

A Hybrid Subband Adaptive System for Speech Enhancement in Diffuse Noise Fields

Hamid Reza Abutalebi, Hamid Sheikhzadeh, *Member, IEEE*, Robert L. Brennan, *Member, IEEE*, and George H. Freeman, *Member, IEEE*

Abstract— Performance of Adaptive Noise Cancellation (ANC) degrades severely when uncorrelated noise components are present at the two inputs. Thus, practical background diffuse noises pose a serious problem for ANC systems. In this research, we propose a new hybrid system that integrates Subband Adaptive Filters (SAFs) and a Wiener filter. The hybrid system is implemented on an oversampled DFT filterbank that efficiently integrates the SAF and the Wiener filter components in the frequency-domain. Performance evaluation of the hybrid system in presence of diffuse noise interference shows that the proposed system is superior to both the Wiener filter and the SAF sub-systems.

Index Terms— Subband Adaptive Filter (SAF), Adaptive Noise Cancellation (ANC), Uncorrelated noise, Coherence, Diffuse noise field, Wiener filter.

I. INTRODUCTION

NOISE CANCELLATION is a common adaptive filtering application [1]. To reduce the adaptive filter length and accelerate the convergence rate, Subband Adaptive Filters (SAFs) have been introduced [2]. However, the use of adaptive filters for noise cancellation (in both fullband and subband implementations) has been limited by some difficulties. A major problem in Adaptive Noise Cancellation (ANC) is weak correlation between noise signals at two input microphones. Clearly, adaptive filters are able to eliminate only the noise signals (at the two microphones) that are correlated. However, in many real-life situations, the inputs of the adaptive filter are partially correlated.

In this letter, we study the problem of uncorrelated noise in subband adaptive noise cancelers. The objective is to present an efficient hybrid system that works in real-life noisy environments. The system is targeted for implementation on an ultra-low resource platform that employs a weighted overlap-add (WOLA) generalized discrete Fourier transform (DFT) filterbank [3].

Manuscript received November 13, 2002; revised January 13, 2003. The associate editor coordinating the review of this manuscript and accepting it for publication was Dr. Darren B. Ward.

H. R. Abutalebi is with the Electrical Engineering Department, Amirkabir University of Technology, Tehran, Iran.

H. Sheikhzadeh is with Dspfactory Ltd., Waterloo, ON, N2V 1K8 Canada and also the Electrical and Computer Engineering Department, University of Waterloo, Waterloo, ON, N2L 3G1, Canada. (e-mail: hsheikh@dspfactory.com).

R. L. Brennan is with Dspfactory Ltd., Waterloo, ON, N2V 1K8, Canada (e-mail: rob.brennan@dspfactory.com)

G. H. Freeman is with the Electrical and Computer Engineering Department, University of Waterloo, Waterloo, ON, N2L 3G1, Canada.

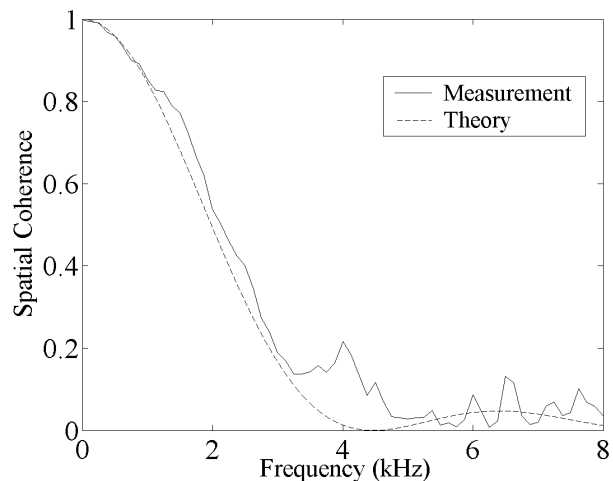


Fig. 1. Spatial coherence of (solid line) the generated diffuse noise field compared to (dashed line) the theoretical curve.

II. SPATIAL COHERENCE AND DIFFUSE NOISE FIELD

The most commonly used feature to characterize the correlation of two (noise) signals x and y (received at two microphones), is the spatial coherence that is defined based on cross- and auto-spectral density functions as [4]

$$\Gamma_{xy}^2(f) = \frac{|P_{xy}(f)|^2}{P_{xx}(f) \cdot P_{yy}(f)}. \quad (1)$$

In a diffuse noise field, the two microphones receive equal-amplitude and random phase noise signals from all directions. This results in a spatial coherence of the $\text{sinc}^2(\cdot)$ form for the 3-dimensional diffuse noise field [4],[5]

$$\Gamma_{xy}^2(f) = \frac{\sin^2(2\pi fd/c)}{(2\pi fd/c)^2} = \text{sinc}^2\left(\frac{2fd}{c}\right), \quad (2)$$

where d is the microphone spacing and c is the sound velocity ($c = 340$ m/s).

The spatial coherence is widely used in the ANC literature to obtain the noise reduction factor of adaptive filtering as [5]

$$\text{NR}(f) = \frac{\text{input noise power at frequency } f}{\text{output noise power at frequency } f} = \frac{1}{1 - \Gamma_{xy}^2(f)}. \quad (3)$$

According to (2) and (3), increasing microphone spacing (d) will limit the noise reduction capability of the adaptive filter to lower frequency bands. Although a decrease in microphone

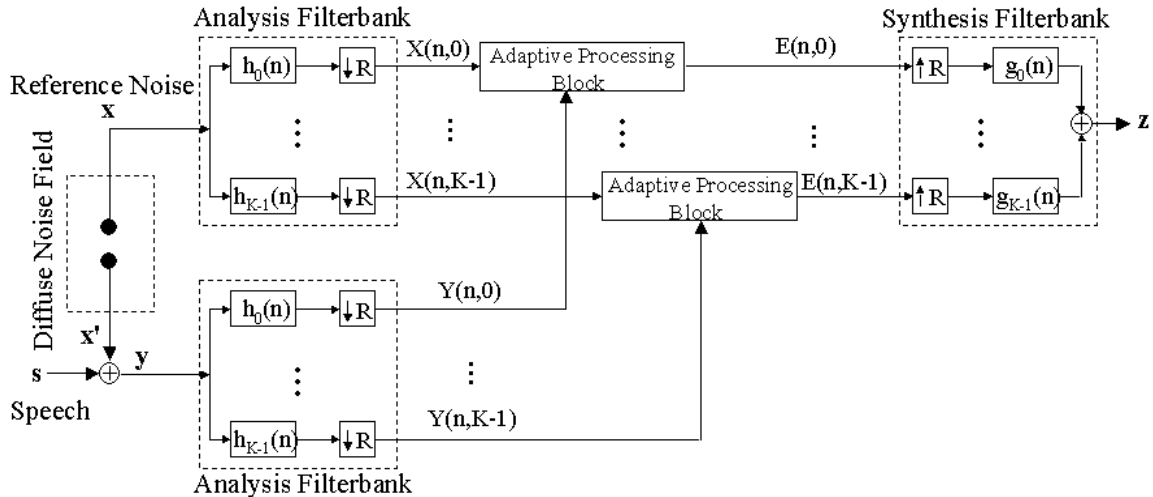


Fig. 2. Block diagram of the SAF system.

separation alleviates this problem, it aggravates the cross-talk problem. As a result, other solutions should be investigated.

In this letter, we have used two input noise signals recorded in a diffuse field at a sampling rate of 16 kHz. The noise field is generated by playing 4 independent pink noise signals in an isolated non-reverberant sound room using 4 loudspeakers located at various locations. With a microphone spacing of $d = 38$ mm, the spatial coherence of the two recorded noises has the form shown in Fig. 1. It closely matches the theoretical curve computed using (2) for $d = 38$ mm. Thus, the noise field can be regarded as a diffuse field in the area close to the two microphones.

III. SUBBAND ADAPTIVE FILTERING

Fig. 2 displays block diagram of the employed SAF system in noise cancellation setup. The analysis filterbanks split the noisy and reference inputs into K frequency bands. After decimation of each subband signal by a factor of R , adaptive processing blocks cancel the noise in the output signal using the Least Mean Square (LMS) algorithm. The SAF system employs an oversampled filterbank ($R < K$) to minimize aliasing error and to maintain a low processing delay [3,6]. Each adaptive processing block represents an adaptive filter that works on a specific frequency band. Finally, the synthesis filterbank merges the subband-enhanced signals to obtain the time-domain enhanced speech.

When adaptive filters are employed in low-coherence subbands, the filters change only slightly from their zero initial values due to the low level of correlated components. This leads to a small noise reduction factor, as predicted by (3). This is demonstrated in Fig. 3 by the noise reduction factor curves of the subband adaptive noise canceler. The solid curve displays the ratio of the measured input to output noise power when two input microphones are located in the generated diffuse noise field. This curve consistently follows the theoretical (dashed) curve computed by inserting (2) into (3).

IV. THE HYBRID SYSTEM

In addition to ANC, several other approaches have been proposed for dual-microphone speech enhancement. Fixed and

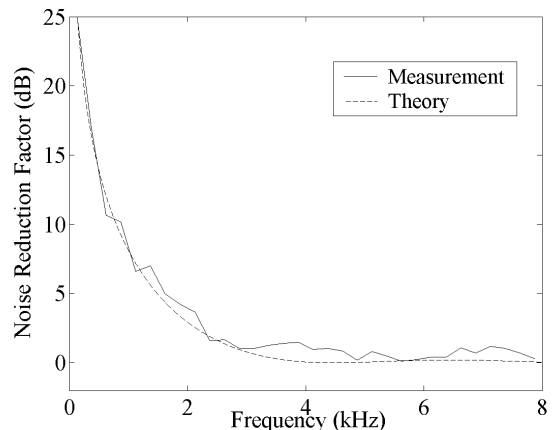


Fig. 3. Noise reduction factor of the SAF system in (solid line) the generated diffuse noise field and (dashed line) the theoretical noise reduction factor of adaptive filtering in diffuse field calculated by inserting (2) in (3).

adaptive (delay and sum) beamformers are common choices. However, in diffuse fields, they offer up to 3 dB of noise reduction at higher bands and almost no noise reduction in low frequencies where the spatial coherence is high [7]. Employing an adaptive post filter can improve the performance only at high frequencies [8]. Also, the Generalized Sidelobe Canceller performs well for coherent noise fields but fails in diffuse noise fields [7].

Single-microphone methods are other possible candidates for speech enhancement in diffuse noise fields, Wiener filtering being one of the most widely used approaches [9]. Although Wiener filtering is a low complexity and robust technique, it significantly suffers from signal distortion caused by inadequate estimations of the noise spectrum. Specifically, in low SNRs and lowpass noises, the performance of this method degrades considerably.

To compensate for the inability of SAF in eliminating uncorrelated noises, we combine an SAF system with a Wiener filtering technique, presenting a hybrid system shown in Fig. 4. The figure displays the noise reduction structure in each subband of the proposed hybrid system (called SAFWF here). This

structure is used in each adaptive processing block of the system shown in Fig. 2.

A Voice Activity Detector (VAD) has been used to determine the noise-only portions (pauses) in the primary input. We have used a modified version of the ETSI AMR-2 VAD [10] that has been implemented on the same oversampled WOLA filterbank [11]. When the VAD detects a pause, the adaptive filter is allowed to adapt and the Wiener filter can update its noise spectrum estimate.

The primary and reference inputs shown in Fig. 4 are complex (frequency-domain) subband sequences. For each subband k , the complex adaptive filter interprets the reference input as a time-sequence signal and convolves it with the adaptive filter impulse response. For the second stage, a generalized Wiener filter is applied by multiplying the adaptive filter outputs $E(n,k)$ by a time varying real gain $G(n,k)$ i.e.,

$$Z(n,k) = G(n,k) \cdot E(n,k), \quad (4)$$

$$G(n,k) = \left[1 - \frac{\beta |\hat{N}(n,k)|^\alpha}{|E(n,k)|^\alpha} \right]^\frac{1}{\alpha}, \quad (5)$$

where α and β are Wiener filter parameters [9] that can be optimized for a particular application. Considering the noise reduction and signal distortion tradeoff, we have set α to 1 and β to 1.5. Also, $\hat{N}(n,k)$ is an estimate of uncorrelated noise amplitude and is computed as $\hat{N}(n,k) = E(n,k)|_{VAD=0}$. In this hybrid system, the Wiener filter is employed to remove the residual uncorrelated noise components. Since the filter is applied on $E(n,k)$ instead of $Y(n,k)$ (as it is done in standard Wiener filtering), the amount of signal distortion is significantly reduced. This will be discussed more in Section V.

The proposed hybrid system has some commonalities with the method described in [12] since both apply a Wiener post-filter after an adaptive stage. However, while [12] uses both time and frequency domain processes, our proposed method is fully implemented on a subband structure. Furthermore, [12] employs a time-domain EM algorithm for adaptation but we use much more efficient subband adaptive filters on an oversampled filterbank.

V. SYSTEM EVALUATION

The noise reduction performances of the proposed SAFWF system, the SAF system and the standard Wiener filtering (STDWF) are compared in a diffuse noise field. We used several sentences from the TIMIT database as our speech material. From the two recorded noise inputs (generated as described in Section II), one was employed as the reference signal, and the other was added to speech at various SNRs. The results of system evaluations for one sentence are displayed by the spectrograms in Fig. 5. (Since it is expected that the SAFWF especially exhibit a superior performance in lower subbands, spectrograms are shown in 0-2 kHz range.). Depicted in rows 1 to 4 are the spectrograms for the clean speech, the noisy (primary) input (in 0 dB SNR pink noise), the signal enhanced by the SAFWF method (with $\beta = 1.5$), and the signal enhanced by the STDWF method (with $\beta = 1.5$),

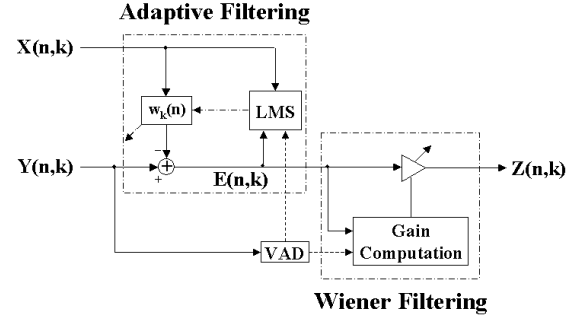


Fig. 4. Noise reduction structure in each subband of the proposed hybrid system (SAFWF).

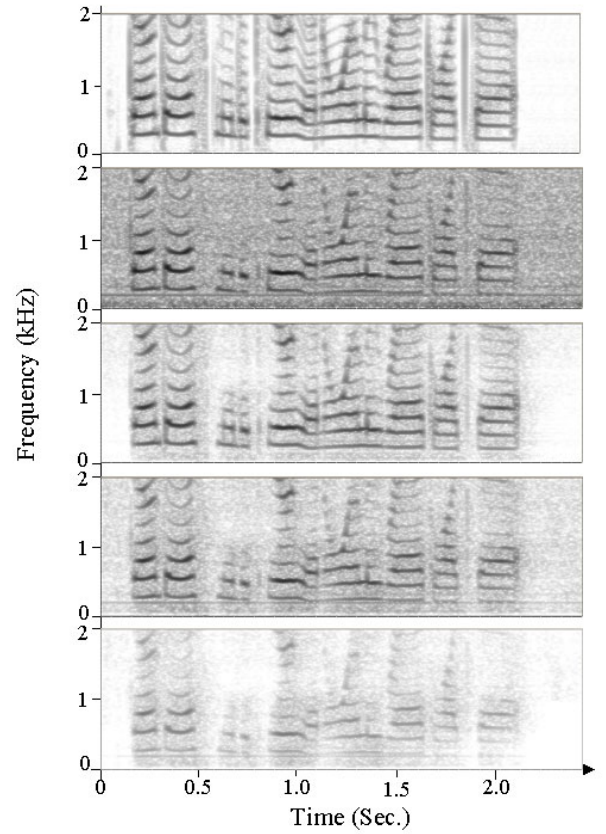


Fig. 5. Spectrograms of (1) clean speech, (2) noisy input, (3) SAFWF output ($\beta = 1.5$), (4) STDWF output ($\beta = 1.5$), (5) STDWF output ($\beta = 2.0$). (Frequency range limited to 0-2 kHz).

respectively. As shown, the STDWF has more residual background noise than the SAFWF. Increasing β (to $\beta = 2.0$) in the STDWF reduces the background noise as evident from the fifth row in Fig. 5. However, this reduction comes at the cost of increased speech cancellation and distortion, evident by comparing rows 3 and 5.

The performance of the proposed system in reducing the uncorrelated noise components is further demonstrated by spectrograms (using the full band of 0-8 KHz frequency range) displayed in Fig. 6. Shown, are the spectrograms of the clean

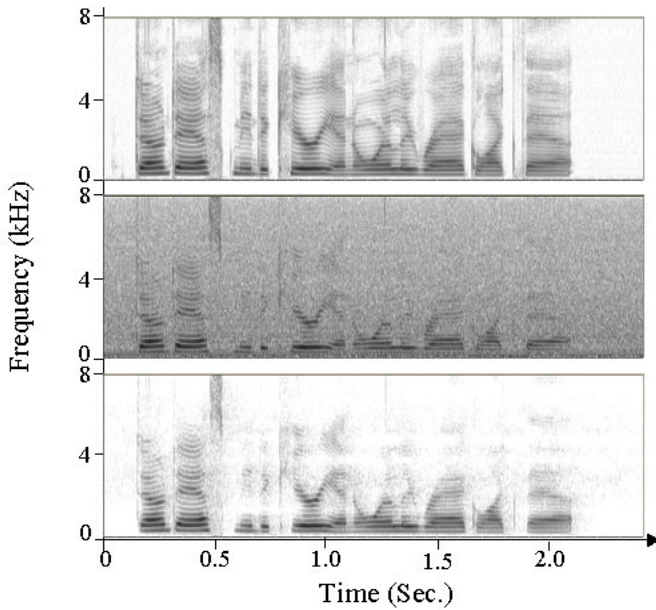


Fig. 6. Spectrograms of (1) clean speech, (2) noisy input, (3) SAFWF output ($\beta = 1.5$) (Fullband).

speech (the same as in Fig. 5), noisy input and the SAFWF enhanced signals, respectively. As it is clear from higher frequency bands of the outputs, Wiener filtering has compensated for the inability of the adaptive filter to remove uncorrelated noise components.

For objective evaluation of speech quality, we have employed the log area ratio (LAR) distance that has been shown to have the highest correlation with subjective quality assessments among all frequency invariant distance measures [13]. (In [13], a correlation of magnitude of 0.62 is reported between the LAR distance and subjective evaluations.) As shown in Fig. 7, the SAFWF outperforms the SAF and STDWF systems especially at low SNRs. Informal listening test results are consistent with the LAR distances, however, formal listening tests are needed to confirm the results.

VI. CONCLUSION

In this research, we addressed the issue of the noise reduction in diffuse noise fields. We proposed, implemented and evaluated a subband hybrid system combining SAFs and Wiener filtering. The proposed system employs ANC to take advantage of high coherence at low frequencies, and Wiener filtering to eliminate the residual noise particularly at higher frequencies. In the SAFWF, the adaptive filter cleanly reduces the noise in the lower subbands, while in the STDWF inaccurate estimation of noise spectrum significantly distorts the speech signal. Subsequent Wiener filtering in the hybrid system is thus performed with higher SNRs. Considering the perceptual importance of the lower subbands, the hybrid technique is subjectively preferred.

Furthermore, the proposed technique is desirable from a computational point of view since both subband adaptive and

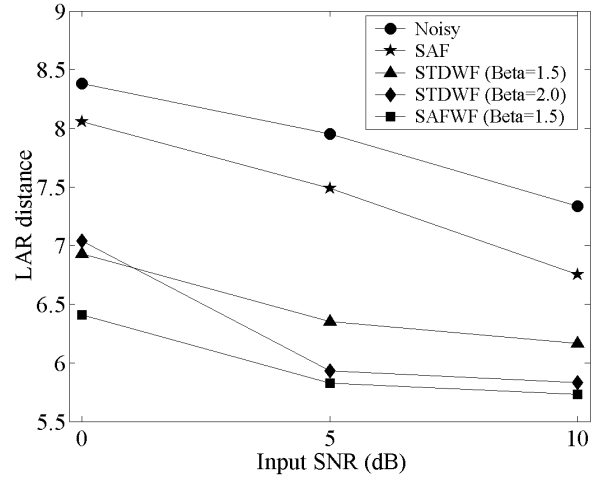


Fig. 7. LAR distance for noisy input, and outputs of SAF, STDWF ($\beta = 1.5$), STDWF ($\beta = 2.0$) and SAFWF ($\beta = 1.5$) for various input SNRs.

Wiener filtering are applied in the subband domain. The complete enhancement system containing the SAF, the Wiener filter and the VAD, is efficiently implemented on a fast oversampled WOLA filterbank [3].

REFERENCES

- [1] S. Haykin, *Adaptive Filter Theory*, 3rd ed., Prentice Hall Inc., 1996.
- [2] A. Gilloire, and M. Vetterli, "Adaptive filtering in subbands with critical sampling: analysis, experiments and application to acoustic echo cancellation," *IEEE Trans. Signal Processing*, vol. SP-40, no. 8, pp. 1862-1875, Aug. 1992.
- [3] R. Brennan, and T. Schneider, "A flexible filterbank structure for extensive signal manipulation in digital hearing aids," in *Proc. IEEE Int. Symp. Circuits and Systems*, pp. 569-572, 1998.
- [4] M. Tohyama, H. Suzuki, and Y. Ando, *The Nature and Technology of Acoustic Space*, Academic Press, 1995.
- [5] M. M. Gouling, "Speech enhancement for mobile telephony," *IEEE Trans. Veh. Technol.*, vol. 39, no. 4, pp. 316-326, Nov. 1990.
- [6] K. Tam, H. Sheikhzadeh, and T. Schneider, "Highly oversampled subband adaptive filters for noise cancellation on a low-resource DSP system," in *Proc. ICSLP*, 2002.
- [7] J. Bitzer, K. U. Simmer, and K. D. Kammeyer, "Multichannel noise reduction – algorithms and theoretical limits," in *Proc. EUSIPCO*, vol. 1, pp. 105-108, 1998.
- [8] C. Marro, Y. Mahieux, and K. U. Simmer, "Analysis of noise reduction and dereverberation techniques based on microphone arrays with postfiltering," *IEEE Trans. Speech Audio Processing*, vol. 6, no. 3, pp. 240-259, May 1998.
- [9] J. S. Lim, and A. V. Oppenheim, "Enhancement and bandwidth compression of noisy speech," *Proc. IEEE*, vol. 67, no. 12, pp. 1586-1604, Dec 1979.
- [10] *Universal Mobile Telecommunication Systems (UMTS); Mandatory Speech Codec speech processing functions, AMR speech codec; Voice Activity Detector (VAD)*, (3GPP TS 26.094 version 4.0.0 Release 4), ETSI TS 126 094 V4.00 (2001-03).
- [11] E. Cornu, *et al.* "ETSI AMR-2 VAD: evaluation and ultra low-resource implementation," in *Proc. ICASSP*, 2003.
- [12] M. Feder, A. V. Oppenheim, and E. Weinstein, "Maximum likelihood noise cancellation using the EM algorithm," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. 37, no. 2, pp. 204-216, Feb. 1989.
- [13] S. R. Quakenbusch, T. P. Barnwell III, and M. A. Clements, *Objective Measures of Speech Quality*, Prentice-Hall, Englewood Cliffs, NJ, 1988.