DOI: 10.61782/fa.2025.0316

FORUM ACUSTICUM EURONOISE 2025

DYNAMIC ROOM IMPULSE RESPONSE MEASUREMENTS WITH A ROBOTIC ARM

Samuel A. Verburg[1*] **Odei Diaz-Gaztelumendi**[1] **Yves Pene**[2]
Efren Fernandez-Grande[3]
[1] Technical University of Denmark (DTU), Denmark
[2] L-Acoustics, France
[3] Universidad Politecnica de Madrid (UPM), Madrid, Spain

ABSTRACT

Measuring room impulse responses (RIRs) over space is central to many applications, including sound field control, room compensation, and loudspeaker system calibration. Typically, RIRs are measured with a single microphone or a microphone array positioned at static, known locations. To extend the measurement area, the microphones must be sequentially re-positioned to measure RIRs at new locations, a process that can be time-consuming and experimentally cumbersome. In this work, we investigate dynamic measurements using a robotic arm. In a dynamic measurement, a microphone is continuously moved along a known trajectory as it captures a continuous excitation signal. Dynamic measurements can therefore cover large areas of space while being potentially easier to deploy. We describe a processing method based on an elementary wave expansion to estimate the RIRs over space from dynamic measurements. The use of a robotic arm allows for accurately tracking the microphone trajectory and provides a way to obtain ground truth measurements to evaluate the estimations. Results from experimental data show successful RIR estimations, particularly in the low-frequency range (20-500 Hz).

Keywords: *room impulse response, dynamic measurement, robotic arm, microphone array*

**Corresponding author*: *saveri@dtu.dk.*

1. INTRODUCTION

Characterizing the sound field inside rooms is essential in many acoustics and audio engineering applications, such as sound field control [1], room response equalization [2], loudspeaker system calibration [3], 6DoF immersive audio [4], and sound field analysis [5]. The sound field is normally characterized by measuring room impulse responses (RIRs) between a loudspeaker and a set of microphones. Typically, the measurements are performed sequentially by acquiring a set of RIRs at a time and then re-positioning the microphones for the next measurement [6]. However, sequential measurement require considerable experimental effort, specially when sampling large areas. Ultimately, positioning many microphones in a sequential manner becomes time consuming, impractical and costly.

Dynamic measurements offer an attractive alternative to sequential static measurments. In a dynamic measurement, a microphone (or a compact array) is continuously moved within the measurement area as it records the sound pressure. The recorded signals are then processed to estimate the sound field over the region of interest. Compared to stationary measurement, dynamic measurements can potentially reduce the acquisition time and experimental effort required to characterize a sound field.

Several dynamic measurements methods have been developed, including techniques for uniform [7] and non-uniform [8, 9] microphone trajectories, methods to measure binaural RIRs [10], head related transfer functions [11], and RIRs from multiple sources simultaneously [12], models that account for the Doppler shift introduced by the microphone movement [13] as well as models based on spherical harmonic expansions [14], compressed sensing [15], and Bayesian inference [16]. While there is a rich literature, dynamic measurements methods have mostly been tested in simulation studies, with the exception of measurements along circular trajectories [7,10,17]. Therefore, there is a lack of experimental validation with real data in the existing literature.

In this study, dynamic measurements with a robotic arm are conducted. The use of a robotic arm makes it possible to define arbitrary trajectories and accurately retrieve the microphone position during the measurement. The robotic arm is also used to perform conventional static RIR measurements, providing a ground truth to evaluate the estimations. A plane wave expansion is used for modeling the sound field. Plane wave expansions are a robust and effective model for representing sound fields in enclosures, particularly in at low and mid frequencies [6]. This work presents one of the first experimental studies of dynamic RIR measurements along arbitrary trajectories with real data.

2. THEORY

The dynamic sampling of RIRs is briefly described in this section since more detailed derivations can be found elsewhere [14]. To dynamically measure RIRs, a static loudspeaker is driven with an excitation signal (such as pseudo-random noise) and the resulting sound pressure is recorded with a microphone moving across the room. The excitation signal, $s(n)$, is defined at times $t_n = n/f_s$ for $n = 0, \dots, N-1$, where f_s is the sampling frequency. The position of the microphone at t_n is denoted $\mathbf{r}_n$. It is assumed that the microphone trajectory, i.e., $\mathbf{r}_0, \dots, \mathbf{r}_{N-1}$, is known. The acoustic response of the room at position $\mathbf{r}$ is described by the RIR, $h(\mathbf{r}, n)$, for $n = 0, \dots, L-1$. Here it is assumed that the RIR is negligibly small for $n > L-1$. The signal measured by the microphone, $y(\mathbf{r}_n, n)$, is given by the convolution of the excitation signal with the RIR [14],

$$y(\mathbf{r}_n, n) = \sum_{m=0}^{L-1} s(n-m) h(\mathbf{r}_n, m) + e(n), \qquad (1)$$

where $e(n)$ is measurement noise. The RIR can be written as the inverse Fourier transform of the frequency response of the room, H,

$$h(\mathbf{r}, n) = \frac{1}{L} \sum_{l=0}^{L-1} H(\mathbf{r}, l) \mathrm{e}^{\mathrm{j}2\pi \frac{l}{L} n}. \qquad (2)$$

Combining equations (1) and (2) results in

$$\begin{aligned} y(\mathbf{r}_n, n) &= \sum_{m=0}^{L-1} s(n-m) \frac{1}{L} \sum_{l=0}^{L-1} H(\mathbf{r}_n, l) \mathrm{e}^{\mathrm{j}2\pi \frac{l}{L} m} + e(n) \\ &= \sum_{l=0}^{L-1} \left[\frac{1}{L} \sum_{m=0}^{L-1} s(n-m) \mathrm{e}^{\mathrm{j}2\pi \frac{l}{L} m} \right] H(\mathbf{r}_n, l) + e(n) \\ &= \sum_{l=0}^{L-1} S_{n,l} H(\mathbf{r}_n, l) + e(n), \end{aligned} \qquad (3)$$

where the term in brackets in the second line of equation (3) is the transpose of the matrix resulting from taking the centered short-time Fourier transform of the excitation signal with a window length of L and overlap of $L-1$, denoted $S_{n,l}$.

The acoustic response of the room is modeled as the superposition of propagating plane waves [6]

$$H(\mathbf{r}, l) \approx \sum_{q=0}^{Q-1} x_{q,l} \mathrm{e}^{\mathrm{j}k_l \mathbf{u}_q \cdot \mathbf{r}}, \qquad (4)$$

where Q is the number of waves considered in the expansion, $\mathbf{u}_q$ is a unitary vector pointing in the direction of the qth wave, k_l is the wavenumber at the sampled frequencies, i.e., $k_l = 2\pi l f_s/(Lc)$, and c is the speed of sound. The unknown coefficient $x_{q,l}$ is the amplitude of the wave with direction $\mathbf{u}_q$ and wavenumber k_l. Combining equations (3) and (4), the measured pressure can be expressed as

$$y(\mathbf{r}_n, n) = \sum_{l=0}^{L-1} S_{n,l} \sum_{q=0}^{Q-1} x_{q,l} \mathrm{e}^{\mathrm{j}k_l \mathbf{u}_q \cdot \mathbf{r}_n} + e(n). \qquad (5)$$

or, algebraically,

$$\mathbf{y} = \mathbf{A}\mathbf{x} + \mathbf{e} \qquad (6)$$

where $\mathbf{y} \in \mathbb{R}^N$ is the recorded pressure expressed as a vector. The matrix $\mathbf{A} \in \mathbb{C}^{N \times QL}$ is defined as $\mathbf{A} \equiv [\mathbf{S}_0 \odot \mathbf{A}_0, \dots, \mathbf{S}_{L-1} \odot \mathbf{A}_{L-1}]$, where $\mathbf{S}_l \in \mathbb{C}^{N \times N}$ is a diagonal matrix with entries $S_{0,l}, \dots, S_{N-1,l}$ in its diagonal, and $\mathbf{A}_l \in \mathbb{C}^{N \times Q}$ is the matrix of plane waves with elements $a_{l(n,q)} = \mathrm{e}^{\mathrm{j}k_l \mathbf{u}_q \cdot \mathbf{r}_n}$. Therefore, the entire trajectory of the moving microphone, i.e., $\mathbf{r}_0, \dots, \mathbf{r}_{N-1}$, is contained in the matrix $\mathbf{A}$ through the plane wave matrix elements, $\mathrm{e}^{\mathrm{j}k_l \mathbf{u}_q \cdot \mathbf{r}_n}$. The vector $\mathbf{x} \in \mathbb{C}^{QL}$ is defined as $\mathbf{x} \equiv [\mathbf{x}_0^T, \dots, \mathbf{x}_{L-1}^T]^T$, where $\mathbf{x}_l \in \mathbb{C}^Q$ is the vector of wave coefficients $[x_{0,l}, \dots, x_{Q-1,l}]^T$. The vector $\mathbf{e} \in \mathbb{R}^N$ contains the measurement noise.

Equation (6) represents a linear system with N equations (one for each time sample) and QL unknowns (one for each plane wave in the expansion). Once the system is solved, the 'static' frequency and impulse responses at any position $\mathbf{r}$ in the measurement area can be computed from the estimated wave coefficients using equations (4) and (2), respectively. In addition, since the plane waves used to represent the sound field are global functions, the pressure can also be extrapolated outside the measurement area to some extent.

2.1 Solving the linear system

Solving the system in equation (6) can be challenging. Typically, the characterization of sound fields from static measurements is done independently for each frequency bin. Therefore the matrices involved are of small to medium size (at most $M \times Q$ where M is the number of static positions), making it possible to apply matrix decomposition-based methods to solve the system of equations. In the dynamic case, however, since all the samples along the trajectory are used to estimate the wave coefficients, the problem cannot be split as easily, and the matrix involved is typically very large. For example, for a one minute recording at $f_s = 2048$ samples/second, 0.5 seconds impulse response and 50 waves in the expansion, the size of the matrix $\mathbf{A}$ is 122880×51150 which is clearly not amenable to decomposition or even to store it in memory. An iterative solver, LSQR [18], is used in this study for approximating the solution to the least squares problem

$$\tilde{\mathbf{x}} = \arg \min_{\mathbf{x} \in \mathbb{C}^{QL}} \|\mathbf{A}\mathbf{x} - \mathbf{y}\|. \tag{7}$$

The LSQR solver is based on the conjugate gradient method, and being iterative approximates the solution in a finite number of steps. Inside the LSQR algorithm the matrix $\mathbf{A}$ only appears in matrix-vector products, therefore it is not necessary to store the full matrix in memory. The regularization applied is determined by the number of iterations. If run for too many iterations, there is the risk of fitting the noise, while if the number of iterations is not enough the algorithm might not have converged.

3. EXPERIMENT

The dynamic sampling of RIRs is tested in an experimental study. The setup is shown in figure 1. Measurements were performed in a IEC standardized listening room with average reverberation time of 0.4 s in the frequency range

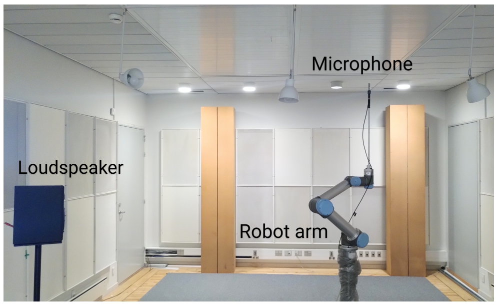

Figure 1. Experimental setup.

of interest and dimensions $7.5 \times 4.74 \times 2.8$ m^3. A Dynaudio BM6 loudspeaker was placed in the room, along with a Universal Robots UR5 robotic arm. A measurement microphone was attach to a 0.7 m metal rod, which in turn was mounted on the robotic arm. A PC and sound card placed outside the room were used to generate the excitation signal, drive the loudspeaker, record the sound pressure, and control the robotic arm. The pressure was recorded at $f_s = 2048$ samples/second during approximately 63 seconds. The excitation signal was a repeated maximum length sequence [19] of length $L = 1023$ samples. The number of plane waves per frequency was $Q = 50$, and the directions were uniformly sampled on the surface of a sphere with unit radius.

The acquired data was processed in MATLAB. The native implementation of the LSQR algorithm was used to approximate the solution to (7). The algorithm was run over one hundred iterations. The total processing time was approximately two hours on a HP Z4-G4 workstation.

The microphone trajectory is shown in figure 2. The trajectory defined was a Lissajous curve. The curve avoids the central part, where the base of the arm is placed, since the robot arm cannot reach this area. To accurately track the microphone position, the robot controller was queried at a 100 Hz rate while performing the dynamic sampling. The positioning data provided by the controller was then upsampled using linear interpolation and used for the RIR reconstruction as described in equations (1) to (6). The position and pressure data were synchronized by a trigger that signaled when the robot arm started to move.

Extensive validation data was additionally acquired with the same setup. In this case the robot was programmed to move the microphone sequentially at $M = 880$ positions defined on a plane that crossed the trajectory of the dynamic measurement (see gray dots in figure 2). The RIR at each of these locations was measured using

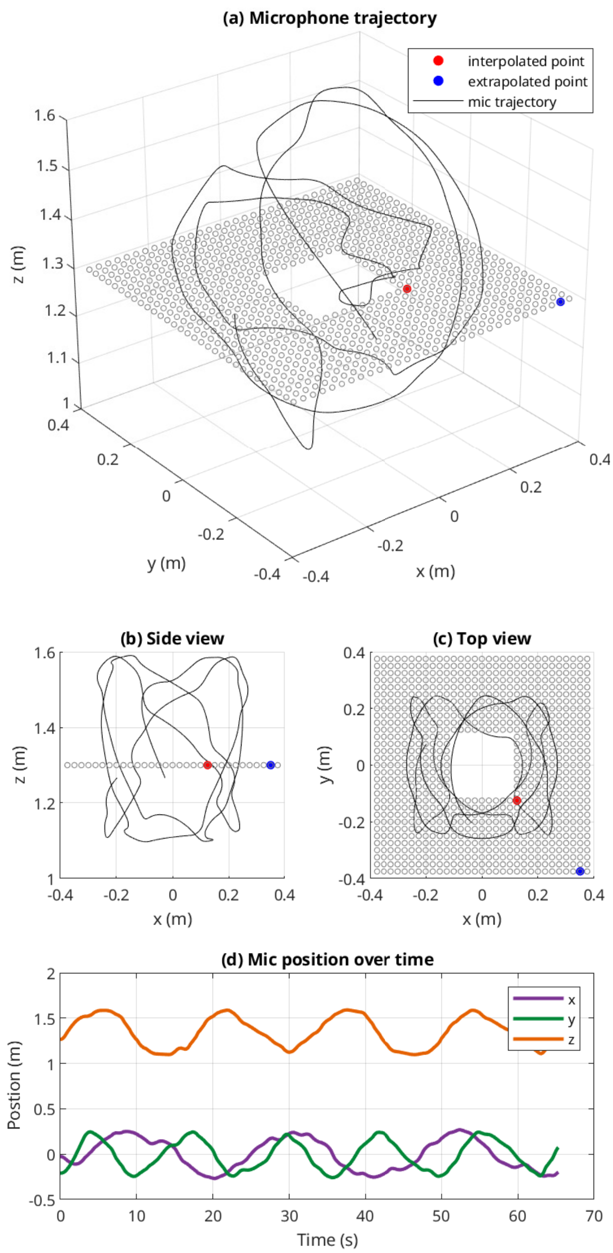

Figure 2. (a) Microphone trajectory and positions of reference measurements. (b) Side view. (c) Top view. (d) x, y, z coordinates of the microphone position as a function of time.

a 5 seconds logarithmic sine sweep, with a total measurement time of approximately 2 hours for the entire plane.

Impulse responses reconstructed from the dynamic measurements as well as the corresponding reference ones are shown in figure 3. Figure 3(a) shows the response at an interpolated position (marked in red in figure 2), i.e., a position within the measurement volume, but not directly on the trajectory of the moving microphone. The reconstruction is very accurate along the entire RIR both in terms of amplitude and phase, although the high frequencies present in the direct sound are slightly underestimated. Figure 3(b) shows the response at an extrapolated position (marked in blue in figure 2). The reconstruction accuracy degrades compared with the interpolated point, as this position is relatively far from the trajectory. The direct sound and early reflections are distinguishable and the phase is recovered to some extent, yet their amplitude is overestimated. In addition, some ringing at the beginning of the impulse response is observable, probably due to the bandlimitation of the reconstruction.

The reconstruction error in the reference plane is quantitatively analyzed in terms of the mean squared error

$$10 \log \frac{1}{M} \sum_{m=0}^{M-1} |H(\mathbf{r}_m, l) - \tilde{H}(\mathbf{r}_m, l)|^2, \tag{8}$$

and the spatial correlation

$$\frac{\left|\sum_{m=0}^{M-1} H^*(\mathbf{r}_m, l)\tilde{H}(\mathbf{r}_m, l)\right|}{\left(\sum_{m=0}^{M-1} H^*(\mathbf{r}_m, l)H(\mathbf{r}_m, l)\right)\left(\sum_{m=0}^{M-1} \tilde{H}^*(\mathbf{r}_m, l)\tilde{H}(\mathbf{r}_m, l)\right)}, \tag{9}$$

where $\tilde{H}(\mathbf{r}_m, l)$ and $H(\mathbf{r}_m, l)$ are the reconstructed and reference frequency responses, respectively, at the reference plane positions $\mathbf{r}_m$ (gray dots in figure 2). The spatial correlation takes values between one (indicating perfect correlation between reference and reconstruction) and zero.

Figure 4(a) and (b) show the mean squared error and the spatial correlation as a function of frequency, respectively. At frequencies below 500 Hz, the mean squared error is below -10 dB and the spatial correlation is above 0.8, indicating very accurate reconstructions. As the frequency increases the spatial distribution of sound pressure becomes more complex, which results in a higher mean squared error and lower spatial correlation.

The sound pressure over space is further analyzed in figure 5. The figure shows the magnitude of the frequency response on the reference plane at 200, 400, 500 and 750

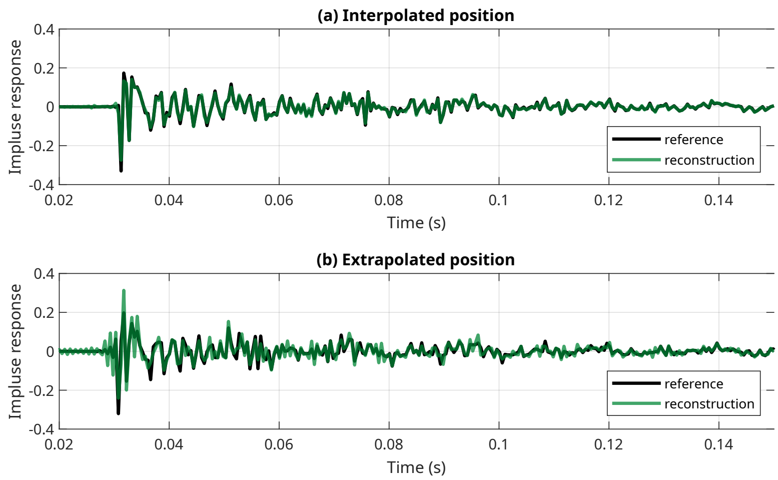

Figure 3. Impulse response. (a) For a point within the trajectory (red dot in figure 2) (b) For a point outside the trajectory (blue dot in figure 2).

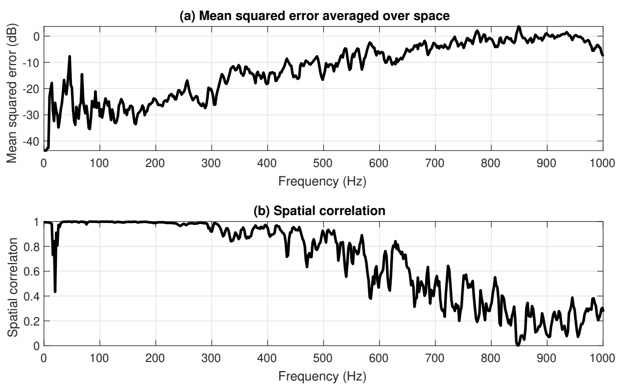

Figure 4. Error of the reconstructed responses at the reference plane as a function of frequency (a) Mean squared error. (b) Spatial correlation.

Hz. For 500 Hz and below the spatial distribution of sound pressure is well recovered, including the areas of small pressure that result from mode interference. Even if at 750 Hz the reconstruction degrades, the pressure distribution is still plausible, specially closer to the area covered by the trajectory.

4. DISCUSSION

The results show that accurate RIRs reconstructions can be obtained from real dynamic measurement data. In particular, the areas close the microphone trajectory are very well recovered and, for the parameters in our experiment, frequencies below 500 Hz are correctly estimated. Nonetheless, the frequency range and reconstruction area could be extended by increasing the length of the trajectory and measurement time. The measurement time of our experiment was approximately one minute to measure on an area of approximately 0.5 m^3, which is very short compared with the total time that this type of RIR measurements normally entail. The reduction in acquisition time offered by dynamic measurements becomes even more interesting when measuring over very large areas, e.g., across several hundred meters. The accuracy of the results could also be improved by optimizing the microphone trajectory and speed.

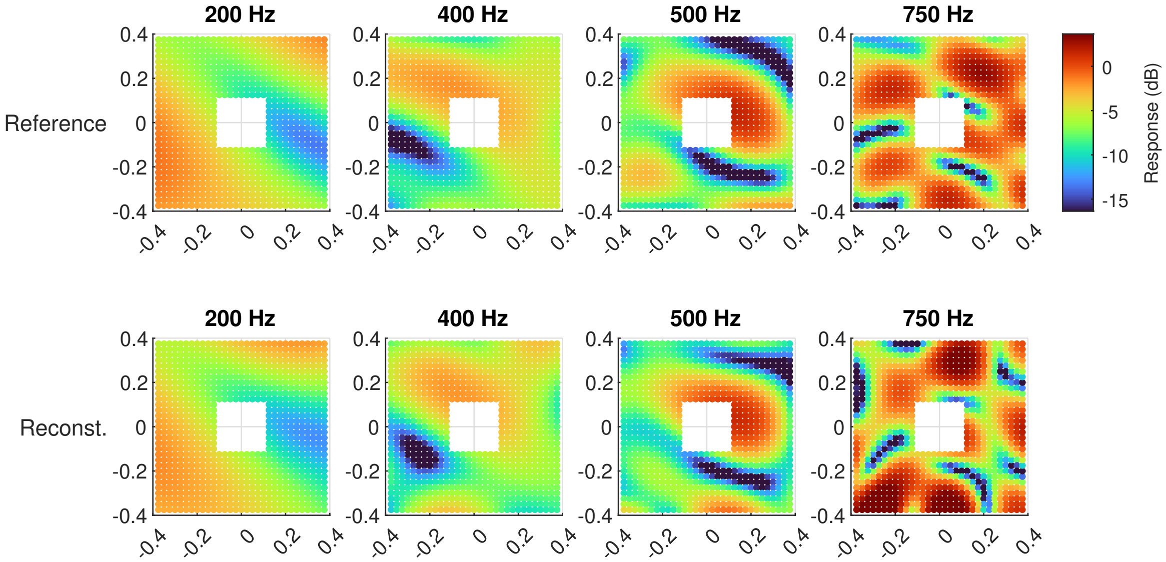

Figure 5. Magnitude of the frequency response on the reference plane (gray dots in figure 2) at 200, 400, 500 and 750 Hz. Top row: reference. Bottom row: reconstruction.

5. CONCLUSION

This work presents one of the first experimental studies of dynamic RIR measurements with real data and in a real room. A robotic arm was used to continuously move a microphone while acquiring pressure data. The experimental results demonstrate that dynamic measurements can achieve accurate RIR estimations over space, particularly in the low frequency range. These findings indicate that dynamic measurements can significantly reduce the experimental effort compared to traditional static methods, with potential applications in sound field control, room compensation, and loudspeaker system calibration.

The use of a robotic arm not only made it possible to track the microphone position accurately but also provided ground truth measurements for evaluating the accuracy of the reconstructions. In conclusion, dynamic RIR measurements with a robotic arm offer a promising alternative to conventional methods, providing efficient and accurate characterization of sound fields.

6. REFERENCES

[1] P. A. Nelson, "Active control of acoustic fields and the reproduction of sound," *Journal of Sound and Vibration*, vol. 177, no. 4, pp. 447–477, 1994.

[2] S. Cecchi, A. Carini, and S. Spors, "Room response equalization—a review," *Applied Sciences*, vol. 8, no. 1, p. 16, 2017.

[3] S. J. Elliott and P. A. Nelson, "Multiple-point equalization in a room using adaptive digital filters," *AES: Journal of the Audio Engineering Society*, vol. 37, no. 11, pp. 899–907, 1989.

[4] J. G. Tylka and E. Y. Choueiri, "Soundfield navigation using an array of higher-order ambisonics microphones," in *2016 AES International Conference on Audio for Virtual and Augmented Reality*, 2016.

[5] I. B. Witew, M. Vorländer, and N. Xiang, "Sampling the sound field in auditoria using large natural-scale array measurements," *The Journal of the Acoustical Society of America*, vol. 141, no. 3, pp. EL300–EL306, 2017.

[6] S. A. Verburg and E. Fernandez-Grande, "Reconstruction of the sound field in a room using compressive sensing," *The Journal of the Acoustical Society of America*, vol. 143, no. 6, pp. 3770–3779, 2018.

[7] 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*, vol. 122, no. 3, pp. 1636–1645, 2007.

[8] F. Katzberg, R. Mazur, M. Maass, P. Koch, and A. Mertins, "Sound-field measurement with moving microphones," *The Journal of the Acoustical society of America*, vol. 141, no. 5, pp. 3220–3235, 2017.

[9] N. Hahn and S. Spors, "Continuous measurement of spatial room impulse responses using a non-uniformly moving microphone," in *2017 IEEE Workshop on Applications of Signal Processing to Audio and Acoustics (WASPAA)*, pp. 205–208, IEEE, 2017.

[10] N. Hahn, W. Hahne, and S. Spors, "Dynamic measurement of binaural room impulse responses using an optical tracking system," in *Proceedings of the International Conference on Spatial Audio*, pp. 16–21, 2017.

[11] J. He, R. Ranjan, W.-S. Gan, N. K. Chaudhary, N. D. Hai, and R. Gupta, "Fast continuous measurement of HRTFs with unconstrained head movements for 3d audio," *Journal of the Audio Engineering Society*, vol. 66, no. 11, pp. 884–900, 2018.

[12] N. Hahn and S. Spors, "Simultaneous measurement of spatial room impulse responses from multiple sound sources using a continuously moving microphone," in *2018 26th European Signal Processing Conference (EUSIPCO)*, pp. 2180–2184, IEEE, 2018.

[13] F. Katzberg, M. Maass, and A. Mertins, "Doppler frequency analysis for sound-field sampling with moving microphones," *Frontiers in Signal Processing*, vol. 4, p. 1304069, 2024.

[14] F. Katzberg, M. Maass, and A. Mertins, "Spherical harmonic representation for dynamic sound-field measurements," in *ICASSP 2021-2021 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pp. 426–430, IEEE, 2021.

[15] F. Katzberg, R. Mazur, M. Maass, P. Koch, and A. Mertins, "A compressed sensing framework for dynamic sound-field measurements," *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, vol. 26, no. 11, pp. 1962–1975, 2018.

[16] J. Brunnström, M. B. Møller, and M. Moonen, "Bayesian sound field estimation using moving microphones," *IEEE Open Journal of Signal Processing*, vol. 6, pp. 312–322, 2025.

[17] F. Katzberg, M. Maass, R. Pallenberg, and A. Mertins, "Positional tracking of a moving microphone in reverberant scenes by applying perfect sequences to distributed loudspeakers," in *2022 International Workshop on Acoustic Signal Enhancement (IWAENC)*, pp. 1–5, IEEE, 2022.

[18] C. C. Paige and M. A. Saunders, "LSQR: An algorithm for sparse linear equations and sparse least squares," *ACM Transactions on Mathematical Software (TOMS)*, vol. 8, no. 1, pp. 43–71, 1982.

[19] D. D. Rife and J. Vanderkooy, "Transfer-function measurement with maximum-length sequences," *Journal of the Audio Engineering Society*, vol. 37, no. 6, pp. 419–444, 1989.