WIDEBAND ADAPTIVE BEAMFORMING SYSTEM FOR SPEECH RECORDING

Sergey Timofeev, Ahmad Bahai and Pravin Varaiya

Department of Electrical Engineering and Computer Science
University of California, Berkeley, CA, USA

ABSTRACT

An adaptive beamforming algorithm that enhances the quality of the signal recorded by a microphone array is presented. The system described here attenuates the interference from spatially scattered sources while preserving the desired signal quality in the speech frequency range. A novel method for defining additional constraints within the adaptation process and performing beamforming in parallel across several frequency bands achieves undistorted recording even with slight errors in localizing the speaker. Simulation results show nearly optimal improvement of speech quality in various noisy environments.

The algorithm is modified to perform in a real environment by introducing two modes of adaptation that switch according to the speaker activity. The implementation of this modified system demonstrates SNR gain close to that of the original simulated system. It is robust to localization errors and non-ideal environmental effects.

Index Terms— Array signal processing, speech processing, adaptive signal processing, acoustic arrays

1. INTRODUCTION

This paper describes a hands-free voice recording system that employs adaptive acoustic beamforming to enhance the quality of the recorded speech in noisy environments.

The speaker who is the user of the system and whose voice is recorded will be referred to as "desired speaker" or "desired source" for the remainder of this paper. The desired speaker's location is assumed to be estimated by the same microphone array that performs the beamforming. All other acoustic signals present in the environment are regarded as interference. In a typical situation the interference is not distributed uniformly in space around the speaker, but has few point sources, whose locations are unknown to the recording device. This prompts the usage of an adaptive beamforming structure that adjusts itself to suppress the strongest interference sources.

This problem falls in the realm of wideband beamforming, as the recorded speech bandwidth is designed to match that of the telephone line: 200 - 3400 Hz [1]. Previous work on time-domain beamforming algorithms has either focused on the narrowband signals (mainly with applications to radar) [2, 3], or produced solutions that do not employ adaptive interference reduction [4, 5]. The main challenge in wideband beamforming is consistency of the beamformer characteristics in the entire frequency band of interest. Narrowband adaptive beamforming solutions do not perform well for wideband signals for several reasons. First, the interference has both frequency and spatial dimensions, which requires more degrees of freedom from the adaptive filter to cancel it. Second, even if a desired beam pattern is achieved at a certain frequency, it quickly deteriorates once the frequency is changed. This paper describes a system that yields consistent operation in the entire frequency range, which employs adaptive interference cancellation techniques to provide high attenuation to the point interference sources.

The approach presented in this paper relies on several previous results. The iterative adaptive algorithms for time-domain beamforming structures have been studied [2, 6]. A method of placing additional constraints on the beamformer response at points in space and frequency has been developed by Buckley [7]. The method of simulating the desired signal to drive the adaptation has appeared in [8].

2. WIDEBAND ADAPTIVE BEAMFORMER

2.1. Signal Modeling

We consider an environment with one desired signal $s(t)$ and P interference signals $i_p(t),\quad p = 1,..,P$. These signals are being recorded by a planar array of K microphones. We assume a situation where the received reflections of the desired signal are insignificant, and there is only one strong direct propagation path from the desired source to the microphones. This assumption is shown to be valid in the experiments described in section 4. The continuous-time signal received by the k^{th} microphone is given by:

$$r_k(t) = s(t - \tau_{s,k}) + \underbrace{\sum_{p=1}^{P} i_p(t - \tau_{p,k})}_{I_k(t)}, \tag{1}$$

in which $\tau_{s,k}$ and $\tau_{p,k},\ p = 1,..,P$ are the propagation delays to the k^{th} microphone from the desired source and interferers respectively. The signal strength is taken to be the same at every microphone. After sampling at frequency F_s the recorded signals are

$$x_k(n) = s_k(n) + I_k(n), \quad k = 1,..,K, \tag{2}$$

in which $s_i(n)$ and $s_j(n),\quad i,j = 1,..,K$ differ from each other only by a relative delay $F_s(\tau_{s,i} - \tau_{s,j})$ in samples, which is not necessarily an integer. Two assumptions are made:

1. Direction of the desired source, and thus the relative delays, are known; and

2. The desired signal and interference are statistically uncorrelated.

2.2. Adaptive LCMV beamformer

The system is based on the Linearly Constrained Minimum Variance (LCMV) structure shown in Fig. 1. The steering delay block compensates the relative propagation delay from the source to the microphones, such that $s_i(n) = s_j(n)\ \forall\, i,j$ in equation (2). The steering delays are in general fractional, and can be split into two parts: the integer part $\lfloor\Delta_k\rfloor$ and fractional part $\delta_k = \Delta_k - \lfloor\Delta_k\rfloor$.

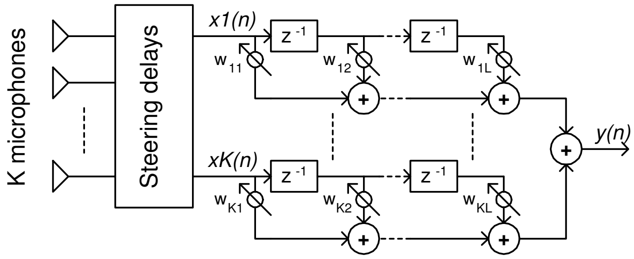


Fig. 1. Block diagram of the LCMV beamformer

The first delay is implemented by a conventional delay line, and an all-pass filter with the appropriate group delay is used for the fractional component.

We define input and weight vectors as follows:

$$\mathbf{X}(n) = [x_1(n), .., x_K(n), .., x_1(n-L+1), .., x_K(n-L+1)]^T \tag{3}$$

$$\mathbf{W} = [w_{1,1}, .., w_{K,1}, .., w_{1,L}, .., w_{K,L}]^T \tag{4}$$

Then the output $y(n) = \mathbf{W}^T\mathbf{X}(n)$. We want to choose such $\mathbf{W}$ that minimizes the output power $E[y(n)^2] = \mathbf{W}^T\mathbf{R}_{XX}\mathbf{W}$ under a set of linear constraints $\mathbf{C}^T\mathbf{W} = \mathbf{f}$. The component of $\mathbf{X}(n)$ due to the desired signal (which after steering is synchronized on all input lines) must be unchanged by the filter, which means

$$\sum_{i=1}^{K} w_{i,j} = \begin{cases} 1, & j = 1 \\ 0, & j = 2, .., L \end{cases} \tag{5}$$

Equation (5) defines some of the columns of the constraint matrix $\mathbf{C}$, while others are defined by the wideband constraints described in the next section. The solution to the optimization problem is

$$\mathbf{W}_{opt} = \mathbf{R}_{XX}^{-1}\mathbf{C}[\mathbf{C}^T\mathbf{R}_{XX}^{-1}\mathbf{C}]^{-1}\mathbf{f}. \tag{6}$$

This method of obtaining the coefficient vector is impractical for several reasons. First, it assumes knowledge of $\mathbf{R}_{XX}$, the autocorrelation matrix of the input vector $\mathbf{X}(n)$. $\mathbf{X}(n)$ therefore must be wide-sense stationary, which in case of speech is true for approximately 30 ms of the signal. Second, the K spatial channels are strongly correlated, resulting in $\mathbf{R}_{XX}$ that is close to singular, yielding incorrect matrix inverses.

Instead, $\mathbf{W}_{opt}$ is approximated iteratively, based on the constrained LMS algorithm developed by Frost [2]:

$$\begin{aligned} \mathbf{F} &\triangleq \mathbf{C}[\mathbf{C}^T\mathbf{C}]^{-1}\mathbf{f} \\ \mathbf{P} &\triangleq \mathbf{I} - \mathbf{C}[\mathbf{C}^T\mathbf{C}]^{-1}\mathbf{C}^T \\ \mathbf{W}(0) &= \mathbf{F} \end{aligned} \tag{7}$$

$$\mathbf{W}(n+1) = \mathbf{P}[\mathbf{W}(n) - \frac{\mu y(n)\mathbf{X}(n)}{\|\mathbf{X}(n)\|^2})] + \mathbf{F}. \tag{8}$$

Both $\mathbf{X}(n)$ and output $y(n)$ are speech signals, so their amplitude varies significantly over time. For this reason the normalized LMS (NLMS) algorithm is used here to ensure uniform adaptation rate.

2.3. Wideband constraints

The mainlobe width of a beamformer decreases with increasing frequency. This means that if the desired source direction is estimated erroneously, high frequency content of the signal will be regarded as interference by the beamformer and attenuated significantly. The proposed algorithm employs additional linear constraints on $\mathbf{W}$ to create uniform beamwidth across the entire frequency range. A pair of constraints fixes the beamformer response in one particular direction at one frequency [7]. This pair can be expressed as

$$[\mathbf{c}(\bar{\omega}), \quad \boldsymbol{s}(\bar{\omega})]^T\mathbf{W} = [A\cos(\alpha), \quad A\sin(\alpha)]^T, \tag{9}$$

in which $\mathbf{c}(\bar{\omega})$ and $\boldsymbol{s}(\bar{\omega})$ are vectors of cosines and sines respectively, whose arguments depend on the difference between the propagation delays for the constraint direction and the steering delays of the beamformer, as well as the frequency at which the constraint is implemented.

To fix the width of the beam at a particular frequency, several directions close to steering are picked, and the beamformer gain is fixed to be 1 at those directions. Based on the computed gain patterns, a configuration was chosen in which three directions around steering are used, each forming an angle of 3° with the steering direction, placed uniformly around it. Suppose the unit-magnitude steering vector $\mathbf{v}$ is represented as $(1, \theta, \phi)$ in spherical or $[\sin(\theta)\cos(\phi), \sin(\theta)\sin(\phi), \cos(\theta)]^T$ in Cartesian coordinates. Let the three extra constraint directions be $\mathbf{a}$, $\mathbf{b}$ and $\mathbf{c}$. Vector $\mathbf{a}$ is chosen as follows: $\mathbf{a} = [\sin(\theta+3^\circ)\cos(\phi), \sin(\theta+3^\circ)\sin(\phi), \cos(\theta+3^\circ)]^T$. Vectors $\mathbf{b}$ and $\mathbf{c}$ are computed as follows:

$$\mathbf{b} = \mathcal{R}_{\mathbf{v},\frac{2\pi}{3}}\mathbf{a}; \quad \mathbf{c} = \mathcal{R}_{\mathbf{v},\frac{2\pi}{3}}\mathbf{b} \tag{10}$$

where $\mathcal{R}_{\mathbf{v},\frac{2\pi}{3}}$ is the 3×3 rotation matrix by angle $\frac{2\pi}{3}$ around the axis $\mathbf{v}$ in $\mathbb{R}^3$. Then the propagation delays are computed for directions $\mathbf{a}$, $\mathbf{b}$ and $\mathbf{c}$ to construct additional constraint vectors as described in equation (9), with parameters $A = 1$ and $\alpha = 0$.

Each such set of constraints controls the beamwidth at a certain frequency. Experiments demonstrate that for consistent beamwidth across the entire band these constraints need to be placed at 500 Hz intervals for frequencies up to 2 kHz and at 300 Hz intervals for higher frequencies. However, placing more constraints reduces the degrees of freedom of the coefficient vector $\mathbf{W}$, which decreases the interference reduction capabilities of the beamformer. This issue is addressed by designing a bank of bandpass filters to split the input into frequency subbands, each having its own beamformer with the wideband constraints covering only its respective subband. Four subbands, each 1 kHz wide, were used in the simulation of the algorithm.

3. SIMULATION OF THE ALGORITHM

A simulation of the proposed algorithm with $K = 8$ microphones and $L = 20$ taps was performed. Far-field propagation model was assumed. The observed performance of several aspects of the system is presented below.

3.1. Interference reduction

The desired signal was set to be a segment of previously recorded speech, and the interference was band-limited white noise with bandwidth of about 200 Hz and center frequency of 1.5 kHz. Performance was measured by computing the MSE between the beamformer output and the desired signal and comparing it to the power of the added interference signal. Applying the proposed adaptive algorithm resulted in interference reduction of 18 dB on average, with the worst performance (10 dB reduction) at the onset of voiced speech segments when the amplitude of the signal and its statistics change

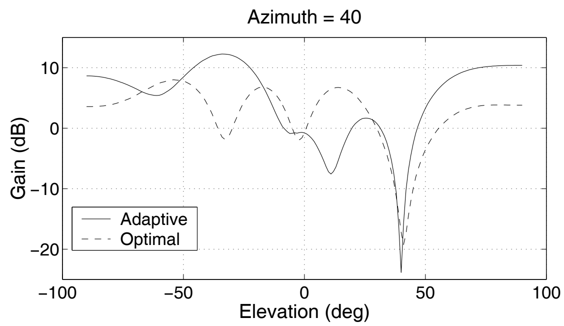


Fig. 2. Spatial response at 1.5 kHz, with the interference at azimuth 40° and elevation 40°

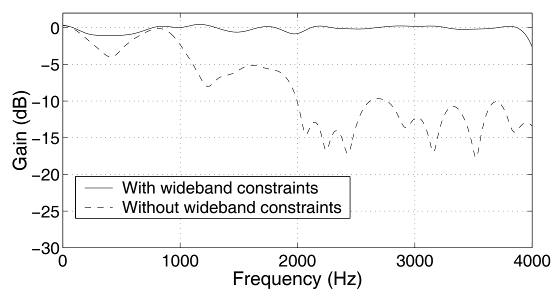


Fig. 3. Frequency response to the off-steering desired signal

rapidly, and the best performance (20 dB reduction) during pauses in desired speech when the received signal only consisted of interference. The optimal solution (6) produced interference reduction of 21 dB.

The gain patterns of the adaptive and optimal beamformers are shown in Fig. 2. There is a null at the direction of the interference (elevation 40°). However, the response at angles which have no interference source present shows less attenuation or even amplification of the signal. The LCMV beamformer attempts to minimize the global interference energy, which is unaffected by amplification of the regions where interference is absent, and thus the response is unconstrained in these regions. Interference reduction was found to be virtually independent of how close the interference and steering directions are, as long as the interference is outside of the main beam: the difference in performance between elevations 10° and 90° is about 1 dB.

The algorithm was also simulated with different types of interference: white noise was reduced by 17 dB, and using a sum of sinusoids as the interference signal produced gain of 19 dB.

3.2. Wideband performance

Wideband constraints have been implemented to ensure that the beamformer does not treat signals that are slightly off steering as interference. Frequency response to a signal arriving at 3° away from the steering direction is shown in Fig. 3 for the beamformers with and without wideband constraints.

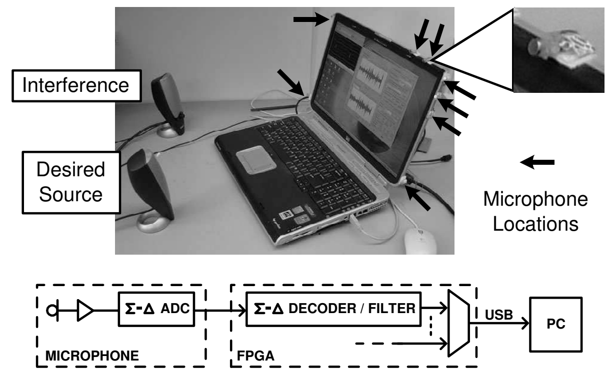


Fig. 4. Platform for voice enhancement

4. IMPLEMENTATION OF THE ALGORITHM

4.1. Platform description

The platform for algorithm implementation was chosen to be a laptop, in which the digital microphones are mounted along the edges of the screen. Acoustic data are sampled at 16 kHz and the processing is done by the laptop. Fig. 4 depicts the system configuration. There are several reasons for the asymmetric microphone placement: avoidance of spatial aliasing dictates that at least one pair of microphones must be within 5 cm of each other (in both vertical and horizontal directions for a planar array), but the beamformer performs better at low frequencies if the microphones are far apart.

4.2. Non-ideal environment

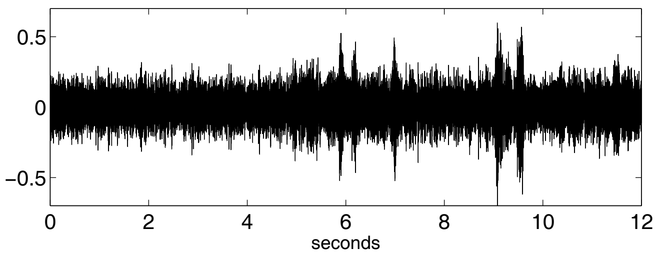
The main difficulty that one faces in this testing environment is the variation of the microphone response. This variation is due to both the internal microphone parameters and the location of the microphone. The far-field model does not hold in the test setting, since the sources of sound are within 1 meter of the laptop screen. In addition to varying signal attenuation due to near-field propagation paths, the desired signal is distorted differently at each microphone. Under these conditions the desired signal can no longer be preserved by the constraint (5), causing it to be suppressed as interference. Several modifications needed to be made to the algorithm before its performance was acceptable.

4.3. Modifications to the algorithm

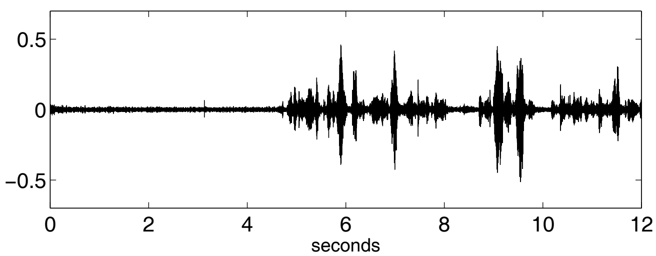
The near-field propagation model is used. Amplitude attenuation of sound is assumed to be inversely proportional to the distance it travels from the source. The microphones have been calibrated to equalize their internal gain. The frequency response differences were not considered due to the difficulty of their estimation in the presence of channel effects. Additionally a high-pass filter with 200 Hz cutoff is applied to the input signals to suppress the strong interference at 60 Hz and its weaker harmonics at 120 Hz and 180 Hz.

The main modification is applied to the beamforming algorithm. In order to avoid cancellation of the desired signal during adaptation, two modes of adaptation are introduced. The weight update (8) is performed when the desired signal is inactive, and only the interference sources are present. In this first mode the presence of the desired signal is simulated by generating a random signal and adding it to all channels after steering. This simulated desired signal is subtracted from the output before the it is fed back into the NLMS update. The mean of each coefficient is recorded; these settle to constant values as the filter converges. When the desired signal becomes active, the adaptation switches to the second mode, wherein there is no simulated desired signal, and the coefficients are still updated according to (8), but are only allowed to vary within a narrow range around their mean. This lets the beamformer track small variations in local interference statistics, but the response of the beamformer does not change enough to cancel the desired signal. This also eliminates the need for wideband constraints. This method performs well as long as the interference remains statistically and spatially stationary during the second mode.

The results of the algorithm testing have shown that desired signals originating from a 40° sector around the steering direction experience no noticeable distortion. The space around the beamformer is then divided into sectors with associated sets of steering delays that have been precalculated and stored in memory. This calculation was done by estimating the time-of-arrival difference based on cross-correlation of received signals when the desired source was placed in the center of the sector.



(a) Recorded by one of the microphones



(b) Processed by adaptive beamformer

Fig. 5. Processed recorded signal with bandlimited noise interference

4.4. Performance

As a consequence of the near-field model different microphones will have different SNRs. Therefore we choose to compare the proposed algorithm's performance to the performance of the delay-sum beamformer, where every output sample is the average of all 8 input samples. The delay-sum beamformer by itself improves the SNR, but usually only 1-2 dB above that of the best microphone. All the gain values that appear in the rest of this section take the delay-sum result as 0 dB.

The modified algorithm was tested in a typical room environment. Two sources of sound (computer speakers) were used to generate the desired signal and interference. First the interference signal was activated, and shortly thereafter, the desired source. In the processing stage the first mode of adaptation was performed on the initial 3 seconds of the signal, and during the rest of the signal the second mode was used.

The time-domain plots of the desired speech signal with bandlimited white noise are shown in Fig. 5. The SNR improvement for the adaptive beamformer was measured to be 19 dB. Filtering the same signal with the computed optimal weight vector (6) resulted in SNR improvement of 20 dB.

Various other types of interference were tested. The performance was measured to be 4 dB better if the interference is a sum of sinusoids, and 5 dB worse if the interference is white noise.

5. CONCLUSION

An adaptive wideband beamforming algorithm was developed for the recording of speech signals. The algorithm is able to attenuate point interference sources by 18-21 dB without prior knowledge of the interference location and characteristics, achieving nearly optimal performance. The beamformer achieves consistent beamwidth across the entire frequency range of interest.

The modified algorithm employing two modes of adaptation achieved significant speech quality improvement in real environments. Desired signal within a 40° sector was preserved, and interference was reduced by 14-23 dB, which approached optimal performance. The modified algorithm is robust to non-ideal conditions, but in turn requires knowledge of desired speaker activity to switch between the adaptation modes.

6. ACKNOWLEDGEMENT

The authors express their appreciation to National Semiconductor for providing funds and equipment, as well as to the engineers there, most notably Dr. Wei Ma and Lin Sun, for sharing their expertise and helping resolve hardware-related issues.

7. REFERENCES

[1] B. Gold and N. Morgan, *Speech and Audio Signal Processing.* John Wiley & Sons, New York, 2000.

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

[3] J. Li and P. Stoica, *Robust Adaptive Beamforming.* John Wiley & Sons, Hoboken, New Jersey, 2005.

[4] M. A. Lehr, B. Widrow, "Directional hearing system," U.S. Patent 5793875, Aug. 11, 1998.

[5] I. Tashev, H. S. Malvar, "A new beamformer design algorithm for microphone arrays," *Proceedings of ICASSP* vol. 3, pp 101-104, Mar. 2005.

[6] L. J Griffiths and C.W. Jim, "An alternative approach to linearly constrained adaptive beamforming," *IEEE Trans. Antennas Propag.*, vol. AP-30, pp. 27-34, Jan. 1982.

[7] K. M. Buckley, "Spatial/spectral filtering with linearly constrained minimum variance beamformers," *IEEE Trans. ASSP* vol. ASSP-35, pp 249-266, Mar. 1987.

[8] Y. Kaneda, J. Ohga, "Adaptive microphone-array system for noise reduction," *IEEE Trans. ASSP* vol. ASSP-34, no. 6, pp 1391-1400, Dec. 1986.