A SUBBAND HYBRID BEAMFORMING FOR IN-CAR SPEECH ENHANCEMENT

Charles Fox⋆†*, Guillaume Vitte*†*, Maurice Charbit*⋆*, Jacques Prado*⋆*, Roland Badeau*⋆*, Bertrand David*⋆

⋆ Institut Telecom, Telecom Paristech, CNRS-LTCI, † Parrot S.A

(charles.fox, guillaume.vitte)@parrot.com, (maurice.charbit, jacques.prado, roland.badeau, bertrand.david)@telecom-paristech.fr

ABSTRACT

In this paper, a multichannel speech enhancement system is presented, dedicated to in-car communication. An experimental study of the acoustic field inside the car interior leads us to propose a hybrid beamforming algorithm, taking two frequency ranges into account, according to the multichannel noise signal coherence. While in the high frequency range, the noise signals for different microphones show little coherence, the same signals are strongly coherent in the low frequency range. This observation leads us to design a hybrid system where an adaptive Minimum Variance Distortionless Response (MVDR) beamforming is used in the high frequency range, while a Speech Distortion Weighted Multichannel Wiener Filter (SDW-MWF) is applied in the low frequency range. The performance of the algorithm is evaluated by objective performance measurements on real in-car audio data, and shows promising results.

Index Terms— Speech Enhancement, Multi-Sensor, In-car Noise Reduction

1. INTRODUCTION

Speech enhancement for communication systems is still an important issue in the automotive environment. Even though, in the past decade, many studies have been dedicated to noise reduction [1, 2, 3] and while several solutions have emerged, the problem remains partially unsolved. The major difficulty lies in the fact that the speech signal and the environmental noise overlap in both time and spectral domains. Initially, only one microphone was used and most algorithms were based on spectral subtraction or spectral amplitude estimation [4]. The latter has been widely implemented in hands-free telephone systems [5].

One way of separating the target signal and the noise is to take into account different spatial characteristics that can be obtained through a multi-microphone array for instance. Recently, several studies have indeed been dedicated to multi-microphone noise reduction [2, 6], and this is also the general framework in which the work presented in this paper is developed.

The MVDR beamformer has been widely studied for multisensor acoustic noise reduction, since it is part of the multichannel Minimum Mean Squared Error (MMSE) Short Time Spectral Amplitude (STSA) estimator [7], which is commonly used in many communication devices for speech enhancement. One of the key difficulties related to this method is the estimation of the acoustical channel between the source of interest and the microphone array, which is challenging in a reverberant environment such as a car interior.

An adaptive approach, based on target signal prediction and Speech Presence Probability (SPP) estimation, has been developed in [8]. This method is efficient when the noise is non-coherent between 2 different microphones, since the prediction of the target signal can be disturbed if the noise is predictable as well as the target signal. As shown hereafter, the noise acoustic field shows strong inter-microphone coherence in the low frequency range for typical array dimensions. Hence, a good performance is not expected from the method developed in [8] in this range.

For strongly coherent noise, the noise recorded by different sensors differs only by a linear filtering. In this case, an Adaptive Noise Cancellation (ANC) [9] algorithm can be used for noise removal, since it estimates the transfer function between two sensors. However, this approach affects the speech intelligibility in the case of in-car noise reduction. This is due to the unavailability of a noise reference in the car interior. To take this distortion into account, a formulation of the Speech Distortion Weighted Multichannel Wiener Filter (SDW-MWF) close to the ANC [1] can be implemented. Therefore, the noise can be reduced without subtracting a critical amount of voice in the low frequencies, where the automotive noise shows strong coherence between microphones.

In section 2, an experimental study of the acoustical properties of the typical automotive noise field is presented. We also present a justification for using different algorithmic strategies in the low and high frequency ranges. In section 3 we develop the two denoising strategies combined in the proposed algorithm. Finally, we present objective evaluation results in section 4 and we conclude in section 5.

2. MOTIVATIONS AND FRAMEWORK

2.1. Properties of in-car noise acoustic field

In this section, the results of an experimental study are presented. Noise has been recorded in a car interior while running at 130 km/h (80 mph), in a typical highway situation. The noise is then produced by different sources such as aerodynamical perturbation outside, wheel/road contact, engine noise... The main purpose was to validate basic properties of noise observed on different microphone types and locations, more particularly the correlation between microphones located in different places in the car.

On the one hand, a qualitative evaluation from noise spectrograms, about 3 seconds long, reported in Figure 1, shows as expected a good stationarity in the whole frequency band. Also, most of the energy is located under 1kHz.

On the other hand, the Mean-Squared Coherence (MSC) is computed between two signals recorded by two microphones located at different positions in order to evaluate the ability to predict noise from one microphone to another. The MSC is defined for two wide-sense jointly stationary zero-mean random processes $x(t)$ and $y(t)$ by

$$\mathrm{C}_{xy}(f) = \frac{|S_{xy}(f)|^2}{S_{xx}(f)S_{yy}(f)} \tag{1}$$

where $S_{xy}(f)$ denotes the Fourier Transform of the covariance function $R_{xy}(\tau) = \mathrm{E}\left[x(t)y(t-\tau)\right]$, where $\mathrm{E}\left[\,\right]$ denotes the mathematical expectation, assumed not to depend on t. The MSC is less than

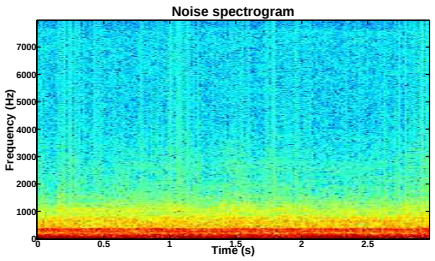

Fig. 1. Spectrogram of a typical car noise

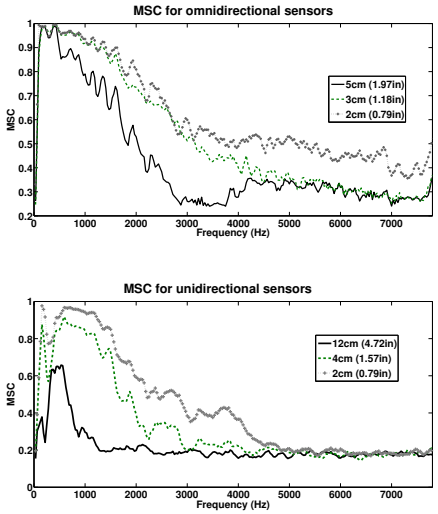

Fig. 2. Mean-Squared Coherence (MSC) for several distances between two microphones.

or equal to 1 and the equality is reached if and only if there is a filter h such that $y = h \star x$, where $\star$ denotes the convolution operation. Therefore a value of MSC close to 1 means that one process can be accurately and linearly predicted by the other one.

The MSCs as functions of frequency are reported in Figure 2 for omnidirectional and unidirectional microphones, and for different distances between them. It appears that for unidirectional microphones, the coherence is very low regardless of the distance between the microphones for most of the frequency bands. For omnidirectional microphones the coherence is above 0.8 in the low frequency band. This phenomenon has already been observed in [6]. It is also shown that, under simple assumptions, the shape of the MSC is $|\frac{\sin 2\pi fd/c}{2\pi fd/c}|^2$ [6] where d denotes the distance between microphones, c the sound velocity and f the frequency. This is in agreement with the curves reported in Figure 2 for omnidirectional sensors: the first lobe's width is a decreasing function of the inter-distance.

To summarize, two different behaviors have been identified: in high frequencies (more than 1kHz), noises are uncorrelated regardless of the microphones type, whereas in low frequencies the noise components between two distant microphones are coherent. We have also verified that the speech signals observed on any pair of microphones are coherent on the full frequency bandwidth. This can be explained by the fact that they are produced by a single acoustic

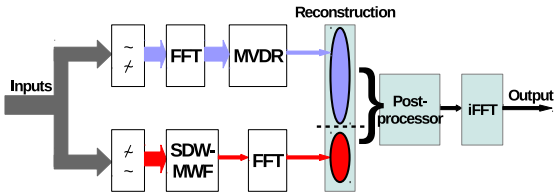

Fig. 3. Structure of the hybrid algorithm. The large arrows are for multichannel signals.

source well-located inside the car [2]. This result suggests using two different strategies depending on the frequency band and based on different types of microphones: for the lower band, omnidirectional microphones take advantage of coherence and for the upper band, unidirectional microphones take advantage of the non-correlation of noises.

2.2. Description of the complete system

The block diagram of the system is depicted in Figure 3. A cosine-modulated filterbank has been designed in order to separate the two considered frequency ranges. After the beamformer, a fullband reconstruction in the frequency domain is computed. A monochannel post-processing based on a spectral amplitude estimator [3] is then applied to the reconstructed signal. Finally, an inverse Fourier Transform is computed in order to obtain the estimated speech signal.

3. NOISE REDUCTION TECHNIQUES

As reported in Figure 4, the system consists of M sensors for a single signal of interest (SOI), usually referred as Single Input Multiple Output (SIMO) model. Under the assumption that the SOI is affected by a simple filtering propagation effect and additive noise, the discrete-time signal on the m-th sensor writes:

$$x_m(t) = h_m(t) \star s(t) + v_m(t) \quad (2)$$

where $s(t)$ denotes the discrete time SOI, $h_m(t)$ are the impulse responses of the different propagation channels between the speaker and the microphones and $v_m(t)$ are the additive noises. We also let $s_m(t) = h_m(t) \star s(t)$. In equation (2), $s(t)$ and $h_m(t)$ are unknown, it follows that $h_m(t)$ can only be identified up to a transfer function. Consequently, we choose a reference channel and label it channel 1. If we assume that the bandwidth of the SOI is fully included in the bandwidth of $h_1(t)$, we may set without loss of generality that $h_1(t) \star s(t) = s(t)$.

3.1. MVDR for high frequency range

3.1.1. System modelling

The proposed denoising process for the high frequency range is based on Short Time Fourier Transform (STFT). Consequently, the data is broken up into overlapping frames using Hamming windows. Applying the STFT on the n-th frame, the mixing model in the frequency domain writes:

$$X_{m,n}(f) = H_m(f)S_n(f) + V_{m,n}(f)^1.$$

[1]This is an approximation, assuming that the length of $h_m(t)$ is low in comparison to the frame length.

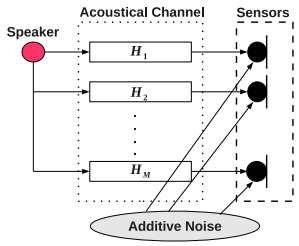

Fig. 4. Input model for MVDR beamforming

Using a compact vector notation, with respect to the M sensors, this reduces to

$$\mathbf{X}_n(f) = \mathbf{H}(f)S_n(f) + \mathbf{V}_n(f) \tag{3}$$

where bold letters stand for column vectors: for example, $\mathbf{X}_n(f) = [X_{1,n}(f) \dots X_{M,n}(f)]^T$, where the superscript ^{T} denotes the transposition. Denoting the conjugate transpose by the superscript $(^H)$, the following assumptions are made:

- $S_n(f)$ is a random circular complex Gaussian variable with zero mean and variance $\sigma_n^2(f)$.
- $\mathbf{V}_n(f)$ is a random circular complex Gaussian variable vector with zero mean and spectral covariance matrix $\Sigma_n(f) = \mathrm{E}\left[\mathbf{V}_n(f)\mathbf{V}_n^H(f)\right]$.
- $S_n(f)$ and $\mathbf{V}_n(f)$ are uncorrelated.
- $\mathrm{E}\left[S_n(f)S_n(f')\right] = 0$ for any couple $f \neq f'$.
- $\mathrm{E}\left[\mathbf{V}_n(f)\mathbf{V}_n^H(f')\right] = 0$ for any couple $f \neq f'$.

We use a Minimum Mean Squared Error Short-Time Spectral Amplitude estimator (MMSE-STSA). The MMSE-STSA multichannel estimator factorises into a monochannel spectral amplitude estimator applied to the output of a MVDR beamforming [7]. The output of the well-known MVDR beamforming is given by:

$$\mathrm{MVDR}_n^{out}(f) = \frac{\mathbf{H}^H(f)\Sigma_n^{-1}(f)\mathbf{X}_n(f)}{\mathbf{H}^H(f)\Sigma_n^{-1}(f)\mathbf{H}(f)} \tag{4}$$

The estimation of the acoustical channel transfer function $\mathbf{H}$ and the noise spectral covariance matrices Σ_n is described in the following section.

3.1.2. Parameter estimation

In this section, the adaptive estimation of the acoustical channel response $\mathbf{H}_n$ is considered. If we assume that the additive noise is incoherent, the only coherent signal between two sensors is the target speech signal. Therefore the Ratio Transfer Function (RTF) between sensors $\mathbf{H}_n$ can be estimated with respect to sensor 1.

$\mathbf{H}_n$ is estimated using a frequency-domain Block-LMS algorithm [2], as shown in Figure 5. The reference channel is delayed by Δ samples in order to overcome causality issues: the parameter Δ allows a non-causal estimation, in order to take into account delays due to sensor placement and reverberation. Its value is a compromise between the number of reflexions used in the estimation, and the global processing latency. The superscript $(^*)$ denotes the complex conjugate. The updating equation for the m-th channel response at frequency bin f is [8]:

$$H_{m,n}(f) = \tag{5}$$
$$H_{m,n-1}(f) + \mu X_{1,n}^*(f)(X_{1,n}(f) - H_{m,n-1}(f)\, X_{m,n}(f))$$

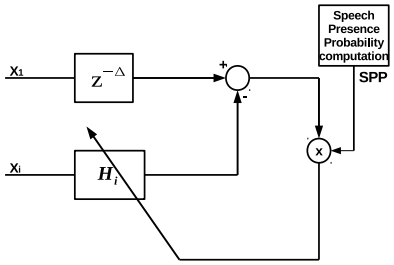

Fig. 5. Estimation of acoustical channel with a Block-LMS algorithm.

where

$$\mu = \mu_0 \frac{SPP_n(f)}{\phi_n^2(f)} \tag{6}$$

and SPP stands for the Speech Presence Probability, computed using [3], μ_0 being a constant stepsize parameter which is chosen experimentally. $\phi_n^2(f) = \mathrm{E}[|X_{1,n}(f)|^2]$ is estimated recursively using:

$$\phi_n^2(f) = \beta\phi_{n-1}^2(f) + (1-\beta)|X_{1,n}(f)|^2 \tag{7}$$

where β is a forgetting factor chosen experimentally.

The noise spectral covariance matrix $\Sigma_n(f)$ is estimated using the same SPP with the following update equation:

$$\Sigma_n(f) = \alpha_n(f)\Sigma_{n-1}(f) + (1-\alpha_n(f))\mathbf{X}_n(f)\mathbf{X}_n^H(f) \tag{8}$$

$$\alpha_n(f) = \alpha_0 + (1-\alpha_0)SPP_n(f) \tag{9}$$

where α_0 is a forgetting factor chosen experimentally.

3.2. SDW-MWF for the low frequency range

3.2.1. Signal Model

For the low frequency range, a reference sensor k (here, $k = 1$) is used to estimate a delayed version of the noise. The estimation of the target speech signal is then performed by subtracting this noise reference:

$$\widehat{s}_k(t-\Delta) = x_k(t-\Delta) - \mathbf{W}_k^T\mathbf{x}(t).$$

where $\mathbf{x}(t)$ is defined by the following stacked vector notation:

- $\mathbf{x}_m(t)$ is the vector $[x_m(t-L+1) \quad \dots \quad x_m(t)]^T$
- $\mathbf{x}(t) = [\mathbf{x}_1(t)^T \quad \mathbf{x}_2(t)^T \quad \dots \quad \mathbf{x}_M(t)^T]^T$

and $\mathbf{W}_k$ is a column vector implementing a spatio-temporal filter of length $L \times M$. The parameter Δ is chosen as in 3.1.2.

3.2.2. Noise estimation

The filter $\mathbf{W}_k$ can be written as the solution minimizing the prediction error:

$$\widehat{\mathbf{W}_k} = \min_{\mathbf{w}} \mathrm{E}\left[|v_k(t-\Delta) - \mathbf{w}^T\mathbf{x}(t)|^2\right]. \tag{10}$$

In the following, $\mathbf{s}_m(t)$, $\mathbf{s}(t)$, $\mathbf{v}_m(t)$ and $\mathbf{v}(t)$ are defined in the same way as $\mathbf{x}_m(t)$ and $\mathbf{x}(t)$. The target signal and the noise are assumed to be independent, which leads to the decomposition of the cost function into two terms:

$$\mathrm{E}\left[|v_k(t-\Delta) - \mathbf{w}^T\mathbf{x}(t)|^2\right] = \tag{11}$$
$$\underbrace{\mathrm{E}\left[|\mathbf{w}^T\mathbf{s}(t)|^2\right]}_{e_s^2} + \underbrace{\mathrm{E}\left[|v_k(t-\Delta) - \mathbf{w}^T\mathbf{v}(t)|^2\right]}_{e_b^2}$$

where e_s^2 is the speech distortion and e_b^2 is the residual noise. A more general criterion $J_\gamma(\mathbf{w})$ can be obtained by including a weighting factor $\gamma > 0$ on the residual noise [1]:

$$J_\gamma(\mathbf{w}) = \mathrm{E}\big[|\mathbf{w}^T\mathbf{s}(t)|^2\big] + \gamma\mathrm{E}\big[|v_k(t-\Delta) - \mathbf{w}^T\mathbf{v}(t)|^2\big]. \quad (12)$$

The solution minimizing (12) is given by

$$\widehat{\mathbf{W}_k} = \Big[\frac{1}{\gamma}\mathrm{E}\big[\mathbf{s}(t)\mathbf{s}(t)^H\big] + \mathrm{E}\big[\mathbf{v}(t)\mathbf{v}(t)^H\big]\Big]^{-1}\mathrm{E}\big[\mathbf{v}(t)v_k(t-\Delta)\big]$$

and is known as the Speech Distortion Weighted Multichannel Wiener Filter (SDW-MWF)[1].

3.2.3. Adaptive Implementation

The SDW-MWF is implemented in a stochastic gradient algorithm [1]. Since the cost function is given by equation (12), its instantaneous gradient estimate is given by :

$$\frac{\delta J_\gamma}{\delta \mathbf{w}}(t) = 2\big[\mathbf{R}_s(t) + \gamma\mathbf{R}_v(t)\big]\mathbf{w} - 2\gamma E\big[\mathbf{C}_v(t)\big] \quad (13)$$

where $\mathbf{R}_s(t) = \mathrm{E}\big[\mathbf{s}(t)\mathbf{s}(t)^H\big]$, $\mathbf{R}_v(t) = \mathrm{E}\big[\mathbf{v}(t)\mathbf{v}(t)^H\big]$ and $\mathbf{C}_v(t) = \mathrm{E}\big[\mathbf{v}(t)v_k^*(t-\Delta)\big]$.

It follows that the update equation for $\mathbf{W}_k(t)$ is:

$$\mathbf{W}_k(t) = \mathbf{W}_k(t-1) - \rho\Big(\big[\mathbf{R}_s(t) + \gamma\mathbf{R}_v(t)\big]\mathbf{W}_k(t-1) - \gamma\mathbf{C}_v(t)\Big)$$

where $\rho > 0$ is the algorithm stepsize, and is normalized as $\rho = \frac{\rho_0}{\mathbf{x}(t)^H\mathbf{x}(t)}$.

Figure 6 shows the generalized scheme for SDW-MWF.

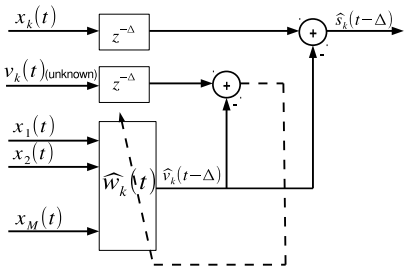

Fig. 6. SDW-MWF for the low frequency range

In order to estimate the matrix $\mathbf{R}_v(t)$, a Voice Activity Detector (VAD) is used to detect whether the target speech is present or not at time t:

$$\mathbf{R}_v(t) = \quad (14)$$
$$\begin{cases} \lambda\mathbf{R}_v(t-1) + (1-\lambda)\mathbf{x}(t)\mathbf{x}(t)^H & \text{when no speech is detected} \\ \mathbf{R}_v(t-1) & \text{otherwise} \end{cases}$$

where $\lambda \in]0\ 1[$ is a forgetting factor. The vector $\mathbf{C}_v(t)$ is estimated the same way, since it is the column indexed $(k-1)L+\Delta$ of matrix $\mathbf{R}_v(t)$ (if $\Delta < L$).

Because the target speech and the noise are uncorrelated, $\mathbf{R}_s(t)$ can be estimated as:

$$\mathbf{R}_s(t) = \mathbf{R}_x(t) - \mathbf{R}_v(t). \quad (15)$$

where $\mathbf{R}_x(t) = E\big[\mathbf{x}(t)\mathbf{x}(t)^H\big]$ is estimated the same way as $\mathbf{R}_v(t)$, with no condition on speech presence:

$$\mathbf{R}_x(t) = \lambda\mathbf{R}_x(t-1) + (1-\lambda)\mathbf{x}(t)\mathbf{x}(t)^H.$$

4. EXPERIMENTS AND RESULTS

4.1. Validation of the approach on synthetic signals

To validate the method, both algorithms are applied at first on synthetic data as follows. Two omnidirectional microphones are used to record separately the speech signal in a silent (not noisy) environment, at a sample rate of 8kHz. Synthesized noise signals, with input SNRs in $\mathcal{S} = [-5, 0, 5, 10, 15]$ dB, were added to each channel considering the two extreme cases: (i) the coherent case, for which the two noises had a MSC of 1 in the full frequency bandwidth and (ii) the non-coherent case, for which the two noises had a MSC of 0. For evaluation, the output segmental SNR is given by:

$$\mathrm{SNR}_{seg} = \frac{10}{M}\sum_{m=0}^{M-1}\log_{10}\frac{\sum_{t=Nm}^{Nm+N-1}s^2(t)}{\sum_{t=Nm}^{Nm+N-1}(s(t)-\hat{s}(t))^2} \quad (16)$$

where M is the number of frames, corresponding to approximately 6s of signal, N the size of the window, corresponding to 128 ms, s the SOI (here, the clean speech on input 1) and $\hat{s}$ the estimated output. As usual in this context, SNRs outside the range -10 dB to 30 dB have been omitted [10].

The average value, over $\mathcal{S}$, of the output SNR_{seg} are reported in Table 1 for both algorithms acting on the two types of noise. In agreement with the theoretical development, the MVDR performs slightly better for the non-coherent case and the SDW-MWF better for the coherent one. Therefore, we may expect the proposed hybrid strategy to bring some improvement in an automotive environment.

Algorithm	Coherent case	Non-coherent case
MVDR	0.27	**0.43**
SDW-MWF	**0.99**	0.25

Table 1. Averaged Segmental SNR in dB for the two algorithms and the two noise cases

4.2. Implementation

In order to compute the SDW-MWF, we used the VAD described in [11] (see equation (14)). For frequency-domain processing, frames are 256 sample-long with 50% overlap and Hamming windowing. Since the lower frequency band (0 to 1 kHz) has to be separated for the hybrid algorithm (see section 2), we used a 16-th order cosine-modulated filterbank.

The inputs were generated using recorded speech on two unidirectionnal microphones in a silent car at a sampling frequency of 8kHz. Real in-car noise, recorded using the same setup with the car moving in a normal freeway situation, was then added to the clean speech in order to obtain realistic in-car noisy speech signals. The distance between the two microphones is 4 cm (1.57 inches), so that the inter-microphones noise coherence is the same as in Figure 2.

The parameters used for the experiment are reported in Table 2.

MVDR		SDW-MWF		Both
$\alpha_0 = 0.75$	$\beta = 0.8$	$\gamma = 2.3$	$\rho_0 = 0.1$	$\Delta = 16$
$\mu_0 = 0.5$		$\lambda = 0.95$	$L = 64$	

Table 2. Parameters used for experiments

4.3. Results

We use 3 different multichannel algorithms for results comparison:

MVDR An MVDR beamforming (as described in section 3.1) followed by an Optimally Modified-Log Spectral Amplitude Estimator (OM-LSA)[3] post-processing.

SDW-MWF An SDW-MWF algorithm (as described in section 3.2) followed by an OM-LSA post-processing.

Combined A low subband SDW-MWF algorithm and an MVDR beamforming in the high subband, followed by an OM-LSA post-processing.

The performance criteria are the output segmental SNR (see equation (16)), and the Perceptual Evaluation of Speech Quality (PESQ, ITU-T P.862), measured on a scale from 0 to 5 (to be compared on a Mean-Opinion Score (MOS) scale). These criteria are measured on a range of input SNR from -5 to 15 dB (which corresponds to an input Segmental SNR of -8 to 1 dB). The results are reported in Figure 7. These results show a better performance in all conditions

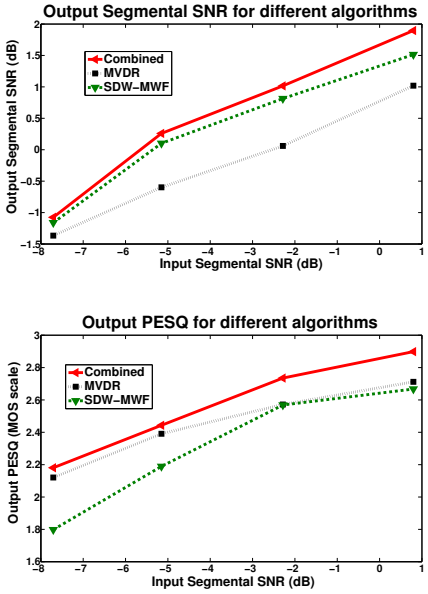

Fig. 7. Results for different input SNRs

for the combined algorithm. It takes advantages of the different noise conditions in the high and low frequency ranges to perform better than the fullband non-hybrid strategies.

5. CONCLUSION

In this paper, we presented a hybrid multichannel speech enhancement method for in-car environments. This choice was motivated by an experimental study of the noise field properties, specifically the inter-microphone noise coherence. The results on real in-car audio data are promising, when using objective performance evaluation methods. Since the latter do not evaluate perceived speech intelligibility, a subjective evaluation will be developed in future work.

6. REFERENCES

[1] A. Spriet, M. Moonen, and J. Wouters, "Stochastic gradient-based implementation of spatially preprocessed speech distortion weighted multichannel Wiener filtering for noise reduction in hearing aids," *IEEE Transactions on Signal Processing*, vol. 53, pp. 911–925, Mar. 2005.

[2] J. Freudenberger, S. Stenzel, and B. Venditti, "An FLMS based two-microphone speech enhancement system for in-car applications," in *Statistical Signal Processing, IEEE/SPS 15th Workshop on*, Sept. 2009, pp. 705 –708.

[3] I. Cohen, "Optimal speech enhancement under signal presence uncertainty using log-spectral amplitude estimator," *Signal Processing Letters, IEEE*, vol. 9, no. 4, pp. 113 –116, Apr. 2002.

[4] Y. Ephraim and D. Malah, "Speech enhancement using a minimum-mean square error short-time spectral amplitude estimator," *Acoustics, Speech and Signal Processing, IEEE Transactions on*, vol. 32, no. 6, pp. 1109 – 1121, Dec. 1984.

[5] Chang Huai You, Soo Ngee Koh, and S. Rahardja, "Adaptive beta-order MMSE speech enhancement application for mobile communication in a car environment," in *Information, Communications and Signal Processing, Proceedings of the Fourth International Conference on*, Dec. 2003, vol. 3, pp. 1629 – 1632.

[6] J. Meyer and K.U. Simmer, "Multi-channel speech enhancement in a car environment using Wiener filtering and spectral subtraction," in *Acoustics, Speech, and Signal Processing, IEEE International Conference on*, Apr. 1997, vol. 2, pp. 1167 –1170.

[7] R.C. Hendriks, R. Heusdens, U. Kjems, and J. Jensen, "On optimal multichannel mean-squared error estimators for speech enhancement," *Signal Processing Letters, IEEE*, vol. 16, no. 10, pp. 885 –888, Oct. 2009.

[8] Min-Seok Choi, Chang-Hyun Baik, Young-Cheol Park, and Hong-Goo Kang, "A soft-decision adaptation mode controller for an efficient frequency-domain generalized sidelobe canceller," in *Acoustics, Speech and Signal Processing, 2007. ICASSP 2007. IEEE International Conference on*, Apr. 2007, vol. 4, pp. 893–896.

[9] B. Widrow, Jr. Glover, J.R., J.M. McCool, J. Kaunitz, C.S. Williams, R.H. Hearn, J.R. Zeidler, Jr. Eugene Dong, and R.C. Goodlin, "Adaptive noise cancelling: Principles and applications," *Proceedings of the IEEE*, vol. 63, no. 12, pp. 1692 – 1716, Dec. 1975.

[10] John H. L. Hansen and Bryan L. Pellom, "An effective quality evaluation protocol for speech enhancement algorithms," in *Proceedings of the International Conference on Speech and Language Processing*, 1998, pp. 2819–2822.

[11] Jongseo Sohn and Wonyong Sung, "A voice activity detector employing soft decision based noise spectrum adaptation," in *Acoustics, Speech and Signal Processing, 1998. Proceedings of the 1998 IEEE International Conference on*, May 1998, vol. 1, pp. 365 –368.