DOI: 10.61782/fa.2025.0379

FORUM ACUSTICUM
EURONOISE 2025

EXPERIMENTAL VALIDATION OF SOUND FIELD ESTIMATION METHODS USING MOVING MICROPHONES

Jesper Brunnström[1*] **Martin Bo Møller**[2] **Toon van Waterschoot**[1]
Marc Moonen[1] **Jan Østergaard**[3]
[1] Department of Electrical Engineering (ESAT), STADIUS, KU Leuven, Leuven, Belgium
[2] Bang & Olufsen, Acoustics R&D, Struer, Denmark
[3] Department of Electronic Systems, Aalborg University, Aalborg, Denmark

ABSTRACT

Sound field estimation in a region of a room is essential for many sound field reproduction tasks. Although most estimation methods require microphones to be stationary, a few recently developed methods allow for flexible estimation with moving microphones, which can relax the equipment constraints of the estimation process. These methods have been evaluated on simulated data, leaving unanswered questions about their robustness in realistic environments. In this paper an experimental evaluation is made of sound field estimation methods using moving microphones, in particular those based on spherical harmonics. Experiments are conducted using sound field data collected with two microphones rotating around concentric circles, and 60 stationary microphone measurements in the interior of the circles. Position information is available for both moving and stationary microphones, as it is essential for sound field estimation. The data are released publicly to facilitate further experiments. The experiments indicate that given similar conditions, the moving microphone methods perform similarly well to the stationary microphone methods. However, due to the motor associated with the moving microphones, the signals are noisier, which leads to the stationary microphone methods performing better on the real data.

Keywords: sound field estimation, dataset, moving microphone, dynamic sampling, spherical harmonics

*__Corresponding author__: jesper.brunnstroem@kuleuven.be.

1. INTRODUCTION

Sound field reproduction tasks such as sound zone control generally require knowledge of the sound field generated by the available loudspeakers for one or more regions in the room [1–4]. For accurate reconstruction, the region of interest must be densely sampled compared to the wavelength of the highest frequency, which means that often many spatial samples must be collected.

To reduce the number of required spatial samples, many sound field estimation methods have been proposed, based on spherical harmonics [5], Gaussian processes [6, 7], kernel ridge regression [8, 9], and machine learning [10], among others. A central assumption among these methods is that the microphones are stationary during the measurement process. Using only one microphone, the measurement process must be repeated many times. Alternatively, a microphone array can be used, which is more expensive and cumbersome than a single microphone.

Although more difficult from a signal processing perspective, allowing the microphones to be moving can simplify the estimation process for the user. Some methods have been proposed for the purpose, with different properties and restrictions [11–14]. These methods have thus far been evaluated on simulated data, leaving unanswered questions about the robustness to noise, calibration errors, and model misspecifications that can be present in realistic environments [15, 16].

To provide an account on the real-world applicability of such methods, an experimental validation is made using real sound field data collected with moving and stationary microphones. The data is made available publicly [1], and contains moving microphone signals together

[1] Available at `github.com/pvjesper/romms`

Table 1. Available options in the RoMMS dataset.

Options	
Microphones	60 stationary, 2 moving
Acoustic conditions	4 reverberation times
Maximum frequency	500 Hz, 1000 Hz, 2000 Hz, 4000 Hz
Sequence length	0.5 s, 1.0 s

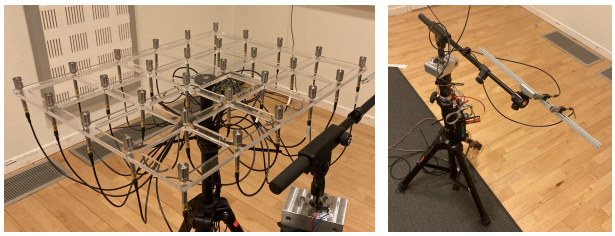

Figure 1. A picture of the stationary microphone array (left) and the moving microphones (right).

with both source signal and position as a function of time. In addition, measurements from a grid of stationary microphones is provided, which can be used to evaluate the sound field estimates. A particular focus is put on the methods [13, 14], which uses spherical harmonic sound field models. While both methods allow for essentially arbitrary trajectories, [14] allows for directional microphones and avoids truncating the sequence of spherical harmonic coefficients.

The experiments using real data are supplemented using simulated data, to investigate the influence of different error sources. In particular, the effects of noise in the microphone signals and error in the position estimates are investigated.

2. DATASET DESCRIPTION

The Rotating Moving Microphone Sound field (RoMMS) dataset is available at `github.com/pvjesper/romms`. A summary of the dataset features is shown in Table 1. The data was recorded with a RME Fireface UFX+ interface at a sampling rate of 48 kHz, using its internal inputs for the moving microphones and 4 RME Micstasy units for the stationary microphones. The loudspeaker was a Genelec 1032A studio monitor, and the microphones were GRAS 40AZ 1/2" low frequency microphones with GRAS 26CC 1/4" preamps, which were calibrated with a B&K 4231 field calibrator.

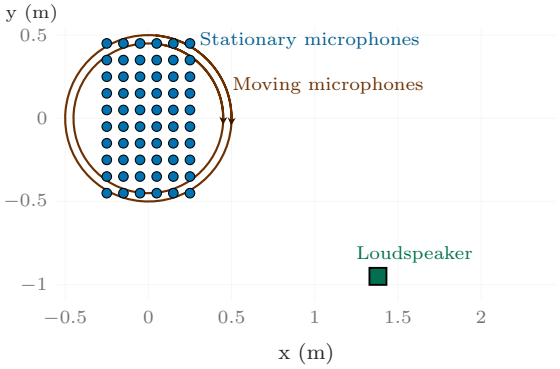

Figure 2. Geometry of the microphones and loudspeaker in the dataset.

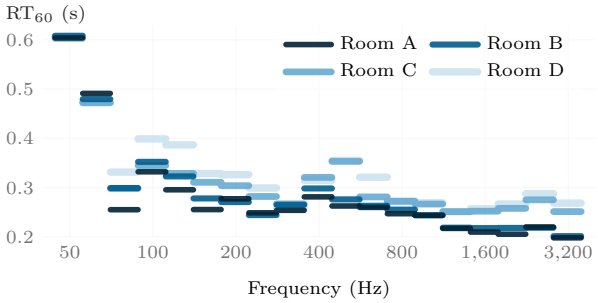

Figure 3. The reverberation time in third-octave bands for each of the four acoustic conditions.

2.1 Geometry

The measurements were made using a single loudspeaker, two moving microphones, and a microphone array with 30 stationary microphones, the geometry of which are shown in Fig. 2. The microphones and the loudspeaker were placed 1.03 m above the floor. The stationary microphone array was placed at two locations to obtain a 10×6 grid with 10 cm spacing, covering an area of 90×50 cm.

2.2 Acoustic conditions

The dataset was recorded in four different acoustic conditions. This can be useful to investigate how to adapt to changing acoustic conditions [17]. The acoustic condition was changed by varying the number of absorbent panels on the walls, thereby changing the reverberation time. The acoustic conditions are referred to as room A, B, C and D, listed from most to fewest acoustic panels.

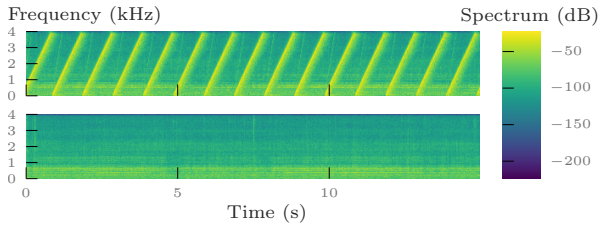

Figure 4. The recorded signal of the outer moving microphone (top), and only background noise recorded by the same microphone (bottom).

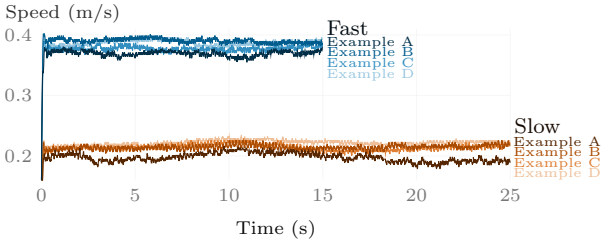

Figure 5. The speed of the outer moving microphone over time.

The reverberation time in terms of the RT_{60} was computed for third-octave bands between 40–4000 Hz by estimating the RT_{20}. The mean RT_{60} of the stationary microphones is shown in Fig. 3. As expected, room A has the shortest and room D has the longest reverberation time.

2.3 Loudspeaker signal

The loudspeaker signal used was a perfect periodic sweep, which is well suited for acoustic measurements, and has been used for moving microphones previously [12, 14]. The period of the signal should be chosen as equal length or longer than the room impulse response (RIR), otherwise a wraparound error is incurred. Recordings were made using a period length of 1 s as well as 0.5 s. Considering the reverberation time shown in Fig. 3, a period length of 1 s can be expected to be clearly sufficient, while 0.5 s could incur a small error. Because the RIRs will have decayed close to 60 dB after 0.5 s, high signal-to-noise ratio (SNR) is required for the error incurred by the shorter period length to be significant.

The periodic sweep has a frequency range between 0 Hz and f_{max}. Depending on the application, different frequency ranges can be of interest. Therefore, recordings were made with $f_{\text{max}} = 500$, 1000, 2000, and 4000 Hz.

For the remainder of this section, an example recording will be considered, with parameters $f_{\text{max}} = 4000$ Hz, sequence length 1 s, and slow speed, referred to as example A, B, C or D, depending on which acoustic environment is used. A spectrogram of the outer moving microphone signal of example A can be seen in Fig. 4, clearly showing the sweep used.

2.4 Moving microphones

As seen in Fig. 1, the two moving microphones are placed on a microphone boom arm fitted with a DC motor. As the boom arm of the microphone stand rotates, the microphones trace circles of radii 0.45 and 0.5 m. The data are collected for slightly less than two revolutions of the boom arm. Shortly before the end of the second revolution, a switch is closed to stop the rotation, generating a clicking noise. This noise along with the signals recorded afterwards is excluded from the data, explaining why the data does not contain two full rotations.

The encoder used for the motor produces a square wave signal as it rotates, which is recorded simultaneously with the microphone signals using the same signal chain, ensuring synchronicity between the position information and acoustic signal from the moving microphones. The square wave pulses are counted from the recorded signal, producing a sequence of timesteps where the estimated position of the microphones is incremented along the circle from their starting position. The encoder has a resolution of 500 pulses per revolution, and a gear with ratio 81.37 was used, which means that 40685 pulses are emitted for a single revolution around the circle. Equivalently, the resolution is 1.54×10^{-4} rad. Between the time of each such measured pulse, the angle is linearly interpolated to produce a smooth trajectory.

The moving microphone signals were recorded at a fast and slow speed. The mean speed over all recordings for the fast setting are 0.334 m/s and 0.376 m/s, and for the slow setting 0.186 m/s and 0.206 m/s, for the inner and outer microphone respectively. The speed over time for the example data can be seen in Fig. 5, which shows that the speed of the rotation varies slightly.

To obtain insights into the error characteristics of the position estimate, a closer look will be taken at the fast trajectory with $f_{\text{max}} = 500$ Hz, sequence length 0.5 s, in room A, which is used in Section 3. In Fig. 6, the deviations between the angles representing the estimated positions and the angles representing a constant speed trajec-

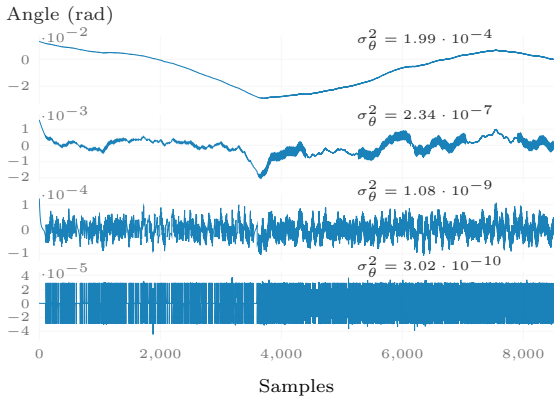

Figure 6. The angle of the microphone as a function of time, after removing the mean speed. The angle deviation is highpassed at 0.1, 1, 10, and 100 Hz.

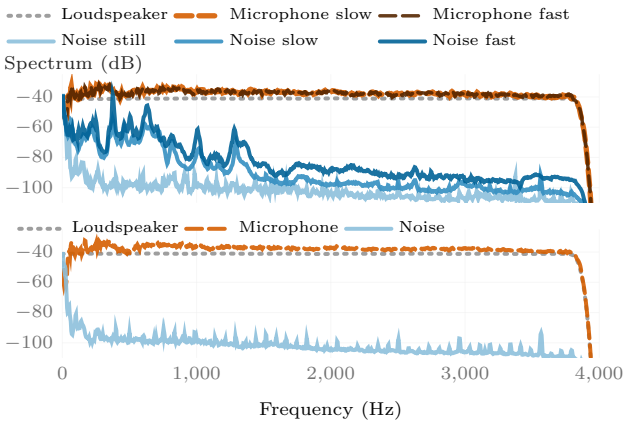

Figure 7. The spectrum of the moving microphone (top) and the stationary microphone (bottom) signals.

tory at the mean speed is shown. The deviations are high passed at 0.1, 1, 10, and 100 Hz using a first-order Butterworth filter, which reveals different characteristics for different frequency ranges. The size of deviations decrease as the frequency increases. Due to the inertia of the microphones and boom arm, the high frequency deviations can be reasonably assumed to represent errors in the position estimates, while the lower frequency deviations are more difficult to predict whether they correspond to the true position or not. Whether even the largest deviations are large enough to affect the sound field estimation performance considerably remains to be shown, and will be considered in Section 4.2.

2.5 Noise

For each of the acoustic conditions, the ambient noise level was recorded with the stationary microphones for 30 out of the 60 positions, the positions with a positive y-coordinate, i.e. for one placement of the 30-microphone array. Noise was also recorded with the moving microphones at fast and slow speed, as well as staying still at the initial point of the trajectory.

Spectrum estimates of the signals recorded by both the stationary and moving microphones in example A are shown in Fig. 7. The spectrum of the loudspeaker signal has an ambiguous scaling, as it is the loudspeaker signal prior to transmission. The spectrum estimates are computed using the Welch method, with a Hann window of length 2048 samples, 50 % overlap, and as the mean over all microphones of the same type. The dominant noise source for the moving microphones is the motor, which leads to higher noise levels for a fast speed over a slow speed, which in turn is higher than when still. The stationary microphone signals have a relatively constant noise level as a function of frequency, while the noise level for the moving microphone signals is highly frequency dependent. Fig. 7 shows that the amplitude of the microphone signal decreases drastically for frequencies below 50 Hz, and for frequencies close to the Nyquist frequency. The reduced amplitude for the low frequencies is due to the loudspeaker, and for the high frequencies due to the construction of the periodic sweep signal. Therefore, accurate estimates cannot be expected at the highest and lowest frequencies, due to the inevitably poor SNR.

2.6 Delay correction

The stationary microphone and the moving microphone signals are recorded through two different signal chains, that while being clock-synchronized introduces slightly different amounts of delay, enough to hinder a comparison between estimated sound pressures from the two sets of microphones. Therefore, the data from the moving microphones have been delayed in order to align the signals with the stationary microphones.

The positions of the loudspeaker and the microphones are known along with the speed of sound, which means that the propagation times from the loudspeaker to the microphones are known. This time can then be compared to the direct path component in the estimated RIRs from the stationary and moving microphones. The time for the

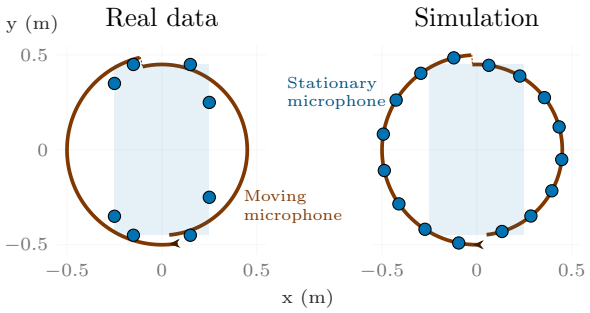

Figure 8. The position of the stationary microphones and the moving microphones in the experiments.

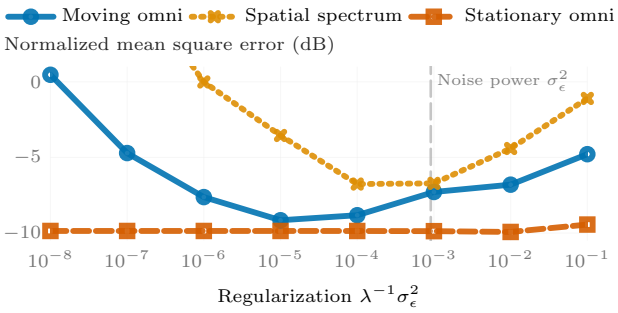

Figure 9. The estimation error as a function of the regularization parameter.

direct path component in the estimated RIRs is estimated by choosing the largest peak of the RIR. As the quantity of interest is the relative delay between the stationary and moving microphones, more sophisticated methods are not deemed necessary. This process is performed for each $f_{\max}$, sequence length, and acoustic condition, using the fast speed. The mean relative delay is 0.259 ms, therefore the moving microphone signals are delayed by that amount, rounded to the nearest sample at a sampling rate of 48 kHz.

3. SOUND FIELD ESTIMATION WITH REAL DATA

The moving microphone data will be used to estimate the sound field in the interior of the disc traced by the circle. The parameters for the data are $f_{\max} = 500$ Hz, sequence length 0.5 s, fast speed, and room A. The trajectory is chosen as one full revolution around the circle, using half the data from each moving microphone, as shown in Fig. 8. This choice avoids the forbidden-frequency problem, as opposed to using a full revolution of one microphone [18, Section 8.10.2]. The eight stationary microphones shown in Fig. 8 are used for sound field estimation with stationary microphones.

The choice of stationary microphones is fairly arbitrary, and the choice leads to better or worse estimation performance compared to the moving microphone methods. Therefore, the microphones are chosen such that the moving and stationary microphones methods achieve a similar lowest estimation error. The choice in Fig. 8 provides a lesser amount of data for the stationary microphones, but instead has higher SNR and spatial samples located closer to the evaluation points.

The first considered sound field estimation method is *moving omni*, the proposed method of [14] using moving omnidirectional microphones. *Spatial spectrum* is the method from [13] using moving microphones, with implementation details described in [14, Section V.B.4]. The moving microphone methods are compared against one method using stationary microphones, which is *stationary omni*, the method from [5] which can be viewed as the stationary microphone equivalent of *moving omni*. *Stationary omni* is also equivalent to [8] for omnidirectional microphones, which is the case in these experiments.

3.1 Regularization

Both *moving omni* and *spatial spectrum* require a regularization matrix to be chosen. In the Bayesian approach of [14], the interpretation of this regularization matrix is $\lambda^{-1}\boldsymbol{\Sigma}$, where $\boldsymbol{\Sigma}$ is the covariance of the measurement noise, and λ is the prior variance of the harmonic coefficients. In these experiments, $\boldsymbol{\Sigma}$ is chosen as $\sigma_\epsilon^2\boldsymbol{I}$, corresponding to an assumption that the noise is temporally white. The parameter σ_ϵ^2 is the noise power calculated from the noise-only recordings in the dataset. The regularization in *spatial spectrum* is chosen as the same value.

While the noise power can be relatively easily measured, the same is not true for the prior variance of the harmonic coefficients λ. In Fig. 9 the effect of the regularization parameter $\lambda^{-1}\sigma_\epsilon^2$ on the estimation error is shown. To obtain a comparatively low error for all methods, the prior variance parameter is set to $\lambda^{-1} = 0.1$, and the noise variance parameter σ_ϵ^2 is set according to the data.

3.2 Experimental results

The estimation performance in terms of the normalized mean square error (NMSE) is shown in Fig. 10 as a func-

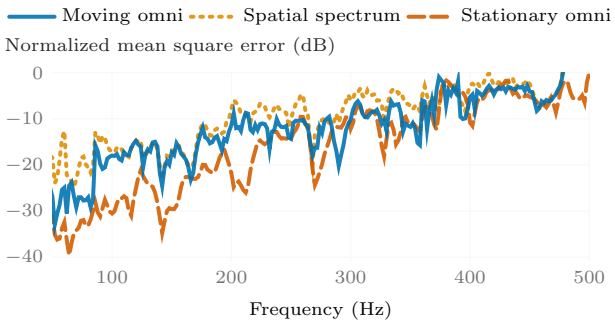

Figure 10. The NMSE of the considered methods as a function of frequency for real data.

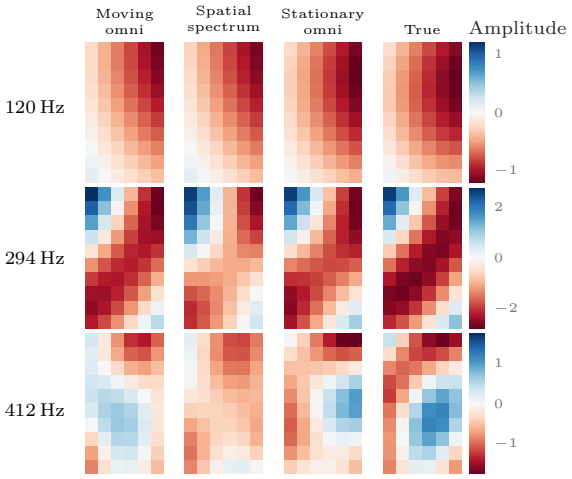

Figure 11. The real part of the sound field estimates for different frequencies. The sound field is shown at the positions of the stationary microphones in Fig. 2.

tion of frequency. *Stationary omni* outperforms *moving omni*, which in turn outperforms *spatial spectrum*. However, the SNR is lower in the stationary microphone data, as well as some of the microphones being placed closer to the evaluation points compared to the moving trajectory. The NMSE increases close to the Nyquist frequency, which is due to the true value approaching zero, as can be seen in Fig. 7, and is therefore not an indication of an increased absolute error.

The real part of the sound field estimates for several frequencies are shown in Fig. 11 as a function of space. It is again apparent that the accuracy for all sound field estimation methods decreases as the frequency increases.

4. INVESTIGATION OF ERROR SOURCES

To provide insights into the factors leading to the loss of estimation accuracy between the moving and stationary microphone methods, the influence of the background noise and microphone position error will be investigated in a simulation study. Considering experiments presented in [14], it can be expected that under ideal conditions the stationary and moving microphones methods should perform similarly well.

A room of size 7×3.2×2.5 m was simulated using the image-source method [19, 20], with a reverberation time of 0.36 s. Unless stated otherwise, experiment parameters are chosen identically to the experiment described in Section 3. The trajectory for the moving microphones is constructed identically to Fig. 2, but with the microphones moving at a constant speed, chosen as the mean speed of the real data used in Section 3. For a fair comparison, the stationary microphones are positioned on the trajectory of the moving microphones, shown in Fig 8, with one microphone per period of the periodic sweep. The evaluation points covers the same area as in Fig. 2, but with a higher resolution at a microphone distance of 5 cm.

The regularization of *spatial spectrum* is chosen as in Section 3.1, but no smaller than 10^{-3}, as numerical issues then cause the estimation error to drastically increase.

4.1 Microphone signal noise

As shown in Fig. 7, the moving microphones have a lower SNR due to the noise of the motor, which means that robustness to noise can be important. Therefore, the relationship between SNR and estimation performance will be investigated.

A noise-only recording from the real moving microphone data is added to the simulated data for both moving and stationary microphones, with a scaling to produce different SNRs. In Fig. 12 the estimation performance is shown as a function of the SNR. The NMSE starts to increase for *moving omni* and *stationary omni* once the SNR is below 40 dB, which can be compared against the SNR of the real data which is approximately 21 dB for the data used in Section 3. The figure shows that given the same SNR, *moving omni* and *stationary omni* provides almost identical estimation performance. These results indicate that the higher noise level in the moving microphone signals in the real data is at least partly responsible for the degraded estimation performance. The estimation error

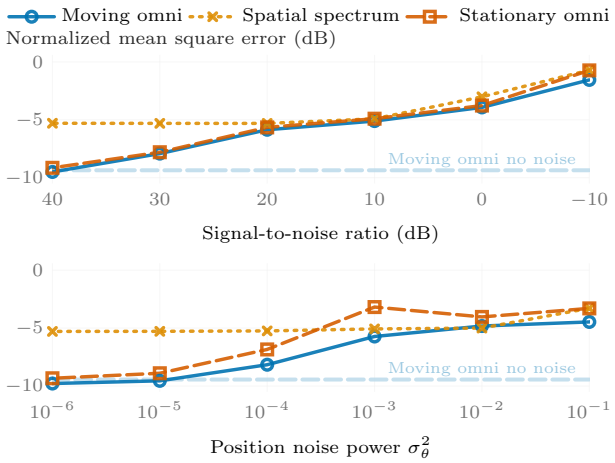

Figure 12. The estimation performance for different SNRs (top) and position noise powers (bottom).

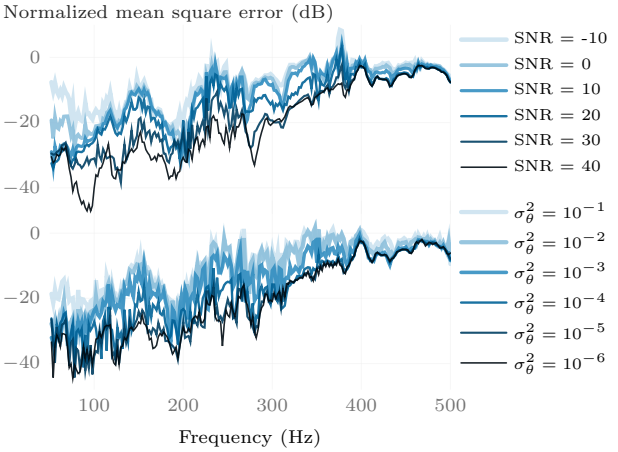

Figure 13. The estimation performance of *moving omni* as a function of frequency for different levels of noise in the microphone signals (top) and the position estimates (bottom).

of *moving omni* per frequency is shown in Fig. 13. The degradation for the low frequencies in particular is considerable as the SNRs decreases.

4.2 Position error

Another potential error source is the microphone position estimation error. In this experiment the sound field estimates are computed using perturbed position estimates. The angle at each sample is perturbed by a white zero-mean Gaussian noise with variance σ_θ^2. These perturbed estimates will always be on the circle.

In contrast to the microphone signal noise, *moving omni* does not include an explicit uncertainty in the position estimates. Still, to obtain good estimates, the regularization parameter should be chosen comparatively large if there is a significant position error, even if the noise in the microphone signal is negligible. Therefore a heuristic strategy for selecting the regularization parameter is used, following the strategy described in Section 3.1, but using $10\sigma_\theta^2$ instead of σ_ϵ^2.

The NMSE for varying levels of position error is shown in Fig. 12. The added position error can be compared against the variability in the position of the real data shown in Fig. 6. Only the lowest frequency deviations in Fig. 6 are high enough amplitude that Fig. 12 would indicate an effect on the estimation performance. This indicates that only the lowest frequency parts of the deviations in Fig. 6 are likely to be a contributor.

5. CONCLUSION

An experimental validation of sound field estimation methods using rotating moving microphones has been made. The effect of noise in the signal and position estimates has been investigated, indicating that the noise in the moving microphone signal from its associated motor is a likely significant source of performance degradation in the real data. For similar noise conditions, similar quality of estimates were obtained from stationary and moving microphones. The moving microphone methods were thereby shown to be competitive on real data, indicating that sound field estimation with moving microphones is a promising approach for sound field measurements.

6. ACKNOWLEDGEMENTS

This research work was carried out at the ESAT Laboratory of KU Leuven, in the frame of Research Council KU Leuven C14-21-0075 ”A holistic approach to the design of integrated and distributed digital signal processing algorithms for audio and speech communication devices”. This project has received funding from the European Union’s Horizon 2020 research and innovation programme under the Marie Skłodowska-Curie grant agreement No. 956369: ’Service-Oriented Ubiquitous Network-Driven Sound — SOUNDS’. The scientific responsibility is assumed by its authors.

7. REFERENCES

[1] W. Zhang, P. N. Samarasinghe, H. Chen, and T. D. Abhayapala, "Surround by sound: A review of spatial audio recording and reproduction," *Appl. Sci.*, vol. 7, May 2017. Art. no. 532.

[2] T. Lee, L. Shi, J. K. Nielsen, and M. G. Christensen, "Fast generation of sound zones using variable span trade-off filters in the DFT-domain," *IEEE/ACM Trans. Audio, Speech, Lang. Process.*, vol. 29, pp. 363–378, Dec. 2020.

[3] J. Zhang, L. Shi, M. G. Christensen, W. Zhang, L. Zhang, and J. Chen, "Robust pressure matching with ATF perturbation constraints for sound field control," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process. (ICASSP)*, pp. 8712–8716, May 2022.

[4] J. Brunnström, T. van Waterschoot, and M. Moonen, "Signal-to-interference-plus-noise ratio based optimization for sound zone control," *IEEE Open J. Signal Process.*, vol. 4, pp. 257–266, Feb. 2023.

[5] N. Ueno, S. Koyama, and H. Saruwatari, "Sound field recording using distributed microphones based on harmonic analysis of infinite order," *IEEE Signal Process. Lett.*, vol. 25, pp. 135–139, Jan. 2018.

[6] D. Caviedes-Nozal, N. A. B. Riis, F. M. Heuchel, J. Brunskog, P. Gerstoft, and E. Fernandez-Grande, "Gaussian processes for sound field reconstruction," *J. Acoust. Soc. Am.*, vol. 149, pp. 1107–1119, Feb. 2021.

[7] D. Caviedes-Nozal and E. Fernandez-Grande, "Spatio-temporal bayesian regression for room impulse response reconstruction with spherical waves," *IEEE/ACM Trans. Audio, Speech, Lang. Process.*, vol. 31, pp. 3263–3277, 2023.

[8] N. Ueno, S. Koyama, and H. Saruwatari, "Kernel ridge regression with constraint of Helmholtz equation for sound field interpolation," in *Proc. Int. Workshop Acoust. Signal Enhancement (IWAENC)*, pp. 436–440, IEEE, Sept. 2018.

[9] J. Brunnström, S. Koyama, and M. Moonen, "Variable span trade-off filter for sound zone control with kernel interpolation weighting," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process. (ICASSP)*, pp. 1071–1075, May 2022.

[10] E. Fernandez-Grande, X. Karakonstantis, D. Caviedes-Nozal, and P. Gerstoft, "Generative models for sound field reconstruction," *J. Acoust. Soc. Am.*, vol. 153, pp. 1179–1190, Feb. 2023.

[11] F. Katzberg, R. Mazur, M. Maass, P. Koch, and A. Mertins, "A compressed sensing framework for dynamic sound-field measurements," *IEEE/ACM Trans. Audio, Speech, Lang. Process.*, vol. 26, pp. 1962–1975, Nov. 2018.

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

[13] F. Katzberg, M. Maass, and A. Mertins, "Spherical harmonic representation for dynamic sound-field measurements," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process. (ICASSP)*, pp. 426–430, June 2021.

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

[15] 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 *Proc. Int. Workshop Acoust. Signal Enhancement (IWAENC)*, Sept. 2022.

[16] F. Katzberg, M. Maass, and A. Mertins, "Doppler frequency analysis for sound-field sampling with moving microphones," *Front. Signal Process.*, vol. 4, Apr. 2024.

[17] J. Brunnström, M. B. Møller, J. Østergaard, and M. Moonen, "Bayesian sound field estimation using uncertain data," in *Proc. Int. Workshop Acoust. Signal Enhancement (IWAENC)*, pp. 329–333, Sept. 2024.

[18] E. G. Williams and J. A. Mann, *Fourier Acoustics: Sound Radiation and Nearfield Acoustical Holography*, vol. 108. Academic Press, 1999.

[19] J. B. Allen and D. A. Berkley, "Image method for efficiently simulating small-room acoustics," *J. Acoust. Soc. Am.*, vol. 65, pp. 943–950, Apr. 1979.

[20] R. Scheibler, E. Bezzam, and I. Dokmanić, "Pyroomacoustics: A Python package for audio room simulations and array processing algorithms," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process. (ICASSP)*, pp. 351–355, Apr. 2018.