SPACE CONSTRAINED BEAMFORMING WITH SOURCE PSD UPDATES

Hai Quang Dam *Siow Yong Low* *Hai Huyen Dam* *Sven Nordholm*

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

ABSTRACT

This paper presents a new space constrained adaptive beamformer employing an updated source power spectral density (PSD). The space constraints are used to capture the target signal spatially and to provide robustness against steering error vectors. The PSD update on the other hand ensures that the spectral information of the desired source is reflected continuously on the space constraints. As such, target signal extraction can be achieved with minimum distortion. The beamformer operates in a subband structure to allow time-frequency operation for each channel, yielding a combination of weighted spatial and temporal filters. Evaluations on real car data show that the proposed algorithm significantly improves the speech intelligibility with noise suppression level up to 21 dB.

1. INTRODUCTION

The explosive growth of hands-free communication systems has spurred a very intensive research and development in the area of speech enhancement. Countless enhancement methods have been proposed throughout the years, with beamforming based techniques form the most promising choice [1, 2]. This is due to the fact that beamforming exploits not only temporal information but also spatial diversity [3, 4]. In other words, adaptive beamformers can achieve notable interference suppression by using the fact that the origins of the desired and interfering signals originate from different locations in space. The drawback of these beamformers is that they introduce target signal cancellation/distortion such as in a generalized sidelobe canceller (GSC) [1]. This is due to the vulnerability of the model based adaptive beamformer with respect to the steering vector errors.

This paper proposes a novel space constrained subband adaptive beamformer employing a source power spectral density (PSD) update. The space constraints guard against steering vector errors by taking into account of an area of point sources rather than a point in space. Essentially, the beamformer is constrained to extract the target signal in a pre-defined area. The estimated source PSD update, on the other hand, is embedded in the optimum Wiener solution in each subband to fully utilize the time-frequency information of the target signal. Here, the source PSD is updated using a least squares criterion and acts as a source spectral moulder. In other words, it tracks the variation in the spectral content of the target signal continuously, yielding a spectrally optimized constraint for each time instant. The integration of both the space constraints and the PSD in its formulation results in an efficient spatio-temporal beamformer. Therefore, a maximally background noise suppression is achieved whilst maintaining excellent target signal integrity. Evaluations in a real car hands-free using a six-element array reveal an impressive noise suppression level of up to 21 dB whilst maintaining excellent timbre of the target signal.

2. THE PROPOSED BEAMFORMER

The block diagram of the proposed beamformer is shown in Fig. 1. Initially, the received signal is decomposed into M subbands with a decimation factor D by using an analysis filter bank. Each subband is then processed independently and finally the synthesis filter bank reconstructs all the subband signals into fullband representation.

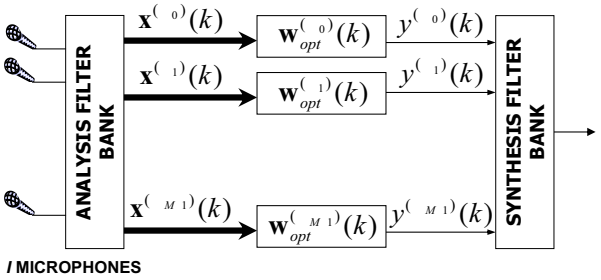

Fig. 1. *The proposed subband beamformer with the analysis and synthesis filter banks.*

*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. A00105530.

2.1. The Source Space Constraints

Consider a linear microphone array with I microphones. The target signal in this case is a person speaking, which can be modelled as an infinite number of point sources clustered closely in space. This space is modelled as a circular area $\mathbf{A}$ with radius r and a distance h from the array, see Fig. 2. The advantage of having the source constrained region as opposed to a point source is consistent with the fact that errors in the response vector cause large radial vectors in the corresponding source location [1, 5]. These errors are typically due to sensor misplacement and gain variations in the microphones.

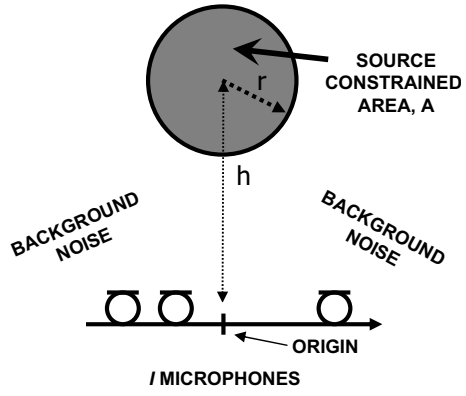

Fig. 2. *Configuration of the linear microphone array and the source constrained area.*

Let us denote $S^{(\Omega)}$ as the PSD of the source in the pre-defined area $\mathbf{A}$ at frequency Ω. The spatio-temporal covariance matrix of the source in the spectral band $[\Omega_a, \Omega_b]$ is given as

$$\mathbf{R}_s = \int_{\Omega_a}^{\Omega_b} \iint_{\mathbf{A}} S^{(\Omega)} \mathbf{d}^{(\Omega)}(\overrightarrow{\mathbf{a}}) \mathbf{d}^{(\Omega)}(\overrightarrow{\mathbf{a}})^H d\overrightarrow{\mathbf{a}} d\Omega, \quad (1)$$

where $\overrightarrow{\mathbf{a}}$ is a point source localisation vector and $(\cdot)^H$ denotes the Hermitian transposition. The response vector $\mathbf{d}^{(\Omega)}(\overrightarrow{\mathbf{a}})$ is defined as

$$\mathbf{d}^{(\Omega)}(\overrightarrow{\mathbf{a}}) = \left[\frac{1}{R_1} e^{-j\Omega\tau_1(\overrightarrow{\mathbf{a}})}, \ldots, \frac{1}{R_I} e^{-j\Omega\tau_I(\overrightarrow{\mathbf{a}})} \right]^T, \quad (2)$$

where $\tau_i(\overrightarrow{\mathbf{a}})$ is the time delay from a point source in the pre-defined area to sensor i, and R_i is the distance between the source and sensor i and $[\cdot]^T$ denotes the transposition operator. The reference point for the beamformer response is defined at the origin of the coordinates.

For a frequency Ω, we define the spatial source covariance matrix as,

$$\mathbf{R}_s^{(\Omega)} = S^{(\Omega)} \bar{\mathbf{R}}_s^{(\Omega)} \quad (3)$$

where

$$\bar{\mathbf{R}}_s^{(\Omega)} = \iint_{\mathbf{A}} \mathbf{d}^{(\Omega)}(\overrightarrow{\mathbf{a}}) \mathbf{d}^{(\Omega)}(\overrightarrow{\mathbf{a}})^H d\overrightarrow{\mathbf{a}}. \quad (4)$$

The spatial source cross covariance vector on the other hand is given by $\mathbf{r}_s^{(\Omega)} = S^{(\Omega)} \bar{\mathbf{r}}_s^{(\Omega)}$ where

$$\bar{\mathbf{r}}_s^{(\Omega)} = \iint_{\mathbf{A}} \mathbf{d}^{(\Omega)}(\overrightarrow{\mathbf{a}}) d\overrightarrow{\mathbf{a}}. \quad (5)$$

2.2. The Source PSD Estimation

Assuming that the source and the background noise are uncorrelated, the received signal covariance matrix $\mathbf{R}^{(\Omega)}$, can be decomposed as

$$\mathbf{R}^{(\Omega)} = \mathbf{R}_s^{(\Omega)} + \mathbf{R}_n^{(\Omega)}, \quad (6)$$

where $\mathbf{R}_n^{(\Omega)}$ denotes the noise covariance matrix. The covariance matrix in (6) can be estimated from K samples of received data with index $\mathcal{K}$ during source silence period when the noise is active alone as,

$$\mathbf{R}_n^{(\Omega)} = \frac{1}{K} \sum_{k \in \mathcal{K}} \mathbf{x}_n^{(\Omega)}(k) \mathbf{x}_n^{(\Omega)}(k)^H, \quad (7)$$

where $\mathbf{x}_n^{(\Omega)}(k)$ is the noise only received data. The noise only period is typically identified using a voice activity detector (VAD). Alternatively, the noise statistics in (7) can be estimated using a calibration process such as in [6].

The spatial source covariance matrix can be written in terms of $\mathbf{R}^{(\Omega)}$ and $\mathbf{R}_n^{(\Omega)}$ by substituting (3) into (6) as,

$$S^{(\Omega)} \bar{\mathbf{R}}_s^{(\Omega)} = \mathbf{R}^{(\Omega)} - \mathbf{R}_n^{(\Omega)}. \quad (8)$$

The received signal covariance matrix can be estimated at each iteration k as follows,

$$\mathbf{R}^{(\Omega)}(k) = \frac{1}{L} \sum_{l=0}^{L-1} \mathbf{x}^{(\Omega)}(k-l) \mathbf{x}^{(\Omega)}(k-l)^H, \quad (9)$$

where $\mathbf{x}^{(\Omega)}(k)$ is the received signal vector and L is the length of the summation. Since $\mathbf{R}_n^{(\Omega)}$ and $\mathbf{R}^{(\Omega)}(k)$ are readily estimated from (7) and (9), our task is to find a non-negative PSD, $S^{(\Omega)}(k)$, which optimizes the following problem:

$$\min_{S^{(\Omega)}, S^{(\Omega)} > 0} \| \mathbf{R}^{(\Omega)}(k) - \mathbf{R}_n^{(\Omega)} - S^{(\Omega)} \bar{\mathbf{R}}_s^{(\Omega)} \|_{\mathcal{F}}, \quad (10)$$

where $\| \cdot \|_{\mathcal{F}}$ is the Frobenius norm. Let us denote the (p, q) complex element of the matrix $\mathbf{R}^{(\Omega)}(k) - \mathbf{R}_n^{(\Omega)}$ as $a_{pq} + j b_{pq}$ where $1 \leq p \leq I$ and $1 \leq q \leq I$. Likewise, each of the complex element of $\bar{\mathbf{R}}_s^{(\Omega)}$ is denoted as $c_{pq} + j d_{pq}$. The cost function in (10) can be reduced to the following

$$\| \mathbf{R}^{(\Omega)}(k) - \mathbf{R}_n^{(\Omega)} - S^{(\Omega)}\bar{\mathbf{R}}_s^{(\Omega)} \|_{\mathcal{F}}$$
$$= \left(\sum_{p=1}^{I}\sum_{q=1}^{I} \left| (a_{pq} + jb_{pq}) - S^{(\Omega)}(c_{pq} + jd_{pq}) \right|^2 \right)^{1/2}$$
$$= \left(\sum_{p=1}^{I}\sum_{q=1}^{I} (a_{pq} - S^{(\Omega)}c_{pq})^2 + (b_{pq} - S^{(\Omega)}d_{pq})^2 \right)^{1/2}. \quad (11)$$

By setting the first derivative of (11) to zero, the estimated PSD can be obtained as,

$$S^{(\Omega)}(k) = \max \left\{ 0, \frac{\sum_{p=1}^{I}\sum_{q=1}^{I} (a_{pq}c_{pq} + b_{pq}d_{pq})}{\sum_{p=1}^{I}\sum_{q=1}^{I} c_{pq}^2 + d_{pq}^2} \right\}. \quad (12)$$

This PSD is estimated at every iteration of the received signal covariance matrix to provide a spectrally optimized constraint on the source. In simple terms, it attempts to preserve the spectra of the source like a spectra moulder.

2.3. The Wiener Solution

We now formulate the Wiener solution using the information aforementioned. Let $\mathbf{w}_{opt}^{(\Omega)}(k)$ be the optimum weight vector for each frequency Ω,

$$\mathbf{w}_{opt}^{(\Omega)}(k) = [w_1^{(\Omega)}(k), \ldots, w_I^{(\Omega)}(k)]^T. \quad (13)$$

The optimum weight vector for each frequency Ω is given by the Wiener solution as,

$$\mathbf{w}_{opt}^{(\Omega)}(k) = \left[S^{(\Omega)}(k)\bar{\mathbf{R}}_s^{(\Omega)} + \mathbf{R}_n^{(\Omega)} \right]^{-1} S^{(\Omega)}(k)\bar{\mathbf{r}}_s^{(\Omega)}. \quad (14)$$

From the pre-defined source area model $\mathbf{A}$, the matrix $\bar{\mathbf{R}}_s^{(\Omega)}$ in (3) of a frequency Ω has non-zero determinant and is therefore a full rank matrix[1]. Thus, this matrix can be decomposed as follows,

$$\bar{\mathbf{R}}_s^{(\Omega)} = \mathbf{V}^{(\Omega)}\mathbf{\Lambda}^{(\Omega)}\mathbf{V}^{(\Omega)H}, \quad (15)$$

where $\mathbf{V}^{(\Omega)} = [\mathbf{v}_1^{(\Omega)}, \cdots, \mathbf{v}_I^{(\Omega)}]$ is a matrix that contains the eigenvectors and $\mathbf{\Lambda}^{(\Omega)} = \text{diag}\{\lambda_1^{(\Omega)}, \cdots, \lambda_I^{(\Omega)}\}$ is a diagonal matrix that consists of the eigenvalues. Substituting (15) into (16), we have

$$\mathbf{w}_{opt}^{(\Omega)}(k) = \left[S^{(\Omega)}(k)\mathbf{V}^{(\Omega)}\mathbf{\Lambda}^{(\Omega)}\mathbf{V}^{(\Omega)H} + \mathbf{R}_n^{(\Omega)} \right]^{-1} S^{(\Omega)}(k)\bar{\mathbf{r}}_s^{(\Omega)}. \quad (16)$$

For computational savings, the spatial source covariance matrix can be estimated using the largest eigenvalue, $\lambda_{max}^{(\Omega)}$ with its corresponding eigenvectors $\mathbf{v}_{max}^{(\Omega)}$. By using the *Matrix Inversion Lemma* [7], the inversion in (16) can be avoided and the optimum weight vector can be reduced to

$$\mathbf{w}_{opt}^{(\Omega)}(k) = \frac{S^{(\Omega)}(k)(\mathbf{R}_n^{(\Omega)})^{-1}\bar{\mathbf{r}}_s^{(\Omega)}}{1 + S^{(\Omega)}(k)\lambda_{max}^{(\Omega)}\mathbf{v}_{max}^{(\Omega)H}(\mathbf{R}_n^{(\Omega)})^{-1}\mathbf{v}_{max}^{(\Omega)}}. \quad (17)$$

Finally, the beamformer output at frequency Ω is calculated as

$$y^{(\Omega)}(k) = \mathbf{w}_{opt}^{(\Omega)}(k)^H \mathbf{x}^{(\Omega)}(k). \quad (18)$$

3. EVALUATIONS

The evaluation of the proposed beamformer is made in a real car hands-free situation. A six-sensor array is mounted on the visor at the passenger side in a Volvo station wagon. Data are gathered on a multi-channel DAT-recorder with a sampling rate of 12 kHz and the car is moving at a constant speed of 110 km/h. The circular area of the target signal is 30 cm from the center of the array with a radius of 10 cm.

The performance of the proposed microphone array is measured in terms of the suppression level, defined as

$$SP = 10\log_{10}\left(\frac{\int_{-\pi}^{\pi} \hat{P}_{in,n}(\omega)d\omega}{\int_{-\pi}^{\pi} \hat{P}_{out,n}(\omega)d\omega} \right) - 10\log_{10}(C_d) \quad (19)$$

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 and C_d is a constant to normalize the source's gain. The performance is also given in terms of the distortion measure, defined as

$$DS = 10\log_{10}\left(\frac{1}{2\pi} \int_{-\pi}^{\pi} |(1/C_d)\hat{P}_{in,s}(\omega) - \hat{P}_{out,s}(\omega)|d\omega \right) \quad (20)$$

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 suppression and distortion levels for a noisy environment with a signal to noise ratio (SNR) of -7 dB and the number of subbands increases from 16 to 128. The decimation factor in this case is made over-sampled and fixed at $D = \frac{M}{2}$ for all number of subbands. The purposes of over-sampling are to reduce the aliasing effects between the adjacent subbands and to ensure the sufficiency of data in estimating the statistics. It can be seen from the table that the proposed structure achieves a significant suppression level while maintaining a small distortion level. Moreover, the suppression and distortion levels are significantly improved when the number of subbands increases from 16 to 64. These levels are, however, approximately the same when the number of subbands increases from 64 to 128. In view of this, the following plots are for the case with 64 subbands.

[1] Depending on how much of the space it spans, it will have a few dominating eigenvalues.

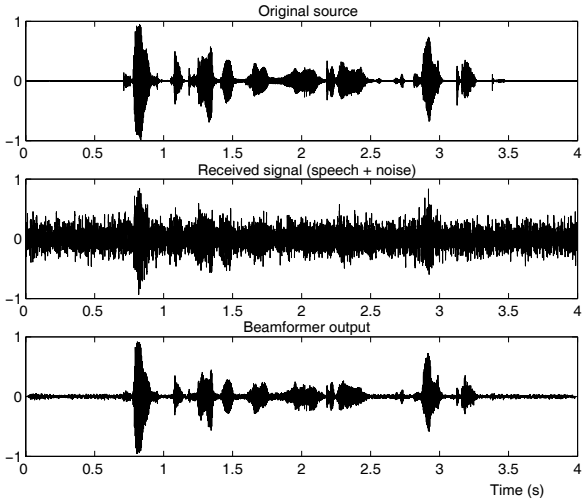

Fig. 3. *Time domain plots of the original source, received signal and the beamformer output.*

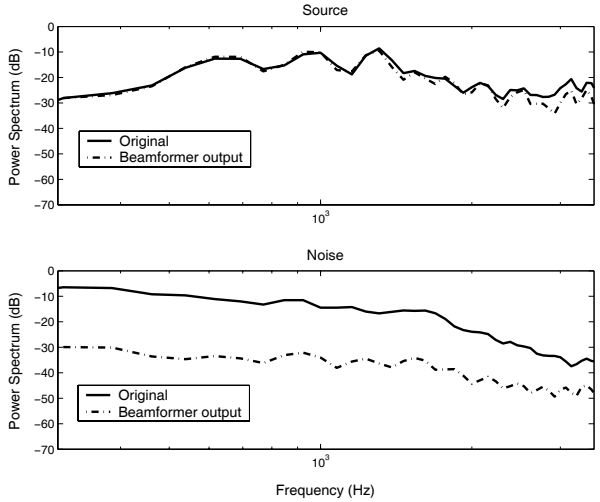

Fig. 4. *Output powers of the source and the noise before and after processing.*

No. of subbands	Suppression, SP (dB)	Distortion, DS (dB)
16	18.99	-25.96
32	21.28	-25.38
64	21.61	-27.87
128	21.48	-27.85

Table 1. *Suppression and distortion levels for different number of subbands.*

Fig. 3 shows the time domain plots of the original speech, noisy speech and the beamformer output. Evidently even in such adverse condition, the proposed beamformer still manages to suppress the background noise significantly whilst maintaining good target signal integrity.

Fig. 4 plots the output powers for both the source and noise before and after the processing. The plot shows that noise is suppressed uniformly across the frequency and the processed target signal remains a faithful replica of the original source.

4. CONCLUSIONS

A new space constrained adaptive beamformer with PSD update has been presented. The novelty of the structure lies in the PSD update of the target signal. By doing so, the PSD of the beamformer's output signal is weighted efficiently in the temporal domain, thus yielding a very good target signal integrity. Besides that, the incorporation of the space constraints in its formulation offers robustness and accurately captures the source of interest. The combination of both the PSD and the space constraints make full use of the available spatio-temporal domain. Results show that the beamformer manages to achieve an impressive noise suppression level up to 21 dB in a real car environment.

5. REFERENCES

[1] O. Hoshuyama, A. Sugiyama, and A. Hirano, "A robust adaptive beamformer for microphone arrays with a blocking matrix using constrained adaptive filters," *IEEE Transactions on Signal Processing*, vol. 47, no. 10, pp. 2677–2684, Jun. 1999.

[2] S. Affes and Y. Grenier, "A signal subspace tracking algorithm for microphone array processing of speech," *IEEE Transactions on Speech and Audio Processing*, vol. 5, no. 5, pp. 425–437, Sep. 1997.

[3] B. D. Van Veen and K. M. Buckley, "Beamforming: A versatile approach to spatial filtering," *IEEE Acoustics, Speech and Signal Processing Magazine*, vol. 5, pp. 4–24, Apr. 1988.

[4] D. H. Johnson and D. E. Dudgeon, *Array Signal Processing: Concepts and Techniques*, Prentice Hall Signal Processing Series. Prentice Hall, New Jersey 07632, 1993.

[5] H. Q. Dam, S. Nordholm, N. Grbic, and H. H. Dam, "Speech enhancement employing adaptive beamformer with recursively updated soft constraints," *International Workshop on Acoustics Echo and Noise Cancellation*, pp. 307–310, Sep. 2003.

[6] S. Nordholm, I. Claesson, and M. Dahl, "Adaptive microphone array employing calibration signals: Analytical evaluation," *IEEE Transactions on Speech and Audio Processing*, vol. 7, no. 3, pp. 241–252, May 1999.

[7] S. Haykin, *Adaptive Filter Theory*, Prentice Hall, Upper Saddle River, New Jersey 07458, 4th edition, 2002.