Multi-channel Wiener Filter for Speech Dereverberation in Hearing Aids – Sensitivity to DoA Errors

Adam Kuklasiński[1,2], Simon Doclo[3], Søren H. Jensen[2], Jesper Jensen[1,2]

[1]*Oticon A/S, 2765 Smørum, Denmark*

[2]*Aalborg University, Dept. of Electronic Systems, Signal and Information Processing Group, 9220 Aalborg, Denmark*

[3]*University of Oldenburg, Dept. of Medical Physics and Acoustics, and Cluster of Excellence Hearing4all, Oldenburg, Gemany*

Correspondence should be addressed to Adam Kuklasiński (adku@oticon.com)

ABSTRACT

In this paper we study the robustness of a recently proposed Multi-channel Wiener Filter-based speech dereverberation algorithm to errors in the assumed direction of arrival (DoA) of the target speech. Different subsets of microphones of a pair of behind-the-ear hearing aids are used to construct various monaural and binaural configurations of the algorithm. Via a simulation experiment with frontally positioned target it is shown, that when correct DoA is assumed binaural configurations of the algorithm almost double the improvement of PESQ measure over monaural configurations. However, in conditions where the assumed DoA is increasingly incorrect, the performance of the binaural configurations is shown to deteriorate more quickly than that of the monaural configurations. In effect, for large DoA errors it is the simpler, monaural configurations that perform better.

1. INTRODUCTION

Hearing aids (HAs) are becoming more and more powerful with every new generation. Thanks to the progress in energy-efficiency of integrated circuits (including Digital Signal Processors – DSPs) increasingly sophisticated signal processing algorithms are becoming viable candidates for use in HA systems, see e.g. an overview of the state-of-the-art in [1, 2]. One of the algorithms that currently attracts the attention of researchers in the field is the Multi-channel Wiener Filter (MWF) [3, 4].

The MWF has a number of advantages which make it a particularly good fit for HA applications. Under certain conditions, the MWF can be decomposed into a beamformer and a post-filter which simplifies the implementation and allows the tradeoff between the speech distortion and the interference rejection to be easily controlled [4]. Moreover, the MWF algorithm can be used with arrays of any number of microphones and does not require any specific arrangement of these microphones. For example, as shown in [5, 6] the performance of the MWF with application to speech dereverberation generally increases with the number of microphones used with it.

The technological advancements which led to the increase of the energy-efficiency of DSPs, have also decreased the power requirements of wireless transmitters and receivers. This allows for more data to be transferred between the left and the right HA. In fact, it potentially allows transmission of the microphone signals between the HAs. Thanks to the fact that the MWF readily accommodates any number of microphone inputs, it is straightforward to incorporate the additional signal(s) received from the contralateral HA into an existing MWF scheme. The potential gain in the performance of the MWF is achieved at a cost of power needed to transmit the signals and to process the increased amount of data. Clearly, it is of interest to evaluate the expected benefit of implementing the binaural link in a HA system using the MWF.

In this paper we are focusing on the relationship between the number of microphones, their location, and the performance of a Multi-channel Wiener Filter-based speech dereverberation algorithm proposed in [6]. As we will show later, the sensitivity of the MWF to errors in the assumed direction of arrival (DoA) of the target speech

greatly increases with the addition of contralateral microphones. Hence, we will evaluate the performance of the MWF not only as a function of the number and location of the microphones, but also as a function of the DoA error.

This paper is organized as follows. In Section 2 we briefly describe the evaluated MWF-based algorithm and the assumptions made in its derivation. Section 3 contains the description and technical details of the conducted experiments and Section 4 contains the obtained results and their interpretation. Section 5 concludes the paper.

2. EVALUATED ALGORITHM

In the following we outline the basic statistical assumptions and the structure of the evaluated algorithm. For a fully detailed description and explanation of the assumptions the reader is referred to [6]. Additional details on the performance of the algorithm in the situation where the correct DoA is assumed have been given in [7].

2.1. Signal model

The considered algorithm [6] operates on the signals received by an array of M microphones located in a reverberant room with a single active talker. The signals are spectrally analyzed using the Short Time Fourier Transform (STFT). The obtained complex STFT coefficients are stacked in an $M \times 1$ vector $\mathbf{y}(n)$ where n denotes the STFT time frame index (the frequency bin index is omitted without loss of generality). The input signal is assumed to be composed of a target speech and a reverberation component, hence:

$$\mathbf{y}(n) = \mathbf{s}(n) + \mathbf{r}(n), \tag{1}$$

where $\mathbf{s}(n)$ and $\mathbf{r}(n)$ denote the target speech and the reverberation components of $\mathbf{y}(n)$, respectively. Because $\mathbf{s}(n)$ and $\mathbf{r}(n)$ are assumed to be uncorrelated, the $M \times M$ cross-correlation matrix of $\mathbf{y}(n)$ may be written as:

$$\begin{aligned}\mathbf{\Phi}_{\mathbf{y}}(n) &= E\{\mathbf{y}(n)\mathbf{y}^H(n)\} \\ &= E\{\mathbf{s}(n)\mathbf{s}^H(n)\} + E\{\mathbf{r}(n)\mathbf{r}^H(n)\} \\ &= \mathbf{\Phi}_{\mathbf{s}}(n) + \mathbf{\Phi}_{\mathbf{r}}(n).\end{aligned} \tag{2}$$

The target source (i.e. the talker) is modeled as a point-source. This implies that $\mathbf{s}(n)$ may be decomposed as:

$$\mathbf{s}(n) = s(n)\mathbf{d}, \tag{3}$$

where the scalar signal $s(n)$ represents the target signal at a certain reference position and the elements of the vector $\mathbf{d}$ represent relative transfer functions (RTFs) of the target signal from the reference position to all of the microphones. In general, the target signal may include the speech propagating via the direct path and (some of) the early reflections present in the room impulse response (RIR). In this case, elements of $\mathbf{d}$ are computed as complex Fourier coefficients of the relevant part of the RIRs from the reference position to the individual microphones. Including early reflections in $\mathbf{d}$ may become problematic in scenarios where the RIRs are variable and/or unknown. In such cases it may be more practical to define target signal as the direct path speech only. In this scenario $\mathbf{d}$ is solely a function of the Direction of Arrival (DoA) of the direct speech and not the particular room the system is situated in. This simplified and more practically operational definition of the target signal has been used in [6, 7] and will be used in the present study as well.

In the evaluated algorithm the reverberation $\mathbf{r}(n)$ is assumed to be cylindrically isotropic. Taking this and (3) into account, (2) can be rewritten as:

$$\mathbf{\Phi}_{\mathbf{y}}(n) = \underbrace{\phi_s(n)\mathbf{d}\mathbf{d}^H}_{\mathbf{\Phi}_{\mathbf{s}}(n)} + \underbrace{\phi_r(n)\mathbf{\Gamma}_{\text{iso}}}_{\mathbf{\Phi}_{\mathbf{r}}(n)}, \tag{4}$$

where $\phi_s(n)$ and $\phi_r(n)$ are, respectively, (scalar and time-varying) spectral variances of the speech and of the reverberation component at the reference position. The matrix $\mathbf{\Gamma}_{\text{iso}}$ is the normalized covariance matrix of the isotropic sound field, and similarly to $\mathbf{d}$, is assumed to be known and constant. Thanks to the isotropy of the reverberation, $\mathbf{\Gamma}_{\text{iso}}$ depends only on the geometry of the microphone array and the HA user's head. It does not depend on the position or orientation of the microphone array in the room.

2.2. Multi-channel Wiener Filter

The structure of the MWF-based algorithm is depicted in Fig. 1. In the considered signal model, the MWF may be factored into a Minimum Variance Distortionless Response (MVDR) beamformer and a single channel Wiener post-filter, so that the MWF may be expressed

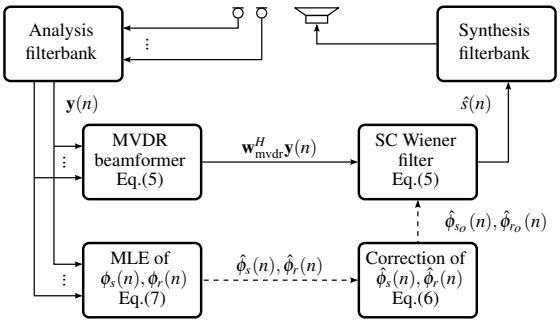

Fig. 1: Block diagram of the evaluated algorithm.

as [3,4]:

$$\hat{s}(n) = \mathbf{w}_{\text{mwf}}^H(n)\mathbf{y}(n), \text{ where} \tag{5a}$$

$$\mathbf{w}_{\text{mwf}}(n) = \underbrace{\left[\frac{\phi_{s_o}(n)}{\phi_{s_o}(n)+\phi_{r_o}(n)}\right]}_{g_{\text{sc}}(n)} \underbrace{\frac{\mathbf{\Gamma}_{\text{iso}}^{-1}\mathbf{d}}{\mathbf{d}^H\mathbf{\Gamma}_{\text{iso}}^{-1}\mathbf{d}}}_{\mathbf{w}_{\text{mvdr}}}. \tag{5b}$$

In (5b), the vector of MVDR beamformer coefficients and the Wiener post-filter gain are denoted as $\mathbf{w}_{\text{mvdr}}$ and $g_{\text{sc}}(n)$, respectively. $\phi_{s_o}(n)$ and $\phi_{r_o}(n)$ denote the spectral variances of the speech and the reverberation at the output of the MVDR beamformer and are related to $\phi_s(n)$ and $\phi_r(n)$ by:

$$\phi_{s_o}(n) = \phi_s(n), \tag{6a}$$

$$\phi_{r_o}(n) = \phi_r(n)(\mathbf{d}^H\mathbf{\Gamma}_{\text{iso}}^{-1}\mathbf{d})^{-1}, \tag{6b}$$

i.e. the MVDR beamformer does not distort the variance of the speech (6a), but the variance of the reverberation has to be corrected by the beamformer suppression factor (6b). The MVDR coefficient vector $\mathbf{w}_{\text{mvdr}}$ is completely determined by $\mathbf{d}$ and $\mathbf{\Gamma}_{\text{iso}}$ whereas the Wiener post-filter $g_{\text{sc}}(n)$ depends on the spectral variances $\phi_{s_o}(n)$ and $\phi_{r_o}(n)$ which are unknown and have to be estimated from the reverberant observations $\mathbf{y}(n)$.

2.3. PSD estimators

Estimation of $\phi_s(n)$ and $\phi_r(n)$ in speech dereverberation and in noise reduction contexts is an active topic of research, e.g. [5–8]. The algorithm from [6] uses Maximum Likelihood Estimators (MLEs) from [9] which may be expressed as:

$$\hat{\phi}_r(n) = \frac{1}{M-1}\,\text{tr}\left\{\left(\mathbf{I}-\mathbf{d}\mathbf{w}_{\text{mvdr}}^H\right)\hat{\mathbf{\Phi}}_{\mathbf{y}}(n)\,\mathbf{\Gamma}_{\text{iso}}^{-1}\right\}, \tag{7a}$$

$$\hat{\phi}_s(n) = \mathbf{w}_{\text{mvdr}}^H\left(\hat{\mathbf{\Phi}}_{\mathbf{y}}(n) - \hat{\phi}_r(n)\,\mathbf{\Gamma}_{\text{iso}}\right)\mathbf{w}_{\text{mvdr}}, \tag{7b}$$

where $\hat{\mathbf{\Phi}}_{\mathbf{y}}(n)$ denotes the estimate of the covariance matrix of $\mathbf{y}(n)$, and $\text{tr}\{\cdot\}$ denotes the matrix trace operator.

3. SIMULATION EXPERIMENTS

In this section we describe two experiments that we conducted to demonstrate the influence of the number and location of the microphones on the performance of the MVDR beamformers and the MWFs in HA systems. In the second experiment we also evaluate the robustness of these algorithms to errors in DoA estimation.

In both experiments we compare four different configurations of the microphones of a pair of two-microphone behind-the-ear HAs (inter-microphone distance $\sim$1 cm) The first configuration uses only the two microphones of a single HA on the left ear of the user. We denote this configuration graphically by $\overset{\otimes}{\times}\!\bigcirc$, where $\bigcirc$ symbolizes the HA user's head and $^\times$ is used to illustrate the number and position of the microphones. Symbol $^\otimes$ is used for the reference position used in the calculation of $\mathbf{d}$ (we always use the front microphone of the left HA). The three remaining microphone array configurations: $^\otimes\!\bigcirc^\times$, $\overset{\otimes}{\times}\!\bigcirc^\times$, and $\overset{\otimes}{\times}\!\bigcirc\!\overset{\times}{\times}$, use microphones from both left and right HA and consist of two, three, and four microphones, respectively. The configurations $\overset{\otimes}{\times}\!\bigcirc$ and $^\otimes\!\bigcirc^\times$ both use only two microphones, but while $\overset{\otimes}{\times}\!\bigcirc$ uses the microphones of a single HA, $^\otimes\!\bigcirc^\times$ uses only the front microphones of both the left and right HA. Comparison of the results obtained using these two configurations will allow us to differentiate between the influence of the *number* and the *location* of the microphones used in a HA system.

3.1. Beampatterns of the MVDR beamformer

The first of the conducted experiments focused on the relationship between the microphone array configurations and the beampatterns of the corresponding MVDR beamformers.

In order to calculate the MVDR coefficient vector $\mathbf{w}_{\text{mvdr}}$ the vector $\mathbf{d}$ and the matrix $\mathbf{\Gamma}_{\text{iso}}$ are needed. We derived the RTF vector $\mathbf{d}$ from measurements made in an anechoic chamber using the microphones of a pair of behind-the-ear HAs placed on the ears of a Head and Torso Acoustic Simulator (HATS), and a loudspeaker positioned directly in front of the HATS. The matrix $\mathbf{\Gamma}_{\text{iso}}$ was derived from measurements using an analogous HA/HATS setup in a simulated cylindrically isotropic sound field. The beamformer coefficients were computed according to the definition of $\mathbf{w}_{\text{mvdr}}$ given in (5b). The beampatterns of the computed beamformers are presented in Fig. 2.

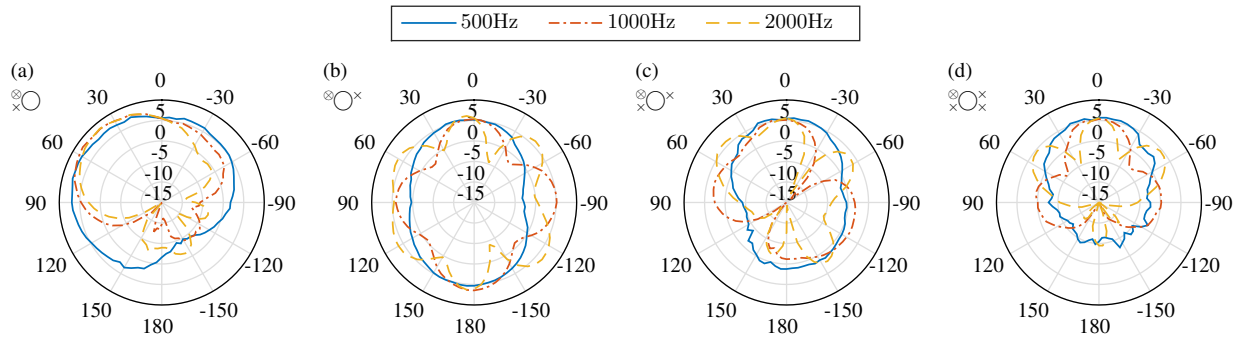

Fig. 2: Beampatterns of the MVDR beamformer using different microphone configurations at three different frequencies. Steering direction is 0°, i.e. towards the front. Positive values of the azimuth angle correspond to the left half of the horizontal plane. Radial coordinate of the above polar plots is in decibels.

Comparison of the beampatterns in Fig. 2a, 2c, and 2d ($\overset{\otimes}{\times}\!\!\bigcirc$, $\overset{\otimes}{\times}\!\!\bigcirc^{\times}$, and $\overset{\otimes}{\times}\!\!\bigcirc^{\times}_{\times}$, respectively) reveals that the suppression of sounds coming from directions other than the target direction (0°) increases with each added microphone. Moreover, the width of the main lobe of the beampattern decreases with the addition of contralateral microphones, especially for high frequencies. Comparing the beampatterns obtained using configurations $\overset{\otimes}{\times}\!\!\bigcirc$ and $\overset{\otimes}{\times}\!\!\bigcirc^{\times}$ (Fig. 2a and 2b) one can conclude that although both of them use the same number of microphones, $\overset{\otimes}{\times}\!\!\bigcirc$ has a wide main lobe and good suppression of the interference, while $\overset{\otimes}{\times}\!\!\bigcirc^{\times}$ has a narrow main lobe but does not suppress the interference impinging from the back.

In many applications the narrowness of the main lobe of the beampattern is a desirable property. In HAs, however, it can have a negative effect on the usability of a given processing algorithm. For example, in HA applications it is not an uncommon practice to assume that the target speaker is situated directly in front of the user's head. This assumption is reasonable since the HA user is usually looking at the person he/she is listening to. Based on this assumption the vector **d** corresponding to the DoA equal to 0° is frequently used. In monaural HAs ($\overset{\otimes}{\times}\!\!\bigcirc$) this method may be used successfully because the beamwidth of the MVDR beamformer allows for considerable deviations of the talker position from the assumed DoA (cf. Fig. 2a). In binaural HAs (i.e. using microphones of the contralateral HA: $\overset{\otimes}{\times}\!\!\bigcirc^{\times}$, $\overset{\otimes}{\times}\!\!\bigcirc^{\times}$, and $\overset{\otimes}{\times}\!\!\bigcirc^{\times}_{\times}$) the main lobe of the beampattern is narrower, which limits the usefulness of this method. A possible solution to that problem is to use other beamformer designs which are implicitly robust to DoA errors. Alternatively, an online estimator of **d** could be used. In this case the error of the used DoA estimator would need to be low enough such that the target speaker is always within the main lobe of the beamformer.

Since the MWF is equivalent to an MVDR beamformer followed by a post-filter, which in [6] also depends on the MVDR coefficients (cf. (7)), it is reasonable to expect that the issues pointed out above will affect the MWF as much, or perhaps even more, than they do the MVDR beamformer. For this reason in the second experiment we evaluated not only the performance of the MWF-based speech dereverberation algorithm from [6] when the assumed DoA is correct, but also how this performance depends on the DoA error. The adopted experimental method described in the following section is an attempt to make this assessment.

3.2. Performance and robustness to DoA mismatch – the MWF and the MVDR beamformer

We implemented the second experiment as a computer simulation in which the input signal was prepared by convolving the International Speech Test Signal (ISTS) [10] with multi-channel Impulse Responses (IRs). Two reverberant conditions were simulated.

The first of the reverberant conditions was denoted "Office" and the corresponding IRs were recorded in a real room using the HA/HATS setup described in Section 3.1. The room used for the IR recording had a rectangular shape and the walls were made of highly reflective materials (painted concrete and glass). This resulted in modal resonances and relatively long reverberation time (1.4s) in this condition. The direct-to-reverberant ratio (DRR)

was equal to 2.3 dB

The second reverberant condition was denoted "Isotropic" and it simulated cylindrically isotropic, exponentially decaying reverberation. The direct path response of the synthesized IR was simulated using the impulse response of the HA/HATS microphone array recorded in an anechoic chamber with a sound source directly in front of the HATS (i.e. analogously to the measurement used to compute **d** in Section 3.1). The isotropic reverberation tail was simulated by a superposition of 72 exponentially decaying white noise sequences filtered by the impulse responses of the HA/HATS array measured for 72 equally spaced horizontal directions (every 5°). The simulated reverberation time was 1 s and the DRR was set to 0 dB.

In all iterations of the experiment the position of the simulated speaker was directly in front of the HATS and the reverberation was assumed to be isotropic (both in "Isotropic" and in "Office" conditions). The sensitivity of the MWF and the MVDR beamformers to errors in the DoA was assessed by repeating the experiment for different *assumed* DoAs of the target speech. In other words, the actual target source position was always in the front, but in each iteration an anechoic vector **d** corresponding to a different DoA was used. PESQ [11] and STOI [12] were used as performance measures.

4. RESULTS

The results of the experiment are presented in Figure 3 as a function of the assumed DoA. The actual DoA is marked with an arrow. Along the PESQ and STOI scores obtained from the outputs of the MWFs and the MVDR beamformers, the scores computed from the unprocessed sound (denoted "Unprocessed") are also included for reference. As expected, all four configurations of both the MVDR beamformer and the MWF resulted in an increase of the used performance measures when the assumed DoA was close to the correct one. In both reverberant conditions the greatest PESQ and STOI improvements were obtained by using the four microphone configuration ($^{\otimes}_{\times}\bigcirc^{\times}_{\times}$) of either the MVDR beamformer or the MWF. The obtained scores of the configurations $^{\otimes}_{\times}\bigcirc^{\times}$, $^{\otimes}_{\times}\bigcirc$, and $^{\otimes}\bigcirc^{\times}$ were progressively worse (in that order), which is in line with the earlier findings that the speech dereverberation performance of the MWF increases with every added microphone (cf. [5, 6]).

Comparing the results obtained using the MVDR beamformers and the MWFs it may be concluded that the scores obtained using the MWFs depend on the assumed DoA in a similar way as the scores obtained by using the corresponding MVDR beamformers. In other words, the MWFs appear to be equally sensitive to incorrect DoAs as the MVDR beamformers.

The configurations $^{\otimes}_{\times}\bigcirc^{\times}_{\times}$ and $^{\otimes}_{\times}\bigcirc^{\times}$ of the MVDR beamformer and of the MWF appear to be superior to $^{\otimes}_{\times}\bigcirc$ only for relatively small deviations of the assumed DoA (approximately $\pm 15°$). It is an important result with significant consequences for the practical use of the MWF in binaural HAs. Specifically, the range of DoAs where the binaural MWF offers any advantage over the simpler monaural implementation may be too narrow to justify the computational and implementational cost of the binaural link if a constant *a priori* DoA is used (like in [5, 6]). The use of the binaural MWF may only be justified if a good estimator of the DoA is implemented and the error of estimation is lower than $\pm 15°$. The PESQ and STOI scores obtained by the monaural configuration $^{\otimes}_{\times}\bigcirc$ of the algorithms exhibit a wide plateau. This suggests that in this configuration the assumed DoA does not have to match the actual DoA accurately; even with only a rough estimate of the DoA good results can be achieved. Compared to other configurations $^{\otimes}\bigcirc^{\times}$ performed poorly. This is due to the fact that the DoAs from the front and the back can not be distinguished using only one microphone from each HA. This is analogous to the front-back confusion observed in human subjects.

5. CONCLUSION

In this paper we have evaluated the influence of erroneous DoA estimates on the performance of an MWF-based speech dereverberation algorithm in monaural and binaural hearing aids. The results indicate that binaural configurations of the MWF are far more sensitive to errors in DoA than the monaural configurations. In result, although sometimes not necessary in monaural hearing aids, in binaural configurations of the MWF the use of an accurate on-line DoA estimator seems necessary in order to achieve the full potential of this method.

6. REFERENCES

[1] V. Hamacher et al. Signal processing in high-end hearing aids: State of the art, challenges, and future trends. *EURASIP J. Appl. Signal Process.*, 2005:2915–2929, January 2005.

[2] V. Hamacher et al. Binaural signal processing in hearing aids: technologies and algorithms.

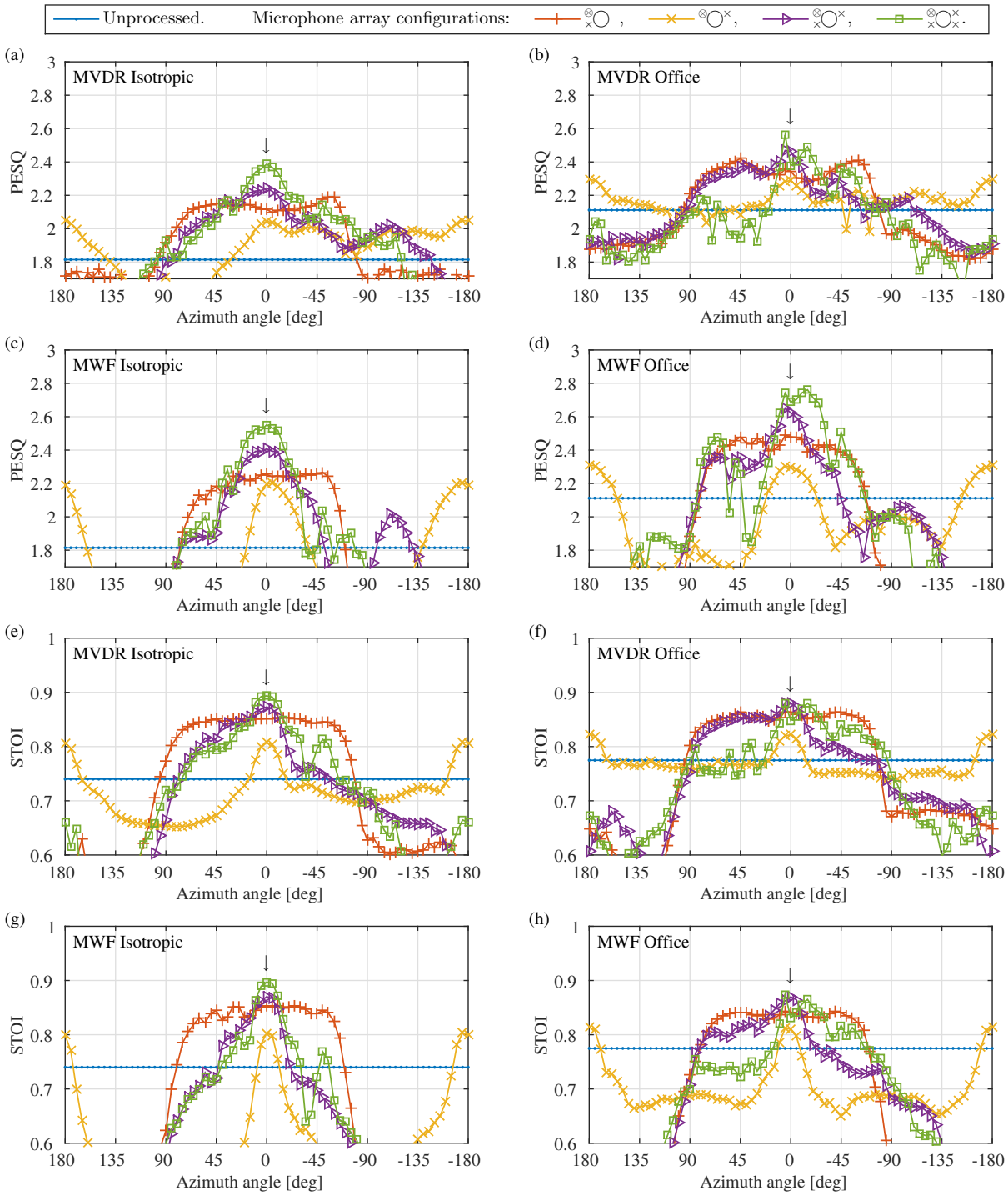

Fig. 3: PESQ (a–d) and STOI (e–h) scores obtained by different configurations of the MVDR beamformer and the MWF in the Isotropic and in the Office reverberation conditions as a function of the assumed DoA. The actual position of the source is marked with an arrow.

In Rainer Martin, Ulrich Heute, and Christiane Antweiler, editors, *Advances in Digital Speech Transmission*, chapter 14, pages 401–429. Wiley, 2008.

[3] S. Doclo et al. Acoustic beamforming for hearing aid applications. In S. Haykin and K. J. Ray Liu, editors, *Handbook on Array Processing and Sensor Networks*, pages 269–302. Wiley, 2008.

[4] S. Doclo et al. Frequency-domain criterion for the speech distortion weighted multichannel Wiener filter for robust noise reduction. *Speech Communication*, 49(7-8):636–656, JUL-AUG 2007.

[5] S. Braun and E.A.P. Habets. Dereverberation in noisy environments using reference signals and a maximum likelihood estimator. In *Signal Processing Conference (EUSIPCO), Proceedings of the 21st European*, pages 1–5, Marrakech, Morocco, 2013.

[6] A. Kuklasiński et al. Maximum likelihood based multi-channel isotropic reverberation reduction for hearing aids. In *Signal Processing Conference (EUSIPCO), Proceedings of the 22nd European*, pages 61–65, Lisbon, Portugal, 2014.

[7] A. Kuklasiński et al. Multi-channel PSD estimators for speech dereverberation – a theoretical and experimental comparison. In *Acoustics, Speech and Signal Processing, IEEE International Conference on*, pages 91–95, Brisbane, Australia, 2015.

[8] U. Kjems and J. Jensen. Maximum likelihood based noise covariance matrix estimation for multi-microphone speech enhancement. In *Signal Processing Conference (EUSIPCO), Proceedings of the 20th European*, pages 295–299, Bucharest, Romania, 2012.

[9] H. Ye and R.D. DeGroat. Maximum likelihood DOA estimation and asymptotic Cramér-Rao bounds for additive unknown colored noise. *IEEE Trans. Signal Process.*, 43(4):938–949, 1995.

[10] I. Holube et al. Development and analysis of an international speech test signal (ists). *International Journal of Audiology*, 49(12):891–903, 2010.

[11] Perceptual evaluation of speech quality: an objective method for end-to-end speech quality assessment of narrow-band telephone networks and speech codecs. *ITU-T Rec. P. 862*, 2001.

[12] C.H. Taal et al. An algorithm for intelligibility prediction of time-frequency weighted noisy speech. *IEEE Trans. Audio, Speech, Language Process.*, 19(7):2125–2136, 2011.