Spatio-Temporal Audio Enhancement Based on IAA Noise Covariance Matrix Estimates

Nørholm, Sidsel Marie; Jensen, Jesper Rindom; Christensen, Mads Græsbøll

Published in:
2014 Proceedings of the 22nd European Signal Processing Conference (EUSIPCO 2014)

Publication date:
2014

Document Version
Early version, also known as pre-print

SPATIO-TEMPORAL AUDIO ENHANCEMENT BASED ON IAA NOISE COVARIANCE MATRIX ESTIMATES

Sidsel Marie Nørholm, Jesper Rindom Jensen, and Mads Græsbøll Christensen

Audio Analysis Lab, AD:MT, Aalborg University, {smn, jrj,mgc}@create.aau.dk

ABSTRACT

A method for estimating the noise covariance matrix in a multichannel setup is proposed. The method is based on the iterative adaptive approach (IAA), which only needs short segments of data to estimate the covariance matrix. Therefore, the method can be used for fast varying signals. The method is based on an assumption of the desired signal being harmonic, which is used for estimating the noise covariance matrix from the covariance matrix of the observed signal. The noise covariance estimate is used in the linearly constrained minimum variance (LCMV) filter and compared to an amplitude and phase estimation (APES) based filter. For a fixed number of samples, the performance in terms of signal-to-noise ratio can be increased by using the IAA method, whereas if the filter size is fixed and the number of samples in the APES based filter is increased, the APES based filter performs better.

Index Terms— Speech enhancement, iterative adaptive approach, multichannel, covariance estimates, harmonic signal model.

1. INTRODUCTION

In many applications such as teleconferencing, surveillance systems and hearing aids, it is desirable to extract one signal from an observation of the desired signal buried in noise. This can be done in several ways, in general separated in three groups: the spectral-subtractive methods, the statistical-model-based methods and the subspace methods [1]. In this work, we focus on the filtering methods, which are in the group of statistical-model-based methods. A filter will, preferably, pass the desired signal undistorted, whereas the noise is reduced. In the design of the filter, an estimate of the noise statistics is often needed. Therefore, this is a widely studied problem in the single-channel case, and several methods for estimating the noise statistics exist [2–6]. In the multi-channel case the problem is more difficult due to the cross-correlation between microphones. Some methods are proposed in [7–11]: in [7–10], the cross correlation elements are only updated in periods of unvoiced speech, which can be problematic in the case of non-stationary noise, whereas, in [11], the elements are updated continuously under the assumption that the position of the source is known. However, this is done by steering a null in the direction of the source which means that the filtering has to be done in two steps; a spatial filtering followed by a temporal. Another approach, used in the present work, is to take advantage of the nature of the desired signal. This signal is often voiced speech or musical instruments which is quasi-periodic, and, therefore, the focus in this paper is signals that can be modelled using the harmonic signal model. For speech signals, voiced/unvoiced detectors [12] make it possible to use the approach only on the voiced segments, which are the primary components of a speech signal. Knowing the parameters of the harmonic model, the noise statistics can be estimated by subtracting the desired signal contribution from the statistics of the observed signal. This approach is also taken in the amplitude and phase estimation (APES) filter [13–15]. However, since the APES filter is based on the sample covariance matrix, the number of samples has to be large, a problem which is even more pronounced in the multichannel setup. This can cause problems if the signal is fast varying and, therefore, not stationary over the interval used for estimating the sample covariance matrix.

In the present paper, the multichannel noise covariance matrix is estimated by the iterative adaptive approach (IAA) [16, 17], and the need for a high number of samples is, therefore, not present.The IAA covariance matrix estimate is modified according to the harmonic signal model to get an estimate of the noise covariance matrix and compared to an APES based filter for a harmonic signal.

The rest of the paper is organised as follows: in Section 2, the signal model is set up in the multichannel case. In Section 3, the used filtering method and the sample covariance matrix are introduced, elaborating the motivation for the IAA method. In Section 4, the IAA method for noise covariance matrix estimation is explained. Section 5 shows results, and Section 6 ends the work with a discussion.

2. SIGNAL MODEL

Considering an array of N_s microphones, the observed signal measured by the n_s'th microphone, for time index $n_t = 0, ..., N_t - 1$ and microphone $n_s = 0, ..., N_s - 1$ is: $x_{n_s}(n_t) = s_{n_s}(n_t) + v_{n_s}(n_t)$, where $s_{n_s}(n_t)$ is the desired signal and

This research was funded by the Villum Foundation and the Danish Council for Independent Research, grant ID: DFF 1337-00084

$v_{n_s}(n_t)$ is the noise. If the desired signal is harmonic, it can be written as a sum of complex sinusoids:

$$s_{n_s}(n_t) = \sum_{l=1}^{L} \alpha_l e^{jl\omega_t n_t} e^{-jl\omega_s n_s}, \tag{1}$$

where L is the number of harmonics in the signal, α_l is the complex amplitude of the l'th harmonic, ω_t is the temporal and ω_s is the spatial frequency. If the signal is real it can easily be transformed to its complex counterpart by use of the Hilbert transform [18]. In this paper, we assume anechoic far field conditions and sampling by a uniform linear array (ULA) with an equal spacing, d, between the microphones. Thereby, the relation between the temporal and spatial frequency is $\omega_s = \omega_t f_s c^{-1} d \sin\theta$, for the temporal sampling frequency f_s, the speed of sound in air c, and the direction of arrival (DOA) $\theta \in [-90°; 90°]$.

The processing of the observed signal is done on a subset of M_t observations in time and M_s observations in space defined by the matrix:

$$\mathbf{X}_{n_s}(n_t) = \begin{bmatrix} x_{n_s}(n_t) & \dots & x_{n_s}(n_t - M_t') \\ \vdots & \ddots & \vdots \\ x_{n_s+M_s'}(n_t) & \dots & x_{n_s+M_s'}(n_t - M_t') \end{bmatrix}, \tag{2}$$

with $M_t' = M_t - 1$ and $M_s' = M_s - 1$. The matrix is then put into vector format using the column-wise stacking operator $\text{vec}\{\cdot\}$, i.e., $\mathbf{x}_{n_s}(n_t) = \text{vec}\{\mathbf{X}_{n_s}(n_t)\}$.

3. FILTERING

To obtain an estimate of the desired signal, $\tilde{s}(n_t)$, from measurements of the noisy observation, $\mathbf{x}_{n_s}(n_t)$ is filtered by the filter $\mathbf{h}_{\omega_{t,s}}$, optimised for a harmonic signal with temporal fundamental frequency ω_t and spatial frequency ω_s. The spatio-temporal linearly constrained minimum variance (LCMV) filter is a good choice for filtering of periodic signals since the filter gain can be chosen to be one at the harmonic frequencies at the DOA of the observed signal whereas the overall output power of the filter is minimised. The filter is the solution to the minimisation problem [19]

$$\min_{\mathbf{h}} \mathbf{h}_{\omega_{t,s}}^H \mathbf{R} \mathbf{h}_{\omega_{t,s}} \quad \text{s.t.} \quad \mathbf{h}_{\omega_{t,s}}^H \mathbf{a}_{l\omega_{t,s}} = 1 \quad \text{for} \quad l = 1, ..., L. \tag{3}$$

Here, $\{\cdot\}^H$ denotes complex conjugate transpose, $\mathbf{R}$ is the covariance matrix of $\mathbf{x}_{n_s}(n_t)$, i.e., $\mathbf{R} = \text{E}\{\mathbf{x}_{n_s}(n_t)\mathbf{x}_{n_s}^H(n_t)\}$, and

$$\mathbf{a}_{l\omega_{t,s}} = \mathbf{a}_{l\omega_t} \otimes \mathbf{a}_{l\omega_s}, \tag{4}$$

$$\mathbf{a}_{\omega} = \begin{bmatrix} 1 & e^{-j\omega} & \dots & e^{-j\omega M'} \end{bmatrix}^T, \tag{5}$$

with $\otimes$ denoting the Kronecker product and $\{\cdot\}^T$ the transpose. The solution is given by:

$$\mathbf{h}_{\omega_{t,s}} = \mathbf{R}^{-1}\mathbf{A}_{\omega_{t,s}}(\mathbf{A}_{\omega_{t,s}}^H \mathbf{R}^{-1}\mathbf{A}_{\omega_{t,s}})^{-1}\mathbf{1}, \tag{6}$$

where $\mathbf{1}$ is an $L \times 1$ vector containing ones and $\mathbf{A}_{\omega_{t,s}}$ is the spatio-temporal steering matrix

$$\mathbf{A}_{\omega_{t,s}} = \begin{bmatrix} \mathbf{a}_{\omega_{t,s}} & \dots & \mathbf{a}_{L\omega_{t,s}} \end{bmatrix}. \tag{7}$$

The covariance matrix is an unknown quantity and is most often replaced by the sample covariance matrix

$$\widehat{\mathbf{R}} = \sum_{p=0}^{N_t-M_t} \sum_{q=0}^{N_s-M_s} \frac{\mathbf{x}_q(n_t-p)\mathbf{x}_q^H(n_t-p)}{(N_t-M_t')(N_s-M_s')}. \tag{8}$$

If the covariance matrix in (3) is replaced by the noise covariance matrix, only the noise power output, and not the overall output power, will be minimised. This will, most often, give better filtering results since perturbations in DOA and fundamental frequency estimates cause a mismatch between the DOA and fundamental frequency of the signal and those used for constraining the LCMV filter, leading to badly regularised filters and signal cancellation. The noise covariance matrix can, for example, be estimated by an amplitude and phase estimation (APES) based approach, as in [20], where a spatio-temporal form of the APES filter [14] is derived. A harmonic signal model is assumed for the desired signal and the part of the sample covariance matrix resembling this signal is then subtracted to give an estimate of the noise covariance matrix. One drawback of both the sample covariance estimate and the APES based covariance estimate is that, in order to make the covariance matrix full rank, the following relation between N_t, N_s, M_t and M_s has to be fulfilled: $(N_t - M_t + 1)(N_s - M_s + 1) \geq M_t M_s$. Normally, there will be a restriction on the number of microphones available, and N_s will, therefore, be fairly small. In order to get a good spatial resolution it is then desirable to choose M_s close or equal to N_s, thereby forcing N_t to be very large compared to M_t. This can be problematic if the signal is not stationary for longer periods of time. Therefore, an alternative method for estimation of the covariance matrix is proposed, where, preferably, $M_t = N_t$ and $M_s = N_s$.

4. IAA COVARIANCE MATRIX ESTIMATES

The iterative adaptive approach (IAA) is a method for estimating the spectral amplitudes, $\alpha_{\Omega_{g,k}}$, in the observed signal for temporal and spatial frequency bins:

$$\mathbf{\Omega}_G = \begin{bmatrix} 0 & 2\pi\frac{1}{G} & \dots & 2\pi\frac{G-1}{G} \end{bmatrix}, \tag{9}$$

$$\mathbf{\Omega}_K = \begin{bmatrix} 0 & 2\pi\frac{1}{K} & \dots & 2\pi\frac{K-1}{K} \end{bmatrix}, \tag{10}$$

initialisation

$$\widetilde{\alpha}_{\Omega_{g,k}} = \frac{\mathbf{a}^H_{\Omega_{g,k}} \mathbf{x}_{n_s}(n_t)}{\mathbf{a}^H_{\Omega_{g,k}} \mathbf{a}_{\Omega_{g,k}}},$$

$$g = 0, ..., G-1, \quad k = 0, ..., K-1.$$

repeat

$$\widetilde{\mathbf{R}} = \sum_{g=0}^{G-1} \sum_{k=0}^{K-1} |\widetilde{\alpha}_{\Omega_{g,k}}|^2 \mathbf{a}_{\Omega_{g,k}} \mathbf{a}^H_{\Omega_{g,k}},$$

$$\widetilde{\alpha}_{\Omega_{g,k}} = \frac{\mathbf{a}^H_{\Omega_{g,k}} \widetilde{\mathbf{R}}^{-1} \mathbf{x}_{n_s}(n_t)}{\mathbf{a}^H_{\Omega_{g,k}} \widetilde{\mathbf{R}}^{-1} \mathbf{a}_{\Omega_{g,k}}},$$

$$g = 0, ..., G-1, \quad k = 0, ..., K-1.$$

until (convergence)

Table 1: IAA for spatio-temporal covariance matrix estimation.

where G and K are the temporal and spatial frequency grid sizes. Element g and k in (9) and (10) are denoted as Ω_g and Ω_k, respectively, and a combination of frequencies Ω_g and Ω_k is denoted by $\Omega_{g,k}$. The amplitudes are estimated by minimisation of a weighted least squares (WLS) cost function [17, 20]

$$J_{\text{WLS}} = \left[\mathbf{x}_{n_s}(n_t) - \alpha_{\Omega_{g,k}} \mathbf{a}_{\Omega_{g,k}}\right]^H \mathbf{Q}^{-1}_{\Omega_{g,k}} \left[\mathbf{x}_{n_s}(n_t) - \alpha_{\Omega_{g,k}} \mathbf{a}_{\Omega_{g,k}}\right], \tag{11}$$

where $\mathbf{a}_{\Omega_{g,k}}$ is given by (4) and (5) for $l = 1$, and $\mathbf{Q}_{\Omega_{g,k}}$ is the noise covariance matrix

$$\mathbf{Q}_{\Omega_{g,k}} = \mathbf{R} - |\alpha_{\Omega_{g,k}}|^2 \mathbf{a}_{\Omega_{g,k}} \mathbf{a}^H_{\Omega_{g,k}}. \tag{12}$$

The covariance matrix, $\mathbf{R}$, is not known, but is estimated as

$$\widetilde{\mathbf{R}} = \sum_{g=0}^{G-1} \sum_{k=0}^{K-1} |\alpha_{\Omega_{g,k}}|^2 \mathbf{a}_{\Omega_{g,k}} \mathbf{a}^H_{\Omega_{g,k}}. \tag{13}$$

The solution to the minimisation of (11) is [17, 20]

$$\widetilde{\alpha}_{\Omega_{g,k}} = \frac{\mathbf{a}^H_{\Omega_{g,k}} \mathbf{R}^{-1} \mathbf{x}_{n_s}(n_t)}{\mathbf{a}^H_{\Omega_{g,k}} \mathbf{R}^{-1} \mathbf{a}_{\Omega_{g,k}}}. \tag{14}$$

Since the estimate of the spectral amplitudes depends on the estimate of the covariance matrix and vice versa, they are estimated by iterating between (13) and (14). Typically, 10 to 15 iterations are sufficient for convergence [21]. The process is summarised in Table 1. With the IAA covariance matrix as a starting point, we find the noise covariance matrix as

$$\mathbf{Q}_{\omega_{t,s}} = \mathbf{R} - \sum_{l=1}^{L} |\alpha_{l\omega_{t,s}}|^2 \mathbf{a}_{l\omega_{t,s}} \mathbf{a}^H_{l\omega_{t,s}}. \tag{15}$$

Since the covariance matrix is estimated with a limited number of samples, the desired signal will leak into neighbouring frequency components. Therefore, we estimate the noise covariance matrix by also subtracting the neighbouring grid points to those corresponding to the harmonic frequencies:

$$\widetilde{\mathbf{Q}}_{\omega_{t,s}} = \widetilde{\mathbf{R}} - \sum_{l=1}^{L} \sum_{y=g_l-\delta}^{g_l+\delta} \sum_{z=k_l-\delta}^{k_l+\delta} |\widetilde{\alpha}_{\Omega_{y,z}}|^2 \mathbf{a}_{\Omega_{y,z}} \mathbf{a}^H_{\Omega_{y,z}},$$

where g_l and k_l are the grid indices corresponding to the l'th harmonic and 2δ is the number of subtracted neighbouring frequency grid points.

5. RESULTS

The IAA noise covariance estimates are tested by use of a synthetic harmonic signal with $\omega_t = 0.5027$ (corresponding to 200 Hz), $f_s = 2500$ Hz, $L = 5$, $\theta = 10°$ and $\alpha_l = 1 \, \forall \, l$. The speed of sound is set to $c = 343.2$ m/s and $d = c/f_s$. The individual microphone signals are artificially delayed according to d and θ. Noise is added to give a desired average input signal-to-noise ratio (SNR). The noise is white Gaussian noise passed through a 10'th order auto-regressive filter made using a harmonic signal with seven harmonics and a fundamental frequency of 137 Hz. For the IAA estimate $N_t = M_t = 20$, $N_s = M_s = 10$. To decrease computational complexity, the grid is modified to make a uniform grid containing the harmonic frequencies, and, thereby, the number of grid points can be decreased, here, $G = 400$ and $K = 71$, and the number of iterations is 10. Alternatively, if the harmonics are not placed on the grid, the relaxation in [22] can be utilised. When the covariance matrices of consecutive samples are estimated, the first estimate is initialised as in Table 1, the rest are initialised with the former estimate of the covariance matrix, and only one iteration is made [21]. The number of subtracted neighbouring frequency grid points is set to eight since this was observed to give the highest SNR.

The performance after filtering is measured by means of the output SNR, $\text{oSNR}(\mathbf{h}) = \frac{\sigma^2_{s,\text{nr}}}{\sigma^2_{v,\text{nr}}}$, with $\sigma^2_{s,\text{nr}}$ and $\sigma^2_{v,\text{nr}}$ being the variances of signal and noise after noise reduction. The variances are computed over 50 consecutive samples and the resulting output SNR is averaged over 100 runs.

The IAA noise covariance estimate, $\widetilde{\mathbf{Q}}_{\omega_{t,s}}$ ($\text{IAA}_{\widetilde{\mathbf{Q}}_{\omega_t,\omega_s}}$) is compared to the IAA covariance estimate $\widetilde{\mathbf{R}}$ ($\text{IAA}_{\widetilde{\mathbf{R}}}$), the IAA noise covariance estimate based on the clean noise signal ($\text{IAA}_{\mathbf{Q}}$) and to the APES based estimate with two different configurations. In the first (APES_1), the number of samples is the same as for the IAA filter whereas the filter length is shorter, $N_t = 20$, $M_t = 10$, $N_s = 10$ and $M_s = 5$. In the second (APES_2), the filter length is the same as in the IAA, but longer data segments are used, $N_t = 224$, $M_t = 20$, $N_s = 10$ and $M_s = 10$. The methods are compared by using the covariance matrix estimates in the LCMV filter. Examples of filter responses are shown in Fig. 1 for an average input SNR of 10 dB. Comparing (a) to (b), it is seen that taking account for the desired signal in the generation of the filter

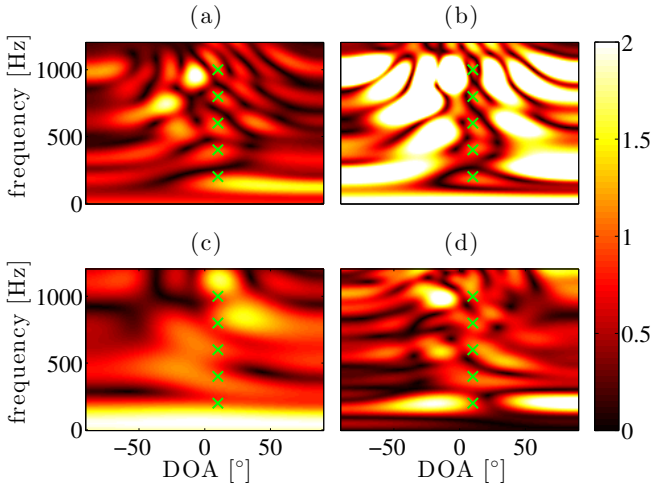

Fig. 1: Filter responses for (a) IAA based on noise covariance matrix estimate $\widetilde{\mathbf{Q}}_{\omega_{t,s}}$ (b) IAA based on covariance matrix estimate, $\widetilde{\mathbf{R}}$ (c) APES based estimate with $N_t = 20$, $M_t = 10$, $N_s = 10$, and $M_s = 5$ (d) APES based estimate with $N_t = 224$, $M_t = 20$, $N_s = 10$, and $M_s = 10$. Harmonics of desired signal are marked by green crosses. The average input SNR is 10 dB.

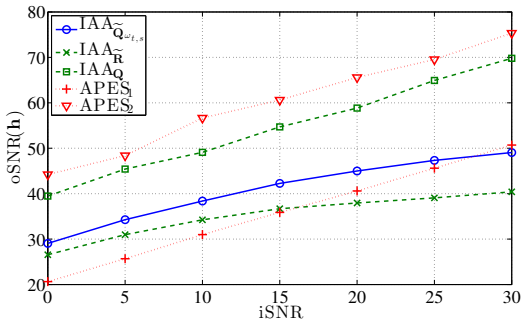

Fig. 2: Output SNR as a function of the input SNR.

gives a much more well conditioned filter. Comparing to (c), (a) has more attenuation at other DOAs and frequencies than the ones of the desired signal, whereas it is difficult to say whether the filter in (a) or (d) will have the best performance.

The output SNR are shown as a function of the input SNR in Fig. 2. For input SNRs from 0 to 10 dB, a gain in SNR of approximately 8 dB can be obtained compared to APES$_1$. At higher input SNRs, the gain decreases. If more samples are available, APES$_2$ outperforms IAA, but then the noise covariance matrix has been estimated on the basis of 4480 samples of the signal compared to only 200 with the IAA method.

The IAA method is tested on a piece of a speech signal sampled at 8 kHz. The fundamental frequency is estimated from the desired signal with an approximate nonlinear least squares estimator [23], and the model order is set to 18. Due to the high model order, here $N_t = M_t = 50$. The DOA, N_s, M_s, c and d are the same as before. Based on the fundamental frequency estimate, we design the grid at each time

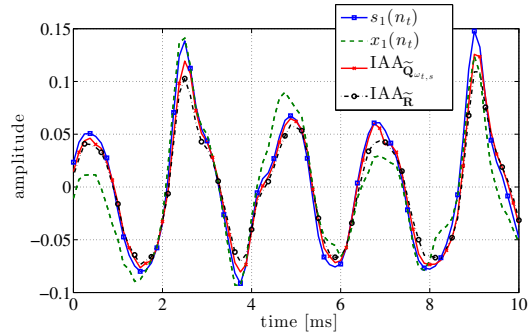

Fig. 3: Reconstructed signal using IAA$_{\widetilde{\mathbf{Q}}_{\omega_{t,s}}}$ compared to reconstruction using IAA$_{\widetilde{\mathbf{R}}}$ and the desired and noisy signal from the first microphone.

instance such that the harmonics lie on the grid, which means that the grid size varies slightly over time, with approximate values of $G = 400$ and $K = 100$. The ten microphone recordings are made using the room impulse response generator [24] under anechoic conditions with a distance of 5 m between source and microphone array. Babble noise from the AURORA database [25] is added to the microphone signals to give an average input SNR of 10 dB.

A short segment of the noisy, desired and estimated signal using, respectively, the proposed IAA noise covariance matrix estimate, $\widetilde{\mathbf{Q}}_{\omega_{t,s}}$, and the IAA covariance matrix estimate, $\widetilde{\mathbf{R}}$, are plotted in Fig. 3. It is seen in the figure that IAA$_{\widetilde{\mathbf{Q}}_{\omega_{t,s}}}$ gives a good estimate of the desired signal and follows the desired signal more closely than the IAA$_{\widetilde{\mathbf{R}}}$ estimate.

6. DISCUSSION

In the present paper, we suggest a method for estimation of the noise covariance matrix based on the iterative adaptive approach (IAA). The method only needs a single snapshot of data to estimate the covariance matrix. This makes it advantageous when fast varying signals are considered. In speech enhancement, IAA has formerly been used for fundamental frequency estimation [20] and joint direction of arrival (DOA) and fundamental frequency estimation [22], both assumed known in the present paper. Here, the covariance matrix estimate from the IAA is modified, under the assumption of a harmonic desired signal, to give an estimate of the noise covariance matrix. This estimate is then used in the linearly constrained minimum variance (LCMV) filter and compared to a spatio-temporal APES based filter proposed in [15]. The proposed method shows better performance in terms of signal-to-noise ratio (SNR) when the number of samples is limited, whereas the APES based filter has a better performance when the number of samples is not an issue. Compared to [11], where the filtering has to be done in two steps, the work presented here does the spatial and temporal filtering jointly.

REFERENCES

[1] P. Loizou, *Speech Enhancement: Theory and Practice*, CRC Press, 2007.

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

[3] I. Cohen and B. Berdugo, "Noise estimation by minima controlled recursive averaging for robust speech enhancement," *IEEE Signal Process. Lett.*, vol. 9, no. 1, pp. 12–15, Jan. 2002.

[4] L. Lin, W. H. Holmes, and E. Ambikairajah, "Subband noise estimation for speech enhancement using a perceptual Wiener filter," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process.*, Apr. 2003, vol. 1, pp. 80–83.

[5] R. C. Hendriks, R. Heusdens, and J. Jensen, "MMSE based noise psd tracking with low complexity," in *IEEE Trans. Acoust., Speech, Signal Process.*, 2010, pp. 4266–4269.

[6] D. Ealey, H. Kelleher, and D. Pearce, "Harmonic tunnelling: tracking non-stationary noises during speech," in *Proc. Eurospeech*, Sep. 2001, pp. 437–440.

[7] R. L. Bouquin-Jeannès, A. A. Azirani, and G. Faucon, "Enhancement of speech degraded by coherent and incoherent noise using a cross-spectral estimator," *IEEE Trans. Speech Audio Process.*, vol. 5, no. 5, pp. 484–487, Sep. 1997.

[8] X. Zhang and Y. Jia, "A soft decision based noise cross power spectral density estimation for two-microphone speech enhancement systems," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process.*, Mar. 2005, vol. 1, pp. 813–816.

[9] M. Rahmani, A. Akbari, B. Ayad, M. Mazoochi, and M. S. Moin, "A modified coherence based method for dual microphone speech enhancement," in *Proc. IEEE Int. Conf. Signal Process. Commun.*, Nov. 2007, pp. 225–228.

[10] J. Freudenberger, S. Stenzel, and B. Venditti, "A noise PSD and cross-PSD estimation for two-microphone speech enhancement systems," in *Proc. IEEE Workshop Statist. Signal Process.*, Aug. 2009, pp. 709–712.

[11] R. C. Hendriks and T. Gerkmann, "Noise correlation matrix estimation for multi-microphone speech enhancement," *IEEE Trans. Audio, Speech, and Language Process.*, vol. 20, no. 1, pp. 223–233, Jan. 2012.

[12] K.I. Molla, K. Hirose, N. Minematsu, and K. Hasan, "Voiced/unvoiced detection of speech signals using empirical mode decomposition model," in *Int. Conf. Information and Communication Technology*, March 2007, pp. 311–314.

[13] J. Li and P. Stoica, "An adaptive filtering approach to spectral estimation and SAR imaging," *IEEE Trans. Signal Process.*, vol. 44, no. 6, pp. 1469–1484, Jun. 1996.

[14] P. Stoica and R. Moses, *Spectral Analysis of Signals*, Pearson Education, Inc., 2005.

[15] J. R. Jensen, M. G. Christensen, and S. H. Jensen, "An optimal spatio-temporal filter for extraction and enhancement of multi-channel periodic signals," in *Rec. Asilomar Conf. Signals, Systems, and Computers*, Nov. 2010, pp. 1846–1850.

[16] T. Yardibi, J. Li, P. Stoica, M. Xue, and A. B. Baggeroer, "Source localization and sensing: A nonparametric iterative adaptive approach based on weighted least squares," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 46, no. 1, pp. 425–443, Jan. 2010.

[17] W. Roberts, Petre Stoica, Jian Li, T. Yardibi, and F.A. Sadjadi, "Iterative adaptive approaches to mimo radar imaging," *IEEE Journal of Selected Topics in Signal Processing*, vol. 4, no. 1, pp. 5–20, Feb. 2010.

[18] M. G. Christensen and A. Jakobsson, "Multi-pitch estimation," *Synthesis Lectures on Speech and Audio Processing*, vol. 5, no. 1, pp. 1–160, 2009.

[19] O. L. Frost, III, "An algorithm for linearly constrained adaptive array processing," *Proc. IEEE*, vol. 60, no. 8, pp. 926–935, Aug. 1972.

[20] J. R. Jensen, M. G. Christensen, and S. H. Jensen, "A single snapshot optimal filtering method for fundamental frequency estimation," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process.*, May 2011, pp. 4272–4275.

[21] G. O. Glentis and A. Jakobsson, "Time-recursive IAA spectral estimation," *IEEE Signal Process. Lett.*, vol. 18, no. 2, pp. 111–114, 2011.

[22] Z. Zhou, M. G. Christensen, J. R. Jensen, and H. C. So, "Joint DOA and fundamental frequency estimation based on relaxed iterative adaptive approach and optimal filtering," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process.*, May 2013.

[23] M. G. Christensen, P. Stoica, A. Jakobsson, and S. H. Jensen, "Multi-pitch estimation," *Elsevier Signal Process.*, vol. 88, no. 4, pp. 972–983, Apr. 2008.

[24] E. A. P. Habets, "Room impulse response generator," Tech. Rep., Technische Universiteit Eindhoven, 2010, Ver. 2.0.20100920.

[25] D. Pearce and H. G. Hirsch, "The AURORA experimental framework for the performance evaluation of speech recognition systems under noisy conditions," in *Proc. Int. Conf. Spoken Language Process.*, Oct 2000.