SSB SUBBAND ECHO CANCELLER USING LOW-ORDER PROJECTION ALGORITHM

Shoji Makino *Josef Noebauer* *Yoichi Haneda* *Akira Nakagawa*

NTT Human Interface Laboratories
3-9-11, Midori-cho, Musashino-shi, Tokyo, 180 Japan
Tel: +81 422 59 3992, Fax: +81 422 60 7811
Email: makino@splab.hil.ntt.jp

ABSTRACT

This paper proposes a new subband echo canceller that almost fully whitens the received input by using the low-order projection algorithm, or the affine projection algorithm. Since the projection algorithm can fully whiten the received input of a small-tap adaptive filter with a relatively small projection order, the proposed subband projection echo canceller achieves nearly maximum convergence with a small projection order. By reflecting the frequency characteristics of the speech and of the echo path, it allows a different projection order to be chosen in different subbands. This gives the proposed method optimum cost performance, which is promising for implementation.

1. INTRODUCTION

Acoustic echo cancellers are widely used for teleconferencing and hands-free telecommunication systems to overcome acoustic feedback, making conversation more comfortable.

Subband echo cancellers, which divide signals into smaller frequency subbands and independently cancel echoes in each subband, have been studied [1]. Since the narrower frequency subbands have a smaller eigenvalue spread compared to the fullband for speech input, the convergence speed can be doubled. Since downsampling expands the sampling interval and reduces the number of taps needed for the adaptive filter, the subband echo canceller is computationally efficient. The major drawback of the subband structure is the delay that is introduced by the filter banks.

In the conventional subband echo canceller, the frequency spectrum of speech remains unchanged in each subband. Moreover, when the downsampling rate becomes smaller than the critical downsampling rate, the frequency spectrum of the band-pass filter has more influence on each subband. In other words, the received input in each subband is bandlimited and not fully whitened, which results in slow convergence.

The projection algorithm, or affine projection algorithm [2] has also been proposed to improve convergence of echo cancellers for speech input. The projection algorithm whitens the received input, *i.e.*, it removes the correlation between consecutive received input vectors. This process is especially effective for speech, which has a highly non-white spectrum.

Josef Noebauer was an exchange student from Fachhochschule Regensburg, Germany. All simulations were done by him.

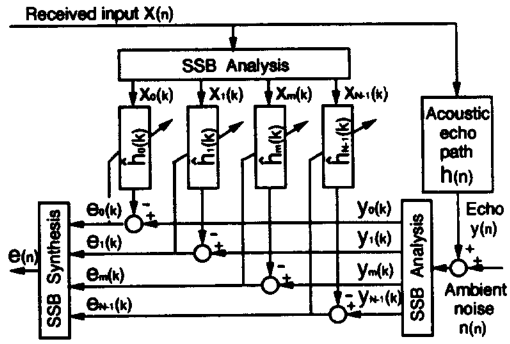


Figure 1: Configuration of a subband echo canceller.

The whitening effect of the projection algorithm is related to the number of taps of the adaptive filter. For a fullband echo canceller, where the number of taps is 1000, the convergence speed can be doubled with a projection order of two for speech input. However, little further improvement is achieved with a relatively small projection order. Even a projection order of 50 is not sufficient for nearly maximum convergence.

This paper proposes a new subband echo canceller that almost completely whitens the received input by using the low-order projection algorithm. By reflecting the short adaptive filter length of the subband echo canceller, the *16th*-order projection algorithm was shown to achieve nearly ultimate convergence. A different projection order can be chosen in different subbands by reflecting the frequency characteristics of speech and of the echo path. This enables optimum cost performance, which is promising for practical implementation.

2. SUBBAND ECHO CANCELLER

2.1. SSB subband echo canceller

The configuration of the subband echo canceller is shown in Fig. 1. For subband analysis and synthesis, a polyphase filterbank is used.

The implementation of the projection algorithm with a real valued signal is easier than with a complex valued one. Therefore a single side band (SSB) modulation [3] is performed on the complex subbands to obtain real valued subbands.

An efficient realization of the SSB modulation process is used with a multiplication by $(-1)^k$. For this purpose, the SSB-modulated subband signals are oversampled by a factor two. As a result, the downsampling rate R is given by

$$R = \frac{N}{4} \tag{1}$$

where N is the number of subbands (0-2π).

2.2. Imperfect whitening of subband echo canceller

The subband echo canceller reduces the eigenvalue spread by dividing signals into smaller frequency subbands compared to the fullband, *i.e.*, the subband echo canceller 'whitens' the received input. Consequently, convergence speed can be doubled in comparison to the fullband for speech input.

However, in the conventional subband echo canceller, the frequency spectrum of speech remains unchanged in each subband. Moreover, when the downsampling rate becomes smaller than the critical downsampling rate, the frequency spectrum of the band-pass filter has more influence on each subband. In other words, the received input in each subband is bandlimited and not fully whitened, which results in slow convergence.

3. PROJECTION ALGORITHM

3.1. Projection algorithm and ES projection algorithm

The p-th order projection algorithm, or affine projection algorithm, updates filter coefficient $\hat{\mathbf{h}}(k+1)$, which satisfies (2) where $p < L$.

$$\hat{\mathbf{h}}(k+1)^T \mathbf{x}(k-i+1) = y(k-i+1) \quad (i = 1, 2, \cdots, p) \tag{2}$$

Equation (2) shows that if $\mathbf{x}(k-i+1)$ is input, then filter $\hat{\mathbf{h}}(k+1)$ outputs the correct value $y(k-i+1)$. The number of taps L is larger than the number of equations p, *i.e.*, the set of equations is underdetermined. Therefore, the projection algorithm chooses the minimum-norm solution.

The ES (exponentially weighted stepsize) projection algorithm reflects the variation characteristics of a room impulse response on the projection algorithm [4]. The expected variation of a room impulse response becomes progressively smaller along the series by the same exponential ratio as the impulse response when a person moves or the environment changes [5]. Therefore, the ES projection algorithm adjusts coefficients with large errors in large steps and coefficients with small errors in small steps. For this purpose, it uses a different stepsize for each coefficient of an adaptive transversal FIR filter by introducing a stepsize matrix $\mathbf{A}$ with a diagonal form instead of the scalar stepsize of the projection algorithm. The diagonal components of this stepsize matrix are time-invariant and are set proportional to the expected variation of the room impulse response.

The p-*th* order ES projection algorithm updates the filter coefficient vector $\hat{\mathbf{h}}(k)$ as follows. When $\mathbf{A} = \mathbf{I}$ (unit matrix), the following equations correspond to the projection algorithm.

$$\begin{aligned}\hat{\mathbf{h}}(k+1) &= \hat{\mathbf{h}}(k) + \mu \mathbf{A}\mathbf{X}(k)[\mathbf{X}(k)^T \mathbf{A}\mathbf{X}(k) + \delta \mathbf{I}]^{-1}\mathbf{e}(k) \\ &= \hat{\mathbf{h}}(k) + \mu \mathbf{A}[\beta_1(k)\mathbf{x}(k) + \beta_2(k)\mathbf{x}(k-1) \\ &\quad + \cdots + \beta_p(k)\mathbf{x}(k-p+1)]\end{aligned} \tag{3}$$

$$\beta(k) = [\mathbf{X}(k)^T \mathbf{A}\mathbf{X}(k) + \delta \mathbf{I}]^{-1}\mathbf{e}(k) \tag{4}$$

$$\mathbf{X}(k) = [\mathbf{x}(k), \mathbf{x}(k-1), \cdots, \mathbf{x}(k-p+1)] \tag{5}$$

$$\begin{aligned}\mathbf{e}(k) &= \mathbf{y}(k) - \mathbf{X}(k)^T \hat{\mathbf{h}}(k) + \mathbf{n}(k) \\ &\cong [e(k), (1-\mu)e(k-1), \cdots, (1-\mu)^{p-1}e(k-p+1)]^T\end{aligned} \tag{6}$$

$$\beta(k) = [\beta_1(k), \beta_2(k), \cdots, \beta_p(k)]^T \tag{7}$$

$$\mathbf{x}(k) = [x(k), x(k-1), \cdots, x(k-L+1)]^T \tag{8}$$

$$\mathbf{y}(k) = [y(k), y(k-1), \cdots, y(k-p+1)]^T \tag{9}$$

$$\mathbf{n}(k) = [n(k), n(k-1), \cdots, n(k-p+1)]^T, \tag{10}$$

where

$$\mathbf{A} = \begin{pmatrix} \alpha_1 & & & 0 \\ & \alpha_2 & & \\ & & \ddots & \\ 0 & & & \alpha_L \end{pmatrix} \tag{11}$$

and

- μ: scalar stepsize ($0 < \mu < 2$),
- δ: small positive constant,
- $\alpha_i = \alpha_0 \gamma^{i-1} \quad (i = 1, \cdots, L)$,
- γ: exponential attenuation ratio of room impulse responses ($0 < \gamma \leq 1$).

The computational complexity of the ES projection algorithm is reduced by introducing an intermediate variable, $\mathbf{z}(k)$:

$$\mathbf{z}(k+1) = \mathbf{z}(k) + \mu \mathbf{A}[\beta_1(k-p+1) + \cdots + \beta_p(k)]\mathbf{x}(k-p+1) \tag{12}$$

$$\hat{y}(k) = \mathbf{z}(k)^T \mathbf{x}(k) + \mu \mathbf{r}(k)^T \mathbf{s}(k-1) \tag{13}$$

$$\begin{aligned}\mathbf{r}(k) &= [\mathbf{x}(k)^T \mathbf{A}\mathbf{x}(k-1), \mathbf{x}(k)^T \mathbf{A}\mathbf{x}(k-2), \\ &\quad \cdots, \mathbf{x}(k)^T \mathbf{A}\mathbf{x}(k-p+1)]^T\end{aligned} \tag{14}$$

$$\mathbf{s}(k-1) = \begin{pmatrix} \beta_1(k-1) \\ \beta_2(k-1) + \beta_1(k-2) \\ \cdots \\ \beta_{p-1}(k-1) + \beta_{p-2}(k-2) + \cdots + \beta_1(k-p+1) \end{pmatrix}. \tag{15}$$

Intermediate variable $\mathbf{z}(k)$ is related to impulse response replica $\hat{\mathbf{h}}(k)$:

$$\mathbf{z}(k) = \hat{\mathbf{h}}(k) - \mu \mathbf{A}[\mathbf{x}(k-1), \mathbf{x}(k-2), \cdots, \mathbf{x}(k-p+1)]\mathbf{s}(k-1). \tag{16}$$

Furthermore, by using the sliding windowed FTF (fast transversal filter), the computational complexity can be reduced to $2L + 20p$ multiply-add operations [6]-[8].

Since the ES (exponentially weighted stepsize) method is independent of the whitening of the received input [4][5], we shall focus on the projection algorithm for simplicity.

In the projection algorithm, filter coefficient vector $\hat{\mathbf{h}}(k)$ is adjusted in the direction of the plane produced by $\mathbf{x}(k)$, $\mathbf{x}(k-1)$, $\cdots$, $\mathbf{x}(k-p+1)$. In other words, the projection algorithm whitens the received input according to projection order p. Therefore, convergence can be improved for a colored input signal such as speech, where consecutive input vectors $\mathbf{x}(k)$, $\mathbf{x}(k-1)$, $\cdots$, $\mathbf{x}(k-p+1)$ are highly correlated.

3.2. Projection order and the whitening effect

The whitening effect of the projection algorithm is related to the number of taps of the adaptive filter. For a fullband echo canceller, where the number of taps is 1000, the convergence speed can be doubled with a projection order of two for speech input. However, little further improvement is achieved with a relatively small projection order. Even a projection order of 50 is not sufficient for nearly maximum convergence.

On the other hand, with a small number of taps L, increasing projection order still leads to a large increase in the convergence speed.

The projection algorithm can be explained as whitening by the Gram-Schmidt process and/or whitening by the linear prediction process. From this, it can be shown that the large received input vector space cannot be whitened by a small projection order, and a small received input vector space can be whitened by a small projection order.

4. NEW SUBBAND ECHO CANCELLER

4.1. Proposed subband projection echo canceller

In the conventional subband echo canceller, the received input in each subband is not fully whitened. Therefore, the projection algorithm is expected to whiten the received input in each subband and speed up the convergence.

In the subband echo canceller the number of taps is reduced depending on the downsampling rate R, which increases with increasing number of subbands N. The effectiveness of an increase in the projection order p depends on the number of taps in the adaptive filter.

Since an adaptive filter is downsampled, the number of taps needed in each subband is reduced by a factor of R. Consequently, the proposed subband projection echo canceller almost fully whitens the received input with a small projection order and achieves nearly maximum convergence with low computational complexity.

Furthermore, by reflecting the frequency characteristics of the speech and of the echo path, a different projection order can be chosen for each subband. This achieves optimum cost performance, which is promising for implementation.

5. COMPUTER SIMULATIONS

The impulse response used in computer simulations was measured in a room with a reverberation time of 300 ms and a speaker-microphone distance of 1 m, truncated with 512 taps. The sampling frequency was 16 kHz in the fullband. The number of taps L was also chosen to be 512 for the fullband, and reduced to $L_{SUB} = L/R$ in the subband. Ambient noise with a fixed SNR of 35 dB was added. The parameters μ and δ were adjusted to give a steady-state ERLE (echo return loss enhancement) of 35 dB. ERLE is defined by

$$ERLE = 10 \log_{10} \frac{p_y}{p_e} \quad \text{(dB)}, \qquad (17)$$

where

$$p_y = E[y(k)^2],$$
$$p_e = E[\tilde{e}(k)^2] = E[(y(k) - \hat{y}(k))^2].$$

The signal powers p_y and p_e were estimated from 50 data samples. Each curve is the average of 10 independent results for white noise and 50 for speech.

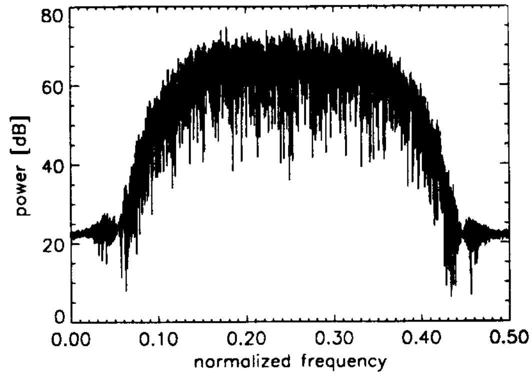


Figure 2: Spectral distribution of the received input in the subband.

5.1. General differences between fullband and subband

The bandlimited subband signals were oversampled by a factor 2 for efficient SSB modulation. Figure 2 shows the real valued subband signal at the reduced sampling rate for white noise input. Similar frequency distributions were obtained with an increase in the number of subbands N. Figure 2 shows that the received input in the subband is no longer equal to a white noise with continuous frequency distribution, but to a bandlimited 'white' noise with non-continuous frequency distribution.

This oversampling yields a higher eigenvalue spread in the subband compared to the fullband for white noise input. As a result, one typical difference between the fullband and subband is the rounded ERLE curve of the subband, due to the bandlimiting and oversampling effect in contrast to the straighter curve of the fullband as shown in Fig. 3 where $N = 1, 4$ and $p = 1$.

The rounded ERLE curve of the subband for a stationary white noise signal is the limit for a non-stationary colored signal like speech. Note that this rounded ERLE curve cannot be avoided with non-critical sampling of the subband signals.

5.2. Convergence curves for white noise input

Figure 4 shows the effectiveness of the projection algorithm in the subband when $N = 8$. An increase in the projection order p to *2nd*, *4th*, *8th* and *16th* order leads to much more rapid convergence. The rounded ERLE curve becomes straighter with increasing projection order p. This is due to the small number of taps in the adaptive filter. An increase to *32nd* order did not yield to a significant improvement compared to the *16th* order, and nearly maximum convergence, *i.e.*, a straight convergence curve like white noise input, was achieved with a projection order of 16, which is far faster than the NLMS ($p = 1$).

5.3. Convergence curves for speech input

Figure 5(a) shows the ERLE convergence for $N = 1$ fullband. It shows that the convergence speed can be more than doubled at a convergence level of 30 dB when using the *2nd* order projection algorithm. The plot shows that a further increase in p to *4th*, *8th*, *16th* and *32nd* order results only in slight further improvement. The projection order of 32 is not sufficient to fully whiten the received input and achieve ultimate convergence.

Figure 5(b) shows ERLE convergence for the $N = 32$ subband. In this case convergence speed is more than double for the $4th$ order compared to NLMS ($p = 1$). The convergence speed is still doubled for $p = 16$ in comparison with $p = 8$. The rounded ERLE curve becomes straighter with increasing projection order p. This is due to the small number of taps in the adaptive filter. An increase to $32nd$ order did not yield to a significant improvement compared to the $16th$ order, and nearly maximum convergence, *i.e.*, a straight convergence curve like white noise input, was achieved with a projection order of 16, which is far faster than the NLMS ($p = 1$).

6. CONCLUSIONS

By reflecting the short adaptive filter length of the subband echo canceller, the low-order projection algorithm almost fully whitens the received input in each subband. Computer simulation results show that nearly maximum convergence can be achieved with a projection order of 16. A different projection order can be chosen in different subbands by reflecting the frequency characteristics of speech and of the echo path. This gives our method optimum cost performance. This subband projection echo canceller has fast convergence and efficient computation, so it is promising for implementation.

7. ACKNOWLEDGEMENTS

We would like to thank Dr. N. Kitawaki, Mr. J. Kojima, Dr. Y. Kaneda, Mr. M. Tanaka, and Mr. S. Shimauchi for many fruitful discussions. We also thank Dr. W. Kellermann for his collaboration and detailed discussions.

8. REFERENCES

[1] W. Kellermann, "Analysis and design of multirate systems for cancellation of acoustical echoes," *Proc. ICASSP88*, pp. 2570-2573, Apr. 1988.

[2] K. Ozeki and T. Umeda, "An adaptive filtering algorithm using an orthogonal projection to an affine subspace and its properties," *Trans. IEICE Japan*, vol. J67-A, pp. 126-132, Feb. 1984 (in Japanese).

[3] R. Crochiere and L. Rabiner, *Multirate Digital Signal Processing*, Englewood Cliffs, NJ: Prentice-Hall, 1983.

[4] S. Makino and Y. Kaneda, "Exponentially weighted stepsize projection algorithm for acoustic echo cancellers," *Trans. IEICE Japan*, vol. E75-A, no. 11, pp. 1500-1508, Nov. 1992.

[5] S. Makino, Y. Kaneda, and N. Koizumi, "Exponentially weighted stepsize NLMS adaptive filter based on the statistics of a room impulse response," *IEEE Trans. Speech and Audio*, vol. 1, no. 1, pp. 101-108, Jan. 1993.

[6] M. Tanaka, Y. Kaneda, S. Makino, and J. Kojima, "Fast projection algorithm and its step size control," *Proc. ICASSP95*, pp. 945-948, May 1995.

[7] S. Gay and S. Tavathia, "The fast affine projection algorithm," *Proc. ICASSP95*, pp. 3023-3026, May 1995.

[8] M. Tanaka, Y. Kaneda, S. Makino, and J. Kojima, "A fast projection algorithm for adaptive filtering," *Trans. IEICE Japan*, vol. E78-A, no. 10, pp. 1355-1361, Oct. 1995.

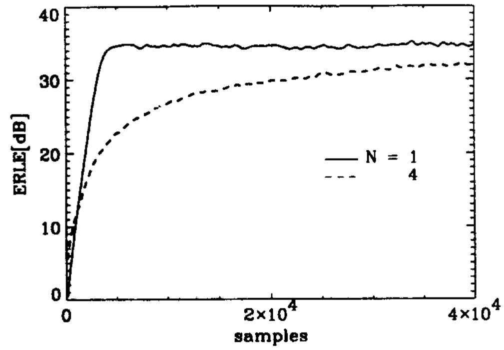


Figure 3: Comparison of fullband and subband convergence curves for a white noise input when $N = 1, 4$ and $p = 1$.

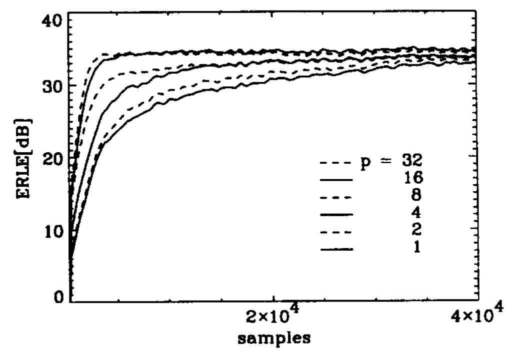
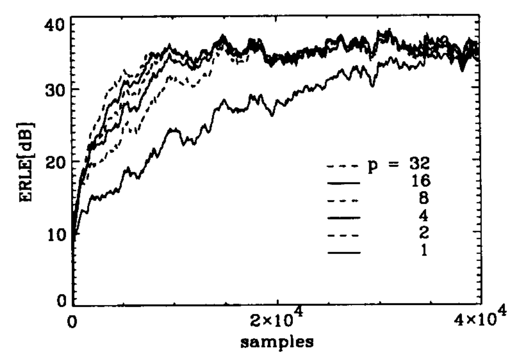
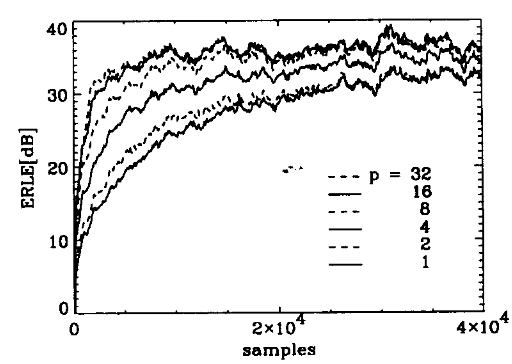


Figure 4: ERLE convergence for a white noise input when $N = 8$.



(a) $N = 1$ fullband



(b) $N = 32$ subband

Figure 5: ERLE convergence for speech input.