

Design and implementation of a hybrid sub-band acoustic echo canceller (AEC)

Mingsian R. Bai*, Cheng-Ken Yang, Ker-Nan Hur

Department of Mechanical Engineering, National Chiao-Tung University, 1001 Ta-Hsueh Road, Hsin-Chu 300, Taiwan

Received 27 April 2008; received in revised form 23 September 2008; accepted 24 September 2008

Handling Editor: J. Lam

Available online 20 November 2008

Abstract

An efficient method is presented for implementing an acoustic echo canceller (AEC) that makes use of hybrid sub-band approach. The hybrid system is comprised of a fixed processor and an adaptive filter in each sub-band. The AEC aims at reducing the echo resulting from the acoustic feedback in loudspeaker-enclosure-microphone (LEM) systems such as teleconferencing and hands-free systems. In order to cancel the acoustical echo efficiently, various processing architectures including fixed filters, hybrid processors, and sub-band structure are investigated. A double-talk detector is incorporated into the proposed AEC to prevent the adaptive filter from diverging in double-talk situations. A de-correlation filter is also used alongside sub-band processing in order to enhance the performance and efficiency of AEC. All algorithms are implemented and verified on the platform of a fixed-point digital signal processor (DSP). The AECs are evaluated in terms of cancellation performance and computation complexity. In addition, listening tests are conducted to assess the subjective performance of the AECs. From the results, the proposed hybrid sub-band AEC was found to be the most effective among all methods in terms of echo reduction and timbral quality.

© 2008 Elsevier Ltd. All rights reserved.

1. Introduction

The problem of acoustic echo arises wherever a loudspeaker and a microphone are placed together in a system, or the loudspeaker-enclosure-microphone (LEM) system illustrated in Fig. 1(a) [1]. The microphone picks up not only the local speech (s) and background noise (b) but also the echoes (d) from the loudspeaker through the so-called echo path. The problem exists in many LEM scenarios where users hear their own delayed voice, which can be disturbing enough to hamper communicational quality. Acoustic echoes can also cause unstable howling in Karaoke machines and hearing aids. Other examples of acoustic echo problem can be found in the application of mobile phones, Bluetooth[®] earphones, teleconferencing, hands-free car kits, etc. To combat this problem, the attenuation of the echo path between the loudspeaker and microphone must be made as high as possible through ingenious acoustical design. In addition to acoustical design which is often insufficient, an electronic acoustic echo canceller (AEC) must be used to reduce the excessive echoes picked up by the microphone.

*Corresponding author.

E-mail address: msbai@mail.nctu.edu.tw (M.R. Bai).

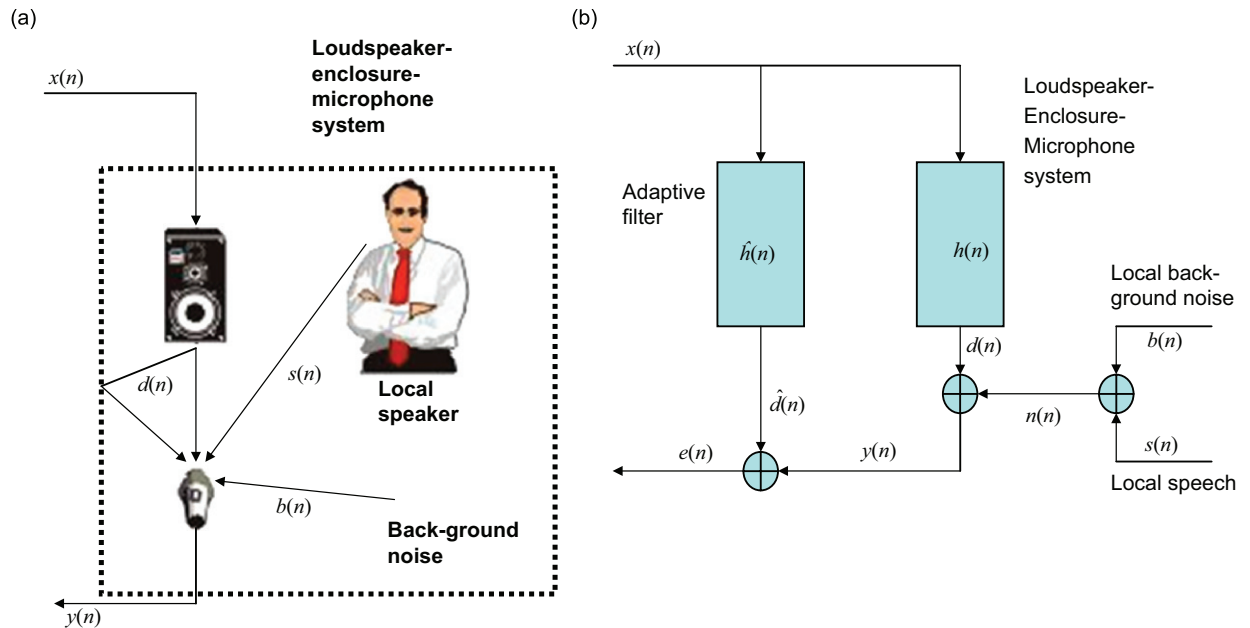


Fig. 1. Schematic diagram of an acoustic echo cancellation problem: (a) loudspeaker-enclosure-microphone (LEM) system and (b) block diagram of an AEC.

Conventional way of dealing with acoustic echo problem is to utilize an adaptive filter [1–6] to subtract the echo from the microphone output, as shown in Fig. 1(b). Thus, the problem is equivalent to a system identification problem of the echo path in the context of adaptive filters. The aim here is to approximate the impulse response $h(n)$ of the echo path by using a digital filter $\hat{h}(n)$ such that the far-end input signal $x(n)$ and the matching error $e(n)$ will be perfectly uncorrelated. To cope with the time-varying characteristics of echo path due to changed relative positions of transducers, moving objects, or any other altered room conditions, the digital filters are generally updated using some sort of adaptive algorithm. The least-mean-squares (LMS) algorithm and the normalized-least-mean-squares (NLMS) algorithm [1–5,7–9] are two well-known approaches of AEC. However, these adaptive algorithms could be computationally intensive in particular for large rooms with long reverberation time. Another problem with the conventional approaches is that these adaptive algorithms could diverge because of extraneous disturbances such as background noise or local speech.

The objective of this study is to find an AEC that is computationally efficient and robust in attaining high echo reduction. It is observed that, under many circumstances, the characteristics of echo path do not vary as much to warrant a fully adaptive filter. For example, the echo path will not change much in a video-conferencing speakerphone, where the locations of the loudspeaker and microphone are usually fixed. In this paper, a ‘hybrid’ approach is presented for applications in which echo paths are not drastically varying. One of the examples of applications of this nature is the Bluetooth headsets for mobile phones, where the microphones and the loudspeakers are closely spaced and relative positions are fixed. In the hybrid approach, fixed filters and adaptive filters are combined in a serial architecture. The main portion of acoustic echo has been cancelled using the fixed filter, whereas the system perturbations are accommodated by the adaptive filter.

To enhance the computational efficiency and the cancellation performance of the hybrid approach, sub-band filtering [10–12] and de-correlation [6,11,13,14] are employed. In each sub-band, different number of taps and step size can be used for the adaptive filter to optimize the performance. De-correlation is also used as preprocessing to ‘whiten’ the highly correlated speech signals such that the adaptive algorithm will converge uniformly. In addition, a double-talk detector based on auto-correlation [1,15–18] is incorporated into the AEC to prevent the adaptive filter from diverging in double-talk situations.

In the paper, assessment was made for the AEC in terms of cancellation performance and computation complexity. The fully adaptive filter, the fixed filter and the hybrid sub-band method are compared by experiments according to the standard, ITU-T Recommendation G.168 [22]. In addition, listening tests are also conducted to compare the subjective performance of the AEC. The data of subjective listening tests were analyzed by using the multianalysis of variance (MANOVA) [21]. The simulation and experimental results are discussed and summarized in conclusions.

2. Conventional AEC approaches

In this section, conventional AEC will be briefly reviewed. Discrete-time representations will be used in what follows. In the LEM system of Fig. 1, the echo path $h(n)$ is defined as the system between the loudspeaker input and the microphone output. The far-end signal $x(n)$ is broadcast by the loudspeaker and fed to the microphone through the echo path to become the echo signal $d(n)$. The echo signal is potentially mixed with the local speech $s(n)$ and the background noise $b(n)$ to form the total microphone output $y(n)$. In order to eliminate the echo, a digital filter $\hat{h}(n)$ is used as the replica of the echo path. The problem *per se* is equivalent to a system identification problem. The digital filter can be fixed or adaptive. Nearly perfect cancellation of echoes is possible for the near end by subtracting the estimated echo signal $\hat{d}(n)$ from the microphone output signal $y(n)$.

2.1. Fixed filter method

Straightforward implementation of a fixed filter method is illustrated in Fig. 2 with the switch opened. The system dynamics of the echo path need to be identified off-line prior to the implementation of the fixed filter. The required filter length [1,3] is strongly dependent on the reverberation time T_{60}

$$N_{\text{AEC}} \approx \frac{l_e}{60} f_s \cdot T_{60}, \quad (1)$$

where f_s is the sampling rate in Hz and l_e is the desired echo loss in dB. For example, if $l_e = 20$ dB, $f_s = 8$ kHz and $T_{60} = 400$ ms, then $N_{\text{AEC}} \approx 1000$.

Despite the simplicity of the fixed filter, the cancellation performance could be significantly degraded due to environmental changes.

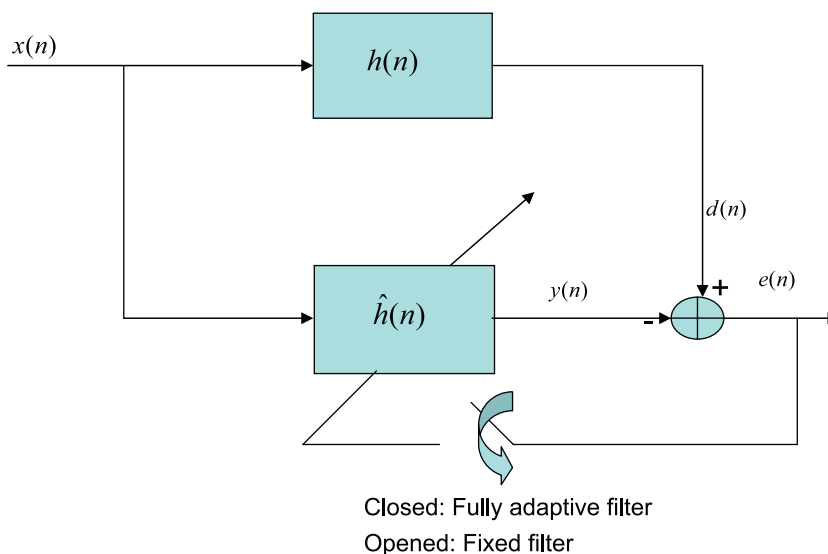


Fig. 2. Block diagram of the fixed filter and the adaptive filter.

2.2. Fully adaptive filter

Fig. 2 shows the block diagram of the fully adaptive filter when the switch is closed. The far-end speech signal $x(n)$ passes the echo path represented by the impulse response function $h(n)$ to mix with the filter output. The adaptive filter $\hat{h}(n)$ attempts to approximate the echo path by minimizing the matching error $e(n)$ between the filter output $y(n)$ and the desired signal $d(n)$. The well-known adaptive filter to achieve this purpose is the LMS algorithm, where the filter coefficient updates according to [1–5,7–9]

$$\mathbf{w}(n+1) = \mathbf{w}(n) + \mu e(n)\mathbf{x}(n), \quad (2)$$

where $\mathbf{w}(n)$ is the filter coefficient and μ is step size. The step size μ determines the convergence behavior and is chosen according to the criterion $0 < \mu < 2/LP_x$, where L is filter length, and P_x is the power of $x(n)$.

The NLMS algorithm [1–5,7–9] is a modified version of the LMS algorithm which takes into account the variation in the input power

$$\mathbf{w}(n+1) = \mathbf{w}(n) + \frac{\mu}{\mathbf{x}^T(n)\mathbf{x}(n) + \beta} e(n)\mathbf{x}(n), \quad (3)$$

where μ and β are parameters to control convergence behavior.

3. Hybrid sub-band implementation of AEC

Although the aforementioned adaptive filters have commonly used in practice, they are generally computationally expensive, especially for large rooms with long reverberation time. In addition, the algorithms could run into slow convergence or even instability problems in the presence of extraneous disturbances such as double talk. Cancellation performance is not uniform throughout the bandwidth for correlated input such as speech signals. To cope with these problems, a modified approach is suggested in this section.

3.1. Sub-band filtering

To enhance the computational efficiency and the cancellation performance, the proposed AEC is realized in sub-band structures [10–12], as shown in Fig. 3(a). Using the analysis filter bank, the far-end signal $x(n)$ and the near-end microphone signal $y(n)$ are split into sub-bands. For example, the frequency response function of an eight-band filter bank is shown in Fig. 3(b). The eight-band filter bank is designed using the Kaiser window approach [10]. Note that different number of taps and step size can be used for different bands in the adaptive filter such that the cancellation performance can be optimized with respect to frequency. To ease the processing, down-sampling and up-sampling can be inserted between the analysis and synthesis filter banks, with the aid of the polyphase representation [10]. Since the adaptive filters are operated at a lower sample rate after down-sampling, the length of the adaptive filters can also be reduced. The error signal $e(n)$ is obtained by recombining and up-sampling the sub-band signals from the filter outputs.

Figs. 3(c) and (d) illustrate the structures of the analysis filter bank and the synthesis filter bank, respectively, where the parameter r is the up/down-sampling ratio. In Fig. 3(c), there are M decimation filters with the common input $y(n)$, whereas in Fig. 3(d), there are M interpolation filters with inputs $y_M(n)$, and their outputs are summed to give the interpolated output. In this paper, the cosine-modulated filter bank (CMFB) [12] is used. In the CMFB design, the prototype filter $p_0(n)$ is a real coefficient linear phase FIR low-pass filter with cutoff $\pi/2M$. The required filter length of the prototype filter N_p is approximately [7]

$$N_p \approx \frac{A_s - 7.95}{14.36\Delta f}, \quad (4)$$

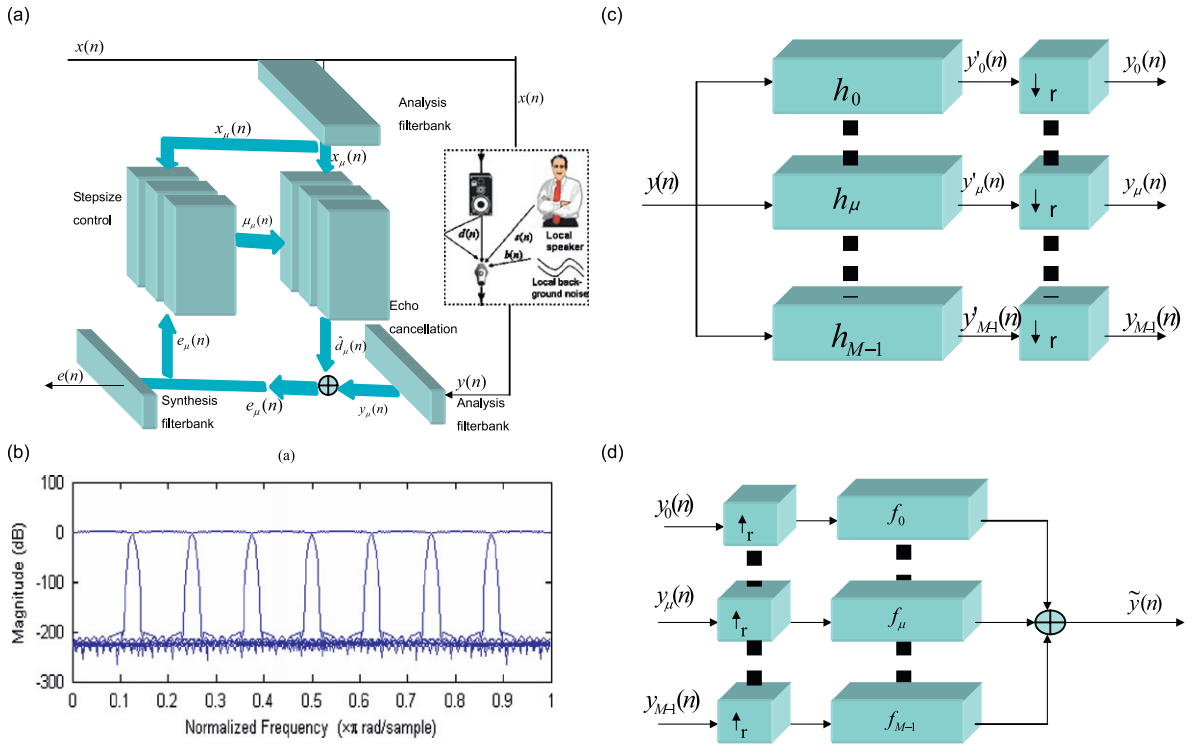


Fig. 3. Schematic diagrams of the sub-band AEC: (a) block diagram of the sub-band AEC, (b) frequency response functions of the cosine-modulated filter bank, (c) structure of the analysis filter bank and (d) structure of synthesis filter bank.

where A_s is the attenuation specification (dB) and Δf is the normalized transition bandwidth (Hz). The impulse response function of the analysis filter banks is given by

$$h_k(n) = 2p_0(n) \cos \left(\frac{\pi}{M(k + 0.5) \left(n - \frac{N}{2} + 1 \right)} + (-1)^k \frac{\pi}{4} \right), \tag{5}$$

where n is the time index, $h_k(n)$ is the coefficient of the k -th analysis filter [12], $p_0(n)$ is the coefficient of the prototype filter, and N is the order of the analysis filter [12]. On the other hand, the synthesis filter bank is given by

$$f_k(n) = h_k(N + 2 - n), \tag{6}$$

where $f_k(n)$ is the coefficient of the k -th synthesis filter [12].

3.2. Hybrid filter structure

Many of the AEC problems involve echo paths that are not drastically time varying. This is particularly true for systems in which the loudspeaker and microphone are fixed in positions and close to each other. For scenarios as such, a hybrid approach that combines a fixed filter $f(n)$ and an adaptive filter $w(n)$ is presented in this section. These two filters are connected in series, as shown in Figs. 4(a) and (b). The adaptive filter is a FIR filter that contains a nominal term $w_0(n) = 1$ and perturbation terms $\Delta \mathbf{w}(n) = [w_1(n) \ w_2(n) \ w_3(n) \ \dots \ w_{L-1}(n)]$, where L is filter length. That is, the overall cancelling path can be written as

$$\hat{h}(n) = f(n) * [\delta(n) + \Delta w(n)], \tag{7}$$

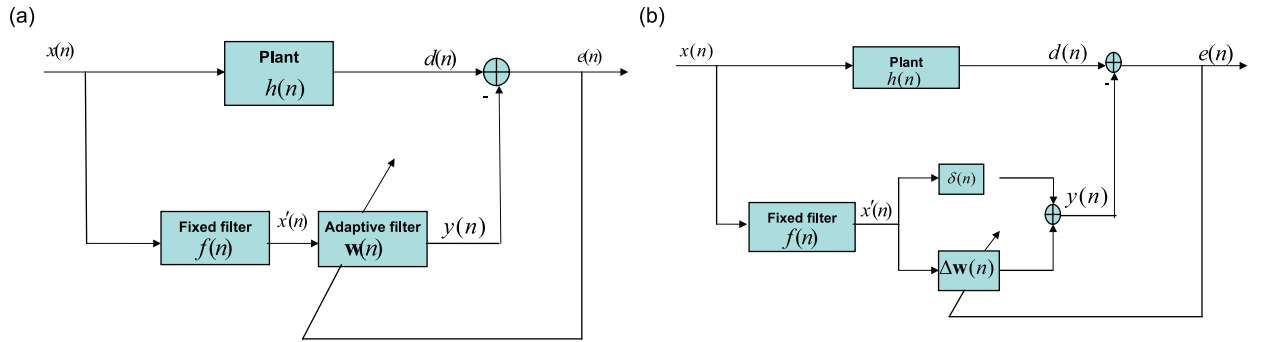


Fig. 4. Block diagram of the hybrid AEC: (a) a fixed filter and an adaptive filter cascaded in series and (b) set the leading coefficient of the adaptive filter to be unity, i.e., $w_0(n) = 1$.

where $\delta(n)$ is the unit pulse sequence, or in the z -domain,

$$\begin{aligned}\hat{H}(z) &= F(z)[1 + \Delta W(z)] \\ &= F(z)[1 + w_1(n)z^{-1} + w_2(n)z^{-2} + \dots + w_{L-1}(n)z^{L-1}].\end{aligned}\quad (8)$$

Using this filter architecture, the nominal term will guarantee the main performance of the fixed filter that was designed off-line, while the perturbation terms will adapt to the small deviations of the echo path. The hybrid approach combines the merits of the fixed filter and the adaptive filter. Very low computation cost is required because the filter $\Delta \mathbf{w}(n)$ is usually very short. Hence, more aggressive step size can be used without destabilizing the adaptive filter.

The remaining problem is how to update the coefficients of the adaptive filter. In the following, the filter update equation will be derived [13,14]. We begin with expressing the error signal by

$$e(n) = d(n) - y(n) = d(n) - \mathbf{w}^T(n)[f(n)*\mathbf{x}(n)],\quad (9)$$

where $f(n)$ denotes the impulse response of fixed filter at the time n ,

$$\mathbf{w}(n) = [w_1(n) \dots w_{L-1}(n)]^T\quad (10)$$

denotes the coefficient vector of the adaptive filter at the time n ,

$$\mathbf{x}(n) = [x(n)x(n-1) \dots x(n-L+1)]^T\quad (11)$$

is the input vector at time n , and L is the order of the adaptive filter. The optimal filter can be obtained by minimizing the instantaneous error squares

$$\hat{\xi}(n) = e^2(n).\quad (12)$$

Instead of a fixed filter, we seek to find the optimal filter coefficients by using gradient search

$$\mathbf{w}(n+1) = \mathbf{w}(n) - \frac{\mu}{2} \nabla \hat{\xi}(n),\quad (13)$$

where μ is the step size and $\nabla \hat{\xi}(n)$ denotes the instantaneous estimate of the gradient at the time n . Note that

$$\nabla \hat{\xi}(n) = \nabla e^2(n) = 2[\nabla e(n)]e(n)\quad (14)$$

and

$$\nabla e(n) = \nabla \{d(n) - \mathbf{w}^T(n)[f(n)*\mathbf{x}(n)]\} = -f(n)*\mathbf{x}(n).\quad (15)$$

Combining Eqs. (13)–(15) leads to the following filter update equation

$$\mathbf{w}(n+1) = \mathbf{w}(n) + \mu e(n)[f(n)*\mathbf{x}(n)].\quad (16)$$

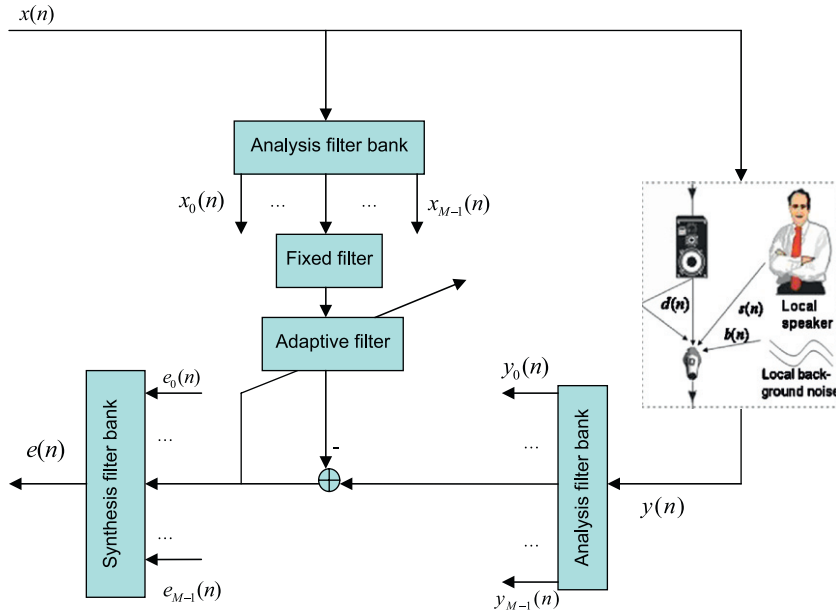


Fig. 5. Block diagram of the hybrid sub-band AEC, where M is the number of sub-bands.

The convergence criterion for the step size is given by [3]

$$0 < \mu < \frac{2}{(L + \Delta)r_{\text{rms}}^2}, \tag{17}$$

where L is filter length, Δ is plant delay, and r_{rms}^2 is the power of the filtered signal $\mathbf{x}'(n) = f(n)*\mathbf{x}(n)$. This resembles the filtered-x LMS algorithm widely used in active noise control [3]. The algorithm can also be extended to the filtered-x NLMS by replacing the step size with

$$\mu' = \frac{\mu}{\mathbf{x}'^T(n)\mathbf{x}'(n) + \beta}. \tag{18}$$

The hybrid filter approach can be combined with the preceding sub-band implementation to achieve maximal computation efficiency and cancellation performance. The hybrid sub-band AEC is illustrated in the block diagram of Fig. 5. M -band analysis and synthesis banks are indicated, while the up/down-sampling modules are not shown ($M = 8$ in the following experiments).

3.3. Double-talk detection

A frequently encountered problem that could destabilize the adaptive filters is ‘double talk.’ When both sides are talking simultaneously, the near-end speech would behave like extraneous noise that makes the adaptive filters diverge. Thus, a double-talk detector [16–18] is required in AEC, as shown in Fig. 6. Double talk can be detected, with the aid of the following correlation coefficient between the microphone signal $y(n)$ and the output of the adaptive filter $\hat{d}(n)$ [1]:

$$\rho_{\text{CL}}(n) = \frac{\left| \sum_{i=0}^{N_C-1} \hat{d}(n-i)y(n-i) \right|}{\sum_{i=0}^{N_C-1} |\hat{d}(n-i)y(n-i)|} \tag{19}$$

whose value is always between 0 and 1. The lower the correlation coefficient, the more likely is the double talk. Once the alarm of double talk is triggered when the correlation drops below certain preset threshold, the adaptation of filters is halted to avoid further divergence.

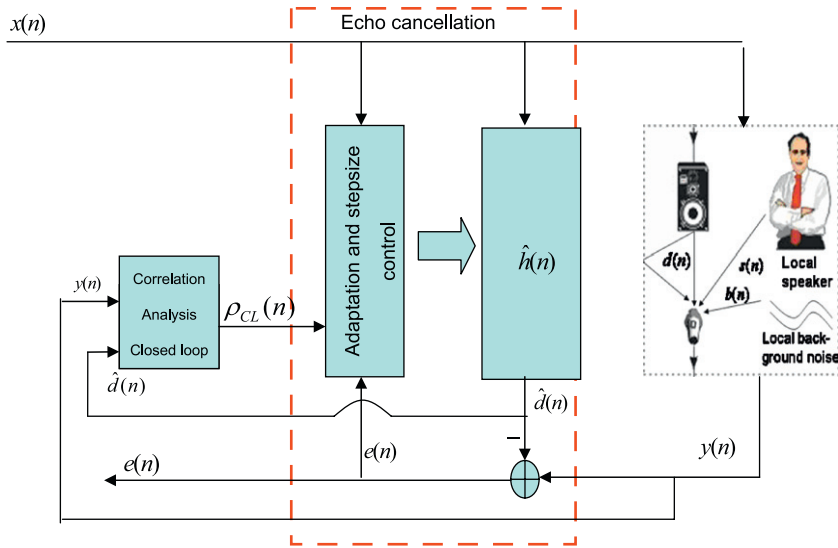


Fig. 6. Double-talk detector embedded in an AEC.

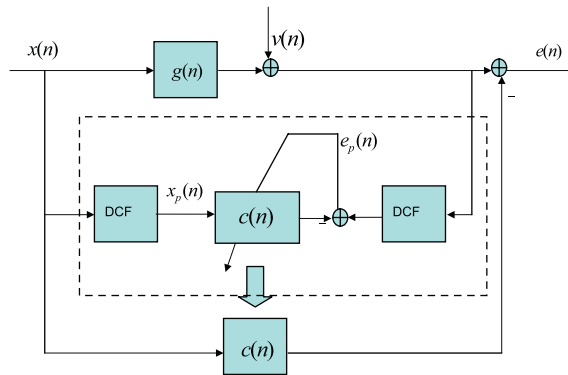


Fig. 7. The improved AEC that makes use of de-correlation filters (DCF).

In practical applications, the above-mentioned double-talk detector tends to be oversensitive. False alarms may be triggered by abrupt fluctuations of the correlation coefficient. To address the issue, the following two-sided single-pole recursion is employed [15]:

$$\bar{\rho}_{CL}(n) = \alpha \bar{\rho}_{CL}(n - 1) + (1 - \alpha) \rho_{CL}(n), \tag{20}$$

where $\bar{\rho}_{CL}(n)$ is the ‘smoothed’ correlation coefficient and

$$\alpha = \begin{cases} \alpha_a, & |\rho(n)| \geq \bar{\rho}_{CL}(n), \\ \alpha_d, & |\rho_{CL}(n)| < \bar{\rho}_{CL}(n). \end{cases} \tag{21}$$

In general, $0 < \alpha_d \ll \alpha_a < 1$, e.g. In this study, α_a is larger than α_d such that the filter adaptation can be halted quickly when double talk occurs and resumed slowly when returning to single-talk scenario. We found it appropriate to select empirical values $\alpha_a = 0.7$ and $\alpha_d = 0.05$ as the *attack* and *decay* time constants, respectively, in the following experiments. Thus, adaptation is quickly halted when correlation drops below

the threshold in double-talk scenario, while adaptation is slowly resumed when the correlation exceeds the threshold in the single-talk scenario. This is found to be a crucial step in stable and robust operation of AEC.

3.4. De-correlation filter

The cancellation performance of the adaptive algorithm can be very poor for speech signals due to eigenvalue disparity or strong correlation of the speech spectrum [3,6]. One way to overcome the problem is to pre-whiten, or de-correlate, the input signal using a de-correlation filter (DCF) [9,11,19,20]. Either simple low-order fixed filters or adaptive filters can be used as DCFs (in the work, we use the former). Fig. 7 illustrates an AEC in conjunction with DCF. The filter update equation of NLMS reads

$$\mathbf{c}(n + 1) = \mathbf{c}(n) + \mu \mathbf{X}(n) \frac{\mathbf{a}(n)\mathbf{a}^T(n)}{\mathbf{x}_f^T \mathbf{x}_f(n)} e(n), \tag{22}$$

where $\mathbf{a}(n)$ denotes the coefficient vector of the DCF, $\mathbf{c}(n)$ denotes the coefficient vector of the adaptive filter, $e(n)$ is the error signal,

$$\mathbf{X}(n) = [\mathbf{x}(n), \mathbf{x}(n - 1), \dots], \tag{23}$$

$$\mathbf{x}_f(n) = \mathbf{X}(n)\mathbf{a}(n). \tag{24}$$

4. Numerical simulations

4.1. The referred standard

Performance of the AECs proposed in this paper are evaluated according to the standard ITU-T Recommendation G.168 [22]. With reference to Fig. 8, the following terms of echo cancellation are defined.

4.1.1. Performance definitions

ERL (dB): the attenuation of a signal from the receive-out port (R_{out}) to the send-in port (S_{in}) of an echo canceller, due to transmission and hybrid loss, i.e., the natural loss in the (cancelled end) echo path.

ERLE (dB): the attenuation of the echo signal as it passes through the send path of an echo canceller. This definition specifically excludes any nonlinear processing (NLP) at the output of the canceller to provide further attenuation. For the LEM system with an echo-cancellation filter (ECF) shown in Fig. 1(b), *ERLE* can be

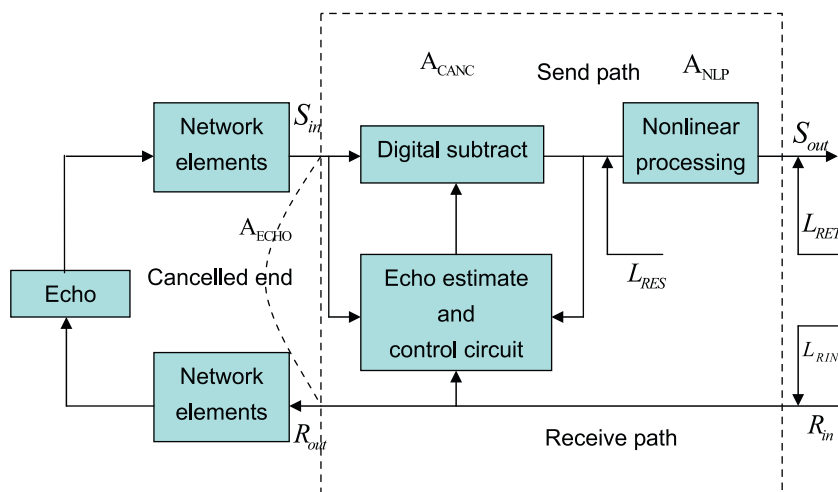


Fig. 8. Schematic diagram of the definitions for echo reduction given in ITU-T Recommendation G.168 [22].

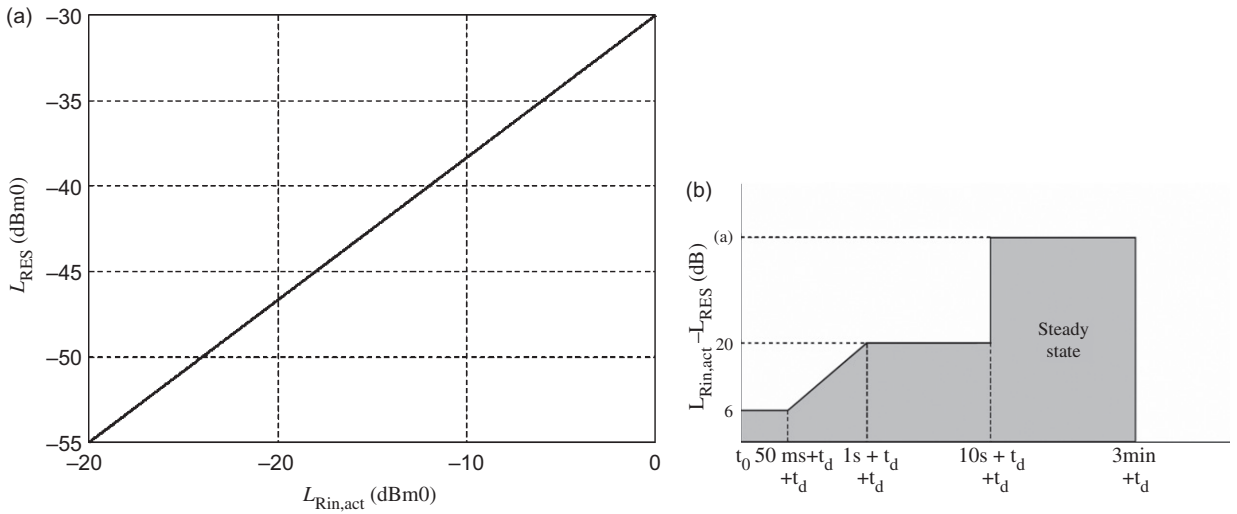


Fig. 9. The performance requirement of AEC given in ITU-T Recommendation G.168 [22]: (a) relationship between received input level ($L_{in,act}$) and residual echo level (L_{RES}) with NLP disabled and (b) convergence characteristic with NLP disabled.

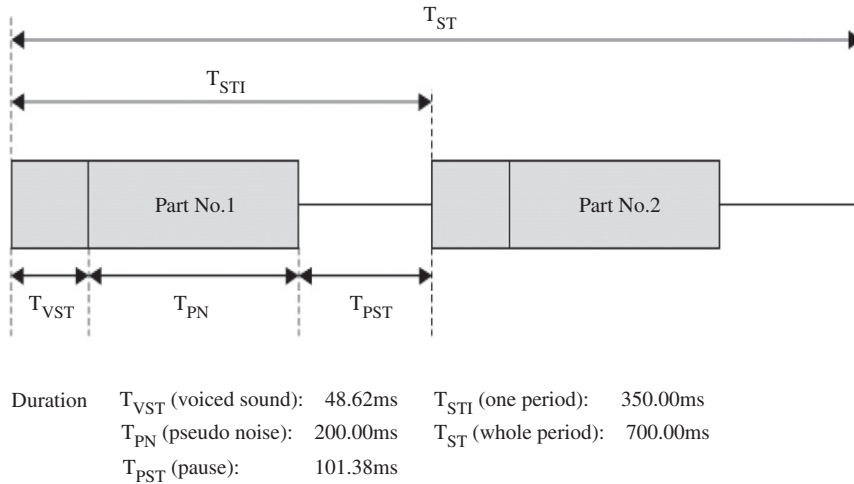


Fig. 10. The composition of the composite source signal (CSS) [22].

calculated as

$$ERLE = 10 \log_{10} \frac{E[d^2(n)]}{E[(d(n) - \hat{d}(n))^2]} \text{ (dB)}, \quad (25)$$

where $E[\bullet]$ denotes the expected value that can be estimated using, say, 100 averages of data.

4.1.2. Requirement of standard

Only the performance of AEC processing is examined in the paper. Any NLP or residual noise reduction at the output of the canceller is disabled. Performance requirement stated in ITU-T Recommendation G.168 is illustrated in Fig. 9. Fig. 9(a) shows the relationship between the received input level ($L_{in,act}$) and the residual echo level (L_{RES}) with NLP disabled. Fig. 9(b) shows the convergence characteristic with NLP disabled. For all values of $L_{Rin,act} \geq -30$ dB and ≤ 0 dB and for all values of $ERL \geq 6$ dB and echo path delay, $t_d \leq \Delta$ ms, the

loss $L_{R_{in,act}} - L_{RES}$ should be greater than or equal to that shown in Fig. 9(b), where Δ represents the maximum delay of the echo path and t_0 is the delay of the receive path. After $10 + t_d + t_0$ seconds, the loss $L_{R_{in,act}} - L_{RES}$ should be greater than or equal to that shown in Fig. 9(a). From Fig. 9(b), the convergence time is defined as the time required for the sum of ERL and $ERLE$ to exceed 20 dB. In summary, if $L_{R_{in,act}} = 0$ dB, and $ERL \geq 6$ dB, then $ERLE$ must be greater than 24 dB, with convergence time less than 1 s.

4.1.3. Composite source signal

According to ITU-T Recommendation G.168, a special test signal, the composite source signal (CSS), is used for testing the AEC under single- and double-talk conditions. CSS consists of two bursts of voiced sound, pseudo noise, and pause, as shown in Fig. 10.

The voiced signal part of CSS is the conditioning signal intended to activate possible speech detectors in voice-controlled systems and to reproduce voiced sounds of real speech in general. The voiced sound has the spectrum extending over approximately 200–4000 Hz. The pseudo noise is created using a random-phase generator with flat spectrum. The duration of the pause is between 100 and 150 ms. To achieve a long term offset free sequence, the CSS should be inverted in amplitude (phase shifted by 180°). More details of how to implement CSS can be found in ITU-T Recommendation G.168 [22].

4.2. Simulation of the hybrid AEC

Simulations are carried out for validating the hybrid AEC and the de-correlation method. Fig. 11 presents the results of the hybrid approach. In this simulation of hybrid method, we test for whether hybrid method is more robust or not. In this simulation, a 512-tapped FIR filter is used to represent an echo path measured for a speakerphone in an office. Another 512-tapped FIR filter is used as the fixed filter in the hybrid AEC. For white noise input, $ERLE = 53$ dB is attained at the sampling rate 48 kHz. Next, we alter the echo path slightly by a filter $\Delta P(z) = 1 + z^{-1} + z^{-2} + z^{-3}$ to simulate the perturbation to the echo path. AEC performance is evaluated with white noise input. In Fig. 11, the dotted line represents the $ERLE$ obtained using the fixed filter, which stays at about 32 dB. The performance is degraded apparently because of the echo path perturbation. The dashed line represents the $ERLE$ obtained using the 512-tapped fully adaptive filter, which settles at

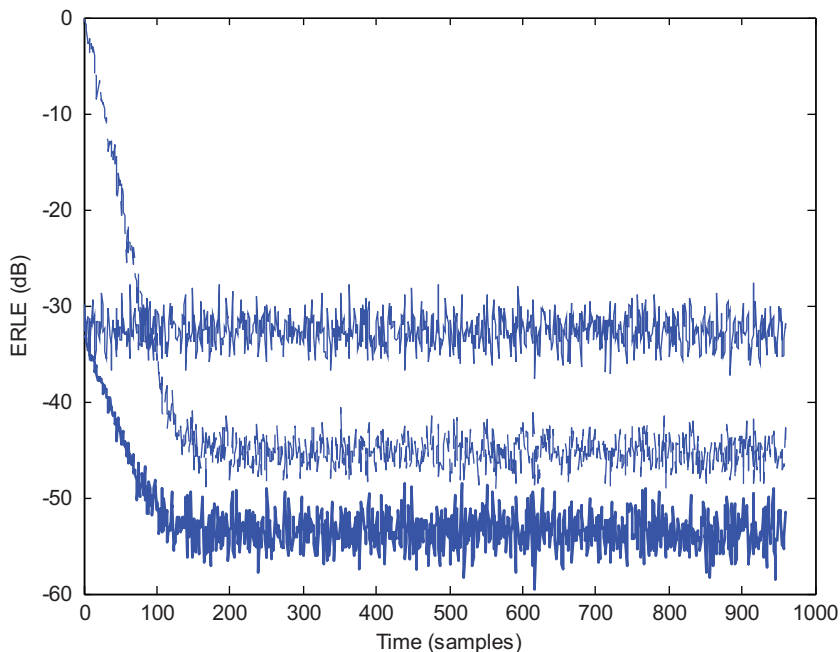


Fig. 11. The simulation results of the hybrid AEC. The echo path is altered by a three-tapped FIR filter (dotted line: fixed filter method; dash line: pure adaptive method; solid line: hybrid method).

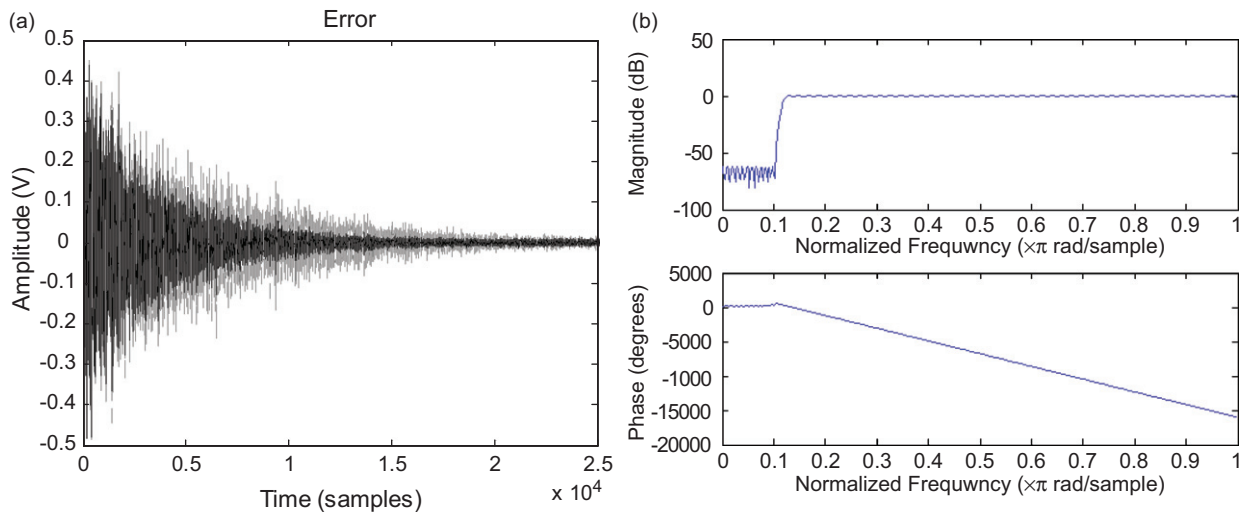


Fig. 12. The simulation of the de-correlation filter: (a) the error signal obtained using the DCF (dashed line: NLMS; solid line: de-correlated NLMS) and (b) the frequency response of the fixed DCF.

around 45 dB, slightly better than that of the fixed filter. Finally, the hybrid AEC along with a three-tapped adaptive filter is tested. The solid line represents the *ERLE* obtained using the hybrid AEC, which settles at around 53 dB, best performance attainable for the scenario without echo path perturbations.

On the other hand, the simulation is also performed to investigate the de-correlation method. The same echo path is used in this simulation. A clip of female speech is used as the test signal. We here track the echo path by using a 1024-tapped adaptive filter with the step size $\mu = 0.01$. Fig. 12(a) shows the results obtained using a linear phase DCF with high-pass characteristics shown in Fig. 12(b). It can be observed from the error signal of Fig. 12(a) that the de-correlated NLMS algorithm converges faster than the ordinary NLMS algorithm.

5. Objective and subjective experiments

Objective and subjective experiments were conducted for validating the proposed AEC. Fig. 13 shows the experimental arrangement, where a PC multimedia loudspeaker and a microphone spaced apart by 40 cm in an office. Fig. 14 shows the impulse response of the echo path. The LEM has reverberation time 458 ms and natural attenuation, $ERL = 6$ dB. All AEC algorithms are implemented on the platform of a fixed-point DSP (ADI, BF533), operated at the sampling rate 48 kHz. In subjective experiments, subjects communicate with each other by using a peer-to-peer Internet telephony network, Skype[®]. The echo cancellation function of Skype[®] is disabled before connecting to our DSP-based AEC module.

5.1. Objective experiments

5.1.1. Double-talk detector

In this section, the double-talk detector is investigated. The test signals are two excerpts of speech. The female speech filtered by echo path is used as the acoustic echo from the far end, whereas the male speech is used as the local speech at the near end. Fig. 15 shows the correlation coefficient calculated using Eq. (21) and the two-sided one-pole recursion. The threshold for judging double talk is selected to be 0.8. Fig. 15 illustrates a switching single (female speech only)-double (circled by red)-single talk scenario, where the correlation coefficient drops to mostly below the threshold during the double talk. Adaptive filters should be disabled whenever double talk is detected.

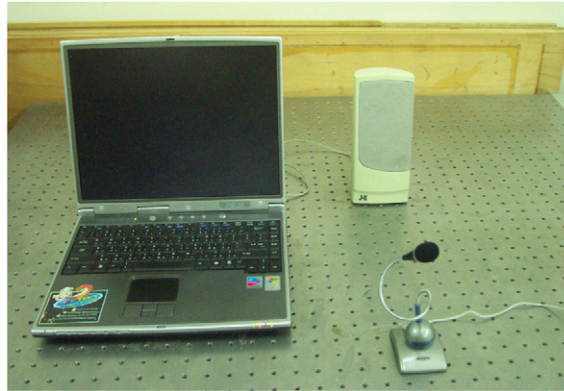


Fig. 13. The experimental arrangement in an office showing the relative positions of the loudspeaker and the microphone for experiments.

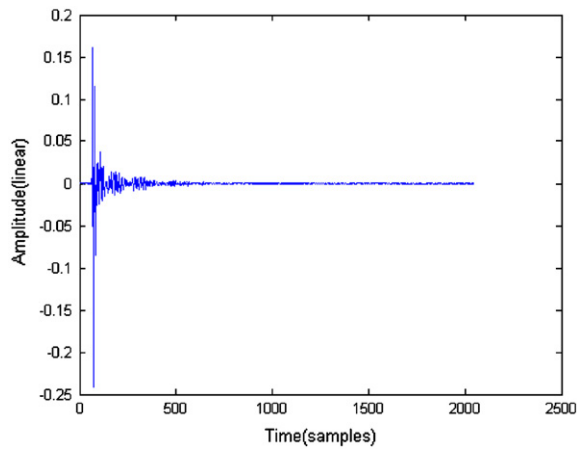


Fig. 14. The impulse response of the echo path for the experiment of the double-talk detector.

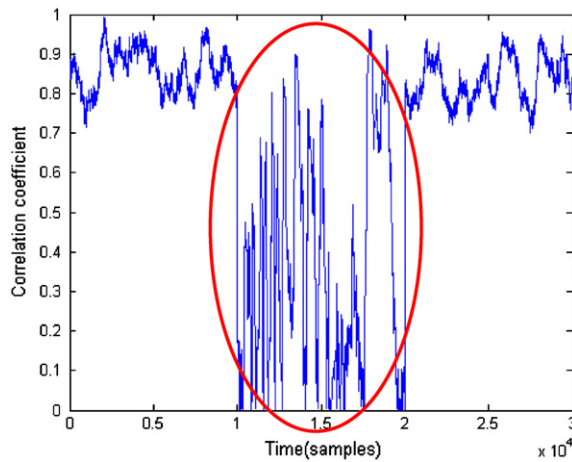


Fig. 15. The smoothed correlation coefficient obtained using the double-talk detector. The red circle indicates the double-talk region.

5.1.2. The hybrid sub-band AEC versus fully adaptive AEC

In this section, the performance of the AECs using the hybrid sub-band method and the fully adaptive method will be compared. The test signal is the female speech with the spectrum shown in Fig. 16. Since the

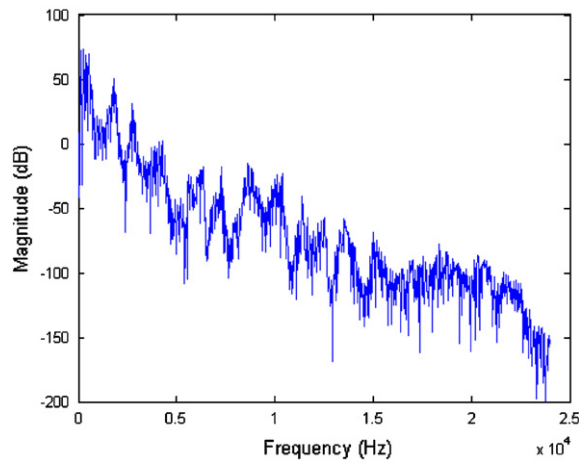


Fig. 16. The spectrum of a female speech signal for the experiment of the hybrid sub-band AEC.

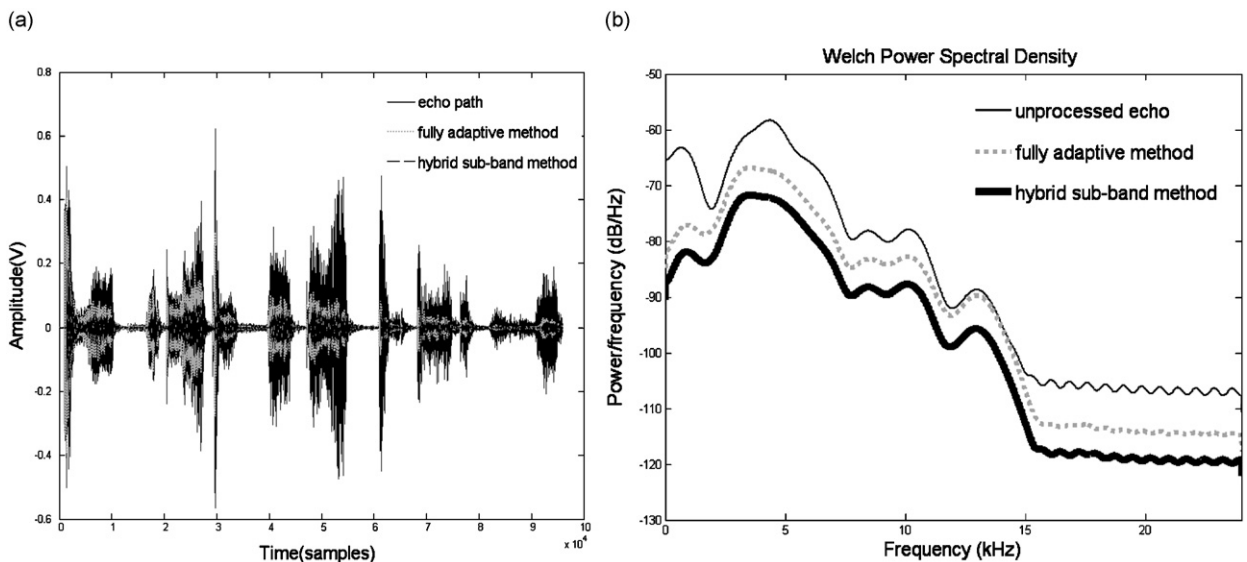


Fig. 17. The echo reduction performance of AEC. The test signal is the female speech: (a) the time-domain waveform and (b) the power spectrum.

test signal concentrates in a frequency range of 0–2.5 kHz, we down-sampled the signals to 8 kHz. Next, we divide the full band into eight sub-bands. Adaptive filters were implemented in the 1st–8th sub-bands. The adaptive filter of pure adaptive structure is 1024-tapped and step size is 0.1. The fixed filters of the hybrid AEC are 160-tapped for the 1st–5th sub-bands, and 75-tapped for the 6th–8th sub-bands. The step size of all adaptive LMS filters is 0.01 and all adaptive filters are 16-tapped. The error signals obtained using the fully adaptive AEC (black dotted line) and the hybrid sub-band AEC (gray dotted line) are compared in Fig. 17(a). The power spectrum is shown in Fig. 17(b). From the results, the performance of the hybrid sub-band AEC ($ERLE = 30.5$ dB) is significantly better than the fully adaptive AEC ($ERLE = 22.4$ dB).

Next, the CSS was used as the test signal, with the other conditions being equal. The error signals obtained using the fully adaptive AEC (black solid line) and the hybrid sub-band AEC (gray dotted line) are compared in Fig. 18(a). The power spectrum is shown in Fig. 18(b). From the results, the performance of the hybrid sub-band AEC ($ERLE = 28.6$ dB) again is better than the fully adaptive AEC ($ERLE = 20.4$ dB).

Fig. 19 plots the $ERLE$ of the AECs versus time. Test signal is the female speech. From the plot, the converging time of the fully adaptive AEC and the hybrid sub-bands AEC was found to be 392 and 280 ms,

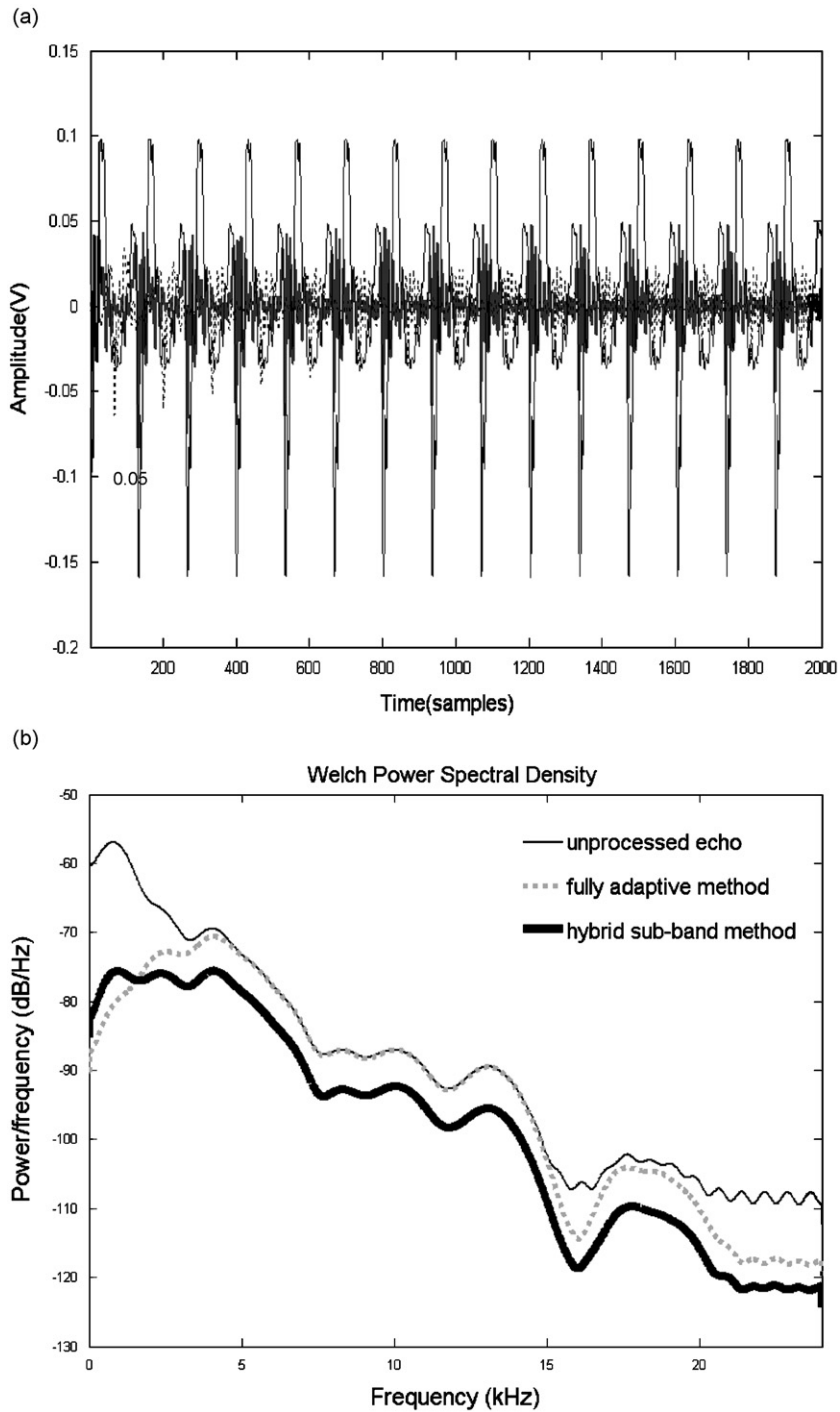


Fig. 18. The echo reduction performance of AEC. The test signal is the CSS: (a) the time-domain waveform and (b) the power spectrum.

respectively. The convergence time is defined as the time required for the sum of *ERL* and *ERLE* to exceed 20 dB. The hybrid sub-band AEC has a faster convergence speed than that of the fully adaptive AEC. Overall, the hybrid sub-band AEC is superior to the conventional fully adaptive AEC in terms of cancellation performance and convergence speed.

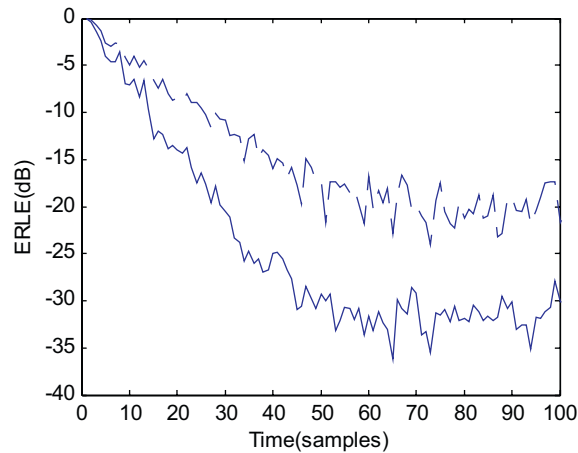


Fig. 19. The *ERLE* curves to compare the speed of convergence for the hybrid sub-band AEC and the fully adaptive method.

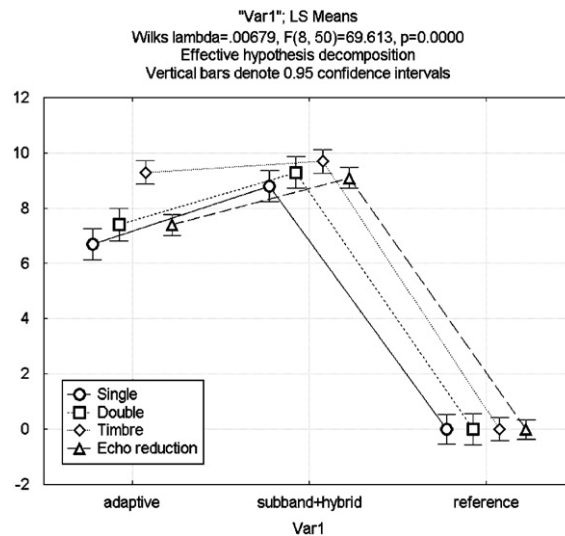


Fig. 20. The results of the listening-only test, with mean and spread (5–95 percent) of the score indicated.

5.1.3. Cancellation performance in the face of echo path perturbation

In this section, the cancellation of the proposed hybrid sub-band algorithm in response to a sudden change of the echo path is examined. Fig. 22(a) shows the experimental arrangement for the echo path perturbation. The echo path perturbation is created by placing a plywood board on the side of the line connecting the loudspeaker and the microphone. Marked difference of the frequency response functions between the original and the perturbed echo paths are clearly visible in Fig. 22(b). The test signal is an 8 s clip of female speech. The settings of hybrid sub-band method remain the same as the previous section. During the operation of the hybrid AEC, a plywood board was suddenly placed at the side of the line connecting the loudspeaker and the microphone. The residual error of the cancelled echo signal is compared with the unprocessed signal in Fig. 23(a). The power spectrum of the residual error after the adaptive filter converges in also is compared with the unprocessed signal in Fig. 23(b). Despite the sudden perturbation of the echo path, the hybrid AEC is still able to converge rapidly to the minimal residual error. One can barely notice the characteristics of the acoustical system have ever changed. Some 25 dB maximal reduction can be achieved using such hybrid

sub-band structure. This promising result reveals the potential of the proposed hybrid sub-band AEC in coping with echo path perturbations in practical applications.

5.2. Subjective experiments

In addition to objective experiments, listening tests were undertaken to assess the subjective performance of the AEC. Two kinds of listening tests, the ‘listening-only’ test and the ‘conversational’ test [19] were carried out in the experiment. Twenty subjects participated in the listening tests. In the listening-only test, subjects give their rating of perceived echo reduction and timbral quality at the near end by listening to the demonstration files. In the conversational test, subjects on both near and far ends give their rating of perceived echo reduction and timbral quality while communicating with each other via Skype®.

In the listening tests, the hybrid sub-band AEC was compared with the fully adaptive AEC. The fixed filter of hybrid structure was 160-tapped for the 1st–5th sub-bands, whereas the fixed filter of hybrid structure was 75-tapped for the 6th–8th sub-bands. The step size of all adaptive filters used in the hybrid AEC was 0.01 and

Table 1
The MANOVA results of the listening-only test.

Test module	Significance value			
	Echo reduction	Single talk	Double talk	Timbre
Listening-only test	0.000050	0.000454	0.001311	0.291625

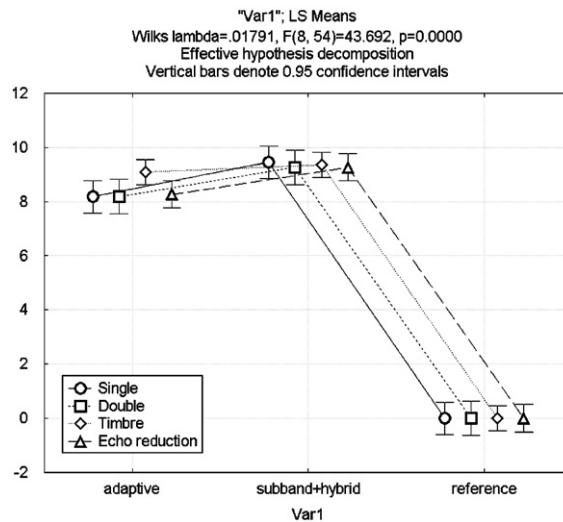


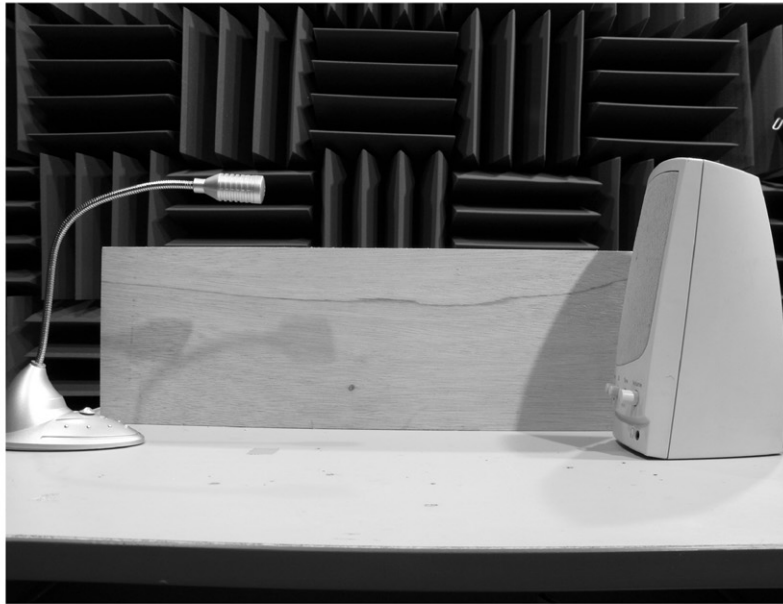
Fig. 21. The results of the conversational test, with mean and spread (5–95 percent) of the score indicated.

Table 2
The MANOVA results of the conversational test.

Test module	Significance value			
	Echo reduction	Single talk	Double talk	Timbre
Conversational test	0.030764	0.020945	0.056050	0.501421

all adaptive filters were 16-tapped. On the other hand, the fully adaptive filter was 1024-tapped and step size 0.1 was used. All algorithms are implemented on the platform of a fixed-point DSP, operated at the sampling rate 48 kHz. However, the speech signals were down-sampled to 8 kHz. Four subjective indices in

(a)



(b)

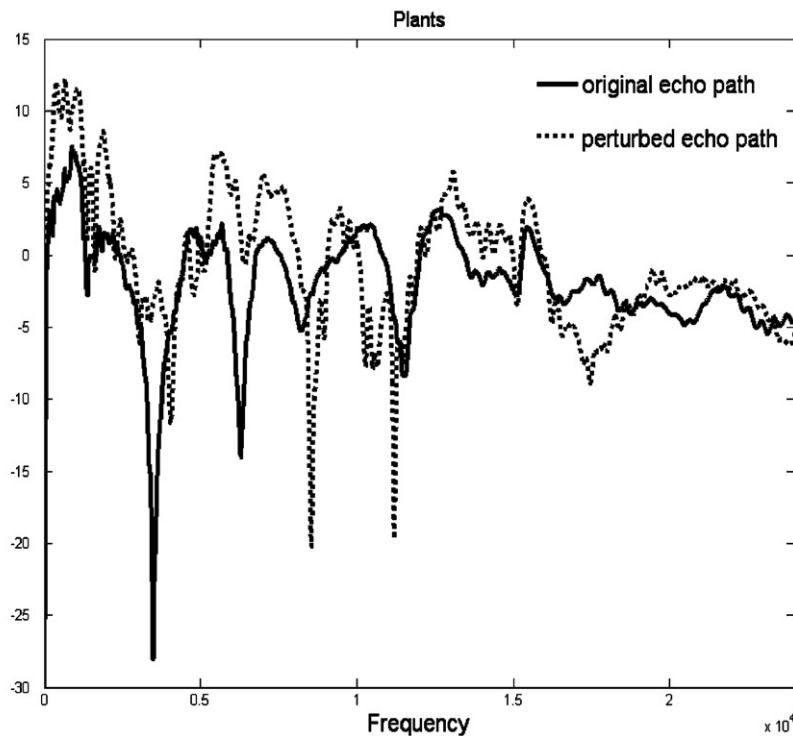


Fig. 22. The experimental arrangement to simulate the echo path perturbation. The echo path is perturbed by placing a plywood board on the side of the line connecting the microphone and the loudspeaker: (a) the photo of the experimental arrangement and (b) the frequency response functions of the original and the perturbed echo paths.

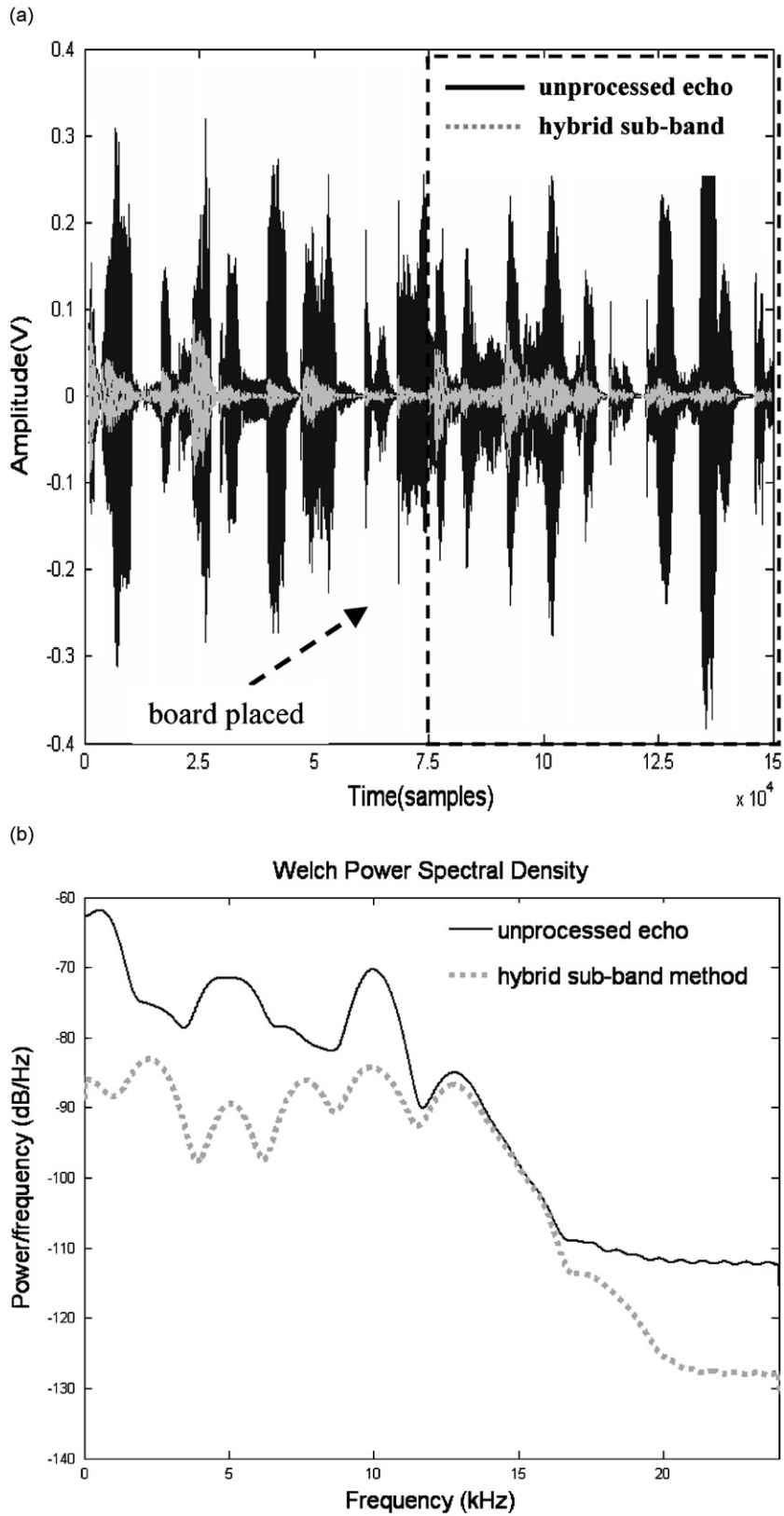


Fig. 23. The echo reduction performance of the echo path perturbation obtained using the hybrid sub-band method. The test signal is a female speech: (a) the time-domain waveform and (b) the power spectrum.

Table 3

Comparison of computational requirement and objective performance of the fully adaptive AEC and the hybrid sub-band AEC.

	Additions	Multiplications	Convergence time (ms)	ERLE (dB)
Hybrid sub-band	894	900	280	30.5
Fully adaptive	2048	2048	392	22.4

the test: (1) echo reduction in the single-talk scenario; (2) echo reduction in the double-talk scenario; (3) timbral quality of the local speech; and (4) overall preference of echo reduction. The subjects participating in the tests were instructed with definitions of the subjective indices before the listening tests. During the tests, the subjects were asked to respond on a questionnaire with the subjective indices placed on the scale from 1 to 10. Fig. 20 shows the result of the listening-only test. The scores from all subjects were also processed by using the MANOVA [21] to justify the statistical significance of the test results. The average, 5–95 percent bracket and the significance level of the grades were shown in the analysis. Cases with significance levels below 0.05 indicate that statistically significant difference exists among methods. From Fig. 20 and Table 1, there is no significant difference in timbral quality between the hybrid sub-band AEC and the fully adaptive AEC. For the other indices, however, the hybrid sub-band AEC predominantly outperformed the fully adaptive AEC.

Fig. 21 shows the result of the conversational test, with MANOVA summarized in Table 2. Again, we found no significant difference in timbral quality between the hybrid sub-band AEC and the fully adaptive AEC. For the other indices, the hybrid sub-band AEC predominantly outperformed the fully adaptive AEC, albeit the score of echo reduction in the double-talk scenario was marginally significant (Figs. 22 and 23).

It can be concluded from the listening-only test and the conversational test that the hybrid sub-band AEC has attained better performance than the fully adaptive AEC in all subjective attributes, with comparable timbral quality.

6. Conclusions

An efficient hybrid sub-bands AEC has been presented for reducing acoustic echoes in LEM systems with stationary characteristics. In the AEC, the fixed filter has accounted for the major portion of echo reduction, while the adaptive filter deals with the residual echo due to the perturbed echo path. The efficiency and performance have been maximized by using the sub-band structure in which different length and step size of filter can be used independently for each sub-band. For example, because the spectrum of most speech signals concentrate in the band 0–2.5 kHz, longer filters can be used in this frequency range.

Objective tests were undertaken to compare the hybrid sub-band AEC with the fully adaptive AEC in terms of performance and convergence time. Computational requirement and objective performance of the fully adaptive AEC and the hybrid sub-band AEC are compared in Table 3. Using the proposed AEC, computational efficiency, convergence time and echo reduction performance are significantly improved over the conventional approach.

Subjective listening tests were conducted to assess the performance of the AECs according to the standard ITU-T Recommendation G.168 and the data were analyzed by using the MANOVA method. The listening tests have revealed that the hybrid sub-band AEC is subjectively superior to conventional fully adaptive AEC in many aspects of echo reduction with comparable timbral quality.

In the future, we wish to combine the present AEC and noise reduction and microphone array technologies into an all-in-one system appropriate for video conferencing and hands-free car kit systems. It is hoped that the results presented in this paper would shed some light on the design strategies for the future AEC systems that meet the ever increasing needs of telecommunication.

Acknowledgment

The work was supported by the National Science Council of Taiwan, Republic of China, under the project number NSC 95-2221-E-009-179.

References

- [1] E. Hansler, G. Schmidt, *Acoustic Echo and Noise Control: A Practical Approach*, Wiley, New York, 2004.
- [2] K.J. Astrom, B. Wittenmark, *Adaptive Control*, Addison-Wesley, New York, 1995.
- [3] S.M. Kuo, D.R. Morgan, *Active Noise Control System*, Wiley, New York, 1996.
- [4] M. Brandstein, D. Ward, *Microphone Arrays*, Springer, New York, 2001.
- [5] Y. Huang, J. Benesty, *Audio Signal Processing. For Next-Generation Multimedia Communication Systems*, Kluwer Academic Publishers, London, 2004.
- [6] C. Breining, P. Dreiscitel, E. Hansler, A. Mader, B. Nitsch, H. Puder, T. Schertler, G. Schmidt, J. Tilp, Acoustic echo control. An application of very-high-order adaptive filters, *IEEE Signal Processing Magazine* 16 (1999) 42–69.
- [7] B. Widrow, S.D. Stearns, *Adaptive Signal Processing*, Prentice-Hall PTR, Englewood Cliffs, NJ, 1985.
- [8] B. Farhang-Boroujeny, *Adaptive Filters Theory and Application*, Wiley, New York, 2000.
- [9] S. Haykin, *Adaptive Filter Theory*, Prentice Hall, Englewood Cliffs, NJ, 1986.
- [10] P.P. Vaidyanathan, *Multirate System and Filter Bank*, Prentice-Hall PTR, Englewood Cliffs, NJ, 1993.
- [11] H. Yasukawa, S. Shimada, An acoustic echo canceller using sub-band sampling and decorrelation methods, *IEEE Transactions on Signal Processing* 4 (1993) 926–930.
- [12] Y.P. Lin, P.P. Vaidyanathan, A Kaiser window approach for the design of prototype filter of cosine modulated filter banks, *IEEE Signal Processing Letters* 5 (1998) 132–134.
- [13] D.R. Morgan, An analysis of multiple correlation cancellation loops with a filter in the auxiliary path, *IEEE Transactions on Acoustics, Speech, and Signal Processing* 28 (1980) 454–467.
- [14] L.J. Eriksson, M.C. Allie, C.D. Bremigan, J.A. Gilbert, Weight vector analysis of an RLMS adaptive filter with on-line auxiliary path modeling, *Proceedings of the ICASSP 89*, IEEE, Glasgow, UK, May, 1989, pp. 2029–2032.
- [15] S.L. Gay, J. Benesty, *Acoustic Signal Processing for Telecommunication*, Kluwer Academic Publishers, London, 2000.
- [16] T. Gansler, M. Hansson, C.J. Ivarsson, G. Salomonsson, A double talk detector based on coherence, *IEEE Transactions on Communications* 44 (11) (1996) 1421–1427.
- [17] P. Heitkamper, An adaptation control for acoustic echo cancellers, *IEEE Signal Process* 4 (1997) 170–172.
- [18] H. Ye, B.X. Wu, A new double talk detection algorithm based on the orthogonality theorem, *IEEE Transactions on Communications* 39 (11) (1991) 1542–1545.
- [19] R. Frenzel, M.E. Heonecke, Using prewhitening and step size control to improve the performance of the LMS algorithm for acoustic echo compensation, *IEEE International Symposium on Circuits and Systems* 4 (1992) 1930–1932.
- [20] S. Yamamoto, S. Kitayama, J. Tamura, H. Ishigami, An adaptive echo canceller with linear predictor, *Transactions of the IECE of Japan* 62 (1979) 851–857.
- [21] G. Keppel, S. Zedeck, *Data Analysis for Research Designs*, W.H. Freeman and Company, New York, 1989.
- [22] ITU-T Rec. G.168, *Transmission systems and media, digital systems and networks*, International Telecommunications Union, Geneva, Switzerland, 2004.