SOFT CONSTRAINED SUBBAND BEAMFORMING FOR HANDS-FREE SPEECH ENHANCEMENT

Nedelko Grbić

Blekinge Institute of Technology
Department of Telecommunications
and Signal Processing
372 25 Ronneby, Sweden

Sven Nordholm

Australian Telecommunications
Research Institute
Curtin University of Technology
Perth, Australia

ABSTRACT

This paper introduces a new constrained adaptive subband beamformer algorithm for speech enhancement in acoustic telecommunication systems. The solution relies on a pre-calculated source covariance matrix and recursive estimates of background noise- and handsfree signal covariance matrices. The constraint acts as an eye-opening in a vicinity of the near-field location of the source and degradations from steering-vector errors can therefor be made small. The algorithm is applied in subbands using a uniform multi channel over-sampled filterbank. Simulations with real speech recorded in an automobile hands-free environment show 19 dB noise reduction and 20 dB hands-free suppression.

1. INTRODUCTION

The increased use of personal communication devices, personal computers and wireless cellular telephones enables the development of new inter-personal communication systems. The merge between computers and telephony technologies brings up the demand for convenient hands-free communications. In such systems the user wish to lead a conversation in much the same way as in a normal person-to-person conversation. However, by installing the microphone far away from the user a number of disadvantages are introduced. These problems are mainly caused by room reverberation, noise and acoustic feedback.

Speech enhancement in hands-free telephony can be performed using spectral subtraction [1] or temporal filtering such as Wiener filtering, noise cancellation and multi microphone methods using a variety of different array techniques [2]. Room reverberation in this context is most effectively handled with array techniques or by proper microphone design and placement. Acoustic feedback for hands-free telephony is usually addressed by conventional echo cancellation techniques [3].

This paper introduces a new constrained subband adaptive beamformer as an alternative to the generalized side-lobe canceler, GSC [4]. All side-lobes are simultaneously suppressed by a soft constrained RLS type of algorithm, individually in each subband. The constraint is calculated from known source position(s) and a known array geometry. The benefit with the proposed method is small target cancellation effects.

The algorithm basically calculates the Wiener solution in each subband individually, where the spatial source auto-covariance matrix and the cross-covariance vector are pre-calculated, while background noise- and hands-free loud-speaker covariance matrices are estimated with the proposed recursive algorithm. Since information about the source position constitutes spatial covariance eigenvectors, it is possible to extend the use of the algorithm by introducing a subspace tracking algorithm [2], and thereby allow for source position tracking.

Simulations in a real car hands-free environment is presented. Results show a significant noise- and hands free-interference reduction within the traditional telephone bandwidth.

2. PROBLEM FORMULATION

We consider a wide band source located in the near-field of a uniform linear array with I microphones. Since the source is assumed to be a person speaking, it is modeled as a infinite number of point sources clustered closely in space within a range of radius $[R_a, R_b]$ and inside the range of angle of arrivals $[\theta_a, \theta_b]$. If $\mathbf{s}$ represents a received array data vector from a desired source having a power spectral density, PSD, $S(\Omega)$ with energy contained in the spectral band $[\Omega_a, \Omega_b]$, the spatial covariance matrix is given by

$$\mathbf{R}_s = \int\int\int_{R_a,\theta_a,\Omega_a}^{R_b,\theta_b,\Omega_b} S(\Omega)\mathbf{d}(R,\theta,\Omega)\mathbf{d}(R,\theta,\Omega)^H dR d\theta d\Omega \quad (1)$$

where the response vector is given by

$$\mathbf{d}(R,\theta,\Omega) = \left[\frac{1}{R_1}e^{-j\Omega\tau_1(R,\theta)}, \frac{1}{R_2}e^{-j\Omega\tau_2(R,\theta)}, \ldots, \frac{1}{R_I}e^{-j\Omega\tau_I(R,\theta)}\right] \quad (2)$$

with $\tau_i(R,\theta)$ denoting the time delay from a point source at radius R and angle θ to sensor i, and R_i is the distance between the source and sensor i.

The background noise statistics and angle of arrivals of distinct interference components are assumed to be unknown. It is often convention in a car hands-free installation to use the existing audio system for the far-end speaker. From this point of view we regard the hands-free speech as several unknown and coherent interference sources with unknown locations in the enclosure.

We consider a setup as illustrated in figure 1, where the constraint region denotes the locations in which the source should be contained. Errors in the response vector, e.g. caused by misplacement and gain variations of the microphones, affects the response in such way that small errors in the response vector causes large radial errors in the corresponding source location. The constraint region is defined as a pie slice region to accommodate for this relation (See figure 1).

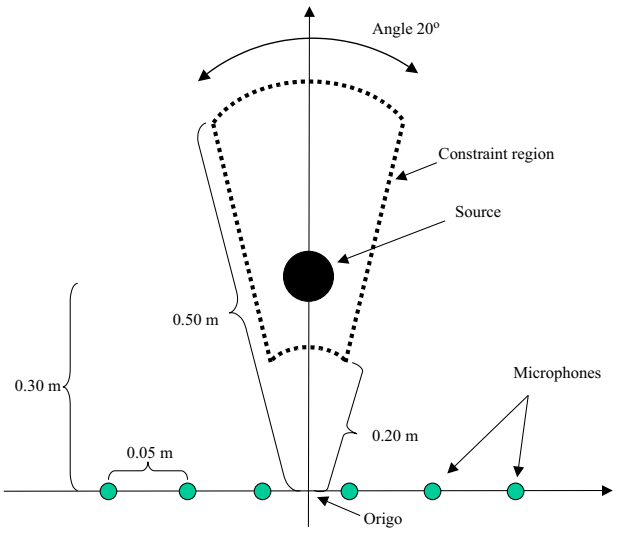

Fig. 1. *Microphone array geometry. The constraint region is pictured as the pie sliced region containing the speech source.*

2.1. Beamformer objective

The objective is formulated in the frequency domain[1] as a combination of least squares and Wiener solution. The source covariance matrix, obtained from a specified constraint region, is calculated as a free-field cluster of point sources, while the interference and noise covariance matrices are estimated from received data.

Given a known array geometry and a corresponding constraint region, our objective is to calculate

$$\mathbf{w}_{opt}^{(\Omega)} = \left[\mathbf{R}_s^{(\Omega)} + \hat{\mathbf{R}}_n^{(\Omega)} + \hat{\mathbf{R}}_j^{(\Omega)}\right]^{-1}\mathbf{r}_s^{(\Omega)} \quad (3)$$

where the array weight vector, $\mathbf{w}_{opt}^{(\Omega)}$, for frequency Ω is defined as

$$\mathbf{w}_{opt}^{(\Omega)} = [w_1^{(\Omega)}\ w_2^{(\Omega)}\ \ldots\ w_I^{(\Omega)}]^T \quad (4)$$

and the source covariance matrix is given by

$$\mathbf{R}_s^{(\Omega)} = \int\int_{R_a,\theta_a}^{R_b,\theta_b} S(\Omega)\mathbf{d}(R,\theta,\Omega)\mathbf{d}(R,\theta,\Omega)^H dRd\theta. \quad (5)$$

The noise covariance matrix, $\hat{\mathbf{R}}_n^{(\Omega)}$, and the interference (jammer) covariance matrix, $\hat{\mathbf{R}}_j^{(\Omega)}$, for frequency Ω are (theoretically) estimates from K samples of stationary received data when each component, noise and interference, are individually active

$$\hat{\mathbf{R}}_n^{(\Omega)} = \frac{1}{K}\sum_{k=1}^{K}\mathbf{x}_n(k)\mathbf{x}_n(k)^H \quad (6)$$

$$\hat{\mathbf{R}}_j^{(\Omega)} = \frac{1}{K}\sum_{k=1}^{K}\mathbf{x}_j(k)\mathbf{x}_j(k)^H. \quad (7)$$

The received array data vectors, $\mathbf{x}_n(k)$ and $\mathbf{x}_j(k)$, essentially contains frequency Ω, when noise and interference sources are active, respectively. The cross covariance vector, $\mathbf{r}_s^{(\Omega)}$, is given by the response vector and the source PSD

$$\mathbf{r}_s^{(\Omega)} = \int\int_{R_a,\theta_a}^{R_b,\theta_b} S(\Omega)\mathbf{d}(R,\theta,\Omega)dRd\theta \quad (8)$$

where the reference point for the beamformer response is defined at the origin of coordinates (See figure 1).

3. A RECURSIVE ALGORITHM

It is desirable to calculate the optimal beamforming weights according to Eq. (3) based on the available data continuously in a recursive way. Also, in order for the array response to be able to track variations in the surrounding environment, the covariance estimates include a forgetting factor. A total covariance matrix, $\mathbf{R}^{(\Omega)}$, for frequency Ω is introduced

$$\mathbf{R}^{(\Omega)} = \mathbf{R}_s^{(\Omega)} + \hat{\mathbf{R}}_n^{(\Omega)} + \hat{\mathbf{R}}_j^{(\Omega)} \quad (9)$$

[1]The representation is made on a finite grid that can be dense. This operation can be an FFT or a filter-bank transformation. The Wiener solution is only preserved if the transform domain produces independent subband signals.

where $\mathbf{R}_s^{(\Omega)}$ is the calculated source covariance matrix from Eq. (5), and where the noise and the interference covariance matrices, defined in Eqs. (6) and (7), are continuously weighted estimates of disturbing sound sources.

It is desired to update the total correlation matrix, $\mathbf{R}^{(\Omega)}$, recursively at each time index k, while maintaining the constant portion corresponding to the pre-calculated source covariance matrix, according to,

$$\mathbf{R}^{(\Omega)}(k) =$$

$$\mathbf{R}_s^{(\Omega)} + \lambda \left[\hat{\mathbf{R}}_n^{(\Omega)}(k-1) + \hat{\mathbf{R}}_j^{(\Omega)}(k-1)\right] + \mathbf{x}(k)\mathbf{x}^H(k) =$$

$$\lambda \mathbf{R}^{(\Omega)}(k-1) + \mathbf{x}(k)\mathbf{x}^H(k) + (1-\lambda)\mathbf{R}_s^{(\Omega)} \tag{10}$$

where λ is a weighting factor and where $\mathbf{x}(k)$ is the received array data vector. The effect of the above update is that the total correlation matrix is weighted and both the rank one "correction term," $\mathbf{x}(k)\mathbf{x}^H(k)$, and the small portion $(1-\lambda)$, of the pre-calculated source covariance matrix, which has been reduced by the weighting factor, are added. Since the pre-calculated source covariance matrix may be rank-deficient, the total correlation matrix is updated by adding scaled eigenvectors belonging to the signal space of the matrix [2]. This will result in several rank one updates as

$$\mathbf{R}^{(\Omega)}(k) = \lambda \mathbf{R}^{(\Omega)}(k-1) + \mathbf{x}(k)\mathbf{x}^H(k) + (1-\lambda)\sum_{p=1}^{P} \gamma_p \mathbf{q}_p \mathbf{q}_p^H \tag{11}$$

where γ_p is the p:th eigenvalue, and $\mathbf{q}_p$ is the p:th ordered eigenvector of the pre-calculated covariance matrix, $\mathbf{R}_s^{(\Omega)}$, and P is the dimension of the signal space, i.e. the effective rank of the matrix. The weighted optimal solution at sample instant k is now given by

$$\mathbf{w}_{opt}^{(\Omega)}(k) = \left[\mathbf{R}^{(\Omega)}(k)\right]^{-1} \mathbf{r}_s^{(\Omega)} \tag{12}$$

where $\mathbf{r}_s^{(\Omega)}$ is the cross covariance vector given in Eq. (8). The inversion of the matrix at each time instant is avoided by making use of the *Matrix-Inversion-Lemma*. One way to reduce the complexity further, at the expense of a small weight perturbation, is to sequentially add one scaled eigenvector at each sample instant in Eq. (11).

3.1. Summary of the Algorithm

The algorithm is stated as an iterative procedure, individually for each subband, indexed $m = 0, 1, \ldots, M-1$. The algorithm is run sequentially with the steps in the operation phase for each frequency $\Omega = 2\pi F_s m/M$, where F_s is the sampling frequency.

Initialization phase:

- Calculate the source covariance matrix and the cross covariance vector according to Eqs. (5) and (8)
- Calculate the eigenvalue decomposition of the source covariance matrix and store the eigenvalues and the eigenvectors
- Initialize the weight vector from Eq. (4) as a zero vector
- Define the inverse covariance matrix and initialize as $\mathbf{P}^{(\Omega)}(0) = \sum_{p=1}^{P} \gamma_p^{-1} \mathbf{q}_p \mathbf{q}_p^H$, and define the same size dummy variable matrix, $\mathbf{D}$.
- Choose a weighting factor λ and a weight smoothing factor α

Operation phase:

for $k = 1, 2 \ldots$

Update the inverse covariance matrix,

$$\mathbf{D} = \lambda \mathbf{P}^{(\Omega)}(k-1) - \frac{\lambda^{-2}\mathbf{P}^{(\Omega)}(k-1)\mathbf{x}(k)\mathbf{x}^H(k)\mathbf{P}^{(\Omega)}(k-1)}{1 + \lambda^{-1}\mathbf{x}^H(k)\mathbf{P}^{(\Omega)}(k-1)\mathbf{x}(k)}$$

$$\mathbf{P}^{(\Omega)}(k) = \mathbf{D} - \frac{\gamma_p(1-\lambda)\mathbf{D}\mathbf{q}_p\mathbf{q}_p^H\mathbf{D}}{1 + \gamma_p(1-\lambda)\mathbf{q}_p^H\mathbf{D}\mathbf{q}_p}$$

where $\mathbf{x}(k)$ is the received array data vector and index $p = k \pmod P$ denotes the index of the eigenvalues and eigenvectors given in Eq. (11).

For each sample instant, the weights are given by

$$\mathbf{w}^{(\Omega)}(k) = \alpha \mathbf{w}^{(\Omega)}(k-1) + (1-\alpha)\mathbf{P}^{(\Omega)}(k)\mathbf{r}_s^{(\Omega)}$$

and the output for frequency Ω is given by

$$y^{(\Omega)}(k) = \mathbf{w}^{(\Omega)}(k)^H \mathbf{x}(k). \qquad \square$$

A parameter α is introduced for weight smoothing and it corresponds to the real valued pole of a first order AR-model. The smoothing is used because the target speech signal adds spatially coherent power to the pre-calculated covariance matrix, and this in turn leads to small weight power fluctuations.

4. SIMULATIONS

4.1. Car Environment

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

4.2. Implementation

A uniform over-sampled DFT filterbank is used to decompose the received array signals into M subband signals. The filterbank is designed with the methodology described in [5], where transformation and reconstruction aliasing effects are minimized.

The integrals in Eq. (5) and Eq. (8) are solved by numerical integration, with the constraint region given in figure 1. The eigenvectors in Eq. (11) are found by SVD, and parameter P is chosen in such way that the eigenvalue spread is limited to 40 dB, for all subbands. This implies that less number of eigenvectors are used in low frequency subbands, since the rank of the corresponding matrices are smaller.

4.3. Results

In order to evaluate the beamformer a set of weights were calculated according to Eq. (3). A sequence of real background noise and hands-free speech were recorded individually and used to calculate the estimated covariance matrices given in Eqs. (6) and (7). Table 1. show suppression levels, normalized to the beamformer source signal gain, for different number of subbands in the structure.

No. of Subbands	Noise Supp.	Interference Supp.
$M = 16$	12.3	13.9
$M = 32$	14.8	15.0
$M = 64$	19.3	20.2
$M = 128$	17.7	18.5
	dB	dB

Table 1. *Suppression levels with different number of subbands.*

The algorithm was also run recursively as given in Sec. 3.1, when all sources were active as in a normal conversation, with 64 number of subbands. Figure 2 shows short time (20 ms) power averages of a single sensor observation, followed by the array output.

Experience show that a smaller constraint region gives better suppression, but at the same time a more noticeable target cancellation. This is related to how large the misplacement and gain variations of the microphones are. By additionally introducing a speech detector, which simply turns off the adaptation during target source periods, one may overcome these problems.

5. CONCLUSIONS

A new constrained adaptive subband beamformer have been presented. The solution consists in combining a pre-calculated spatial covariance matrix with estimated real environment

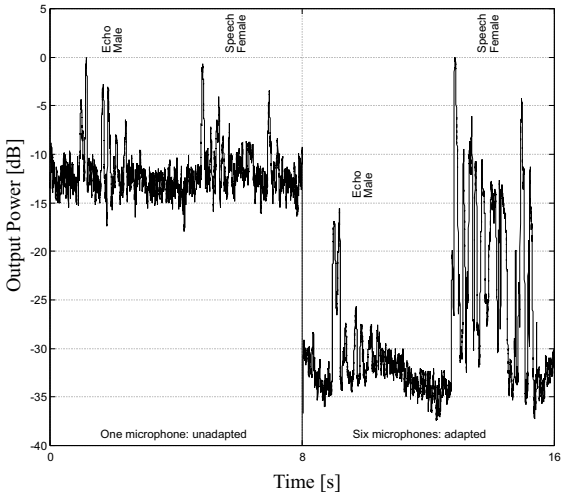

Fig. 2. *Short time (20 ms) power average of unprocessed single microphone observation followed by the beamformer output signal with number of subbands* $M = 64$, $\lambda = 0.99$, $\alpha = 0.01$.

covariance matrices. The algorithm recursively estimates the surrounding noise and interference statistics, while keeping the pre-calculated constraint as a constant part of the solution. A real car hands-free implementation with a linear array show very good noise and interference suppression.

6. REFERENCES

[1] J. R. Deller Jr., J. G. Proakis, and J. H. L. Hansen, *Discrete-Time Processing of Speech Signals*, Macmillan, 2000.

[2] C. Kyriakakis, P. Tsakalides, and T. Holman, "Surrounded by sound," *IEEE Signal Processing Magazine*, pp. 55–66, Jan. 1999.

[3] C. Breining, P. Dreiseitel, E. Hänsler, A. Mader, B. Nitsch, H. Puder, T. Schertler, G. Schmidt, and J. Tilp, "Acoustic echo control, an application of very-high-order adaptive filters," *IEEE Signal Processing Magazine*, pp. 42–69, Jul. 1999.

[4] 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.

[5] 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.