SUBBAND STEREO ECHO CANCELLER USING THE PROJECTION ALGORITHM WITH FAST CONVERGENCE TO THE TRUE ECHO PATH

Shoji Makino *Klaus Strauss* *Suehiro Shimauchi* *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 stereo echo canceller that converges to the true echo path impulse response much faster than conventional stereo echo cancellers. Since signals are bandlimited and downsampled in the subband structure, the time interval between the subband signals become longer, so the variation of the crosscorrelation between the stereo input signals becomes large. Consequently, convergence to the true solution is improved. Furthermore, the projection algorithm, or affine projection algorithm, is applied to further speed up the convergence. Computer simulations using stereo signals recorded in a conference room demonstrate that this method significantly improves convergence speed and almost solves the problem of stereo echo cancellation with low computational load.

1. INTRODUCTION

A stereo teleconferencing system provides a more realistic presence in teleconferencing compared to monaural systems. It helps listeners distinguish who is talking at the other end by means of spatial information. The most significant problem with stereo echo cancellation using the conventional linear combiner structure is that the adaptive filter often misconverges or, if not, its convergence speed is very slow because of the crosscorrelation between the stereo signals [1]. As a result, the conventional stereo echo canceller suffers from variation in both the near-end echo path and the far-end transmission path. There is a strong need for a stereo echo canceller that can identify the true echo path impulse response quickly with low computational complexity.

Several methods for overcoming this problem have been proposed. Some utilize the uncorrelated elements of stereo signals [2][3] and some the variations in the cross-correlation between stereo signals [4][5].

Subband echo cancellers, which divide signals into smaller frequency subbands and independently cancel echoes in each subband, have been studied [6]. Since the narrower frequency subbands have a smaller eigenvalue spread compared to the fullband for speech input, the convergence speed can be improved. Since downsampling expands the sampling interval and reduces the number of taps needed for the adaptive filter, the subband echo canceller is computationally efficient.

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

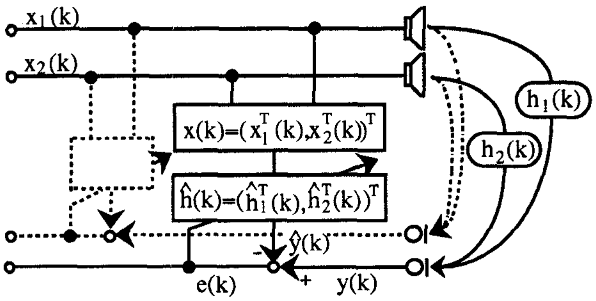


Figure 1: Configuration of stereo echo canceller.

In this paper, the subband structure is analyzed and introduced for stereo echo cancellation. We show that the subband structure improves the convergence to the true echo path impulse response by efficiently emphasizing the variation of the crosscorrelation between the stereo signals. We also show that the projection algorithm, or affine projection algorithm [7][8], can further emphasize the variation of the crosscorrelation between the stereo signals to achieve significantly faster convergence. The proposed method almost solves the problem of stereo echo cancellation with low computational load.

2. STEREO ECHO CANCELLATION

2.1. Problem of stereo echo cancellation

A stereo (two-channel) telecommunication system is shown in Fig. 1. Input signals $\mathbf{x}_1(k)$ and $\mathbf{x}_2(k)$ and filter coefficients $\hat{\mathbf{h}}_1(k)$ and $\hat{\mathbf{h}}_2(k)$ are combined as $\mathbf{x}(k) = [\mathbf{x}_1^T(k), \mathbf{x}_2^T(k)]^T$ and $\hat{\mathbf{h}}(k) = [\hat{\mathbf{h}}_1^T(k), \hat{\mathbf{h}}_2^T(k)]^T$. Thus, stereo echo cancellation is achieved by linearly combining two monaural echo cancellers.

The most significant problem with stereo echo cancellation using the conventional linear combiner structure is that the adaptive filter cannot identify the true echo path impulse response because of the crosscorrelation between the stereo signals [1].

Consequently, talker movement or changes in transmission room are considered as variation in the echo path in the receiving room. Accordingly, the performance of the stereo echo canceller degrades at the instant of abrupt changes in the environment in the transmission room.

However, in practical situations, there are several reasons that make the stereo echo canceller converge to

0-8186-7919-0/97 $10.00 © 1997 IEEE

the true echo path.
(1) The stereo signals $\mathbf{x}_1(k)$ and $\mathbf{x}_2(k)$ contain independent noise.
(2) The length of the adaptive filter $\hat{\mathbf{h}}_1(k)$ and $\hat{\mathbf{h}}_2(k)$ is shorter than that of the impulse response in the transmission room. These truncated components act as independent noise.
(3) The crosscorrelation between the stereo signals $\mathbf{x}_1(k)$ and $\mathbf{x}_2(k)$ varies slightly even when the talker does not move his body or head while speaking.

Since reason (3) is the most promising in practical situations, we focus on the variation of the crosscorrelation between the stereo signals.

2.2. Effect of variation in stereo crosscorrelation

Recently, the effect of crosscorrelation on a stereo echo canceller was studied and it was shown that the variation in crosscorrelation between stereo signals is effective for coefficient error convergence [4][5]. Also, the stereo projection algorithm was shown to be effective in identifying the true echo path impulse response by emphasizing the slightly varying crosscorrelation between stereo signals in actual teleconferencing situations.

3. NEW STEREO ECHO CANCELLER

3.1. Subband effect on the variation of the crosscorrelation

The configuration of the proposed subband stereo echo canceller is shown in Fig. 2. In the subband structure, signals are divided into N smaller frequency subbands and downsampled by a factor R, called the downsampling rate. As a result, the sampling interval is expanded longer than that of the fullband. This procedure emphasizes the variation of crosscorrelation in the stereo signals (we call this effect *de-crosscorrelation*). Consequently, convergence speed to the true echo path impulse response can be improved. Of course, since the subband structure effectively *de-autocorrelates* the two input signals, convergence can be improved for speech input. Since downsampling expands the sampling interval and reduces the number of taps needed for the adaptive filter, the proposed subband stereo echo canceller is computationally efficient.

3.2. Proposed subband stereo projection echo canceller

The stereo projection algorithm has been shown to be effective in fullband stereo echo cancellation [3]-[5]. This algorithm emphasizes the variation of crosscorrelation in the stereo signals and also de-autocorrelates the two input signals, so convergence speed to the true echo path impulse response can be improved. Here, we use the projection algorithm in the subband structure to further improve convergence speed.

The p-th order projection algorithm, or affine projection algorithm, updates filter coefficient vector $\hat{\mathbf{h}}(k)$ as follows [7][8].

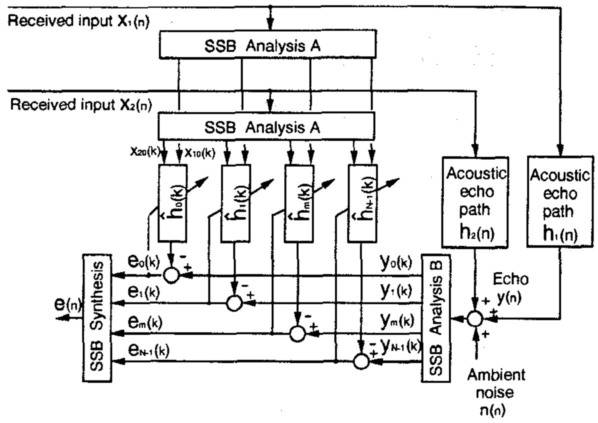


Figure 2: Proposed subband stereo echo canceller.

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

$$[\beta_1(k), \beta_2(k), \cdots, \beta_p(k)]^T = [\mathbf{X}(k)^T\mathbf{X}(k) + \delta\mathbf{I}]^{-1}\mathbf{e}(k) \quad (2)$$

$$\mathbf{X}(k) = [\mathbf{x}(k), \mathbf{x}(k-1), \cdots, \mathbf{x}(k-p+1)] \quad (3)$$

$$\begin{aligned}&\mathbf{e}(k) = \mathbf{y}(k) - \mathbf{X}(k)^T\hat{\mathbf{h}}(k) + \mathbf{n}(k) \\ &= [e(k), (1-\mu)e(k-1), \cdots, (1-\mu)^{p-1}e(k-p+1)]^T \quad (4)\end{aligned}$$

$$\mathbf{x}(k) = [x(k), x(k-1), \cdots, x(k-L+1)]^T \quad (5)$$

$$\mathbf{y}(k) = [y(k), y(k-1), \cdots, y(k-p+1)]^T \quad (6)$$

$$\mathbf{n}(k) = [n(k), n(k-1), \cdots, n(k-p+1)]^T, \quad (7)$$

where

μ: scalar stepsize ($0 < \mu < 2$),
δ: small positive constant.

Since the number of taps needed in each subband is reduced by downsampling with a factor of R, the projection algorithm can decorrelate the received input of a small-tap adaptive filter with a relatively small projection order [6]. The proposed subband stereo projection echo canceller is expected to de-crosscorrelate as well as de-autocorrelate the stereo input signals, and to achieve much faster convergence.

By introducing an intermediate variable and using the sliding windowed FTF (fast transversal filter) [9]-[12], the computational complexity can be reduced to $4L + 32p$ multiply-add operations.

3.3. Relation between the subband effect and the projection algorithm

The p-th order projection algorithm, or affine projection algorithm, updates filter coefficient $\hat{\mathbf{h}}(k+1)$, which satisfies (8) where $p < L$.

$$\hat{\mathbf{h}}(k+1)^T\mathbf{x}(k-i+1) = y(k-i+1) \qquad (i = 1, 2, \cdots, p) \quad (8)$$

Equation (8) shows that if $\mathbf{x}(k-i+1)$ is input, then filter $\hat{\mathbf{h}}(k+1)$ outputs the correct value $y(k-i+1)$.

On the other hand, in subband processing with downsampling rate R, $R-1$ samples are redundant and (8) is automatically satisfied for $i = 1, 2, \cdots, R$. Thus, subband processing with downsampling rate R, filter length L/R and update occasion $1/R$, is equivalent to the *R-th* order projection algorithm with original sampling, filter length L and update occasion $1/R$.

Similarly, the proposed method with downsampling rate R, projection order p, filter length L/R and update occasion $1/R$ is equivalent to the *pR-th* order projection algorithm with original sampling, filter length L and update occasion $1/R$.

4. COMPUTER SIMULATIONS

In our computer simulations, the input stereo signals were speech recorded by two microphones. Two speakers, who remained in their places, spoke alternately. The true echo path impulse responses were measured in a conference room with a reverberation time of 250 ms. The number of taps, L, was 512 in each filter of $\hat{\mathbf{h}}_1(k)$ and $\hat{\mathbf{h}}_2(k)$. The sampling frequency was 8 kHz. Ambient noise with a fixed SNR of 35 dB was added. The stepsizes α and small positive constants δ of the projection algorithm were adjusted to give a steady-state coefficient error of -30 dB. The adaptive filter coefficients in the subbands were synthesized to fullband and the performances were evaluated by the coefficient error in the fullband.

4.1. Effect of projection order when $N = 1$ (fullband)

Figure 3 shows the effectiveness of the stereo projection algorithm in the fullband ($N = 1$). Although the talker did not move his body or head while speaking, the crosscorrelation between the stereo signals varied slightly in this real teleconferencing situation. As a result, the coefficient error converged slowly when $p = 1$ (NLMS algorithm). However the convergence improved considerably with the projection order. An increase in the projection order p to *2nd*, *4th*, *8th* and *32nd* order leads to much more rapid convergence.

4.2. Effect of number of subband when $p = 1$ (NLMS algorithm)

Figure 4 shows coefficient error convergence for $p = 1$ (NLMS). The downsampling rate is given by $R = N/4$. The convergence improved considerably with the number of subband. An increase in the number of subband N to 4, 8 and 32 leads to much more rapid convergence. It can be seen that Figs. 3 and 4 roughly look the same. The subband structure has almost the same effect as the stereo projection algorithm as discussed in section 3.3.

4.3. Convergence curves of the proposed subband stereo projection echo canceller

Figure 5 (a) shows coefficient error convergence for $N = 32$ subbands. Convergence speed can be improved by increasing the projection order. The convergence with a projection order of 32 is more than doubled compared to that with $p = 1$ (NLMS).

Although the coefficient error is specially important for stereo echo cancellation, the residual echo is also important for practical evaluation. Figure 5 (b) shows residual echo convergence for $N = 32$ subbands. This

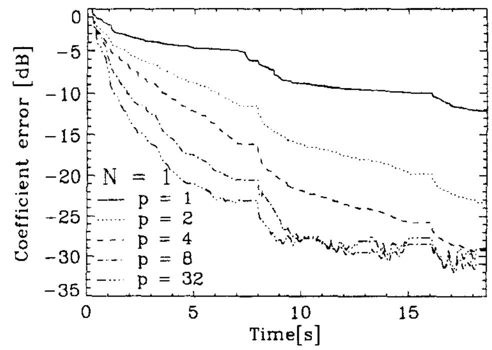


Figure 3: Coefficient error convergence for speech input when $N = 1$ fullband.

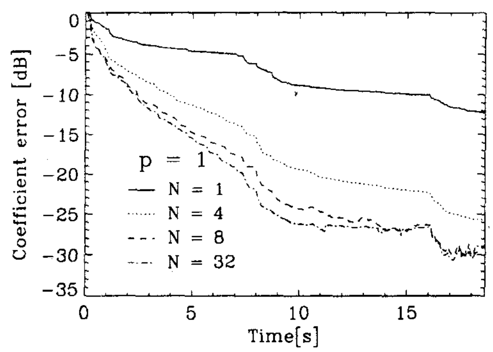


Figure 4: Coefficient error convergence for speech input when $p = 1$ NLMS algorithm.

figure shows that the residual echo is not affected by speaker alternation in the transmission room.

4.4. Comparison of fullband and subband

Figure 6 shows the effectiveness of the subband structure. Instead of using the fullband with a projection order of $p = 32$, which has high computational load, we can use $N = 32$ subbands with the projection order of $p = 8$ and yields low computational load.

4.5. Comparison with independent white noise input

Figure 7 shows that the fastest combination of the proposed method here, with $N = 32$ and $p = 32$, is about one third of the theoretically fastest convergence with independent white noise inputs. Considering the non-stationary characteristic of speech, we could say that the problem of stereo echo cancellation is almost solved by the proposed method.

5. CONCLUSIONS

We proposed a new stereo echo canceller that uses a subband structure and the stereo projection algorithm. The subband structure improves the convergence to the true echo path impulse response by efficiently emphasizing the variation of the crosscorrelation between the stereo signals. The projection algorithm further emphasizes the variation of the crosscorrelation between the stereo signals to achieve significantly faster convergence. Computer simulations showed that it can expedite coefficient error convergence much faster than the conventional stereo NLMS echo canceller and almost solves the problem of stereo echo cancellation.

6. ACKNOWLEDGEMENTS

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

7. REFERENCES

[1] M. Sondhi, D. Morgan, and J. Hall, "Stereophonic acoustic echo cancellation - an overview of the fundamental problem," *IEEE SP letters*, vol. 2, no, 8, pp. 148-151, Aug. 1995.

[2] J. Benesty, F. Amand, A. Gilloire, and Y. Grenier, "Adaptive filtering algorithm for stereophonic acoustic echo cancellation," *Proc. ICASSP95*, pp. 3099-3102, May 1995.

[3] J. Benesty, P. Duhamel, and Y. Grenier, "A multichannel affine projection algorithm with applications to multichannel acoustic echo cancellation," *IEEE SP letters*, vol. 3, no, 2, pp. 35-37, Feb. 1996.

[4] S. Shimauchi and S. Makino, "Stereo projection echo canceller with true echo path estimation," *Proc. ICASSP95*, pp. 3059-3062, May 1995.

[5] S. Shimauchi and S. Makino, "Stereo echo cancellation algorithm using imaginary input-output relationships," *Proc. ICASSP96*, pp. 941-944, May 1996.

[6] S. Makino, J. Noebauer, Y. Haneda, and A. Nakagawa, "SSB subband echo canceller using low-order projection algorithm," *Proc. ICASSP96*, pp. 945-948, May 1996.

[7] 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).

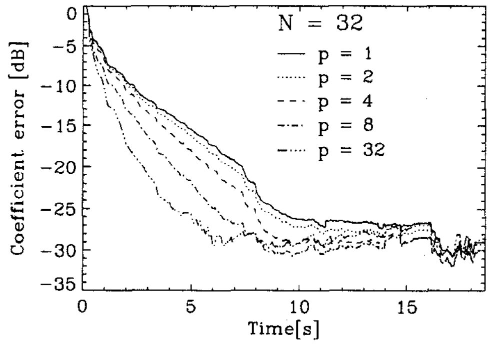
[8] S. Makino and Y. Kaneda, "Exponentially weighted stepsize projection algorithm for acoustic echo cancellers," *Trans. IEICE Japan*, vol. E75-A, no. 11, pp. 1500-1508, Nov. 1992.

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

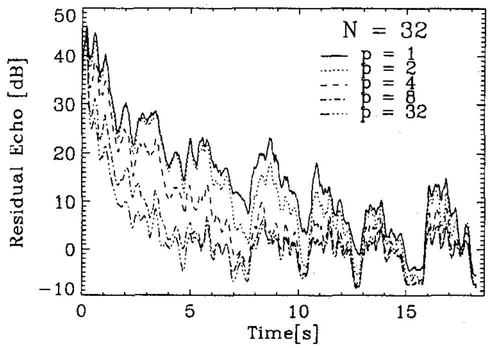
[10] S. Gay and S. Tavathia, "The fast affine projection algorithm," *Proc. ICASSP95*, pp. 3023-3026, May 1995.

[11] 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.

[12] F. Amand, J. Benesty, A. Gilloire, and Y. Grenier, "A fast two-channel projection algorithm for stereophonic acoustic echo cancellation," *Proc. ICASSP96*, pp. 949-952, May 1996.



(a) Coefficient error



(b) Residual echo

Figure 5: Convergence for speech input when $N = 32$ subbands.

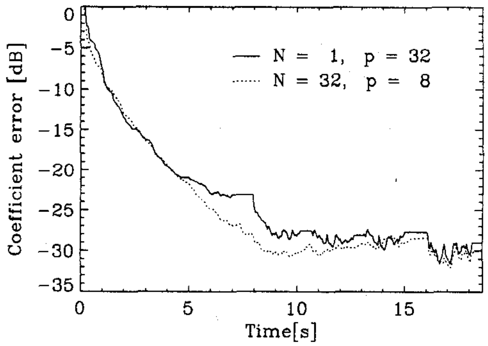


Figure 6: Comparison of fullband and subband convergence curves when $N = 1, p = 32$ and $N = 32, p = 8$.

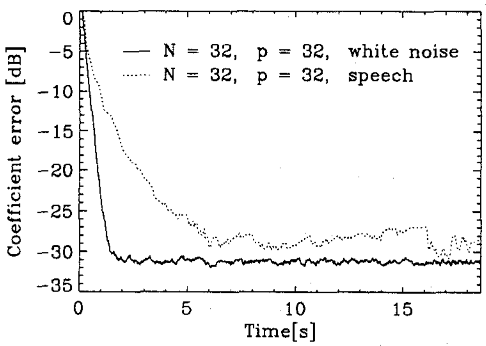


Figure 7: Comparison of convergence curves for an independent white noise inputs when $N = 32, p = 32$.