A Robust Stereophonic Acoustic Echo Canceler Using Delayless Subband Adaptive Filter

*Won-Cheol Lee and *Young-Min Cho

※This research work was supported by Korea Research Foundation under contract no.04-E-0327.

Abstract

This paper proposes a new stereophonic acoustic echo canceler with deploying delayless subband adpative filters. Due to the strong correlation between stereo signals, a stereophonic acoustic echo canceler is suffering from the slow convergence and the misalignment for estimating impulse responses corresponding to true echo paths at receiving room. Specially, dual adaptive filters for echo cancellation are significantly affected by the abrupt change of the transmission room environment so that the impariments on voice communication could be experienced. To prevent these performance degradations, this paper proposes a robust subband echo canceler employing pre-processing block so as to enhance the convergence speed and provide the insusceptibility to the environment change at transmission room.

Ⅰ. Introduction

In recent years, the monoaural acoustic echo canceler has been widely utilized on various hand-free communication units specially in the teleconferencing system. And its major role is the suppression of undesired echo signals which are induced by cross-coupling effect between the speaker and the microphone at the same conferencing site. So if nothing were done to this echo signal, subsequently it is delivered back to the remote site. As a result the remote talker listens his own voice after designated time delay. Eventhough communication channel is characterized as distortion-free, this acoustic echo leads talker to feel discomfort over the remote conversation.

Referring to relevant articles [1]-[4], many schemes corresponding to monoaural echo cancellation have been well exploited. But it is noteworthy that the monoaural teleconferecing system could not provide the naturalness and realistic presence over the voice communication. To enhance the quality of remote conversation, the multichannel version of teleconferencing system becomes preferable because of its capability of creating more realistic sound environments. However in the situation of deploying multi-channel conference system, coupling effect bringing out acoustic echos between speakers and microphones becomes more complicated to be mitigated. As a simpler version of multichannel system, this paper will consider the stereophonic teleconferencing system and its deployable structure of the echo canceler.

*Department of Information and Telecommunication Engineering, Soongsil University
Manuscript Received : January 9, 1998.

As mentioned in [1], unlike monoaural echo canceler, stereophonic echo canceler encounters a variety of inherent problems which are still left to be unsolved. To overcome these fundamental problems such as the misalignment, the slow convergence and the variation of suboptimal solutions to the change of far-end envrioment, several modified schemes have been proposed in [5]-[7].

In this paper, the structure of stereophonic acoustic echo canceler and its functional aspects are discussed in section Ⅱ. And fundamental problems existing on the stereophonic acoustic echo canceler are investigated in section Ⅲ. Section Ⅳ introduces the structure of the full-band stereophonic acoustic echo canceler including the pre-processing block. Section Ⅴ provides the proposed subband version of stereophonic acoustic echo canceler referring to the aspects discussed in section Ⅳ. In section Ⅴ, the superior performances of the proposed subband stereophonic echo canceler will be verified via conducting computer simulations under various situations. Finally the concluding remarks follow in section Ⅵ.

Ⅱ. Conventional Stereophonic Echo Canceler

A stereophonic teleconference system basically consists of two pairs of microphones and speakers at each teleconferencing site and four adaptive filters need to be implemented either at transmission or receiving sites to cancel out unwanted echo signals. The basic structure of the stereophonic echo canceler is depicted in Fig. 1.

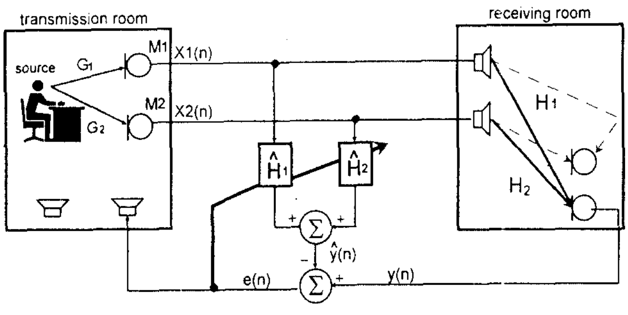


Figure 1. Basic structure of the stereophonic acoustic echo canceler.

Here G_1 and G_2 represent acoustic pickup transfer functions which convey source signal into microphone M_1 and M_2. Here due to the fact that the single source signal impinges two differently located microphones, high degree of correlation between two microphone inputs might be experienced. Furthermore since at receiving room four propagation channels represented by $\{H_{ij}\}_{i,j=1,2}$ arise between speakers and microphones in symmetric fashion, without a loss of generality, two acoustic propagation channels, i.e., $H_1 \triangleq H_{i1}$ and $H_2 \triangleq H_{i2}$ which are formed between a fixed speaker and two distinct microphones, are sufficient to be considered. In Fig. 1, the adaptive filters $\hat{H}_1$ and $\hat{H}_2$ are used to regenerate echo replica via tracking the time-varying true echo path systems H_1 and H_2 respectively. The composite echo signal $y(n)$ in Fig. 1. is formed by adding up two distinct echo path system outputs.

Assume that far-end talker's speech is not in presence, the composite received signal at microphone can be expressed by

$$y(n) = x_1(n) * h_1(n) + x_2(n) * h_2(n) \tag{1}$$

where $*$ is a convolution operator, and $x_1(n)$ and $x_2(n)$ are speaker outputs and $h_1(n)$ and $h_2(n)$ are impulse response sequences associated with corresponding true echo path systems respectively. Suppose that the echo canceler is implemented by FIR-type of filter whose coefficients are updated in recursive manner, the echo replica $\hat{y}(n)$ is comprised of two distinct FIR filter outputs given by

$$\hat{y}(n) = \hat{H}_{1n}^T X_{1n} + \hat{H}_{2n}^T X_{2n} \tag{2}$$

where A^T is a transpose of matrix A, $\hat{H}_{in}$, $i = 1, 2$, is L-dimensional weight vector at the time instant n associated with each adaptive filter, and X_{in} is the reference input vector expressed as

$$X_{in} = [x_i(n), x_i(n-1) \cdots x_i(n-L+1)], \quad i = 1 \rightarrow 2. \tag{3}$$

Then the residual cancellation error signal is represented by

$$e(n) \triangleq y(n) - \hat{y}(n) = y(n) - \hat{H}_n^T X_n \tag{4}$$

where

$$X_n = [X_{1n}^T X_{2n}^T]^T \tag{5}$$

and

$$\hat{H}_n = [\hat{H}_{1n}^T \hat{H}_{2n}^T]^T \tag{6}$$

Thus, in order to satisfy the perfect alignment such that converged adaptive filter weight vectors $\hat{H}_{1n}$ and $\hat{H}_{2n}$ should be identical to the true echo path systems H_1 and H_2 respectively.

Ⅲ. Fundamental Problems on Stereophonic Acoustic Echo Canceler

This section discusses the major disruptive problems which are directly affecting the performance of overall echo cancellation system.

3.1 Misalignment and its Effects

Suppose that the length of impulse responses corresponding to the acoustic pick-up transfer functions and echo path systems are finite. Then speaker output signal vectors $\{X_{1n}, X_{2n}\}$ and transmission room impulse response vectors $\{G_1, G_2\}$ satisfy the following relationship,

$$X_{1n}^T G_2 = X_{2n}^T G_1. \tag{7}$$

A Wiener solution $\hat{H}_{opt} \triangleq [\hat{H}_{1opt}^T \hat{H}_{2opt}^T]$ for dual adaptive filters can be obtained by solving the following system of linear equations given by

$$R\hat{H}_{opt} \triangleq \begin{bmatrix} R_{11} & R_{12} \\ R_{21} & R_{21} \end{bmatrix} \begin{bmatrix} \hat{H}_{1opt} \\ \hat{H}_{2opt} \end{bmatrix} = \begin{bmatrix} P_1 \\ P_2 \end{bmatrix} \tag{8}$$

where, $R=E\{X_n X_n^T\}$ is the reference input correlation matrix, and $P_i=E\{y(n)X_{in}^T\}$, $i=1, 2$ is the cross-correlation vector between the reference input vector and echo signal. To obtain a set of Wiener solution from (8), it is necessary that the input correlation matrix R should be invertible. But with a help of (7), multiplying a nontrivial vector $U=[G_2^T -G_1^T]^T$ from the right side on (8) gives rise to

$$\begin{aligned} RU &= E\left\{\begin{bmatrix} X_{1n} \\ X_{2n} \end{bmatrix} [X_{1n}^T \; X_{2n}^T] \begin{bmatrix} G_2 \\ -G_1 \end{bmatrix}\right\} \\ &= E\left\{\begin{bmatrix} X_{1n} \\ X_{2n} \end{bmatrix} (X_{1n}^T G_2 - X_{2n}^T G_1)\right\} \\ &= 0. \end{aligned} \tag{9}$$

According to (9), it can be noticed that input correlation matrix turns out to be singular. Unfortunately this implies the existence of many solutions satisfying (8) such that adaptive filter possibly misconverges into the unwanted solution. In other words, optimum weight vectors are not uniquely determined, subsequently it results that weight vectors could not be equivalent to actual echo path impulse responses, i.e., $\hat{H}_{1opt} \neq H_1$ and $\hat{H}_{2opt} \neq H_2$. On the contrary, if the input signals are statistically uncorrelated, Wiener solutions for each adaptive filter have the form of $\hat{H}_{1opt} \equiv R_{11}^{-1}P_1$ and $\hat{H}_{2opt} \equiv R_{22}^{-1}P_2$. More interestingly, it can be shown that the resulting optimum weights are coincident with true echo path impulse responses.

3.2 Environment Change and its Effects

Another undesired performance degradation arises at the instant of environment change in the transmission room. To see the effects, using the relations $X_1(z)=S(z)G_1(z)$ and $X_2(z)=S(z)G_2(z)$ together with (1) and (4), the Z-transform of the estimation error signal can be expressed as

$$E(z)=\{\Delta H_1(z)G_1(z)+\Delta H_2(z)G_2(z)\}\, S(z) \tag{10}$$

where $S(z)$ is the Z-transform of remote talker's speech and

$$\Delta H_1(z)=H_1(z)-\hat{H}_1(z), \quad \Delta H_2(z)=H_2(z)-\hat{H}_2(z).$$

Since $G_1(z)$ and $G_2(z)$ are not trivial, in order to satisfy $E(z)\equiv 0$ it only happens when $\Delta H_1(z)\equiv 0$ and $\Delta H_2(z)\equiv 0$. But as discussed in last subsection, if the two speaker outputs are not perfectly uncorrelated, true echo paths could never be restored. Moreover assume that there exist some filters satisfying $E(z)\equiv 0$, the estimates $\hat{H}_1(z)$ and $\hat{H}_2(z)$ are required to be adjusted at the variation of acoustic pickup transfer functions subject to $E(z)\equiv 0$ as in (10). This implies that the performance of acoustic echo cancellation is inevitably corrupted at every instant of change of transmission room environment.

3.3 Slow Convergence and Large Computational Complexity

It is well known that the convergence speed of the adaptive filter directly depends on the eigenvalue spread ratio of reference input autocorrelation matrix. Since the reference signal to each adaptive cancellation filter is speech in general, the adaptive filters suffer from the slow convergence. Specially for stereophonic echo canceler the slow convergence problem becomes evenmore serious because the overall input correlation matrix as in (8) has been characterized as being ill-conditioned because deriving source is in common for both channels. Consequently, the convergence speed for stereophonic echo canceler becomes significantly slower than the monoaural.

In stereophonic conferencing system, four adaptive filters are needed to control the echo signal via tracking established acoustic echo paths at every instant of time. Under the room circumstance, echo path is characterized with long impulse response. Thus, FIR-type adaptive filter having long tap delays is required so as to obtain acceptible cancellation capability without giving any perceptual degradation. Thus, on the implementation, the high degree of computational complexity becomes one of the primary concerns.

IV. Proposed Stereophonic Echo Canceler

The structure of the proposed stereophonic echo canceler is illustrated in Fig. 2. Major difference from the conventional one originates on the usage of preprocessing block which is generating the reference inputs to each echo cancellation filter.

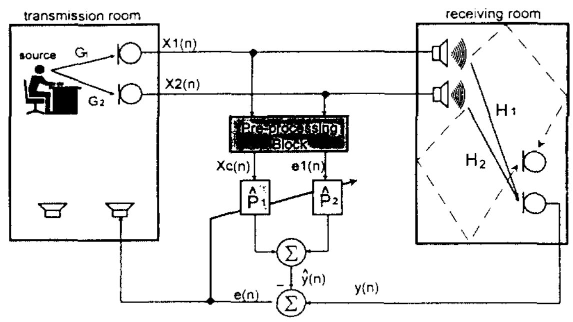


Figure 2. A robust stereophonic acoustic echo canceler.

Here $x_1(n)$ and $x_2(n)$ are the left and the right microphone output signals at transmission site, and $x_C(n)$ is the composite signal, i.e., $x_C(n) = x_1(n) + x_2(n)$. Furthermore let the transfer function between $x_c(n)$ and $x_1(n)$ be $F_1(z)$ and between $x_c(n)$ and $x_2(n)$ be $F_2(z)$. Since the source signal is used in common for producing both microphone output signals $\{x_1(n), x_2(n)\}$, $F_1(z)$ and $F_2(z)$ can be represented by exact formulation in terms of $G_1(z)$ and $G_2(z)$ as follows

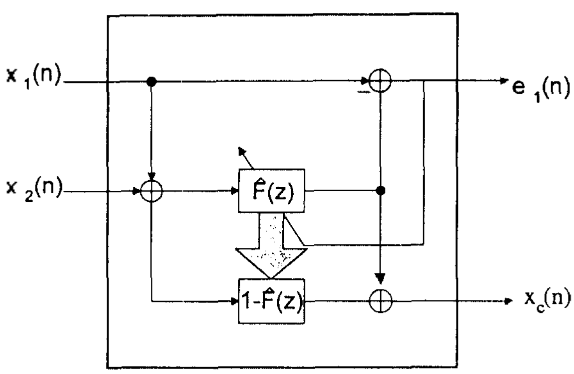
$$F_1(z) = \frac{G_1(z)}{G_1(z) + G_2(z)} \text{ and } F_2(z) = \frac{G_2(z)}{G_1(z) + G_2(z)}. \quad (11)$$

From (11) it can be easily noticed that

$$F(z) \triangleq F_1(z) \equiv 1 - F_2(z) \quad (12)$$

Thus, as shown in Fig. 3, by making use of (12), both $x_1(n)$ and $x_2(n)$ can be estimated from the common reference signal $x_c(n)$ using only single adpative filter $\hat{F}(z) \equiv \hat{F}_1(z) = 1 - \hat{F}_2(z)$. Resulting estimation errors, i.e., $e_1(n) \triangleq x_1(n) - \hat{x}_1(n)$ and $e_2(n) \triangleq x_2(n) - \hat{x}_2(n)$ can be related as follows:

$$e_1(n) + e_2(n) = 0. \quad (13)$$



Pre-processing Block

Figure 3. Proposed pre-processing block.

The reference signals to the dual adaptive filters are the composite signal $X_C(t)$ as well as the error signal $e_1(n)$ which are generated from the pre-processing block. Referring the orthogonality principle, it is noteworthy that $e(n)$ and $x_C(n)$ are statistically *uncorrelated* provided that the adaptive filter employed on the pre-processing block satisfies the Wiener solution. Furthermore the composite echo signal $y(n)$ can be decomposed into two distinct filtered outputs derived by the uncorrelated signals $x_C(n)$ and $e_1(n)$. To see this, with the help of (13), the Z-transform pair of the received signal $y(n)$ is represented in terms of $X_C(z)$ and $E_1(z)$ as below

$$\begin{aligned} Y(z) &= H_1(z)X_1(z) + H_2(z)X_2(z) \\ &= H_1(z)(\hat{X}_1(z) + E_1(z)) + H_2(z)(\hat{X}_2(z) + E_2(z)) \\ &= P_1(z)X_C(z) + P_2(z)E_1(z), \end{aligned} \quad (14)$$

where

$$P_1(z) \triangleq (H_1(z) - H_2(z))\hat{F}(z) + H_2(z) \quad (15)$$

and

$$P_2(z) \triangleq H_1(z) - H_2(z). \quad (16)$$

Equation (14) indicates the composite echo signal $y(n)$ is comprised of filtered outputs derived by the combined signal $x_C(n)$ and the estimation error signal $e_1(n)$ through $P_1(z)$ and $P_2(z)$ respectively rather than $H_1(z)$ and $H_2(z)$. Interestingly, as long as the filter coefficients of $\hat{F}(z)$ near to the optimal in minimum mean square sense, $x_C(n)$ and $e_1(n)$ becomes likely *uncorrelated* with each other. Therefore generating $x_C(n)$ and $e_1(n)$ can be viewed as processing outputs through the decorrelation filter in between highly correlated signals $x_1(n)$ and $x_2(n)$. Thus, deriving $x_C(n)$ and $e_1(n)$ as reference signals to adaptive filters for estimating the desired signal $y(n)$ is more meaningful rather than using $x_1(n)$ and $x_2(n)$ straightforward.

Furthermore, by making use of (15), (16) it is obvious that echo path systems $H_1(z)$ and $H_2(z)$ can be expressed as follows:

$$H_1(z) = P_1(z) + (1 - \hat{F}(z))P_2(z) \quad (17)$$

$$H_2(z) = P_1(z) - \hat{F}(z)P_2(z). \quad (18)$$

Using (17) and (18), the resulting dual adaptive filters $\hat{P}_1(z)$ and $\hat{P}_2(z)$ can be utilized for estimating true echo path systems as follows

$$\hat{H}_1(z) = \hat{P}_1(z) + (1 - \hat{F}(z))\hat{P}_2(z) \quad (19)$$

$$\hat{H}_2(z) = \hat{P}_1(z) - \hat{F}(z)\hat{P}_2(z). \quad (20)$$

Thereby, using (19) and (20), the estimates of true echo path systems $H_1(z)$ and $H_2(z)$ can be obtained indirectly based on estimating $P_1(z)$ and $P_2(z)$ together with $F(z)$. Therefore, if the adaptive filter used in the pre-processing block converges into the Wiener solution, reference inputs to the echo cancellation filter turn out to be nearly uncorrelated with each other, the optimum solutions for $\hat{P}_1(z)$ and $\hat{P}_2(z)$ can be obtained in unique fashion. Accordingly the estimated echo path systems $\hat{H}_1(z)$ and $\hat{H}_2(z)$ could be coincident to the true echo path systems.

Another advantage of employing pre-processing block for stereophonic echo canceler in Fig. 2 is that inherently it could gives rise to some *robustness* to the abrupt environment change in transmission room. As in (11) and (12), the variation of environment is taken into account on the transfer function $F(z)$. Thus, at the change of environment such as talker's movements, the optimum solution of the adaptive filter $\hat{F}(z)$ in the pre-processing block might be changed. Subsequently, the correlation between two reference signals to the echo canceler filters increases so that the performance of the echo cancellation becomes degraded at that instant. As soon as the adaptive filter $\hat{F}(z)$ converges into the newly altered Wiener solution induced by the change of environment, once again dual echo canceler filters begin to track true echo path systems in accurate. Here the change of the enviroment at the transmission room implies the variation of the transfer function $F(z)$, sequentially this provoke modification of the transfer function of $P_1(z)$ regarding to (15). Therefore the adaptive filter $\hat{P}_1(z)$ as the estimate of $P_1(z)$ could reacts exclusively to the change of environment more effectively.

V. Stereophonic Subband Adaptive Echo Canceler

Inherently the acoustic echo paths are oftenly characterized into the time-varying systems whose impulse responses have long residual delays. Thus in order to estimate these long impulse responses with using adaptive filter, applicable echo canceler must have features such as the fast convergence speed [8] and the long tap FIR-type filter so as to track the echo path system in a short amount of time. Here it is well-known that the convergence speed is directly affected by the spectral characteristics of the reference signal to the adaptive filter. But in the echo cancellation system another considerable fact is that the echo signal required to be eliminated has the dynamic spectral characterisitics. Thus, from the above observations, the full-band adaptive echo canceler critically suffers from the large computational burden associated with calculating the long tap filter weights and the slow convergence due to the large dynamic range of the reference input spectrum.

In order to mitigate these problems, recently the subband processing has been regarded as an appropriate technique for an efficient echo cancellation mounted on the monoaural teleconferencing system. Compared with the conventional LMS-type fullband adaptive filtering, the subband scheme possesses at least two adavantages such as the reduced computational complexity due to the adoption of a highly parallel structure with multirate operation, and the improved convergence behavior which is attributed that each subband the adaptation step size can be matched to the energy of the input signal in that band. Figure 4 depicts the configuration of a conventional subband echo canceler deployable to the stereophonic teleconferencing system.

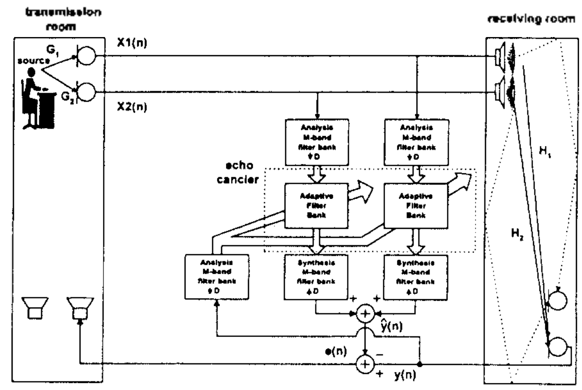


Figure 4. Configuration of a conventional subband echo canceler.

In this structure, the dual reference inputs to echo canceler together with the echo signal are split into M adjacent frequency subbands by identical analysis filter banks. Here, M contiguous single-sideband bandpass filters are utilized whose outputs are down-sampled by a factor of D to produce a series of complex subband signals in sequel. The coefficients of the each subband adaptive filter in different subbands are independently. The outputs obtained from the filter are then up-sampled by the factor D and the echo replica can be obtained by composing two synthesis filter bank outputs. However as mentioned in [10] the conventional subband adaptive filter imposes the undesired signal path delay due to the usage of the analysis and synthesis filter banks simultaneously. In echo cancellation system, this unwanted path delay is regarded as a critical obstacle producing sound impairments over voice communication between far-end listener and near-end talker. To avoid this undesired signal path delay whereas retaining the low compuational complexity and the fast convergence referring to the statements in previous section, this section introduces a new type of stereophonic echo canceler deploying the delayless subband architecture rather than the conventional which is overlaid on the full-band echo cancellation system shown in Fig. 2.

The overall structure of the proposed subband echo canceler is depicted in Fig. 5. Unlike the conventional shown in Fig. 4, this structure does not include the synthesis filter bank so as to prevent the echo cancellation unit from introducing the undesired signal path delay. To make a further progress regarding the content discussed in section Ⅳ, the combined signal $x_C(n)$ and one of speaker output signals $x_1(n)$ are fed into two sets of M-channel analysis filter bank to generate corresponding decimated complex subbands signals, i.e., $x_{C,i}(m)$ and $x_{1,i}(m)$, $i=0 \to M-1$. In order to generate a series of complex subband signal, generally, a bank of single-side bandpass filters is required to be implemented.

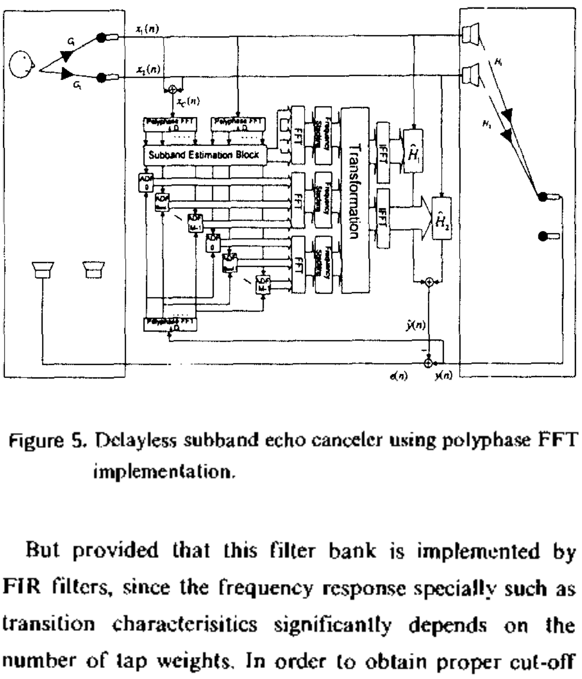


Figure 5. Delayless subband echo canceler using polyphase FFT implementation.

But provided that this filter bank is implemented by FIR filters, since the frequency response specially such as transition characterisitics significantly depends on the number of tap weights. In order to obtain proper cut-off frequency characterisitics, high-degree of bandpass FIR filters are necessary. But to process signals through theses filters, computational complexity becomes considerable so that it is hard to execute the analysis in proper amount of time. To reduce the processing time, the analysis part can be implemented via polyphase type of filter bank [10]-[12]. Toward this the protoytpe filter $H(z)$ subject to given specifications is constructed a priori with employing relatively large number of coefficients. Figure 6 shows the basic procedure for generation kth subband complex signal out of M subband ones by making use of the proto-type low pass filter together with modulating factor $W_M = e^{-j\frac{2\pi}{M}}$.

Here the signal is multiplied by W_M^{nk}, then corresponding kth band signal is down-coverted into the baseband. Then

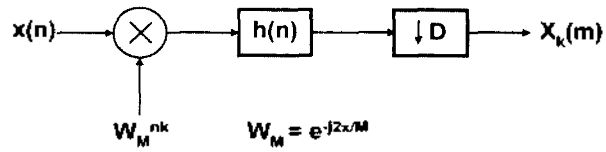


Figure 6. Basic procedure for generation kth subband complex signal.

the signal is processed through low-pass filter where its cutoff frequency is $\frac{f_s}{2M}$, then downsampling by the factor of D. Based on the single channel of a DFT filter bank analysis as in Fig. 6, the output can be expressed as

$$X_k(m) = \sum_{n=-\infty}^{\infty} h(mM-n)x(n)W_M^{nk} \tag{21}$$

It is noteworthy that the number of subband denoted by M is not always equivalent to the decimation factor D. Assume that $M=D$, this is the case that the filter bank is critically sampled. However, in many practical applications, in order to alleviate the aliasing effect which due to the usage of non-ideal lowpass filter, it is desired to have an oversampled filter bank where $D<M$ so that the filter bank signals $X_k(m)$ are sampled at a higher rate. Let $M=DI$ where I is referred to the oversampling ratio, i.e., if $I=1$, it is critically sampled and $I=2$ it is oversampled by a factor of two. In this paper, the oversampled case will be considered.

To develop this class of structure, the following formula are related to the basic mathematical form of the FFT filter bank analyzer.

Here let $n=rM+l$, then (21) becomes for $l=0, 1, \cdots, M-1$,

$$X_k(m) = \sum_{r=-\infty}^{\infty} \sum_{l=0}^{M-1} h(mD-rM-l)x(rM+l)W_M^{kl}. \tag{22}$$

And let $x_l(m) = x(mM-l)$, then

$$X_k(m) = \sum_{r=-\infty}^{\infty} \sum_{l=0}^{M-1} h((m-rI)M-l)x_l(m)W_M^{kl}. \tag{23}$$

Furthermore, let $P_l(m) = h(mM-l)$, then (23) becomes

$$X_k(m) = \sum_{l=0}^{M-1} \left\{ \sum_{r=-\infty}^{\infty} P_l(m-rI)x_l(r) \right\} W_M^{kl}. \tag{24}$$

From the above equation,

$$y_l(m) \triangleq \sum_{r=-\infty}^{\infty} P_l(m-rI)x_l(r). \tag{25}$$

Here in (25), it can be noticed that $y_l(m)$ is the result of l sample interpolator where the polyphase filter $p_l(m)$ acts as the interpolating filter. Then, (24) can be simply expressed by FFT operation as

$$X_k(m)=\sum_{l=0}^{M-1} y_l(m)W_M^{kl}=FFT^*\{y_l(m)\}, \tag{26}$$

where $FFT^*\{\cdot\}$ is the complex conjugate of the output of the FFT output. In sequel, referring to the structure shown in Fig. 5, a set of subband signals, i.e., $X_{1,i}(m)$ and $X_{C,i}(m)$, $i=0, \cdots, M-1$, are processed to generate estimation error signals $E_{1,i}(m)$, $i=0\rightarrow M-1$, corresponding to ith subband which is utilized as one of the reference signals to the individual echo canceler subsequently. Then each subband estimation error signal can be expressed as

$$E_{1,i}(m)=X_{1,i}(m)-\sum_{k=0}^{K-1} f_{i,k}(m)X_{C,i}(m-k),\ i=0\rightarrow M-1, \tag{27}$$

where the $f_{i,l}(m)$, $l=0\rightarrow M-1$ are the complex valued coefficients of i^{th} subband adaptive filter at the time instant m, and K is the length of adaptive filter. In (27) the coefficients of subband estimation filter can be recursively updated as follows

$$\mathbf{f}_i(m)=\mathbf{f}_i(m-1)+\alpha\frac{\mathbf{X}^*_{C,i}(m)e_{1,i}(m)}{\|\mathbf{X}_{C,i}\|} \tag{28}$$

where α is the convergence parameter, $\|\cdot\|$ is the Euclidean norm and $\mathbf{X}_{C,i}(m)$ is the reference input vector given by

$$\mathbf{X}_{C,i}(m)=[X_{C,i}(m), X_{C,i}(m-1), \cdots X_{C,i}(m-N_1+1)], \tag{29}$$

and $\mathbf{f}_i(m)$ is the length-N_1 ith subband coefficient vector composed of $\{f_{i,l}(m)\}_{l=0}^{N_1-1}$ which is updated at the decimated time instant m.

In sequel, in order to track the modified echo path systems $P_{1,i}(z)$ and $P_{2,i}(z)$ for each subband i, two sets of major subband adaptive filters are employed with using $X_{C,i}(m)$ and $E_{1,i}(m)$ as reference signals. As shown in Fig. 5, the decomposed subband desired signals $Y_i(m)$, $i=0\rightarrow M-1$, are generated by passing $y(n)$ through M-channel analysis filter bank. Here the subband adaptive filter coefficients for $\hat{P}_1(z)$ are iteratively updated as follows:

$$\mathbf{P}_{1,i}(m)=\mathbf{P}_{1,i}(m-1)+\mu\frac{\mathbf{X}^*_{C,i}(m)E_i(m)}{\|\mathbf{X}_{C,i}(m)\|+\|\mathbf{E}_{1,i}(m)\|}, \tag{30}$$

and similarly those for $\hat{P}_2(z)$ are renewed by the following recursion:

$$\mathbf{P}_{2,i}(m)=\mathbf{P}_{2,i}(m-1)+\mu\frac{\mathbf{E}^*_{1,i}(m)E_i(m)}{\|\mathbf{X}_{C,i}(m)\|+\|\mathbf{E}_{1,i}(m)\|}, \tag{31}$$

where * denotes complex conjugate operator, $E_i(m)$ is the residual cancellation error associated with i^{th} subband, and the subband estimation error signal vector $\mathbf{E}_{1,i}(m)$ is given by

$$\mathbf{E}_{1,i}(m)=[e_{1,i}(m), e_{1,i}(m-1), \cdots, e_{1,i}(m-N_2+1)]. \tag{32}$$

In (30) and (31), for a convenience, $\mathbf{P}_{1,i}(m)$ and $\mathbf{P}_{2,i}(m)$ are the coefficient vectors having the same filter length of M.

In order to construct length-L wideband echo cancellation filters $\hat{H}_1(z)$, $\hat{H}_2(z)$ using M subbands so as to process the real stereo signals, one-half of the complex subbands need to be processed. Thus $M/2+1$ subband filters are enough to update the wideband filter coefficients. And the length of each subband filter is set to be L/D which is utilized to span length-L wideband filter coefficient in sequel. As shown in Fig. 5, to update the dual wideband echo canceler $\hat{H}_1(z)$ and $\hat{H}_2(z)$, intermediate steps should be incorporated. Towards this, recursively updated coefficient vectors $\{\mathbf{f}_i(m), \mathbf{P}_{1,i}(m), \mathbf{P}_{2,i}(m)\}_{i=0}^{M-1}$ are transformed into frequency bins via L/D-point FFTs to obtain L/D frequency bins per each individual subband. Subsequently, frequency bins of the estimates of true echo path system associated with each subband are constructed by making use of the relationship as in (19) and (20) as follows, for $i=0\rightarrow M/2+1$ and $l=0\rightarrow L/D-1$,

$$\hat{H}_{1,i}(l)=\hat{P}_{1,i}(l)+\hat{P}_{2,i}(l)-\hat{F}_i(l)\hat{P}_{2,i}(l) \tag{33}$$

$$\hat{H}_{2,i}(l)=\hat{P}_{1,i}(l)-\hat{F}_i(l)\hat{P}_{2,i}(l), \tag{34}$$

where $\{\hat{P}_{s,i}(l), F_i(l)\}$, $s=1, 2$, $i=0\rightarrow M-1$, are the FFT pairs to $\{p_{s,i}(m), f_i(m)\}$. Thus, in Fig. 5, after the transformation from the modified into the estimate considering the oversampling process, only $L/2P$ significant frequency bins are stacked to form size-L full-band transformed frequency bins in proper order[10]. Finally, with a help of the frequency stacking, the length-L wideband eacho cancellation filters $\hat{H}_1(z)$ as well as $\hat{H}_2(z)$ are formed in transformed domain. To complete the computation of the finite impulse responses of those filters in temporal domain, L-point IFFTs are utilized for each channel. And these wideband echo cancellation filters are excited by each individual speaker input signals $x_1(n)$ and $x_2(n)$ to generated the echo replica.

Ⅵ. Simulation Results

In order to verify the performance improvement of the proposed stereophonic echo canceler, ERLE(Echo Return Loss Enhancement) curves are examined under different situations. It is well-described that the ERLE is wideley utilized for measuring the effectiveness of the echo canceler in quantitive. Figure 7 and 8 show the ERLE curves while letting the source signal the colored gaussian noise, i.e., AR(6) process, and the speech signal respectively. For a comparison between the performances acheived with and without using the preprocessing block, two ERLE curves are depicted at each figure. Regarding to the ERLE curves shown in Fig. 7 and 8, it can be noticed that the structure with using preprocessing block as in Fig. 7 shows the superior performance rather than one without using preprocessing block for either the colored gaussian noise or the speech case. By inspection, it can be observed that ERLE values are increased more or less by 5dB on both cases.

To see the robustness to the change of the transmission room environment, the acoustic pick-up transfer functions $G_1(z)$ and $G_2(z)$ are intentionally altered at sample instant n = 10000 while letting the true echo path systems $H_1(z)$ and $H_2(z)$ stationary. Figure 9 and 10 show the ERLE curves resulted from the stereophonic subband echo canceler with and without preprocessing block where the source signal has a form of either the AR(6) process or the speech respectively. Similiar to the previous results,

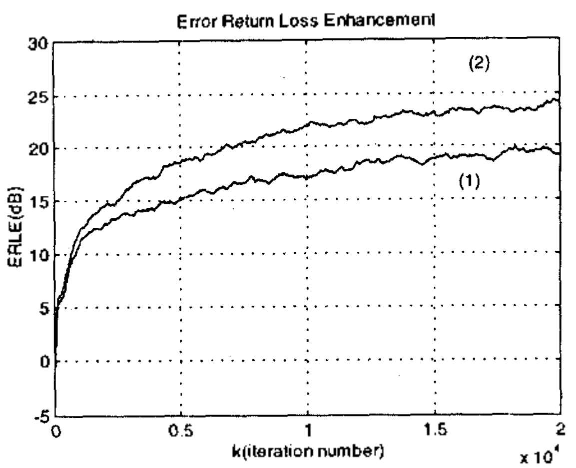


Figure 7. The comparison of ERLE(AR(6));
(1) the conventional method, (2) the proposed method using pre-processing block.

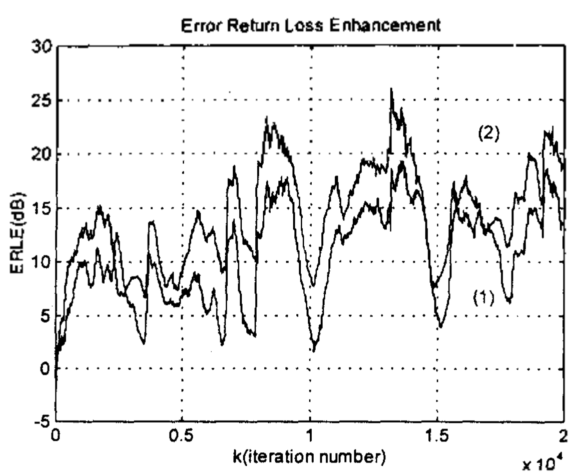


Figure 8. The comparison of ERLE(Speech);
(1) the conventional method, (2) the proposed method using pre-processing block.

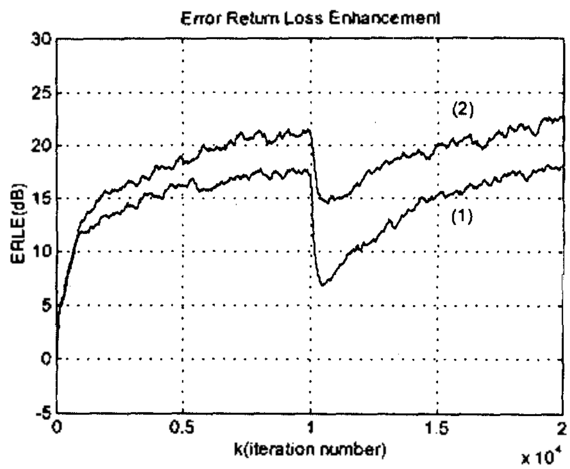


Figure 9. The comparison of ERLE(AR(6)) for environment change of transmission room;
(1) the conventional method, (2) the proposed method using pre-processing block.

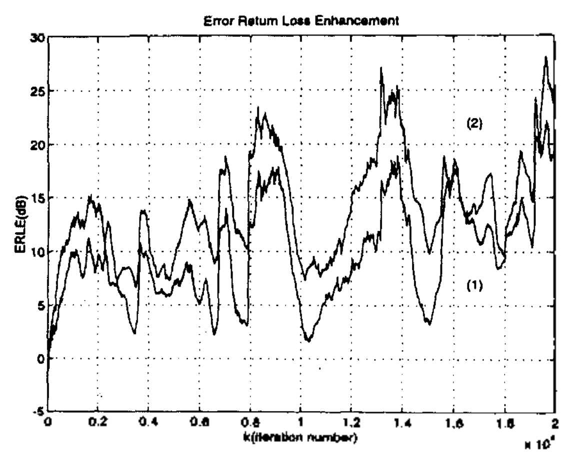


Figure 10. The comparison of ERLE(Speech) for environment change of transmission room;
(1) the conventional method, (2) the proposed method using pre-processing block.

the ERLE curves associated with the structure employing the preprocessing block overwhelm those obtained from the conventional. Therefore it can state that the proposed echo canceler behaves well and shows some robustness even at the instant of transmission room environment change.

To investigate performance from the perfect alignment point of view, the following formula is utilized to measure how the recovered the echo path systems are close to the true echo path systems which is given by

$$Misalignment = 10log_{10}\frac{\langle(\hat{H}_i - H_i)^T(\hat{H}_i - H_i)\rangle}{H_i^T H_i}\ [dB],\ i = 1, 2$$

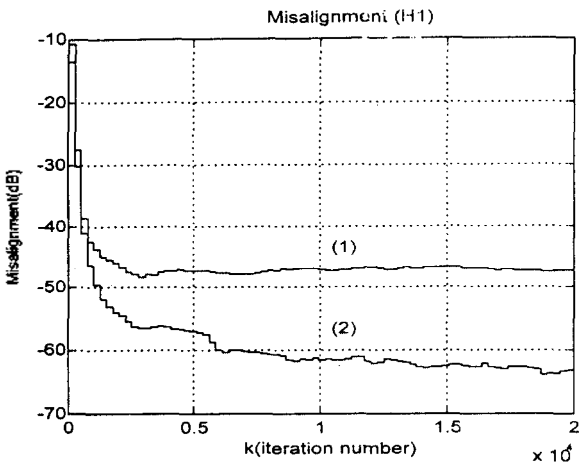


Figure 11. The comparison of Misalignment(H_1);
(1) the conventional method, (2) the proposed method using pre-processing block.

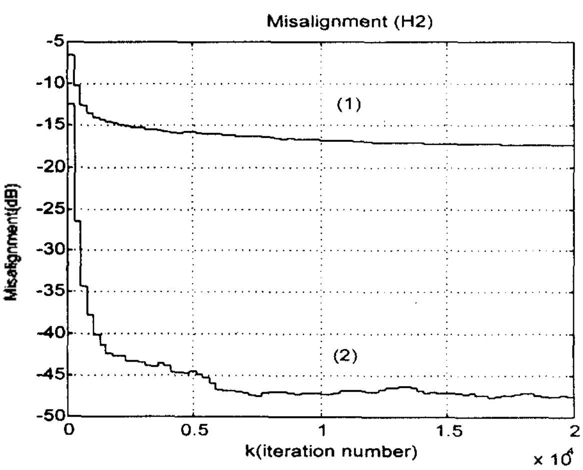


Figure 12. The comparison of Misalignment(H_2);
(1) the conventional method,(2) the proposed method using pre-processing block.

where <> is the time averaged value per each 256 time samples. Figure 11 and 12 corresponds to the misalignment curves which indicate the similiarity of the estimates to the true echo path systems. As a result, it is noteworthy that the estimates obtained from the proposed structure superbly matched to the true echo path systems.

VI. Concluding Remarks

This paper proposed the delayless subband version of the modified full band stereophonic echo canceler deploying the pre-processing block which plays a major role for producing exellent performances. Instead of performing the echo cancellation process through full-band dual adaptive filters, the delayless subband schemes are employed so as to suppress unwanted signal path delay, accerlate the convergence speed and reduce the computational complexity. According to the simulation results, its superior behaviors are verified at the point of the robustness to the change of the transmission room environment and the alignment of the impulse responses of estimates to those of the true echo path systems.

Acknowledgement

The authors wish to acknowledge the financial support of the Korea Research Foundation made in the program year of 1996.

References

1. M. M. Sondhi, D. R. Morgan and J. L. Hall, "Stereophnic Acoustic Echo Cancellation-An Overview of the Fundamental Problem," *IEEE Signal Processing Letters*, vol. 2, no. 8, pp. 148-151, August 1995.
2. M. Miyoshi and N. Koizumi, "NTT's Research on Acoustics for future Telecommunication Services," Applied Acoustics, vol. 36, pp. 307-326, 1992.
3. S. H. Jensen, "Acoustic Echo canceler for Hands-Free Mobile Radiotelephony," Proc. EUSIPCO, Belgium, pp. 1629-1632, 1992.
4. K. Murano, S. Unagami and F. Amano, "Echo Cancellation and Application," IEEE Comm. Mag., pp. 49-55, Jan. 1990.
5. F. Amand, J. Benesty, A. Gilloire and Y. Grenier, "A Fast Two-Channel Projection Algorithm," *Proc. ICASSP'96*, pp. 949-952.
6. S. Minami, "A Stereophonic Echo canceler Using Single Adaptive Filter," *Proc. ICASSP'95*, pp. 3027-3030.
7. A. Hirano and A. Sugiyama, "A Compact Multi-channel Echo canceler with a Single Adaptive Filter per Channel," *Proc. ICASSP'92*, pp. 1922-1925.

8. J. J. Shynk, "Frequency-domain and multirate adaptive filtering," *IEEE Signal Processing Mag.*, vol. 9, pp. 14-37, Jan. 1992.
9. A. Gilloire and M. Vetterli, "Adaptive Filtering in Subbands," *Proc. ICASSP*, vol. 2, pp. 1572-1575, 1988.
10. D. R. Morgan and J. C. Thi, "A Delayless Subband Adaptive Filter Architecture," *IEEE Trans. on Signal Processing*, vol. 43, no. 8, pp. 1819-1830, August 1995.
11. N. J. Fliege, *Multirate digital Signal Processing*, JOHN WILEY and SONS, 1994.
12. Ronald E. Crochiere, Lawrence R. Rabier, *Multirate digital Signal Processing*, Prentice-hall, 1983.

▲Won-Chul Lee

Won-Chul Lee received the B.S. degree in electronic engineering from Seogang University in 1986 and M.S. degree from Yonsei University and Ph.D. degree from Polytechnic University, New York, in 1988, 1994 respectively. He joined Polytechnic University as a Post-Doctoral Fellow from July. 1994 to July. 1995. From Jan. 1994 to Dec. 1994, he was a member of the examining committee of IEEE Trans. on Signal Processing. He has been an assistant professor of the department of Information and Telecommunication, Soongsil University since Sep. 1995. He has been working in the Research Institute of Yonsei University on signal processing as a researcher since Sep. 1995.

He has been an editor of the Acoustical Society of Korea and executive board since Sep. 1995. and Jan. 1998 respectively. He is currently an editor of the KICS(Korea Institute of Communication Sciences). His major research area includes digital system identification, speech signal coding and mobile communication system, radar signal processing.

▲Young-Min Cho

Young-Min Cho was born in Seoul, Korea on May 22, 1974. He received the B.S. degree in the Information and Telecommunication Engineering from Soongsil University, Seoul, in 1997. He is working on his M.S. degree in same university. His current interests are in adaptive signal processing and mobile communication system.