Direction-of-arrival and power spectral density estimation using a single directional microphone

Elisa Tengan, Maja Taseska[1], Thomas Dietzen and Toon van Waterschoot
Dept. of Electrical Engineering (ESAT-STADIUS)
KU Leuven
Leuven, Belgium
Email: elisa.tengan@esat.kuleuven.be, taseska.maja@gmail.com,
thomas.dietzen@esat.kuleuven.be, toon.vanwaterschoot@esat.kuleuven.be

***Abstract*—A method is proposed for estimating direction-of-arrival (DOA) and power spectral density (PSD) of stationary point source signals using a single, rotating, directional microphone. By considering different microphone orientations for different time frames, the DOA is estimated by locating the maxima in an estimated PSD vector obtained by solving a group-sparsity constrained optimization problem using a dictionary composed of the known microphone response sampled on an angular grid. The estimated DOAs are then used for obtaining an overdetermined least squares problem with a nonnegativity constraint for re-estimating the PSD of the point source signals. The DOA estimation performance is compared between cases of different frequency-dependent microphone directivity patterns, as well as with the MUSIC algorithm for 6-element uniform linear and circular microphone arrays. The proposed stationary point source PSD estimation using DOA information is compared with traditional single-channel methods for PSD estimation employing minimum statistics and MMSE-based approaches for a rotating microphone setup, one speech source and one stationary interfering point source.**

***Index Terms*—Direction-of-arrival estimation, power spectral density estimation, single-channel, speech enhancement**

I. Introduction

Multichannel noise reduction is known to show better performance than single-channel approaches, due to the spatial diversity offered by microphone arrays [1]. However, due to computational complexity and hardware design restrictions, multichannel speech enhancement is not always able to reach its theoretical potential in practice [2], [3]. In this paper, we provide a different perspective on capturing and estimating spatial information of audio signals for noise reduction while maintaining a simple hardware setup, by proposing a method for power spectral density (PSD) estimation based on signal direction-of-arrival (DOA) estimation with a single, rotating, directional microphone. Some examples of applications that could benefit from the proposed approach include those involving devices that can present spatially dynamic behavior, such as hearing aids, smartphones and cameras. Applications involving the use of off-the-shelf multi-microphone mobile devices, where the microphone array geometry might be unknown, could also benefit from a single-microphone fallback strategy.

Although single-microphone source localization has already been performed by employing machine learning and scattering structures [4]–[6], as well as by exploiting the Doppler effect obtained from constant circular motion [7], [8], single-channel spatial audio analysis based on movements of a directional microphone remains a largely unexplored problem. Therefore, the method proposed in this paper is developed under a set of considerably strong assumptions, mainly of spatiotemporal stationarity of uncorrelated sources, anechoic conditions and controlled microphone movements. As in most research challenges, these assumptions are expected to be further relaxed in the future while gradually developing more suitable solutions.

The main concept behind the proposed method is that a directional microphone will capture a spatially static and localized sound source with a different response depending on the direction towards which it is oriented. As the microphone orientation varies for different observation time frames, changes in the microphone signal PSD can be analyzed for determining spatial information about the sources generating the observed sound field. More specifically, we show that the DOA of multiple point source signals can be estimated by solving a group-sparsity constrained optimization problem for the direction-dependent PSD values relative to a given angular dictionary, and locating peaks in the estimated PSD vector. Due to the biased nature of the PSD estimation in the sparsity-constrained problem, the estimated DOAs are then used for re-estimating solely the PSD of the located point source signals, by solving an overdetermined least-squares problem with a nonnegativity constraint. The performance of the DOA estimation step is assessed through simulations for different frequency-dependent microphone directivity patterns, and is compared to the MUSIC algorithm for 6-element uniform linear and circular microphone arrays. The proposed PSD estimation method is compared with the traditional single-channel methods employing minimum statistics and MMSE-based approaches for a rotating microphone setup, one speech source and one stationary interfering point source.

This research work was carried out at the ESAT Laboratory of KU Leuven, in the frame of FWO Mandate SB 1S86520N, and VLAIO O&O Project no. HBC.2017.0358 "SPOTT - Tomorrow's Scalable and PersOnalised advertising Technology, Today". The research leading to these results has also received funding from the European Research Council under the European Union's Horizon 2020 research and innovation program/ERC Consolidator Grant: SONORA (no. 773268). This paper reflects only the authors' views and the Union is not liable for any use that may be made of the contained information.

[1] Maja Taseska is currently with Microsoft Applied Sciences.

ISBN: 978-9-0827-9706-0

This paper is organized as follows. In section II, the signal model is defined. In section III, the proposed method is presented by detailing the two-step approach consisting of DOA estimation followed by multiple point source PSD estimation. In section IV, the simulation procedure is described for the evaluation of both estimation steps, and the results obtained are then discussed. Finally, in section V, we conclude with a summary and final remarks on the work presented.

II. Signal model

The signal recorded by a single, rotating and directional microphone is modeled in the short-time Fourier transform (STFT) domain as:

$$Y(k,n) = \sum_{p=1}^{P} a(k,\theta_{S_p} - \gamma_n) H(k,\theta_{S_p}) S(k,n,\theta_{S_p}) + D(k,n) \tag{1}$$

where k corresponds to the discrete frequency index and n corresponds to the discrete time frame index. We assume that there are a total of P point sources in the far field, denoted as $S_1, S_2, \ldots, S_P$, and that P is known. The expression in (1) describes that while considering a two-dimensional scenario, on the horizontal plane, the STFT of the resulting signal recorded by the microphone $Y(k,n)$ is a sum of the STFT of the P point source signals $S(k,n,\theta_{S_p})$ arriving from P distinct directions, θ_{S_1} to θ_{S_P}, multiplied by the direction-dependent microphone response $a(k,\theta_{S_p} - \gamma_n)$, relative to the microphone orientation γ_n, and by the room transfer function (RTF) $H(k,\theta_{S_p})$, added to diffuse or sensor noise $D(k,n)$.

If we consider that the recording is performed in anechoic conditions, then $H(k,\theta_p) = 1$, $\forall\, k$ and for $p = 1,\ldots,P$. Assuming that the source signals are stationary, the PSDs do not vary from one time frame n to another. Moreover, if we assume that the source signals are uncorrelated, and that the microphone response is real-valued, then the microphone signal PSD $\phi_Y(k,n)$ can be described as follows:

$$\phi_Y(k,n) = \sum_{p=1}^{P} a^2(k,\theta_p - \gamma_n)\phi_S(k,\theta_p) + \phi_D(k,n) \tag{2}$$

where $\phi_D(k,n)$ is the noise PSD for frequency k and time frame n, and $\phi_S(k,\theta_p)$ is the PSD for frequency k corresponding to a source at position θ_p. As the directional microphone is oriented towards different directions γ_n for different time frames n, the resulting microphone signal PSD presents variations with n, since the relative positions of the sound sources with respect to the microphone do not remain the same, and consequently their PSD values are multiplied with different squared microphone response coefficients over time.

III. Proposed method

A. DOA estimation based on group-sparsity regularization

Firstly, we define the vector $\widetilde{\boldsymbol{\phi}}_Y \in \mathbb{R}^N$, whose elements correspond to the sum of PSD values over all K frequency bins in each observation time frame, for a total of N different frames:

$$\widetilde{\boldsymbol{\phi}}_Y = \left[\sum_{k=1}^{K} \phi_Y(k,1) \;\; \ldots \;\; \sum_{k=1}^{K} \phi_Y(k,N)\right]^\top \tag{3}$$

with $[\cdot]^\top$ as the transpose operator. We also define a dictionary matrix $\mathbf{A} \in \mathbb{R}^{N\times KL}$ containing the squared microphone response coefficients for each frame and candidate source direction from a grid of L uniformly distributed angles:

$$\mathbf{A} = \begin{bmatrix} \boldsymbol{a}^2(\theta_1 - \gamma_1) & \ldots & \boldsymbol{a}^2(\theta_L - \gamma_1) \\ \vdots & \ddots & \vdots \\ \boldsymbol{a}^2(\theta_1 - \gamma_N) & \ldots & \boldsymbol{a}^2(\theta_L - \gamma_N) \end{bmatrix} \tag{4}$$

with:

$$\boldsymbol{a}^2(\theta_i - \gamma_n) = \left[a^2(1,\theta_i - \gamma_n) \;\; \ldots \;\; a^2(K,\theta_i - \gamma_n)\right] \tag{5}$$

It is assumed that the angular dictionary contains θ_{S_1} to θ_{S_P}. We now define the following linear system of equations:

$$\widetilde{\boldsymbol{\phi}}_Y = \mathbf{A}\,\widetilde{\boldsymbol{\phi}}_S + \widetilde{\boldsymbol{\phi}}_D \tag{6}$$

where the vector $\widetilde{\boldsymbol{\phi}}_S \in \mathbb{R}^{KL}$ contains the PSD values of the point source signals in all candidate directions ranging from θ_1 to θ_L:

$$\widetilde{\boldsymbol{\phi}}_S = \left[\boldsymbol{\phi}_S^\top(\theta_1) \;\; \ldots \;\; \boldsymbol{\phi}_S^\top(\theta_L)\right]^\top \tag{7}$$

with $\boldsymbol{\phi}_S(\theta_i) \in \mathbb{R}^K$ defined as:

$$\boldsymbol{\phi}_S(\theta_i) = \left[\phi_S(1,\theta_i) \;\; \ldots \;\; \phi_S(K,\theta_i)\right]^\top \tag{8}$$

The vector $\widetilde{\boldsymbol{\phi}}_D \in \mathbb{R}^N$, similarly to $\widetilde{\boldsymbol{\phi}}_Y$, is defined as:

$$\widetilde{\boldsymbol{\phi}}_D = \left[\sum_{k=1}^{K} \phi_D(k,1) \;\; \ldots \;\; \sum_{k=1}^{K} \phi_D(k,N)\right]^\top \tag{9}$$

By ensuring that $\gamma_1 \neq \gamma_2 \neq \ldots \neq \gamma_N$, with $0 \leq \gamma_n \leq 2\pi$, $\forall\, n$, and assuming that $\widetilde{\boldsymbol{\phi}}_Y$ and $\mathbf{A}$ are known, source localization can be achieved by solving the proposed linear system of equations, which would allow us to identify from the estimated vector $\widetilde{\boldsymbol{\phi}}_S$ in which direction within the angular dictionary there are peaks in power, indicating the point source DOAs. Since the linear system in (6) is underdetermined, the following Group Lasso optimization problem is considered:

$$\underset{\widetilde{\boldsymbol{\phi}}_S}{\text{minimize}} \quad \frac{1}{2}\left\|\widetilde{\boldsymbol{\phi}}_Y - \mathbf{A}\widetilde{\boldsymbol{\phi}}_S\right\|_2^2 + \lambda \sum_{i=1}^{L} \left\|\boldsymbol{\phi}_S(\theta_i)\right\|_2 \tag{10}$$

$$\text{subject to} \quad \widetilde{\boldsymbol{\phi}}_S \geq 0 \tag{11}$$

The nonnegativity constraint in (11) is necessary for complying with the intrinsic nonnegativity property of PSD values [9]. The Group Lasso formulation includes the regularization term $\lambda \sum_{i=1}^{L} \|\boldsymbol{\phi}_S(\theta_i)\|_2$, which enforces sparsity between so-called different groups [10]. The group sparsity penalty resonates with the assumption that only a limited number of point sources are present in space, and consequently, only a few of the shorter vectors $\boldsymbol{\phi}_S(\theta_i)$ composing $\widetilde{\boldsymbol{\phi}}_S$ should be nonzero.

After solving the optimization problem (10)-(11) and obtaining an estimate of $\widetilde{\boldsymbol{\phi}}_S$, and therefore, of $\boldsymbol{\phi}_S(\theta_1), \ldots, \boldsymbol{\phi}_S(\theta_L)$, the PSD values are averaged over the K frequency bins for each of the L candidate directions, allowing for DOA estimation by finding the indices of θ for which there are peaks in the average PSD exceeding a predetermined threshold. For a total of P sources assumed to be present, P peaks should then be identified.

B. Stationary point source power spectral density estimation

Although the previous step is sufficient for modelling and estimating angular peaks in power and effectively estimating the source DOAs, due to the use of a sparsity constraint in the formulation of the optimization problem, the resulting PSD estimates for all directions in the angular dictionary are inherently biased [11]. Hence, a re-estimation step for the PSD values using the previously estimated DOAs is proposed.

Using the PSD signal model in (2), a new linear system of equations for the microphone signal PSD can be formulated, for each frequency bin, as:

$$\boldsymbol{\phi}_Y(k) = \mathbf{A}_S(k)\boldsymbol{\phi}_S(k) + \boldsymbol{\phi}_D(k) \tag{12}$$

The matrix $\mathbf{A}_S(k) \in \mathbb{R}^{N\times P}$ now contains squared microphone response coefficients for only the directions where the P sources are assumed to be located, based on the preceding DOA estimation, denoted by $\hat{\theta}_{S_1}, \ldots, \hat{\theta}_{S_P}$:

$$\mathbf{A}_S(k) = \begin{bmatrix} a^2(k, \hat{\theta}_{S_1} - \gamma_1) & \ldots & a^2(k, \hat{\theta}_{S_P} - \gamma_1) \\ \vdots & & \vdots \\ a^2(k, \hat{\theta}_{S_1} - \gamma_N) & \ldots & a^2(k, \hat{\theta}_{S_P} - \gamma_N) \end{bmatrix} \tag{13}$$

The new PSD vectors are defined as follows:

$$\boldsymbol{\phi}_Y(k) = [\phi_Y(k,1) \;\; \ldots \;\; \phi_Y(k,N)]^\top \tag{14}$$

$$\boldsymbol{\phi}_S(k) = \left[\phi_S(k,\hat{\theta}_{S_1}) \;\; \ldots \;\; \phi_S(k,\hat{\theta}_{S_P})\right]^\top \tag{15}$$

$$\boldsymbol{\phi}_D(k) = [\phi_D(k,1) \;\; \ldots \;\; \phi_D(k,N)]^\top \tag{16}$$

If $P \leq N$, an ordinary least-squares approach can be used for solving the overdetermined linear system with a nonnegativity constraint and estimating the PSD values of the point sources:

$$\underset{\boldsymbol{\phi}_S(k)}{\text{minimize}} \quad \frac{1}{2}\left\|\boldsymbol{\phi}_Y(k) - \mathbf{A}_S(k)\boldsymbol{\phi}_S(k)\right\|_2^2 \tag{17}$$

$$\text{subject to} \quad \boldsymbol{\phi}_S(k) \geq 0 \tag{18}$$

Hence, in this re-estimation step, we avoid the bias induced by the Group Lasso formulation presented in the DOA estimation step and allow a more accurate PSD estimation for the stationary point sources.

IV. Simulations

A. Setup

In order to evaluate the proposed method, simulations were carried out such that it was possible to first test the DOA estimation separately, and then evaluate the performance of the stationary point source PSD estimation following such preliminary step. When evaluating the DOA estimation, two point sources of white Gaussian noise of equal power, denoted σ_S^2, are placed in the far field with an angular separation varying from 30° to 180°. Additive spatially white Gaussian noise of power σ_D^2 is also included and the signal-to-noise ratio (SNR), defined as σ_S^2/σ_D^2, is set to $0\,\mathrm{dB}$ and $15\,\mathrm{dB}$ in two different simulations. When evaluating the PSD estimation, since the main motivation for performing such step is to later allow noise reduction and speech enhancement, we employ one speech source and one stationary point source classified as an interference, even though one of the assumptions implied in the proposed method is set to be violated (i.e., one of the sources is not stationary). The speech source of average power σ_S^2 is simulated with a recording of a male speaker from Music for Archimedes [12], and the interfering point source is simulated with a white Gaussian noise of power σ_I^2. They have their positions fixed at 0° and 180°, respectively, with a signal-to-interference ratio (SIR), defined as σ_S^2/σ_I^2, of $3\,\mathrm{dB}$. We seek to estimate the PSD of the interfering point source when there is no additive noise.

We then simulate a microphone recording of the resulting signal with the microphone oriented towards the directions of 0°, 60°, 120°, 180°, 240°, and 300°, by multiplying the simulated source signal with the known microphone responses $a(k, \theta_{S_p} - \gamma_n)$ resulting in $N = 6$ different observation frames. The duration of each frame is $500\,\mathrm{ms}$. The sampling frequency is $16\,\mathrm{kHz}$, and the microphone remains static during one time frame. The PSD $\phi_Y(k,n)$ is estimated using Welch's method, considering $K = 512$ frequency bins, employing a Hann window and 50% overlap. The angular dictionary for building the matrix $\mathbf{A}$ presents a 5° resolution, resulting in a grid with $L = 72$ candidate directions for the point sources. In all simulations, the true DOAs are chosen to be on grid.

The optimization problems defined in (10)-(11) and (17)-(18) are solved using CVX [13], with the regularization parameter λ in (10) set to $0.01\|\mathbf{A}^\top\widetilde{\boldsymbol{\phi}}_Y\|_\infty$, with $\|\cdot\|_\infty$ as the l_∞-norm, and a total of 100 realizations simulated for each of the scenarios previously described. We evaluate the performance of the proposed DOA estimation method in terms of root-mean-square error (RMSE), computed as:

$$\text{RMSE} = \sqrt{\frac{1}{N_r}\sum_{l=1}^{N_r}\frac{\sum_{p=1}^{P}(\theta_{S_p} - \hat{\theta}_{S_p}^l)^2}{P}} \tag{19}$$

where N_r is the total number of realizations, θ_{S_p} is the true DOA of source S_p, and $\hat{\theta}_{S_p}^l$ is the estimated DOA of source S_p in realization l.

Various cardioid microphone responses are simulated, both with a flat frequency response, as well as with distinct frequency-dependent directivity patterns to simulate more realistic conditions where a microphone becomes more directional for higher frequencies, allowing a performance comparison between different responses. For a normalized frequency value $f \in [0,1]$, and a direction $\theta \in [0, 2\pi]$, the frequency-dependent directivity patterns, denoted as *Sub-to-cardioid* and *Omni-to-*

cardioid, are defined as a linear combination of two directivity functions:

$$a(f,\theta) = (1-f)a_L(\theta) + f a_H(\theta) \tag{20}$$

where:

$$\begin{aligned} a_H(\theta) &= 0.5 + 0.5\cos(\theta) \\ a_L(\theta) &= 0.75 + 0.25\cos(\theta) \end{aligned} \qquad \text{Sub-to-cardioid} \tag{21}$$

or:

$$\begin{aligned} a_H(\theta) &= 0.5 + 0.5\cos(\theta) \\ a_L(\theta) &= 1 \end{aligned} \qquad \text{Omni-to-cardioid} \tag{22}$$

We also compare the RMSE obtained when using the simulated cardioid microphone of flat frequency response with the wideband MUSIC algorithm [14], [15] with harmonic averaging [16] applied to a simulated 6-element uniform linear array (ULA) and to a simulated 6-element uniform circular array (UCA), with a microphone spacing of $5\,\mathrm{cm}$.

For evaluating the accuracy of the interfering stationary point source PSD estimation, we compute the normalized mean-square error (NMSE) per frequency bin as:

$$\mathrm{NMSE}(k) = \frac{1}{N_r}\sum_{l=1}^{N_r}\left(\frac{\phi_{S_I}(k) - \hat{\phi}^l_{S_I}(k)}{\phi_{S_I}(k)}\right)^2 \tag{23}$$

where $\phi_{S_I}(k)$ is the true interfering source PSD and $\hat{\phi}^l_{S_I}(k)$ is the estimated PSD in realization l. We compare the NMSE obtained with the proposed method with the NMSE obtained when employing the single-channel methods based on minimum statistics [17], [18] and on the unbiased minimum mean-square error (MMSE) estimator [19], with implementations from Voicebox [20], while using the same simulated output from the cardioid microphone that is oriented towards different directions for different time frames.

B. Results

The RMSE values resulting from the DOA estimation of two sources for different SNRs and angular separation values obtained with the proposed method and different microphone responses are presented in Fig. 1. We can observe that for the proposed method, larger angular separation between sources contributes to lower RMSE values. We also notice a performance improvement when the microphone response presents a higher degree of directivity for larger frequency ranges, with the best case being the ideal cardioid pattern with a flat frequency response. The performance comparison between the proposed method for a cardioid microphone with a flat frequency response and the MUSIC algorithm applied to the linear and circular arrays in terms of RMSE for DOA estimation is presented in Fig. 2. We observe that the proposed method presents overall a lower error than MUSIC for both array geometries and SNR levels under the condition of sufficient angular separation between sources, which situates between $60°$ and $90°$ for $0\,\mathrm{dB}$ SNR and between $30°$ and $60°$ for $15\,\mathrm{dB}$ SNR. The resulting NMSE values per frequency bin for the interfering stationary point source PSD estimation obtained with minimum statistics (MS), minimum mean-square error estimator (MMSE) and the proposed method are shown in Fig. 3. We can observe that for all methods, the error for frequencies up to approximately $2000\,\mathrm{Hz}$ is higher than for the remaining frequency bands, most likely due to the speech signal power being mostly concentrated in lower frequency bands. In addition, due to the total observation time considered for a set of 6 different microphone orientations reaching $3\,\mathrm{s}$, we violate the assumption of signal stationarity made in the signal model in (2), since a speech signal's average stationarity window is around 20 ms [21]. Consequently, the DOA estimation can be affected by the variations in the speech PSD values between successive frames, consequently also affecting the PSD estimation of the stationary point source. Nevertheless, we can observe that overall, the proposed method still presents a higher estimation accuracy than the traditional single-channel methods considered for the given setup.

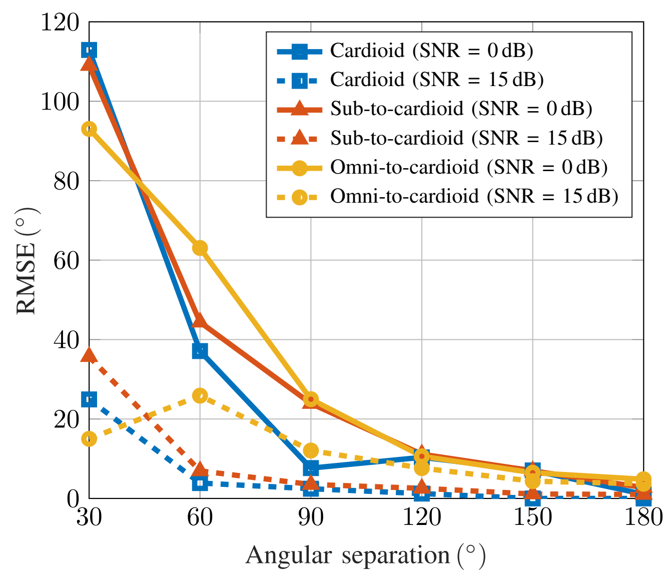

Fig. 1: RMSE of DOA estimation of two stationary point source signals for varying angular separations and different SNRs, obtained with different microphone responses.

V. Conclusion

In this paper, a method for estimating the PSD of a stationary point source while exploiting spatial information in terms of the DOA of signals arriving at a single directional microphone was proposed. With the microphone being able to record while successively facing distinct directions, a group-sparsity constrained optimization problem was solved, allowing the DOA estimation for point source signals via the identification of peaks in PSD levels over a given angular dictionary. The estimated DOAs were used for reducing the linear system of equations resulting from the proposed signal model and re-estimating the point source PSD. Simulation results showed that the proposed method of DOA estimation presents

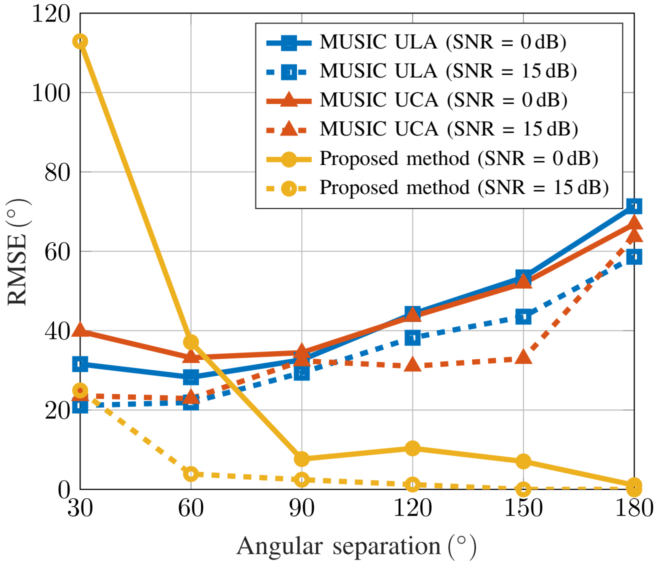

Fig. 2: RMSE of DOA estimation of two stationary point source signals for varying angular separations and different SNRs, obtained with wideband MUSIC and the proposed method.

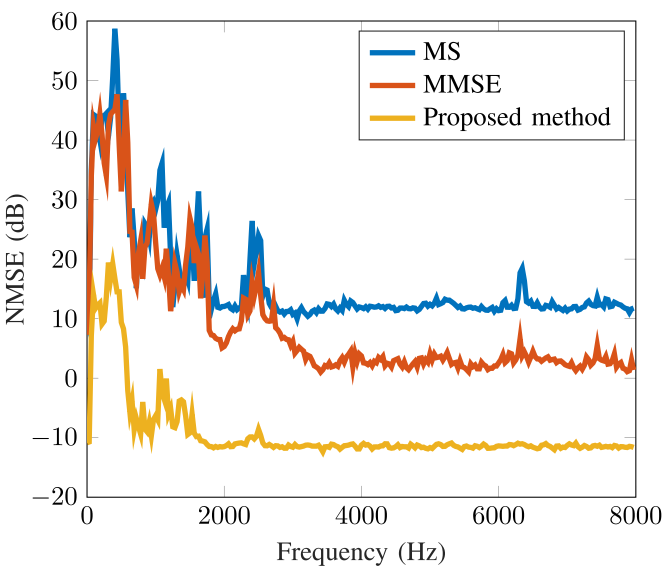

Fig. 3: NMSE of interfering point source PSD estimation per frequency, obtained with minimum statistics (MS), minimum mean-square error (MMSE) estimator and proposed method.

higher accuracy when there is a larger angular separation between sources, as well as a stronger microphone directivity over wider frequency bands. Moreover, the PSD estimation of a stationary interfering point source after estimating the DOA presented higher accuracy than traditional single-channel methods that do not exploit spatial information of audio sources from the rotating microphone recordings. Future work includes experimental tests, expanding the proposed DOA estimation method to off-grid locations by means of integrated wideband dictionaries [22], [23], and considering the presence of reverberation and speech signals into the proposed model.

References

[1] J. Benesty, J. Chen, and Y. Huang, *Microphone Array Signal Processing*, ser. Springer Topics in Signal Processing. Berlin, Heidelberg: Springer Berlin Heidelberg, 2008, vol. 1.

[2] R. C. Hendriks, T. Gerkmann, and J. Jensen, *DFT-Domain Based Single-Microphone Noise Reduction for Speech Enhancement: A Survey of the State of the Art*. Morgan & Claypool, Jan. 2013, vol. 9.

[3] A. Plinge, F. Jacob, R. Haeb-Umbach, and G. A. Fink, "Acoustic Microphone Geometry Calibration: An overview and experimental evaluation of state-of-the-art algorithms," *IEEE Signal Process. Mag.*, vol. 33, no. 4, pp. 14–29, Jul. 2016.

[4] A. Saxena and A. Y. Ng, "Learning sound location from a single microphone," in *2009 IEEE International Conference on Robotics and Automation (ICRA) Proceedings*, 2009, pp. 1737–1742.

[5] D. E. Badawy and I. Dokmanić, "Direction of Arrival with One Microphone, a few LEGOs, and Non-Negative Matrix Factorization," *IEEE/ACM Trans. Audio Speech Lang. Process.*, vol. 26, no. 12, pp. 2436–2446, Dec. 2018.

[6] T. Takiguchi, Y. Sumida, R. Takashima, and Y. Ariki, "Single-Channel Talker Localization Based on Discrimination of Acoustic Transfer Functions," *EURASIP J. Adv. Signal Process.*, vol. 2009, no. 1, p. 918404, Dec. 2009.

[7] A. Schasse, C. Tendyck, and R. Martin, "Source localization based on the doppler effect," in *2012 13th International Workshop on Acoustic Signal Enhancement (IWAENC) Proceedings*, 2012, pp. 1–4.

[8] Y. Hioka, R. Drage, T. Boag, and E. Everall, "Direction of arrival estimation using a circularly moving microphone," in *2018 16th International Workshop on Acoustic Signal Enhancement (IWAENC) Proceedings*, 2018, pp. 91–95.

[9] B. Porat, *A course in digital signal processing*. New York: John Wiley, 1997.

[10] M. Yuan and Y. Lin, "Model selection and estimation in regression with grouped variables," *J Royal Statistical Soc B*, vol. 68, no. 1, pp. 49–67, Feb. 2006.

[11] Hastie, T., Tibshirani, R., and Wainwright, M., *Statistical Learning with Sparsity: The Lasso and Generalizations*. CRC Press LLC, 2015.

[12] Bang and Olufsen, "Music for Archimedes," CD B&O 101, 1992.

[13] M. Grant and S. Boyd, "CVX: Matlab software for disciplined convex programming, version 2.1," http://cvxr.com/cvx, Mar. 2014.

[14] R. Schmidt, "Multiple emitter location and signal parameter estimation," *IEEE Trans. Antennas Propagat.*, vol. 34, no. 3, pp. 276–280, Mar. 1986.

[15] T. E. Tuncer and B. Friedlander, Eds., *Classical and modern direction-of-arrival estimation*. Boston: Academic Press, 2009.

[16] M. R. Azimi-Sadjadi, A. Pezeshki, and N. Roseveare, "Wideband DOA estimation algorithms for multiple moving sources using unattended acoustic sensors," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 44, no. 4, pp. 1585–1599, 2008.

[17] R. Martin, "Noise power spectral density estimation based on optimal smoothing and minimum statistics," *IEEE Trans. Speech Audio Process.*, vol. 9, no. 5, pp. 504–512, Jul. 2001.

[18] ——, "Bias compensation methods for minimum statistics noise power spectral density estimation," *Signal Processing*, vol. 86, no. 6, pp. 1215–1229, Jun. 2006.

[19] T. Gerkmann and R. C. Hendriks, "Unbiased MMSE-Based Noise Power Estimation With Low Complexity and Low Tracking Delay," *IEEE Trans. Audio Speech Lang. Process.*, vol. 20, no. 4, pp. 1383–1393, May 2012.

[20] M. Brookes *et al.*, "Voicebox: Speech processing toolbox for matlab," http://www.ee.ic.ac.uk/hp/staff/dmb/voicebox/voicebox.html, 1997.

[21] E. Hänsler and G. Schmidt, Eds., *Speech and audio processing in adverse environments*, ser. Springer series on signals and communication technology. Berlin: Springer, 2008.

[22] M. Butsenko, J. Sward, and A. Jakobsson, "Estimating Sparse Signals Using Integrated Wideband Dictionaries," *IEEE Trans. Signal Process.*, vol. 66, no. 16, pp. 4170–4181, Aug. 2018.

[23] M. Jälmby, J. Swärd, F. Elvander, and A. Jakobsson, "Computationally efficient estimation of multi-dimensional damped modes using sparse wideband dictionaries," in *2018 26th European Signal Processing Conference (EUSIPCO) Proceedings*, 2018, pp. 1745–1749.