

A - 11 ステレオエコーキャンセラの 解の一意性解析と学習法

On Uniqueness of Filter Coefficients and the Learning Method
in Stereophonic Acoustic Echo Canceller

電子情報システム専攻 田中 正彦

あらまし

現在、ステレオエコーキャンセラでは前処理を用いる学習が有効な結果を示している。しかし、音質劣化を招くなどの問題点を抱えている。そこで本研究では、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最適解以外の解(不定解)が無限に存在する係数不確定性問題がある。これに対してまず、一意解の存在条件 [1][2] を示す。この条件は2つのエコーキャンセラに対して同時に成立しないので、適応フィルタの係数の分割により、この問題を解決する学習法を提案する。計算機実験を行った結果、係数誤差が低減し、不定解を避けられることを示す。

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。ステレオや多チャンネルエコーキャンセラはいまだ開発途上であり、今後発展が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。これを解決する方法として前処理を用いる学習が有効な結果を示しているが、時変のフィルタを挿入するため音質劣化を招くなどの問題点がある。

そこで、前処理を用いない学習法について検討する。まず、係数不確定性問題に対して一意解が存在する条件を示す。しかし、この条件は片方向でしか成立しないため双方向のエコーキャンセラを同時に学習できない。本論文では、双方向の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システムを示す。

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それぞれ同定し、エコーを除去する。

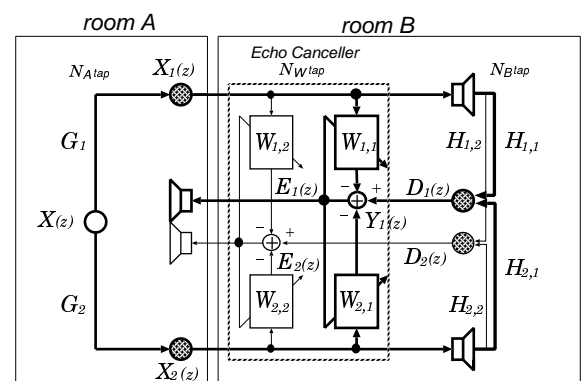


図 1: ステレオエコーキャンセラの構成

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。

2.3 係数の一意解の存在条件

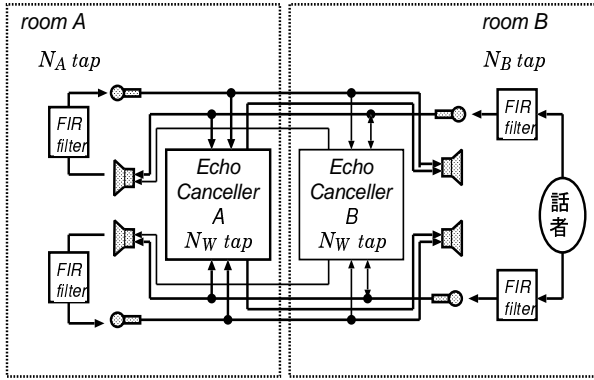


図 2: ステレオエコーキャンセラの構成

図 2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。係数不確実性問題に対して、 $N_A > N_W \geq N_B$ (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り、最適解に収束する。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、双方向のエコーキャンセラを学習できない。双方向について条件を成立させるために N_W を N_A, N_B より短くすることを考える。単に短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、エコー消去機能の劣化を招いてしまう。これを解決する一つの方法として、本論文では、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、

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

$$y_j(n) = \sum_{i=1}^2 w_{i,j}^T(n) x_{i,f}(n) \quad (12)$$

で与えられる。 $w_{i,j,f}(n)$, $h_{i,j,f}$, $h_{i,j,b}$, $w_{i,j,b}(n)$, $x_{i,f}(n)$, $x_{i,b}(n)$ は、

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

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

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

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

$$x_{i,b}(n) = [x_i(n - N_W), \dots, x_i(n - N_B + 1)] \quad (17)$$

で定義される。これより、誤差信号 $e_j(n)$ は、

$$e_j(n) = \sum_{i=1}^2 \{h_{i,j,f} - w_{i,j}(n)\}^T x_{i,f}(n) + \sum_{i=1}^2 h_{i,j,b}^T x_{i,b}(n - N_W) \quad (18)$$

で得られる。ここで x_i を式 (8) から

$$e_j(n) = \sum_{i=1}^2 \{h_{i,j,f} - w_{i,j}(n)\}^T G_{i,f} s_f(n) + \sum_{i=1}^2 h_{i,j,b}^T G_{i,b} s_b(n - N_W) \quad (19)$$

$s_f(n), s_b(n - N_W)$ は

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

$$s_{i,b}(n - N_W) = [s_i(n - N_W), \dots, s_i(n - N_B + 1)] \quad (21)$$

であり、 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、

$$e_j(n) = \begin{bmatrix} h_{1,j,f}^T - w_{1,j}^T(n) \\ h_{2,j,f}^T - w_{2,j}^T(n) \end{bmatrix} \begin{bmatrix} G_{1,f} \\ G_{2,f} \end{bmatrix} s_f(n) + \begin{bmatrix} h_{1,j,b}^T \\ h_{2,j,b}^T \end{bmatrix} \begin{bmatrix} G_{1,b} \\ G_{2,b} \end{bmatrix} s_b(n - N_W) = D_f^T(n) G_f s_f(n) + H_b^T G_b s_b(n - N_W) \quad (23)$$

$$D_f(n) = \begin{bmatrix} h_{1,j,f} - w_{1,j}(n) \\ h_{2,j,f} - w_{2,j}(n) \end{bmatrix}$$

$$H_b = \begin{bmatrix} h_{1,j,b} \\ h_{2,j,b} \end{bmatrix}$$

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

$N_W + N_A - 1$

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

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

これより $e_j(n)$ の二乗平均誤差を求めると

$$\begin{aligned} E[e_j^2(n)] &= E[e_j(n)e_j^T(n)] \\ &= E[\mathbf{D}_f^T(n)\mathbf{G}_f\mathbf{s}_f(n)\mathbf{s}_f^T(n)\mathbf{G}_f^T\mathbf{D}_f(n) \\ &\quad + \mathbf{H}_b^T\mathbf{G}_b\mathbf{s}_b(n - N_W)\mathbf{s}_b^T(n - N_W)\mathbf{G}_b^T\mathbf{H}_b \\ &\quad + 2\mathbf{D}_f^T(n)\mathbf{G}_f\mathbf{s}_f(n)\mathbf{s}_b^T(n - N_W)\mathbf{G}_b^T\mathbf{H}_b] \\ &= \overline{\mathbf{D}_f^T(n)\mathbf{G}_f\mathbf{R}_f\mathbf{G}_f^T\mathbf{D}_f(n)} \\ &\quad + \mathbf{H}_b^T\mathbf{G}_b\mathbf{R}_b\mathbf{G}_b^T\mathbf{H}_b \\ &\quad + 2\overline{\mathbf{D}_f^T(n)\mathbf{G}_f\mathbf{R}_{fb}\mathbf{G}_b^T\mathbf{H}_b} \\ &= \overline{\mathbf{D}_f^T(n)\mathbf{Q}_f\mathbf{D}_f(n)} + \mathbf{H}_b^T\mathbf{Q}_b\mathbf{H}_b \\ &\quad + 2\overline{\mathbf{D}_f^T(n)\mathbf{Q}_{fb}\mathbf{H}_b} \end{aligned} \quad (24)$$

ただし,

$$\begin{aligned} \mathbf{R}_f &= \overline{\mathbf{s}_f(n)\mathbf{s}_f^T(n)}, \quad \mathbf{R}_{fb} = \overline{\mathbf{s}_f(n)\mathbf{s}_b^T(n - N_W)}, \\ \mathbf{R}_b &= \overline{\mathbf{s}_b(n - N_W)\mathbf{s}_b^T(n - N_W)}, \\ \mathbf{Q}_f &= \mathbf{G}_f\mathbf{R}_f\mathbf{G}_f^T, \quad \mathbf{Q}_{fb} = \mathbf{G}_f\mathbf{R}_{fb}\mathbf{G}_b^T, \quad \mathbf{Q}_b = \mathbf{G}_b\mathbf{R}_b\mathbf{G}_b^T \end{aligned}$$

で \mathbf{Q}_f は正方行列である。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 $\frac{\partial E[e_j^2(n)]}{\partial \mathbf{D}_f(n)} = 0$ より

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

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

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 学習方法と収束性

以下に分割学習の手順を示す。

- エコーパスの反響モデルの長さ以上の適応フィルタ係数を用意
- 適応フィルタの係数列 (N_W タップ) を前後に 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

- 一定の時間間隔において前半部分の学習を止め、後半部分の学習
このとき、前半部分のフィルタ係数の収束値を固定して使用

- 前半部分と後半部分を切替えて学習を繰り返す
この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 より短くできる。しかし、実際は適応フィルタのタップ長はエコーパスの長さをカバーしており、解が一意に定まり、かつ最適解が求まる条件を満たす事が可能となる。

ここで、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、式 (28) のようにフィルタ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。学習の切替を行い、後半部分の学習の場合は、係数は最適値に前半部分の誤差を加えた式 (29) の値に収束する。

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

ただし、 $\mathbf{K}_b = \mathbf{Q}_b^{-1}\mathbf{Q}_{fb}^T$ である。そして、二回目の前半学習では、

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

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。

式 (28) から式 (30) までと同様に、学習の切替え回数を n としたときの式を数学的帰納法に基づいて計算すると、

$$\mathbf{W}_f^{(n)} = \mathbf{H}_f + \mathbf{K}_f(\mathbf{K}_b\mathbf{K}_f)^{(n-1)}\mathbf{H}_b \quad (31)$$

$$\mathbf{W}_b^{(n)} = (\mathbf{I} - (\mathbf{K}_b\mathbf{K}_f)^{(n)})\mathbf{H}_b \quad (32)$$

と表される。ここで、 \mathbf{I} は単位行列である。式 (31)、式 (32) において最適解が求まるためには、

$$\lim_{n \rightarrow \infty} (\mathbf{K}_b\mathbf{K}_f)^{(n)} = 0 \quad (33)$$

となる必要がある。つまり、 $\mathbf{K}_b\mathbf{K}_f$ の最大固有値を λ_{max} とすると、

$$|\lambda_{max}| < 1 \quad (34)$$

であれば学習が収束する。

4 実験結果

従来法と提案法との比較を計算機実験により行う。入力信号を白色ガウス雑音とする。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10万サンプル毎とし、切替点までのフィルタ係数の10000サンプルの時間平均をとることで学習を安定させることにする。又、 $N_W = N_A = N_B = 12 \text{ tap}$ とし、 g_i 、 $h_{i,j}$ の特性としては互いに独立な乱数とする。 $K_b K_f$ の最大固有値は0.250であった。入力信号の種、及びフィルタ係数(前半部分)の初期値を変化させて50の集合平均をとったものを結果とする。

結果には出力誤差、フィルタ係数の軌跡、式(35)のような正規化係数誤差ノルム NCEV(Normalized Coefficient Error Vector)

$$NCEV(n) = \frac{\sum_{i=1}^2 \|w_{i,j}(n) - h_{i,j}\|^2}{\sum_{i=1}^2 \|h_{i,j}\|^2} \quad (35)$$

による比較を行い、それぞれを図3、図4、図5に示した。

図3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4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、提案法では係数の前後切替を繰り返しながら、最適解に近付いている様子が分かる。図5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。これより提案法が一意解ではあるが最適値以外の解に収束していることが判断できる。

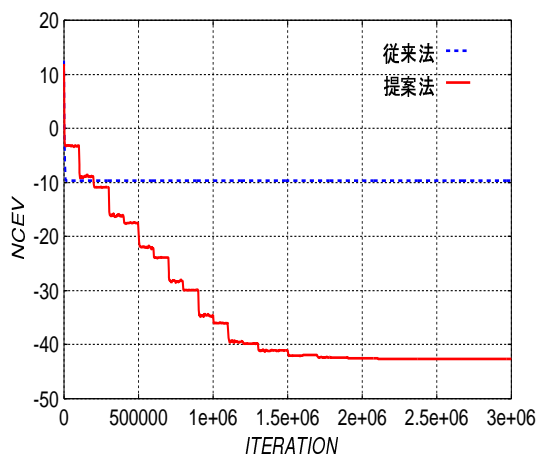


図3: 係数誤差

5 むすび

線形結合形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。

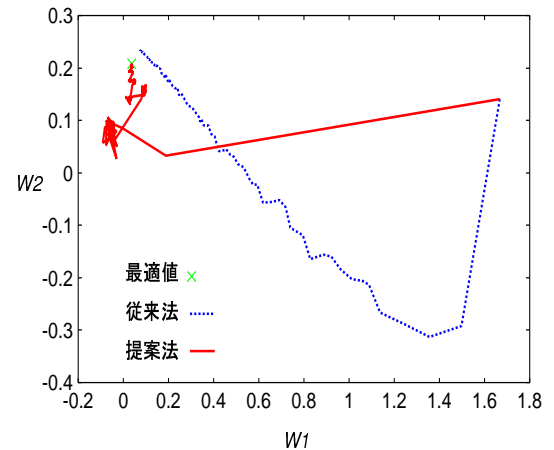


図4: 係数の軌跡

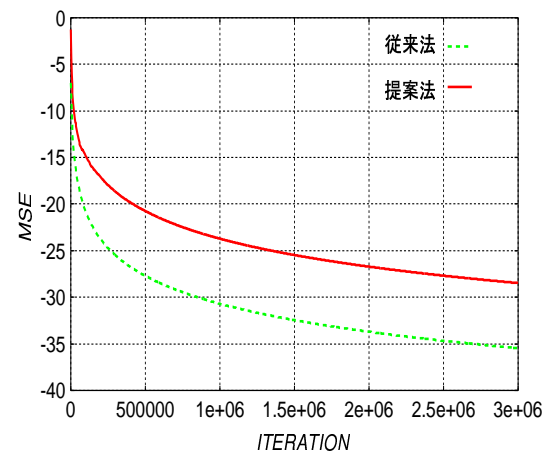


図5: 出力誤差

従来法に比べて収束速度は遅いが係数誤差は低減できた。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確認できた。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少しずれた値に収束してしまうことが分かった。本方法では音質の劣化を招く前処理を用いないで一意解に収束する事が可能であることを示した。

6 口頭論文

[1] 田中正彦, 中山謙二, 平野晃宏: “ステレオエコーキャンセラの解の一意性解析”
平成11年度電気関係学会北陸支部連合大会

[2] 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, Nov 1999

[3] 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適応フィルタの分割学習”, 平成12年度電気関係学会北陸支部連合大会

[4] 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前処理を用いないエコーパス同定”
Proceedings of 15th DSP Symposium, Nov 2000