ONLINE INTER-FRAME CORRELATION ESTIMATION METHODS FOR SPEECH ENHANCEMENT IN FREQUENCY SUBBANDS

Alexander Schasse and Rainer Martin

Institute of Communication Acoustics, Ruhr-Universität Bochum, 44780 Bochum, Germany
Email: {alexander.schasse,rainer.martin}@rub.de

ABSTRACT

In this paper, we propose solutions for the online adaptation of optimal FIR filters for speech enhancement in DFT subbands. An important ingredient to such filters is the estimation of the inter-frame correlation of the clean speech signal. While this correlation was assumed to be perfectly known in former studies, we discuss two online estimation approaches based on a constant noise inter-frame correlation and on the use of a binary mask. We show that a filtering of subband signals based on these estimated quantities outperforms a conventional, instantaneous spectral weighting, such as the frequency-domain Wiener filter at least for high SNR conditions.

Index Terms— Noise reduction, MVDR, Wiener filter, filter-bank system, subband filtering

1. INTRODUCTION

Single-channel noise-reduction (NR) algorithms often process the noisy speech in the time-frequency domain. In many cases, they are based on the discrete short-time Fourier transform (DFT) and an appropriate overlap/add or overlap/save method for resynthesis. NR is then achieved by using instantaneous and real-valued spectral weights, e.g., depending on the *a-priori* SNR [1] in case of a Wiener filter [2, 3]. However, as stated in recent works by Benesty and Huang [4, 5], these approaches do not explicitly exploit the correlation inherent in the signals or caused by overlapping analysis frames. Therefore, it seems promising to interpret the subband signals of a filter-bank system (e.g. based on a DFT) as time domain signals and apply filters instead of instantaneous weights. The main difference of filtering the subband signals instead of using instantaneous real-valued weights lies in the estimation of not only the magnitude of the clean speech DFT coefficients, but also their phase. Furthermore, a longer memory is used in the process. In this work, we consider the filtering of subband signals of a DFT filter-bank system based on the algorithms presented in [4, 5], where the authors develop an MVDR and (time domain) Wiener filter based on the inter-frame correlation (IFC) of the clean speech signal, which was assumed to be perfectly known in [4, 5]. The focus of our contribution lies on this quantity and its blind estimation. We will present two algorithms to estimate the noise IFC, which directly leads to the required speech correlation, and discuss their characteristics. In the first approach we assume the noise IFC to be a fixed quantity. The other one uses a binary mask as an extended voice activity detector (VAD) to update the noise IFC during speech pauses.

The paper is structured as follows. In the next section we describe the signal model and define the noise reduction filters based on temporal filtering of the subband signals. In the third Section, we analyze the IFC of speech and noise and develop estimation algorithms. In Section 4, we evaluate these algorithms and we conclude our work in Section 5.

2. SUBBAND FILTERS

2.1. Signal Model

Throughout this paper we consider an additive noise model

$$y(n) = x(n) + v(n), \tag{1}$$

where $y(n)$ is the noisy speech, $x(n)$ the clean speech and $v(n)$ the noise signal, while n is the discrete time index. Furthermore, we assume statistical independence of the noise and the clean speech signal. In the short-time frequency domain, these signals are represented by

$$Y(k,m) = X(k,m) + V(k,m). \tag{2}$$

Here, k is the frequency bin index and m is the frame index. In each frequency subband we define FIR filters of order L-1 as

$$\begin{aligned}\hat{X}(k,m) &= \sum_{l=0}^{L-1} h^*(k,m,l)Y(k,m-l) &(3)\\ &= \mathbf{h}^H(k,m)\mathbf{y}(k,m). &(4)\end{aligned}$$

The vector $\mathbf{h}(k,m)$ contains the time-varying filter coefficients $h(k,m,l)$ in each subband k and the vector $\mathbf{y}(k,m) = [Y(k,m), Y(k,m-1), \ldots, Y(k,m-L+1)]^T$ of length L contains the history of noisy subband samples. The superscript ^{H} represents the Hermitian transpose operator. A conventional instantaneous spectral weighting could be realized with $L = 1$.

2.2. Inter-Frame Correlation and Filter Design

In this work we use both the MVDR and the Wiener filter [4, 5] in subbands, both of which depend on the inter-frame correlation (IFC). In case of the clean speech signal the IFC is given by

$$\boldsymbol{\gamma}_X(k,m) = \frac{\mathrm{E}\{\mathbf{x}^*(k,m)X(k,m)\}}{\mathrm{E}\{|X(k,m)|^2\}}. \tag{5}$$

Here, $\mathbf{x}(k,m) = [X(k,m), X(k,m-1), \ldots, X(k,m-L+1)]^T$ contains a history of the clean speech subband signal $X(k,m)$. Based on this definition, the two filter functions can be formulated. First, the MVDR filter defined in [4, 5] reduces the output noise power in each subband under the constraint $\mathbf{h}^H(k,m)\mathbf{x}(k,m) = 1$ and is given as

$$\mathbf{h}_{\text{MVDR}}(k,m) = \frac{\mathbf{\Phi}_Y^{-1}(k,m)\boldsymbol{\gamma}_X^*(k,m)}{\boldsymbol{\gamma}_X^T(k,m)\mathbf{\Phi}_Y^{-1}(k,m)\boldsymbol{\gamma}_X^*(k,m)}. \tag{6}$$

Here, $\mathbf{\Phi}_Y(k,m)$ is the covariance matrix of the k-th noisy subband signal and the superscript ^{T} represents the transpose operator. If all

978-1-4799-0356-6/13/$31.00 ©2013 IEEE

quantities are known, this filter provides, at least in theory, noise reduction without target signal distortion. Secondly, the subband time-domain Wiener filter which minimizes the mean-square error is defined based on the speech power $\phi_X(k,m) = \mathrm{E}\left\{|X(k,m)|^2\right\}$ as

$$\mathbf{h}_{\text{Wiener}}(k,m) = \phi_X(k,m)\mathbf{\Phi}_Y^{-1}(k,m)\boldsymbol{\gamma}_X^*(k,m). \tag{7}$$

When comparing the filter definitions in (6) and (7), the filters can be summarized as

$$\mathbf{h}_{\text{MVDR}}(k,m) = c_{\text{MVDR}}(k,m)\mathbf{\Phi}_Y^{-1}(k,m)\boldsymbol{\gamma}_X^*(k,m) \tag{8}$$

$$\mathbf{h}_{\text{Wiener}}(k,m) = c_{\text{Wiener}}(k,m)\mathbf{\Phi}_Y^{-1}(k,m)\boldsymbol{\gamma}_X^*(k,m) \tag{9}$$

with

$$c_{\text{MVDR}}(k,m) = \frac{1}{\boldsymbol{\gamma}_X^T(k,m)\mathbf{\Phi}_Y^{-1}(k,m)\boldsymbol{\gamma}_X^*(k,m)} \tag{10}$$

$$c_{\text{Wiener}}(k,m) = \phi_X(k,m). \tag{11}$$

In case of the Wiener filter, the filter coefficients are set to zero during speech pauses due to the multiplicative effect of the speech power $\phi_X(k,m)$. In case of the MVDR filter, the expression in (10) represents a power normalization. However, in terms of implementation issues, the MVDR filter is more robust, since the inverse of the covariance matrix also appears in the denominator. Therefore, numerical errors (due to the estimation process or the calculation of the inverse) are better balanced.

3. INTER-FRAME CORRELATION ESTIMATION

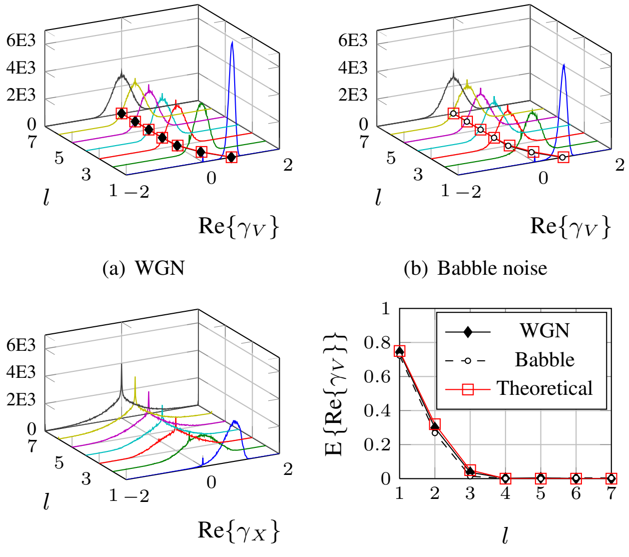
The IFC $\boldsymbol{\gamma}_Y(k,m)$ of the noisy subband signal $Y(k,m)$ can be easily estimated by recursive smoothing. If the IFC of the noise signal is known, $\boldsymbol{\gamma}_X(k,m)$ can then be calculated as [4, 5]

$$\boldsymbol{\gamma}_X(k,m) = \frac{\phi_Y(k,m)}{\phi_Y(k,m)-\phi_V(k,m)}\boldsymbol{\gamma}_Y(k,m) - \frac{\phi_V(k,m)}{\phi_Y(k,m)-\phi_V(k,m)}\boldsymbol{\gamma}_V(k,m). \tag{12}$$

Here, $\phi_Y(k,m)$ and $\phi_V(k,m)$ are the power of the noisy signal and the noise signal, respectively. The noise power can be estimated by well known techniques such as the Minimum Statistics [6] or the MMSE-based approach presented in [7]. Note that (12) corresponds to a Maximum Likelihood (ML) estimate of $\boldsymbol{\gamma}_X(k,m)$ in case of jointly Gaussian distributed signals.

3.1. VAD-Based Estimation

Based on a voice activity detector (VAD), the IFC $\boldsymbol{\gamma}_V(k,m)$ of the noise signal can be updated during speech pauses. However, this approach requires stationarity of the noise signal during speech activity and depends on the performance of the VAD. If the speech signal is corrupted by nonstationary noise, the properties of $\boldsymbol{\gamma}_V(k,m)$ might change during speech activity. Furthermore, if the VAD doesn't work perfectly, it might interpret speech sequences as noise (which leads to speech distortion) or fails to update $\boldsymbol{\gamma}_V(k,m)$. In our work, we follow two different approaches, i.e. (i) we assume the noise IFC to be a constant vector for all frequencies and frames and (ii) extend the idea of VAD to a binary mask approach.



(a) WGN

(b) Babble noise

(c) Speech

(d) Mean IFC comparison

Fig. 1. Histograms of the real part of IFC for two noise types (WGN in (a), babble noise in (b)) and speech (c), estimated by temporal averaging. The respective mean values and the theoretical values based on the frame overlap (red squares) are drawn at the bottoms of the first two plots for both noise types and additionally in (d). The time lag l corresponds to the index of the filter coefficients in (3).

3.2. Constant Noise IFC

If the noise signal is stationary, the expected values needed to calculate its IFC analogue to (5) can be estimated by temporal averages which approach the ML solution for Gaussian processes. Figure 1 shows histograms of the real part of the IFC for white Gaussian noise (WGN), non-stationary babble noise and speech (the respective imaginary parts are similarly distributed). Additionally, the estimated means are plotted for $l = 1, \ldots, L-1$. We observe Gaussian distributions for both noise types while the histogram of the speech IFC shows a super-Gaussian shape and is asymmetric for small time lags l. If we assume that the noise signal is WGN, the IFC $\boldsymbol{\gamma}_V(k,m)$ is solely defined by the frame overlap and the window function $h(n)$ of the analysis filter-bank system. This means, we can calculate an estimate of the l-th component of $\boldsymbol{\gamma}_V(k,m)$ by

$$\hat{\gamma}_{V\{l\}}(k,m) = \frac{\sum_n h(n)h(n+lR)}{\sum_n h^2(n)}, \tag{13}$$

where R represents the frame advance. These theoretical values are also shown in Fig. 1 (a), (b) and (d) as red curves. It is obvious that this estimate works very well for WGN as well as for babble noise and other noise types tested. Table 1 summarizes the small mean square errors (MSE) between the estimator in (13) and the complex-valued IFCs estimated by temporal averaging for different noise signals.

3.3. Estimation Based on a Binary Mask

Following the definition in (5), the IFC of the clean speech signal is not defined during speech pauses since all elements in $\mathbf{x}(k,m)$ are zero. Therefore, it seems reasonable not to detect speech activity globally, i.e. across all subbands, but within each subband. This corresponds to the estimation of a binary mask. The ideal binary

Noise type	MSE
WGN	1.0 E-3
Babble, party	6.6 E-3
Babble, restaurant	6.2 E-3
Traffic	9.0 E-3

Table 1. Mean square error between the theoretical noise IFC based on the frame overlap in (13) and the respective complex-valued IFCs based on time averages.

mask [8] is defined based on the SNR in each point in the time-frequency plane

$$\mathcal{M}(k,m) = \begin{cases} 1 & \text{, if SNR}(k,m) > \delta \\ 0 & \text{, otherwise} \end{cases}. \qquad (14)$$

If all entries of $\mathbf{m}(k,m) = [\mathcal{M}(k,m), \mathcal{M}(k,m-1), \ldots, \mathcal{M}(k,m-L+1)]^T$ in the temporal span of a subband filter are zero, the noise IFC, which is assumed to be defined at all times and frequencies, can be updated based on the input signal. Since we do not apply the binary mask directly to the noisy signal, but only to update the noise IFC, we can set the threshold δ to values that are smaller than 0 dB to increase the robustness and to prevent the inclusion of speech bins in the IFC estimate.

4. EVALUATION

4.1. Evaluation Settings

To evaluate NR based on subband filtering, we chose a uniformly modulated DFT filter-bank system [9] with a frame length and DFT size of K=512 and a down-sampling factor of R=128, i.e. 75% frame overlap, at a sampling rate of 16 kHz. We use square-root-Hann windows for the analysis and synthesis filter-bank system to provide perfect signal reconstruction. The length of the subband filters is set to L=10, and compared to a conventional frequency-domain Wiener filter (i.e. L=1). To estimate the noise power in both cases, we use the approach presented in [10]

$$\phi_Y(k,m) = \alpha_Y \phi_Y(k,m-1) + (1-\alpha_Y)|Y(k,m)|^2 \qquad (15)$$

$$\hat{\phi}_V(k,m) = \min\{\hat{\phi}_V(k,m-1), \phi_Y(k,m)\}(1+\epsilon), \qquad (16)$$

and set $\alpha_Y = 0.85$ and chose ϵ to induce a maximum power increase of 4 dB/s. The spectral gains of the frequency-domain Wiener filter (L=1) are calculated based on the *a priori* SNR as estimated by the *decision-directed* approach [1] with a smoothing parameter of $\alpha_{\text{DD}} = 0.94$. The first noise IFC estimator (Const.) is based on a real-valued and constant vector following (13). The smoothed *a posteriori* SNR [1] is used to calculate the binary mask in case of the second IFC estimator (Bin. Mask) presented in Section 3.3. The threshold δ is set to 0dB. In both cases the speech IFC is then calculated by (12). Furthermore, we use a first order recursion $\mathbf{\Phi}_Y(k,m) = \lambda \mathbf{\Phi}_Y(k,m-1) + \mathbf{y}(k,m)\mathbf{y}^H(k,m)$ and the matrix inversion lemma [11] to derive a robust estimator for the inverse of the covariance matrix of the noisy speech signal

$$\mathbf{k}(k,m) = \frac{\mathbf{\Phi}_Y^{-1}(k,m-1)\mathbf{y}(k,m)}{\lambda + \mathbf{y}^H(k,m)\mathbf{\Phi}_Y^{-1}(k,m-1)\mathbf{y}(k,m)} \qquad (17)$$

$$\mathbf{\Phi}_Y^{-1}(k,m) = \lambda^{-1}\mathbf{\Phi}_Y^{-1}(k,m-1) - \mathbf{k}(k,m)\mathbf{y}^H(k,m)\mathbf{\Phi}_Y^{-1}(k,m-1) \qquad (18)$$

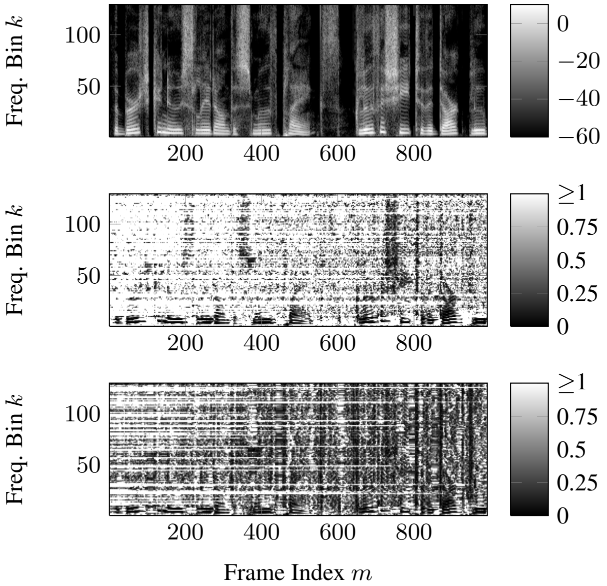


Fig. 2. From top to bottom: Spectrogram of the clean speech signal; MSE of the estimated speech IFC in case of a constant noise IFC; MSE of the estimated speech IFC based on the binary mask approach.

as known, e.g., from the RLS algorithm [12].

As a general observation, we find that the processed signals using the subband filters sound slightly reverberant when compared to the traditional Wiener Filter (L=1). In informal listening tests, this did not lead to a degradation of the perceived signal quality. On the contrary: the speech seems to be more present in these signals. To evaluate the NR performance we calculate SNR improvements (ΔSNR) and the Log-Likelihood Ratio (LLR) [13] as a distance to the clean speech signal. Note that this reverberation effect might bias the objective measures.

4.2. IFC Estimation

To compare the performance of the proposed IFC estimators in terms of their accuracy, we calculate the mean square error (MSE)

$$\text{MSE}(k,m) = \frac{1}{L}\|\boldsymbol{\gamma}_X(k,m) - \hat{\boldsymbol{\gamma}}_X(k,m)\|^2. \qquad (19)$$

Figure 2 shows the MSE at all time-frequency points for both IFC estimators when applied to a noisy signal (WGN, 0dB). To calculate the reference speech IFC $\boldsymbol{\gamma}_X(k,m)$, we estimate the expectations via first order recursive smoothing of the clean speech DFT coefficients. As a reference, the first plot in Fig. 2 shows the spectrogram of the clean speech signal. We observe in both cases that the MSE is small during speech activity, especially at voiced sounds. However, the binary mask based approach achieves much smaller MSEs for unvoiced sounds and during speech pauses.

4.3. NR Performance

Based on the discussion in Section 2.2 and preliminary experimental results we found the MVDR filter to be more robust than the subband Wiener filter. Only if all quantities, especially the speech power $\phi_X(k,m)$, are perfectly known, the Wiener filter eliminates almost all of the noise. Therefore, we focus on the use of the MVDR filter and compare it for both IFC estimation approaches to a conventional frequency-domain Wiener filter (L=1). We designed the

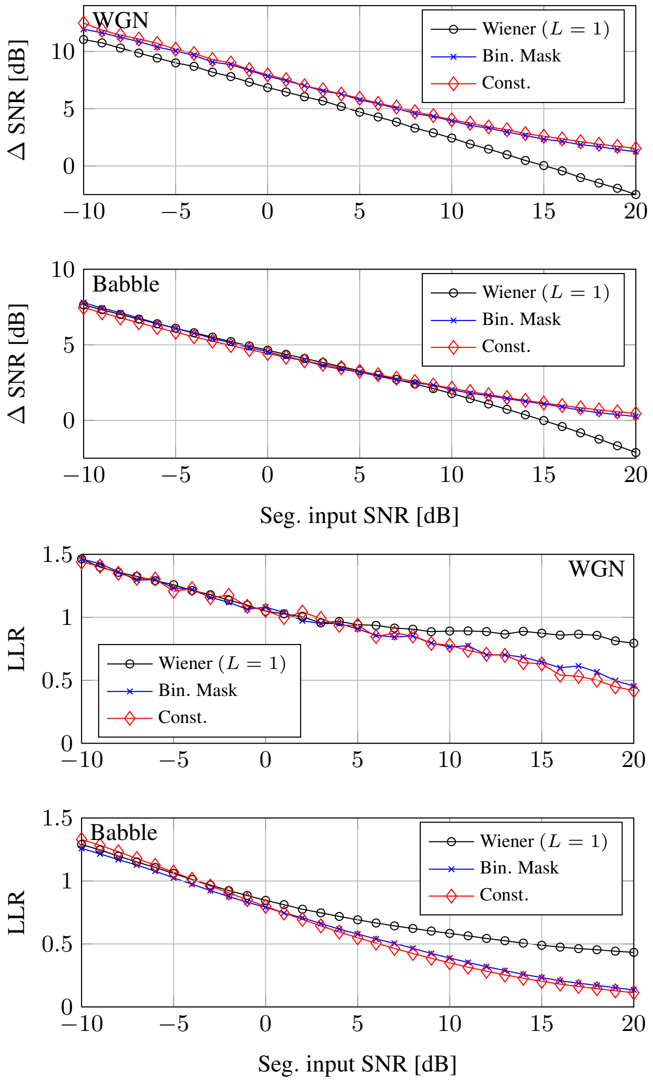


Fig. 3. Simulation results for varying segmental input SNRs and two noise types, white Gaussian noise (WGN) and babble noise. The figures show the SNR improvement and the LLR values for the conventional Wiener filter (L=1) and the subband MVDR filter (L=8) based on the use of the binary mask (Bin. Mask) and of the constant noise IFC (const.) estimators.

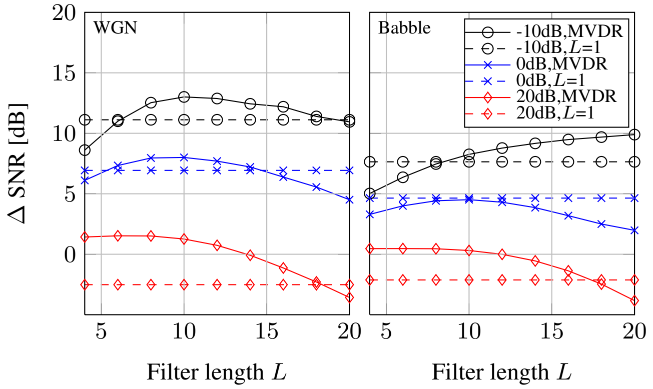


Fig. 4. Influence of the filter length L (MVDR, constant noise IFC) compared to the Wiener filter for WGN (left) and babble noise (right).

NR algorithms such that they achieve the same noise reduction of approximately 20 dB for WGN at an input SNR of 0 dB. Figure 3 shows the resulting SNR gains and LLR values for varying segmental input SNRs and noise types. We see that the MVDR approach (L=10) performs better than the conventional Wiener filter at high input SNRs. Only for WGN, the MVDR filter shows a consistent improvement for all input SNR values. In all cases, both IFC estimators show a similar behavior, while the approach based on the binary mask performs slightly better. This means that the speech is barely effected by the larger MSEs of a constant noise IFC as shown in Fig. 2. The noise reduction itself is mainly achieved by the estimate of the noise power. However, both filtering approaches, i.e. the MVDR filter (L=10) and the Wiener filter (L=1), suffer from noise power estimation errors. These errors lead to residual noise artifacts for babble noise and a decreasing of the signal quality for high input SNRs in case of the Wiener filter. Figure 4 shows the SNR improvements for varying filter lengths L in case of a constant noise IFC for WGN and babble noise. We compare the MVDR filter (solid) and the conventional Wiener filter (dashed) for the input SNRs -10 (black), 0 (blue) and 20 dB (red). Beside the noise power estimate, the filter length L is chosen as a compromise between noise reduction (larger values) and speech distortion (smaller values). The best performance in the context of a blind algorithm is achieved for filter lengths between 8 and 12.

Informal listening tests show that the MVDR filter leads to less musical noise and less modulation artifacts of the speech components compared to the Wiener filter, especially for input SNRs above 0dB. The filter has a distinct smoothing effect which also leads to slightly reverberant signals. However, the speech signal seems to be more present in the processed signals for all input SNRs.

5. RELATION TO PRIOR WORK

The theory of MVDR filters goes back to the Capon filters described in [14]. In multi-channel speech processing they were used in various applications in conjunction with microphone arrays [15, 16, 17] and further discussed in the context of room acoustics in [18]. A comprehensive theory of subband filters can be found in [19]. Also, Kalman filters were used to exploit temporal correlations in subbands in [20, 21] and extended in [22, 23] to the speech and noise components. However for Kalman filtering, a time-varying transition matrix needs to be estimated which is difficult when only a noisy signal is given. Furthermore, such a filtering can be applied in the frequency-domain as well. [24] uses a cascaded filter-bank system to implement a subband frequency-domain Wiener filter to increase the frequency resolution for NR in hearing aids. The authors in [4, 5] assume idealized conditions for the IFC estimation, i.e. they calculate the unknown IFCs based on the known speech and noise signals.

6. CONCLUSIONS

We introduced and discussed two methods to estimate the inter-frame correlation which is needed to implement optimal subband time-domain FIR filters as proposed in [4, 5]. We showed that the speech enhancement performance can be improved especially for high input SNRs by exploiting rather simple assumptions concerning the inter-frame correlation of the noise signal. Setting the noise IFC to a constant vector achieves similar results than using a more complicated approach based on a binary mask.

7. REFERENCES

[1] Y. Ephraim and D. Malah, "Speech enhancement using a minimum-mean square error short-time spectral amplitude estimator," *IEEE Trans. Acoust. and Speech Signal Process.*, vol. 32, no. 6, pp. 1109–1121, 1984.

[2] P. Vary and R. Martin, *Digital Speech Transmission: Enhancement, Coding And Error Concealment*, John Wiley & Sons, 2006.

[3] R. Martin, U. Heute, and C. Antweiler, *Advances in Digital Speech Transmission*, John Wiley & Sons, 2008.

[4] J. Benesty and Y. Huang, "A single-channel noise reduction MVDR filter," in *Proc. IEEE Int. Conf. Acoustics, Speech, and Signal Process. (ICASSP)*, Prague, Czech Republic, May 2011, pp. 273–276.

[5] Y. Huang and J. Benesty, "A multi-frame approach to the frequency-domain single-channel noise reduction problem," *IEEE Trans. Audio, Speech, and Language Process.*, vol. 20, no. 4, pp. 1256 –1269, May 2012.

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

[7] T. Gerkmann and R. C. Hendriks, "Unbiased MMSE-based noise-power estimation with low complexity and low tracking delay," *IEEE Trans. Audio, Speech, and Language Process.*, vol. 20, no. 4, pp. 1383–1393, May 2012.

[8] G. Kim, Y. Lu, Y. Hu, and P. C. Loizou, "An algorithm that improves speech intelligibility in noise for normal-hearing listeners," *J. Acoust. Soc. Am. (JASA)*, vol. 126, no. 3, pp. 1486–94, 2009.

[9] P. P. Vaidyanathan, *Multirate systems and filter banks*, Prentice-Hall, Inc., Upper Saddle River, NJ, USA, 1993.

[10] E. Hänsler and G. Schmidt, *Acoustic Echo and Noise Control: A Practical Approach*, Adaptive and Learning Systems for Signal Processing, Communications and Control Series. John Wiley & Sons, 2005.

[11] G. H. Golub and C. F. Van Loan, *Matrix Computations*, Johns Hopkins Studies in the Mathematical Sciences. Johns Hopkins University Press, 1996.

[12] S. Haykin, *Adaptive filter theory (3rd ed.)*, Prentice-Hall, Inc., Upper Saddle River, NJ, USA, 1996.

[13] Y. Hu and P. C. Loizou, "Evaluation of objective quality measures for speech enhancement," *IEEE Trans. Audio, Speech, and Language Process.*, vol. 16, no. 1, pp. 229–238, 2008.

[14] J. Capon, "High-resolution frequency-wavenumber spectrum analysis," *Proc. IEEE*, vol. 57, no. 8, pp. 1408 – 1418, Aug. 1969.

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

[16] S. Gannot and I. Cohen, "Adaptive beamforming and postfiltering," in *Springer Handbook of Speech Processing*, J. Benesty, M. M. Sondhi, and Y. Huang, Eds., pp. 945–978. Springer Berlin Heidelberg, 2008.

[17] M. Brandstein and D. Ward, *Microphone Arrays: Signal Processing Techniques and Applications*, Digital Signal Processing. Springer, 2010.

[18] E. Habets, J. Benesty, I. Cohen, S. Gannot, and J. Dmochowski, "New insights into the MVDR beamformer in room acoustics," *IEEE Trans. Audio, Speech, and Language Process.*, vol. 18, no. 1, pp. 158 –170, jan. 2010.

[19] K.-A. Lee, W.-S. Gan, and S. M. Kuo, *Subband Adaptive Filtering: Theory and Implementation*, Wiley Publishing, 2009.

[20] W.-R. Wu and P.-C. Chen, "Subband Kalman filtering for speech enhancement," *IEEE Trans. Circuits and Syst. II*, vol. 45, no. 8, pp. 1072 –1083, Aug. 1998.

[21] H. Puder, "Kalman-filters in subbands for noise reduction with enhanced pitch-adaptive speech model estimation," *European Trans. on Telecommunication*, vol. 13, no. 2, pp. 139–148, 2002.

[22] T. Esch and P. Vary, "Speech enhancement using a modified Kalman filter based on complex linear prediction and supergaussian priors," in *Proc. IEEE Int. Conf. Acoustics, Speech, and Signal Process. (ICASSP)*, Las Vegas, Nevada, U.S.A., April 2008, pp. 4877 –4880.

[23] T. Esch and P. Vary, "Exploiting temporal correlation of speech and noise magnitudes using a modified Kalman filter for speech enhancement," in *Proc. ITG Symposium Speech Communication*, Aachen, Germany, Oct. 2008.

[24] A. Schasse, R. Martin, W. Soergel, T. Pilgrim, and H. Puder, "Efficient implementation of single-channel noise reduction for hearing aids using a cascaded filter-bank," in *Proc. ITG Symposium Speech Communication*, Braunschweig, Germany, Sept. 2012, pp. 287–290.