ROBUST BROADBAND BEAMFORMER WITH DIAGONALLY LOADED CONSTRAINT MATRIX AND ITS APPLICATION TO SPEECH RECOGNITION

Wenyi Zhang and Bhaskar D. Rao

Department of Electrical and Computer Engineering
University of California, San Diego
La Jolla, CA 92093-0407, USA
Email: w3zhang@ucsd.edu, brao@ece.ucsd.edu

ABSTRACT

In this paper, we develop a robust wideband adaptive beamforming algorithm for microphone array based speech recognition. We develop a quadratic constraint based approach to deal with the uncertainty in the look direction. In order to address the ill-conditioning associated with the constraint matrix, diagonal loading (DL) is employed. The advantage of adding DL to the constraint matrix is that the constraint matrix is only determined by the geometry of the array thereby allowing the DL level to be chosen offline. We also develop an iterative algorithm (and corresponding adaptive algorithm) to solve for the robust beamformer coefficients. The developed algorithm is applied to the problem of beamforming using microphone arrays for speech recognition and shown to be superior to existing algorithms.

1. INTRODUCTION

We consider the use of microphone arrays for speech recognition. In particular, we consider the development of robust wideband adaptive beamforming algorithms for this purpose. An adaptive beamformer is able to adjust its beam pattern based on the input statistics to place deep nulls in the direction of interferences. Among the broadband adaptive beamformer, the Frost beamformer is one of the most extensively studied [1]. The Frost beamformer has a multichannel tapped-delay-line structure (Fig. 1), and a set of linear constraints are introduced to ensure a desired frequency response in the look direction.

However, the performance of the Frost beamformer can degrade severely in practice when steering vector errors exist, which may be due to look direction error, array sensor position error, and small mismatches in the sensor response. In such cases, the desired signal might be mistaken as an interference signal and be suppressed [2, 4]. Several robust beamforming algorithms have been proposed to address this problem [4, 5, 6, 7]. In [5], the steering vector errors are modelled by "time-delay errors" and compensated for by self-adjusted interpolation filtering. In [6], a method is proposed to optimize the worst-case performance. The problem is formulated as minimizing a quadratic function subject to infinitely many quadratic constraints. It is reduced to a second-order cone programming problem which can be solved by interior point methods.

Er and Cantoni proposed a robust broadband beamforming algorithm which restricts the error between the desired and actual beam pattern of the array over the frequency band of interest and over a small spatial region around the array's look direction, allowing for uncertainty in the look direction [4]. The constraint thus obtained can be a quadratic constraint or reduced to a set of linear constraints. In [7], linear constraints are constructed for a set of sampling points around the look direction, which are reduced to a small number of linear constraints using matrix SVD. The algorithm is similar to [4] if the number of sampling points is large enough, as summation can be viewed as an integration. However, Er's linear constraint algorithm is an approximation because the quadratic constraint is not strictly equal to the set of linear constraints. An additional norm constraint is imposed to overcome this limitation and this complicates the optimization problem.

In this paper, we also consider quadratic constraint based approach to deal with the uncertainty in the look direction. In order to address the ill-conditioning associated with the constraint matrix, a diagonal loading (DL) is added to the constraint matrix thereby ensuring a robust solution to the quadratic constraint beamforming problem. The advantage of adding DL to constraint matrix is that the constraint matrix is only determined by the geometry of the array thereby allowing the DL level to be chosen offline. This is superior to adding DL to the signal covariance matrix where the DL level has to be chosen online. It is shown that the diagonal loading is equivalent to an additional norm constraint without introducing it explicitly. We also develop an iterative algorithm (and corresponding adaptive algorithm) to solve for the robust beamformer coefficients. The developed algorithm is applied to the problem of beamforming using microphone arrays for speech recognition and shown to be superior to existing algorithms.

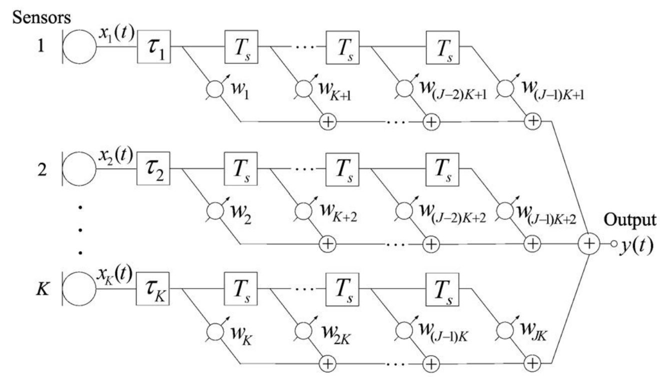


Fig. 1. Broadband beamformer with K sensors and J taps.

This material is based upon work supported by the RESCUE project, under National Science Foundation Award Number 0331690.

2. QUADRATIC CONSTRAINT ROBUST BEAMFORMER

A quadratic constraint broadband beamformer which is robust to DOA error is proposed by Er [4]. Consider a pre-steered Frost beamformer with K sensors and J taps (Fig.1). The weighted square error between desired and actual beam pattern of Frost beamformer over the interested frequency range and a small spatial region, chosen to deal with look direction uncertainty, can be written as

$$e^2 = \int_{\bar{\theta}-\Delta\theta}^{\bar{\theta}+\Delta\theta} \int_{\omega} f(\theta) \left| \mathbf{w}^H V(\theta,\omega) - \mathbf{w}_d^H V(\theta,\omega) \right|^2 d\omega d\theta \tag{1a}$$

$$= (\mathbf{w}-\mathbf{w}_d)^H \int_{\bar{\theta}-\Delta\theta}^{\bar{\theta}+\Delta\theta} \int_{\omega} f(\theta) V(\theta,\omega) V(\theta,\omega)^H d\omega d\theta (\mathbf{w}-\mathbf{w}_d) \tag{1b}$$

$$= (\mathbf{w}-\mathbf{w}_d)^H \Phi (\mathbf{w}-\mathbf{w}_d), \tag{1c}$$

$$\text{with} \;\; \Phi = \int_{\bar{\theta}-\Delta\theta}^{\bar{\theta}+\Delta\theta} \int_{\omega} f(\theta) V(\theta,\omega) V(\theta,\omega)^H d\omega d\theta. \tag{2}$$

$\bar{\theta}$ is the assumed look direction, $\Delta\theta$ is a measure of uncertainty in the assumed look direction, $f(\theta)$ is a spatial weighting function, ω is the frequency variable and $V(\theta,\omega)$ is the array steering vector. Φ is the positive semi-definite constraint matrix which can be calculated by either mathematical integration or by numerical techniques. $\mathbf{w}$ is the pursued beamformer's weight vector, and $\mathbf{w}_d$ is the desired beamformer's weight vector. Generally, Delay and Sum beamformer is used as the desired beamformer because of its robustness in the look direction.

The robust beamforming problem can be mathematically formulated as the following optimization problem

$$\min_{\mathbf{w}} \mathbf{w}^H R_{XX} \mathbf{w}, \quad \text{subject to} \quad (\mathbf{w}-\mathbf{w}_d)^H \Phi (\mathbf{w}-\mathbf{w}_d) \leq \varepsilon \tag{3}$$

where R_{XX} is the correlation matrix of the concatenated data vector X as in [1], and ε is a parameter chosen to control the tightness of the quadratic constraint. The weight vector $\mathbf{w}$ is a real vector, and Φ is a positive definite complex matrix. The constraint function $(\mathbf{w}-\mathbf{w}_d)^H \Phi (\mathbf{w}-\mathbf{w}_d)$ will always be a real number. Let $\Phi = \Phi_r + j\Phi_i$, then the constraint function $(\mathbf{w}-\mathbf{w}_d)^H \Phi (\mathbf{w}-\mathbf{w}_d) = (\mathbf{w}-\mathbf{w}_d)^T \Phi_r (\mathbf{w}-\mathbf{w}_d)$. Hence we always replace Φ with its real part Φ_r and assume no difference between Φ and Φ_r.

Let $\mathbf{w}_e = \mathbf{w} - \mathbf{w}_d$. Problem (3) can be written as

$$\min_{\mathbf{w}_e} (\mathbf{w}_d + \mathbf{w}_e)^T R_{XX} (\mathbf{w}_d + \mathbf{w}_e), \quad \text{subject to} \quad \mathbf{w}_e^T \Phi \mathbf{w}_e \leq \varepsilon \tag{4}$$

Two methods will be developed to solve this quadratic constraint beamforming problem.

2.1. Lagrangian Multiplier Method

The Lagrange function is defined as

$$L(\lambda) = (\mathbf{w}_d + \mathbf{w}_e)^T R_{XX} (\mathbf{w}_d + \mathbf{w}_e) + \lambda (\mathbf{w}_e^T \Phi \mathbf{w}_e - \varepsilon) \tag{5}$$

Taking partial derivative with respect to $\mathbf{w}_e$, the optimal solution can be shown to be

$$\mathbf{w}_e = -(R_{XX} + \lambda\Phi)^{-1} R_{XX} \mathbf{w}_d \tag{6}$$

λ is the Lagrange multiplier and can be obtained via Newton's method. The details are omitted here because of space limitations.

2.2. Iterative Algorithm

We develop an iterative algorithm using the approach discussed in [8]. In [8], an iterative solution (also adaptive algorithm) was developed to solve the following constrained optimization problem:

$$\min_{\mathbf{w}\in S} \left\{ \frac{1}{2}\mathbf{w}^T R \mathbf{w} - \mathbf{w}^T b \right\}, \quad S = \{\mathbf{w} : \|\mathbf{w}\| \leq \varepsilon\} \tag{7}$$

The iterative algorithm is given by

$$\mathbf{w}_{k+1} = P[(\mathbf{I} - \mu R)\mathbf{w}_k + \mu b], \tag{8}$$

where

$$P(\mathbf{w}) = \begin{cases} \mathbf{w}, & \text{if } \|\mathbf{w}\| \leq \varepsilon \\ \mathbf{w}\frac{\sqrt{\varepsilon}}{\|\mathbf{w}\|}, & \|\mathbf{w}\| > \varepsilon. \end{cases} \tag{9}$$

We now manipulate the robust beamforming problem (4) into a form compatible with (8), enabling development of the iterative algorithm. Using eigenvalue decomposition, the matrix Φ in equation (4) can be written as,

$$\Phi = U\Lambda U^T = (U\Lambda^{1/2})(U\Lambda^{1/2})^T \tag{10}$$

Let $\tilde{\mathbf{w}}_e = (U\Lambda^{1/2})^T \mathbf{w}_e$. After some manipulation, equation (4) can be written as

$$\min_{\tilde{\mathbf{w}}_e} \{ \frac{1}{2}\tilde{\mathbf{w}}_e^T \tilde{R} \tilde{\mathbf{w}}_e - \tilde{\mathbf{w}}_e^T \tilde{b} \}, \quad \text{subject to} \quad \tilde{\mathbf{w}}_e^T \tilde{\mathbf{w}}_e \leq \varepsilon \tag{11}$$

where $\tilde{R} = \Lambda^{-1/2} U^T R_{XX} U \Lambda^{-1/2}$, and $\tilde{b} = -\Lambda^{-1/2} U^T R_{XX} \mathbf{w}_d$. Now the iterative algorithm (8) can be applied to solve this problem [8]. The corresponding adaptive algorithm is obtained by substituting the instantaneous estimates $\hat{R}_k$ and $\hat{b}_k$ into (8), where $\hat{R}_k = \Lambda^{-1/2} U^T X_k X_k^T U \Lambda^{-1/2}$ and $\hat{b} = -\Lambda^{-1/2} U^T X_k X_k^T \mathbf{w}_d$, respectively.

3. DIAGONAL LOADED ROBUST BEAMFORMER

In [4], Er proposed a method wherein the original quadratic constraint is replaced by a set of linear constraints. Assume matrix Φ is a low rank matrix which can be represented by its eigen-decomposition $\Phi = U_1 \Lambda_1 U_1^T$. The original quadratic constraint in equation (4) can be approximated by a set of linear constraints $U_1^T \mathbf{w}_e = \mathbf{0}$. The linear constraint algorithm is an approximation because the quadratic constraint $\mathbf{w}_e^T \Phi \mathbf{w}_e \leq \varepsilon$ is not equivalent to the linear constraints $U_1^T \mathbf{w}_e = \mathbf{0}$. Often matrix Φ is only approximated by $U_1 \Lambda_1 U_1^T$ but is not exactly equal to it. In general, $\Phi = U_1 \Lambda_1 U_1^T + U_2 \Lambda_2 U_2^T$. Although the norm of Λ_2 is small, if the norm of $\mathbf{w}$ is big enough, $\mathbf{w}_e^T U_2 \Lambda_2 U_2^T \mathbf{w}_e$ may exceed ε. In such cases, the original quadratic constraint is not satisfied.

In [4], an extra norm constraint on $\mathbf{w}$ was proposed to ensure that the quadratic constraint and the set of linear constraints are equivalent. However, it complicates the optimization problem and was not pursued in detail. The other disadvantage of the algorithm is that to ensure better approximation to matrix Φ, the dimension of Λ_1 has to be increased, which increases the number of linear constraints. Consequently, the degrees of freedom of the weight vector are reduced compromising the ability to suppress interferences.

In view of the low rank property of the matrix Φ, a method which can robustly solve equation (4) is now proposed. The idea is to add a diagonal loading to Φ to restore it to a full rank matrix, i.e. construct a new matrix Φ' such that $\Phi' = \Phi + \lambda I$. The diagonal loading method here is close to the diagonal loading method in [2] except that the diagonal loading is added to the constraint matrix Φ instead of the signal covariance matrix R_{XX}. Substituting Φ' for Φ in (4), the problem can be rewritten as

$$\min_{\mathbf{w}_e}(\mathbf{w}_d + \mathbf{w}_e)^T R_{XX}(\mathbf{w}_d + \mathbf{w}_e), \quad \text{s.t.} \quad \mathbf{w}_e^T \Phi' \mathbf{w}_e \leq \varepsilon \quad (12)$$

where $\Phi' = \Phi + \lambda I$. The above optimization problem can be solved robustly by the Lagrangian method in section 2.1 or by the iterative algorithm in section 2.2. Since Φ is totally determined by the array geometry, it can be calculated beforehand as well as the DL level λ. It can be chosen optimally offline with respect to Φ. In contrast, in diagonal loading of the signal covariance matrix R_{XX}, the DL level has to be adjusted online with respect to different values of R_{XX}. Furthermore, the new well-conditioned constraint matrix Φ' ensures a robust eigenvalue decomposition such that the iterative and corresponding adaptive algorithm are stable, while for the method of adding DL to signal covariance matrix no such adaptive algorithm is available.

To get more insight into problem (12), we expand the constraint

$$\mathbf{w}_e^T \Phi' \mathbf{w}_e = \mathbf{w}_e^T \Phi \mathbf{w}_e + \lambda \mathbf{w}_e^T \mathbf{w}_e. \quad (13)$$

Define $\mathbf{w}_e^T \Phi \mathbf{w}_e = \varepsilon_1$ and $\lambda \mathbf{w}_e^T \mathbf{w}_e = \varepsilon_2$, to ensure the quadratic constraint is satisfied, it must be true $\varepsilon_1 + \varepsilon_2 \leq \varepsilon$. In other words, from the constraint $\mathbf{w}_e^T \Phi' \mathbf{w}_e \leq \varepsilon$, two constraints $\mathbf{w}_e^T \Phi \mathbf{w}_e \leq \varepsilon$ and $\lambda \mathbf{w}_e^T \mathbf{w}_e \leq \varepsilon$ can be met. Consequently, if constraint $\mathbf{w}_e^T \Phi' \mathbf{w}_e \leq \varepsilon$ is satisfied, not only is the constraint in our original optimization problem (4) satisfied, but also a new norm constraint on the weight vector $\mathbf{w}_e$ is introduced. The norm constraint will introduce further robustness as shown before [3]. And λ can be considered as a parameter which leverages constraint $\mathbf{w}_e^T \Phi \mathbf{w}_e = \varepsilon_1$ and the norm constraint. In conclusion, the diagonal loading to the matrix Φ facilitates robustness in the solution of the original optimization problem (4).

4. EXPERIMENTS

A speech recognition experiment is conducted on the Multichannel Overlapping Numbers Corpus (MONC) to test the performances of various beamforming algorithms. MONC is a multichannel speech database recorded in a moderately reverberant 8.2m*3.6m*2.4m rectangular room [9]. The recording scenario was designed to simulate three speakers seated around a 1.2m diameter circular meeting room table. The loudspeakers (L1, L2, L3) were placed at 90° spacings at an elevation of 35cm. An eight-element, 20cm diameter, circular microphone array was placed in the centre of the table. An additional microphone was placed at the centre of the array. The loudspeaker and microphone placement is illustrated in Fig.2.The sampling rate for recording is $8k$Hz.

In applying the beamforming algorithms to the data from the MONC database, the desired source is speaker L1, which is assumed to be a point source coming from the location: angle 180°, radius 70cm, height 35cm. We define the origin of the coordinates to be the center of the circular microphone array, and define angle 0° to be the direction of the 8_{th} microphone. The angle increases counter clockwise, which means the 1_{st} microphone is in the direction of angle 45°. Only microphones 1 to 8 are used in the beamforming.

Speech recognition results using different beamforming algorithms are shown in Fig.3. Five beamforming algorithms and one single microphone based approach are compared. CentreMic means using single centre microphone's signal for speech recognition, without beamforming. DS, Frost, robFrost_orig, robFrost_Er and robFrost_DL represent conventional Delay-and-Sum beamforming, optimal Frost beamforming [1], robust Frost beamforming with quadratic constraint

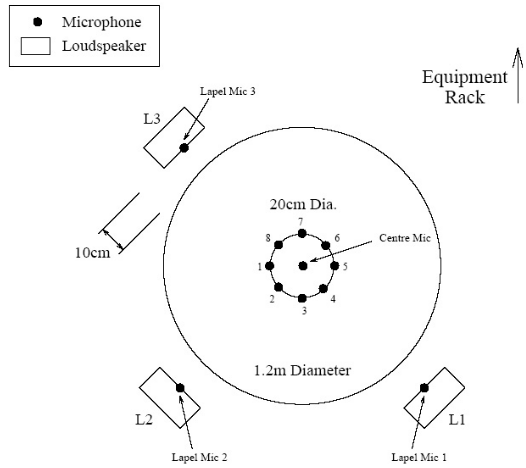


Fig. 2. MONC database experiment scenario

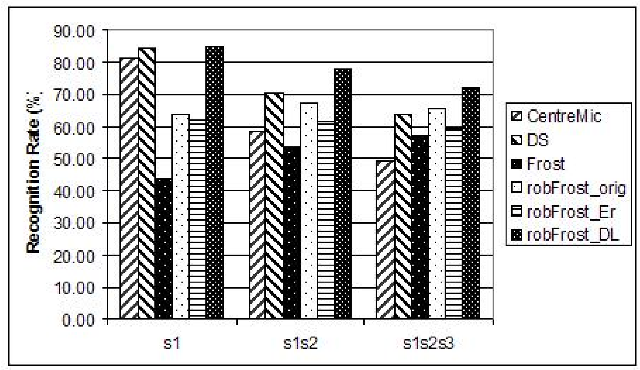
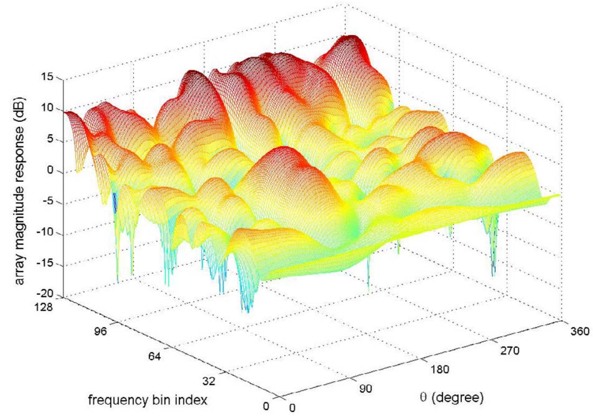
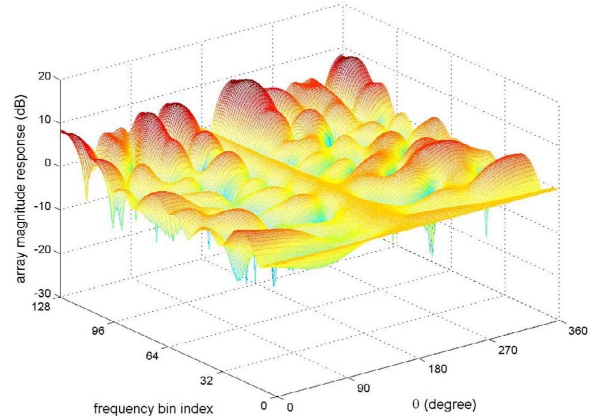


Fig. 3. Speech recognition rate

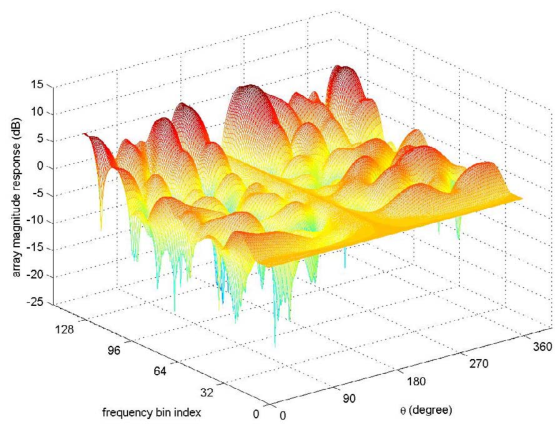
(using original constraint matrix, solved by Lagrangian method (6)), Er's robust Frost beamforming with a set of linear constraints [4] and robust Frost beamforming with quadratic constraint (using diagonal loaded constraint matrix) respectively. For the robFrost_DL algorithm, both the Lagrangian multiplier method and the iterative method yield identical experimental results, hence only the results obtained by the Lagrangian method are shown in Fig.3. The tap delay between each two taps of the Frost structure is the sampling interval, i.e. $0.125ms$. The constraint matrix of the robust Frost beamforming is calculated through integration over the whole frequency band 0–4000Hz and the uncertainty region: angle $180 \pm 3^\circ$, height $35 \pm 5cm$, and radius $70 \pm 3cm$. s1, s1s2 and s1s2s3 represent three different scenarios separately: only speaker s1 is speaking, both s1&s2 are speaking, and all three speakers are speaking. For the proposed algorithm, the DL level λ is chosen to be 10^{-5} times largest eigenvalue of Φ. From the experiments results, it is evident that the proposed beamforming algorithm is significant better than the other algrithms. We also plot the beam pattern of different beamformers using one multichannel sample sentence from the MONC database. Only speakers L1 and L2 are present in the sample sentence. Fig.4 shows the beam pattern of various beamformers over angle θ and frequency. It is apparent that the robust Frost beamforming with diagonal loaded constraint matrix has the best combination of robustness in the look direction (180°) and good suppression in



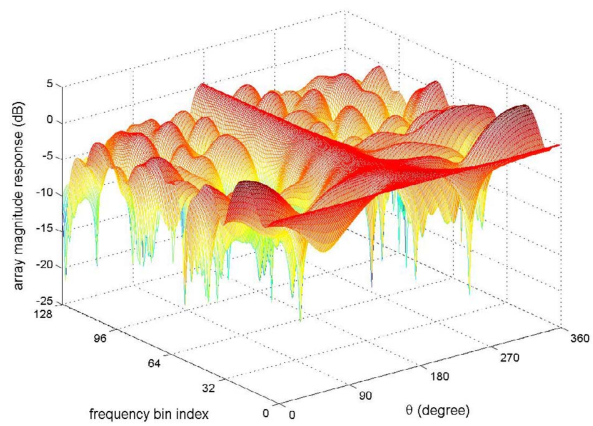
(a) Frost beamformer



(b) robust Frost beamformer with original constraint matrix



(c) Er's robust Frost beamformer with a set of linear constraints [4]



(d) robust Frost beamformer with diagonal loaded constraint matrix

Fig. 4. Beam pattern of various beamformers over angle θ and frequency bins. The radius is fixed at 70cm and height fixed at 35cm. The look direction is 180° and the interference direction is 90°

the interference direction (90°).

5. REFERENCES

[1] O. L. Frost III, "An algorithm for linearly constrained adaptive array processing," *Proc. IEEE,* vol. 60, pp. 926–935, Aug. 1972.

[2] H. Cox, R. M. Zeskind, and M. Owen, "Robust adaptive beamforming," emphIEEE Trans. on ASSP, vol. 35, pp. 1365–1376, 1987.

[3] J. Li, P. Stoica, and Z. Wang, "Doubly constrained robust Capon beamformer," *IEEE Transactions on Signal Processing,* vol. 51, pp. 2407–2423, Sep. 2004.

[4] M. H. Er and A. Cantoni, "A new set of linear constraints for broad-band time domain element space processors," *Antennas and Propagation, IEEE Transactions on,* vol. 34, Issue 3, pp. 320–329, Mar. 1986.

[5] Q. Zou, Z. Yu, and Z. Lin, "A robust algorithm for linearly constrained adaptive beamforming," *Signal Processing Letters, IEEE,* vol. 11, Issue 1, pp. 26–29, Jan. 2004.

[6] A. El-Keyi, T. Kirubarajan, and A.B. Gershman, "Robust wideband beamforming based on worst-case performance optimization," *Proc. IEEE Workshop on Statistical Signal Processing,* Bordeaux, France, Jul. 2005.

[7] Y. R. Zheng, R. A. Goubran and M. El-Tanany, "Robust near-field adaptive beamforming with distance discrimination," *Speech and Audio Processing, IEEE Transactions on* vol. 12, Issue 5, pp. 478–488, Sep. 2004.

[8] A. Krieger and E. Masry, "Constrained adaptive filtering algorithms: asymptotic convergence properties for dependent data," *Information Theory, IEEE Transactions on,* vol. 35, Issue 6, pp. 1166–1176, Nov. 1989.

[9] Multichannel Numbers corpus distribution. http://www.idiap.ch/speech/.