WEAVER SSB SUBBAND ACOUSTIC ECHO CANCELLER

Peter L. Chu

PictureTel Corporation
One Corporation Way
Peabody, MA 01960
chu@pictel.com

ABSTRACT

A Weaver SSB subband structure is used to implement an acoustic echo canceller. The structure has 29 bands of 250 Hz width, covering the audio range from 0 to 7 kHz. The Weaver structure lowers each bandpass region to baseband, allows for oversampling to eliminate aliasing components, and is computationally efficient. The subsampled components are purely real, as compared to the complex components found in some other subband schemes. The adaptive filter update algorithm is a variant of the block NLMS. The double-talk, divergence, echo suppression, and noise fill-in algorithms all fully exploit the bandpass structure to achieve performance difficult to attain in full-band or two-hand acoustic echo cancellers. The acoustic echo canceller has been extensively field tested and has been shown to be robust.

1. INTRODUCTION

For the past two years, a robust subband acoustic echo canceller has been shipped with the PictureTel system 4000 VCS (videoconferencing system). There are currently 3000 installed VCS's, all reporting very satisfactory audio performance as far as the echo cancellation is concerned. The installed base forms an existence proof of the practicality and usefulness of subband acoustic echo cancellation. The echo cancellation algorithm automatically adapts to a wide range of acoustic environments. No training sequences are used because the speed of convergence is fast enough so that the length of time echo is heard is not objectionable. The echo canceller resides on an ATT DSP16A along with a telephone line echo canceller, far-end to near-end sampling rate converter, and a host-controlled four input-output audio mixer with arbitrary mixing gains, making for a very cost-effective implementation.

The advantages of subband echo cancellation vs. full band echo cancellation have been well established in the literature. The advantages are three-fold,

- For the same acoustic filter length in time, the number of filter computations is substantially less for the subband scheme as compared to the full band scheme [1],[2].
- The bandpassed speech components are whiter and less correlated than the original speech, and lower energy bands of the speech can have a higher gradient step size, since the adaptive filter algorithm operates independently in each band, allowing for faster overall convergence of the adaptive filter gradient descent algorithm [1],[2].
- Double-talk detection, echo suppression, and noise fill-in are all much enhanced in performance by exploiting the subband structure.

2. DESCRIPTION OF THE MODULES

In this section we shall describe the main modules of the acoustic canceller– the bandpass structure, the adaptive filters, the double-talk detector, the echo suppression, and the noise fill-in. The complete block diagram in shown in figure 1, where the place of each module in the entire structure is shown.

2.1. Bandpass Filter Structure

The acoustic echo canceller is implemented with a 29 channel bandpass filter bank structure, each band having a -3 dB width of 250 Hz, except for the lowest frequency band, which is -3 dB at 125 Hz. The bandpass channels should be subsampled as much as possible with the restriction that aliasing components are not created by the subsampling. The original sampling rate is 16 kHz, and the subsampled rate of the bandpass channels is 1 kHz. Other solutions to this problem have used subband structures with complex outputs [4],[3], or additional crossband adaptive filters to remove alias effects [2]. This structure uses the Weaver SSB modulation scheme proposed in [5], to shift a band of frequencies to baseband. The subsampled baseband signals have purely real components. Although the total number of multiplies is equal between the real vs. complex scheme, the real scheme is easier to implement since all signal samples may be handled identically, i.e., the code does not have to handle real and imaginary terms. Details of the Weaver SSB scheme are shown in the block diagram for the analysis bandpass filter structure in figure 2, while the block diagram for the synthesis bandpass filter structure is shown in figure 3. The bandpass filters are implemented efficiently using a polyphase structure. The lowpass filters shown in the block diagrams are designed to have as much attenuation as possible from 500 Hz on up, to prevent aliasing.

2.2. Adaptive Filter

The adaptive filter derives its output from

$$e_b(i) = m_b(i) - \sum_{j=0}^{L-1} \alpha_b(j) s_b(i-j), b = 1, \ldots, M \qquad (1)$$

where $m_b(i)$ is the subsampled microphone signal in bandpass channel b, $s_b(i)$ is the subsampled loudspeaker signal in the same band, M is the total number of bandpass channels, and L is the total number of adaptive filter taps for the band. The filter adapts the weights $\alpha_b(j)$ to minimize the average energy $E(e_b^2(i))$. The number of taps required for the adaptive filter in band b depends on the reverberance of the room in that spectral band, which in turn depends on the size of the room and losses due to absorption.

The method used to update the adaptive filter taps is a variant of the "Block LMS" [6]. This algorithm is used because its convergence

speed is equal to that of the usual NLMS, but only a small fraction of the filter coefficents need to be updated every sample tick. Because the DSP16A must write new taps to external memory through its parallel port, a slow process, the Block LMS is much faster computationally on the DSP16A than the usual LMS. The update equation is

$$\alpha_b^{i+K+1}(j) = \alpha_b^i(j) + \frac{\beta}{E_b(i)} \sum_{k=0}^{K-1} e_b(i+K-k)s_b(i+K-k-j) \tag{2}$$

where $\alpha_b(j)$ is the jth adaptive filter tap weight for band b, i indexes time, K is the thinning ratio (K=8 for this implementation, so that 1/8 of the filter taps are updated every sample tick), $e_b(i)$ is the adaptive filter output defined in (1), $s_b(i)$ is the loudspeaker signal, β is a constant controlling the convergence time, and $E_b(i)$ is an estimate of the loudspeaker energy in channel b at time i. In general, larger values of β yield faster adaptation speeds at the expense of worse cancellation of the echo once the adaptive filter has converged.

Because speech varies so much in amplitude over short intervals of time, instead of finding $E_b(i)$ as an average, $\sum_i s_b^2(i)$, an estimate yielding better results is $E_b(i) = \max\{s_b^2(i-j), j = 0, ..., L-1\}$, in (2). This maximum bandpass signal search is restricted to the last L samples because these are the only loudspeaker samples which affect the tap update equation, (2). Typically, a single large peak in the loudspeaker signal causes all taps to diverge. By using the maximum value, the effect of any large peak is eliminated. With this modification of the Block NLMS, the adaptation convergence time can be sped up by a factor of two without risk of divergence.

It would be a large computational burden to calculate the actual maximum of the bandpass signal over the last L samples, since L will be fairly large. An alternative is to use a "running maximum" calculated as follows:

0. Initialize values--only at the very start of the program. b refers to the bandpass channel number.

$$\begin{aligned} max_b &\leftarrow 0 \\ age_b &\leftarrow 0 \\ temp_b &\leftarrow 0 \end{aligned}$$

1.

$$\begin{aligned} &\text{If } |s_b(i)| > max_b \text{ then} \\ &\{ \\ &max_b \leftarrow |s_b(i)| \\ &age_b \leftarrow 0 \\ &temp_b \leftarrow 0 \\ &\} \\ &\text{Else } age_b \leftarrow age_b + 1 \end{aligned}$$

where $|s_b(i)|$ is the magnitude of the current sample, max_b is the running maximum, age_b is the length of time that the running maximum has been in effect, and $temp_b$ is an intermediate variable.

2.

$$\begin{aligned} &\text{If } (age_b > L_1) \text{ and } (|s_b(i)| > temp_b) \text{ then} \\ &temp_b \leftarrow |s_b(i)| \end{aligned}$$

3.

$$\begin{aligned} &\text{If } age_b > L_2 \text{ then} \\ &\{ \\ &max_b \leftarrow temp_b \\ &temp_b \leftarrow 0 \\ &age_b \leftarrow L_1 \\ &\} \end{aligned}$$

In this implememtation, $L_1 = L/2$ and $L_2 = 3 \cdot L/2$, where L is equal to the number of taps in the adaptive filter. With these settings of L_1 and L_2 the actual maximum of the past L samples is guaranteed to be less than or equal to the running maximum, max_b.

2.3. Doubletalk Detector

Ideally, the adaptive filter would only update its taps when the sound at the microphone is primarily due to the loudspeaker alone. If a major portion of the sound at the microphone is from people talking in the room, then if left to itself, the adaptive filter taps will diverge. The doubletalk detector attempts to detect this situation and disable the adaptive filter tap update when the situation occurs. A good double talk detector must detect all instances of doubletalk, since lack of such detection introduces divergence in the adaptive filter. Occasional false alarms are permissible, i.e., claiming that there is doubletalk when there is sound from the loudspeaker only.

A commonly used algorithm [7] is to declare doubletalk whenever

$$|m(i)| > \frac{1}{2} \max\{|s(i)|, |s(i-1)|, ..., |s(i-L-1)|\} \tag{3}$$

where L is the number of taps in the adaptive filter, $m(i)$ is the microphone signal, and $s(i)$ is the loudspeaker signal. This algorithm requires that the SPL (sound pressure level) from people in the room be 6 dB higher than the SPL from the loudspeaker at the microphone. Unfortunately, this requirement is not satisfactory, since a decent loudspeaker volume and long-range pickup microphone requires that the doubletalk detector work with at least a 0 dB difference between loudspeaker and people SPL at the microphone.

The doubletalk detector of this article works by comparing the magnitude spectral shape of the loudspeaker signal to that of the microphone signal, thereby gaining a significant improvement in detection capability over the detector of (3). For each subband, b, the following test is performed,

$$|m_b(i)| \overset{?}{>} GD(s_b(i)) \tag{4}$$

where G is the speaker-to-mic gain and $D(s_b(i))$ is a decaying maximum of $s_b(i)$, calculated as follows:

$$\begin{aligned} &\text{If } |s_b(i)| > D \text{ then} \\ &D \leftarrow |s_b(i)| \\ &Else \\ &D \leftarrow \gamma D \end{aligned} \tag{5}$$

where D is initialized to zero and is the returned value of the function $D(\cdot)$, and γ is chosen so that D decays with a time constant equal to that of the decay of the acoustic energy of the room. Typically, γ = .995. The test of (4) is performed indepedently on each bandpass

channel b. If the test is positive for any bandpass channel, then doubletalk is declared for all channels. With this bandpass doubletalk detector, reliable doubletalk detection occurs with a 0 dB difference between the loudspeaker and people SPL's at the microphone.

2.4. Bandpass Echo Suppression

In a typical conference room of 25 by 20 by 9 foot dimensions, the bandpass adaptive filters with 128 ms worth of adaptive filter taps produce 20 dB of loudspeaker attenuation in the microphone signal. This amount of attenuation is insufficient for videoteleconferencing because of round trip time delays. Essentially, the loudspeaker component in the microphone signal must be reduced to inaudible levels. Nonlinear echo suppression takes place, independently on each band, to reduce the loudspeaker component. On each band b,the following actions take place to find a band gain term, μ_b:

$$\begin{aligned} \text{If } |e_b(i)| > T, \mu_b &\leftarrow \mu_b + \delta \\ \text{Else } \mu_b &\leftarrow \mu_b - \delta \\ \text{If } \mu_b > 1, \mu_b &\leftarrow 1 \\ \text{If } \mu_b < 0, \mu_b &\leftarrow 0 \end{aligned} \tag{6}$$

where $e_b(i)$ is the echo canceller output for channel b, μ_b is defined as the gain for band b, and

$$T = H\,D(s_b(i)) \tag{7}$$

where H is the speaker-to-adaptive-filtered-mic gain, $D(\cdot)$ is as defined in (5), and δ typically is .05. The band gain μ_b is used to modify the channel amplitude,

$$c_b(i) = \mu_b e_b(i) \tag{8}$$

where $c_b(i)$ is the output of the bandpass echo suppression routine. The band gain term, μ_b, gets small when it is likely that the microphone signal is solely due to to the loudspeaker signal. If human speech is present which would mask the adaptive filter attenuated loudspeaker signal in channel b, then μ_b increases in value. This form of gradual gain change in subbands produces no hard clipping artifacts as does conventional centerclipping, yet sucessfully removes the presence of loudspeaker signal. During doubletalk, spectral bands slowly fade in and out of the microphone speech, which is not subjectively objectionable for moderate amounts of doubletalk.

2.5. Bandpass Noise Fill-in

When the near end is speaking and the far end is silent, the near end will hear background noise disappear and reappear in synchronization with his speech as a result of the far end's nonlinear echo suppression. This change in background noise is particularly disturbing, because subjectively we can ignore a constant background noise but not a constantly changing background noise. To eliminate this artifact, noise is added onto each band in a fashion complementary ot the amount of gain reduction the band experiences in (8),

$$d_b(i) = c_b(i) + (1 - \mu_b)n_b(i) \tag{9}$$

where $d_b(i)$ is the output of the noise fill-in routine, $c_b(i)$ is the nonlinear echo suppressed bandpass signal defined in (8), μ_b is the nonlinear echo suppression gain as defined in (6), and $n_b(i)$ is a synthetically generated white noise whose energy is chosen to match the background noise energy in channel b.

3. SUMMARY

A wideband subband acoustic echo canceller has been implemented on a single ATT DSP16A and tested in the field for two years in a videoconferencing system. The delay introduced by the subband filters is not a negative factor since this delay is small compared to the delays of the video and audio compression-decompression. The subband acoustic echo canceller has proven itself to be robust and quickly adaptable, so that no training sequences are ever used. It is the author's belief that similar performance from a single band acoustic echo canceller would be extremely difficult to achieve.

References

1. Gilloire, A., "Experiments with Sub-band Acoustic Echo Cancellers for Teleconferencing", *Proceedings ICASSP 1987*, Dallas, Texas, pp. 2141-2144.
2. Gilloire, A., and Vetterli, M., *Proceedings ICASSP 1988*, New York City, New York, pp. 1572-1575.
3. Amano, F., and Perez, H., "A New Subband Echo Canceller Structure", *Proceedings ICASSP 1991*, Toronto, Canada, pp. 3585-3588.
4. Gay, S.L., "Fast Converging Subband Acoustic Echo Cancellation Using RAP on the WE DSP16A", *Proceedings ICASSP 1990*, Albuquerque, New Mexico, pp. 1141-1144.
5. Crochiere, R.E., and Rabiner, L.R., *Multirate Digital Signal Processing*, Prentice Hall, Englewood Cliffs, New Jersey, 1983.
6. Gingell, M.J., et. al., "A Block Mode Update Echo Canceller Using Custom LSI", /it GLOBECOM Conference Record, vol. 3, Nov., 1983, pp. 1394-97.
7. Duttweiler, D.L., "A Twelve-Channel Digital Voice Echo Canceller", *IEEE Transactions on Communications*, COM-26, no. 5, May, 1978, pp. 647-653.

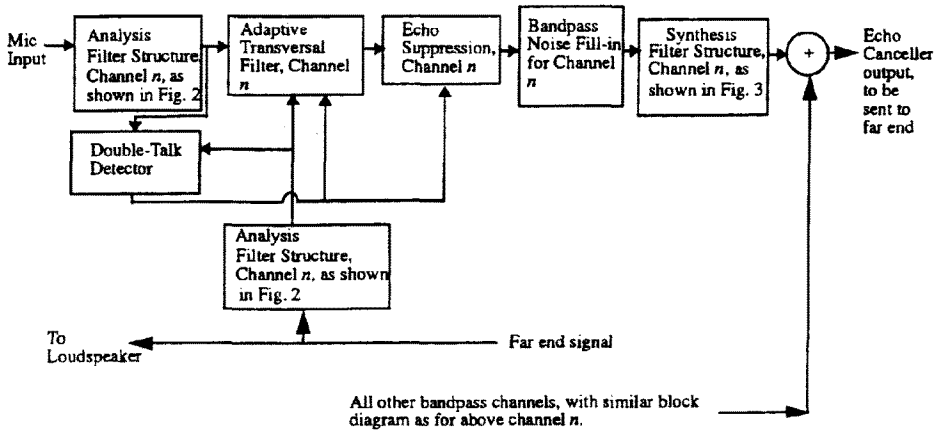

Figure 1. Block Diagram of Subband Acoustic Echo Canceller, using the Analysis and Synthesis Bandpass Filter Structures shown in Fig.2 and Fig. 3

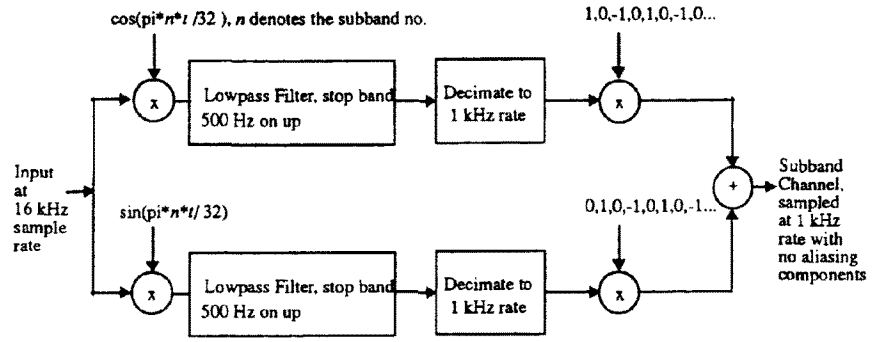

Figure 2. Analysis Bandpass Filter Structure for subband channel n, n =0,1,2,..28, t denotes sample time.

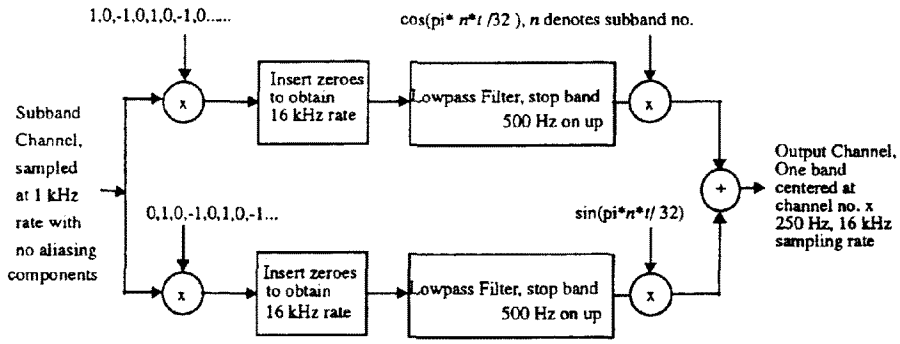

Figure 3. Synthesis Bandpass Filter Structure for channel n, n denotes the subband channel number. n = 0,1,2,...28, and t denotes the sample time.