

ステレオエコーキャンセラにおける 前処理を用いないエコーパス同定 Echo Path Identification without Pre-Processing in Stereophonic Acoustic Echo Canceller

田中 正彦 平野 晃宏 中山 謙二
Masahiko Tanaka Akihiro Hirano Kenji Nakayama

金沢大学大学院 自然科学研究科 電子情報システム専攻
Dept. of Electrical and Computer Eng., Faculty of Eng., Kanazawa Univ.
E-mail: nakayama@t.kanazawa-u.ac.jp

あらまし

ステレオエコーキャンセラにおける前処理を用いないエコーパス同定法を提案する。ステレオエコーキャンセラでは不定解が無限に存在する係数不確定性の問題がある。これに対して一意解の存在条件はあるが、両方のエコーキャンセラには成立しない。そこで本稿では、フィルタ係数を分割することで2つのエコーキャンセラについて解が一意に定まる学習法を提案する。計算機シミュレーションを行った結果、係数誤差が-10dB低減し、不定解を避けられることを示した。

ABSTRACT

This paper proposes echo path identification method without pre-processing in stereophonic acoustic echo cancellers. In stereophonic acoustic echo cancellers, the uniqueness problem in which no unique solution for filter coefficients exists is the most serious problem. Regarding this problem, there are conditions for unique solutions, but don't satisfy for both echo cancellers simultaneously. So this paper proposes the learning method that the filter coefficients have unique solutions on both echo cancellers by dividing filter coefficients. Computer simulation results shows that proposed method reduces coefficient error by 10dB and can avoid unsettled solutions.

1 まえがき

TV 会議システムやハンズフリー通信では、スピーカからマイクロホンへの音声の回り込みによって音響

エコーが発生し、快適な会話の妨げとなっている。この音響エコーを除去するために音響エコーキャンセラが広く用いられている。現在、TV 会議システムにおいてより臨場感溢れるものが求められており、そのために音声の多チャンネル化が検討されている [1]-[11]。ステレオや多チャンネルエコーキャンセラの開発はいまだ途上であり、今後発展の期待される分野である。

ステレオ音声信号においては、両チャンネルの信号間に相互相関がある場合が多い。ステレオ TV 会議において一人の話者だけが発言している場合には、両チャンネルの信号は同じ音声成分のみとなり、相互相関は非常に強くなる。このような相互相関がステレオエコーキャンセラの収束特性に大きな影響を与えることが知られている。特に話者が一人である場合には、エコーを消去できるフィルタ係数の解は一意に定まらないという、解の不確定性問題が起こるとされている [3], [6], [7], [8]。これに対して、一意解が存在する条件は明らかにされた [10] が、この条件は片方向でしか成立しないため 2 つのエコーキャンセラを同時に学習する事ができない。本論文では、双方向で一意解の存在条件を仮想的に満たす学習法を提案する。

2 ステレオエコーキャンセラ

2.1 基本構成

本稿では、最も一般的なステレオエコーキャンセラである線形結合形 [1], [2] を対象とする。図 1 に線形結合形ステレオエコーキャンセラを用いた音声会議を示す。

ステレオエコーキャンセラは 4 つの適応フィルタから成り、各適応フィルタは対応するエコーパス特性を

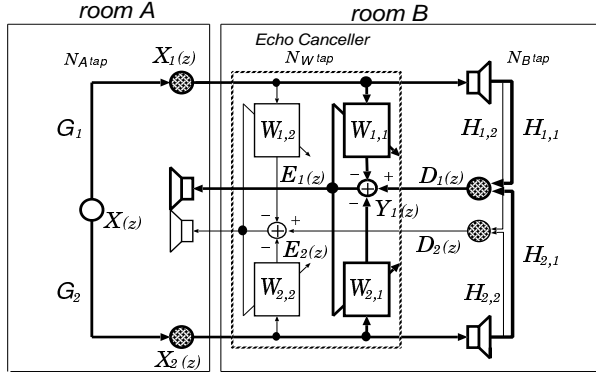


図 1: ステレオエコーキャンセラの構成

それぞれ同定し、エコーを除去する。

図 1において、第 i チャンネルの参照入力信号を $X_i(z)$ 、第 i 番目のスピーカから第 j 番目のマイクロホンに至るエコーパスを $H_{i,j}(z)$ とすると、第 j チャンネルに混入するエコー $D_j(z)$ は、

$$D_j(z) = H_{1,j}(z)X_1(z) + H_{2,j}(z)X_2(z) \quad (1)$$

となる。同様に、 $H_{i,j}(z)$ を推定する適応フィルタの伝達関数を $W_{i,j}(z)$ とすると、疑似エコー $Y_j(z)$ は、

$$Y_j(z) = W_{1,j}(z)X_1(z) + W_{2,j}(z)X_2(z) \quad (2)$$

である。誤差信号 $E_j(z)$ は、

$$E_j(z) = D_j(z) - Y_j(z) \quad (3)$$

で求められる。 $E_j = 0$ となる、すなわち任意の参照入力信号に対してエコー信号が消去される最適解は、

$$W_{i,j}(z) = H_{i,j}(z) \quad (4)$$

である。

2.2 係数不確定性問題 [3],[7]

話者の音声 $X(z)$ が部屋の伝達関数 $G_i(z)$ ($i = 1, 2$) を経てマイクロホンへの入力として $X_i(z)$ になるとする。 $X_i(z)$ は、

$$X_i(z) = G_i(z)X(z) \quad (5)$$

で与えられる。このとき、誤差信号は

$$E_j(z) = \{(H_{1,j}(z) - W_{1,j}(z))G_1(z) + (H_{2,j}(z) - W_{2,j}(z))G_2(z)\}X(z) \quad (6)$$

となる。エコーを消去できる条件は、 $E_j(z) = 0$ より、

$$G_1(z)(H_{1,j}(z) - W_{1,j}(z)) + G_2(z)(H_{2,j}(z) - W_{2,j}(z)) = 0 \quad (7)$$

となる。1つの方程式に2個の未知数 $W_{1,j}(z)$ と $W_{2,j}(z)$ が存在するため、フィルタ係数の解が一意に定まらず、解が無数個あることが分かる。これが解の不確定性と呼ばれる問題で、ステレオエコーキャンセラにおける最大の問題となっている [7]。

2.3 係数の一意解の存在条件

図 1において部屋 A および部屋 B の伝達特性は各々 $N_A - 1$ 次および $N_B - 1$ 次の FIR フィルタで表現できるものとし、適応フィルタのタップ数を N_W とする。前節の係数不確定性問題に対して、既に $N_A > N_W \geq N_B$ (A の部屋で話者が発言した場合) の条件の下で解が一意に定まる [10] ことが知られている。しかし、逆方向 (B の部屋で話者が発言した場合) については $N_B > N_W \geq N_A$ という条件が必要になる。この2つの条件は明らかに同時に成立しないため双方向の学習をする事ができない。双方向について条件を成立させるには N_W を N_A, N_B より小さくする必要がある。単に小さくするだけではエコーパスより適応フィルタのタップ数が不足するためエコー消去機能が劣化する。これを解決する一つの方法として、分割学習を提案する。

3 分割学習

3.1 誤差信号

まず、 $N_B > N_W$ で適応フィルタのタップ数が足りない場合の誤差信号を調べる。

時刻 n における話者の音声を $s(n)$ とし、部屋 A における話者から第 i チャンネルのマイクロホンに至る経路のインパルス応答ベクトルを g_i とすると、第 i チャンネルの参照入力信号 $x_i(n)$ は、

$$x_i(n) = g_i^T s(n) \quad (8)$$

$$g_i = [g_{i,0}, g_{i,1}, \dots, g_{i,N_A-1}]^T \quad (9)$$

$$s(n) = [s(n), \dots, s(n - N_A + 1)]^T \quad (10)$$

となる。上付文字 T は転置を意味する。

第 i 番目のスピーカから第 j 番目のマイクロホンに至るエコーパスのインパルス応答ベクトルおよび適応フィルタの係数ベクトルを各々 $h_{i,j,f}$ 、 $w_{i,j}(n)$ とし、フィルタのタップ数が足りない部分に対応するエコーパスのインパルス応答ベクトルを $h_{i,j,b}$ とすると、エコー $d_j(n)$ および疑似エコー $y_j(n)$ は、

$$d_j(n) = \sum_{i=1}^2 \{h_{i,j,f}^T x_{i,f}(n) + h_{i,j,b}^T x_{i,b}(n - N_W)\} \quad (11)$$

$$y_j(n) = \sum_{i=1}^2 \mathbf{w}_{i,j}^T(n) \mathbf{x}_{i,f}(n) \quad (12)$$

で与えられる。 $\mathbf{w}_{i,j,f}(n)$, $\mathbf{h}_{i,j,f}$, $\mathbf{h}_{i,j,b}$, $\mathbf{w}_{i,j,b}(n)$, $\mathbf{x}_{i,f}(n)$, $\mathbf{x}_{i,b}(n)$ は,

$$\mathbf{w}_{i,j}(n) = [w_{i,j,0}(n), \dots, w_{i,j,N_W-1}(n)]^T \quad (13)$$

$$\mathbf{h}_{i,j,f} = [h_{i,j,0}, \dots, h_{i,j,N_W-1}]^T \quad (14)$$

$$\mathbf{h}_{i,j,b} = [h_{i,j,N_W}, \dots, h_{i,j,N_B}]^T \quad (15)$$

$$\mathbf{x}_{i,f}(n) = [x_i(n), \dots, x_i(n - N_W + 1)] \quad (16)$$

$$\mathbf{x}_{i,b}(n) = [x_i(n - N_W), \dots, x_i(n - N_B)] \quad (17)$$

で定義される。これより、誤差信号 $e_j(n)$ は,

$$e_j(n) = \sum_{i=1}^2 \{ \mathbf{h}_{i,j,f} - \mathbf{w}_{i,j}(n) \}^T \mathbf{x}_{i,f}(n) + \sum_{i=1}^2 \mathbf{h}_{i,j,b}^T \mathbf{x}_{i,b}(n - N_W) \quad (18)$$

で得られる。ここで \mathbf{x}_i を式 (8) から

$$e_j(n) = \sum_{i=1}^2 \{ \mathbf{h}_{i,j,f} - \mathbf{w}_{i,j}(n) \}^T \mathbf{G}_{i,f} \mathbf{s}_f(n) + \sum_{i=1}^2 \mathbf{h}_{i,j,b}^T \mathbf{G}_{i,b} \mathbf{s}_b(n - N_W) \quad (19)$$

$\mathbf{s}_f(n), \mathbf{s}_b(n - N_W)$ は

$$\mathbf{s}_{i,f}(n) = [s_i(n), \dots, s_i(n - N_W + 1)] \quad (20)$$

$$\mathbf{s}_{i,b}(n - N_W) = [s_i(n - N_W), \dots, s_i(n - N_B)] \quad (21)$$

であり、 \mathbf{G}_i は式 (22) のような行列である。さらに行列を用いて整理すると、

$$e_j(n) = \begin{bmatrix} \mathbf{h}_{1,j,f}^T - \mathbf{w}_{1,j}^T(n) \\ \mathbf{h}_{2,j,f}^T - \mathbf{w}_{2,j}^T(n) \end{bmatrix} \begin{bmatrix} \mathbf{G}_{1,f} \\ \mathbf{G}_{2,f} \end{bmatrix} \mathbf{s}_f(n) + \begin{bmatrix} \mathbf{h}_{1,j,b}^T \\ \mathbf{h}_{2,j,b}^T \end{bmatrix} \begin{bmatrix} \mathbf{G}_{1,b} \\ \mathbf{G}_{2,b} \end{bmatrix} \mathbf{s}_b(n - N_W) = \mathbf{D}_f^T(n) \mathbf{G}_f \mathbf{s}_f(n) + \mathbf{H}_b^T \mathbf{G}_b \mathbf{s}_b(n - N_W) \quad (23)$$

$$\mathbf{D}_f(n) = \begin{bmatrix} \mathbf{h}_{1,j,f} - \mathbf{w}_{1,j}(n) \\ \mathbf{h}_{2,j,f} - \mathbf{w}_{2,j}(n) \end{bmatrix}$$

$$\mathbf{H}_b = \begin{bmatrix} \mathbf{h}_{1,j,b} \\ \mathbf{h}_{2,j,b} \end{bmatrix}$$

これより $e_j(n)$ の二乗平均誤差を求めると

$$E[e_i^2(n)] = E[e_i(n)e_i^T(n)] = E[\mathbf{D}_f^T(n) \mathbf{G}_f \mathbf{s}_f(n) \mathbf{s}_f^T(n) \mathbf{G}_f^T \mathbf{D}_f(n)] \quad (24)$$

$$+ \mathbf{H}_b^T \mathbf{G}_b \mathbf{s}_b(n - N_W) \mathbf{s}_b^T(n - N_W) \mathbf{G}_b^T \mathbf{H}_b + 2\mathbf{D}_f^T(n) \mathbf{G}_f \mathbf{s}_f(n) \mathbf{s}_b^T(n - N_W) \mathbf{G}_b^T \mathbf{H}_b = \overline{\mathbf{D}_f^T(n) \mathbf{G}_f \mathbf{R}_f \mathbf{G}_f^T \mathbf{D}_f(n)} + \mathbf{H}_b^T \mathbf{G}_b \mathbf{R}_b \mathbf{G}_b^T \mathbf{H}_b + 2\overline{\mathbf{D}_f^T(n) \mathbf{G}_f \mathbf{R}_{fb} \mathbf{G}_b^T \mathbf{H}_b} = \overline{\mathbf{D}_f^T(n) \mathbf{Q}_f \mathbf{D}_f(n)} + \mathbf{H}_b^T \mathbf{Q}_b \mathbf{H}_b + 2\mathbf{D}_f^T(n) \mathbf{Q}_{fb} \mathbf{H}_b \quad (25)$$

ただし、

$$\mathbf{R}_f = \overline{\mathbf{s}_f(n) \mathbf{s}_f^T(n)}, \quad \mathbf{R}_{fb} = \overline{\mathbf{s}_f(n) \mathbf{s}_b^T(n - N_W)},$$

$$\mathbf{R}_b = \overline{\mathbf{s}_b(n - N_W) \mathbf{s}_b^T(n - N_W)},$$

$\mathbf{Q}_f = \mathbf{G}_f \mathbf{R}_f \mathbf{G}_f^T$, $\mathbf{Q}_{fb} = \mathbf{G}_f \mathbf{R}_{fb} \mathbf{G}_b^T$, $\mathbf{Q}_b = \mathbf{G}_b \mathbf{R}_b \mathbf{G}_b^T$ で \mathbf{Q}_f は正方形行列である。上付文字 $\overline{\quad}$ は期待値を意味する。

これより、二乗平均誤差が最小となる条件は

$$\frac{\partial e_j^2(n)}{\partial \mathbf{D}_f(n)} = 0 \text{ より}$$

$$\frac{\partial e_j^2(n)}{\partial \mathbf{D}_f(n)} = 2\mathbf{Q}_f \overline{\mathbf{D}_f(n)} + 2\mathbf{Q}_{fb} \mathbf{H}_b = 0 \quad (26)$$

$$\overline{\mathbf{D}_f(n)} = \mathbf{Q}_f^{-1} \mathbf{Q}_{fb} \mathbf{H}_b \quad (27)$$

上式の右辺がタップ数が足りないために生じる誤差になる。この誤差を減少させるため分割学習を行う。

3.2 学習方法

以下に分割学習の手順を示す。

- 適応フィルタの係数列 (N_W タップ) を前後に等しく 2 分割
- 最初に分割した前半部分 ($N_W/2$ タップ) を学習
このとき、後半部分 ($N_W/2$ タップ) は 0 のままで学習は行わない
- 一定の時間間隔をおいて前半部分の学習を止め、後半部分の学習
- 前半部分と後半部分を切替えて学習を繰り返す

この 2 分割学習により適応フィルタのタップ長は仮想的に $N_W/2$ とみなすことができ、 N_A, N_B より低減できる。式 (27) を変形すると

$$\mathbf{W}_f^{(1)} = \mathbf{H}_f + \mathbf{K}_f \mathbf{H}_b \quad (28)$$

となる。ただし、 $\mathbf{K}_f = \mathbf{Q}_f^{-1} \mathbf{Q}_{fb}$ である。従って、最初の前半部分の学習で係数は最適値にタップ数の不足による誤差を加えた値に収束する。次の後半部分の学習で係数は最適値に前半部分の誤差を加えた次式の値に収束する。

$$\mathbf{G}_{i,p} = \left[\begin{array}{cccccccc} g_{i,p,0} & g_{i,p,1} & \cdots & g_{i,p,N_A-1} & 0 & \cdots & \cdots & 0 \\ 0 & g_{i,p,0} & g_{i,p,1} & \cdots & g_{i,p,N_A-1} & \ddots & \vdots & \vdots \\ \vdots & 0 & \ddots & \ddots & \cdots & \ddots & \ddots & \vdots \\ \vdots & \vdots & \ddots & \ddots & \ddots & \cdots & \ddots & 0 \\ 0 & \cdots & \cdots & 0 & g_{i,p,0} & g_{i,p,1} & \cdots & g_{i,p,N_A-1} \end{array} \right] \Bigg\} N \quad (22)$$

$N_W + N_A - 1$

$$p = f \text{ or } b$$

$$\text{if } p = f: N = N_W, \quad \text{if } p = b: N = N_B - N_W$$

$$\mathbf{W}_b^{(1)} = \mathbf{H}_b + \mathbf{K}_b(\mathbf{H}_f - \mathbf{W}_f^{(1)}) \quad (29)$$

ただし、 $\mathbf{K}_b = \mathbf{Q}_b^{-1} \mathbf{Q}_{bf}$ である。そして、次の学習では

$$\mathbf{W}_f^{(2)} = \mathbf{H}_f + \mathbf{K}_f(\mathbf{H}_b - \mathbf{W}_b^{(1)}) \quad (30)$$

に収束する。式(30)において $(\mathbf{H}_b - \mathbf{W}_b^{(1)})$ が \mathbf{H}_b より減少していれば誤差は前後の切替え学習を繰り返すことで徐々に減少できると考えられる。

4 計算機シミュレーション

前章の解析結果を元に従来法と提案法との比較を計算機シミュレーションにより行った。評価尺度として、正規化係数誤差ノルム NCEN(Normalized Coefficient Error Norm)

$$NCEN(n) = \frac{\sum_{i=1}^2 \|w_{i,j}(n) - h_{i,j}\|^2}{\sum_{i=1}^2 \|h_{i,j}\|^2} \quad (31)$$

を用いた。適応アルゴリズムには NLMS アルゴリズム、学習の切替間隔を 2000 サンプル毎とし、入力信号を白色ガウス雑音とした。又、今回は $N_W = N_A = N_B = 12 \text{ tap}$ とした。フィルタ係数(前半部分)の初期値を変えて、いくつかシミュレーションを行った。結果の一つを、係数誤差を図2に、フィルタ係数の軌跡を図3に、出力誤差を図4に示した。

図2において、従来法が直ちに不定解に収束しているのに対し、提案法は係数誤差が-10dB以上低減されている。又、従来法はフィルタ係数の初期値に依存して収束値が異なるのに対して提案法はすべての初期値に対して同じ収束特性を示した。従って、従来法は不定解に、提案法は一意解に収束していると判断できる。従来法に比べて収束速度が遅いのは分割するためにフィルタ係数の半分しか学習しないためであり、さらに係数の切替間隔による影響も考えられる。

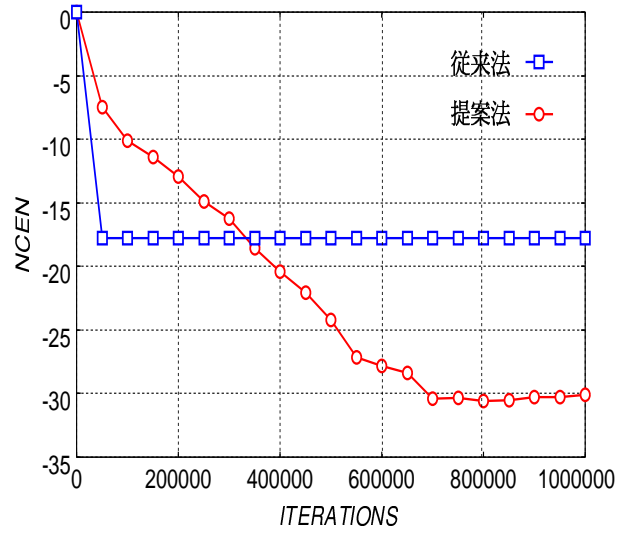


図 2: 係数誤差

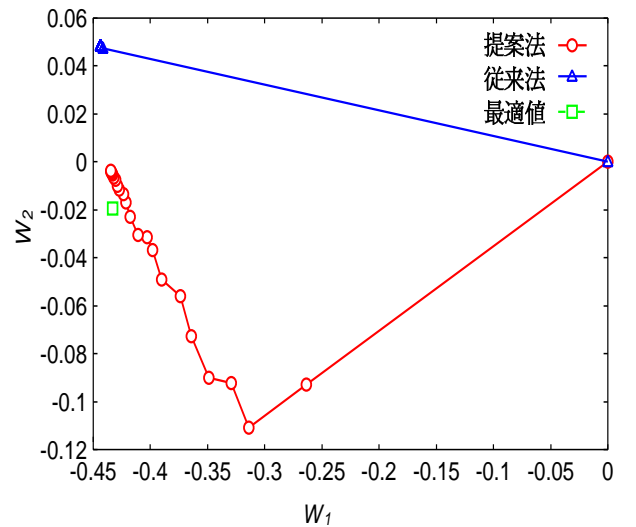


図 3: 係数の軌跡

図3において、従来法は勾配法に従って不定解へと直線的に収束しているが、提案法では不定解を避け、最適解に近付いている様子が分かる。

図4において、提案法は従来法に比べて出力誤差が大きくなっている。これより提案法が不定解でない値に収束していることが判断できる。

5 むすび

線形結合型ステレオエコーキャンセラにおいて、2つあるエコーキャンセラの内、両方に対して解を一意に決定させるために適応フィルタの分割学習を提案し、計算機シミュレーションにより従来法との比較を行った。従来法に比べて係数誤差が-10dB以上小さくなった。フィルタ係数の初期値を変化させても同じ結果である事から提案法では不定解を避けて一意解に収束していることが言える。しかし、出力誤差が従来法に比べ大きくなっている事から一意解ではあるが最適解からずれた値に収束していることが分かった。

今後の課題として、提案法の学習において収束性が保証されているのか理論的に解析する必要があること、又、今回は単にフィルタ係数を前後に等しく分割したが他の分割法などによる学習や係数の切替間隔について検討していく必要がある。提案法による収束速度の改善も検討が必要である。

参考文献

[1] 藤井哲郎, 島田正治, “多チャンネルエコーキャンセラの一構成法,” 信学技報, CS-84-178, pp. 7-14, 1984.

[2] M. M. Sondhi and D. R. Morgan, “Acoustic echo cancellation for stereophonic teleconferencing,” Proc. of IEEE ASSP Workshop Applied Signal Processing Audio Acoustics, May 9, 1991.

[3] A. Hirano and A. Sugiyama, “Convergence characteristics of a multi-channel echo canceller with strongly cross-correlated input signals — Analytical Results —,” Proc. of 6th DSP Symposium, pp. 144-149, November 1991.

[4] M. M. Sondhi and D. R. Morgan, “Acoustic echo cancellation for stereophonic teleconferencing,” presented at the 1991 IEEE ASSP Workshop Appl. Singal Processing Audio Acoustics, News Paltz, NY, Oct. 4-8, 1991.

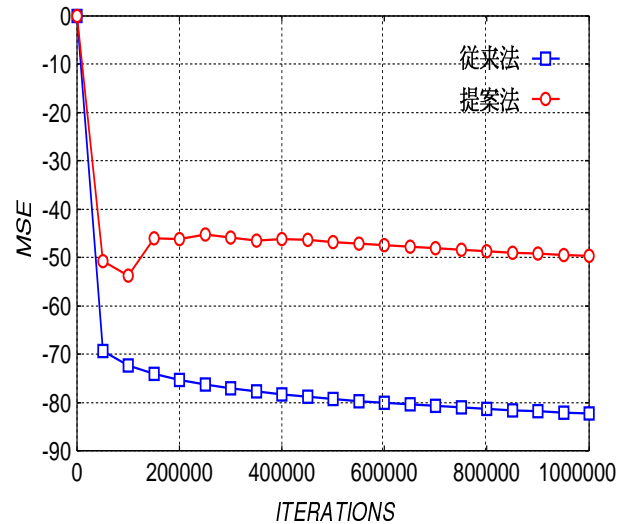


図4: 出力誤差

[5] A. Hirano and A. Sugiyama, “A compact multi-channel echo canceller with a single adaptive filter per channel,” Proc. of ISCAS '92, pp. 1922-1925, 1992.

[6] M. M. Sondhi and D. R. Morgan, “Stereophonic acoustic echo cancellation — an overview of the fundamental problem,” IEEE SP Letters, vol. 2, no. 8, pp. 148-151, Augst 1995.

[7] 杉山昭彦, “マルチチャネルエコーキャンセラにおける課題と研究動向,” 応用信号処理研究分科会第15回研究会資料, pp. 1-8, 1996

[8] A. Hirano, “Convergence analysis of a multi-channel acoustic echo canceller,” Proceedings of 12th DSP Symposium, pp. 521-526, Nov 1997

[9] Y. Joncour and A. Sugiyama, “A stereo echo canceler with pre-processing for correct echo path identification,” Proc. of ICASSP '98, pp. 3677-3680, May 1998.

[10] 田中正彦, 中山謙二, 平野 晃宏, “ステレオエコーキャンセラにおける解の一意性解析” Proceedings of 14th DSP Symposium, pp. 31-36, Nov 1999

[11] S. Haykin, “Adaptive Filter Theory,” Prentice Hall, 1996