EVALUATION OF A HYBRID SUBBAND ADAPTIVE SYSTEM FOR NOISE REDUCTION IN MOBILE AND VEHICULAR APPLICATIONS

H. R. Abutalebi[1,3], H. Sheikhzadeh[2,3], R. L. Brennan[2], and G. H. Freeman[3]

[1]EE Dept., Amirkabir University of Technology, Tehran, Iran
[2]Dspfactory Ltd., Waterloo, Ontario, Canada
[3]E&CE Dept., University of Waterloo, Waterloo, Ontario, Canada

E-mails: abutalebi@aut.ac.ir, {hsheikh, rob.brennan}@dspfactory.com, freeman@pce.uwaterloo.ca

ABSTRACT

In this research, we have evaluated the performance of a new hybrid noise reduction system in various real-life noise fields occurring in mobile and vehicular applications. The noise reduction block is composed of a Subband Adaptive Filter (SAF) followed by Wiener filtering. For system evaluation, we have used recordings of real-life noises to contaminate speech materials chosen from the TIMIT database. It is shown that all the recorded noises are samples of diffuse noise fields. As a result, the hybrid system outperforms both the SAF and standard Wiener filtering in all sets of the recordings. The preference of this hybrid system is especially noted in the case of lowpass noise and intense noise conditions.

1. INTRODUCTION

Subband Adaptive Filters (SAFs) have shown good performance in noise cancellation when the two input noises are correlated [1]. However, many real-life noise fields in mobile and vehicular applications are approximately diffuse [2].

Diffuse noise fields are mathematically characterized through the spatial coherence function commonly used to specify the correlation of two noise signals x and y recorded at two input microphones in a noise field. The spatial coherence function is defined based on the cross- and auto-spectral densities as [3]:

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

For diffuse noise fields, the spatial coherence, obtained by averaging (1) over the spherical coordinates, drops with frequency, following a $\mathrm{sinc}^2(.)$ form [2-3]:

$$\Gamma_{xy}^2 = \frac{\sin^2(2\pi f d/c)}{(2\pi f d/c)^2} = \mathrm{sinc}^2(2fd/c)\ , \tag{2}$$

where c is the sound velocity (c = 340 m/s) and d is the distance between the two input microphones. Considering (2), it is obvious that the SAF can only eliminate the noise in the lower frequency regions where there is a high correlation (coherence) between the two microphone signals.

To compensate for the inability of the SAF to eliminate noise in diffuse noise fields, we have proposed a hybrid system integrating the SAF system and Wiener filtering (called SAFWF here), and have examined its performance in an isolated non-reverberant sound room [4]. Based on the promising results obtained, we further evaluate the performance of the SAFWF system in real-life mobile and vehicular noisy environments.

This paper is organized as follows. The structure of the hybrid system is described in the next section. Sec. 3 describes the employed noise and speech materials. System evaluations are reported in Sec. 4, and finally the conclusions of this work are presented in Sec. 5.

2. THE HYBRID SYSTEM

To improve the performance of SAF systems in diffuse noise fields, we have already proposed a hybrid system that takes advantage of the complementary characteristics of subband adaptive and Wiener filtering, resulting in a much higher noise reduction performance for diffuse noise fields [4].

Shown in Fig. 1 is the block diagram of the employed enhancement system. The analysis filterbanks split the primary (noisy) and reference (noise) inputs into K subbands. After decimation of each subband by a factor of R, a noise reduction block reduces the noise in the corresponding frequency band. Finally, the synthesis filterbank combines the subband enhanced signals to obtain a time-domain output. To limit the aliasing distortion while maintaining a low processing delay, oversampled DFT filterbanks are used as analysis/synthesis filterbanks. Based on Weighted OverLap-Add (WOLA) analysis/synthesis, the filterbanks are implemented on an ultra low-resource platform [5].

In the original SAF system, an adaptive filter is used as the noise reduction block. In the SAFWF, this block is replaced with an adaptive filter cascaded with a Wiener filter as depicted in Fig. 2. After elimination of the correlated noise components by the adaptive filter, the Wiener filter further processes the error signal reducing the remaining uncorrelated noise. Since the quality of the error signal is already improved through adaptive filtering prior to Wiener filtering, much fewer artifacts are expected compared to Wiener filtering alone [4]. Objective evaluations reported in Sec. 4 also confirm this conclusion.

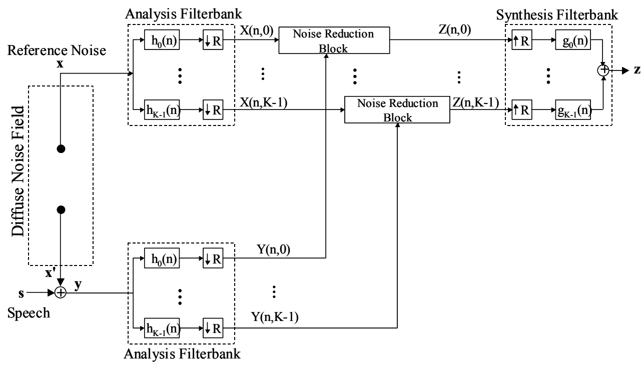

Fig. 1. Block diagram of the SAF system.

Wiener filtering is implemented using a frequency-domain generalized Wiener filter [6]. As shown in Fig. 2, the (subband) adaptive filter outputs are multiplied by a time-varying real gain G(n,k); i.e., $Z(n,k)=G(n,k)\cdot E(n,k)$, where

$$G(n,k)=\left[1-\frac{\beta\,|\,\hat{N}(n,k)\,|^{\alpha}}{|\,E(n,k)\,|^{\alpha}}\right]^{\frac{1}{\alpha}}, \tag{3}$$

and $\alpha=1$ and $\beta=1.5$ are Wiener filter parameters that have been optimized for this application [4] .Also, $\hat{N}(n,k)$ is the uncorrelated noise spectrum estimated during speech pauses.

A Voice Activity Detector (VAD) has been employed to detect the noise-only portions of the primary (noisy) input. The VAD is a modified version of the ETSI AMR-2 VAD [7] that has been implemented on the same oversampled WOLA filrterbank [8]. When a speech pause is detected, subband adaptive filters are adapted, and the noise spectrum estimate of the Wiener filter is updated.

The complete system (the SAFs, the Wiener filter, and the VAD) is efficiently implemented on the oversampled WOLA filterbank [5].

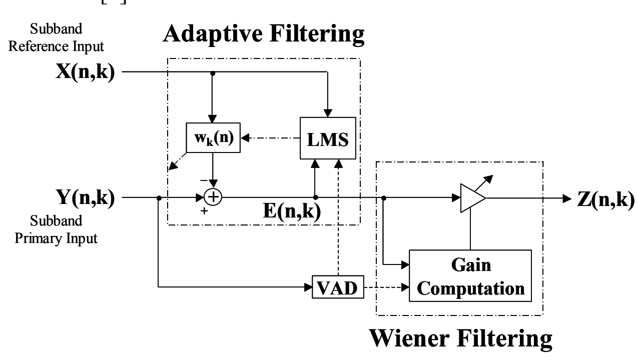

Fig. 2. Block diagram of the noise reduction block.

3. NOISE AND SPEECH MATERIALS FOR SYSTEM EVALUATION

In this research, we have evaluated the performance of the hybrid SAFWF system employing several noises recorded by two input microphones in real-life situations. The sampling frequency was 16 kHz, and the microphone spacing was set to $d=38$ mm (a typical value for boomless headsets). Recordings were done in the following situations:

(1) Sitting in a shopping mall (Sit-Mall1)

(2) Inside a working car parked next to a highway (HWY)

(3) Inside a moving car with open windows (CarWOpen)

(4) Inside a moving car with closed windows (CarWClose)

In each recording set, the first and second microphone inputs are considered as primary and reference noises, respectively. Figs. 3(a)-(d) display the average Power Spectral Density (PSD) of the primary noise in each case. While all noises have lowpass spectra, the PSDs of the two Car noises (CarWOpen and CarWClose) fall much faster with frequency. The effect of engine noise appears as two local peaks at about 2.7 kHz and 5.4 kHz in Figs. 3(b)-(d).

The spatial coherence curves of the first and second microphone signals in above recording sets are plotted in Figs. 4(a)-(d), respectively. The curves closely match theoretical spatial coherence computed for diffuse noise by (2) with d = 38 mm. This observation is consistent with the diffuse noise assumption usually considered for most environmental noise fields in vehicular applications [2]. Also, due to its directivity, the engine noise has acted more as a coherent source evident from the peaks in the coherence function at 2.7 kHz and 5.4 kHz (in Figs. 4(b)-(d)).

Several sentences from the TIMIT database have been used as the speech material. For each set of the noise recordings, the primary noise was added to speech signal at 0 dB SNR and used as the noisy input.

4. EVALUATION OF THE HYBRID SYSTEM

Through subjective and objective evaluations, the performance of the SAFWF system is compared to that of the SAF approach. Also, we have examined the performance of single-microphone standard Wiener filter (STDWF) method where the filter is directly applied to the primary noisy input.

To objectively measure the system performance, we use the Log Area Ratio (LAR) metric which has been shown to have the highest correlation with subjective assessments among all frequency-invariant distance measures [9]. Given the reflection (PARCOR) coefficients $K(m,l), l=1,...,L$ of the m^{th} frame, the corresponding Area Ratio (AR) is defined as [9]:

$$AR(m,l)=\frac{1+K(m,l)}{1-K(m,l)}. \tag{4}$$

Here, we have set the number of the AR coefficients to L = 16. Given $AR_s(m,l)$ and $AR_z(m,l)$ for m^{th} frame of signals s (clean speech) and z (enhanced output) , the LAR distance between the m^{th} frames of two signals is calculated as [9]:

$$LAR_{sz}(m)=\left\{\frac{1}{L}\sum_{l=1}^{L}\left|20\log_{10}\left[\frac{AR_s(m,l)}{AR_z(m,l)}\right]\right|^2\right\}^{\frac{1}{2}}. \tag{5}$$

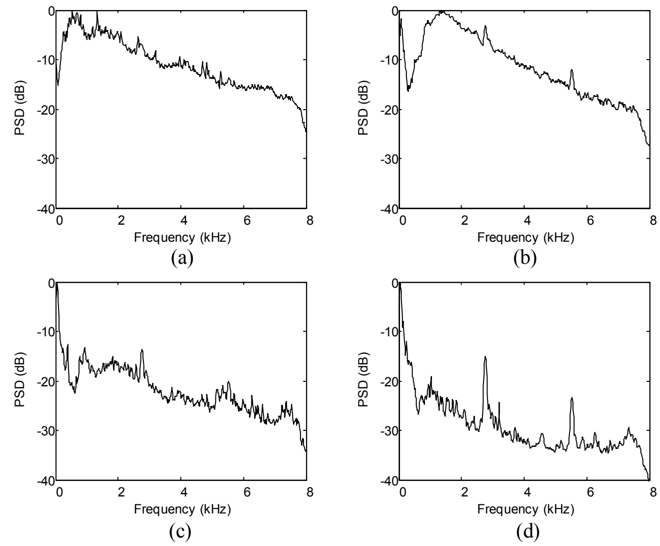

Fig. 3. Power Spectral Density (PSD) of the noises recorded by two microphones in various situations: (a) Sit-Mall1, (b) HWY, (c) CarWOpen, (d) CarWClose.

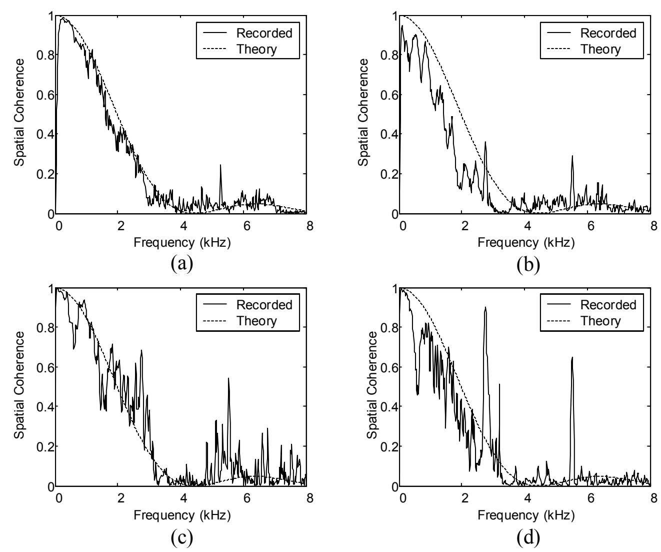

Fig. 4. Spatial coherence of the noises recorded by two microphones in various situations: (a) Sit-Mall1, (b) HWY, (c) CarWOpen, (d) CarWClose.

In order to remove frames with unrealistically high LAR distances, we compute the overall LAR distance by first discarding frames with the top 5% LAR values, and then averaging (5) over the remaining frames (as suggested in [10]).

The results of objective evaluations are shown in Fig. 5. The low LAR-distance improvement obtained by the SAF indicates this method has difficulty rejecting noise in a diffuse noise environment. Evidently, SAF has better performance in the third and forth environmental conditions (CarWOpen and CarWClose). This can be justified by considering Figs. 3 and 4 as follows: 1) In the CarWOpen and CarWClose cases, the noise spectrum is dominated by lowpass and coherent components. This results in improved noise reduction by adaptive filtering. 2) In particular, there are highly coherent engine-related noises in these two cases that are efficiently canceled by the adaptive filters.

As is evident from Fig. 5, the hybrid system (SAFWF) outperforms both the SAF and the STDWF systems for all four test sets. Especially in CarWOpen and CarWClose cases, the SAFWF method produces better LAR distance improvements. This demonstrates that an improvement in the adaptive filer performance leads to better performance of the Wiener filter since Wiener filters typically generate less speech distortion at high input SNRs.

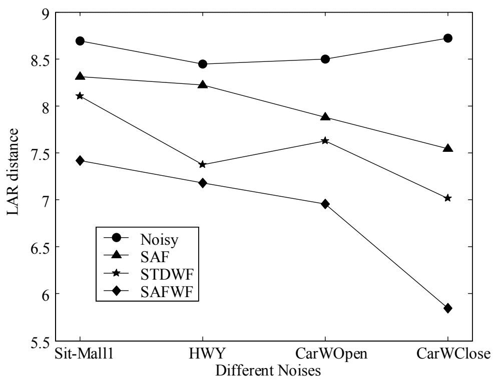

Fig. 5. LAR-distances for the noisy inputs and the outputs of the SAF, STDWF and SAFWF techniques in four noisy environments.

In order to examine the effect of input SNR, we have repeated the objective assessments for the forth set of noise recordings (CarWClose) at input SNRs of 0, 5, and 10 dB. As depicted in Fig. 6, the SAFWF offers a better performance in all cases, however its superiority is more evident at low input SNRs.

To further verify the hybrid system performance, we repeated the objective LAR tests using four other noise sets at 0 dB SNR. Two were recorded while walking (Walk-Mall), and sitting (Sit-Mall2) in a shopping mall. Two different office noise sets (Office1 and Office2) were also recorded. Recoding methods were exactly those used in Sec. 3. The objective evaluation results (keeping the same system parameters set in previous evaluations) are shown in Fig. 7 where the noises (from left to right) are sorted from less lowpass to more lowpass. As evident from Fig. 7, the results are consistent with those presented in Fig. 5.

Also, we have done some informal listening tests confirming the objective assessments. The artifacts produced by the STDAW technique are considerably reduced by applying the Wiener filter after the adaptive filtering.

5. CONCLUSION

Experimental results confirm the diffuse model for the noise field in vehicular and mobile applications. In this research, we evaluated the performance of a hybrid subband adaptive and Wiener filtering structure for noise reduction in diffuse fields.

The hybrid system benefits from the advantages of adaptive filters in coherent bands and also utilizes the Wiener filter to remove uncorrelated components. This way, it considerably reduces the inherent artifacts of the STDWF method.

Objective and subjective assessments confirm the superiority of the hybrid system for noise reduction in different real-life vehicular and mobile fields. Considering the spectral characteristics of the employed noises, it is clear that there are larger improvements for lowpass noise sources particularly at low SNRs.

6. REFERENCES

[1] K. Tam, H. Sheikhzadeh, and T. Schneider "Highly oversampled subband adaptive filters for noise cancellation on a low-resource DSP system," in *Proc. of the ICSLP,* 2002.

[2] M. M. Goulding, "Speech enhancement for mobile telephony," *IEEE Trans. Vehicular Tech.*, vol. 39, no. 4, pp. 316-326, Nov.1990.

[3] M. Tohyama, H. Suzuki, and Y. Ando, *The Nature and Technology of Acoustic Space*, Academic Press, 1995.

[4] H. R. Abutalebi, H. Sheikhzadeh, R. L. Brennan, and G. H. Freeman, "A hybrid subband system for speech enhancement in diffuse noise fields," Accepted for publication in *IEEE Signal Processing Letters,* 2003.

[5] 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.

[6] 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.

[7] ETSI TS 126 094 V4.00 (2001-03), "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)".

[8] E. Cornu, *et al*. "ETSI AMR-2 VAD: evaluation and ultra low-resource implementation," in *Proc. ICASSP*, 2003.

[9] S. R. Quakenbusch, T. P. Barnwell III, and M. A. Clements, *Objective Measures of Speech Quality*, Prentice-Hall, Englewood Cliffs, NJ, 1988.

[10] J. H. L. Hansen, and B. L. Pellom, "An effective quality evaluation protocol for speech enhancement algorithms," in *Proc. of the ICSLP*, vol. 7, pp. 2819-2822, Dec. 1998.

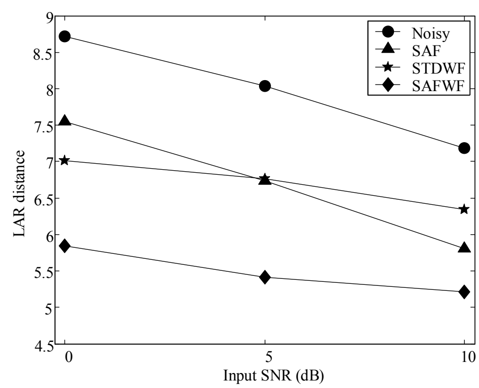

Fig. 6. LAR-distances for the noisy inputs and the outputs of the SAF, STDWF and SAFWF techniques for CarWClose case and different input SNRs.

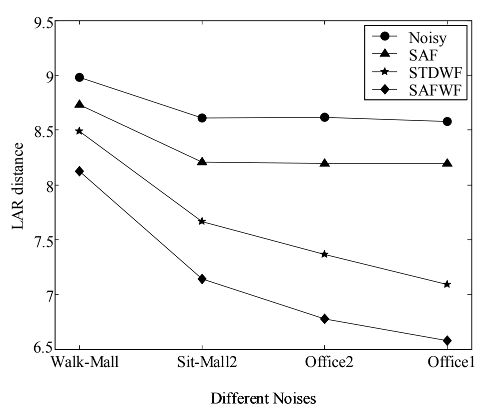

Fig. 7. LAR-distances for the noisy inputs and the outputs of the SAF, STDWF and SAFWF techniques in four new recording environments.