On the Bias of Direction of Arrival Estimation Using Linear Microphone Arrays

Florian Jacob, Reinhold Haeb-Umbach

Department of Communications Engineering, Paderborn University, 33098 Paderborn, Germany
Email: {jacob, haeb}@nt.uni-paderborn.de
Web: nt.uni-paderborn.de

Abstract

This contribution investigates Direction of Arrival (DoA) estimation using linearly arranged microphone arrays. We are going to develop a model for the DoA estimation error in a reverberant scenario and show the existence of a bias, that is a consequence of the linear arrangement and limited field of view (FoV) that is associated with it. Two reasons are identified to cause the bias: First, the limited FoV leading to a clipping of the measurements, and, second, the angular distribution of the signal energy of the reflections being non-uniform. Since both issues are a consequence of the linear arrangement of the sensors, the bias arises largely independent of the kind of DoA estimator. The experimental evaluation demonstrates the existence of the bias for a selected number of DoA estimation methods and proves that the prediction from the developed theoretical model matches the simulation results.

1 Introduction

Microphone arrays are of great importance for many signal processing applications, such as speech recognition, speech enhancement, advanced teleconferenceing systems, distributed heading-aids and many more, since they achieve superior performance compared to a single microphone. One particular task is the estimation of the Direction of Arrival (DoA) of an impinging signal. DoA estimates are used to steer a beamformer to a given direction, perform blind source separation (BSS) [1], localize speakers [2] and events [3] or run geometry calibration algorithms [4, 5].

Since DoA estimation is a fundamental task in array signal processing, there exists a wide variety of methods. Commonly used techniques are subspace based algorithms, e. g. the multiple signal classification (MUSIC) algorithm [6] and its variants like ESPRIT [7]. Furthermore, ICA-based methods have been proposed, that can identify the delay corresponding to the impinging direction [8]. Another popular algorithm is steered response power with phase transform (SRPPhat) [9], which is the generalization of generalized cross correlation with phase transform (GCCPhat) [10] for more than two microphones. GCCPhat allows for a direct computation of the DoA using the time delay belonging to the maximum of the cross correlation function of the signals. SRPPhat evaluates the cross correlation functions of all pairs of microphones for each candidate direction, adds up the GCCPhat-scores and finally performs a maximum search to obtain an estimate of the DoA.

Another group of algorithms devises a distance measure between an observation and all possible candidate directions. [1] employs a cosine distance between the observed phase and the candidate models, i.e., the expected observations in an anechoic sound field originating from a given direction. With the cosine distance, [1] specifically accounts for spatial aliasing, which arises for all frequencies, where the sensor distance is larger than half a wave length. The approach proposed in [11] accounts for spatial aliasing, too, however the introduced distance measure was derived from a statistical model using the complex Watson distribution.

Many of the previously mentioned techniques are able to estimate the DoA, even if more than one source is prominent. Besides, [12] accomplishes this goal even in an underdetermined case, e. g., if there are fewer microphones than impinging directions to be estimated.

A performance evaluation of DoA estimators is mostly carried out using circular arrays, which consist of at least 3 microphones [11–13]. Contrary to that, applications that build on top of DoA estimates, often use linear arrays [2, 14, 15], due to physical constraints of the carrying device. Thus, this paper investigates the particular characteristics of DoA estimation based on linear microphone arrays.

In order to characterize the error of the DoA estimation process a model for a reverberant scenario will be devised. This model predicts the existence of a systematic estimation error, i.e., a bias, which is triggered by two effects, both being related to the limited field of view (FoV), that again is a consequence of the linear arrangement. The limited FoV results in a restriction of the measurements to the half-plane, which will cause a bias. A second contribution to the bias comes from the reverberation profile in a linear microphone arrangement. While the impinging directions of all reflected signals are uniformly distributed on a circle, the limited FoV causes a mapping to one half-plane (semicircle), that introduces a bias term, too.

In this contribution we applied SRPPhat, which is known to be relatively robust against reverberation, although the design did not explicitly tackle the multipath problem in a reverberant scenario. But we also observed systematic errors being present in other DoA estimation methods that only consider the a direct propagation path like [1], [11] and [16]. Thus the bias is a consequence of the linear arrangement, rather than an issue of a particular DoA estimator.

The remainder of this paper is organized as follows. Sec. 2 introduces DoA estimation using GCCPhat and SRPPhat respectively. Sec. 3 derives a model for the impact of the restriction of the Time Difference of Arrival (TDoA) measurements to the limited FoV of a linear array. Sec. 4 considers the effect of reverberation on the estimation process. Afterwards, Sec. 5 discusses why the bias vanishes if the microphones are not arranged on a line. Sec. 6 eventually shows the existence of a bias for various DoA estimators by simulations. Finally Sec. 7 concludes this paper.

2 Linear array DoA estimation

Let us consider a reverberated speech signal $s(t)$, that is captured by an array of D microphones. The received signal of the dth microphone can be expressed as

$$x_d(t) = (h_{d,\mathrm{LOS}}(t) + h_{d,\mathrm{rev}}(t)) * s(t) + n(t), \qquad (1)$$

where $h_{d,\mathrm{LOS}}(t)$ covers the delay and attenuation introduced by the line-of-sight (LOS) component of the room impulse response (RIR), $h_{d,\mathrm{rev}}(t)$ comprises of the reverberation and $n(t)$ models the sensor noise. This time domain model can be approximated in the short-time Fourier transform (STFT) domain by a multiplicative transfer function

$$X_d(m,k) = H_d(m,k) \cdot S(m,k) + \tilde{N}_d(m,k), \qquad (2)$$

where m denotes the frame and k the frequency bin index. Here, $X_d(m,k)$ is the STFT of the microphone signal, $H_d(m,k)$ corresponds to the transfer function of the LOS propagation path and $S(m,k)$ is STFT of the speech signal. The term $\tilde{N}_d(m,k)$ covers the sensor noise and those parts of the microphone signal $x_d(t)$ that are caused by $h_{d,\mathrm{rev}}(t)$.

In the following we will recap the concept of GCCPhat [10] for the two-element microphone array and its extension SRPPhat, that can handle larger configurations. At first GCCPhat estimates the time delay between the signals of the microphones d and l. In order to do so, the cross power spectrum weighted by the energy is processed by the inverse discrete Fourier transform (IDFT), resulting in

$$\mathrm{phat}_{l,d}(m,\lambda) = \mathrm{IDFT}\left(\frac{X_l(m,k)X_d^*(m,k)}{|X_l(m,k)|\,|X_d(m,k)|}\right), \qquad (3)$$

where λ denotes the signal delay in samples. Omitting the index of the STFT frame m, the TDoA is recovered by

$$\hat{\tau}_{l,d} = \underset{\lambda}{\mathrm{argmax}}\left\{\ \mathrm{phat}_{l,d}(\lambda)\ \right\} \cdot f_\mathrm{s}^{-1}, \qquad (4)$$

where f_s is the sampling rate. In practice the sampling rate is often too low to achieve high-resolution estimates, thus an interpolation is required to obtain precise estimates [17, 18].

To reveal the desired DoA from the TDoA the sound source is assumed to be located in the far field of the array. Using the far field assumption the wave front impinging on the array can be approximated by a plane wave. Fig. 1 shows this situation for a two-element microphone array.

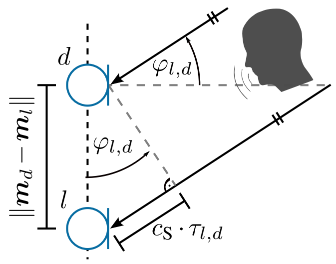

Figure 1: DoA estimation based on the far field assumption for two-element microphone array.

Now, the DoA with respect to the broadside of the array consisting of the microphones d and l, located at the positions $\boldsymbol{m}_d$ and $\boldsymbol{m}_l$ respectively, can be obtained via

$$\hat{\varphi}_{l,d} = \arcsin(\hat{\tau}_{l,d}/\tau_{\max}) \text{ with } \tau_{\max} = \|\boldsymbol{m}_l - \boldsymbol{m}_d\| \cdot c_\mathrm{S}^{-1}, \qquad (5)$$

where c_S denotes the speed of sound. For SRPPhat, that builds on top of GCCPhat, Eq. (5) can be omitted. Instead, SRPPhat computes the expected far field TDoA for the microphone pair d, l for each candidate direction φ by

$$\tilde{\tau}_{l,d}(\varphi) = \begin{bmatrix}\cos(\varphi)\\ \sin(\varphi)\end{bmatrix}^\mathrm{T} \cdot (\boldsymbol{m}_l - \boldsymbol{m}_d) \cdot c_\mathrm{S}^{-1} \qquad (6)$$

and afterwards evaluates the corresponding score

$$P_\mathrm{SRPPhat}(\varphi) = \sum_{d=1}^{D} \sum_{l=d+1}^{D} \mathrm{phat}_{l,d}(\tilde{\tau}_{l,d}(\varphi) \cdot f_\mathrm{s}). \qquad (7)$$

Here, D denotes the total number of microphones. Finally, a maximum search of Eq. (7) reveals the DoA estimate $\hat{\varphi}$.

3 Model for TDoA clipping

As indicated above, TDoA estimates are the basis of the DoA estimation. Thus, the precision of the DoA depends on the quality of the corresponding TDoA. The quality of the TDoA estimation itself is influenced by many factors. In principle, the TDoA cannot be larger than $\tau_{\max}$ (cf. Eq. (5)):

$$-\tau_{\max} \leq \tau \leq \tau_{\max}. \qquad (8)$$

In practice, however, due to reverberation, noise or other causes, the TDoA estimate, as computed, e.g., by Eq. (4), can very well exceed these bounds.

The resulting TDoA estimation error is often assumed to follow a zero-mean normal distribution [19]. Hence, the TDoA estimate $\hat{\tau}$ can be modeled by the normal distribution $\mathcal{N}(\hat{\tau};\tau,\sigma_\tau^2)$, where τ indicates the ground truth delay and σ_τ the standard deviation. Thus TDoA estimate is no longer constrained to the interval given in Eq. (8). Nevertheless, the estimate needs to be bound to this interval to be able to compute the DoA based on Eq. (5). Therefore, the estimate $\hat{\tau}$ given by Eq. (4) needs to be clipped:

$$\hat{\tau}_\mathrm{c} = \max(-\tau_{\max},\ \min(\hat{\tau}, \tau_{\max})). \qquad (9)$$

This clipping affects the distribution of the TDoA, and turns the normal distribution into

$$\begin{aligned}\mathrm{p}_{\hat{\tau}_\mathrm{c}}(\hat{\tau}_\mathrm{c}) = & c_1 \cdot \delta(\hat{\tau}_\mathrm{c} + \tau_{\max}) + c_2 \cdot \delta(\hat{\tau}_\mathrm{c} - \tau_{\max}) \qquad (10)\\ & + \mathrm{rect}(\hat{\tau}_\mathrm{c}/(2 \cdot \tau_{\max})) \cdot \mathcal{N}(\hat{\tau}_\mathrm{c};\tau,\sigma_\tau^2),\end{aligned}$$

since the parts of the probability mass that fall outside $\pm\tau_{\max}$ are mapped to the borders of the interval. In Eq. (10), $\delta(\cdot)$ denotes the Dirac-Delta impulse and $\mathrm{rect}(\cdot)$ is a rectangular function with unit width. The constants c_1 and c_2 contain the clipped mass of the normal distribution and they are given by the Q-function [20], which is complementary cumulative distribution function of the standard normal distribution:

$$c_1 = 1 - \mathrm{Q}\left(\frac{-\tau_{\max} - \tau}{\sigma_\tau}\right) \text{ and } c_2 = \mathrm{Q}\left(\frac{\tau_{\max} - \tau}{\sigma_\tau}\right). \qquad (11)$$

An exaggerated example of the resulting distribution $\mathrm{p}_{\hat{\tau}_\mathrm{c}}(\hat{\tau}_\mathrm{c})$ is shown in Fig. 2a.

The impact of the clipping on the DoA estimate can be derived from a random variable transformation. Propagating the probability density function (PDF) from Eq. (10) through the non-linearity given in Eq. (5) leads to:

$$\begin{aligned}\mathrm{p}_{\hat{\varphi}}(\hat{\varphi}) = & c_1 \cdot \delta(\hat{\varphi} + \pi/2) + c_2 \cdot \delta(\hat{\varphi} - \pi/2) \qquad (12)\\ & + \tau_{\max} \cdot \mathcal{N}(\sin(\hat{\varphi}) \cdot \tau_{\max};\tau,\sigma_\tau^2) \cdot \cos(\hat{\varphi}).\end{aligned}$$

Due to the limited FoV of linearly arranged microphones the angle $\hat{\varphi}$ is limited to the interval $[-\pi/2;\pi/2]$. A visualization of the resulting distribution is shown in Fig. 2b. The PDFs shown in Fig. 2 indicate the existence of a bias. From Fig. 2b it can be seen that the mean of the distribution $\mathrm{p}_{\hat{\varphi}}(\hat{\varphi})$ is not equal to the ground truth direction φ anymore.

The bias that is a consequence of the clipping and the nonlinear transformation from TDoA to DoA is given by

$$b = E\left[\mathrm{p}_{\hat{\varphi}}(\hat{\varphi})\right] - \varphi = \int_{-\frac{\pi}{2}}^{\frac{\pi}{2}} \hat{\varphi} \cdot \mathrm{p}_{\hat{\varphi}}(\hat{\varphi})\,\mathrm{d}\hat{\varphi} - \varphi, \qquad (13)$$

where $E[\cdot]$ denotes the expectation operator. Please note that the bias b depends on the ground truth direction, since $\mathrm{p}_{\hat{\varphi}}(\hat{\varphi})$ is a function of delay τ of the ground truth direction. A numerical evaluation of the integral leads to the bias depicted in Fig. 3. For a presentation, which is independent of a particular microphone distance, the standard deviation σ_τ of the TDoA is measured in terms of a fraction of $\tau_{\max}$.

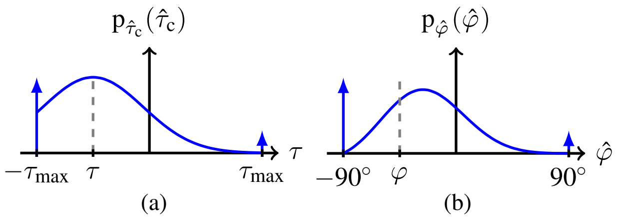

Figure 2: PDF of the clipped TDoA (a) and of the corresponding DoA (b).

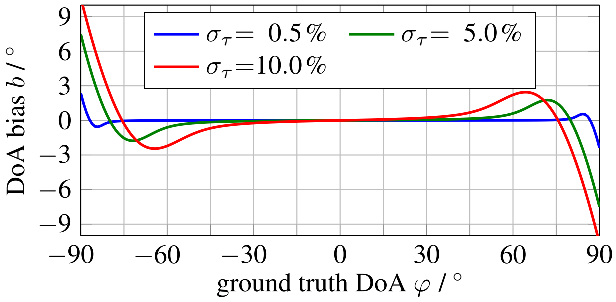

Figure 3: DoA bias caused by the clipping of the normally distributed TDoA measurements.

4 Impact of reverberation

The previous section outlined the impact of the clipping that is necessary due to the limited FoV of a linearly arranged microphone array. This section covers the effect of reverberation on the DoA estimation process.

In statistical room acoustics the phase of the plane waves representing late reverberation is modelled as a random process. It is assumed to be uniformly distributed, if the boundary conditions given in [21, 22] are met. Taking into account the uniform distribution of the phase, the distribution of the impinging directions of the late reverberation follow a uniform distribution, too and thus the expected value of the corresponding directions evaluates to zero.

However, the isotropic model takes only the late reverberation into account and neglects the impact of the early reflections. Furthermore the amplitudes of the reflections matter, since, the the first- and second-order reflections exhibit higher energy than higher-order reflections, and therefore have a larger effect on the DoA. Thus, a realistic model needs to consider the amplitudes as well.

Moreover the limited FoV of a linear array makes it impossible to distinguish between a sound wave impinging from the front-half-plane or the rear-half-plane (cf. Fig. 1). This ambiguity is equivalent to a mapping of the global DoA φ to a DoA of the linear array

$$\varphi_{\text{Linear}} = \begin{cases} \varphi & \text{for } |\varphi| \leq +\frac{\pi}{2} \\ \operatorname{sgn}(\varphi)\cdot\pi - \varphi & \text{otherwise} \end{cases}, \tag{14}$$

where the function $\operatorname{sgn}(\varphi)$ extracts the sign of φ.

The overall distribution of the DoA in a reverberant scenario $p'_{\hat{\varphi}}(\hat{\varphi})$ is the weighted superposition of the PDF of LOS-component (cf. Eq. (12)) and the PDF of the reflections $p_{\text{rev}}(\hat{\varphi})$. In order to estimate $E[p'_{\hat{\varphi}}(\hat{\varphi})]$ we incorporate the RIR. The RIR from the source to the dth microphone can be modeled by a weighted sum of $R = R_{\max}$ time-delayed impulses

$$h_{d,\text{LOS}}(t) + h_{d,\text{rev}}(t) = \sum_{r=0}^{R} \beta_{d,r}\delta(t - \xi_{d,r}), \tag{15}$$

where $\beta_{d,r}$ describes the absorption on the rth propagation path, while $\xi_{d,r}$ denotes the delay of this path relative to the LOS component ($r = 0$).

If the delays $\xi_{d,r}$ and $\xi_{l,q}$ that are corresponding to the RIR to the dth and lth microphone respectively as well as the conforming absorptions coefficients $\beta_{d,r}$ and $\beta_{l,q}$ are known, the average TDoA of the RIR can be computed by

$$\hat{\tau}^{\text{RIR}}_{d,l} = \sum_{r=1}^{R}\sum_{q=1}^{R} \beta_{d,r}\beta_{l,q}(\xi_{d,r} - \xi_{l,q}) \bigg/ \left(\sum_{r=1}^{R}\sum_{q=1}^{R} \beta_{d,r}\beta_{l,q}\right). \tag{16}$$

Using Eq. (5), the average DoA $\hat{\varphi}_{\text{RIR}}$ of the RIR can be obtained from $\hat{\tau}^{\text{RIR}}_{d,l}$. This DoA can be considered an approximation of $E[p'_{\hat{\varphi}}(\hat{\varphi})]$, which incorporates the effect of reverberation, while neglecting the clipping.

The computation of Eq. (16) requires the knowledge of the RIR. Hence the image method (IM) [23] was employed. It computes the parameters of the model introduced in Eq. (15) by mirroring the source position at the walls of the enclosure (see Fig. 5). The distance from the image position to the microphone leads to the delays $\xi_{d,r}$ and $\xi_{l,q}$, while parameters $\beta_{d,r}$ and $\beta_{l,q}$ depend on the number of walls the signal component was reflected from.

A set of random scenarios, which are described in Sec. 6, was simulated to estimate $E[p'_{\hat{\varphi}}(\hat{\varphi})]$ via Eq. (16). The results are depicted in Fig. 4 as a function of the ground truth DoA, on the hand side considering the complete RIR ($R = R_{\max}$) and on the other hand only for the first- and second-order reflections ($R = 2$). It shows that the estimated DoA tends to have a smaller absolute value than the true DoA. Moreover, the illustrated bias depends on both the ground truth DoA and the reverberation time. The significant impact of early reflections can be seen in case of $T_{60} = 0.2\,\text{s}$, since the bias of $R = 2$ equals the result if the complete RIR is considered ($R = R_{\max}$). If T_{60} increases a difference can be seen, however, it vanishes for $R = 5$.

A plausibility argument for the existence of this bias can be given by looking at the geometric configuration shown in Fig. 5. Due to the linear arrangement the FoV is limited to 180°, thus the image source in left mirror room is mapped to an equivalent position in the lower right corner (unfilled head). If the four image sources in front of the microphone array are considered, it can be seen that three of them cause a delay, that is smaller than the delay of the LOS component and only one has a larger delay. This irregular distribution results in the bias.

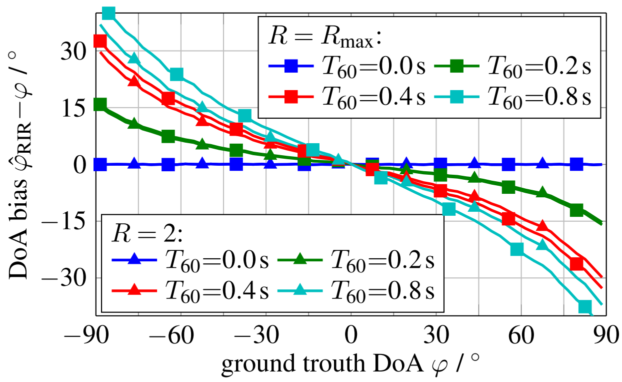

Figure 4: Bias of the DoA estimation introduced by the reflections, either considering the complete RIR ($R = R_{\max}$) or first- and second-order reflections ($R = 2$).

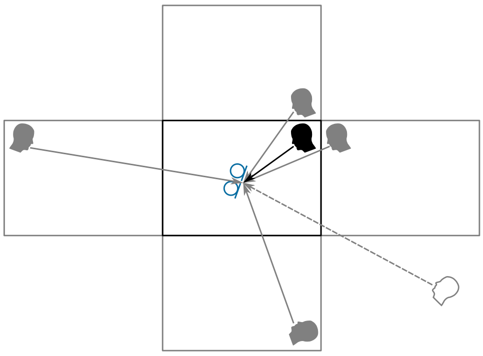

Figure 5: Microphone array (blue), that is capturing the signal from the source (black) and the first order images (grey). The unfilled head indicates the equivalent position of the image source from the left side for a linear array.

5 Microphones not arranged in a line

So far we showed two factors contributing to a bias if a two-element microphone array was used for DoA estimation. On the one hand, the limited FoV requires a clipping of the TDoA measurements at the borders of the FoV, which results in biased estimates in particular for large ground truth DoA values. On the other hand the mapping of the reverberation into one half-plane (see Eq. (14)) causes a bias that strongly depends on the reverberation time.

Please note that both effects are not a peculiarity of SRPPhat, in fact they are a consequence of the linear arrangement. If a microphone array consists of more than two microphones, that are still arranged on a line, both effects are still present and the bias will remain. Furthermore, many other estimators, like [1, 11], also constrain their search space in case of linearly arranged arrays, which is similar to the clipping and will therefore result in a bias, too. The main contribution of the bias is caused by reflections and thus is also not related to a particular estimator.

Many estimators, especially those who are evaluated in Sec. 6, claim to be robust to reverberation, but they do not explicitly take reverberation into account. A possible solution to avoid the reverberation-related bias might be the usage of estimators that explicitly model reflections, e.g. by estimating the DoA of the direct path as well as the DoA of the early reflections [24]. However, the only reason for the bias is the linear arrangement. If the microphones are not arranged on a line, the FoV increases to 360°. Therefore, the clipping (cf. Sec. 3) is not needed anymore, and also the mapping (cf. Eq. (14)) will not be present. Reconsidering Fig. 5 the first order reflections are given by the original image sources (gray heads) and the average direction need to be computed by taking the circular property of the directions into account. Thus the reflections are approximately uniformly distributed and the bias vanishes completely.

6 Simulations

The proposed model for the bias of linear arrays is evaluated through simulations. The speech signals are generated using samples from the TIMIT database [25], that are convolved with RIRs created by our own implementation of the IM [23]. In order to mitigate artifacts introduced by particular geometric configurations we sampled the room size as well as the microphone location from a uniform distribution. The room size varies between $4.0 \times 4.0\,\mathrm{m}^2$ and $8.0 \times 8.0\,\mathrm{m}^2$, while the height is 3.0 m. The arrays are located somewhere in a square of 0.5 m edge length at the room center and the array consists of 2 microphones arranged at a distance of 0.05 m. To obtain the results presented in the following 10 000 random configurations were simulated. The bias of SRPPhat measured in the simulations is shown in Fig. 6.

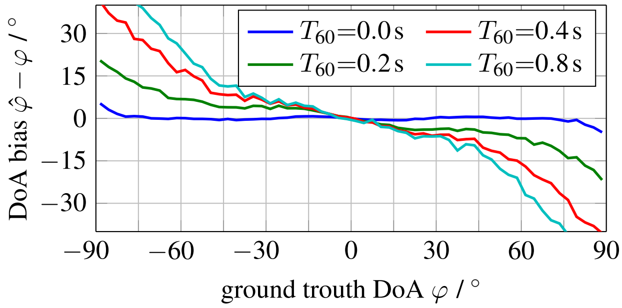

Figure 6: DoA estimation bias of SRPPhat using a linear two-element microphone array.

The similarity with the reverberation-related part of the bias (cf. Fig. 4) can clearly be seen. The impact of the clipping-related bias is much smaller, but especially for T_{60}=0.0 s it is present. The deviation between the model and the simulations can be explained by the assumed normal PDF of the TDoA error, which may not hold in practice.

Fig. 7 demonstrates that the observed bias is present also for other DoA estimators. For all four estimators shown, the bias was approximately the same.

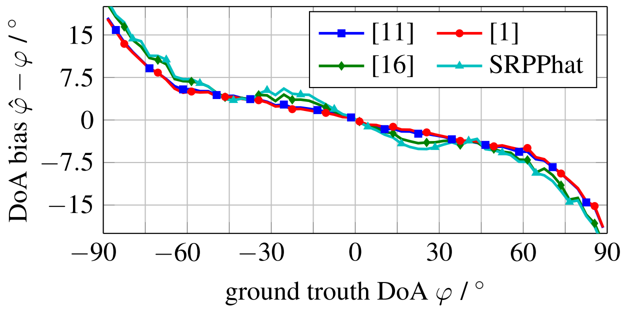

Figure 7: Bias of selected DoA estimators ($T_{60} = 0.2\,\mathrm{s}$).

7 Conclusions

In this paper we investigated DoA estimation using a linear microphone array. It turned out that the linear arrangement, which causes a limited FoV, is the reason for a significant bias, that depends on the impinging direction. On the one hand side the limited FoV asks for a clipping of the TDoA estimates, resulting in a small bias, which is mostly relevant for signals arriving from the end-fire directions. Moreover, the arrangement leads to a mapping of the DoA of the reflections to one half-plane. This mapping results in an underestimation of the DoA, that has the largest impact at the end-fire directions and vanishes for signals coming from broadside. Since the bias is a consequence of the linear arrangement it is not an issue a particular DoA estimator. However, the bias can be prevented if the microphones are not arranged on a line. If the estimator explicitly accounts for reverberation, as in recent work of [24], at least the second, dominant, cause of bias, can be avoided. Here, the bias has been estimated from the RIR. Thus, we are, unfortunately, not able to predict (and thus compensate) the bias, unless we have knowledge of the RIR.

References

[1] B. Loesch and B. Yang, "Blind source separation based on time-frequency sparseness in the presence of spatial aliasing<," in *Latent Variable Analysis and Signal Separation*, pp. 1–8, Springer Berlin Heidelberg, 2010.

[2] T. Nishiura, S. Nakamura, and K. Shikano, "Talker Localization in a Real Acoustic Environment Based on DoA Estimation and Statistical Sound Source Identification," in *IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP 2002)*, vol. 1, pp. 893–896, May 2002.

[3] R. Levorato and E. Pagello, "DOA Acoustic Source Localization in Mobile Robot Sensor Networks," in *IEEE International Conference on Autonomous Robot Systems and Competitions (ICARSC 2015)*, pp. 71–76, Apr. 2015.

[4] 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 Processing Magazine*, vol. 33, pp. 14–29, July 2016.

[5] F. Jacob, J. Schmalenstroeer, and R. Haeb-Umbach, "DOA-based Microphone Array Position Self-calibration Using Circular Statistics," in *IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP 2013)*, pp. 116–120, May 2013.

[6] R. Schmidt, "Multiple emitter location and signal parameter estimation," *IEEE Transactions on Antennas and Propagation*, vol. 34, pp. 276–280, Mar. 1986.

[7] R. Roy and T. Kailath, "ESPRIT-estimation of signal parameters via rotational invariance techniques," *IEEE Transactions on Acoustics, Speech, and Signal Processing*, vol. 37, pp. 984–995, July 1989.

[8] H. Sawada, R. Mukai, and S. Makino, "Direction of arrival estimation for multiple source signals using independent component analysis," in *International Symposium on Signal Processing and its Applications (ISSPA 2003)*, vol. 2, pp. 411–414 vol.2, July 2003.

[9] J. DiBiase, H. Silverman, and M. Brandstein, "Robust Localization in Reverberant Rooms," in *Microphone Arrays* (M. Brandstein and D. Ward, eds.), Digital Signal Processing, pp. 157–180, Springer Berlin Heidelberg, 2001.

[10] C. Knapp and G. Carter, "The generalized correlation method for estimation of time delay," *IEEE Transactions on Acoustics, Speech, and Signal Processing*, vol. 24, pp. 320–327, Aug. 1976.

[11] L. Drude, F. Jacob, and R. Hab-Umbach, "DoA-Estimation based on a Complex Watson Kernel Method," in *European Signal Processing Conference (EUSIPCO 2015)*, pp. 255–259, Aug. 2015.

[12] S. Araki, H. Sawada, R. Mukai, and S. Makino, "DOA estimation for multiple sparse sources with arbitrarily arranged multiple sensors," *Journal of Signal Processing Systems*, vol. 63, pp. 265–275, June 2011.

[13] A. Karbasi and A. Sugiyama, "A new DOA estimation method using a circular microphone array," in *European Signal Processing Conference (EUSIPCO 2007)*, pp. 778–782, Sept. 2007.

[14] O. Thiergart, M. Taseska, and E. A. P. Habets, "An informed MMSE filter based on multiple instantaneous direction-of-arrival estimates," in *European Signal Processing Conference (EUSIPCO 2013)*, pp. 1–5, Sept. 2013.

[15] J. Schmalenstroeer, F. Jacob, R. Haeb-Umbach, M. Hennecke, and G. Fink, "Unsupervised Geometry Calibration of Acoustic Sensor Networks Using Source Correspondences," in *Interspeech 2011*, pp. 597–600, Aug. 2011.

[16] J. Schmalenstroeer and R. Haeb-Umbach, "Online Diarization of Streaming Audio-Visual Data for Smart Environments," *IEEE Journal of Selected Topics in Signal Processing*, vol. 4, pp. 845–856, Oct. 2010.

[17] S. Tervo and T. Lokki, "Interpolation Methods for the SRP-PHAT Algorithm," in *International Workshop on Acoustic Echo and Noise Control (IWAENC 2008)*, 2008.

[18] X. Lai and H. Torp, "Interpolation Methods for Time-Delay Estimation Using Cross-Correlation Method for Blood Velocity Measurement," *IEEE Transactions on Ultrasonics, Ferroelectrics, and Frequency Control*, vol. 46, no. 2, pp. 277–290, 1999.

[19] Y. Oualil, F. Faubel, M. Doss, and D. Klakow, "A TDOA Gaussian mixture model for improving acoustic source tracking," in *European Signal Processing Conference (EUSIPCO 2012)*, pp. 1339–1343, Aug. 2012.

[20] S. Kay, *Intuitive probability and random processes using MATLAB®*. Springer Science & Business Media, 2006.

[21] T. C. Lawin-Ore and S. Doclo, "Using statistical room acoustics for analysing the output SNR of the MWF in acoustic sensor networks," in *European Signal Processing Conference (EUSIPCO 2012)*, pp. 1259–1263, Aug. 2012.

[22] M. R. Schroeder, "Frequency-Correlation Functions of Frequency Responses in Rooms," *The Journal of the Acoustical Society of America*, vol. 34, no. 12, pp. 1819–1823, 1962.

[23] J. B. Allen and D. A. Berkley, "Image method for efficiently simulating small-room acoustics," *The Journal of the Acoustical Society of America*, vol. 65, pp. 943–950, Apr. 1979.

[24] J. R. Jensen, J. K. Nielsen, R. Heusdens, and M. G. Christensen, "DoA Estimation of Audio Sources in Reverberant Environments," in *IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP 2016)*, pp. 176–180, March 2016.

[25] J. S. Garofolo, L. D. Consortium, *et al.*, *TIMIT: acoustic-phonetic continuous speech corpus*. Linguistic Data Consortium, 1993.