MULTI-CHANNEL ADAPTIVE BEAMFORMING WITH SOURCE SPECTRAL AND NOISE COVARIANCE MATRIX ESTIMATIONS

Hai Quang Dam, Sven Nordholm, Hai Huyen Dam and Siow Yong Low

Western Australian Telecommunications Research Institute (WATRI) *,
The University of Western Australia, Crawley, WA 6009, Australia

ABSTRACT

This paper introduces a subband adaptive beamformer equipped with a source spectral estimation and an updating scheme for the noise covariance matrix. This scheme is employed to effectively estimate and track the noise statistics such that it is continuously evaluated in the solution. More specifically, the noise covariance matrix is updated, depending on the instantaneous source spectral estimation. Experimental result shows that the proposed method achieves a high suppression level and low source distortion on real car data.

1. INTRODUCTION

During the past decade, many research efforts have been devoted to the speech enhancement techniques, especially for hands-free communication, teleconferencing and speech recognition. Those applications have been developed to bring the human communications to a higher level of comfort. Generally speaking, speech enhancement techniques can be classified into single-channel and multi-channel beamforming techniques [1]. Single-channel techniques only exploit spectral and temporal differences between the speech and the noise signals to suppress the noise. Multi-channel beamforming techniques employing microphone array with certain geometric structure to exploit spatial diversity in addition to spectral and temporal information of the desired and undesired signals. Various beamforming techniques have been proposed in the literature for speech enhancement applications [2, 3]. Among them, the generalized sidelobe canceller (GSC) [4] has attracted considerable interest and references therein. The GSC offers good interference suppression but succumbs to target signal cancellation due to the presence of steering vector errors and reverberation effects.

Recently, soft constrained adaptive beamforming has been suggested for microphone arrays [5]. In this approach, the source covariance matrix is assumed to be known from a model or calibration phase. The beamforming weights are obtained by minimizing the mean square error (MSE), under a quadratic constraint. This work was further extended in [6, 7] to allow for constraint adaptation according to the current target spectrum. The main idea is to employ the target source power spectral density (PSD) updates to extract the target signal. The PSD updates are then incorporated in the MSE solution for each subband to fully utilize the time-frequency information of the target signal. By doing so, the noise covariance matrix needs to be estimated before or during the beamforming process. In [7], a noise-only detection was proposed to estimate the noise-only covariance matrix. Like other voice or noise detections, this detection requires a large computational burden. To overcome that, this paper proposes a new update method based on the spectral information of the source to efficiently estimate the noise statistics. By fully utilizing spectral information, good noise suppression can be achieved with minimal distortion on the source of interest. Evaluations in a real car hands-free environment using a six element microphone array reveal that the proposed method achieves significant noise suppression level up to 18.7 dB whilst maintaining excellent timbre of the target signal.

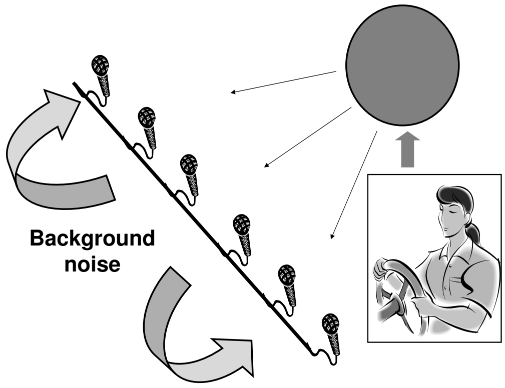

Fig. 1. *Configuration of the linear microphone array and the source of interest.*

*WATRI is a joint venture between The University of Western Australia and Curtin University of Technology. The work has also been sponsored by ARC under grant no. DP0451111.

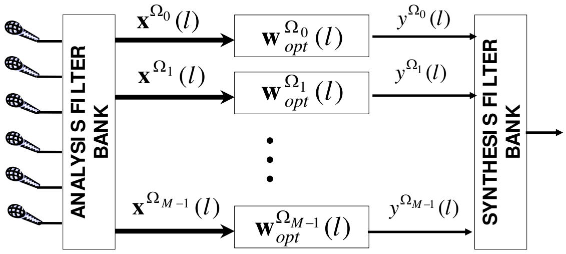
ANALYSIS FILTER BANK
$\mathbf{x}^{\Omega_0}(l)$
$\mathbf{x}^{\Omega_1}(l)$
$\mathbf{x}^{\Omega_{M-1}}(l)$
$\mathbf{w}_{opt}^{\Omega_0}(l)$
$\mathbf{w}_{opt}^{\Omega_1}(l)$
$\mathbf{w}_{opt}^{\Omega_{M-1}}(l)$
$y^{\Omega_0}(l)$
$y^{\Omega_1}(l)$
$y^{\Omega_{M-1}}(l)$
SYNTHESIS FILTER BANK

Fig. 2. *Subband beamformer with the analysis and synthesis filter banks.*

2. PROBLEM FORMULATION

Consider a linear microphone array with I microphones as depicted in Fig. 1. The received signal consists of the target signal and the background noise, the target signal in this case is a person speaking. This signal is decomposed into M subbands by using an oversampled analysis filter bank. Finally, the synthesis filter bank reconstructs all the processed subband signals into fullband representation.

A block diagram of the subband beamformer is given in Fig. 2. For each frequency $\Omega \in [\Omega_0, \cdots, \Omega_{M-1}]$, let us denote $\mathbf{x}^{\Omega}(k)$ as the received signal from the *I* microphones, where *k* is the time index. This signal can be written as

$$\mathbf{x}^{\Omega}(k) = \mathbf{x}_s^{\Omega}(k) + \mathbf{x}_v^{\Omega}(k), \tag{1}$$

where $\mathbf{x}_s^{\Omega}(k)$ and $\mathbf{x}_v^{\Omega}(k)$ are the contribution from the original source and the noise, respectively. By assuming that the source and the noise are statistically independent with the covariance matrices, $\mathbf{R}_s^{\Omega}$ and $\mathbf{R}_v^{\Omega}$ respectively, the covariance matrix for the received signal $\mathbf{R}^{\Omega}$ can be given as

$$\mathbf{R}^{\Omega} = \mathbf{R}_s^{\Omega} + \mathbf{R}_v^{\Omega}. \tag{2}$$

By defining p_s^{Ω} as the power spectral amplitude of the source for each frequency Ω, equation (2) can be rewritten as

$$\mathbf{R}^{\Omega} = p_s^{\Omega}\bar{\mathbf{R}}_s^{\Omega} + \mathbf{R}_v^{\Omega}, \tag{3}$$

where $\bar{\mathbf{R}}_s^{\Omega}$ is the spatial covariance matrix for the source. This matrix relies only on the position of the original source and can be calculated from training snapshots during a calibration process prior to the beamforming period or by introducing a constrained model [5].

3. ADAPTIVE ALGORITHM WITH THE NOISE COVARIANCE MATRIX UPDATED

In the following adaptive algorithm, the covariance matrix for the received signal $\mathbf{R}^{\Omega}$ is estimated at each time instant l as follows

$$\mathbf{R}^{\Omega}(l) = \frac{1}{N}\sum_{k=l-N+1}^{l} \mathbf{x}^{\Omega}(k)\mathbf{x}^{\Omega}(k)^{H}, \tag{4}$$

where $(\cdot)^H$ denotes the complex conjugate and N is sample size for estimating the covariance matrix. Thus at each time instant l, (3) can be given as follows:

$$\mathbf{R}^{\Omega}(l) = p_s^{\Omega}(l)\bar{\mathbf{R}}_s^{\Omega} + \mathbf{R}_v^{\Omega}, \tag{5}$$

where $p_s^{\Omega}(l)$ is the power spectral amplitude of the source at time instant l. During the beamforming process, the matrix $\mathbf{R}_v^{\Omega}$ may vary depending on the noise property. Thus, a fixed noise covariance matrix can cause high error in the beamforming result. To cope with that, a noise covariance matrix update is proposed in the Subsection 3.2 for re-estimating the covariance matrix $\mathbf{R}_v^{\Omega}$ after certain time period.

The objective here is to efficiently estimate $p_s^{\Omega}(l)$ while the noise covariance matrix is periodically estimated.

3.1. Source power spectral estimation

Equation (5) can be written as

$$\mathbf{R}^{\Omega}(l) - p_s^{\Omega}(l)\bar{\mathbf{R}}_s^{\Omega} - \mathbf{R}_v^{\Omega} = 0. \tag{6}$$

Here, $\mathbf{R}^{\Omega}(l)$ is readily obtained from (4) while $\bar{\mathbf{R}}_s^{\Omega}$ is precalculated before the beamforming process and $\mathbf{R}_v^{\Omega}$ is evaluated by the scheme proposed in the subsection 3.2. Obviously, equation (6) might not have the exact solution due to the high variance of the covariance matrix evaluated from a restricted number of samples. Thus, the following estimation for $p_s^{\Omega}(l)$ is proposed. The power spectral amplitude, $p_s^{\Omega}(l)$ is estimated as a non-negative solution which optimizes the following problem

$$p_s^{\Omega}(l) = \arg\min_{p_s^{\Omega}\geq 0} \| \mathbf{R}^{\Omega}(l) - p_s^{\Omega}\bar{\mathbf{R}}_s^{\Omega} - \mathbf{R}_v^{\Omega} \|_{\mathcal{F}}^2, \tag{7}$$

where $\| \cdot \|_{\mathcal{F}}$ is the Frobenius norm. The source power spectral amplitude is estimated at every iteration to provide a spectrally optimized constraint on the source. In simple terms, it attempts to preserve the spectra of the source like a spectra mouldier.

3.2. Noise covariance matrix estimation

Based on the assumption that the noise is long-term stationary [7], the noise covariance matrix $\mathbf{R}_v^{\Omega}$ is not required to be estimated at each time instant. Thus, we evaluate $\mathbf{R}_v^{\Omega}$ at every L time instants to reduce the computational burden. So, in the beginning the matrix $\mathbf{R}_v^{\Omega}(n)$, i.e., $n = 0$, is initialized to a zero matrix and subsequently updated at each time instant nL $(n = 1, 2, ...)$.

The matrix $\mathbf{R}_v^\Omega(n)$ is being used in place of $\mathbf{R}_v^\Omega$ during the period $[nL+1,(n+1)L]$ for the optimization problem (7).

To perform the estimation of the noise covariance matrix, a weight value $\eta(l)$ is proposed to be calculated at each time instant l based on the source power spectral estimation.

$$\eta(l) = \begin{cases} e^{-\alpha p_s^\Omega(l)}, & \text{if} \quad e^{-\alpha p_s^\Omega(l)} > \beta \\ \beta, & \text{otherwise} \end{cases} \tag{8}$$

where $\alpha > 0$ is the adjusting parameter and $0 < \beta < 1$ is the lower bound for the weight value. Since $p_s^\Omega(l)$ is positive, we have $\beta \leq \eta(l) \leq 1$ for all l. The weight value $\eta(l)$ presents the noise information of the matrix $\mathbf{R}^\Omega(l)$ and the exponential function serves as a regulator of how much the noise information is updated. If noise dominates at frequency Ω then the error from the previous estimation of noise covariance matrix may cause high error in the source power spectral estimation. Thus, the lower bound β is proposed for reserving the update portion in this case.

The noise covariance matrix $\mathbf{R}_v^\Omega(n)$ is evaluated as follows

$$\mathbf{R}_v^\Omega(n) = \frac{E[\eta(l)\mathbf{R}^\Omega(l)]}{E[\eta(l)]} \quad \text{for} \quad l \in [(n-1)L+1, nL], \tag{9}$$

where $E[\,\cdot\,]$ is the expectation operator. Since L is a restricted number, equation (9) can be approximated as

$$\mathbf{R}_v^\Omega(n) = \frac{\sum\limits_{l=(n-1)L+1}^{nL} \eta(l)\mathbf{R}^\Omega(l)}{\sum\limits_{l=(n-1)L+1}^{nL} \eta(l)}. \tag{10}$$

3.3. Wiener solution

Adaptive beamforming is employed for each frequency Ω after the received signals passed through the analysis filter bank. At each time instant l, the Wiener solution can be calculated as follows

$$\mathbf{w}_{opt}^\Omega(l) = \left[p_s^\Omega(l)\bar{\mathbf{R}}_s^\Omega + \mathbf{R}_v^\Omega(\,\mathsf{floor}(\frac{l}{L})\,)\right]^{-1} \mathbf{d}_s p_s^\Omega(l), \tag{11}$$

where the operator $\mathsf{floor}(a)$ rounds the value a to the nearest integer towards minus infinity. The vector $\mathbf{d}_s$ is the spatial cross covariance vector for the source of interest and is calculated as the normalized covariance between a reference microphone and all microphones of the array.

Since the matrix $\mathbf{R}_v^\Omega(\,\mathsf{floor}(\frac{l}{L})\,)$ is unchanged during the period $l \in [(n-1)L+1, nL]$ $(n = 1, 2, ...)$, low rank updates can be used by the inverse matrix lemma [1] for ease of computation. The output of beamformer at frequency Ω is calculated as follows

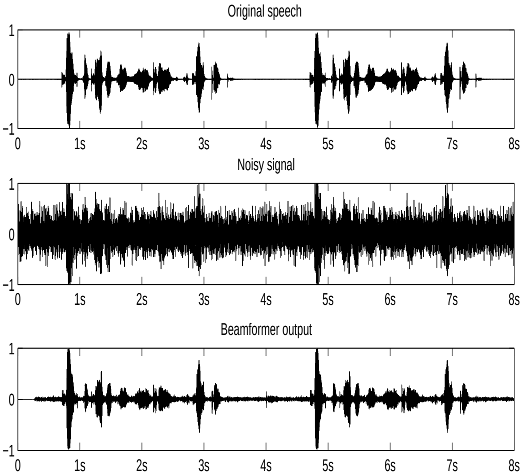

Fig. 3. *Plots of the (a) source, (c) received signal and (d) beamformer output with $M = 64$ subbands.*

$$y^\Omega(l) = \mathbf{w}_{opt}^\Omega(l)^H \mathbf{x}^\Omega(l). \tag{12}$$

4. SIMULATION

The performance evaluation of the beamformer was made in a car hands-free situation where a six sensor microphone array was mounted on the visor at the passenger side in a Volvo station wagon with one hands-free loudspeaker to simulate a real hands-free situation. Data was gathered on a multichannel DAT-recorder with a sampling rate of 12 kHz and a 300-3400 Hz bandwidth. The car was running at the speed of 110 km/h on a paved road.

A uniform over-sampled DFT filterbank is used to decompose the received array signals into M subband signals [8]. The over-sampled filterbank is designed to be compliant with minimal transformation and reconstruction aliasing effects. The spatial source covariance matrix has been estimated from training signals during calibration process prior the beamforming period. The noise covariance matrix is initialized to a zero matrix at the beginning. The values α, β, N and L are set to 1, 0.02, 40 and 100, respectively.

The performance of the beamformer is measured by using the noise suppression and the speech distortion. The noise

M	Noise suppression (dB)	Source distortion (dB)
16	15.9	-26.6
32	18.4	-26.7
64	18.7	-29.9
128	17.5	-30.9

Table 1. *Noise suppression levels and source distortion measures for different number of subbands.*

suppression level is defined as,

$$SP = 10\log_{10}\left(\frac{\int_{-\pi}^{\pi}\hat{P}_{in,n}(\omega)d\omega}{\int_{-\pi}^{\pi}\hat{P}_{out,n}(\omega)d\omega}\right) - 10\log_{10}(C_d) \tag{13}$$

where $\hat{P}_{in,n}(\omega)$ and $\hat{P}_{out,n}(\omega)$ are the spectral power estimates of the reference sensor observation and the output respectively, when the noise is active alone. The constant C_d normalizes to the source's gain. The source distortion measure is given as

$$DS = \\ 10\log_{10}\left(\frac{1}{2\pi}\int_{-\pi}^{\pi}|(1/C_d)\hat{P}_{in,s}(\omega) - \hat{P}_{out,s}(\omega)|d\omega\right) \tag{14}$$

where $\hat{P}_{in,s}(\omega)$ and $\hat{P}_{out,s}(\omega)$ are the spectral power estimates of the reference sensor observation and the output respectively, when the source is active alone.

Table 1 shows the noise suppression levels and the source distortion measures for the number of subbands increases from 16 to 128. Fig. 3 shows the time domain plots for the source, the noisy signal and the beamformer output and Fig. 4 shows the PSDs of the source and the noise before and after the beamformer in the case of $M = 64$ subbands.

Results show that the beamformer can achieve a high noise suppression level up to 18.7 dB and a low source distortion -29.9 dB in the case of $M = 64$ subbands. The low source distortion levels especially in the cases 64 and 128 subbands show the effectiveness of the proposed noise updating scheme.

5. CONCLUSIONS

In this paper, a new beamformer has been presented to enhance a speech source in a noisy environment. The beamformer includes the estimation process for the noise covariance matrix without implementing any kind of detection. This scheme is suitable for the case where the noise is slowly varying. The simulation results in a car environment show a high noise suppression level is achieved whilst maintaining low source distortion.

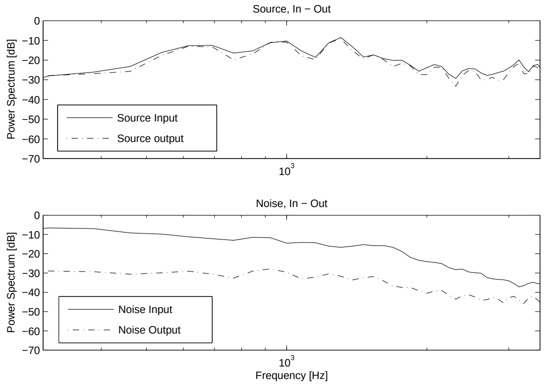

Fig. 4. *PSDs of the source and the noise before and after the beamformer with $M = 64$ subbands.*

6. REFERENCES

[1] S. Haykin, *Adaptive Filter Theory*, Prentice Hall, Upper Saddle River, New Jersey, 4th edition, 2002.

[2] M. Brandstein and D. Ward (Eds), *Microphone Arrays: Techniques and Applications*, Springer Verlag, 2001.

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

[4] L. J. Griffiths and C. W. Jim, "An alternative approach to linearly constrained adaptive beamforming," *IEEE Transactions on Antennas and Propagation*, vol. 30, no. 1, pp. 27–34, January 1982.

[5] N. Grbić and S. Nordholm, "Soft constrained subband beamforming for handsfree speech enhancement," *IEEE International Conference on Acoustics, Speech, and Signal Processing*, vol. 1, pp. 885–888, 2002.

[6] H. Q. Dam, S. Nordholm, N. Grbić, and H. H. Dam, "Speech enhancement employing adaptive beamformer with recursively updated soft constraints," *International Workshop on Acoustic Echo and Noise Control*, pp. 307–310, September 2003.

[7] H. Q. Dam, S. Y. Low, S. Nordholm, and H. H. Dam, "Adaptive microphone array with noise statistics updates," *IEEE International Symposium on Circuits and Systems*, vol. 3, pp. 433–436, May 2004.

[8] P. P. Vaidyanathan, *Multirate Systems And Filter Banks*, Prentice Hall, 1993.