SPATIAL FILTER BANK DESIGN FOR SPEECH ENHANCEMENT BEAMFORMING APPLICATIONS

Zohra Yermeche, Per Cornelius, Nedelko Grbić and Ingvar Claesson

Blekinge Institute of Technology
School of Engineering
37225 Ronneby, Sweden

ABSTRACT

In this paper, a new spatial filter bank design method for speech enhancement beamforming applications is presented. The aim of this design is to construct a set of different filter banks that would include the constraint of signal passage at one position (and closing in other positions corresponding to known disturbing sources). By performing the directional opening towards the desired location in the fixed filter bank structure, the beamformer is left with the task of tracking and suppressing the continuously emerging noise sources. This algorithm has been implemented in MATLAB and tested on real speech recordings conducted in a car hands-free communication situation. Results show that a reduction of the total complexity can be achieved while maintaining the noise suppression performance and reducing the speech distortion.

1. INTRODUCTION

Microphone arrays can be exploited for speech enhancement in order to extract a speaker while suppressing interfering speech and background noise. These arrays are used in conjunction with digital beamforming, a technique performing spatial filtering to separate signals that have overlapping frequency content but are originated from different spatial locations. A microphone array consists of a set of acoustic sensors placed at different locations in order to spatially sample the sound pressure field. It offers a directivity gain proportional to the number of sensors. Thus, adaptive array processing, i.e. beamforming, of the spatial microphone samples allows time-variant control of spatial and spectral selectivity [1]. Several beamforming techniques have been suggested in order to enhance the desired speech source [2, 3, 4]. A Constrained Adaptive Subband Beamformer has been evaluated in [5] for speech enhancement in hands-free communication situations. The adaptive beamformer optimizes the array output by adjusting the weights of finite length digital filters so that the combined output contains minimal contribution from noise and interference. A soft constraint, formed from calibration data, secures the spatio-temporal passage of the desired source signal, without the need of any speech detection. The computational complexity of the finite impulse response filters is substantially reduced by introducing a subband beamforming scheme [6].

The weight update equation for the constrained adaptive beamformer implies the calculation and the use of a combined covariance matrix at each iteration. From the observation that the combined covariance matrix comprises a pre-calculated fixed part and a recursively updated part, the beamforming problem can be divided into a fixed part and an adaptive part. The objective in this paper is to transfer the a-priori known portion of the optimisation process into the filter bank structure (fixed part of the system).

Information about the desired speech location is used in the filter bank design by adding a spatial decomposition of the multichannel data for each subband. This spatial decomposition takes the form of a spatial transformation matrix, and it is extracted from correlation function estimates.

2. SUBBAND BEAMFORMING

Figure 1 illustrates the overall architecture of the microphone array speech enhancement system, based on the constrained adaptive subband beamformer. The structure includes a multichannel uniform over-sampled analysis filterbank used to decompose the received array signals into a set of subband signals and a set of adaptive beamformers, each adapting on the multichannel subband signals. The outputs of the beamformers are reconstructed by a synthesis filter-bank in order to create a time domain output signal. The spatial characteristics of the input signal are maintained when using the same modulated filter bank. The filter banks are defined by two prototype filters, which leads to efficient polyphase realisations [7].

The source is assumed to be a wideband source, as in the case of a speech signal, located in the near field of a uniform linear array of number I microphones. The filtering operations of the beamformer are formulated in the frequency domain as multiplications with number I complex

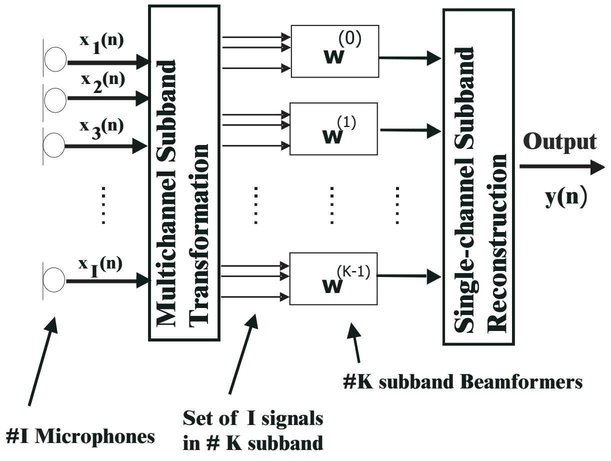
x 1(n)
x 2(n)
x 3(n)
x I(n)
Multichannel Subband Transformation
w(0)
w(1)
w(K-1)
Single-channel Subband Reconstruction
Output
y(n)
#I Microphones
Set of I signals in # K subband
#K subband Beamformers

Fig. 1. *Structure of the subband beamformer.*

frequency domain representation weights, $w_i^{(f)}(n)$, for each frequency. For a specific frequency, f, the output is given by

$$y^{(f)}(n) = \sum_{i=1}^{I} w_i^{(f)}(n) x_i^{(f)}(n) \tag{1}$$

where the signals, $x_i^{(f)}(n)$ are digitally sampled microphone observations and $y^{(f)}(n)$ corresponds to the beamformers output. These time domain signals are narrow band, containing essentially components with frequency f.

The objective of the beamformer is formulated in the frequency domain as a calibrated weighted recursive least square solution, where the optimal weight vectors $\mathbf{w}_{ls,opt}^{(f)}(n)$ are calculated by

$$\mathbf{w}_{ls,opt}^{(f)}(n) = \left[\hat{\mathbf{R}}_{ss}^{(f)} + \hat{\mathbf{R}}_{ii}^{(f)} + \hat{\mathbf{R}}_{xx}^{(f)}(n)\right]^{-1} \hat{\mathbf{r}}_s^{(f)} \tag{2}$$

where an initial calibration procedure is used to calculate source correlation estimates, i.e. the correlation matrix estimate $\hat{\mathbf{R}}_{ss}^{(f)}$ and the cross correlation vector estimate $\hat{\mathbf{r}}_s^{(f)}$, for microphone observations when the source signal of interest is active alone, as well as the interference correlation matrix estimate, $\hat{\mathbf{R}}_{ii}^{(f)}$, when the known source interferences are active alone [5].

Conversely, the correlation estimates, $\hat{\mathbf{R}}_{xx}^{(f)}(n)$, are continuously calculated from observed data by

$$\hat{\mathbf{R}}_{xx}^{(f)}(n) = \sum_{p=0}^{n} \lambda^{n-p} \mathbf{x}^{(f)}(p) \mathbf{x}^{(f)^H}(p) \tag{3}$$

where

$$\mathbf{x}^{(f)}(n) = [x_1^{(f)}(n), \quad x_2^{(f)}(n), \quad \dots \quad x_I^{(f)}(n)]^T$$

and λ is a forgetting factor, with the purpose of tracking variations in the surrounding noise environment. The initially precalculated correlation estimates constitutes a soft constraint in the recursive update of the beamforming weights.

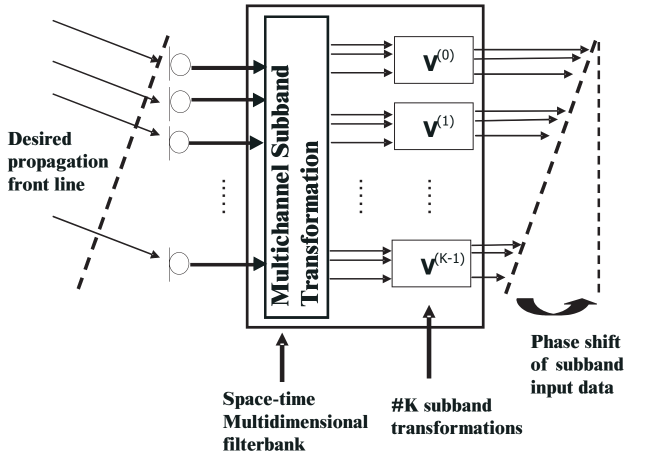
Desired propagation front line
Multichannel Subband Transformation
V(0)
V(1)
V(K-1)
Phase shift of subband input data
Space-time Multidimensional filterbank
#K subband transformations

Fig. 2. *Structure of the multidimensional space-time filterbank (The output data of the filter bank are phase-shifted to be in-phase for the source propagation direction, and out-of-phase for interference propagation directions).*

3. SUGGESTED METHOD

In the previous structure of the constrained subband beamformer, both the fixed pre-calculated source correlation estimates and the updated data correlation estimates are used at each iteration for the update of the beamformers weight vectors (see (2)).

In this section a more efficient design method is introduced. The spatial information carried by the source correlation estimates is used to process the data prior to the beamformers, through a matrix transformation, based on matrices $\mathbf{V}^{(f)}$ (see Fig. 2), undergoing a spatial decomposition.

The resulting subband signal vector $\mathbf{x}'^{(f)}(n)$ is given by

$$\mathbf{x}'^{(f)}(n) = \mathbf{V}^{(f)H} \mathbf{x}^{(f)}(n). \tag{4}$$

By this method, the spatial information carried by the input vector to the beamformer is transformed in such a way to direct the array towards the source position, and close its opening in directions of known interfering sources.

The objective is therefore to maximise the quadratic ratio between the source signal power and the interference signal power according to

$$\mathbf{v_{max}}^{(f)} = \arg\max_{\mathbf{v^{(f)}}} \left\{ \frac{\mathbf{v}^{(f)H} \hat{\mathbf{R}}_{ss}^{(f)} \mathbf{v}^{(f)}}{\mathbf{v}^{(f)H} \hat{\mathbf{R}}_{ii}^{(f)} \mathbf{v}^{(f)}} \right\} \tag{5}$$

The source correlation matrix, $\hat{\mathbf{R}}_{ss}^{(f)}$, and the interference correlation matrix, $\hat{\mathbf{R}}_{ii}^{(f)}$, are estimated from received data when each component, source and interference, are individually active.

The solution of this optimisation problem is accomplished by the eigenvectors of the composed matrix $\hat{\mathbf{R}}_{ss}^{(f)}\hat{\mathbf{R}}_{ii}^{(f)^{-1}}$, in the order of decreasing corresponding eigenvalues, i.e. the optimal solution is the eigenvector belonging to the maximum eigenvalue.

Hence, the transformation matrix $\mathbf{V}^{(f)}$ is chosen to be the eigenvector matrix of the matrix $\hat{\mathbf{R}}_{ss}^{(f)}\hat{\mathbf{R}}_{ii}^{(f)^{-1}}$.

One way to reduce the complexity of the problem, without any significant loss of information, is to reduce the number of eigenvectors in $\mathbf{V}^{(f)}$ such that the most significant eigenvectors are used. As a result, the dimension of the input vector, and consequently the dimension of the correlation matrix and weight vector in (2), is reduced.

4. SIMULATIONS AND RESULTS

The performance of the beamformer was evaluated in a car hands-free telephony environment with a linear microphone array mounted on the visor at the passenger side, see Fig. 3. The measurements were performed in a Volvo station wagon. The speech originating from the passenger position constitutes the desired source signal and the hands-free loudspeaker emission is the source interfering signal, while the ambient noise received in a moving car constitutes the background noise. A loudspeaker was mounted to the passenger seat to simulate a real person engaging a conversation. The sensors used in this evaluation were six high quality Sennheiser microphones uniformly spaced in-line with 5 cm spacing. The microphone-array was positioned at a distance of 35 cm from the artificial speaker. Data was gathered on a multichannel DAT-recorder with a sampling rate of 12 KHz, and with a 300-3400 Hz bandwidth.

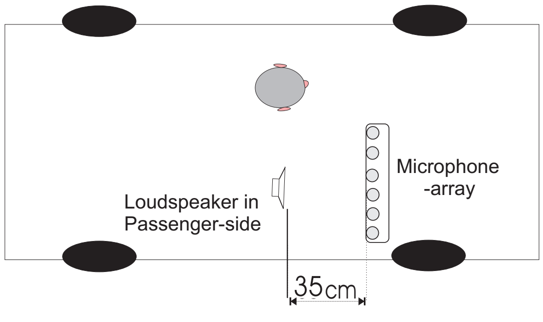

Fig. 3. *Placement of the microphone-array in the car. The distance between microphone centers is 5 cm.*

The desired source calibration signals were initially recorded when a speech sequence was emitted from the artificial talker, in a non-moving car with the engine turned off. Similarly, interference calibration signals were recorded by emitting a different speech sequence, from the hands-free loudspeaker alone, within the bandwidth.

In order to evaluate the proposed beamformer's new structure, a set of weights were calculated according to (2), based on correlation estimates calculated from source input data. The performance evaluation includes source speech distortion and suppression of both background noise and hands-free loudspeaker interference as well as computational complexity.

In the particular case of a car scenario, the interference sound power originated from the hands-free loudspeaker is low compared to the noise generated by the wind, the car engine and the tire frictions. Based on this observation, the interfering source signal is considered as part of the surrounding noise, and the optimisation of (5) is simplified by replacing the interference correlation matrix estimate $\hat{\mathbf{R}}_{ii}^{(f)}$ by the identity matrix of size I. Consequently, the transformation matrix $\mathbf{V}^{(f)}$ is chosen to be the eigenvector matrix of the source calibration signals, $\hat{\mathbf{R}}_{ss}^{(f)}$.

The original Constrained Subband beamformer is compared in Fig. 4 to the reduced-rank beamformer based on the proposed space-time filter-bank structure described in Sec. 3, when it comes to noise suppression (top subplot), interference suppression (middle subplot) and speech distortion (bottom subplot). The effect of reducing the rank of the beamformer, with one or two dimensions, on the resulting performances is also presented. The figure shows that by using the spatial decomposition of the input data (following (4)) prior to the beamforming process, the distortion of the speech is considerably decreased while the noise and interference suppression is relatively unchanged. Furthermore, the reduction from six to four dimensions for this scenario maintains the performances of the algorithm.

In Fig. 5, the computational complexity gain of the proposed method is evaluated when reducing the rank of the beamformer algorithm. A considerable reduction of computational complexity is obtained when less dimensions are used.

5. CONCLUSION

A new beamforming algorithm based on a spacial filter bank design method has been presented and evaluated on real-world recordings in a car hands-free situation.

Results with the new method were compared to the ones obtained from the original constrained subband beamformer. The results clearly show that by directing the beamformer input-vector towards the source propagation direction, prior to the beamformer, the proposed method maintains the noise and interference suppression performances of the original subband beamformer, while decreasing the speech distor-

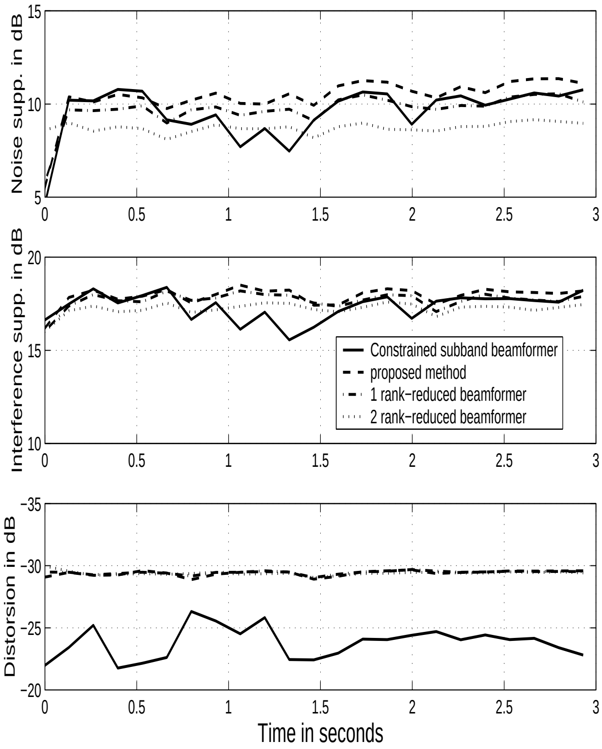

Fig. 4. *Performance evaluation of the spatial subband beamformer, using an array of six microphones, in the case of 0, 1 and 2 dimensions reduced in the beamformer calculations (i.e. using a transformation matrix composed of, respectively, the 6, 5 and 4 most significant eigenvectors).*

tion with approximately 5 dB and reducing its computational complexity significantly.

6. REFERENCES

[1] D. Johnson and D. Dudgeon, *Array Signal Processing - Concepts and Techniques,* Prentice Hall, 1993.

[2] D. A. Florêncio and H. S. Malvar, "Multichannel filtering for optimum noise reduction in microphone arrays," in *IEEE International Conference on Acoustics, Speech and Signal Processing*, May 2001, vol. 1, pp. 197–200.

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

[4] F. Asano, S. Hayamizu, T. Yamada, and S. Nakamura, "Speech enhancement based on the subspace method," *IEEE Trans. Acoust. Speech Signal Processing*, vol. 8, no. 5, pp. 497 – 507, Sep. 2000.

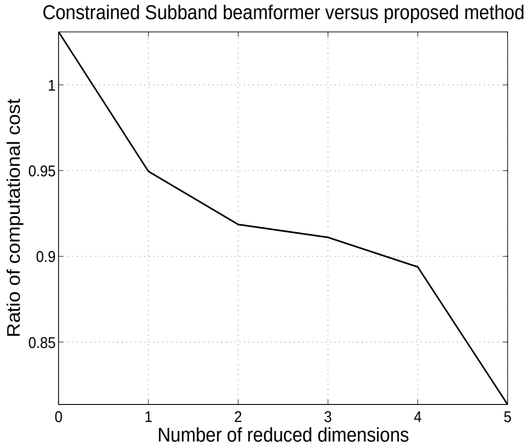

Fig. 5. *Ratio of the computational cost between the reduced-rank proposed beamformer and the Constrained Subband beamformer. It can be seen that the proposed subband beamformer even with full-rank presents less than 5 per cent increase in computational cost, when compared to the original subband beamformer, while gaining considerably in performance.*

[5] Z. Yermeche, P Márquez Garcia, N. Grbić and I. Claesson, "A Calibrated Subband Beamforming Algorithm for Speech Enhancement," in *IEEE Sensor Array and Multichannel Signal Processing Workshop*, August 2002.

[6] N. Grbić and S. Nordholm, "Soft constrained subband beamforming for hands-free speech enhancement," in *IEEE International Conference on Acoustics, Speech and Signal Processing*, vol. 1, pp.885-888, May 2002.

[7] J. M. de Haan, N. Grbić, I. Claesson, and S. Nordholm, "Design of oversampled uniform dft filter banks with delay specifications using quadratic optimization," in *IEEE International Conference on Acoustics, Speech and Signal Processing*, May 2001, vol. VI, pp. 3633–3636.