Int. Conf. on Spoken Language Processing (ICSLP 2002), pp. 773-776, Denver, Colorado, Sept. 2002

A REAL-TIME ACOUSTIC HUMAN-MACHINE FRONT-END FOR MULTIMEDIA APPLICATIONS INTEGRATING ROBUST ADAPTIVE BEAMFORMING AND STEREOPHONIC ACOUSTIC ECHO CANCELLATION

W. Herbordt[†], *J. Ying*[‡], *H. Buchner*[†], *W. Kellermann*[†] [*]

[†]Telecommunications Laboratory, University Erlangen-Nuremberg
Cauerstraße 7, 91058 Erlangen, Germany
{herbordt,buchner,wk}@LNT.de
[‡]Intel Architecture Development Co., Ltd
Beijing Kerry Centre, Ganghua Lu 1, 100020 Beijing
ying.jia@intel.com

ABSTRACT

Joint beamforming microphone arrays and multi-channel acoustic echo cancellation (AEC) can be ef£ciently applied for hands-free speech communication. Especially, systems relying on adaptive generalized sidelobe canceller (GSC) structures are very promising, since they combine high noise-reduction performance with computational ef£ciency. So far, robustness of the GSC was very challenging, due to reverberation and non-stationarity of desired signals and interferers.

In this contribution, we present for the £rst time a real-time system which integrates GSC and stereophonic AEC. It is robust against desired signal cancellation while highly suppressing interference and acoustic echoes. The realization on a low-cost PC platform, with the microphone array connected directly to the universal serial bus (USB), provides maximum hardware and software compatibility for personalized mobile computing devices and desktop PCs.

1. INTRODUCTION

With the need for natural and comfortable speech communication gaining more and more importance, dialogue systems, videoconferencing, voice over internet protocol (VoIP), and other multimedia services call for high-quality hands-free acoustic interfaces. For that, the user should be allowed to move freely without wearing or holding any microphone device. Multi-channel sound reproduction is necessary to enhance sound realism.

Solutions which ensure minimum cost, convenient usage, and optimum hardware and software compatibility, e.g., between mobile PCs, desktop PCs, and personal digital assistants (PDAs), are highly desirable.

For optimum recording quality, the signals of interest should be free from any kind of impairment, i.e., noise, reverberation, local interferers, and echoes of the loudspeaker signals.

Acoustic echo cancellation (AEC) is used whenever a reference of the loudspeaker signals is accessible, since it allows maximum suppression of these interferers. With personalized devices, echoes from the loudspeakers that are part of the device fall into this category. For minimum complexity, multi-channel AEC may be realized very ef£ciently in the frequency domain [1].

The use of microphone arrays gives one the opportunity to exploit spatial separation of the desired speaker and the noise sources. Here, a robust generalized sidelobe canceller (GSC) [2] seems to be very promising. On the one hand, it allows high interference rejection and the desired speaker may move freely within a so-called tracking region without requiring to estimate the current speaker position. On the other hand, however, adaptation is delicate, and often leads to target signal cancellation or transient interferer signals when the speaker of interest and the interferers are present simultaneously (double-talk). These problems may be ef£ciently addressed by realizing the robust GSC in the frequency-domain (FGSC) using frequency-domain adaptive £lters (FDAFs) [3] while reducing the computational complexity considerably [4].

For high output signal quality, it is thus desirable to reconcile FGSC and multi-channel AEC with exploitation of optimum positive synergies [5]: Placing AECs directly into the sensor channels (FAEGSC) yields optimum synergies at the expense of high computational load. One AEC after the GSC reduces computational complexity, however, the AEC is almost inef£cient due to the strong time-variance of the GSC. This problem can be avoided by embedding the AEC into the GSC (FGSAEC). It equally requires only one AEC for an arbitrary number of sensors and preserves most of the synergies of FAEGSC [6] (see Figure 1).

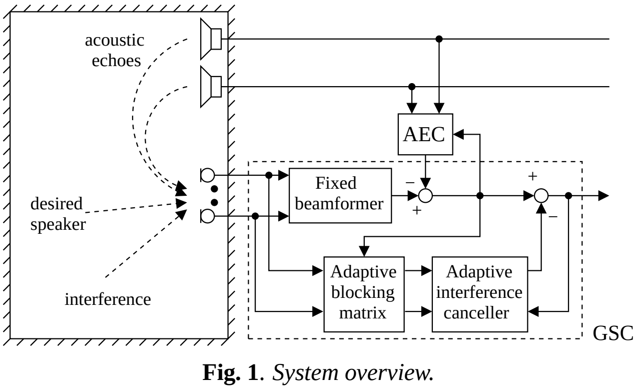


Fig. 1. *System overview.*

In the following, we show how the FGSAEC algorithm embedding a stereophonic AEC can be ef£ciently realized in real-

[*]This work was supported by a grant from Intel Corp., Hillsboro, OR.

time. The advantages of low-cost PC hardware are exploited by integrating the entire speech capture unit into a single USB device, instead of using expensive multi-channel sound cards. Our acoustic front-end may then be deployed in combination with all computing devices having a USB. In Section 2, we describe the FGSAEC signal processing. Section 3 briefly presents the realization on the PC platform. Finally, experimental results in Section 4 show the efficiency of our approach.

2. FREQUENCY-DOMAIN REALIZATION OF FGSAEC

2.1. Notations

Uppercase symbols denote frequency-domain[1] variables, lowercase symbols stand for time-domain variables, and the boldface font indicates a vector or matrix quantity. Superscript ^{T} and ^{H} represent transpose and complex conjugate transpose, respectively. The number of microphones is denoted by M.

The Discrete Fourier Transform (DFT) length for the GSC and for the AEC is $2L_\mathrm{g}$ and $2L_\mathrm{h}$, respectively. $\mathbf{F}_\mathrm{g}$ and $\mathbf{F}_\mathrm{h}$ are the $2L_\mathrm{g} \times 2L_\mathrm{g}$ and the $2L_\mathrm{h} \times 2L_\mathrm{h}$ DFT matrices, respectively. The parameters L_g and L_h are identical to the number of filter taps of GSC and AEC adaptive filters, respectively. For better tracking behavior of the FDAFs, block overlaps by factors α_g and α_h are introduced in the GSC and AEC input signal blocks, respectively [3]. The discrete time variable is n. We further use the time index $k = n\,\alpha_\mathrm{g}/L_\mathrm{g}$ that reflects the discrete time in numbers of blocks of length $L_\mathrm{g}/\alpha_\mathrm{g}$. For our discussion, we assume that $L_\mathrm{h}/\alpha_\mathrm{h}$ is an integer multiple of $L_\mathrm{g}/\alpha_\mathrm{g}$, which is reasonable for our system. For a better reading, we define $Q = \frac{L_\mathrm{h}\alpha_\mathrm{g}}{L_\mathrm{g}\alpha_\mathrm{h}}$, and the time index $q = n\,\alpha_\mathrm{h}/L_\mathrm{h}$. GSC and AEC adaptive filters are updated at time stamps k and q, respectively.

The vectors **0** and **1** are vectors with zeroes and ones, respectively.

2.2. Fixed beamformer (FBF)

The fixed beamformer attenuates interference components relative to the desired signal. In the general case, the FBF may be realized as a filter&sum beamformer for giving the reference path a desired fixed directivity pattern, which influences the directivity pattern of the GSC [4]. Using a simple delay&sum beamformer (see Figure 2), the FBF output signal can be written as $y_\mathrm{f}(n) = \frac{1}{M}\sum_{m=0}^{M-1} x_m(n)$, where $x_m(n)$, $m = 0,\,1,\,\ldots,\,M-1$ are the sensor signals $x_m(n)$ and where the array is assumed to be steered to the desired speaker position.

2.3. Stereophonic acoustic echo canceller (AEC)

The stereophonic AEC identifies the impulse responses between the loudspeakers and the FBF output using adaptive FIR filters. An estimate of the acoustic echoes is given by the outputs of the FIR filters, which are then subtracted from the FBF output to cancel the acoustic echoes in the GSC reference path. The specific problems of multi-channel AEC include all those known for mono AEC, i.e., convergence speed, tracking, double-talk (e. g. [7]), but in addition to that, multi-channel AEC has to cope with the high correlation of the loudspeaker signals [7]. In our implementation, these problems are addressed by decorrelating the loudspeaker signals by some nearly inaudible preprocessing [7] and by applying a two-channel frequency-domain algorithm, which takes the cross-correlation into account [1]. Double-talk is detected by methods, which are presented in [7].

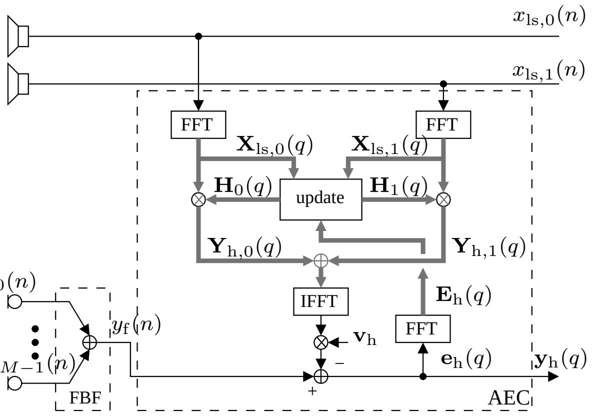

$x_{\mathrm{ls},0}(n)$
$x_{\mathrm{ls},1}(n)$
FFT
FFT
$\mathbf{X}_{\mathrm{ls},0}(q)$
$\mathbf{X}_{\mathrm{ls},1}(q)$
$\mathbf{H}_0(q)$
update
$\mathbf{H}_1(q)$
$\mathbf{Y}_{\mathrm{h},0}(q)$
$\mathbf{Y}_{\mathrm{h},1}(q)$
$x_0(n)$
IFFT
$\mathbf{E}_\mathrm{h}(q)$
$y_\mathrm{f}(n)$
$\mathbf{v}_\mathrm{h}$
FFT
$x_{M-1}(n)$
$-$
$\mathbf{e}_\mathrm{h}(q)$
$\mathbf{y}_\mathrm{h}(q)$
$+$
FBF
AEC

Fig. 2. *Stereophonic acoustic echo canceller (AEC) in the fixed reference path of the generalized sidelobe canceller.*

In the following we describe the basic stereophonic AEC signal processing without considering the double-talk detection. We capture the last $2L_\mathrm{h}$ samples of the loudspeaker signals $x_{\mathrm{ls},i}(n)$, $i = 0,\,1$ in vectors, and we define matrices of frequency-domain loudspeaker signals as:

$$\mathbf{X}_{\mathrm{ls},i}(q) = \mathrm{diag}\left\{ \mathbf{F}_\mathrm{h} \begin{pmatrix} x_{\mathrm{ls},i}(q\frac{L_\mathrm{h}}{\alpha_\mathrm{h}} - 2L_\mathrm{h} + 1) \\ x_{\mathrm{ls},i}(q\frac{L_\mathrm{h}}{\alpha_\mathrm{h}} - 2L_\mathrm{h} + 2) \\ \vdots \\ x_{\mathrm{ls},i}(q\frac{L_\mathrm{h}}{\alpha_\mathrm{h}}) \end{pmatrix} \right\}, \quad (1)$$

$$\mathbf{X}_\mathrm{ls}(q) = (\mathbf{X}_{\mathrm{ls},0}(q),\,\mathbf{X}_{\mathrm{ls},1}(q))\,. \quad (2)$$

With the $2L_\mathrm{h} \times 1$ vectors of adaptive filter transfer functions $\mathbf{H}_i(q)$, $i = 0,\,1$, captured in a matrix

$$\mathbf{H}(q) = \left(\mathbf{H}_0^T(q),\,\mathbf{H}_1^T(q)\right)^T, \quad (3)$$

and with $\mathbf{v}_\mathrm{h} = (\mathbf{0}_{1\times L_\mathrm{h}},\,\mathbf{1}_{1\times L_\mathrm{h}})^T$, the time-domain error signal $\mathbf{e}_\mathrm{h}(q)$ between the sum of the adaptive filter outputs and the FBF output signal is obtained as

$$\mathbf{e}_\mathrm{h}(q) = \begin{pmatrix} \mathbf{0}_{L_\mathrm{h}\times 1} \\ y_\mathrm{f}(q\frac{L_\mathrm{h}}{\alpha_\mathrm{h}} - L_\mathrm{h} + 1) \\ y_\mathrm{f}(q\frac{L_\mathrm{h}}{\alpha_\mathrm{h}} - L_\mathrm{h} + 2) \\ \vdots \\ y_\mathrm{f}(q\frac{L_\mathrm{h}}{\alpha_\mathrm{h}}) \end{pmatrix} - \mathbf{v}_\mathrm{h}\mathbf{F}_\mathrm{h}^{-1}\mathbf{X}_\mathrm{ls}(q)\mathbf{H}(q)\,. \quad (4)$$

The frequency-domain error signal which is required for the filter update is obtained by $\mathbf{E}_\mathrm{h}(q) = \mathbf{F}_\mathrm{h}\mathbf{e}_\mathrm{h}(q)$. The filter update equations may be written as

$$\mathbf{H}(q+1) = \mathbf{H}(q) + \mu_\mathrm{h}\mathbf{S}^{-1}(q)\mathbf{X}_\mathrm{ls}^H(q)\mathbf{E}_\mathrm{h}(q)\,. \quad (5)$$

μ_h is a stepsize parameter, $\mathbf{S}(q)$ is a recursive estimate of the cross-power spectral density matrix of the loudspeaker signals:

$$\mathbf{S}(q) = \lambda_\mathrm{h}\,\mathbf{S}(q) + (1-\lambda_\mathrm{h})\mathbf{X}_\mathrm{ls}^H(q)\mathbf{X}_\mathrm{ls}(q)\,, \quad (6)$$

[1] Here, frequency domain corresponds to the Discrete Fourier Transform (DFT) domain.

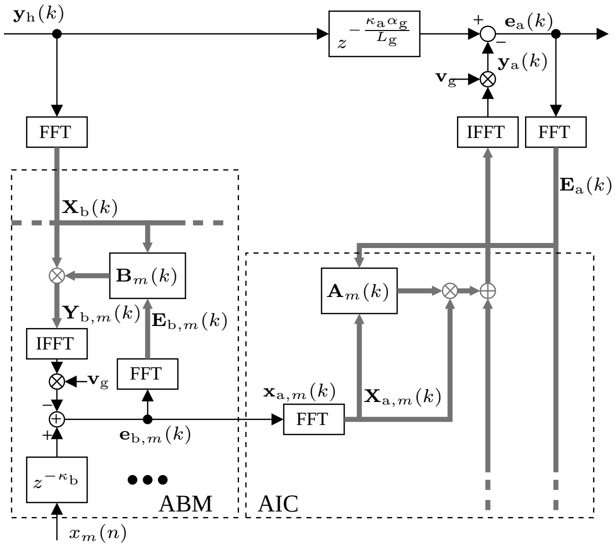
with the forgetting factor $0 < \lambda_\mathrm{h} < 1$.

One block of length $L_\mathrm{h}/\alpha_\mathrm{h}$ of the AEC output signal is finally given by the last $L_\mathrm{h}/\alpha_\mathrm{h}$ samples of the error signal $\mathbf{e}_\mathrm{h}(q)$. These signal blocks are by a factor Q larger than the signal blocks which are required for the GSC. We therefore split $\mathbf{e}_\mathrm{h}(q)$ into Q blocks $\mathbf{x}_\mathrm{b}(k-i)$, $i = 0, 1, \ldots, Q-1$ of length L_h/Q. $Q-1$ blocks of $\mathbf{x}_\mathrm{b}(k-i)$ are buffered until they are used by the GSC.

2.4. Adaptive blocking matrix (ABM)

The adaptive blocking matrix consists of adaptive FIR filters between the AEC output and the sensor channels: It adaptively subtracts the signal of interest from the adaptive sidelobe cancelling path in order to prevent the desired signal to be cancelled by the AIC.

In Figure 3, the reference path with one signal path of the adaptive sidelobe cancelling path is depicted for simplicity.



$\mathbf{y}_\mathrm{h}(k)$
$z^{-\frac{\kappa_\mathrm{a}\alpha_\mathrm{g}}{L_\mathrm{g}}}$
+
$\mathbf{e}_\mathrm{a}(k)$
−
$\mathbf{y}_\mathrm{a}(k)$
$\mathbf{v}_\mathrm{g}$
FFT
IFFT
FFT
$\mathbf{E}_\mathrm{a}(k)$
$\mathbf{X}_\mathrm{b}(k)$
$\mathbf{B}_m(k)$
$\mathbf{Y}_{\mathrm{b},m}(k)$
$\mathbf{E}_{\mathrm{b},m}(k)$
IFFT
FFT
$\mathbf{v}_\mathrm{g}$
−
+
$\mathbf{e}_{\mathrm{b},m}(k)$
$\mathbf{A}_m(k)$
$\mathbf{x}_{\mathrm{a},m}(k)$
$\mathbf{X}_{\mathrm{a},m}(k)$
FFT
$z^{-\kappa_\mathrm{b}}$
ABM
AIC
$x_m(n)$

Fig. 3. *FGSAEC: adaptive blocking matrix (ABM) and adaptive interference canceller (AIC).*

The time delay κ_b ensures causality of the ABM adaptive filters. For applying the overlap-save method to the ABM adaptive filter inputs in the frequency domain, we have to transform $2\alpha_\mathrm{g}$ subsequent blocks of the AEC output signal $\mathbf{y}_\mathrm{h}(k)$ into the frequency domain. That is,

$$\mathbf{X}_\mathrm{b}(k) = \mathrm{diag}\left\{ \mathbf{F}_\mathrm{g} \begin{pmatrix} \mathbf{y}_\mathrm{h}(k-2\alpha_\mathrm{g}+1) \\ \mathbf{y}_\mathrm{h}(k-2\alpha_\mathrm{g}+2) \\ \vdots \\ \mathbf{y}_\mathrm{h}(k) \end{pmatrix} \right\} . \tag{7}$$

The $2L_\mathrm{g}$-by-1 vectors of ABM adaptive filter transfer functions are denoted by $\mathbf{B}_m(k)$, $m = 0, 1, \ldots, M-1$. The ABM filter input $\mathbf{X}_\mathrm{b}(k)$ is filtered by the adaptive filters $\mathbf{B}_m(k)$, yielding

$$\mathbf{Y}_{\mathrm{b},m}(k) = \mathbf{X}_\mathrm{b}(k)\mathbf{B}_m(k) . \tag{8}$$

For the adaptation algorithm, frequency-domain error signals $\mathbf{E}_{\mathrm{b},m}(k)$ are required which are free of circular convolution effects. Or, the time-domain error signals $\mathbf{e}_{\mathrm{b},m}(k)$ have to be constrained in such a way that the first block of L_g samples is discarded and that the second block of L_g samples is saved. That is,

$$\mathbf{e}_{\mathrm{b},m}(k) = \begin{pmatrix} \mathbf{0}_{L_\mathrm{g}\times 1} \\ x_m(k\frac{L_\mathrm{g}}{\alpha_\mathrm{g}} - \kappa_\mathrm{b} - L_\mathrm{g} + 1) \\ x_m(k\frac{L_\mathrm{g}}{\alpha_\mathrm{g}} - \kappa_\mathrm{b} - L_\mathrm{g} + 2) \\ \vdots \\ x_m(k\frac{L_\mathrm{g}}{\alpha_\mathrm{g}} - \kappa_\mathrm{b}) \end{pmatrix} - \mathbf{v}_\mathrm{g}\mathbf{F}_\mathrm{g}^{-1}\mathbf{X}_\mathrm{b}(k)\mathbf{B}_m(k) , \tag{9}$$

where $\mathbf{v}_\mathrm{g} = \left(\mathbf{0}_{1\times L_\mathrm{g}}, \mathbf{1}_{1\times L_\mathrm{g}}\right)^T$. The update equation for the m-th adaptive filter reads [2]

$$\mathbf{B}_m(k+1) = \mathbf{B}_m(k) + \mu_\mathrm{b}\mathbf{G}_\mathrm{g}\mathbf{X}_\mathrm{b}^H(k)\mathbf{E}_{\mathrm{b},m}(k) . \tag{10}$$

The parameter μ_b is the stepsize. The matrix $\mathbf{G}_\mathrm{g}$ realizes a constraint that prevents circular convolution effects in the filter update equations. That is,

$$\mathbf{G}_\mathrm{g} = \mathbf{F}_\mathrm{g}\mathrm{diag}\left\{\left(\mathbf{1}_{1\times L_\mathrm{g}}, \mathbf{0}_{1\times L_\mathrm{g}}\right)\right\}\mathbf{F}_\mathrm{g}^{-1} . \tag{11}$$

In contrast to the AEC (see Section 2.3), we need the circular convolution constraints for the ABM, since the impulse responses of optimum ABM filters are generally much longer than the length of the adaptive filters. Circular convolution effects thus cannot be disregarded.

One block of length $L_\mathrm{g}/\alpha_\mathrm{g}$ of the time-domain AIC input signal $\mathbf{x}_{\mathrm{a},m}(k)$ is obtained by saving the last $L_\mathrm{g}/\alpha_\mathrm{g}$ samples of the error signal vector $\mathbf{e}_{\mathrm{b},m}(k)$.

2.5. Adaptive interference canceller (AIC)

The AIC adaptively subtracts signal components from the AEC output which are correlated with the AIC filter inputs. In Figure 3, the time delay κ_a is introduced for causality reasons.

The frequency-domain adaptive filter inputs $\mathbf{X}_{\mathrm{a},m}(k)$ are obtained in the same way as in Eq. 7 with $\mathbf{X}_\mathrm{b}(k)$ and $\mathbf{y}_\mathrm{h}(k)$ replaced by $\mathbf{X}_{\mathrm{a},m}(k)$ and $\mathbf{x}_{\mathrm{a},m}(k)$, respectively. The $2L_\mathrm{g}$-by-1 vectors of adaptive FIR filter transfer functions are denoted by $\mathbf{A}_m(k)$, $m = 0, 1, \ldots, M-1$. The time-domain AIC output signal is calculated as

$$\mathbf{y}_\mathrm{a}(k) = \mathbf{v}_\mathrm{g}\mathbf{F}_\mathrm{g}^{-1}\sum_{m=0}^{M-1}\mathbf{X}_{\mathrm{a},m}(k)\mathbf{A}_m(k) . \tag{12}$$

Capturing L_g samples of $\mathbf{y}_\mathrm{h}(k)$ in a vector, the AIC error signal $\mathbf{e}_\mathrm{a}(k)$ can be written as

$$\mathbf{e}_\mathrm{a}(k) = \begin{pmatrix} \mathbf{0}_{L_\mathrm{g}\times 1} \\ \mathbf{y}_\mathrm{h}(k - \frac{\kappa_\mathrm{a}\alpha_\mathrm{g}}{L_\mathrm{g}} - \alpha_\mathrm{g} + 1) \\ \mathbf{y}_\mathrm{h}(k - \frac{\kappa_\mathrm{a}\alpha_\mathrm{g}}{L_\mathrm{g}} - \alpha_\mathrm{g} + 2) \\ \vdots \\ \mathbf{y}_\mathrm{h}(k - \frac{\kappa_\mathrm{a}\alpha_\mathrm{g}}{L_\mathrm{g}}) \end{pmatrix} - \mathbf{y}_\mathrm{a}(k) , \tag{13}$$

The filter update equations of the adaptive filters are given by

$$\mathbf{A}_m(k+1) = \mathbf{A}_m(k) + \mu_\mathrm{a}\mathbf{G}_\mathrm{g}\mathbf{X}_{\mathrm{a},m}^H(k)\mathbf{E}_\mathrm{a}(k) , \tag{14}$$

where μ_a is a stepsize parameter and where $\mathbf{E}_\mathrm{a}(k)$ is the frequency-domain AIC error signal. One block of length $L_\mathrm{g}/\alpha_\mathrm{g}$ of the GSC output signal is finally obtained by saving the last $L_\mathrm{g}/\alpha_\mathrm{g}$ samples of $\mathbf{e}_\mathrm{a}(k)$.

[2] Coefficient constraints for improved robustness against cancellation of desired signal components may be introduced according to [2, 4].

2.6. Adaptation control (AC)

The adaptation control of the GSC can be summarized as follows: The ABM(AIC) is only adapted when the signal-to-noise ratio (SNR) is high(low) in order to prevent interference components to be cancelled by the ABM and in order to prevent desired signal cancellation, respectively. In our implementation, we modified the GSC adaptation control after [2] using a fullband SNR estimate to a DFT-bin-wise operation for better interference rejection and improved robustness against target signal cancellation during double-talk [4].

3. REALIZATION ON A PC PLATFORM

For illustrating the efficiency of our approach, we implemented the FGSAEC algorithm on a PC platform in real-time. The multi-channel audio capture unit is realized as separate hardware integrating the microphones, the preamplifiers, the A/D conversion, and the microphone calibration. The digitized sensor data is fed into the PC via a standard USB port with specific drivers for the microphone array. Compared to standard multi-channel audio capture modules or DSP-based systems, our solution ensures greater flexibility and lower hardware cost. It is especially suited to portable computing devices due to the small package size and since no additional power supply is necessary.

Our experiments are conducted on an Intel Pentium IV 1.4 GHz processor at a sampling rate of 12 kHz. For an efficient implementation of the FGSAEC algorithm, we made use of the vector-based Intel Signal Processing Libraries[3]. In an environment with 300 ms reverberation time and an 8-element microphone array, our present FGSC and FGSAEC realizations are running with 28% and 50% CPU load, respectively. For reducing the computational complexity (for e.g. PDAs), usage of less microphones or usage of Fast Fourier transforms that explicitely exploit overlapping input signal segments may be considered.

4. EXPERIMENTAL RESULTS

For evaluating the proposed real-time system experimentally, we compare the average interference rejection (IR) and the average echo-return-loss enhancement (ERLE) of FGSAEC, FGSC, FAEGSC, and TGSAEC, the time-domain equivalent of FGSAEC, for only interference and double-talk between interference and desired speaker. Since it is difficult to study IR and ERLE separately for real-time scenarios, we illustrate the results that we obtained with recorded signals in simulations. Audio examples which illustrate the performance of the real-time system can be found in [8]. We used a linear microphone array with 8 equally spaced, broadside steered sensors with 4 cm spacing in an office environment with 300 ms reverberation time. The male desired speaker and the male interferer are located in the array look-direction and 30 degrees off the array axis, respectively. The stereophonic loudspeakers, which are playing music, are placed to the left and to the right of the microphone array. All distances to the array center are 60 cm. The results are depicted in Table 1.

For only interference, IR and ERLE are higher than for the double-talk case, since the ABM is fixed and since the AIC can be adapted permanently over the entire frequency range, yielding optimum tracking capability of non-stationary interference. The performance of TGSAEC and FGSAEC is identical. During double-talk, IR and ERLE is considerably improved for FGSAEC relative to TGSAEC, as controlling the adaptation in individual frequency bins still allows tracking of the transient ABM and of non-stationary interference at frequencies with low SNR [4]. FGSAEC clearly improves the suppression of acoustic echoes relative to FGSC, however, optimum performance of FAEGSC cannot be obtained due to leakage effects across the sidelobe cancelling path.

	Interference only		**Double-talk**	
(in dB)	IR	$ERLE$	IR	$ERLE$
TGSAEC	22.4	26.0	5.6	12.3
FGSAEC	21.1	25.6	14.7	21.0
FGSC	20.7	21.9	14.5	14.7
FAEGSC	22.0	30.5	14.9	28.1

Table 1. *Performance evaluation.*

5. CONCLUSIONS

In this contribution, we presented a real-time implementation of a computationally efficient combination of robust GSC and stereophonic AEC with high output signal quality on a low-cost PC platform. Hardware requirements are minimized and maximum hardware and software compatibility is ensured by utilization of a USB microphone array.

6. ACKNOWLEDGMENTS

The authors would like to thank the speech communication group of Intel Corp., Beijing, China, for the fruitful collaboration during the realization of the real-time demonstrator, David Graumann, Intel Corp., Hillsboro, OR, and Bob Liang, Intel Corp., Santa Clara, CA, for their support and encouragement, and Jacob Benesty, Bell Labs, Murray Hill, NJ, for stimulating discussions.

7. REFERENCES

[1] H. Buchner, W. Herbordt, and W. Kellermann, “An efficient combination of multi-channel acoustic echo cancellation with a beamforming microphone array,” *Conf. Rec. 1st Intl. Workshop on Hands-Free Speech Communication (HSC)*, pp. 55–58, April 2001.

[2] O. Hoshuyama, A. Sugiyama, and A. Hirano, “A robust adaptive beamformer for microphone arrays with a blocking matrix using constrained adaptive filters,” *IEEE Trans. on Signal Processing*, vol. 47, no. 10, pp. 2677–2684, October 1999.

[3] E. Moulines, O. A. Amrane, and Y. Grenier, “The generalized multidelay adaptive filter: Structure and convergence analysis,” *IEEE Trans. on Signal Processing*, vol. 43, no. 1, pp. 14–28, January 1995.

[4] W. Herbordt and W. Kellermann, “Frequency-domain integration of acoustic echo cancellation and a generalized sidelobe canceller with improved robustness,” *European Transactions on Telecommunications (ETT)*, vol. 13, no. 2, 2002.

[5] M. S. Brandstein and D. B. Ward, Eds., *Microphone Arrays: Signal Processing Techniques and Applications*, Springer Verlag, 2001.

[6] W. Herbordt and W. Kellermann, “Limits for generalized sidelobe cancellers with embedded acoustic echo cancellation,” *Proc. Int. Conf. on Acoustics, Speech, and Signal Processing (ICASSP)*, vol. 5, Mai 2001.

[7] S. L. Gay and J. Benesty, Eds., *Acoustic Signal Processing for Telecommunications,* Kluwer Academic Publishers, Boston, 2000.

[8] W. Herbordt, http://www.LNT.de/∼herbordt.

[3] Can be found on the Intel web site http://developer.intel.com.