Microphone Array Designed for a Conference Room

Anders Johansson, Sven Nordholm, Erik Östlin

Australian Telecommunications Research Institute,
Curtin University of Technology,
Wark Ave. Bentley, WA Australia

Abstract

Conference telephony is very much dependent on hands-free operation because it offers convenience, flexibility, and ease of use. In many conference rooms the background noise and reverberation level can be quite high. It is usually not enough to use a single microphone to obtain a desired quality. However, an array of microphones will be able to suppress reverberations, background noise and near-end echoes. This paper shows a study performed using real data from ATRI's board room. The results show a potential suppression of background noise and disturbing sources of more than 10 dB over the whole frequency range. The system is calibrated and designed in-situ using Matlab and a Texas Instruments based DSP-system. Different filter-lengths have been studied ranging from 16-256 long and a 6 microphone element array have been used.

1. Introduction

Conference telephony and video conference are growing areas of communication. It is important to maintain a good sound quality even if the system is operated in hands-free mode. Another important aspect is ease of use. Thus using close talk microphones or headsets for each user would be inconvenient. The two key advantages of hands-free communication are convenience and flexibility. In a hands-free environment, microphones are placed at a remote distance from the speakers, [1], [2], [3], [4], [5]. The loudspeaker/loudspeakers are placed and the volume is set to a level such that the sound will be clearly audible to all the users. However, this means that the microphones will pick-up background noise as well as the signal from the loudspeaker, thus near-end echo. Without suitable signal processing the result will be poor sound-quality and acoustic feedback for the far-end user. The quality of the speech signal will be affected by the room acoustics. To obtain sound quality similar to or better than obtained using hand-held telephone sets, filtering is required to suppress the loudspeaker signal, as well as background noise and room reverberation, without causing speech distortion. Furthermore, hands-free equipment must handle full-duplex operation. Future voice communication will also include stereo information which invokes new problems for echo cancellation [6]. This will add considerably more complexity and it has not been considered in this study.

The procedure proposed in this paper is based on a multi-dimensional filter design. The idea has been to gather calibration signals in the room from positions corresponding to the speaker positions and loudspeaker positions. From these calibration signals an LS optimization problem is formulated and solved. This gives FIR filter weights which are used in the DSP-system and will provide a multi-dimensional filtering given the calibration data. By doing so one can get calibrated responses from different speaker localizations. The calibration signals used are a broad-band noise, from a number of positions in the room, which are corresponding to the speakers positions, the obtained filters are optimized for the given positions. The procedure must be connected to a localization procedure, depending on which speaker that is active. This is currently investigated.

2. Signal Model

The calibration signals are gathered from 12 positions according to Fig. 1. The different signal sources are denoted $s_m(t), m = 1, 2, \ldots, M$ and are assumed to be mutually uncorrelated. These signals impinge on an array of I microphone elements, each corrupted with noise $\eta_i(t)$. This noise can be considered to contain electronic noise and background noise from air-conditioning etc. The transfer function between source no. m and array element no. i is denoted $G_{m,i}(\omega)$. The input signals, $x_i(t)$ are

$$x_i(t) = \sum_{m=1}^{M} s_m(t) * g_{m,i}(t) + \eta_i(t) \quad ; i = 1, 2, \ldots, I \quad (1)$$

In the sequel all signals are assumed to be band limited and sampled.

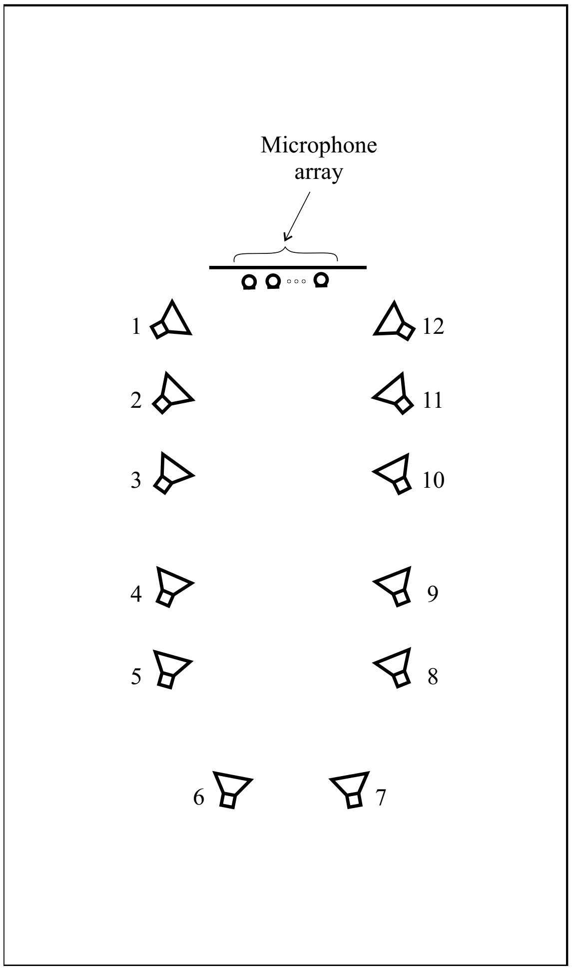

Figure 1: Measurement setup.

3. Theory

A broadband spatial temporal filter has been used to solve the multi-dimensional filtering problem, a schematic diagram can be seen in Fig. 2. The output, $y(n)$ from the filter is given by

$$y(n) = \sum_{i=1}^{I} \sum_{j=0}^{L-1} w_i(j) x_i(n-j) \tag{2}$$

where L is the number of filter taps.

A new set of weights is designed for each source position. The desired signal $d(n)$ used for the design is obtained from the calibration signal $s_m(n)$ by filtering it with $f(n)$. In this study only a delay were used, but if some reverberation in the desired signal is requested this can be achived by choosing an appropriate filter. The delay consists of the actual delay between the loudspeaker and the microphone array (τ, no. of

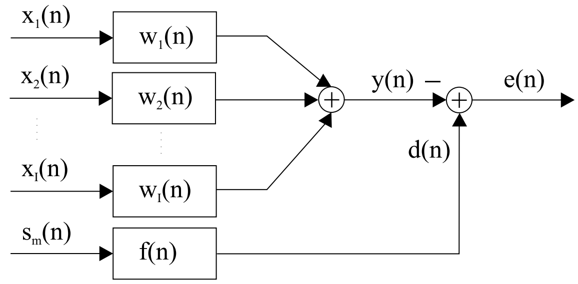

Figure 2: The broadband spatial temporal filter.

samples) and half the filter-length ($L/2$, no. of samples).

When designing the spatial temporal filter the mean square error has been minimized as

$$\mathbf{w}_{opt} = \arg\min_{\mathbf{w}} \frac{1}{N} \sum_{n=1}^{N} |d(n) - y(n)|^2 \tag{3}$$

where parameter N is the total number of time samples used in the averaging. The resulting filters are conventional FIR-filters where the weights are found by solving the normal equations.

The total correlation matrix $\mathbf{R_{xx}}$, when formulating the normal equations, is defined as

$$\mathbf{R_{xx}} = \begin{pmatrix} \mathbf{R}_{\mathbf{x}_1\mathbf{x}_1} & \mathbf{R}_{\mathbf{x}_1\mathbf{x}_2} & \dots & \mathbf{R}_{\mathbf{x}_1\mathbf{x}_I} \\ \mathbf{R}_{\mathbf{x}_2\mathbf{x}_1} & \mathbf{R}_{\mathbf{x}_2\mathbf{x}_2} & \dots & \mathbf{R}_{\mathbf{x}_2\mathbf{x}_I} \\ \vdots & \vdots & \ddots & \\ \mathbf{R}_{\mathbf{x}_I\mathbf{x}_1} & \mathbf{R}_{\mathbf{x}_I\mathbf{x}_2} & \dots & \mathbf{R}_{\mathbf{x}_I\mathbf{x}_I} \end{pmatrix} \tag{4}$$

where

$$\mathbf{R}_{\mathbf{x}_i\mathbf{x}_j} = \begin{pmatrix} r_{ij}(0) & r_{ij}(1) & \dots & r_{ij}(L-1) \\ r_{ij}(1) & r_{ij}(0) & \dots & r_{ij}(L-2) \\ \vdots & \vdots & \ddots & \vdots \\ r_{ij}(L-1) & r_{ij}(L-2) & \dots & r_{ij}(0) \end{pmatrix} \tag{5}$$

and

$$r_{ij}(k) = \frac{1}{N} \sum_{n=0}^{N-k-1} x_i(n) x_j(n+k) \quad ; k = 0, 1, \dots, L-1 \tag{6}$$

which is the cross-correlation estimator.

The filter vectors $\mathbf{w}_i$, where $i = 1, 2 \dots I$, are arranged as a column-vector

$$\mathbf{w} = \left[\begin{array}{cccc} \mathbf{w}_1^T & \mathbf{w}_2^T & \dots & \mathbf{w}_I^T \end{array} \right]^T \tag{7}$$

and the cross-correlation vector, the correlation between the desired signal $d(n)$ and the input signals $\mathbf{x}_i(n)$, where $i = 1, 2 \dots I$, are arranged as

$$\mathbf{r}_{\mathbf{x}d} = \left[\begin{array}{cccc} \mathbf{r}_{\mathbf{x}_1 d} & \mathbf{r}_{\mathbf{x}_2 d} & \dots & \mathbf{r}_{\mathbf{x}_I d} \end{array} \right]^T \tag{8}$$

where

$$\mathbf{r}_{\mathbf{x}_i d} = \begin{bmatrix} r_{x_i d}(0) & r_{x_i d}(1) & \ldots & r_{x_i d}(L-1) \end{bmatrix}^T \quad (9)$$

and

$$r_{x_i d}(k) = \frac{1}{N} \sum_{n=0}^{N-k-1} x_i(n) d(n+k) \qquad ; k = 0, 1, \ldots, L-1 \quad (10)$$

The correlation matrices and vectors are averaged from the broadband noise calibration signals, using raw correlation. The length of the recordings (no. of samples) are $> 10IL$.

The spatial temporal filtering weights are found from the solution of the least squares problem.

$$\mathbf{w} = \mathbf{R_{xx}}^{-1} \mathbf{r}_{\mathbf{x}d} \quad (11)$$

A new set of weights is found for each desired source.

If there is a fixed noise/interference source in a known direction, then we can force attenuation in that direction by adding the correlation matrix, $\mathbf{R}_{nn}$, to $\mathbf{R}_{xx}$. $\mathbf{R}_{nn}$ is estimated in the same way as $\mathbf{R}_{xx}$. The optimal least square solution to the problem becomes

$$\mathbf{w} = [\mathbf{R_{xx}} + \mathbf{R_{nn}}]^{-1} \mathbf{r}_{\mathbf{x}d} \quad (12)$$

thus one will be able to handle interference.

4. Filter Design Procedure

The calibration and filter design procedure is given as:

- Gather signals from the 12 positions, see Fig. 1, and save the data from source m into a data matrix $\mathbf{X}_m$, synchronously gather the signal $s_m(n)$ into a vector, $\mathbf{s}_m$.
- Select the signals used in the filter design. Thus the source of interest, m, determines which of the calibration signals that should be used.
- Create $\mathbf{R_{xx}}$ and $\mathbf{r}_{\mathbf{x}d}$ and solve for $\mathbf{w}$. This is done off-line in Matlab. After the design is performed the weights are downloaded to the DSP for processing.

5. Hardware Description

The DSP used is a Texas Instruments TMS320C6701 on a Spectrum Signal Processing Daytona PCI board, plugged into a PC. The data acquisition board is a General Standards Corporation PMC-ADADIO which communicates with the DSP via a local PCI bus on the Daytona board. The ADCs and DACs on the data acquisition board uses 16-Bit resolution and 8kHz sample frequency. The microphone system from Larson-Davis uses microphones of 1/2" free field type and the array spacing was 40mm. The realtime software running in the DSP is written in assembly and C and the PC front-end for the DSP program is written in LabVIEW.

6. Results

A comparison between two strategies has been done. Two sets of spatial temporal filters were designed from the calibration recordings.

- Including only desired signal and one disturbance when forming $\mathbf{w}$.
- Desired signal and all disturbances are used when forming $\mathbf{w}$.

Speaker position number 3 was used as desired in both designs and the speaker in position 9 was used as disturbance in the first design. The two sets contains filters ranging from 16 to 256 taps in length. The resulting filters were loaded into the DSP-system for evaluation. The performance was recorded by measuring the output signal using a dynamic signal analyzer.

Above filter designs were used to perform two measurements. The first measurement compares the suppression of speaker number 9 relative to speaker number 3 for different filter-lengths and both designs. The second measurement compares the suppression of many positions relative speaker number 3 for the filter length 256 and both designs. The suppression ratio is calculated as

$$\mathit{Suppression\ Ratio}\,[dB] = 10 \log \left(\frac{\int_{-\pi}^{\pi} \hat{P}_{y_3}(\omega) d\omega}{\int_{-\pi}^{\pi} \hat{P}_{y_m}(\omega) d\omega} \right) \quad (13)$$

where $\hat{P}_{y_m}(\omega)$ and $\hat{P}_{y_3}(\omega)$ denote the spectral densities for the disturbing and the desired location respectively, measured at the output of the DSP-system. The index $m \in [1, M]$ denotes the position of the disturbance. The results from the measurements are shown in Figs. 3 and 4.

From the first measurement it can be seen, see Fig. 3, that there is little to be gained by using filters longer than 64 taps.

The second measurement, see Fig. 4, shows that the designs suppressing only one position will perform poor in any other position. Figs. 5 and 6 also show the suppression for different frequencies for the same measurement.

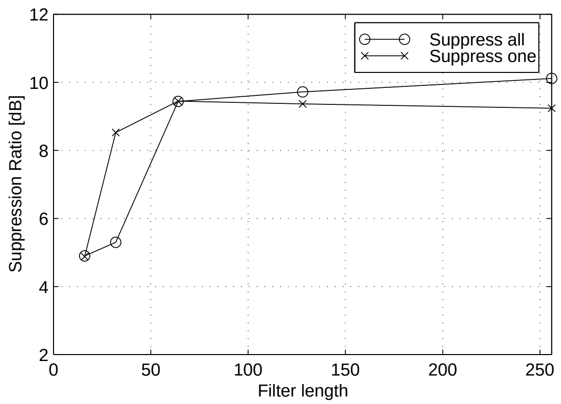

Figure 3: Average suppression versus filter length.

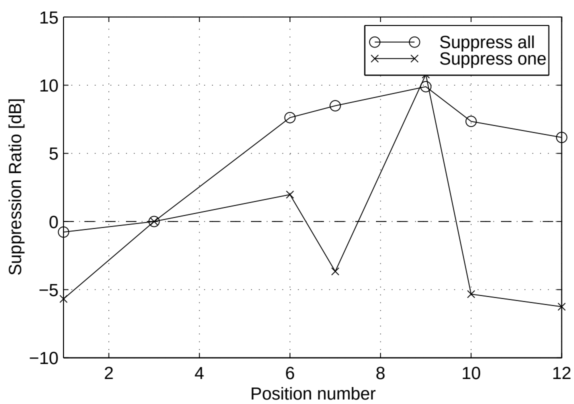

Figure 4: Average suppression versus position, filter length 256 taps.

7. Conclusions and future work

A microphone array has been designed using calibration data from 12 different positions. One source is considered as desired and the other sources should be suppressed. It is possible to suppress undesired background noise as well as undesired sources by 10 dB. The real-time implementation show some degradation compared to Matlab implementation but this may be improved by using better reconstruction filters.

Next phase would be to integrate source localization and adaptive processing of the array, as well as stereo sound.

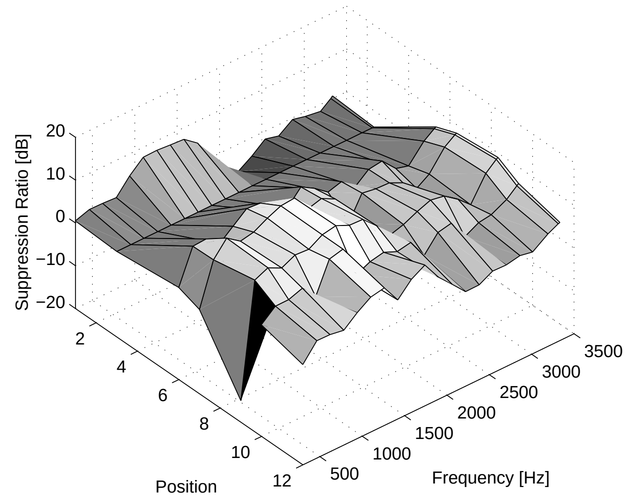

Figure 5: Suppression for all positions versus position and frequency, filter length 256 taps.

References

[1] H.F. Silverman, W.R. Patterson, J.L. Flanagan, D. Rabinkinn, "A digital processing system for source location and sound capture by large microphone arrays", Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing, pp. 251-254 vol.1, 1997.

[2] J. Flanagan, D. Berkeley, G. Elko, J. West, M. Sondhi,"Autodirective Microphone Arrays", Acustica, vol 73, pp. 58-71, 1991.

[3] H.F. Silverman, W.R.-III Patterson, J.M. Sachar, "First measurements of a large-aperture microphone array system for remote audio acquisition", Proceedings of the 2000 IEEE International Conference on Multimedia and Expo. ICME2000, pp. 823-826 vol.2, 2000.

[4] W. Kellermann, "Strategies for combining acoustic echo cancellation and adaptive beamforming microphone arrays", Proceedings of the 1997 IEEE International Conference on Acoustics, Speech, and Signal Processing, pp. 219-222, 1997.

[5] S. Nordholm, I. Claesson, M. Dahl,"Adaptive Microphone Array Employing Calibration Signals. An Analytical Evaluation.", IEEE Transaction on Speech and Audio Processing, vol. 7, no. 3, pp. 241-252, May 1999.

[6] J. Benesty, D. Morgan, M. Sondhi, "A Better Understanding and an Improved Solution to the Specific Probelems of Stereophonic Acoustic Echo Cancellation", IEEE Transactions on Speech and Audio Processing, vol. 6, no. 2, pp. 156-165, March 1998.

[7] N. Grbić, S. Nordholm, I. Claesson, "Optimal and Adaptive Beamforming for Speech Signals in a mixture of Spatially Coherent and Incoherent Noise

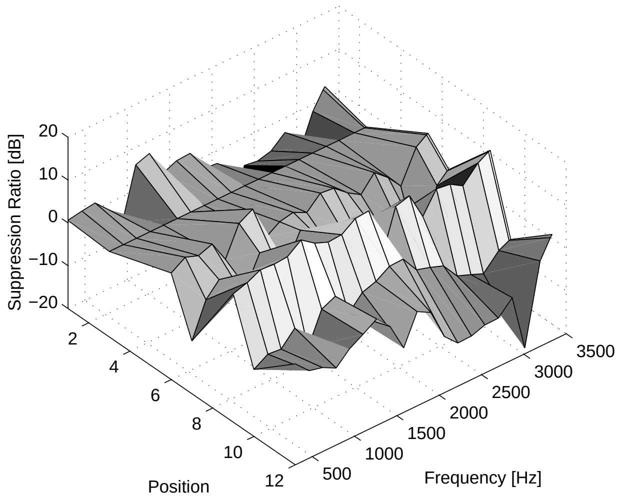

Figure 6: Suppression for one position versus position and frequency, filter length 256 taps.

Fields", Submitted for publication IEEE transactions on Speech and Audio Processing.

[8] S. Nordholm, I. Claesson, N. Grbić, "Optimal and Adaptive Michrophone Arrays for Speech Input in Automobiles", Submitted for publication IEEE transactions on Speech and Audio Processing.

[9] B. D. van Veen, K. M. Buckley, "Beamforming: A Versatile Approach to Spatial Filtering", IEEE ASSP Magazine, April 1988.

[10] S. Nordholm, I. Claesson, "Adaptive Microphone Array Employing Calibration Signals. An Analytical Evaluation.", Research Report 7/95, ISSN 1103-1581, ISRN HKR-RES–95/7–SE, July 1995.