpolation) are compared for different SNRs. The anti-aliasing angular speed ($\Omega_0 \approx 24.57\ °/s$) is indicated by dashed vertical lines.

Evaluation

In this section, the continuous measurement of spatial room impulse responses is simulated for the configuration in Fig. 1(a). The plane wave propagates parallel to the xy-plane with an angle of $\phi_{\mathrm{pw}} = 270°$. The impulse response at (r_0, ϕ) thus reads

$$h(\phi, t) = \delta\left(t - \frac{r_0}{c}\cos(\phi - \phi_{\mathrm{pw}})\right), \qquad (13)$$

as shown in Fig. 2. The sampling frequency is $f_s = 16$ kHz and the speed of sound is assumed to $c = 343$ m/s. The plane wave is driven by a perfect sweep with a period of $N = 1600$ corresponding to 0.1 s. The microphone is assumed to be omni-directional. Non-integer delays were implemented with fractional delay filters [16]. According to (11), the anti-aliasing angular speed is $\Omega_0 \approx 24.57\ °/s$. The angular speed and the signal-to-noise ratio at the microphone were varied:

$$\Omega = \frac{360}{7}, \frac{360}{8}, \ldots, \frac{360}{28}\ °/s,$$
$$\mathrm{SNR} = 40, 60, 80\ \mathrm{dB}.$$

The captured signal $s(n)$ is simulated, and the sound field on the circle is reconstructed by using different interpolation methods:

- Nearest neighbour (NN): equivalent to the NLMS algorithm with step size 1 [17]
- Linear interpolation (LI)
- (periodic) Sinc interpolation (SI): equivalent to the approach based on the projection-slice theorem [4]

Finally, the impulse responses are obtained by (4).

The accuracy of the measurement is evaluated in terms of normalized system distance (SD) defined as

$$\mathrm{SD}(\phi) = \left(\frac{\sum_{n=0}^{N-1} |h(\phi, n) - \hat{h}(\phi, n)|^2}{\sum_{n=0}^{N-1} |h(\phi, n)|^2}\right)^{1/2} \qquad (14)$$

where $h(\phi, n)$ denotes the original impulse response and $\hat{h}(\phi, n)$ the estimated impulse response.

In Fig. 3, the performance of the employed methods is shown for different angular speeds ($\Omega = 15, 24, 45\ °/s$). Generally, a slowly moving microphone achieves better performance. In other words, the reconstruction error is reduced by increasing the number of sampling points. If the anti-aliasing condition is not fulfilled, as in Fig. 3(a), there is no benefit of using a higher-order interpolation. The slight improvements around $\phi = 90, 270°$ are at-

tributed to the piecewise constant value of the time delay $\frac{r}{c}\cos(\phi - \phi_{\text{pw}})$ (see Fig. 4), where the system is nearly time-invariant.

In Fig. 3(b), the angular speed of the microphone is slightly below the anti-aliasing speed, $\Omega < \Omega_0$. The sinc interpolation clearly outperforms the other methods. However, the corresponding system distance (blue curve) still exhibits angular dependencies, meaning that the performance depends on the time variance of the system. Therefore, the value of Ω_0 seems to be a little optimistic, due to the crude approximation of the spatial bandwidth of the sound field (9).

The angular speed is further decreased in Fig. 3(c). The sinc interpolation is able to achieve a low system distance that does not depend on ϕ. The effect of the time variability is thus perfectly compensated. The achievable accuracy (minimum system distance) is limited by the SNR.

In Fig. 4, the system distances are averaged over ϕ. For $\Omega > \Omega_0$, the performance is governed by the time variability of the system. While there is no significant difference among the methods, the linear interpolation is slightly better than the others. For $\Omega < \Omega_0$, the sinc interpolation achieves apparently the best performance. The corresponding system distance exhibits a dramatic decrease until it reaches the noise floor. The system distances for NN and LI decrease monotonically irrespective to the relation of Ω and Ω_0.

Conclusion

The continuous measurement of impulse responses is considered as a sound field interpolation problem. The signal captured by the microphone is interpreted in terms of a spatio-temporal sampling of the sound field. The original sound field is interpolated from the sampled values. The impulse responses are then obtained by computing the circular cross-correlation of the estimated sound field and the excitation signal.

By numerical simulations, the influence of the microphone speed on the performance of a continuous measurement is investigated. The microphone speed was varied and the performance was compared for different interpolation methods. The sinc interpolation achieves the lowest system distance, provided that the angular speed of the microphone fulfills the anti-aliasing condition. This validates the interpretation of the continuous measurement as an interpolation problem.

It was pointed out that the anti-aliasing condition introduced by earlier studies is rather optimistic. To assure a better performance, the anti-aliasing condition has to be derived based on a better approximation of the spatial bandwidth. If the anti-aliasing condition cannot be met or if the spatial bandwidth is not known, linear interpolation may be preferred.

Acknowledgements

References

[1] B. Rafaely, *Fundamentals of Spherical Array Processing.* Springer, 2015.

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

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

[4] 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.

[5] C. Antweiler, A. Telle, P. Vary, and G. Enzner, "Perfect Sweep NLMS for Time-variant Acoustic System Identification," in *Proc. of IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, Kyoto, Japan, Mar. 2012.

[6] 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.

[7] F. Katzberg, R. Mazur, M. Maass, P. Koch, and A. Mertins, "Measurement of Sound Fields Using Moving Microphones," in *Proc. of the 42nd IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP)*, New Orleans, USA, Mar. 2017.

[8] 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.

[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, "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.

[11] H. D. Lüke, *Korrelatioknssignale.* Springer, 1992.

[12] 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.

[13] N. Hahn and S. Spors, "Comparison of Continuous Measurement Techniques for Spatial Room Impulse Responses," in *Proc. of the European Signal Processing Conference (EUSIPCO)*, Budapest, Hungary, Aug. 2016.

[14] ——, "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.

[15] M. Abramowitz and I. A. Stegun, *Handbook of Mathematical Functions: with Formulas, Graphs, and Mathematical Tables.* Courier Corporation, 1964.

[16] 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.

[17] 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*, Erlangen, Germany, 2014.