ADAPTIVE BEAMFORMING METHODS FOR DYNAMICALLY STEERED MICROPHONE ARRAY SYSTEMS

Ville Myllylä

Nokia Technology Platforms
Visiokatu 1, 33720 Tampere, Finland
ville.myllyla@nokia.com

Matti Hämäläinen

Nokia Research Center
Visiokatu 1, 33720 Tampere, Finland
matti.s.hamalainen@nokia.com

ABSTRACT

This paper introduces two methods for the integration of a dynamically steered beamforming filter with acoustic echo cancellation (AEC) and adaptive Generalized Sidelobe Canceler (GSC) based beamforming methods. We evaluate the performance of a beamforming system with moving and changing source positions. Individual contributions of adaptive beamforming and steering independent AEC processing methods are evaluated for high level echo cancellation in a typical office environment. The results show that the proposed adaptive beamforming method increases the overall AEC performance even if GSC adaptation would be disturbed by dynamic beam steering.

Index Terms— Beamforming, Beam steering, Adaptive Echo Cancellation, Adaptive Interference Cancellation, Polynomial Beamforming Filter

1. INTRODUCTION

Current commercial teleconference and communication systems apply acoustic echo cancellation (AEC) and noise reduction techniques to enable fluent communication with a variable degree of success. Typical meeting rooms and office spaces are fairly challenging acoustic environments for high quality hands-free communication. Room reverberation and the need for high loudspeaker levels make the echo cancellation more challenging. The loudspeaker-to-microphone distance is typically small in portable devices causing high levels of acoustic echo in the microphone signals. On the other hand, background noise is not as severe a problem as it is, for example, in the car environment. Microphone arrays and beamforming (BF) techniques open up new possibilities for improving the performance of traditional single channel communication systems [1].

We consider a teleconference system for multiple simultaneous users in a typical office environment. The system has a microphone array front-end capable of steering towards the active speaker. Conversations involving multiple speakers lead to active beam steering operation with unpredictable changes in the look direction of the microphone array. Beam steering introduces changes in the signal path from the acoustic source to the beamformer output.

Adaptive beamforming techniques are known for their good performance in suppressing non-diffuse interfering sound fields. A commonly applied class of adaptive beamforming methods is based on Linearly-Constrained Minimum Variance (LCMV) beamforming, which can be implemented efficiently using the Generalized Sidelobe Canceler (GSC) filter structure [2]. Since then, many GSC techniques have been proposed to improve the robustness for real systems, applications and environments [1, in ch. 5],[3].

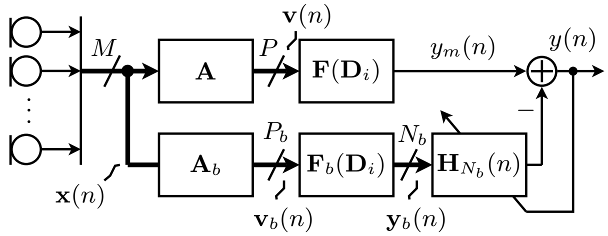

Fig. 1. Modified GSC implementation with a PBF front-end.

Integration of multiple audio technologies increases the technical complexity, since AEC, beamforming and beam steering algorithms tend to affect each other's performance. In the teleconferencing application it is favorable if both AEC and GSC perform echo cancellation, since echo is typically the most dominating and disturbing factor in such systems. In [3], several integrated AEC and adaptive beamforming methods were evaluated with different echo and interference levels. The study included also a dynamic beam steering experiment with random target source movement in the range of $\pm 10^\circ$. This type of source position movements did not cause problems to the studied methods. In [4], the interaction between AEC and dynamic beamforming was studied with more relaxed beam steering assumptions. The paper proposed a steering independent AEC structure for teleconference systems supporting changing speaker locations.

In this paper we derive dynamically steered beamforming techniques based on the GSC structure. In Fig. 1, a Polynomial Beamforming Filter (PBF) [4, 5] is applied to a conventional GSC structure [6] by replacing a fixed Filter-and-Sum (FSB) beamforming filters with corresponding PBF filter blocks. In the PBF, processing is divided in to fixed beamformer $\mathbf{A}$ providing signal decomposition into P intermediate signals and beam steering filter $\mathbf{F}(\mathbf{D}_i)$. The blocking matrix PBF filter stages are denoted by a subscript $_b$. The blocking matrix outputs N_b noise reference signals, which are input to an N_b-channel adaptive filter $H_{N_b}(n)$. This adaptive noise reduction filter is commonly called as Adaptive Interference Canceler (AIC). In the steering independent AEC integration AEC filter is connected between $\mathbf{A}$ and $\mathbf{F}(\mathbf{D}_i)$ as shown in Fig. 2 and Fig. 3. In this configuration the number of parallel AEC filters can be reduced ($P < M$) without compromising the AEC performance. In the following section we introduce two adaptive beamforming systems that are based on the steering independent AEC and the GSC structure illustrated in Fig. 2.

2. DYNAMICALLY STEERED ADAPTIVE BEAMFORMING SYSTEMS

In this section we derive two GSC-based filter structures integrating PBF beamformer front-end, steering independent AEC, and two alternative AIC configurations having one or P parallel AIC filters, which is noted by a subscript $_P$ (in experiments $P = 5$). The integrated AEC and AIC system is referred as AIEC. System diagrams are illustrated in Fig. 2 and Fig. 3.

2.1. PBF front-end and blocking matrix implementation

The PBF filter can be considered as a Filter-and-Sum Beamforming (FSB) filter with filter coefficients $h_{j,k}(\mathbf{D}_i)$ defined as a polynomial function of desired steering parameters $\mathbf{D}_i$. A good choice for polynomial basis is Chebyshev polynomials $F_p(\mathbf{D}_i) = \cos(p \cdot \arccos(\mathbf{D}_i))$, where the steering parameter could be, e.g., a normalized steering angle $\mathbf{D}_i = \theta_i/180^\circ$, where $-180^\circ < \theta_i \leq 180^\circ$.

The PBF output is defined as

$$y_m(n, \mathbf{D}_i) = \sum_{p=0}^{P-1} F_p(\mathbf{D}_i) \sum_{j=1}^{M} \sum_{k=0}^{L-1} a_p(j,k) x_j(n-k), \quad (1)$$

which can be written in a matrix form

$$y_m(n, \mathbf{D}_i) = \mathbf{F}(\mathbf{D}_i)\mathbf{A}\mathbf{x}(n). \quad (2)$$

The PBF has been partitioned into two stages where the first stage $\mathbf{A}$ is time-invariant and the second stage provides the time-variant beam steering filter $\mathbf{F}(\mathbf{D}_i)$. The PBF filter is particularly suitable for the calculation of multiple beams since $\mathbf{F}(\mathbf{D}_i)$ can be calculated very efficiently with the complexity of $O(P)$. A more detailed derivation of the PBF is available in [4, 5].

In the GSC structure depicted in Fig. 1, the steering filter $\mathbf{F}(\mathbf{D}_i)$ steers the maximum sensitivity of the beamformer towards the desired source direction $\mathbf{D}_i$ and the blocking filter $\mathbf{F}_b(\mathbf{D}_i)$ steers a spatial null towards the same source direction.

In the case of symmetric ring array, it is possible to design a PBF filter approximating the same beam response to all look directions. In this case we can reuse the signal decomposition filter $\mathbf{A}$ also for the blocking filter ($\mathbf{A} = \mathbf{A}_b$) and implement the blocking filter by rotating the steering filter to the opposite look direction. With these assumptions the blocking matrix output can be defined as

$$\begin{aligned} y_b(n, \mathbf{D}_i) &= \mathbf{F}_b(\mathbf{D}_i)\mathbf{A}_b\mathbf{x}(n) \\ &\approx \mathbf{F}\left((((\mathbf{D}_i + 1)(\mathrm{mod}\ 2)) - 1\right)\mathbf{A}\mathbf{x}(n). \quad (3) \end{aligned}$$

2.2. Steering independent AEC

Steering independent AEC structure utilized in this paper was proposed in [4]. In Fig. 2 and Fig. 3, the AEC generates a replica $\hat{\mathbf{d}}(n)$ of the echo components $\mathbf{d}(n)$ within the intermediate signal vector $\mathbf{v}(n)$ composed by speech, $\mathbf{s}(n)$, echo, $\mathbf{d}(n)$, and noise, $\mathbf{n}(n)$. After reducing the estimated echo $\hat{\mathbf{d}}(n)$ from intermediate signal $\mathbf{v}(n)$, the output becomes

$$\mathbf{e}(n) = \left[\mathbf{A}\mathbf{x}(n) - \mathbf{W}_P(n)\mathbf{u}(n)\right], \quad (4)$$

where $\mathbf{u}(n) = [u(n), \ldots, u(n - L_{AEC} + 1)]^T$ and L_{AEC} is the AEC filter length. The output signal of the PBF filter becomes

$$y_m(n, \mathbf{D}_i) = \mathbf{F}(\mathbf{D}_i)\mathbf{e}(n), \quad (5)$$

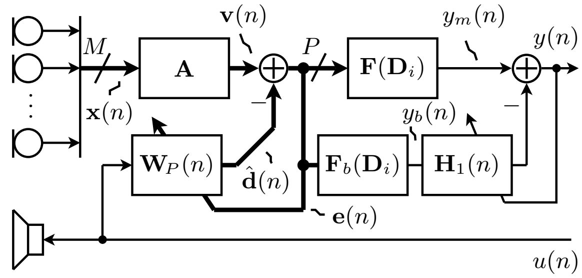

Fig. 2. Dynamically steerable adaptive beamforming system with PBF, AEC and single channel AIC (referred as AIEC). In the AIC configuration without AEC processing (referred as AIC), the AEC block $\mathbf{W}_P(n)$ becomes inactive. For the sake of clarity the steering variable $\mathbf{D}_i$ has been omitted from the output signals names.

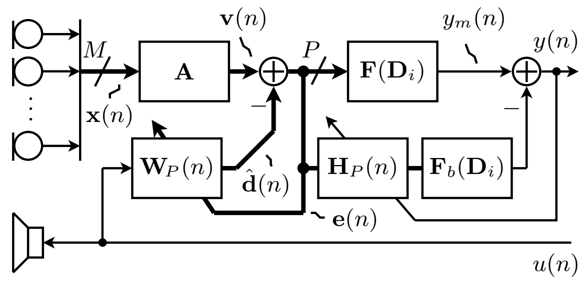

Fig. 3. Dynamically steerable adaptive beamforming system with PBF, AEC and P-channel AIC, referred as AIEC$_P$. The system with $\mathbf{W}_P(n)$ inactive is referred as AIC$_P$. Other details as in Fig. 2.

where $\mathbf{e}(n)$ also represent the signal components for near-end speech and noise. The adaptation error term $\mathbf{e}(n)$ is independent of the beam shape filter $\mathbf{F}(\mathbf{D}_i)$. Therefore, the AEC configuration is not affected by the beam steering.

2.3. Dynamically steered AIC processing

A single channel AIC implementation in Fig. 2 is a straightforward single channel textbook adaptive filter implementation. Therefore, we will concentrate on the derivation of the multi-channel AIC.

The input signal data for AIC processing is collected in to a matrix

$$\mathbf{E}(n) = [\mathbf{e}_1(n), \mathbf{e}_2(n), \ldots, \mathbf{e}_P(n)], \quad (6)$$

where

$$\mathbf{e}_j(n) = [e_j(n), e_j(n-1), \ldots, e_j(n - L_{AIC} + 1)]^T,\ j = 1\ldots, P, \quad (7)$$

represents each P intermediate signals and L_{AIC} is the AIC filter length. When the AIC filter coefficient matrix is defined by

$$\mathbf{H}_P(n) = [\mathbf{h}_1(n), \ldots, \mathbf{h}_P(n)], \quad (8)$$

the system output is then achieved as

$$y(n) = y_m(n) - \mathbf{F}_b(\mathbf{D}_i)\mathrm{diag}[\mathbf{H}_P^{\mathbf{T}}(n)\mathbf{E}(n)]. \quad (9)$$

Adaptation of the AICs follows the NLMS rule according to

$$\mathbf{h}_j(n+1) = \mathbf{h}_j(n) - \alpha_j(n)\mathbf{e}_j(n)/\|\mathbf{e}_j(n)\|^2 y(n),\ j = 1, ..., P, \quad (10)$$

where the error $y(n)$ (from the NLMS point of view) is adjusted for

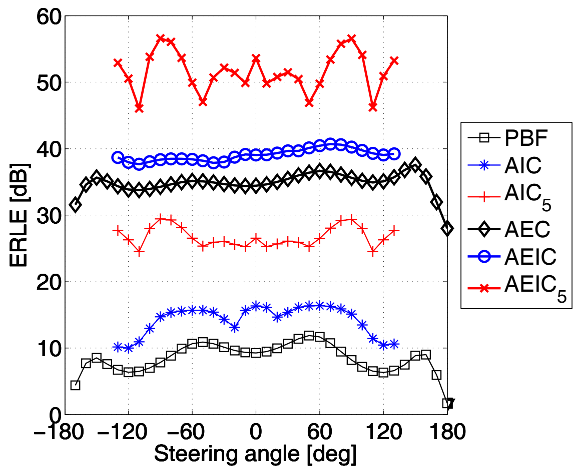

Fig. 4. Steady state ERLE performance. Look directions with AICs restricted to $\pm 135°$ range, while the loudspeaker is at 180 degrees.

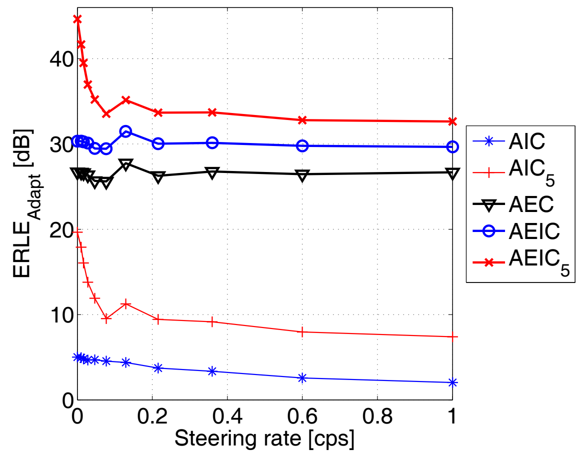

Fig. 5. Average disturbance reduction with the loudspeaker distance of 225mm.

each P AIC adaptation with the step-size

$$\alpha_j(n) = \text{sign}[F_{b\,j}(\mathbf{D}_i)]\mu/P,\ j = 1, ..., P, \tag{11}$$

where $F_{b\,j}(\cdot)$ refers to the blocking filter coefficient applied to the jth input channel and μ is a common step-size factor.

3. EXPERIMENTS

Audio recordings were done in an office environment illustrated in Fig. 6. A circular microphone array of 8 omni-directional microphones[1] was mounted on the surface of a table by drilling holes through the table. A dome shaped PC loudspeaker[2] propagates sound against the table from the height of 1cm producing radially symmetric sound field. The loudspeaker was used as a downlink signal source at distance 225mm from the center of the microphone array. Five 2-way studio loudspeakers[3] {S1,...,S5} were placed around the table for the simulation of near-end speakers. The ambient noise level was 39dBA at the location of the microphone array.

3.1. Simulation set-up

All AIC and AEC filters used in simulations were of NLMS type and equal length 1000 (at sampling rate 8kHz), to ease the comparison of echo reduction performance of AEC and AIC. AIC filters had a fixed step-size of value 0.5. In the steady-state and rotating beam experiments, we applied white noise as a loudspeaker excitation signal and the local disturbances were minimized. Under these circumstances a fixed step-size (of value 0.3) for the AEC adaptation was adequate. In the teleconferencing experiment with recorded speech, a proper adaptation control was applied [7]. Furthermore, the adaptation of AIC filters was inhibited during the near-end speech activity to prevent near-end speech attenuation. PBF filter was designed with 4th other ($P = 5$) Chebyshev polynomials providing continuous ($\theta_i = \pm 180$) steering.

[1]Sennheiser KE-4-211-2
[2]Philips model SBA1500
[3]Genelec 1029A

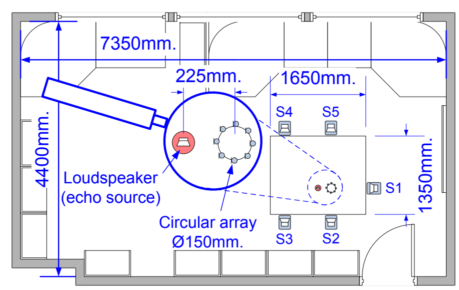

Fig. 6. Illustration of the room environment and loudspeaker microphone array configuration at the distance of 225mm. The room has a reverberation time $T_{60} \approx 320$ ms.

A set of Echo Return Loss Enhancement (ERLE) metrics were derived to measure AEC performance:

$$\text{ERLE} = 10\log_{10}\frac{E\{\|\mathbf{x}(n)\|^2\}}{M\cdot E\{y^2(n)\}},\ \text{and} \tag{12}$$

$$\text{ERLE}_{\text{Adapt}} = 10\log_{10}\frac{E\{y^2_{\text{PBF}}(n)\}}{E\{y^2(n)\}}. \tag{13}$$

With different configurations, $y(n)$ in (12) and (13) is replaced by the corresponding output signal. $\text{ERLE}_{\text{Adapt}}$ measures AEC performance without the contribution of PBF.

3.2. Steady-state

In this experiment we measured the ERLE (12) while the PBF look direction was fixed. Measures were repeated for several look directions to cover whole circle at 10 degree steps. In the case of configurations applying AIC the steering angle range was limited to $\pm 135°$ to eliminate the case of having the interfering source at the desired look direction. Results are depicted in Fig. 4. AEC configuration with a known reference signal provides the best echo signal cancellation. The performance can be further improved by AIC processing. The additional degrees of freedom improve the performance of five channel AIC considerably compared to the single AIC following the blocking filter.

3.3. Rotating beam

For the evaluation of algorithm sensitivity for constant rate beam steering, a beam rotation experiment was performed to quantify the degradation of ERLE performance against constant velocity beam steering in cycles per second (cps). In this case one cycle corresponds to $270^\circ = 2 \times 135^\circ$ azimuth rotation. The resulting $\text{ERLE}_{\text{Adapt}}$ figures, along (13), measured for several steering speeds are given in Fig. 5. The results demonstrate that steering independent AEC provides the basis for good system performance. AIC processing can increase the echo reduction and multi-channel AIC outperforms the single channel AIC in echo signal attenuation.

3.4. Teleconferencing

Our teleconferencing test case is introduced in Fig. 7. The active near-end speakers are shown in Fig. 7(a). The far-end speaker is frequently active, Fig. 7(b), resulting lots of doubletalk. The speech levels were subjectively adjusted to mimic ordinary conversation.

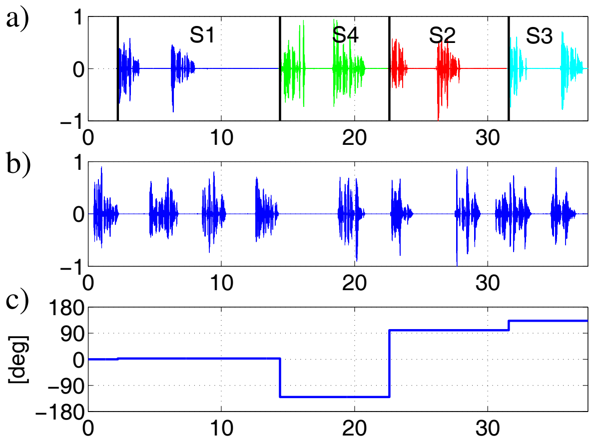

Fig. 7. Echo reduction performance of the AEC in teleconferencing situation. a) Near-end speakers, b) Echo, c) Look directions

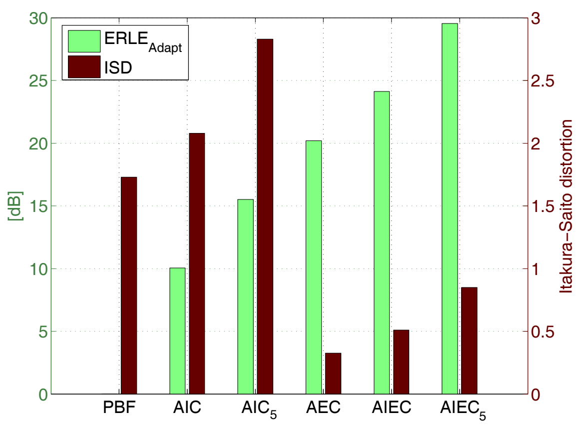

Fig. 8. System performance in terms of average $\text{ERLE}_{\text{Adapt}}$, (13), and Itakura-Saito distortion (ISD) measure [8] in realistic teleconferencing situation.

On average, the signal to echo ratio at the microphone positions was -7 dB due to the loudspeaker proximity. Beam steering was idealized by providing an accurate source direction for the system. The look direction is illustrated in Fig. 7(c).

The teleconferencing performance results are shown in Fig. 8. As a near-end speech distortion measure, we used an average of the Itakura-Saito distortion (ISD) [8] computed over the active speech frames of length 20ms, while the PBF-processed clean speech was used as a reference. Speech distortion originates from the residual echo, background noise, and attenuation by the AIC stage. On the contrary to ISD, $\text{ERLE}_{\text{Adapt}}$ was computed as an average when only echo and noise was present.

Methods with AEC processing have naturally higher $\text{ERLE}_{\text{Adapt}}$ levels. AEC can attenuate echo much better than AIC also during the double-talk without near-end speech attenuation and, therefore, it provides much lower speech distortion. While the ordinary AIC adapts slower and have less degrees of freedom than the five channel AIC, it also causes lower near-end speech attenuation and thus achieves better ISD results. Steering independent AEC provides the best stand-alone performance but AIC processing can further improve the total AEC performance of the system close to 10dB (AIEC_5).

4. CONCLUSIONS

In this paper we studied the adaptive beamforming (AIC) and echo cancellation and their combined AEC performance in the presence of strongly time-variant echo path changes caused by dynamic beam steering. Beam steering is a common function in teleconference applications applying microphone arrays, and theretofore audio algorithms should be able to cope with acoustic changes originating from beam steering. The system performance was evaluated mainly from an echo cancellation performance point of view as it is one of the most critical interfering sources in a typical office environment. These results suggest the AIC techniques can be applied relatively efficiently even in the case of active beam steering. The evaluated systems applied relatively simple control logic and further improvements can be expected with more advanced control logic.

References

[1] M. Brandstein and D. Ward, Eds., *Microphone arrays: Signal Processing Techniques and Applications*, Springer, 2001.

[2] L. J. Griffiths and C. W. Jim, “An alternative approach to linearly constrained adaptive beamforming,” *IEEE Trans. Antennas and Propag.*, vol. AP-30, no. 1, pp. 27–34, Jan. 1982.

[3] W. Herbordt, W. Kellermann, and S. Nakamura, Joint optimization of acoustic echo cancellation and adaptive beamforming, in E. Hänsler, G. Schmidt (Eds.), *Topics in Acoustic Echo and Noise Control*. Berlin, Germany: Springer, 2006.

[4] M. Hämäläinen and V. Myllylä, “Acoustic echo cancelation for dynamically steered microphone array systems,” *Proc. WASPAA*, New Paltz, NY, October 2007.

[5] M. Kajala and M. Hämäläinen, “Filter-and-sum beamformer with adjustable filter characteristics,” in *Proc. ICASSP*, July 2001, pp. 2917–2920.

[6] S. Nordebo, I. Claesson, and S. Nordholm, “Broadband adaptive beamforming: A design using 2-D spatial filters,” in *Ant. and Propag. Soc. Int. Symp.*, vol. 2, June 1993, pp. 702–705.

[7] V. Myllylä, “Residual echo filter for enhanced acoustic echo control,” *Signal Proc.*, vol.86, no.6, pp. 1193–1205, 2006.

[8] L. Rabiner and B.-H. Juang, *Fundamentals of Speech Recognition*. Prentice Hall, pp. 171-177, 1993.