

ステレオエコーキャンセラにおける解の一意性解析

On Uniqueness of Filter Coefficients
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: tanaka@leo.ec.t.kanazawa-u.ac.jp

あらまし

遠端側で話者一人が発言している場合におけるステレオエコーキャンセラのフィルタ係数の一意性について検討する。音声はMAモデルによって生成され、部屋の音響特性がFIRフィルタでモデル化される場合に、エコーを消去できる条件を示す。これより、フィルタ係数の解が一意に定まる可能性があることを示し、解が一意に定まる条件を明確にする。また、解が一意に定まる場合と定まらない場合の具体例を示す。計算機シミュレーションにより、解析の妥当性を示す。

ABSTRACT

This paper investigates uniqueness of filter coefficients in a stereophonic acoustic echo canceller for a single-talker case. Assuming that a talker speech is modeled by an MA model and that room acoustic characteristics are modeled by FIR filters, the condition for echo cancellation is clarified. Analytical results show that the filter coefficients may have unique solutions even in a single-talker case. The condition for unique solution is derived. Examples for the unique and non-unique cases are also shown. Computer simulation results validate the analysis.

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]-[10]。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]。

しかし、フィルタ係数の解の分布はごく簡単な場合しか解析されていない[3]。

本論文では、単一話者時におけるステレオエコーキャンセラの解の一意性を解析する。音声はMAモデルによって生成され、部屋の音響特性がFIRフィルタでモデル化される場合に、エコーを消去できる条件を示す。これより、フィルタ係数の解が一意に定まる可能性があることを示し、解が一意に定まる条件を明確にする。また、解が一意に定まる場合と定まらない場合の具体例を示す。計算機シミュレーションによって、解析の妥当性を確認する。

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

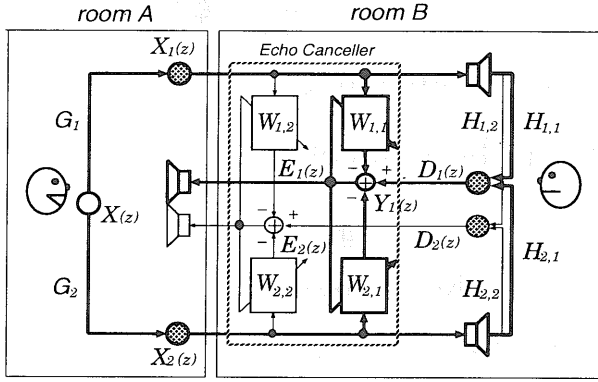


図 1: ステレオエコーキャンセラの構成

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除去される最適解は、

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

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]。

3 解の一意性解析

3.1 仮定

話者の音声は L 次の MA モデルによって生成されるものとする。AR モデルに対しては、十分次数の高い MA モデルで近似する。部屋 A および部屋 B の伝達特性は各々 M 次および N 次の FIR フィルタで表現できるものとする。適応フィルタのタップ数はエコーパスのインパルス応答長 N と等しいものとする。エコーパスのインパルス応答長の方が短い場合には、エコーパスのインパルス応答の後半は0であるとみなせば良い。適応フィルタのタップ数が不足する場合には問題がさらに複雑になるため、本稿では取り扱わないものとする。

3.2 誤差信号

時刻 n における MA モデルへの入力信号を $v(n)$ 、MA モデルの係数ベクトルを \mathbf{b} とすると、話者の音声 $x(n)$ は、

$$x(n) = \mathbf{b}^T \mathbf{v}(n) \quad (8)$$

で生成される。 \mathbf{b}^T および $\mathbf{v}(n)$ は、

$$\mathbf{b} = [b_0, b_1, \dots, b_{L-1}]^T \quad (9)$$

$$\mathbf{v}(n) = [v(n), \dots, v(n-L+1)]^T \quad (10)$$

である。上付文字 T は行列の転置を示す。

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 \mathbf{g}_i とすると、第 i チャンネルの参照入力信号 $\mathbf{x}_i(n)$ は、

$$\mathbf{x}_i(n) = \mathbf{g}_i^T \mathbf{x}(n) \quad (11)$$

$$\mathbf{g}_i = [g_{i,0}, g_{i,1}, \dots, g_{i,M-1}]^T \quad (12)$$

$$\mathbf{x}(n) = [x(n), \dots, x(n-M+1)]^T \quad (13)$$

となる。第 i 番目のスピーカから第 j 番目のマイク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 $\mathbf{h}_{i,j}$, $\mathbf{w}_{i,j}(n)$ とすると、エコー $d_j(n)$ および疑似エコー $y_j(n)$ は、

$$d_j(n) = \mathbf{h}_{1,j}^T \mathbf{x}_1(n) + \mathbf{h}_{2,j}^T \mathbf{x}_2(n) \quad (14)$$

$$y_j(n) = \mathbf{w}_{1,j}^T(n) \mathbf{x}_1(n) + \mathbf{w}_{2,j}^T(n) \mathbf{x}_2(n) \quad (15)$$

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

$$\mathbf{h}_{i,j} = [h_{i,j,0}, \dots, h_{i,j,N-1}]^T \quad (16)$$

$$\mathbf{w}_{i,j}(n) = [w_{i,j,0}(n), \dots, w_{i,j,N-1}(n)]^T \quad (17)$$

$$\mathbf{x}_i(n) = [x_i(n), \dots, x_i(n-M+1)] \quad (18)$$

で定義される。これより、誤差信号 $e_j(n)$ は、

$$e_j(n) = (\mathbf{h}_{1,j} - \mathbf{w}_{1,j}(n))^T \mathbf{x}_1(n) + (\mathbf{h}_{2,j} - \mathbf{w}_{2,j}(n))^T \mathbf{x}_2(n) \quad (19)$$

で得られる。ここで、

$$\begin{aligned} \mathbf{a}_{i,j}^T &= \mathbf{h}_{i,j}^T - \mathbf{w}_{i,j}^T \\ &= [a_{i,j,0}, a_{i,j,1}, \dots, a_{i,j,N-1}] \end{aligned} \quad (20)$$

とにおいて \mathbf{x}_j を \mathbf{b} , $\mathbf{v}(n)$, \mathbf{g} を用いて整理すると

$$\begin{aligned} e_j(n) &= \mathbf{a}_{1,j}^T \mathbf{x}_1(n) + \mathbf{a}_{2,j}^T \mathbf{x}_2(n) \\ &= (\mathbf{a}_{1,j}^T \mathbf{G}_1 + \mathbf{a}_{2,j}^T \mathbf{G}_2) \mathbf{B} \mathbf{V}(n) \\ &= [\mathbf{a}_{1,j}^T \quad \mathbf{a}_{2,j}^T] \begin{bmatrix} \mathbf{G}_1 \\ \mathbf{G}_2 \end{bmatrix} \mathbf{B} \mathbf{V}(n) \\ &= [\mathbf{a}_{1,j}^T \quad \mathbf{a}_{2,j}^T] \mathbf{G} \mathbf{B} \mathbf{V} \end{aligned} \quad (21)$$

となる。ここで、 $\mathbf{V}(n)$ は

$$\mathbf{V}(n) = [v(n) \cdots v(n-N-M-L+2)]^T \quad (22)$$

であり、 \mathbf{G}_i および \mathbf{B} は式 (23), (24) のように定義される。

3.3 エコーパス同定条件

(21) 式より任意の入力信号 $\mathbf{V}(n)$ に対してエコーパスを同定できる条件は

$$[\mathbf{a}_{1,j}^T \quad \mathbf{a}_{2,j}^T] \mathbf{G} \mathbf{B} = \mathbf{0} \quad (25)$$

となる。ここで、一般性を失うことなく $b_0 \neq 0$ と仮定できるので、 \mathbf{B} の一般化逆行列 \mathbf{B}^+ が存在することは容易に確かめられる。したがって、(25) 式の右側から \mathbf{B}^+ を乗ずることにより、

$$[\mathbf{a}_{1,j}^T \quad \mathbf{a}_{2,j}^T] \mathbf{G} = \mathbf{0} \quad (26)$$

を得る。これは、エコーを消去できるフィルタ係数の条件は、遠端話者の音声 $x(n)$ が白色であるか有色であるかには依存しないことを示している。

(26) 式の行列のサイズを調べると、 $2N$ 個の未知数 $\mathbf{w}_{i,j}(n)$ を含む $N+M-1$ 個の連立方程式であることがわかる。これより、エコーを除去できるフィルタ係数が一意に定まる必要条件として、

$$M-1 \geq N \quad (27)$$

が導出される。 M , N は整数であるから、(27) 式は

$$M > N \quad (28)$$

と等価である。これは、部屋 A の残響時間が部屋 B よりも長いという条件である。ここで、ステレオ音声会議においては、フィルタ係数が一意に定まる条件は片方の部屋に対してのみ成立し、反対側の部屋では成立しないことに注意する必要がある。 $M > N$ であれば、部屋 B で発生したエコーを消去するエコーキャンセラの係数は一意に定まるが、部屋 A のエコーキャンセラの係数は一意には定まらない。

方程式の数が未知数よりも多い場合、すなわち $M > N$ である場合、エコーを除去できるフィルタ係数が一意に定まる必要十分条件は、行列 \mathbf{G} の一般化逆行列が存在することである。これより、フィルタ係数が一意に定まる必要十分条件

$$\text{rank}(\mathbf{G}) = 2N \quad (29)$$

が導出される。

4 具体例

ここでは、いくつかの具体的な例を挙げて、フィルタ係数の一意性を検討する。これにより、フィルタ係数が一意に定まる場合が存在することを明らかにする。

4.1 g_2 が g_1 の定数倍である場合

g_2 が g_1 の定数倍である場合を考える。行列 \mathbf{G} は

$$\mathbf{G} = \begin{bmatrix} \mathbf{G}_1 \\ \alpha \mathbf{G}_1 \end{bmatrix} \quad (30)$$

となる。また、周波数領域においては $G_2(z) = \alpha G_1(z)$ となる。明らかに、行列 \mathbf{G} の階数は N である。したがって、解は一意には定まらない。

$$G_i = \left[\begin{array}{cccccccc} g_{i,0} & g_{i,1} & \cdots & g_{i,M-1} & 0 & \cdots & \cdots & 0 \\ 0 & g_{i,0} & g_{i,1} & \cdots & g_{i,M-1} & \ddots & \ddots & \vdots \\ \vdots & 0 & \ddots & \ddots & \cdots & \ddots & \ddots & \vdots \\ \vdots & \vdots & \ddots & \ddots & \cdots & \ddots & \ddots & 0 \\ 0 & \cdots & \cdots & 0 & g_{i,0} & g_{i,1} & \cdots & g_{i,M-1} \end{array} \right] \left. \vphantom{\begin{array}{c} \\ \\ \\ \\ \end{array}} \right\} N \quad (23)$$

$\underbrace{\hspace{15em}}_{N+L+M-2}$

$$B = \left[\begin{array}{cccccccc} b_0 & b_1 & \cdots & b_{L-1} & 0 & \cdots & \cdots & 0 \\ 0 & b_0 & b_1 & \cdots & b_{L-1} & \ddots & \ddots & \vdots \\ \vdots & 0 & \ddots & \ddots & \cdots & \ddots & \ddots & \vdots \\ \vdots & \vdots & \ddots & \ddots & \cdots & \ddots & \ddots & 0 \\ 0 & \cdots & \cdots & 0 & b_0 & b_1 & \cdots & b_{L-1} \end{array} \right] \left. \vphantom{\begin{array}{c} \\ \\ \\ \\ \end{array}} \right\} N+M-1 \quad (24)$$

$\underbrace{\hspace{15em}}_{N+L+M-2}$

4.2 1要素以外 g_2 が g_1 の定数倍である場合

g_2 が g_1 の定数倍で、さらに第1要素に定数を加えたものである場合、すなわち、

$$g_2 = [\alpha g_{1,0} + \beta \quad \alpha g_{1,1} \quad \cdots \quad \alpha g_{1,M-1}]^T \quad (31)$$

である場合を考える。明らかにこの場合の $\text{rank}(G)$ は、

$$g_2 = [\beta \quad 0 \quad \cdots \quad 0]^T \quad (32)$$

とした場合と一致する。したがって、 $M > N$ であり、かつ、 $g_{1,i}$ ($i = N, \dots, M-1$) のうち1個でも0でないものが存在すれば、フィルタ係数は一意に定まることがわかる。例えば、 $M = N+1$ である場合、

$$\text{rank}(G) = \text{rank} \left(\begin{bmatrix} O & g_{1,N} I \\ \beta I & O \end{bmatrix} \right) \quad (33)$$

となる。ここで、 I は単位行列である。これより、フィルタ係数が一意に定まる条件は、

$$\beta \neq 0 \quad (34)$$

$$g_{1,N} \neq 0 \quad (35)$$

となる。周波数領域においても、

$$G_2(z) = \alpha G_1(z) + \beta \quad (36)$$

より、同様の結果が得られる。

5 計算機シミュレーション

以上の解析結果を検証するために計算機シミュレーションを行った。評価尺度として、正規化係数誤差ノルム NCEN (Normalized Coefficient Error Norm)

$$NCEN(n) = \frac{\sum_{i=1}^2 \|w_{i,j}(n) - h_{i,j}\|^2}{\sum_{i=1}^2 \|h_{i,j}\|^2} \quad (37)$$

を用いた。適応アルゴリズムには NLMS アルゴリズムを用いた。

まず、入力信号 $x(n)$ が白色ガウス雑音である場合のシミュレーションを行った。フィルタ係数が一意に定まらない場合として g_i のタップ数 $M = 4$ 、エコーパスと適応フィルタのタップ数 $N = 4$ を、一意に定まる場合として $M = 5$ 、 $N = 4$ を選んだ。 g_i 、 $h_{i,j}$ の値としては互いに独立な乱数とし、乱数の種を変化させていくつかのシミュレーションを行った。 $M = 5$ 、 $N = 4$ である場合には、乱数で作成した g_i の全てが (29) 式を満たすものであった。正規化係数誤差ノルム $NCEN(n)$ の収束特性を図2に、フィルタ係数の収束値を図3に示す。

解が一意に定まらない場合には、いずれの結果もフィルタ係数が最適解であるエコーパスのインパルス応答に収束したものはなかった。一方、解が定まる場合には、すべての誤差が十分小さな値となっているので最適解に収束したと判断できる。しかし、 g_i の値によって学習速度は大きく異なっていた。

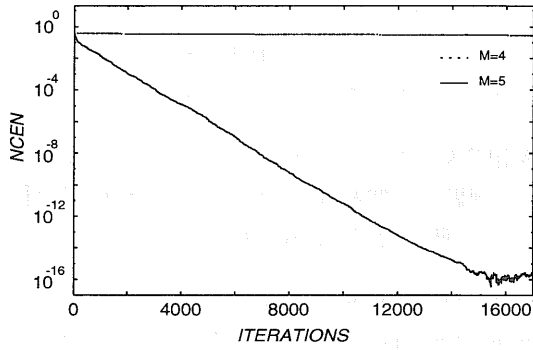


図 2: 白色雑音入力時の正規化係数誤差ノルム

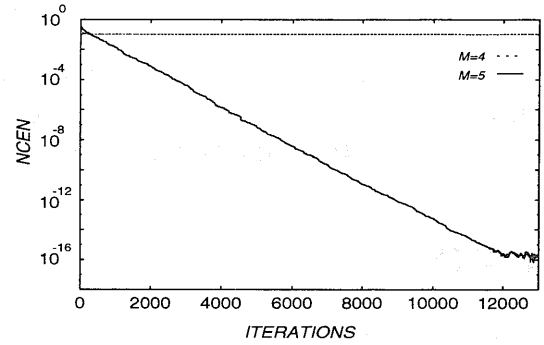


図 4: 有色雑音入力時の正規化係数誤差ノルム

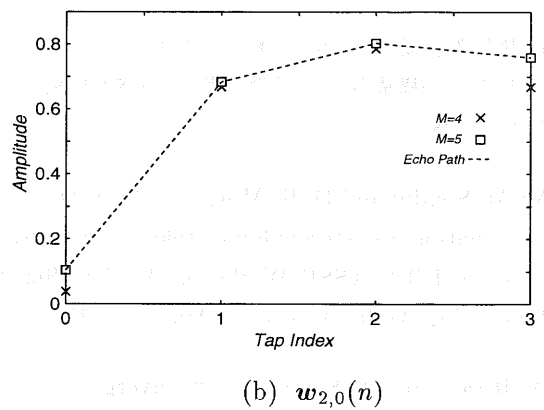
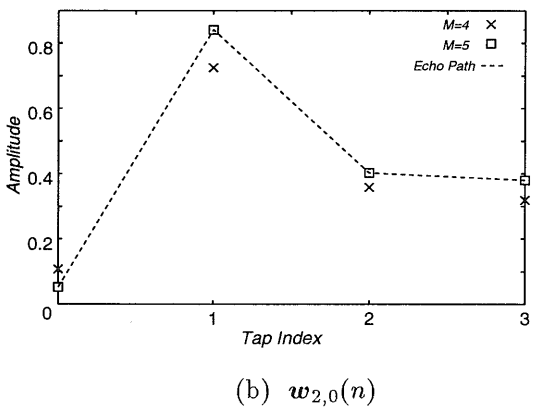
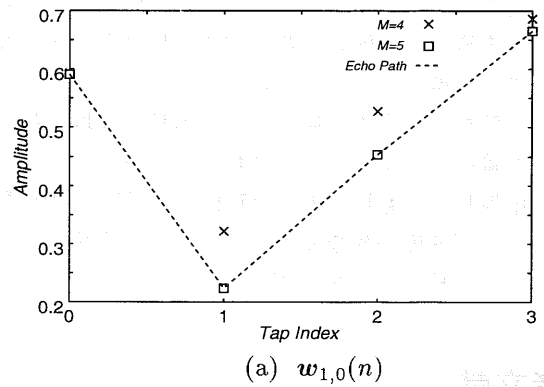
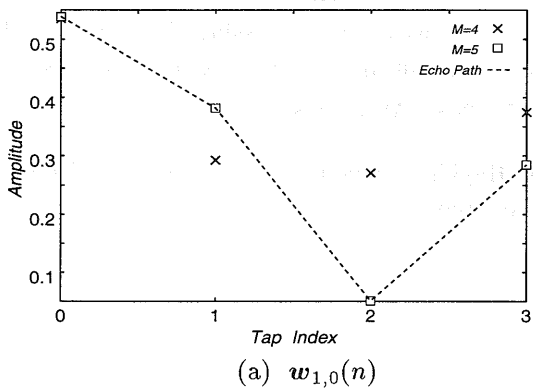


図 3: 白色雑音入力時のフィルタ係数収束値

図 5: 有色雑音入力時のフィルタ係数収束値

次に、入力信号がMAモデルで生成された有色信号である場合についてもシミュレーションを行った。移動平均フィルタのタップ数は $L=5$ とした。 M と N の値、および $g_i, w_{i,j}$ は白色の場合と同じであり、MA係数 b の値として独立な乱数を使用した。正規化係数誤差ノルム $NCEN(n)$ の収束特性を図4に、フィルタ係数の収束値を図5に示す。

解が一意に定まらない場合では、フィルタ係数が最適解に収束したものは全くなかった。解が一意に定まる場合には、すべての結果が最適解に収束した。

6 むすび

単一話者時におけるステレオエコーキャンセラの解の一意性を解析した。遠端側の残響時間が近端側より長い場合にはエコーを除去できるフィルタ係数の解が一意に定まる可能性があることを示し、解が一意に定まる条件を導いた。解が一意に定まる場合と定まらない場合の具体例を示した。計算機シミュレーションによって、解析の妥当性を確認した。本稿では取り扱わなかった適応フィルタのタップ数がエコーパスのインパルス応答長よりも短い場合の解析や、行列 G の性質によって学習速度が異なること、実空間の音響特性を用いるなどの検討は今後の課題である。

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 Appls. Singal Processing Audio Acoustics, News Paltz, NY, Oct. 4-8, 1991.
- [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] S. Haykin, “Adaptive Filter Theory,” Prentice Hall, 1996