REFERENCE MICROPHONE SELECTION FOR THE WEIGHTED PREDICTION ERROR ALGORITHM USING THE NORMALIZED L-P NORM

Anselm Lohmann[1]*, Toon van Waterschoot*[2]*, Joerg Bitzer*[3]*, Simon Doclo*[1,3]

[1]Carl von Ossietzky Universität Oldenburg, Dept. of Medical Physics and Acoustics, Germany
[2]KU Leuven, Department of Electrical Engineering (ESAT-STADIUS), Leuven, Belgium
[3]Fraunhofer IDMT, Project Group Hearing, Speech and Audio Technology, Oldenburg, Germany
anselm.lohmann@uni-oldenburg.de

ABSTRACT

Reverberation may severely degrade the quality of speech signals recorded using microphones in a room. For compact microphone arrays, the choice of the reference microphone for multi-microphone dereverberation typically does not have a large influence on the dereverberation performance. In contrast, when the microphones are spatially distributed, the choice of the reference microphone may significantly contribute to the dereverberation performance. In this paper, we propose to perform reference microphone selection for the weighted prediction error (WPE) dereverberation algorithm based on the normalized ℓ_p-norm of the dereverberated output signal. Experimental results for different source positions in a reverberant laboratory show that the proposed method yields a better dereverberation performance than reference microphone selection based on the early-to-late reverberation ratio or signal power.

Index Terms— Dereverberation, weighted prediction error, acoustic sensor networks, reference microphone selection

1. INTRODUCTION

Microphone recordings of a speech source inside a room are typically degraded by reverberation, i.e. acoustic reflections against walls and objects in the room. While early reflections may improve speech intelligibility, late reverberation typically reduces both speech intelligibility as well as automatic speech recognition performance [1, 2]. Therefore, effective speech dereverberation is required for many applications, including voice-controlled systems, hearing aids and hands-free telephony [3–8]. A popular blind multi-channel dereverberation algorithm is the weighted prediction error (WPE) algorithm [7, 8], which is based on multi-channel linear prediction (MCLP). WPE performs dereverberation in a chosen reference microphone by estimating the late reverberant component using a prediction filter and subtracting this estimate from the reference microphone signal.

When performing multi-microphone speech enhancement using compact microphone arrays, the choice of the reference microphone typically does not have a large influence on the quality of the output signal. However, when considering spatially distributed microphones, there may be large differences in the early-to-late reverberation ratio (ELR) and signal power in each microphone. Hence, the choice of the reference microphone may significantly contribute to the speech enhancement performance [9–11]. In [9] and [10], the reference microphone selection problem was formulated for different multi-microphone noise reduction algorithms by maximizing the output signal-to-noise ratio. In [11], different reference microphone selection methods were proposed for speech enhancement in meeting recognition scenarios when considering different microphone sensitivities. However, to the best of the authors' knowledge, no work exists on reference microphone selection for multi-microphone dereverberation.

In this paper, we propose to perform reference microphone selection for the WPE algorithm based on the normalized ℓ_p-norm of the dereverberated output signal. From the WPE optimization problem, it may appear logical to formulate the reference microphone selection problem as an ℓ_p-norm minimization problem. However, since the ℓ_p-norm depends on the signal power, which may greatly vary for spatially distributed microphones, we propose to normalize for the output signal power, leading to a selection based on the ratio of the ℓ_p-norm and the ℓ_2-norm [12–14]. Experimental results for several source positions and spatially distributed microphones in a reverberant laboratory show that the dereverberation performance using the proposed reference microphone selection method is larger than the performance when selecting the reference microphone based on the estimated ELR [15] or signal power [9] [11]. Furthermore, similar performance can be achieved for the proposed method using only a small number of WPE iterations.

2. SIGNAL MODEL

We consider a scenario where a single speech source is captured in a room by M spatially distributed microphones. Similarly as in [7, 8], we consider a static scenario without additive noise. In the short-time Fourier transform (STFT) domain, let $s(f, n)$ denote the clean speech signal with $f \in \{1, ..., F\}$ the frequency bin index and $n \in \{1, ..., N\}$ the time frame index, where F and N denote the number of frequency bins and time frames, respectively. The reverberant

This work has received funding from the European Union's Horizon 2020 research and innovation programme under the Marie Skłodowska-Curie grant agreement No. 956369 and KU Leuven Internal Funds C14/21/075, and from the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) under Germany's Excellence Strategy - EXC 2177/1 - Project ID 390895286.

signal at the m-th microphone $x_m(f,n)$ can be written as

$$x_m(f,n) = \sum_{l=0}^{L_h-1} h_m(f,l)s(f,n-l) + e_m(f,n), \quad (1)$$

where $h_m(f,l)$ denotes the subband convolutive transfer function with length L_h between the speech source and the m-th microphone, and $e_m(f,n)$ denotes the subband modelling error [16]. In the remainder of the paper, the frequency bin index f will be omitted where possible. Assuming the subband modelling error to be 0, the dereverberation problem in microphone r (referred to as the reference microphone) can be formulated as (see Fig. 1)

$$d_r(n) = x_r(n) - u_r(n). \quad (2)$$

The desired component $d_r(n) = \sum_{l=0}^{L_d-1} h_r(l)s(n-l)$ consists of the direct path and early reflections in the reference microphone signal $x_r(n)$, where L_d denotes the temporal cut-off between early and late reflections. The undesired component $u_r(n) = \sum_{l=L_d}^{L_h-1} h_r(l)s(n-l)$, which we aim to estimate, is the late reverberant component in the reference microphone signal $x_r(n)$. It should be noted that for spatially distributed microphones, the power of the desired component $d_r(n)$ and the power of the undesired component $u_r(n)$ may greatly depend on the choice of reference microphone r.

Using the MCLP model [7], the late reverberant component $u_r(n)$ can be written as the sum of filtered delayed versions of all reverberant microphone signals. Whereas for compact microphone arrays the same prediction delay is typically used for each microphone, it has been shown in [17] that for spatially distributed microphones it is beneficial to use a microphone-dependent prediction delay, i.e.

$$u_r(n) = \sum_{m=1}^{M} \sum_{l=0}^{L_g-1} g_{m,r}(l)x_m(n-\tau_{m,r}-l), \quad (3)$$

where $g_{m,r}(l)$ denotes the m-th prediction filter of length L_g and $\tau_{m,r}$ denotes the prediction delay for the m-th microphone. Using (3), the signal model in (2) can be rewritten in vector notation as

$$\mathbf{d}_r = \mathbf{x}_r - \mathbf{X}_{\boldsymbol{\tau},r}\mathbf{g}_r, \quad (4)$$

with

$$\mathbf{d}_r = \begin{bmatrix} d_r(1) & \cdots & d_r(N) \end{bmatrix}^T \in \mathbb{C}^N, \quad (5)$$

$$\mathbf{x}_r = \begin{bmatrix} x_r(1) & \cdots & x_r(N) \end{bmatrix}^T \in \mathbb{C}^N. \quad (6)$$

The multi-channel delayed convolution matrix $\mathbf{X}_{\boldsymbol{\tau},r}$ in (4) is defined as

$$\mathbf{X}_{\boldsymbol{\tau},r} = \begin{bmatrix} \mathbf{X}_{\tau_{1,r}} & \cdots & \mathbf{X}_{\tau_{M,r}} \end{bmatrix} \in \mathbb{C}^{N \times ML_g}, \quad (7)$$

where $\mathbf{X}_{\tau_{m,r}} \in \mathbb{C}^{N \times L_g}$ is the convolution matrix of $\mathbf{x}_m$ delayed by $\tau_{m,r}$ frames with τ the prediction delay in the reference microphone and $\mathbf{g}_r \in \mathbb{C}^{ML_g}$ is the stacked vector of all prediction filter coefficients $g_{m,r}(l)$. The dereverberation problem, i.e. estimation of the desired component $\mathbf{d}_r$, is now reduced to estimating the filter $\mathbf{g}_r$ predicting the undesired late reverberant component.

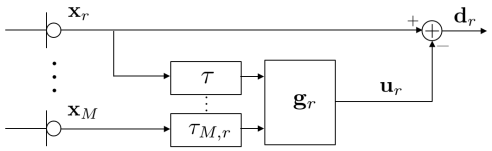

Fig. 1: WPE with microphone-dependent prediction delays [17]

3. WPE ALGORITHM

Since the desired speech component $\mathbf{d}_r$ can be assumed to be sparser than the reverberant microphone signal $\mathbf{x}_r$, it has been proposed in [8] to compute the prediction filter $\mathbf{g}_r$ by minimizing the sparsity-promoting ℓ_p-norm of the output in (4), i.e.

$$\min_{\mathbf{g}_r} J(\mathbf{g}_r) = \|\mathbf{d}_r\|_p^p = \|\mathbf{x}_r - \mathbf{X}_{\boldsymbol{\tau},r}\mathbf{g}_r\|_p^p, \quad (8)$$

where the ℓ_p-norm is defined as $\|\mathbf{d}_r\|_p = (\sum_{n=1}^{N} |d_r(n)|^p)^{1/p}$. For effective dereverberation, the sparsity-promoting parameter p is typically chosen in the range $0 < p < 1$ [8], leading to a non-convex optimization problem in (8).

A popular method for solving non-convex optimization problems such as (8) is the iteratively reweighted least-squares (IRLS) algorithm [18], where the original problem is replaced with a series of convex quadratic problems. Namely, in the i-th iteration the ℓ_p-norm in (8) is approximated by a weighted ℓ_2-norm, i.e.

$$\|\mathbf{d}_r\|_p^p \approx \|\mathbf{d}_r\|_{\mathbf{W}_r^{(i)}}^2 = \mathbf{d}_r^H \mathbf{W}_r^{(i)} \mathbf{d}_r, \quad (9)$$

where $\mathbf{W}_r^{(i)} = \text{diag}\left(\mathbf{w}_r^{(i)}\right)$ is a diagonal matrix of the weight vector in the i-th iteration $\mathbf{w}_r^{(i)}$. Given a previous estimate $\hat{\mathbf{w}}_r^{(i-1)}$ of the weights $\mathbf{w}_r^{(i)}$, the minimization problem in the i-th iteration can be written as

$$\min_{\mathbf{g}_r} \|\mathbf{x}_r - \mathbf{X}_{\boldsymbol{\tau},r}\mathbf{g}_r\|_{\hat{\mathbf{W}}_r^{(i-1)}}^2, \quad (10)$$

yielding a closed-form solution for the prediction filter

$$\hat{\mathbf{g}}_r^{(i)} = \left(\mathbf{X}_{\boldsymbol{\tau},r}^H (\hat{\mathbf{W}}_r^{(i-1)})^{-1} \mathbf{X}_{\boldsymbol{\tau},r}\right)^{-1} \mathbf{X}_{\boldsymbol{\tau},r}^H (\hat{\mathbf{W}}_r^{(i-1)})^{-1} \mathbf{x}_r. \quad (11)$$

The estimated weights $\hat{\mathbf{w}}_r^{(i-1)}$ are subsequently updated such that the approximation in (8) is a first-order approximation [8], i.e.

$$\hat{\mathbf{w}}_r^{(i)} = |\hat{\mathbf{d}}_r^{(i)}|^{2-p}, \quad (12)$$

where the dereverberated output in the i-th iteration $\hat{\mathbf{d}}_r^{(i)} = \mathbf{x}_r - \mathbf{X}_{\boldsymbol{\tau},r}\hat{\mathbf{g}}_r^{(i)}$ with the $|.|$ and $(.)^{2-p}$ operators applied element-wise. To prevent division by zero, a small positive constant ϵ is typically included in the weight update in (12). The initial weights $\hat{\mathbf{w}}_r^{(0)}$ are computed by defining an initial prediction filter $\hat{\mathbf{g}}_r^{(0)} = \mathbf{0}$. In total, I_{WPE} iterations are performed.

4. REFERENCE MICROPHONE SELECTION

When considering spatially distributed microphones, the power of the desired component $\mathbf{d}_r$ and the power of the undesired component $\mathbf{u}_r$ may greatly vary depending on the reference microphone r.

Hence, the quality of the dereverberated output $\hat{\mathbf{d}}_r$ may also depend significantly on the choice of reference microphone. Based on the WPE cost function, in Section 4.1 we first define the reference microphone selection problem as an ℓ_p-norm minimization problem. However, since the differences in signal power between the microphones may be large, the ℓ_p-norm-based reference microphone selection problem may not yield the dereverberated output $\hat{\mathbf{d}}_r$ with the highest signal quality. Hence, in Section 4.2 we propose to normalize for the power in the dereverberated output, leading to reference microphone selection based on the normalized ℓ_p-norm.

4.1. Reference microphone selection using ℓ_p-norm

By considering the WPE cost function in (8), it may appear logical to select the reference microphone as the one minimizing the cost function

$$\min_r \left\| \hat{\mathbf{d}}_r^{(I)} \right\|_p = \left\| \mathbf{x}_r - \mathbf{X}_{\boldsymbol{\tau},r} \hat{\mathbf{g}}_r^{(I)} \right\|_p, \tag{13}$$

where $\hat{\mathbf{d}}_r^{(I)}$ and $\hat{\mathbf{g}}_r^{(I)}$ correspond to the dereverberated output and the prediction filter for reference microphone r after I WPE iterations. Since WPE is run independently per frequency, a different reference microphone may be selected for each frequency when using (13). In order to select a single reference microphone over all frequencies, we propose to minimize the sum over all frequencies, i.e.

$$\boxed{\hat{r}_{\ell_p}^{(I)} = \underset{r}{\operatorname{argmin}} \sum_{f=1}^{F} \left\| \mathbf{x}_r(f) - \mathbf{X}_{\boldsymbol{\tau},r}(f) \hat{\mathbf{g}}_r^{(I)}(f) \right\|_p}, \tag{14}$$

where $\hat{r}_{\ell_p}^{(I)}$ denotes the selected reference microphone based on the ℓ_p-norm.

4.2. Reference microphone selection using normalized ℓ_p-norm

When the differences in signal power are large between the microphones, selecting the reference microphone based on the ℓ_p-norm of the output may not yield the best dereverberated output, but possibly the output with the smallest power (irrespective of the amount of reverberation reduction). In order to normalize for the signal power in the different microphones, we propose to normalize the dereverberated output using the ℓ_2-norm, i.e.

$$\underline{\hat{\mathbf{d}}}_r^{(I)} = \frac{\hat{\mathbf{d}}_r^{(I)}}{\|\hat{\mathbf{d}}_r^{(I)}\|_2}. \tag{15}$$

When inserting the normalized dereverberated output $\underline{\hat{\mathbf{d}}}_r^{(I)}$ into the problem in (14), the modified reference microphone selection problem can be reformulated as a normalized ℓ_p-norm minimization problem, i.e.

$$\boxed{\hat{r}_{\ell_p/\ell_2}^{(I)} = \underset{r}{\operatorname{argmin}} \sum_{f=1}^{F} \frac{\left\| \mathbf{x}_r(f) - \mathbf{X}_{\boldsymbol{\tau},r}(f) \hat{\mathbf{g}}_r^{(I)}(f) \right\|_p}{\left\| \mathbf{x}_r(f) - \mathbf{X}_{\boldsymbol{\tau},r}(f) \hat{\mathbf{g}}_r^{(I)}(f) \right\|_2}}, \tag{16}$$

where $\|.\|_p/\|.\|_2$ and $\hat{r}_{\ell_p/\ell_2}^{(I)}$ denote the normalized ℓ_p-norm [12–14] and the selected reference microphone based on the normalized ℓ_p-norm, respectively. The normalized ℓ_p-norm, also known as the ℓ_p/ℓ_q-norm, is a popular alternative to the ℓ_p-norm due to its scale-invariance [14]. Typically, q is chosen such that $q > 1$, with $q = 2$ being a common choice due to its relation to signal power [12].

5. EXPERIMENTAL EVALUATION

In this section, we evaluate the performance of the proposed WPE reference microphone selection method for spatially distributed microphones in a reverberant room. In Section 5.1, we discuss the considered acoustic scenario and algorithm parameters. In Section 5.2, we present the simulation results and evaluate the performance of the proposed method against a selection based on the ELR and signal power.

5.1. Acoustic setup and algorithm parameters

We consider $M = 8$ spatially distributed microphones and a single static (directional) speech source in a laboratory with dimensions of about 6m×7m×2.7m and reverberation time $T_{60} \approx 1300$ ms. Fig. 2 depicts the position of the microphones and the 12 considered positions of the speech source.

The reverberant microphone signals were generated at a sampling rate of 16 kHz by convolving anechoic speech signals from the TIMIT database [19] with measured room impulse responses from the BRUDEX database [20]. The signals were processed using an STFT framework with frame size of 1024 samples, a frame shift $L_{\text{shift}} = 256$ samples and square-root Hann analysis and synthesis windows.

The WPE algorithm was run using the entire speech utterance (batch processing) and implemented with prediction filter length $L_g = 15$, number of reweighting iterations $I_{\text{WPE}} = 10$, weight regularization parameter $\epsilon = 10^{-7}$ and sparsity-promoting parameter $p \in \{0.05, 0.5, 0.9\}$. The microphone-dependent prediction delays were computed with a prediction delay of $\tau = 2$ for the reference microphone, estimated time-differences-of-arrival using the generalised cross-correlation with phase transform (GCC-PHAT) [21] method and implemented using cross-band filtering [17].

5.2. Simulation results

For the WPE algorithm, we consider the following reference microphone selection methods:

- ℓ_p: reference microphone selection using (14) based on the ℓ_p-norm of the dereverberated output with $I = I_{\text{WPE}}$ WPE iterations
- ℓ_p/ℓ_2: reference microphone selection using (16) based on the ℓ_p/ℓ_2-norm of the dereverberated output with $I = \{0, 1, I_{\text{WPE}}\}$ WPE iterations
- Max-E$\hat{\text{L}}$R: reference microphone selection by choosing the reverberant microphone signal with the largest estimated ELR using the method proposed in [15]

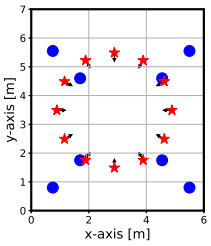

Fig. 2: Positions of $M = 8$ spatially distributed microphones (●) and 12 considered speech source positions (★)

- Max-Power: reference microphone selection by choosing the reverberant microphone signal with the largest average signal power [9] [11]

It should be noted that the (normalized) ℓ_p-norm-based reference microphone selection methods with $I > 0$ require running WPE in each reference microphone, whereas the reference microphone selection methods based on the normalized ℓ_p-norm with $I = 0$ and based on the estimated ELR and signal power do not require any WPE iterations.

In order to compute the performance improvement of the considered reference microphone selection methods, the dereverberation performance, i.e. the quality of the dereverberated output, using the selected reference microphone is evaluated against the average dereverberation performance of all possible reference microphones, i.e.

$$\Delta\text{PESQ} = \text{PESQ}\big(\hat{d}_{\hat{r}}(t), s_{\hat{r}}(t)\big) - \text{PESQ}_{\text{avg}}, \qquad (17)$$

where ΔPESQ denotes the perceptual evaluation of speech quality (PESQ) improvement with $\text{PESQ}\big(\hat{d}_{\hat{r}}(t), s_{\hat{r}}(t)\big)$ and $\text{PESQ}_{\text{avg}} = \frac{1}{M}\sum_{r=1}^{M} \text{PESQ}\big(\hat{d}_r(t), s_r(t)\big)$ denoting the PESQ of the (time-domain) dereverberated output $\hat{d}_{\hat{r}}(t)$ with time index t in selected reference microphone $\hat{r}$ and the average PESQ using all possible reference microphones, respectively. The target signal is the (time-domain) direct speech received at the reference microphone position $s_r(t)$. The improvement in the frequency-weighted segmental signal-to-noise ratio (ΔFWSSNR) is defined similarly as in (17). The above measures [22] are averaged across the 12 considered positions of the speech source.

For the considered reference microphone selection methods[1], Fig. 3 depicts the average performance improvement over all considered source positions in terms of ΔFWSSNR and ΔPESQ using a sparsity-promoting parameter $p = 0.05$, $p = 0.5$ and $p = 0.9$. For all considered values of the sparsity-promoting parameter p, the performance of the proposed method using $I = I_{\text{WPE}}$ WPE iterations is larger than the performance using a selection based on the estimated ELR or signal power. Furthermore, a similar performance can be achieved using the proposed method with only $I = 1$ WPE iteration. Even when $I = 0$, the performance using the proposed method is similar to the performance of the considered estimated ELR and signal power-based reference microphone selection methods. Finally, it can be seen that the normalization of the dereverberated output is required for WPE reference microphone selection as the performance using the ℓ_p-norm-based reference microphone selection method is significantly lower than the performance using the proposed method.

[1]Audio examples available on uol.de/f/6/dept/mediphysik/ag/sigproc/audio/dereverb/wpe-refmic-selection.html

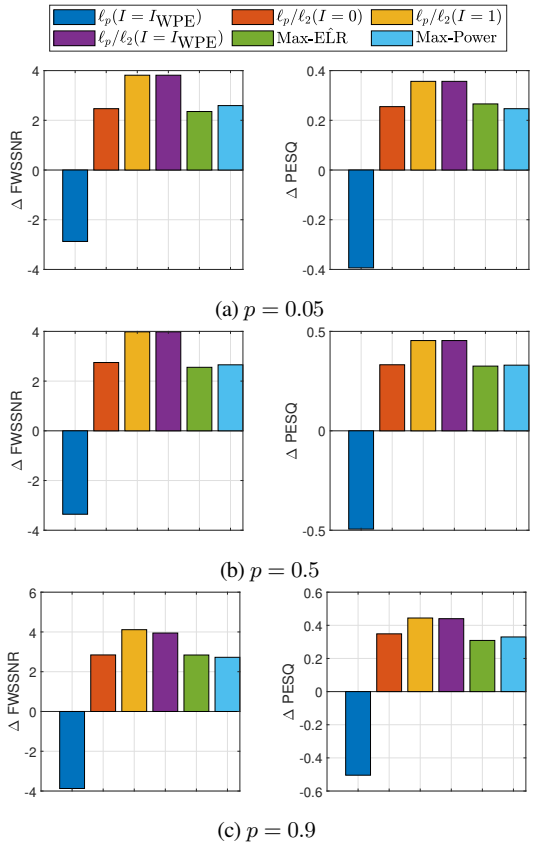

Fig. 3: Average performance improvement in terms of FWSSNR improvement and PESQ improvement for all considered reference microphone selection methods using a sparsity-promoting parameter (a) $p = 0.05$, (b) $p = 0.5$ and (c) $p = 0.9$

6. CONCLUSION

In this paper we have presented a reference microphone selection method for the WPE algorithm. Based on the WPE cost function, we first defined the reference microphone selection as an ℓ_p-norm minimization problem. However, when considering spatially distributed microphones, the differences in signal power between the microphones may be large and the ℓ_p-norm-based reference microphone selection problem may not yield the dereverberated output with the highest signal quality. Hence, we proposed to normalize for the power in the dereverberated output, leading to reference microphone selection based on the normalized ℓ_p-norm. The experimental results showed that the performance of the proposed method is larger than the performance using a selection based on the estimated ELR or signal power. Furthermore, similar performance can be achieved for the proposed method using only a small number of WPE iterations. Investigating the performance of the considered reference microphone selection methods for acoustic scenarios with additive noise, a moving source or different microphone sensitivities are directions for future research.

7. REFERENCES

[1] R. Beutelmann and T. Brand, "Prediction of speech intelligibility in spatial noise and reverberation for normal-hearing and hearing-impaired listeners," *Journal of the Acoustical Society of America*, vol. 120, no. 1, pp. 331–342, 2006.

[2] T. Yoshioka, A. Sehr, M. Delcroix, K. Kinoshita, R. Maas, T. Nakatani, and W. Kellermann, "Making machines understand us in reverberant rooms: Robustness against reverberation for automatic speech recognition," *IEEE Signal Processing Magazine*, vol. 29, no. 6, pp. 114–126, 2012.

[3] E. A. P. Habets and P. A. Naylor, "Dereverberation," in *Audio Source Separation and Speech Enhancement*, E. Vincent, T. Virtanen, and S. Gannot, Eds. Wiley, 2018.

[4] B. Cauchi, I. Kodrasi, R. Rehr, S. Gerlach, A. Jukić, T. Gerkmann, S. Doclo, and S. Goetze, "Combination of MVDR beamforming and single-channel spectral processing for enhancing noisy and reverberant speech," *EURASIP Journal on Advances in Signal Processing*, 2015.

[5] J. Lemercier, J. Thiemann, R. Koning, and T. Gerkmann, "A neural network-supported two-stage algorithm for lightweight dereverberation on hearing devices," *EURASIP Journal on Audio, Speech, and Music Processing*, 2023.

[6] D. S. Williamson and D. Wang, "Time-frequency masking in the complex domain for speech dereverberation and denoising," *IEEE/ACM Trans. on Audio, Speech, and Language Processing*, vol. 25, no. 7, pp. 1492–1501, 2017.

[7] T. Nakatani, T. Yoshioka, K. Kinoshita, M. Miyoshi, and B. H. Juang, "Speech dereverberation based on variance-normalized delayed linear prediction," *IEEE Trans. on Audio, Speech, and Language Processing*, vol. 18, no. 7, pp. 1717–1731, 2010.

[8] A. Jukić, T. van Waterschoot, T. Gerkmann, and S. Doclo, "Multi-channel linear prediction-based speech dereverberation with sparse priors," *IEEE/ACM Trans. on Audio, Speech, and Language Processing*, vol. 23, no. 9, pp. 1509–1520, 2015.

[9] T. Lawin-Ore and S. Doclo, "Reference microphone selection for MWF-based noise reduction using distributed microphone arrays," in *Proc. ITG Conference on Speech Communication*, Braunschweig, Germany, 2012, pp. 1–4.

[10] J. Zhang, H. Chen, L. Dai, and R. Hendriks, "A study on reference microphone selection for multi-microphone speech enhancement," *IEEE/ACM Trans. on Audio, Speech, and Language Processing*, vol. 29, pp. 671–683, 2021.

[11] S. Araki, N. Ono, K. Kinoshita, and M. Delcroix, "Comparison of reference microphone selection algorithms for distributed microphone array based speech enhancement in meeting recognition scenarios," in *Proc. International Workshop on Acoustic Signal Enhancement (IWAENC)*, Tokyo, Japan, 2018, pp. 316–320.

[12] N. Hurley and S. Rickard, "Comparing measures of sparsity," in *2008 IEEE Workshop on Machine Learning for Signal Processing*, Cancun, Mexico, 2008, pp. 55–60.

[13] L. Li, "Sparsity-promoted blind deconvolution of ground-penetrating radar (GPR) data," *IEEE Geoscience and Remote Sensing Letters*, vol. 11, no. 8, pp. 1330–1334, 2014.

[14] X. Jia, M. Zhao, Y. Di, P. Li, and J. Lee, "Sparse filtering with the generalized lp/lq norm and its applications to the condition monitoring of rotating machinery," *Mechanical Systems and Signal Processing*, vol. 102, pp. 198–213, 2018.

[15] F. Xiong, S. Goetze, B. Kollmeier, and B. Meyer, "Joint estimation of reverberation time and early-to-late reverberation ratio from single-channel speech signals," *IEEE/ACM Trans. on Audio, Speech, and Language Processing*, vol. 27, no. 2, pp. 255–267, 2019.

[16] Y. Avargel and I. Cohen, "System identification in the short-time Fourier transform domain with crossband filtering," *IEEE Trans. on Audio, Speech, and Language Processing*, vol. 15, no. 4, pp. 1305–1319, 2007.

[17] A. Lohmann, T. van Waterschoot, J. Bitzer, and S. Doclo, "Dereverberation in acoustic sensor networks using weighted prediction error with microphone-dependent prediction delays," in *Proc. International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, Rhodes, Greece, 2023, pp. 1–5.

[18] I. Daubechies, R. DeVore, M. Fornasier, and C.S. Güntürk, "Iteratively reweighted least squares minimization for sparse recovery," *Communications on Pure and Applied Mathematics: A Journal Issued by the Courant Institute of Mathematical Sciences*, vol. 63, no. 1, pp. 1–38, 2010.

[19] J. S. Garofolo, L. F. Lamel, W. M. Fisher, J. G. Fiscus, D. S. Pallett, and N. L. Dahlgren, "TIMIT acoustic phonetic continuous speech corpus," *Linguistic Data Consortium*, 1993.

[20] W. Middelberg D. Fejgin and S. Doclo, "BRUDEX database: Binaural room impulse responses with uniformly distributed external microphones," in *Proc. ITG Conference on Speech Communication*, Aachen, Germany, 2023, pp. 126–130, https://doi.org/10.5281/zenodo.7986447.

[21] C. Knapp and G. Carter, "The generalized correlation method for estimation of time delay," *IEEE Trans. on Acoustics, Speech, and Signal Processing*, vol. 24, no. 4, pp. 320–327, 1976.

[22] K. Kinoshita, M. Delcroix, S. Gannot, E. A. P. Habets, R. Haeb-Umbach, W. Kellermann, V. Leutnant, R. Maas, T. Nakatani, B. Raj, A. Sehr, and T. Yoshioka, "A summary of the REVERB challenge: state-of-the-art and remaining challenges in reverberant speech processing research," *EURASIP Journal on Advances in Signal Processing*, 2016.