MULTI-MICROPHONE NOISE REDUCTION USING INTERCHANNEL AND INTERFRAME CORRELATIONS

Emanuël A. P. Habets[1], *Jacob Benesty*[2], *Jingdong Chen*[3]

[1] International Audio Laboratories Erlangen[†]
Am Wolfsmantel 33, 91058 Erlangen, Germany
emanuel.habets@audiolabs-erlangen.de

[2] INRS-EMT, University of Quebec
Montreal, Quebec, H5A 1K6, Canada
benesty@emt.inrs.ca

[3] Northwestern Polytechnical University
Xi'an, Shaanxi 710072, China
jingdongchen@ieee.org

ABSTRACT

Multi-microphone noise reduction methods often operate in the time-frequency domain in which a complex gain is applied to each time-frame and subband. These methods can achieve good noise reduction with little speech distortion by exploiting the fact that the desired signal is correlated across the channels. In the context of single-microphone noise reduction, it has been shown recently that the performance in terms of noise reduction and speech distortion can be improved by exploiting the correlation between subsequent time-frames, i.e., by exploiting the interframe correlation. In this paper, we exploit both interchannel and interframe correlations in the context of multi-microphone noise reduction. Now the interframe correlation is taken into account, i.e., a filter is applied in each subband and channel instead of just a gain. The results of our experimental study show that we can improve the fullband signal-to-noise ratios (SNRs) by using interchannel and interframe correlations when dealing with signals, such as speech, that exhibit a sufficiently large interframe correlation.

Index Terms— Noise reduction, microphone array processing, interframe correlation, interchannel correlation, minimum variance distortionless response (MVDR) beamformer.

1. INTRODUCTION

Distant or hands-free speech capture is required in many applications such as hearing aids and teleconferencing. Microphone arrays are often used for the acquisition and consist of sets of microphones that are arranged in specific topologies. The received microphone signals usually consist of a mixture of signals of the desired source and ambient noise. As the ambient noise degrades the quality and intelligibility of the desired source, the received signals are processed in order to extract the desired source signal.

In the last four decades numerous spatio-temporal filters have been proposed to process the received microphone signals (see [1, 2] and the references therein). Multi-microphone noise reduction methods often operate in the time-frequency domain in which a complex gain is applied to each time-frame and subband. These methods can achieve good noise reduction by exploiting the fact that the desired signal is correlated across the channels, i.e., by exploiting the interchannel correlation. In the context of single-microphone noise reduction, the correlation between subsequent time-frames, i.e., interframe correlation, has been successfully taken into account [3, Chap. 17], [4]. A major advantage of such approach is that it enables the development of low-latency algorithms by using short analysis frames while exploiting the temporal characteristics of the desired signal. Recently, Benesty and Huang [5] derived a single-channel MVDR filter with which noise reduction can be achieved without distorting the desired speech signal by exploiting the interframe correlation. In the context of multiple microphone noise reduction the interframe correlation was used in [6] to improve the spatial prediction in highly reverberant environments. In this work it was assumed that the spectral coefficients of the desired signal are mutually uncorrelated across time and frequency, and that the interframe correlation was strictly caused by the acoustic channel. Moreover, a filter was applied in each subband and channel instead of just a gain.

In this paper we propose to exploit both interchannel and interframe correlations in the context of multi-microphone noise reduction. Now the interframe correlation caused by the desired signal is taken into account. The results of our experimental study show that we can improve the fullband SNRs by using interchannel and interframe correlations when dealing with signals, such as speech, that exhibit a sufficiently large interframe correlation.

This paper is organized as follows. Section 2 provides a new signal model, linear array model, definitions, and fundamental assumptions made in this paper. In Section 3, the performance measures are defined. In Section 4, an MVDR beamformer is deduced using the proposed signal model. Section 5 investigates the performance of the MVDR beamformer that exploits both interchannel and interframe correlations. Finally, Section 6 provides some concluding remarks.

2. PROPOSED SIGNAL AND ARRAY MODELS

We consider the well-accepted room acoustics signal model in which an N-element microphone array captures a convolved source signal in some noise field. In the short-time Fourier transform (STFT) domain we can express the spectral coefficients of the received signals at time-frame m and discrete-frequency k as[1]

$$\begin{aligned} Y_n(k,m) &= G_n(k)S(k,m) + V_n(k,m) \\ &= X_n(k,m) + V_n(k,m),\ n = 1,2,\ldots,N, \end{aligned} \tag{1}$$

where $G_n(k)$ is the transfer function from the unknown speech source $S(k,m)$ to the nth microphone that is assumed to be time-invariant, and $V_n(k,m)$ is the additive noise at microphone n. We assume that the spectral coefficients $X_n(k,m) = G_n(k)S(k,m)$ and $V_n(k,m)$ are uncorrelated and zero-mean complex random variables. By definition, $X_n(k,m)$ is coherent across the array.

[†] A joint institution of the University of Erlangen-Nuremberg and Fraunhofer IIS, Germany.

[1] In this work, we assume that the analysis window is sufficiently long such that the multiplicative transfer function approximation [7] holds.

It is more convenient to write the N microphone signals in a vector notation:

$$\begin{aligned}\mathbf{y}(k,m) &= \mathbf{g}(k)S(k,m)+\mathbf{v}(k,m)\\ &= \mathbf{x}(k,m)+\mathbf{v}(k,m)\\ &= \mathbf{d}(k)X_1(k,m)+\mathbf{v}(k,m),\end{aligned} \quad (2)$$

where

$$\begin{aligned}\mathbf{y}(k,m) &= \left[Y_1(k,m)\ Y_2(k,m)\ \cdots\ Y_N(k,m)\right]^T,\\ \mathbf{x}(k,m) &= \left[X_1(k,m)\ X_2(k,m)\ \cdots\ X_N(k,m)\right]^T,\\ &= S(k,m)\left[G_1(k)\ G_2(k)\ \cdots\ G_N(k)\right]^T\\ &= S(k,m)\mathbf{g}(k),\\ \mathbf{v}(k,m) &= \left[V_1(k,m)\ V_2(k,m)\ \cdots\ V_N(k,m)\right]^T,\\ \mathbf{d}(k) &= \left[1\ \frac{G_2(k)}{G_1(k)}\ \cdots\ \frac{G_N(k)}{G_1(k)}\right]^T\\ &= \frac{\mathbf{g}(k)}{G_1(k)},\end{aligned}$$

and superscript ^{T} denotes transpose of a vector or a matrix. The vector $\mathbf{d}(k)$ is termed the steering vector since it determines the direction of the desired signal $X_1(k,m)$. This definition is a generalization of the classical steering vector to a reverberant (multipath) environment.

We can express the output of the beamformer as

$$\begin{aligned}Z(k,m) &= \sum_{l=0}^{L-1}\mathbf{h}_l^H(k,m)\mathbf{y}(k,m-l)\\ &= \underline{\mathbf{h}}^H(k,m)\underline{\mathbf{y}}(k,m),\end{aligned} \quad (3)$$

where L is the number of consecutive time-frames used for each one of the frequency-bins[2], the superscript ^{H} is the transpose-conjugate operator, $\mathbf{h}_l(k,m) = \left[H_{l,1}(k,m)\ H_{l,2}(k,m)\ \cdots\ H_{l,N}(k,m)\right]^T$, $l=0,1,\ldots,L-1$ are FIR filters of length N, and

$$\begin{aligned}\underline{\mathbf{h}}(k,m) &= \left[\mathbf{h}_0^T(k,m)\,\mathbf{h}_1^T(k,m)\ \cdots\ \mathbf{h}_{L-1}^T(k,m)\right]^T\\ \underline{\mathbf{y}}(k,m) &= \left[\mathbf{y}^T(k,m)\,\mathbf{y}^T(k,m-1)\ \cdots\ \mathbf{y}^T(k,m-L+1)\right]^T\end{aligned}$$

are vectors of length NL. The case $L=1$ corresponds to the conventional STFT-domain linear beamforming [1], the case $N=L=1$ corresponds to the classical single-channel noise reduction in the STFT domain with a gain [8], and the case $N=1,\ L>1$ is also the single-channel noise reduction in the STFT domain but with a filter where the interframe correlation is taken into account [5].

Let us now decompose the signal $Z(k,m)$ into the following form:

$$\begin{aligned}Z(k,m) &= \underline{\mathbf{h}}^H(k,m)\underline{\mathbf{x}}(k,m)+\underline{\mathbf{h}}^H(k,m)\underline{\mathbf{v}}(k,m)\\ &= X_{1,\mathrm{f}}(k,m)+V_{\mathrm{rn}}(k,m),\end{aligned} \quad (4)$$

where

$$\underline{\mathbf{x}}(k,m) = \mathbf{x}_1(k,m)\otimes\mathbf{d}(k), \quad (5)$$

$$\mathbf{x}_1(k,m) = \left[X_1(k,m)\,X_1(k,m-1)\ \cdots\ X_1(k,m-L+1)\right]^T,$$

$\otimes$ is the Kronecker product, $\underline{\mathbf{v}}(k,m)$ is defined in a similar way to $\underline{\mathbf{y}}(k,m)$,

$$X_{1,\mathrm{f}}(k,m) = \underline{\mathbf{h}}^H(k,m)\underline{\mathbf{x}}(k,m) \quad (6)$$

is a filtered version of the desired signal at L successive time-frames, and

$$V_{\mathrm{rn}}(k,m) = \underline{\mathbf{h}}^H(k,m)\underline{\mathbf{v}}(k,m) \quad (7)$$

is the residual noise which is uncorrelated with $X_{1,\mathrm{f}}(k,m)$.

At time-frame m, our desired signal is $X_1(k,m)$ [and not the whole vector $\underline{\mathbf{x}}(k,m)$ or $\mathbf{x}_1(k,m)$]. However, the vector $\underline{\mathbf{x}}(k,m)$ in $X_{1,\mathrm{f}}(k,m)$ contains both the desired signal, $X_1(k,m)$, and the components $X_1(k,m-l),\ l\neq 0$, which are not the desired signals at time-frame m but signals that are correlated with $X_1(k,m)$. Therefore, the elements $X_1(k,m-l),\ l\neq 0$, contain both a part of the desired signal and a component that we consider as an interference. This suggests that we should decompose $X_1(k,m-l)$ into two orthogonal components corresponding to the part of the desired signal and interference, i.e.,

$$X_1(k,m-l) = \rho_{X_1}^*(k,m,l)X_1(k,m)+X_{1,\mathrm{i}}(k,m-l), \quad (8)$$

where

$$X_{1,\mathrm{i}}(k,m-l) = X_1(k,m-l)-\rho_{X_1}^*(k,m,l)X_1(k,m), \quad (9)$$

$$E\left[X_1(k,m)X_{1,\mathrm{i}}^*(k,m-l)\right] = 0, \quad (10)$$

and

$$\rho_{X_1}(k,m,l) = \frac{E\left[X_1(k,m)X_1^*(k,m-l)\right]}{E\left[|X_1(k,m)|^2\right]} \quad (11)$$

is the interframe correlation coefficient of the signal $X_1(k,m)$. Hence, we can write the vector $\underline{\mathbf{x}}(k,m)$ as

$$\begin{aligned}\underline{\mathbf{x}}(k,m) &= X_1(k,m)\left[\boldsymbol{\rho}_{X_1}^*(k,m)\otimes\mathbf{d}(k)\right]+\underline{\mathbf{x}}_{\mathrm{i}}(k,m)\\ &= X_1(k,m)\underline{\mathbf{d}}(k,m)+\underline{\mathbf{x}}_{\mathrm{i}}(k,m)\\ &= \underline{\mathbf{x}}_{\mathrm{d}}(k,m)+\underline{\mathbf{x}}_{\mathrm{i}}(k,m),\end{aligned} \quad (12)$$

where

$$\boldsymbol{\rho}_{X_1}(k,m) = \left[1\,\rho_{X_1}(k,m,1)\ \cdots\ \rho_{X_1}(k,m,L-1)\right]^T \quad (13)$$

is the (normalized) interframe correlation vector between $X_1(k,m)$ and $\mathbf{x}_1(k,m)$,

$$\begin{aligned}\underline{\mathbf{x}}_{\mathrm{i}}(k,m) = [&X_{1,\mathrm{i}}(k,m)\,X_{1,\mathrm{i}}(k,m-1)\ \cdots\\ &X_{1,\mathrm{i}}(k,m-L+1)]^T\otimes\mathbf{d}(k)\end{aligned}$$

is the interference signal vector of length NL,

$$\underline{\mathbf{d}}(k,m) = \boldsymbol{\rho}_{X_1}^*(k,m)\otimes\mathbf{d}(k) \quad (14)$$

is a vector of length NL, and

$$\underline{\mathbf{x}}_{\mathrm{d}}(k,m) = X_1(k,m)\underline{\mathbf{d}}(k,m) \quad (15)$$

is the desired signal vector.

Using (12), we can rewrite (4) as

$$Z(k,m) = X_{\mathrm{fd}}(k,m)+X_{\mathrm{ri}}(k,m)+V_{\mathrm{rn}}(k,m), \quad (16)$$

where

$$X_{\mathrm{fd}}(k,m) = X_1(k,m)\underline{\mathbf{h}}^H(k,m)\underline{\mathbf{d}}(k,m) \quad (17)$$

[2] We can use different numbers of consecutive time-frames for different frequencies but to simplify the presentation, we use the same number L.

is the filtered desired signal and

$$X_{\rm ri}(k,m) = \underline{\mathbf{h}}^H(k,m)\underline{\mathbf{x}}_{\rm i}(k,m) \tag{18}$$

is the residual interference. Note that the above decomposition of the signal $X_1(k,m-l)$ is critical in order to properly design optimal multichannel noise reduction filters with the interframe correlation scheme.

The three terms on the right-hand side of (16) are mutually uncorrelated. Therefore, the variance of $Z(k,m)$ is

$$\phi_Z(k,m) = \phi_{X_{\rm fd}}(k,m) + \phi_{X_{\rm ri}}(k,m) + \phi_{V_{\rm rn}}(k,m), \tag{19}$$

where

$$\begin{aligned} \phi_{X_{\rm fd}}(k,m) &= E\left[|X_{\rm fd}(k,m)|^2\right] & (20)\\ &= \phi_{X_1}(k,m)\left|\underline{\mathbf{h}}^H(k,m)\underline{\mathbf{d}}(k,m)\right|^2, \\ \phi_{X_{\rm ri}}(k,m) &= E\left[|X_{\rm ri}(k,m)|^2\right] & (21)\\ &= \underline{\mathbf{h}}^H(k,m)\boldsymbol{\Phi}_{\underline{\mathbf{x}}_{\rm i}}(k,m)\underline{\mathbf{h}}(k,m) \\ &= \underline{\mathbf{h}}^H(k,m)\boldsymbol{\Phi}_{\underline{\mathbf{x}}}(k,m)\underline{\mathbf{h}}(k,m) \\ &\quad - \phi_{X_1}(k,m)\left|\underline{\mathbf{h}}^H(k,m)\underline{\mathbf{d}}(k,m)\right|^2, \\ \phi_{V_{\rm rn}}(k,m) &= E\left[|V_{\rm rn}(k,m)|^2\right] & (22)\\ &= \underline{\mathbf{h}}^H(k,m)\boldsymbol{\Phi}_{\underline{\mathbf{v}}}(k,m)\underline{\mathbf{h}}(k,m), \end{aligned}$$

$\phi_{X_1}(k,m) = E\left[|X_1(k,m)|^2\right]$ is the variance of $X_1(k,m)$, and $\boldsymbol{\Phi}_{\underline{\mathbf{x}}_{\rm d}}(k,m) = \phi_{X_1}(k,m)\underline{\mathbf{d}}(k,m)\underline{\mathbf{d}}^H(k,m)$ is the correlation matrix of $\underline{\mathbf{x}}_{\rm d}(k,m)$, with $\boldsymbol{\Phi}_{\underline{\mathbf{x}}}(k,m)$, $\boldsymbol{\Phi}_{\underline{\mathbf{x}}_{\rm i}}(k,m)$, and $\boldsymbol{\Phi}_{\underline{\mathbf{v}}}(k,m)$ being the correlation matrices of $\underline{\mathbf{x}}(k,m)$, $\underline{\mathbf{x}}_{\rm i}(k,m)$, and $\underline{\mathbf{v}}(k,m)$, respectively.

3. PERFORMANCE MEASURES

We will now define the most important performance measures with respect to the reference microphone (i.e., microphone 1) in the context of interframe and interchannel correlations.

The subband and fullband input SNRs at time-frame m are defined as

$$\text{iSNR}(k,m) = \frac{\phi_{X_1}(k,m)}{\phi_{V_1}(k,m)}, \tag{23}$$

$$\text{iSNR}(m) = \frac{\sum_{k=0}^{K-1}\phi_{X_1}(k,m)}{\sum_{k=0}^{K-1}\phi_{V_1}(k,m)}, \tag{24}$$

where $\phi_{V_1}(k,m) = E\left[|V_1(k,m)|^2\right]$ is the variance of $V_1(k,m)$.

We define the subband output SNR as

$$\begin{aligned} \text{oSNR}\left[\underline{\mathbf{h}}(k,m)\right] &= \frac{\phi_{X_{\rm fd}}(k,m)}{\phi_{X_{\rm ri}}(k,m) + \phi_{V_{\rm rn}}(k,m)} \\ &= \frac{\phi_{X_1}(k,m)\left|\underline{\mathbf{h}}^H(k,m)\underline{\mathbf{d}}(k,m)\right|^2}{\underline{\mathbf{h}}^H(k,m)\boldsymbol{\Phi}_{\rm in}(k,m)\underline{\mathbf{h}}(k,m)}, \end{aligned} \tag{25}$$

where

$$\boldsymbol{\Phi}_{\rm in}(k,m) = \boldsymbol{\Phi}_{\underline{\mathbf{x}}_{\rm i}}(k,m) + \boldsymbol{\Phi}_{\underline{\mathbf{v}}}(k,m) \tag{26}$$

is the interference-plus-noise correlation matrix. For the particular filter $\underline{\mathbf{h}}(k,m) = \mathbf{i}_{NL,1}$, where $\mathbf{i}_{NL,1}$ is the first column of the identity matrix $\mathbf{I}_{NL}$ (of size $NL \times NL$), we have

$$\text{oSNR}\left[\mathbf{i}_{NL,1}(k,m)\right] = \text{iSNR}(k,m). \tag{27}$$

With the identity filter, $\mathbf{i}_{NL,1}$, the SNR cannot be improved.

Using the subband input and output SNRs we define the subband array gain as

$$\mathcal{A}\left[\underline{\mathbf{h}}(k,m)\right] = \frac{\text{oSNR}\left[\underline{\mathbf{h}}(k,m)\right]}{\text{iSNR}(k,m)}, \tag{28}$$

We define the fullband output SNR at time-frame m as

$$\text{oSNR}\left[\underline{\mathbf{h}}(:,m)\right] = \frac{\sum_{k=0}^{K-1}\phi_{X_1}(k,m)\left|\underline{\mathbf{h}}^H(k,m)\underline{\mathbf{d}}(k,m)\right|^2}{\sum_{k=0}^{K-1}\underline{\mathbf{h}}^H(k,m)\boldsymbol{\Phi}_{\rm in}(k,m)\underline{\mathbf{h}}(k,m)}. \tag{29}$$

We also define the fullband array gain as

$$\mathcal{A}\left[\underline{\mathbf{h}}(:,m)\right] = \frac{\text{oSNR}\left[\underline{\mathbf{h}}(:,m)\right]}{\text{iSNR}(m)}. \tag{30}$$

We end this subsection by giving the subband and fullband speech distortion indices:

$$\begin{aligned} \upsilon_{\rm sd}\left[\underline{\mathbf{h}}(k,m)\right] &= \frac{E\left\{|X_{\rm fd}(k,m) - X_1(k,m)|^2\right\}}{\phi_{X_1}(k,m)} \\ &= \left|\underline{\mathbf{h}}^H(k,m)\underline{\mathbf{d}}(k,m) - 1\right|^2, \end{aligned} \tag{31}$$

$$\upsilon_{\rm sd}\left[\underline{\mathbf{h}}(:,m)\right] = \frac{\sum_{k=0}^{K-1}E\left\{|X_{\rm fd}(k,m) - X_1(k,m)|^2\right\}}{\sum_{k=0}^{K-1}\phi_{X_1}(k,m)}. \tag{32}$$

4. MVDR FILTER

By minimizing the subband MSE of the residual interference-plus-noise with the constraint that the desired signal is not distorted, we easily find the MVDR filter:

$$\underline{\mathbf{h}}_{\rm MVDR}(k,m) = \frac{\boldsymbol{\Phi}_{\rm in}^{-1}(k,m)\boldsymbol{\Phi}_{\underline{\mathbf{y}}}(k,m) - \mathbf{I}_{NL}}{\text{tr}\left[\boldsymbol{\Phi}_{\rm in}^{-1}(k,m)\boldsymbol{\Phi}_{\underline{\mathbf{y}}}(k,m)\right] - NL}\mathbf{i}_{NL,1}, \tag{33}$$

that we can rewrite as

$$\underline{\mathbf{h}}_{\rm MVDR}(k,m) = \frac{\boldsymbol{\Phi}_{\underline{\mathbf{y}}}^{-1}(k,m)\underline{\mathbf{d}}(k,m)}{\underline{\mathbf{d}}^H(k,m)\boldsymbol{\Phi}_{\underline{\mathbf{y}}}^{-1}(k,m)\underline{\mathbf{d}}(k,m)}, \tag{34}$$

where $\boldsymbol{\Phi}_{\underline{\mathbf{y}}}(k,m)$ is the correlation matrix of $\mathbf{y}(k,m)$. It should be noted that the filter in (34) is different from the filter proposed in [6].

It is clear that we always have

$$\upsilon_{\rm sd}\left[\underline{\mathbf{h}}_{\rm MVDR}(k,m)\right] = 0,\ \upsilon_{\rm sd}\left[\underline{\mathbf{h}}_{\rm MVDR}(:,m)\right] = 0. \tag{35}$$

The MVDR beamformer rejects the maximum level of noise allowable without distorting the desired signal.

5. PERFORMANCE EVALUATION

The processing was done at 8 kHz in the STFT domain with a window length of 256 points (32 ms), discrete Fourier transform (DFT) size of 512 points, and 50% overlap. In practice the covariance matrix $\boldsymbol{\Phi}_{\underline{\mathbf{v}}}(k,m)$ can be estimated for each time-frequency bin when the desired source is inactive. In this study we have used the signal vector $\underline{\mathbf{v}}(k,m)$ directly in order to put aside the influence of a voice activity detector. The covariance matrices $\boldsymbol{\Phi}_{\underline{\mathbf{y}}}(k,m)$ and $\boldsymbol{\Phi}_{\underline{\mathbf{v}}}(k,m)$ were estimated recursively with a weighting factor λ_y using $\underline{\mathbf{y}}(k,m)$ and $\underline{\mathbf{v}}(k,m)$. Using these covariance matrices, the steering vector $\underline{\mathbf{d}}(k)$ was estimated based on the procedure described in [5].

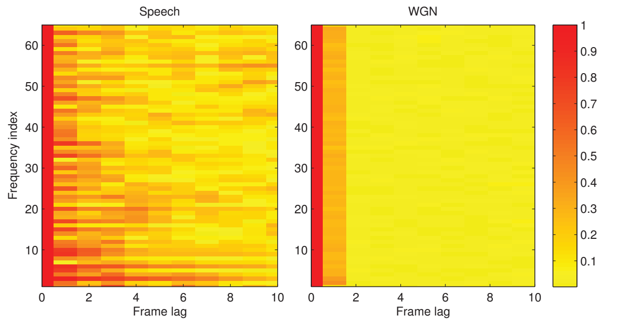

Fig. 1. Magnitude of the correlation coefficient per subband of a 1 second segment speech signal and WGN signal.

First, we investigated the interframe correlation of a 1 second anechoic speech signal and zero-mean white Gaussian noise (WGN) signal. In Fig. 1, the magnitude of the correlation coefficients for different frame lags and subbands are shown. Both signals exhibit some interframe correlation due to the overlapping time-frames (in this case 50%) used to compute the STFT. Especially at low frequencies, we see that the speech signal exhibits significantly larger interframe correlation compared to the WGN signal.

Secondly, we evaluated the performance of the MVDR filter in a simulated reverberant environment of size $5 \times 6 \times 3$ m (length $\times$ width $\times$ height) and a reverberation time of 200 ms. All room impulse responses were computed using the source-image method. The microphone arrays consisted of $N = 3$ microphones with an inter-microphone distance of 5 cm. The desired source was placed at a distance of 1 m. An undesired source was placed at a distance of 1.5 m with an input signal to coherent interference ratio of 20 dB. Finally, spatially WGN noise was added to the microphone signals such that the signal to incoherent noise ratio was equal to 35 dB.

In Fig. 2, the average fullband array gain and speech distortion index are shown for $N \in \{1, 3\}$ and $L \in \{1, 2, 4\}$ as a function of the weighting factor λ_y. The results for $N = 1$ and $L = 1$ are omitted as for these settings the MVDR filter does not provide any noise reduction nor speech distortion. We observe that the array gain can be increased by exploiting the interframe correlation. Using a large λ_y (close to 1), we cannot capture the short-term variation of quasi-stationary speech signals. But with a small λ_y, the covariance matrix $\mathbf{\Phi}_{\underline{\mathbf{y}}}(k, m)$ may become ill-conditioned or singular for some time frames and frequency bins. An informal listening indicated that a moderate value of λ_y produces the best speech quality.

In Fig. 3, the short-term fullband array gain and speech distortion index are shown for 100 time-frames during which speech was present ($N = 3$, $L \in \{1, 2, 4\}$ and $\lambda_y = 0.85$). We observe that the short-term array gain can be increased by exploiting the interframe correlation. In addition, we observe that the speech distortion index increases when L increases. This is most likely caused by estimation errors of the interframe correlations which, compared to the interchannel correlations, are more difficult to estimate accurately. Nevertheless, we observe a significant improvement in the array gain when L is increased.

6. CONCLUSIONS

In this contribution, we exploited interchannel and interframe correlations in the context of multi-microphone noise reduction. We proposed a signal model that takes the successive time-frames into account and can be used to derive optimal noise reduction filters. As an example, we derived an MVDR filter that exploits the interchannel and interframe correlations. The results of our experimental study showed that using this filter, we can increase the array gain while keeping the level of speech distortion low.

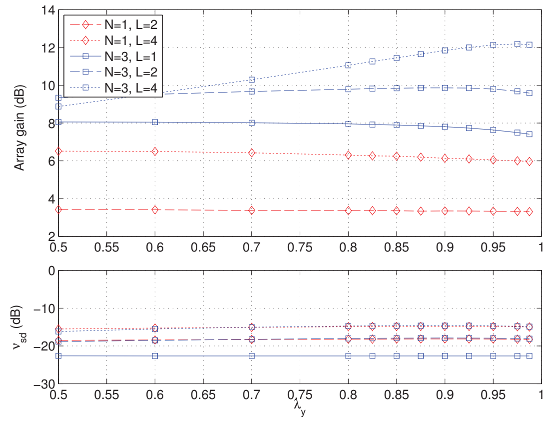

Fig. 2. Average fullband array gain $E\{\mathcal{A}[\underline{\mathbf{h}}(:, m)]\}$ and speech distortion index $E\{\nu_{\text{sd}}[\underline{\mathbf{h}}(:, m)]\}$ as a function of the weighting factor λ_y.

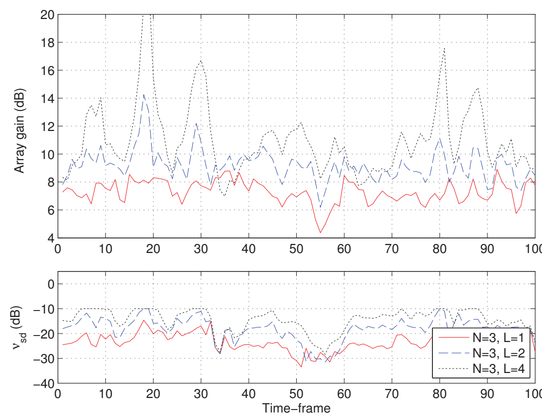

Fig. 3. Fullband array gain and speech distortion index in dB.

7. REFERENCES

[1] J. Benesty, J. Chen, and Y. Huang, *Microphone Array Signal Processing*. Berlin, Germany: Springer-Verlag, 2008.

[2] S. Gannot and I. Cohen, "Adaptive beamforming and postfiltering," in *Springer Handbook of Speech Processing*, J. Benesty, M. M. Sondhi, and Y. Huang, Eds. Springer-Verlag, 2008, ch. 47.

[3] S. V. Vaseghi, *Advanced digital signal processing and noise reduction*, 4th ed. John Wiley & Sons Ltd., 2008.

[4] E. Plourde and B. Champagne, "Multidimensional STSA estimators for speech enhancement with correlated spectral components," *IEEE Trans. Signal Process.*, vol. 59, no. 7, pp. 3013–3024, Jul. 2011.

[5] J. Benesty and Y. Huang, "A single-channel noise reduction MVDR filter," in *Proc. IEEE Intl. Conf. on Acoustics, Speech and Signal Processing (ICASSP)*, 2011, pp. 273–276.

[6] E. Habets, "A distortionless subband beamformer for noise reduction in reverberant environments," in *Proc. Intl. Workshop Acoust. Echo Noise Control (IWAENC)*, Tel-Aviv, Israel, Aug. 2010, pp. 1–4.

[7] Y. Avargel and I. Cohen, "On multiplicative transfer function approximation in the short-time Fourier transform domain," *IEEE Signal Process. Lett.*, vol. 14, no. 5, pp. 337–340, May 2007.

[8] J. Benesty, J. Chen, Y. Huang, and I. Cohen, *Noise Reduction in Speech Processing*. Berlin, Germany: Springer-Verlag, 2009.