Improved Time-Efficiency in Continuous Measurement of Spatial Room Impulse Responses by Dual-Band Excitation

Nara Hahn and Sascha Spors
Institute of Communications Engineering
University of Rostock, Germany
nara.hahn@uni-rostock.de

Introduction

In a continuous measurement, a large number of spatial impulse responses are obtained by using a moving microphone [1, 2, 3]. The sound source is excited typically by a periodic signal and the signal is captured by the microphone moving on a predefined trajectory. The individual impulse responses are then extracted from the recorded signal. As shown in previous studies [4, 5], the signal captured by the microphone is spatio-temporally sampled from the sound field, and thus the system identification constitutes a spatial interpolation process.

The effective number of spatial sampling points is given as the total duration of the measurement divided by the period of the excitation signal. Since the microphone speed determines the density of the sampling points, it has to be chosen carefully by taking the wavelength at the highest frequency into account. This causes an oversampling of the low frequency components. Also, in order to avoid temporal aliasing, the excitation period has to be longer than the length of the impulse response which is determined by low-frequency components. For high-frequencies, where the impulse response decay much faster, the excitation period is usually too long. Satisfying both conditions for spatial and temporal aliasing leads to a slow microphone speed and the time-efficiency of the continuous measurement method is limited.

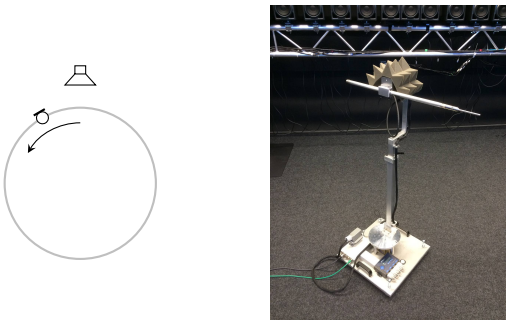
In this paper, the continuous measurement is performed separately for low and high frequency bands to overcome the dilemma. The system is excited by low- and high-pass filtered signals to obtain the respective impulse responses. In low frequencies, the microphone speed is increased without causing spatial aliasing, while in high frequencies, a shorter excitation period is used which allows for a faster movement without losing spatial resolution.

Continuous Measurement

Let assume that the impulse responses are to be measured on a circle as depicted in Fig. 1. An omnidirectional microphone ʊ is moved on a circular trajectory, the polar angle of which is denoted by $\varphi(n)$ with n being the discrete-time index. The signal $s(n)$ captured by the microphone along the trajectory reads

$$s(n) = p\left(\varphi(n), n\right), \tag{1}$$

which constitutes a spatio-temporal sampling of the sound field $p(\phi, n)$. The radius of the circle r is omitted for brevity. Assume that the microphone moves at a



(a) (b)

Figure 1: Impulse response measurement using a moving microphone ʊ. (a) The microphone is moved on a circular trajectory, e.g. by using (b) a motorized arm [6].

constant angular speed Ω in °/s. The sound field is then sampled at equiangular positions,

$$\begin{aligned}\varphi(n) &= \Omega \times \frac{n}{f_s} \\ &= 360 \times \frac{n}{L}, \quad n = 0, 1, \ldots, L-1,\end{aligned} \tag{2}$$

where f_s denotes the sampling frequency. The total number of samples for a 360°-rotation is $L = \frac{360}{\Omega} \times f_s$ which is assumed to be an integer.

For the continuous measurement, a periodic self-orthogonal signal $\psi(n)$ is typically used to excite the system,

$$\sum_{k=0}^{N-1} \psi(k)\psi(k-n) = \sigma_\psi^2 \cdot \delta(n \bmod N), \tag{3}$$

where N denotes the period, $\delta(n)$ the unit impulse function, and σ_ψ^2 the energy of $\psi(n)$ within a period. A signal satisfying (3) is called perfect sequence [7, 8] and has been used often for the identification of linear systems.

Once a sound source is excited by a periodic perfect sequence $\psi(n)$, the reproduced sound field on the circle can be represented with a finite impulse response (FIR) model,

$$p(\phi, n) = \sum_{k=0}^{N-1} \psi(n-k) h(\phi, k), \tag{4}$$

where $h(\phi, k)$ denotes the impulse response at the corresponding position. It is crucial that the excitation period N is longer than any impulse response of the system. Otherwise, the identified impulse response suffers from

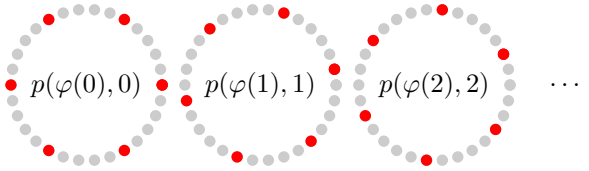


Figure 2: Spatio-temporal sampling of the sound field in a continuous measurement ($L = 30, N = 5$). The captured signal is indicated by gray dots ● at the corresponding positions. The corresponding number of sampling points is $M = 6$. The spatial samples of the sound field is indicated by red dots ● for each $n = 0, 1, 2, \ldots$.

temporal aliasing, meaning that the impulse response is superimposed with time-shifted versions of itself [2].

Considering the orthogonal property (3), the impulse response is given as the circular cross-correlation of the sound field and the excitation signal,

$$h(\phi, n) = \sum_{k=0}^{N-1} p(\phi, k)\psi(k - n). \tag{5}$$

Note that (5) requires N consecutive samples of $p(\phi, n)$, therefore, the sound field at a given position has to be recorded at least as long as the impulse response. However, in a continuous measurement where the microphone constantly moves, only one sample is available at each position.

As shown in (1), the captured signal $s(n)$ corresponds to the n-th sample of the sound field $p(\phi, n)$ at position $\varphi(n)$. Since $p(\phi, n)$ is periodic with the same period as $\psi(n)$, every other N-th samples of $s(n)$ represent the sound field at time n,

$$\begin{aligned} s(n) &= p\left(\varphi(n), n\right) \\ s(n+N) &= p\left(\varphi(n+N), n\right) \\ s(n+2N) &= p\left(\varphi(n+2N), n\right) \\ &\vdots \end{aligned} \tag{6}$$

In other words, the sound field at a given time is spatially sampled at different positions, $\varphi(n+\mu N)$ for $\mu \in \mathbb{Z}$. This is illustrated in Fig. 2 with an example where $L = 30$ and $N = 5$. The gray dots ● indicate the captured signal $s(n)$ along the circle. For a given $n \in [0, N-1]$, the sound field $p(\phi, n)$ is captured at $M = \frac{L}{N}$ equiangular positions,

$$\tfrac{360n}{L}, \tfrac{360(n+N)}{L}, \tfrac{360(n+2N)}{L}, \ldots \tfrac{360(n+(M-1)N)}{L}, \tag{7}$$

which are indicated by the red dots ●. Note also that the distribution of the sampling points is circularly shifted by $\frac{360n}{L}$.

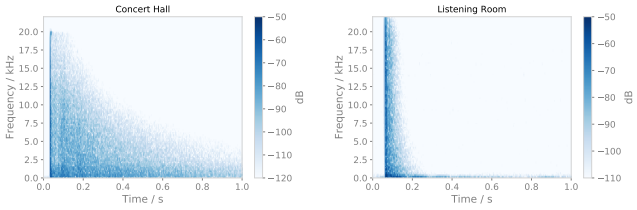
As proposed in previous studies [4, 5], the original sound field on the circle can be interpolated from the spatial samples (6),

$$\hat{p}(\phi, n) = \mathcal{L}\left\{s(n + \mu N), \mu \in \mathcal{Z}\right\}, \tag{8}$$

where $\hat{p}(\phi, n)$ denotes the estimate of the sound field and $\mathcal{L}\{\cdot\}$ the interpolation operator. Once the sound field at a target position is interpolated for $n = 0, 1, \ldots, N-1$, the corresponding impulse response $\hat{h}(\phi, n)$ can be computed by (5). Obviously, the accuracy of the impulse responses heavily depends on the number of sampling points M which is inversely proportional to Ω. Based on the approximated spatial bandwidth in the circular harmonics domain $\lceil \frac{2\pi f}{c} R \rceil$ [9, Sec. 4.2], an anti-aliasing condition for Ω can be derived [5],

$$\begin{aligned} \Omega &\leq \frac{360 f_s}{N} \frac{1}{2\lceil \frac{2\pi f}{c} R \rceil + 1} \\ &\leq \frac{360}{T} \frac{1}{2\lceil \frac{2\pi f}{c} R \rceil + 1} \end{aligned} \tag{9}$$

where T denotes the length of the impulse response in seconds, and c the speed of sound.



(a) (b)

Figure 3: Spectrograms of (a) a concert hall [10] and (b) a listening room.

Dual-Band Excitation

Considering the physical nature of room impulse responses, the anti-aliasing condition discussed in the previous section leads to an inefficient sampling. See Fig. 3 for the spectrograms of two typical room impulse responses, concert hall (left) and listening room (right). The frequency-dependent decay is mainly due to the higher wall/air absorption at high frequencies and the room modes (standing waves) at low frequencies [11, Ch. 3 and 6]. Note from (9) that the lower limit of Ω is determined by the highest frequency of interest and the length of the impulse response T which is dominated by low frequencies. If only high frequencies are concerned, the excitation period can be reduced without causing temporal aliasing, which in turn allows to increase the angular speed. At low frequencies, the excitation period is chosen properly but the number of spatial sampling points is unnecessarily high, thus oversampled. The sampling points can be reduced by increasing Ω as far as spatial aliasing does not occur.

The improved spatial sampling by dual-band excitation is illustrated in Fig. 4. Compared to the full-band case (left), fewer sampling points ● are required at low frequencies (middle) whereas the excitation period ●●●●● is unchanged. For high frequencies (right), still the same number of sampling points ● are needed but the excitation period ●●● is reduced. In both cases, the speed of the microphone is increased, indicated by the larger spacing between the dots compared to the full-band case.

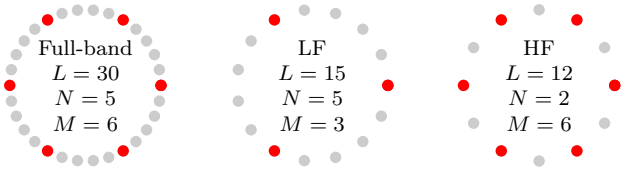


Figure 4: Spatial sampling of the sound field in full-band (left) and dual-band (center, right) excitation. The captured signal is indicated by ●, and the sound field $p(\phi, n)$ at time $n = 0$ by ●.

In order to perform the continuous measurement separately for low and high frequency bands, linear-phase low- and high-pass filters are used. The filters must satisfy

$$g_{\mathrm{LPF}}(n) + g_{\mathrm{HPF}}(n) = \delta(n - \nu), \tag{10}$$

where $g_{\mathrm{LPF}}(n)$ and $g_{\mathrm{HPF}}(n)$ denote the respective FIR coefficients, and ν the group delay of the filters.

The excitation period for high frequencies N_{HF} is determined by the length of the impulse response in the frequency range $f \geq f_{\mathrm{HPF}}^{\mathrm{stop}}$, with $f_{\mathrm{HPF}}^{\mathrm{stop}}$ denoting the stop-band frequency of the high-pass filter. At low frequencies, the excitation period is unchanged $N_{\mathrm{LF}} = N$. By using the filters (10), low- and high-pass filtered perfect sequences, $\psi_{\mathrm{LF}}(n)$ and $\psi_{\mathrm{HF}}(n)$, are generated with period N_{LF} and N_{HF}, respectively. Since the excitation signals are periodic, the filtering is performed by circular convolution. A circular shift of $-\nu$ samples is applied to compensate the group delay.

Based on the parameters for dual-band excitation, a modified anti-aliasing condition can be derived from (9),

$$\Omega_{\mathrm{LF}} \leq \frac{360 f_s}{N_{\mathrm{LF}}} \frac{1}{2\lceil \frac{2\pi f_{\mathrm{LPF}}^{\mathrm{stop}}}{c} R \rceil + 1} \tag{11}$$

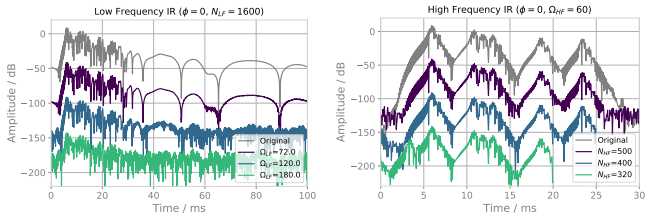
$$\Omega_{\mathrm{HF}} \leq \frac{360 f_s}{N_{\mathrm{HF}}} \frac{1}{2\lceil \frac{\pi f_s}{c} R \rceil + 1}, \tag{12}$$

where f is replaced with the stop-band frequency of the low-pass filter $f_{\mathrm{LPF}}^{\mathrm{stop}}$ and N with N_{HF}.

The time-efficiency of the continuous measurement is improved if the duration for dual-band excitation $\frac{360}{\Omega_{\mathrm{LF}}} + \frac{360}{\Omega_{\mathrm{HF}}}$ is reduced compared to the full-band excitation $\frac{360}{\Omega}$. This can be achieved if the low-frequency response is much longer than the high-frequency response, i.e. the spectrogram has a concave-like shape ◣. This is particularly the case for small rooms where the high frequency response is very short and the late reverberation is dominated by low-frequency room modes ◺, as in Fig. 3(b). Measurement of anechoic impulse responses like head-related impulse responses (HRIRs) is unlikely to benefit from dual-band excitation, because the response at low frequencies is already very short.

Evaluation

In this section, the continuous measurement is simulated for a rectangular room ($5.8 \times 5.0 \times 3.0$ m^3) by



(a) LF (b) HF

Figure 5: Continuously measured impulse response ($\phi = 0$) for different angular speed and excitation period. 50 dB offsets are added between the impulse responses.

using a second-order image source model. Frequency-dependent reflection coefficients are applied in an octave-band filter bank in order to simulate the different decay rate. An omni-directional microphone is moved on a circle of radius $r = 0.2$ m with varying angular speeds $\Omega = 30, \ldots, 180$ °/s. For dual-band excitation, 121-tap linear phase filters are designed by using the Remez algorithm (`scipy.signal.remez`), where the stop-band gain is set to -100 dB. The frequency range is evenly divided by the filters. For low frequencies, the excitation period is fixed to $N_{\mathrm{LF}} = 1600$, whereas for high frequencies, the period is varied, $N_{\mathrm{HF}} = 320, \ldots, 1600$. The captured signal is simulated by fractional delay filters [12, 13]. Gaussian noise (SNR -60 dB) is added to the captured signal. For system identification, the original sound field is computed from the captured signal by using the periodic sinc interpolation. The individual impulse responses are then computed with (5). Finally, the low- and high-pass filtered impulse responses are added together. The sampling frequency is set to $f_s = 16$ kHz and the speed of sound to $c = 343$ m/s.

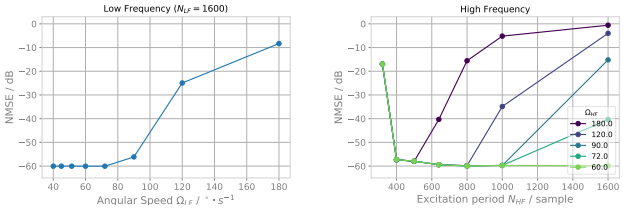
The low-pass filtered impulse responses for $\phi = 0$ are shown in Figure 5(a). Vertical offsets of 50 dB are added between the impulse responses for better visualization. Clearly, the accuracy increases for slower microphone movement which constitutes more sampling points. The noisy components for $\Omega_{\mathrm{LF}} = 120$ and $\Omega_{\mathrm{LF}} = 180$ are attributed to spatial aliasing [14].

Figure 5(b) shows the high-pass filtered impulse responses. Here, the angular speed $\Omega_{\mathrm{HF}} = 60$ °/s is fixed and the excitation period N_{HF} is varied. For $N_{\mathrm{HF}} = 400$ and $N_{\mathrm{HF}} = 320$, the impulse responses suffer from temporal aliasing where the later part of the impulse response is truncated and superimposed to the early part.

The accuracy of the individual impulse responses are quantified by the normalized mean square error (NMSE),

$$\mathcal{E}(\phi) = \frac{\|h(\phi, n) - \hat{h}(\phi, n)\|}{\|h(\phi, n)\|}, \tag{13}$$

where $h(\phi, n)$ and $\hat{h}(\phi, n)$ denote the original and estimated impulse responses, respectively. The norm $\|\cdot\|$ is defined as $\|x(n)\| \equiv \sqrt{\sum_{n=0}^{N-1} |x(n)|^2}$. In order to evaluate the overall performance, the NMSEs for 180 equiangular positions $\phi = 0, 2, \ldots, 358°$ are averaged ($\bar{\mathcal{E}}$). In



(a) (b)

Figure 6: Overall performance (averaged NMSE) of (a) low- and (b) high-pass filtered impulse responses.

Fig. 6(a), the low-frequency performance increases ($\bar{\mathcal{E}}$ decreases) for slower movements until $\bar{\mathcal{E}}$ reaches the noise floor −60 dB. The high-frequency performance is shown in Fig. 6(b) for different combinations of $(\Omega_{\mathrm{HF}}, N_{\mathrm{HF}})$. For each Ω_{HF}, the NMSE decreases as N_{HF} gets smaller which corresponds to more sampling points. One exception is $N_{\mathrm{HF}} = 320$ (leftmost dot) where the NMSE increases abruptly because of temporal aliasing. In terms of measurement time, it is more efficient to choose a higher angular speed and shorter excitation period as long as temporal aliasing does not occur, e.g. $(\Omega_{\mathrm{HF}} = 180, N_{\mathrm{HF}} = 400)$.

Notice that the NMSE is very similar ($\bar{\mathcal{E}} \approx -15$ dB) for

$$(\Omega_{\mathrm{HF}} = 180, N_{\mathrm{HF}} = 800) \text{ and } (\Omega_{\mathrm{HF}} = 90, N_{\mathrm{HF}} = 1600),$$

both of which have the same number of sampling points $M = 40$. This is also the case for

$$(\Omega_{\mathrm{HF}} = 180, N_{\mathrm{HF}} = 640) \text{ and } (\Omega_{\mathrm{HF}} = 72, N_{\mathrm{HF}} = 1600),$$

where $\bar{\mathcal{E}} \approx -40$ dB and $M = 50$. This validates the concept of treating the continuous measurement as a sampling/reconstruction problem.

Conclusion

Continuous measurement techniques have often been claimed to be faster than conventional static measurement. In reverberant spaces, however, the reduction of measurement time is marginal if the anti-aliasing condition is taken into account. The small wavelength at high frequencies leads to an oversampling at low frequencies, whereas the long reverberation at low frequencies forces to use an excitation signal which is too long for high frequencies. In this paper, this problem was tackled by using dual-band excitation where the measurement is performed separately for high and low frequencies. This enabled a more efficient spatial sampling of the sound field at both frequency bands, thereby reducing the measurement time.

The perceptual properties of the continuously measured impulse responses are still under investigation. The influence of spatial and temporal aliasing on the perceived quality needs to be further evaluated. Additional speed-up of the continuous measurement is expected by allowing imperceptible errors.

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

References

[1] T. Ajdler, L. Sbaiz, and M. Vetterli, "Dynamic Measurement of Room Impulse Responses Using a Moving Microphone," *J. Acoust. Soc. Am. (JASA)*, vol. 122, no. 3, pp. 1636–1645, 2007.

[2] C. Antweiler and G. Enzner, "Perfect Sequence LMS for Rapid Acquisition of Continuous-azimuth Head Related Impulse Responses," in *Proc. IEEE Workshop Appl. Signal Process. Audio Acoust. (WASPAA)*, New Paltz, NY, USA, Oct. 2009.

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

[4] N. Hahn and S. Spors, "Identification of Dynamic Acoustic Systems by Orthogonal Expansion of Time-variant Impulse Responses," in *Proc. Int. Sym. Commun. Control Signal Process. (ISCCSP)*, Athens, Greek, May 2014.

[5] ——, "Continuous Measurement of Impulse Responses on a Circle Using a Uniformly Moving Microphone," in *Proc. Eur. Signal Process. Conf. (EUSIPCO)*, Nice, France, Aug. 2015.

[6] B. Bernschütz, C. Pörschmann, S. Spors, and S. Weinzierl, "Entwurf und Aufbau eines variablen sphärischen Mikrofonarrays für Forschungsanwendungen in Raumakustik und virtual Audio," in *Proc. German Annu. Conf. Acoust. (DAGA)*, Berlin, Germany, Mar. 2010.

[7] 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.

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

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

[10] H. Lee and C. Millns, "Microphone Array Impulse Response (MAIR) Library for Spatial Audio Research," in *143rd Audio Eng. Soc. Conv.*, Oct. 2017.

[11] H. Kuttruff, *Room Acoustics*. Crc Press, 2016.

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

[13] J.-P. Berrut and L. N. Trefethen, "Barycentric Lagrange Interpolation," *SIAM Review*, vol. 46, no. 3, pp. 501–517, 2004.

[14] N. Hahn and S. Spors, "Spatial Aliasing in Continuous Measurement of Spatial Room Impulse Responses," in *Proc. German Annu. Conf. Acoust. (DAGA)*, Kiel, Germany, Mar. 2017.