Clock-Offset and Microphone Gain Mismatch Invariant Beamforming

Sofia-Eirini Kotti
TNO
The Hague, The Netherlands
sofia-eirini.kotti@tno.nl

Richard Heusdens
Delft University of Technology/NLDA
Delft, The Netherlands
r.heusdens@tudelft.nl

Richard C. Hendriks
Delft University of Technology
Delft, The Netherlands
r.c.hendriks@tudelft.nl

***Abstract*—The use of wireless acoustic sensor networks (WASNs) has received increased attention over the last decade. The advantages of WASNs over stand-alone multi-microphone devices are that the microphone array is not anymore limited by the dimensions of a single device, and that microphones can be placed at arbitrary locations. One of the disadvantages, however, is that for many applications, like beamforming, the clocks of all devices in the network need to be synchronised and that the microphone gains need to be equalised. In this paper we will prove that a specific class of beamformers is clock-offset and gain mismatch invariant. The parameters for these beamformers (acoustic transfer function and power spectral density matrices) can be estimated directly from the uncalibrated microphone signals, instead of first synchronising the clocks and equalising the gains and then estimating them. The resulting beamformers are applied to the non-calibrated microphone signals. We will substantiate, by means of computer simulations, that the proposed approach gives identical results compared to the setup where microphone signals are first calibrated, so that clock-offset compensation and microphone gain equalisation becomes unnecessary.**

***Index Terms*—Beamforming, clock synchronisation, microphone gain equalisation, wireless acoustic sensor networks**

I. Introduction

Over the last years we have seen a clear shift in research focus from stand-alone multi-microphone noise-reduction algorithms towards noise-reduction algorithms for wireless acoustic sensor networks (WASNs). In WASNs, multiple devices, each equipped with one or multiple microphones, can collaborate by sharing their microphone recordings. Depending on the setup, the calculations can be done centralised using a fusion center, or distributed [1], [2], [3], [4], [5]. The advantages of WASNs over stand-alone multi-microphone devices are that the microphone array is not anymore limited by the dimensions of a single device, and that microphones can be placed at arbitrary positions. However, despite their advantages, WASNs also come with new challenges. Among these is the fact that, in general, the clocks of the devices in the network are not synchronised and the microphone gains are different.

Gain equalisation was already an issue in conventional microphone arrays, but becomes more prominent in heterogeneous WASNs. Typically, gain equalisation is done by active compensation of the microphone gain differences. To estimate these gain differences, many approaches use acoustic test signals in combination with either time difference of arrival [6] or direction of arrival estimation [7]. These approaches thus require additional processing and information like sensor positions or emission time of the calibration signal.

The origin of the clock synchronization problem is twofold. Firstly, each device in the network is turned on at a different moment in time, leading to *clock offset*. Secondly, there might be a sampling-rate mismatch between devices, leading to *clock skew*. As multi-microphone noise-reduction methods heavily rely on the differences in arrival time of the acoustic sources at the microphones, performance of such algorithms will substantially degrade when clock skew and/or clock offset is present. A commonly applied strategy to overcome the above mentioned issues is to estimate the clock skew and clock offset and use these estimates to synchronise the clocks. There are different strategies to estimate the clock parameters known from literature. In [8], [9], [10], the internal clocks are synchronised to a reference or virtual clock by exchanging a series of time stamps. In [11], [12], the clock parameters are estimated by correlating calibration signals with a known reference signal, while in [13], [14], [15], [16], the parameters are estimated by exchanging the recorded audio signals. Once the clocks are synchronised and the gains equalised, the typical approach is to estimate the required beamformer parameters (acoustic transfer function (ATF) and power spectral density matrices) from the noisy recordings, after which the beamformer can be applied to the noisy microphone signals.

Although both clock offset and clock skew are detrimental for the performance of the beamformer, the clock skew can be argued to be of minor importance as devices that are of the same type typically have extremely accurate sampling rates compared to the precision required for beamforming applications [15]. In addition, if we read-out the buffers collecting all incoming data at regular time instances (instead of collecting a fixed number of samples per batch), the clock skew will *not* aggregate; we will only introduce buffer under- or overflows. In the case of an overflow (too many samples) we simply ignore the last incoming sample. In the case of a buffer underflow, we zero-pad the data which results in a frequency-domain interpolation of the buffered data. Clock offsets, on the other hand, are inevitable since they occur as a consequence of different onset times of the devices, or due to different internal sensor delays.

Assuming that clock-skew is negligible within a single analysis frame (typically 20 - 30 ms), we will focus on clock-offset compensation and microphone gain equalisation only. In this paper, we will show that a specific class of beamformers, the so-called low-rank multichannel Wiener filters [17], is clock-offset and gain mismatch invariant so that there is no need for clock synchronisation and gain equalisation. Instead, the beamformer parameters are estimated from the non-calibrated microphone signals, and the resulting beamformers are applied directly to the uncalibrated microphone signals. We will show, by means of computer simulations, that the proposed approach will lead to similar performance compared to the setup where microphone signals are first calibrated, making clock-offset compensation and microphone gain equalisation unnecessary.

II. Preliminaries

Given $A, B \in \mathbb{C}^{m\times m}$, the generalised eigenvalue problem is the problem of finding a nonzero vector $u \in \mathbb{C}^m$ and associated scalar $\lambda \in \mathbb{C}$ such that $Au = \lambda Bu$. The pair (λ, u) is called an eigenpair of the (linear) matrix pencil (A, B). In many practical problems, like the one considered here, the matrix pencil (A, B) is Hermitian definite. That is, A, B are Hermitian and $B \succ 0$ (positive definite). Given the Hermitian-definite matrix pencil (A, B), there exists a non-singular $U = (u_1, \ldots, u_m), u_i \in \mathbb{C}^m$, such that

$$U^H AU = \mathrm{diag}(a_1, \ldots, a_m) \text{ and } U^H BU = \mathrm{diag}(b_1, \ldots, b_m),$$

where the superscript $(\cdot)^H$ denotes matrix conjugate-transposition. Moreover, $Au_i = \lambda_i Bu_i$ for $i = 1, \ldots, m$, where $\lambda_i = a_i/b_i \geq 0$. See [18, Corollary 8.7.2]. This decomposition is known as the *generalised eigenvalue decomposition* (GEVD). We will refer to the vectors u_i as the generalised eigenvectors. Since $B \succ 0$, we have that $B^{-1}Au_i = \lambda_i u_i$ and we conclude that the eigenpairs (λ_i, u_i) are the right eigenpairs of $B^{-1}A$. The vectors u_i do *not* constitute an orthogonal basis for $\mathbb{C}^m$ since $B^{-1}A$ is not Hermitian in general. However, $B^{-1}A = B^{-1/2}SB^{1/2}$ with $S = B^{-1/2}AB^{-1/2}$ Hermitian and $B^{1/2}$ is the unique Hermitian square-root of B, and we conclude that $B^{-1}A$ is similar to a Hermitian matrix and, therefore, has real nonnegative eigenvalues.

Consider an array of m microphones and let the received signal at microphone i, say $y_i(\omega)$ where ω represents the angular frequency, be given by

$$y_i(\omega) = x_i(\omega) + v_i(\omega),$$

where $x_i(\omega)$ and $v_i(\omega)$ are the received target and noise signal[1], respectively, at microphone i. In order to improve the readability, we will drop the frequency variable ω.

Let $w \in \mathbb{C}^m$ denote a beamformer (spatial filter). Stacking the received microphone signals y_i in a vector $y = (y_1, \ldots, y_m)^T \in \mathbb{C}^m$, where the superscript $(\cdot)^T$ denotes matrix transposition, and similarly for the signals x_i and v_i, the beamformer output is given by

$$w^H y = w^H x + w^H v.$$

[1]The noise signal consists of all signals except for the target. This includes microphone self-noise, interferers, background noise, etc.

We will consider the signals to be realisations of zero-mean wide-sense stationary processes, the latter being denoted by the corresponding capital letter. In order to design beamformers, it is convenient to exploit the statistical characteristics of both the target and noise signals. Assuming the noise and target are uncorrelated, the cross-power spectral density (CPSD) matrix of the received process Y is given by

$$R_Y = R_X + R_V,$$

where $R_Y = \mathrm{E}(YY^H)$ and R_X and R_V are defined similarly. The operator $\mathrm{E}(\cdot)$ denotes the expectation operator. Applying the GEVD to the pencil (R_X, R_V) and setting $b_i = 1$ for all i, we have[2]

$$U^H R_X U = \Lambda \text{ and } U^H R_V U = I_m,$$

where $\Lambda = \mathrm{diag}(\lambda_1, \ldots, \lambda_m)$ and I_m is the $m \times m$ identitiy matrix. Since $R_Y = R_X + R_V$, we conclude that

$$U^H R_Y U = \Lambda + I_m. \tag{1}$$

Equation (1) is of practical importance, since it shows that if the pair (λ, u) is an eigenpair of the matrix pencil (R_X, R_V), then $(\lambda+1, u)$ is an eigenpair of the pencil (R_Y, R_V). Hence, in practical applications where we do not have access to R_X directly, we can estimate R_Y and R_V based on observed data, and compute the GEVD using these estimates.

III. Optimal beamformers

Consider the mean squared-error (MSE) between the beamformer output and the desired target signal at a particular reference microphone i. Without loss of generality we will assume $i = 1$. With this we have

$$\begin{aligned}\mathrm{E}|w^H Y - X_1|^2 &= \mathrm{E}|w^H X + w^H V - X_1|^2 \\ &= \mathrm{E}|w^H X - X_1|^2 + \mathrm{E}|w^H V|^2,\end{aligned}$$

where we used the property $\mathrm{E}(XV^H) = 0$. The term $\mathrm{E}|w^H X - X_1|^2$ represents the signal distortion, whereas the term $\mathrm{E}|w^H V|^2$ represents the residual noise variance. We can compromise between signal distortion and noise reduction by defining the constrained optimisation problem [19], [20], [21]

$$\begin{aligned}&\text{minimise} && \mathrm{E}|w^H X - X_1|^2 \\ &\text{subject to} && \mathrm{E}|w^H V|^2 \leq c,\end{aligned} \tag{2}$$

where $0 \leq c \leq \sigma^2_{V_1}$, and $\sigma^2_{V_1}$ is the noise variance at the reference microphone before beamforming.

In order to find the expressions for the different beamformers, we express the beamformer weights in terms of the generalised eigenvectors. That is, we have $w = Ua$ with $a \in \mathbb{C}^m$. Solving the constrained optimisation problem (2), the (unique) a-minimiser a^* is given by [22], [23]

$$a^* = (\Lambda + \mu I_m)^{-1} U^H R_X e_1,$$

and thus

$$w^* = U(\Lambda + \mu I_m)^{-1} U^H R_X e_1, \tag{3}$$

[2]The choice $b_i = 1$ implies that we normalise the (right) generalised eigenvectors such that $u_i^H R_V u_i = 1$.

where $\mu \geq 0$ is a Lagrange multiplier chosen such that[3] $a^H a = c$. The filters thus obtained are referred to as the speech-distortion weighted multichannel Wiener filter (SDW-MWF) [22], [23].

In many applications the rank of R_X is assumed to be lower than m. For example, in a free-field single speech source scenario, we have $\text{rank}(R_X) = 1$. However, in practical applications $\text{rank}(\hat{R}_X) > 1$ due to all kind of disturbances like microphone self-noise, estimation errors in R_Y and R_V, etc. In those cases we would like to replace R_X by a low-rank approximation of it. Following [24], [25], [17], we can find a low-rank approximation of R_X based on the GEVD of the matrix pencil (R_X, R_V). As mentioned before, we do not have access to R_X directly but we can compute the eigenpairs based on R_Y and R_V. Let $U^{-H} = Q = (q_1, \ldots, q_m)$, $q_i \in \mathbb{C}^m$. With this, we can express R_X as

$$R_X = Q\Lambda Q^H = \sum_{i=1}^{m} \lambda_i q_i q_i^H.$$

Note that $Q^H R_V^{-1} R_X = \Lambda Q^H$ and we conclude that $q_1, \ldots, q_m$ are the left eigenvectors of $R_V^{-1} R_X$. A rank-r approximation of R_X can then be computed by selecting the first r left eigenvectors. That is,

$$\hat{R}_X = Q_r \Lambda_r Q_r^H = \sum_{i=1}^{r} \lambda_i q_i q_i^H. \tag{4}$$

Let $e_1 = (1, 0, \ldots, 0)^T \in \mathbb{C}^m$. With this, the optimal filter weights w^* given by (3) become

$$w^* = U_r(\Lambda_r + \mu I_r)^{-1} \Lambda_r Q_r^H e_1 \tag{5}$$

since $U^H Q = I_m$. That is, the left and right eigenvectors are bi-orthogonal. The filters (5) are referred to as the low-rank multichannel Wiener filters (LR-MWF) [17].

Note that many of the existing beamformers can be expressed as (5). The case $\mu = 1$ and $r = m$ gives the classical multi-channel Wiener filter since $R_Y^{-1} = U(\Lambda + I_m)^{-1} U^H$ by (1). In fact, μ can be seen as a trade-off parameter that controls the signal distortion and noise reduction. If we have $r = 1$, we have $w^* = \alpha u_1$ where $\alpha = (\lambda_1 + \mu)^{-1} \lambda_1 \bar{q}_{11} \in \mathbb{C}$. With this, the output signal-to-noise ratio (SNR_{out}) becomes

$$\text{SNR}_{\text{out}} = \frac{w^H R_X w}{w^H R_V w} = \lambda_1,$$

since $R_X u_1 = \lambda_1 R_V u_1$. Hence we conclude that this case leads to the maximum SNR beamformer, independent of the value of μ. The special case in which $\mu = 0$ leads to the MVDR beamformer. See [21] for an complete overview.

IV. Clock-offset and gain compensation

As mentioned in the introduction, even though clock skew can be neglected in many practical scenarios, having a clock offset is inevitable. In addition, the (unknown) microphone gains have to be equalised. Instead of synchronising the microphones to compensate for the different clock offsets and calibrate the gains, we will show that the low-rank multichannel Wiener filters as discussed in the previous section are invariant to clock-offsets and microphone gain differences.

Let τ_i denote the clock offset of the ith microphone with respect to the reference microphone, so that $\tau_1 = 0$. Moreover, let g_i denote the gain of microphone i and assume, without loss of generality, that the gain of the reference microphone is $g_1 = 1$. With this, the uncalibrated microphone signals can be expressed as $\tilde{y} = Ty$ where $T = \text{diag}(1, g_2 e^{j\omega\tau_2}, \cdots, g_m e^{j\omega\tau_m})$. As a consequence, since $y = x + v$, we have $\tilde{y} = T(x + v) = \tilde{x} + \tilde{v}$. Let $\tilde{R}_X$ and $\tilde{R}_V$ denote the CPSD matrices of the uncalibrated target and noise process, respectively. Since $\tilde{X} = TX$ and $\tilde{V} = TV$, we have $\tilde{R}_X = \text{E}(\tilde{X}\tilde{X}^H) = T\text{E}(XX^H)T^H = TR_X T^H$ and similarly we find $\tilde{R}_V = TR_V T^H$. Hence,

$$\tilde{R}_V^{-1} \tilde{R}_X = (TR_V T^H)^{-1}(TR_X T^H) = T^{-H} R_V^{-1} R_X T^H,$$

and we conclude that $\tilde{R}_V^{-1}\tilde{R}_X$ and $R_V^{-1}R_X$ are similar, even though this does not hold for the constituent matrices. We have the following result.

Proposition 1. Let $\tilde{U}^H \tilde{R}_X \tilde{U} = \tilde{\Lambda}$ and $\tilde{U}^H \tilde{R}_V \tilde{U} = I_m$ be the GEVD of the matrix pencil $(\tilde{R}_V, \tilde{R}_X)$. Then $\tilde{\Lambda} = \Lambda$ and $\tilde{U} = T^{-H} U B$, where $B = \text{diag}(B_1, \cdots, B_k)$, $B_i \in \mathbb{C}^{n_i \times n_i}$ unitary, and n_i denotes the algebraic multiplicity of λ_i and k the number of distinct eigenvalues.

Proof. Since $\tilde{R}_V^{-1}\tilde{R}_X = \tilde{U}\tilde{\Lambda}\tilde{U}^{-1}$ is similar to $R_V^{-1}R_X = U\Lambda U^{-1}$, we conclude that $\tilde{\Lambda} = \Lambda$. In addition, since $\tilde{R}_V^{-1}\tilde{R}_X = T^{-H}R_V^{-1}R_X T^H = T^{-H}U\Lambda(T^{-H}U)^{-1}$ and the fact that eigenvectors associated to λ_i are unique up to an invertible transform $B_i \in \mathbb{C}^{n_i \times n_i}$, we conclude that $\tilde{U} = T^{-H}UB$ where $B = \text{diag}(B_1, \cdots, B_k)$, $B_i \in \mathbb{C}^{n_i \times n_i}$ invertible. Moreover, since

$$\tilde{U}^H \tilde{R}_V \tilde{U} = (B^H U^H T^{-1})(TR_V T^H)(T^{-H}UB) = B^H B,$$

we conclude that $B^H B = I_m$, which completes the proof. □

In order to calculate the low-rank multichannel Wiener filters using the left and right eigenvectors of $\tilde{R}_V^{-1}\tilde{R}_X$, we combine (5) and (4) and the fact that $\tilde{Q} = \tilde{U}^{-H} = TQB$, and we obtain[4]

$$\begin{aligned}\tilde{w}^* &= \tilde{U}_r(\Lambda_r + \mu I_r)^{-1}\Lambda_r \tilde{Q}_r^H e_1 \\ &= T^{-H} U_r B_r (\Lambda_r + \mu I_r)^{-1} \Lambda_r B_r^H Q_r^H T^H e_1. \end{aligned} \tag{6}$$

Moreover, since B_r and $(\Lambda_r + \mu I_r)^{-1}\Lambda_r$ have a block-diagonal structure, where the block-entries of $(\Lambda_r + \mu I_r)^{-1}\Lambda_r$ are scaled identities (with scaling factors $\lambda_i/(\lambda_i + \mu)$), they commute and (6) reduces to

$$\begin{aligned}\tilde{w}^* &\overset{(a)}{=} T^{-H} U_r (\Lambda_r + \mu I_r)^{-1} \Lambda_r Q_r^H e_1 \\ &\overset{(b)}{=} T^{-H} w^*,\end{aligned}$$

[3] Since the minimum of (2) is attained on the boundary of the feasible set $\{a \in \mathbb{C}^m : a^H a \leq c\}$, we can replace the inequality constraint by an equality one.

[4] With slight abuse of notation we denote here by B_r the $r \times r$ leading principal submatrix of B.

where (a) uses $T^H e_1 = e_1$ and (b) follows from (5). The output of the beamformer, $\tilde{w}^{*H}\tilde{y}$, then becomes

$$\tilde{w}^{*H}\tilde{y} = w^{*H}T^{-1}Ty = w^{*H}y,$$

and we conclude that the LR-MWF (and thus the SDW-MWF as a special case) are invariant to clock offsets and gain variations and produce the same target estimate as if the clocks were perfectly synchronised and gains were perfectly equalised.

V. Experimental results

In this section we present experimental results obtained by computer simulations to substantiate our claim that the LR-MWF is clock-offset and gain mismatch invariant. To do so, we considered a box-shaped room with dimensions $4 \times 4 \times 3$ m. The target source is centred in the room and an interfering noise source and $m = 7$ microphones are distributed uniformly at random in the room. Room impulse responses (RIRs) were calculated using [26] ($T_{60} = 50$ ms). The target and interferer (both speech) signals, sampled at a sampling frequency of 16 kHz, were taken form the TIMIT database [27]. The signals had a duration of 5 seconds and the signal-to-interferer ratio (SIR) at the reference microphone ($i = 1$) was set to 0 dB. The microphone-self noise was white Gaussian noise with 40 dB SNR. Processing of the signals was done on a frame-by-frame basis using a 30 ms, 50% overlap, Hann window. The covariance matrices R_Y and R_V were estimated by their sample covariance matrix. Clock offsets were introduced in the system by shifting the received microphone signals y_i, $i = 2, \ldots, m$, in time. The beamformer parameters were set to $r = 1$ and $\mu = 0$ which corresponds to the MVDR beamformer which, in this case, can be expressed as

$$w_{\text{MVDR}} = \frac{R_V^{-1}d}{d^H R_V^{-1} d}, \tag{7}$$

where d is the (relative) acoustic transfer function from the target source to the microphones.

The beamformer performance is evaluated in terms of both SNR and STOI [28] scores at the output of the beamformer as a function of the variance of the clock offset, where the SNR is defined as

$$\text{SNR} = 10\log\left(\frac{\|x_1\|_2^2}{\|w^H y - x_1\|_2^2}\right) \quad (\text{dB}).$$

Figure 1 shows the results (averaged over 100 runs) for SNR scores (top subplot) and STOI scores (bottom subplot) as a function of clock offset in the absence of a gain mismatch. Figure 2, on the other hand, shows results as a function of gain mismatch in the absence of clock offset. The blue curves (triangles) represent the performance of the MVDR beamformer implemented as (7), while the red curves (squares) represent the GEVD implementation. We estimated R_V from the received interfering signal and calculated d based on the complete RIRs (of which the lengths exceed the analysis frame length). Note that in practical situations both interferers and d are not available and need to be estimated from the noisy

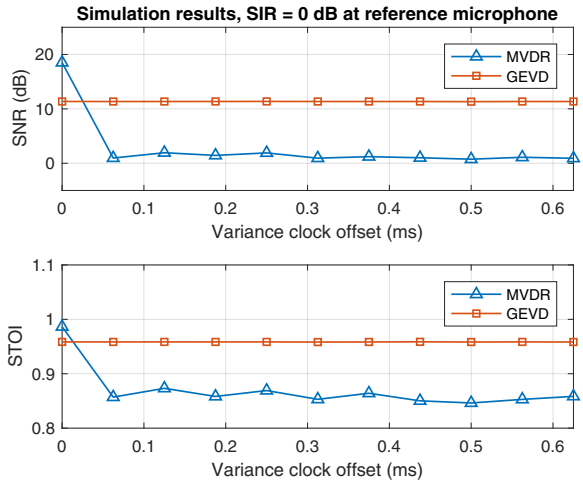

Figure 1. SNR/STOI scores vs. variance of the clock offset.

data or need to be known a priori. As such this experiment represents an idealised situation in order to minimise the effect of imperfections in the estimation of these parameters. By inspection of Figure 1 and 2 we conclude that, independent of the clock offset or gain mismatch, the performance of the GEVD-based beamformer is constant, whereas the performance of the MVDR beamformer as implemented by (7) degrades substantially. The bottom plot of Figure 2, however, shows that a gain mismatch itself has little effect on the STOI scores of the MVDR beamformer, even though the SNR scores drop significantly. The reason for this is that, in the absence of phase errors, the beam is steered in the direction of the target source, while the gain mismatch mainly effects the null-steering of the beamformer [29]. Hence, there is little target signal distortion introduced and as such the intelligibility is not severely degraded. In addition, in the absence of both clock offset and gain mismatch, which corresponds to the intersection points on the vertical axes, the performance of both methods differs. This difference is due to the fact that with the MVDR implementation (7) the true acoustic transfer function is used, whereas the GEVD approach implicitly estimates d by making a rank $r = 1$ approximation of R_X.

References

[1] R. Heusdens, G. Zhang, R. Hendriks, Y. Zeng, and W. Kleijn, "Distributed mvdr beamforming for (wireless) microphone networks using message passing," *Proceedings International Workshop on Acoustic Signal Enhancement (IWAENC)*, 2012.

[2] S. Markovich-Golan, S. Gannot, and I. Cohen, "Distributed multiple constraints generalized sidelobe canceler for fully connected wireless acoustic sensor networks," *IEEE Trans. on Audio, Speech and Language Processing*, vol. 21, no. 2, pp. 343 – 356, Feb. 2013.

[3] A. Bertrand and M. Moonen, "Distributed LCMV beamforming in a wireless sensor network with single-channel per-node signal transmission," *IEEE Trans. on Signal Processing*, vol. 61, no. 13, pp. 3447 – 3459, Jul. 2013.

[4] A. Koutrouvelis, T. Sherson, R. Heusdens, and R. Hendriks, "A low-cost robust distributed linearly constrained beamformer for wireless acoustic sensor networks with arbitrary topology," *IEEE Trans. on Audio, Speech and Language Processing*, vol. 26, no. 8, pp. 1434 – 1448, Aug. 2018.

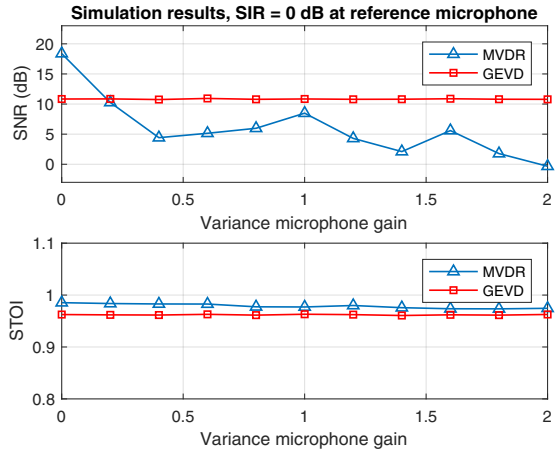

Figure 2. SNR/STOI scores vs. variance of the gain.

[5] J. Zhang, A. Koutrouvelis, R. Heusdens, and R. Hendriks, "Distributed rate-constrained LCMV beamforming," *IEEE Signal Processing Letters*, vol. 26, no. 5, pp. 675 – 679, May 2019.

[6] N. Gaubitch, W. Kleijn, and R. Heusdens, "Calibration of distributed sound acquisition systems using toa measurements from a moving acoustic source," in *Proc. ICASSP 2014*, Florence, Italy, May 5-9 2014.

[7] N. Tashev, "Gain self-calibration procedure for microphone arrays," in *IEEE Int. Conference on Multimedia and Expo (ICME)*, vol. 2, June 2004, pp. 983–986.

[8] R. Rajan and A. van der Veen, "Joint ranging and synchronization for an anchorless network of mobile nodes," *IEEE Trans. on Signal Processing*, vol. 63, no. 8, pp. 1925–1940, Aug. 2015.

[9] L. Schenato and F. Fiorentin, "Average timesynch: A consensus-based protocol for clock synchronization in wireless sensor networks," *Automatica*, vol. 47, no. 9, pp. 1878–1886, Sep. 2011.

[10] J. Schmalenstroeer, P. Jebramcik, and R. Haeb-Umbach, "A combined hardware-software approach for acoustic sensor network synchronization," *Signal Processing*, vol. 107, pp. 171–184, 2015.

[11] R. Lienhart, I. Kozintsev, S. Wehr, and M. Yeung, "On the importance of exact synchronization for distributed audio signal processing," in *Proc. ICASSP 2003*, vol. 4, 2003, pp. IV–840.

[12] S. Wehr, I. Kozintsev, R. Lienhart, and W. Kellermann, "Synchronization of acoustic sensors for distributed ad-hoc audio networks and its use for blind source separation," in *IEEE Sixth International Symposium on Multimedia Software Engineering*, 2004, pp. 18–25.

[13] M. Bahari, A. Bertrand, and M. Moonen, "Blind sampling rate offset estimation for wireless acoustic sensor networks through weighted least-squares coherence drift estimation," *IEEE Trans. on Audio, Speech and Language Processing*, vol. 25, no. 3, pp. 674–686, 2017.

[14] S. Markovich-Golan, S. Gannot, and I. Cohen, "Blind sampling rate offset estimation and compensation in wireless acoustic sensor networks with application to beamforming," in *Proc. IWAENC 2012*, 2012, pp. 1–4.

[15] D. Cherkassky and S. Gannot, "Blind synchronization in wireless acoustic sensor networks," *IEEE Trans. on Audio, Speech and Language Processing*, vol. 25, no. 3, pp. 651–661, 2017.

[16] L. Wang and S. Doclo, "Correlation maximization-based sampling rate offset estimation for distributed microphone arrays," *IEEE Trans. on Audio, Speech and Language Processing*, vol. 24, no. 3, pp. 571–582, 2016.

[17] R. Serizel, M. Moonen, B. van Dijk, and J. Wouters, "Low-rank approximation based multichannel Wiener algorithm for noise reduction with application in cochlear implants," *IEEE Trans. on Audio, Speech and Language Processing*, vol. 22, no. 4, pp. 785–799, 2014.

[18] G. Golub and C. Van Loan, *Matrix Computations*, 3rd ed. Oxford: North Oxford Academic, 1983.

[19] Y. Ephraim and H. van Trees, "A signal subspace approach for speech enhancement," *IEEE Trans. Speech and Audio Processing*, vol. 3, no. 4, pp. 251–266, 1995.

[20] S. Doclo, A. Spriet, J. Wouters, and M. Moonen, "Frequency-domain criterion for the speech distortion weighted multichannel Wiener filter for robust noise reduction," *Speech Communication*, vol. 49, no. 7-8, pp. 636–656, Jul. 2007.

[21] J. Jensen, J. Benesty, and M. G. Christensen, "Noise reduction with optimal variable span linear filters," *IEEE Trans. on Audio, Speech and Language Processing*, vol. 24, no. 4, pp. 631–644, April 2016.

[22] A. Spriet, M. Moonen, and J. Wouters, "Spatially pre-processed speech distortion weighted multichannel Wiener filtering for noise reduction," *Signal Processing*, vol. 84, pp. 2367–2387, 2004.

[23] S. Doclo, J. W. A. Spriet, and M. Moonen, *Speech Distortion Weighted Multichannel Wiener Filtering techniques for Noise Reduction*, ser. Springer Series on Signals and Communication Technology. Springer, 2004, ch. 9, pp. 199–228.

[24] S. Jensen, P. Hansen, S. Hansen, and J. Sørensen, "Reduction of broadband noise in speech by truncated QSVD," *IEEE Trans. on Speech and Audio Processing*, vol. 3, no. 6, pp. 439–448, 1995.

[25] S. Doclo and M. Moonen, "GSVD-based optimal filtering for single and multimicrophone speech enhancement," *IEEE Trans. on Signal Processing*, vol. 50, no. 9, pp. 2230–2244, Sep. 2002.

[26] A. Wabnitz, N. Epain, C. Jin, and A. van Schaik, "Room acoustics simulation for multichannel microphone arrays," *Proceedings of the International Symposium on Room Acoustics (ISRA 2010)*, 29 - 31 August 2010, Melbourne, Australia.

[27] J. Garofolo et al., "TIMIT acoustic-phonetic continuous speech corpus," Philadelphia: Linguistic Data Consortium, 1993, https://catalog.ldc.upenn.edu/LDC93S1.

[28] C. Taal, R. Hendriks, R. Heusdens, and J. Jensen, "An algorithm for intelligibility prediction of time-frequency weighted noisy speech," *IEEE Trans. on Audio, Speech, and Language Processing*, vol. 19, no. 7, pp. 2125 – 2136, Sep. 2011.

[29] O. Bakr and M. Johnson, "Impact of phase and amplitude errors on array performance," Electrical Engineering and Computer Sciences, University of California at Berkeley, Tech. Rep., 2009, technical Report No. UCB/EECS-2009-1.