

エコーと近端音声の相互相関に基づく 可変ステップサイズ音響エコーキャンセラ A Variable Step size Acoustic Echo Canceler Based on Cross-correlation between Echo and Near-end Signal

牛丸 晋一郎 中山 謙二 平野 晃宏
Shinichiro Ushimaru Kenji Nakayama Akihiro Hirano

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: ushimaru@leo.ec.t.kanazawa-u.ac.jp

あらまし

近端音声に遠端音声のエコーが加わった信号からエコーを消去する方法として適応フィルタを用いた音響エコーキャンセラ [1] がある。双方向通話時(ダブルトーク)において、遠端と近端音声の相関性が強い場合に適応フィルタの係数学習が不安定になる。本稿では、遠端音声のエコーと近端音声の相互相関を近似的に導出する方法 [8] に基づいて、相互相関の新しい利用法を提案し可変ステップサイズに応用している。音声信号を用いて計算機シミュレーションによりその有効性を確認する。

ABSTRACT

In an acoustic echo canceler, learning coefficients during double talk causes unstable behavior. Especially, when cross-correlation between the echo and the near-end signal is high, the adaptation is hard to converge. The far-end and near-end signals are usually both speech signals, therefore, the cross-correlation is high. In this paper, a variable step size adaptive filter is proposed. The method is based on an indirect estimation of cross-correlation between the echo and near-end signal[8]. The proposed method is evaluated through computer simulation using the far-end and near-end voice signals. Compared with the fixed step size method, the proposed method can guarantee good convergence.

1 まえがき

TV 会議などの音声伝達システムにおいて回線に遠端音声のエコーが含まれる。このエコーは回線を単独で伝わる場合(シングルトーク)と近端音声と混在して伝わる場合(双方向通話:ダブルトーク)がある。シングルトーク時におけるエコー除去は適応フィルタを用いて比較的容易に行われる。しかしダブルトーク時においては、混在するエコーと他の音声の相互相関により学習の収束性が悪くなる。すなわち、他の音声と混在しているため、エコーのみの特性をとらえることが難しくなる [2]-[4]。相互相関が強い場合にはステップサイズを小さくすることにより収束特性を改善することができる。しかし、収束速度と残留誤差におけるトレードオフが問題となる [5]-[7]。

本研究では、遠端音声のエコーと近端音声の相互相関を近似的に導出する方法 [8] を基に、相互相関の新たな利用法を提案し、可変ステップサイズに応用している。この方式により相関の変化に素早く追従し、安定な学習が可能となる。計算機シミュレーションにより、その有効性を示す。

2 音響エコーキャンセラの基本構成

音響エコーキャンセラの一例として、遠隔会議システムにおける音響エコーキャンセラを図 1 に示す。この方式は、室内音響(エコーパス)をシミュレートする適応フィルタ AF (Adaptive Filter) を挿入することにより、遠端音声を通して疑似エコーを生成し、減算によって実際のエコーを除去することを目的として

いる。

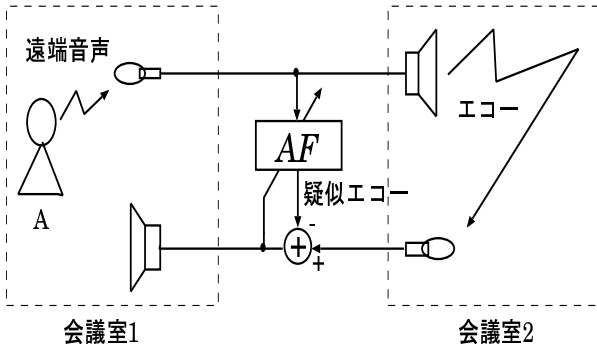


図 1: 遠隔会議システムにおける音響エコーキャンセラ (シングルトーク)

3 ダブルトーク時の収束特性

3.1 ダブルトーク問題

図 2 に上記と同様な遠隔会議システムにおけるダブルトークの様子を示す。

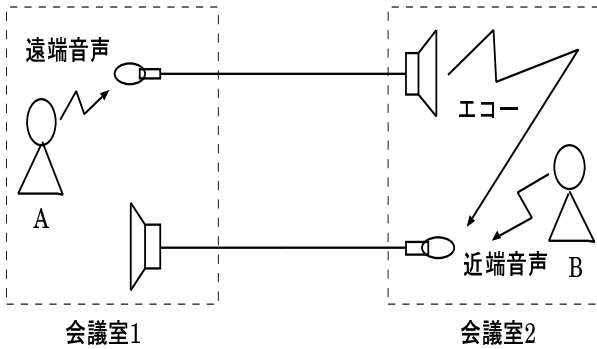


図 2: 遠隔会議システムにおけるダブルトーク

ダブルトークとは図に示すように遠端音声と近端音声と同時に送信される状態をいう。これは会議システムなどにおいては当然考えられ、避けられない状況である。

ダブルトーク時の問題は回線中にエコーと近端音声混在することである。このシステムにおいては近端音声のみ会議室 1 へ伝えるためにエコーだけを除去しなければならない。しかし、適応フィルタに対する参照信号がエコーと近端音声からなる混在信号なのでエコーパスの推定が困難となり、適応動作にも相応の考

慮が必要となってくる。これがダブルトークの問題点である。

3.2 相互相関と収束特性

図 3 にダブルトークを考慮した音響エコーキャンセラシステムを示す。

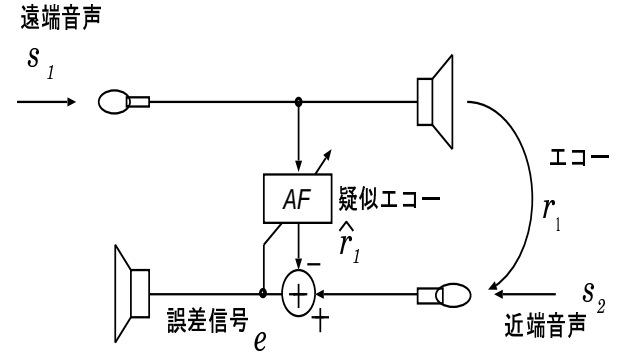


図 3: 音響エコーキャンセラシステム

遠端音声 s_1 が回線を通り、近端側のスピーカから出力される時エコー r_1 が発生する。このエコーを適応フィルタ AF によって生成された疑似エコー \hat{r}_1 により差し引きし除去する。この時、ある値を持った近端音声 s_2 が入力されるとダブルトークとなる。このシステムにおいて、ある時刻 n における誤差信号 $e(n)$ は以下の式で与えられる。

$$e(n) = r_1(n) + s_2(n) - \hat{r}_1(n)$$

適応フィルタは $e(n)$ の二乗平均 $E[e^2(n)]$ を最小にするように係数を学習する。 $E[e^2(n)]$ は次式のように求められる。

$$\begin{aligned} E[e^2(n)] &= E[(r_1(n) - \hat{r}_1(n))^2 + 2(r_1(n) - \hat{r}_1(n))s_2(n) + s_2^2(n)] \\ &= E[(r_1(n) - \hat{r}_1(n))^2] + 2E[(r_1(n) - \hat{r}_1(n))s_2(n)] \\ &\quad + E[s_2^2(n)] \end{aligned}$$

上式において右辺第 3 項は適応フィルタの特性と独立しているので一定とみなすことができる。右辺第 2 項に着目すると $r_1(n) - \hat{r}_1(n)$ と $s_2(n)$ との間に相関がなければ 0 となる。従って $E[e^2(n)]$ を最小にするには右辺第 1 項の最小化を考えればよい。つまりこれは適応フィルタの伝達関数がエコーパスの伝達関数に近づくことを意味している。しかし、 $r_1(n) - \hat{r}_1(n)$ と $s_2(n)$ との間に相関がある場合、第 2 項は 0 にはならない。この状態で $E[e^2(n)]$ の最小化を行なうと、エコーパスの推定が理想値からずれることは明らかであ

る。従って、ダブルトーク時では、遠端音声と近端音声の相互相関の度合に応じた学習法を考慮する必要がある。

4 従来の可変ステップサイズ

4.1 可変ステップサイズの利用

本稿では、勾配法を対象とし、具体的には以下の式で示される NLMS アルゴリズムを考える。

$$\mathbf{w}(n+1) = \mathbf{w}(n) + \mu e(n)\mathbf{x}(n)/\{\mathbf{x}^T(n)\mathbf{x}(n)\}$$

係数更新の過程で変化の度合を制御するのがステップサイズ μ である。この μ の設定が収束速度と安定性に大きく影響する。相互相関が強い場合においても小さいステップサイズを用い、十分な収束時間があればエコーを除去することが可能であるが、効率の良い学習は望めない。

そこで効率の良い学習を行うために、遠端と近端の音声の相互相関の強弱によってステップサイズを変化させる可変ステップサイズが提案されている [8]。これは、相互相関が弱い時はステップサイズを大きく、逆に強い時はステップサイズを小さくすることを目的としている。これにより、小さいステップサイズを終始用いる場合に比べ収束時間の短縮がはかれ、また相互相関の強いエコーも除去できる。

4.2 相互相関の導出

前項で述べた可変ステップサイズを用いるにあたって、信号間の相互相関の値を求める必要がある。遠端音声 s_1 と近端音声 s_2 の相互相関はそれらの積の2乗平均で示される。しかし、回線を伝わるのは近端音声 s_2 とエコー r_1 の混在信号なので s_2 のみを検出することはできない。従って直接この相互相関を求めることは不可能である。そこで以下のようにして相互相関を求めた。(図3参照)

まず、AFにより生成される疑似エコー \hat{r}_1 と混在信号 $r_1 + s_2$ の相互相関を求める。 E は平均を表す。

$$\alpha = E[\hat{r}_1(r_1 + s_2)] \quad (1)$$

次に、疑似エコー \hat{r}_1 の自己相関を求める。

$$\beta = E[\hat{r}_1^2] \quad (2)$$

ここで、

$$E[\hat{r}_1^2] \simeq E[\hat{r}_1 r_1] \quad (3)$$

と近似できると仮定する。式(1)、(2)、(3)より α と β の差 γ をとると近似的に疑似エコー \hat{r}_1 と近端音声 s_2 の

相互相関が得られる。

$$\begin{aligned} \gamma &= \alpha - \beta \\ &= E[\hat{r}_1 r_1] + E[\hat{r}_1 s_2] - E[\hat{r}_1^2] \\ &\simeq E[\hat{r}_1 s_2] \end{aligned} \quad (4)$$

適応動作が進み、疑似エコー \hat{r}_1 がエコー r_1 の特性に近づくにつれ、 γ の値はエコー r_1 と近端音声 s_2 の相互相関に近づく。本研究では、この γ を次式のように遠端音声で正規化した Γ の値を用いた。

$$\Gamma = \frac{\gamma}{E[s_1^2]} \quad (5)$$

4.3 従来法の問題点

従来は、式(6)のように Γ に反比例するステップサイズの更新式を提案し、ステップサイズが固定の場合よりも収束特性が改善されることが確かめられた [8]。しかし、相関の急激な変化に対して十分に追従できない為、相互相関の変化の大きい箇所では効率良く学習できないという問題がある。実際のシミュレーション結果の例を挙げて説明する。図4-aのような入力音声に対し、図4-bのような相関値が導出されると、式(6)に基づいて図4-cのような可変ステップサイズが得られる。しかし、ステップサイズの値が下がるまでに時間を要する為、その間に収束特性が悪化する(図4-d)。つまりこれは、音声間の相互相関の変化に十分に追従できていないことが原因と考えられる。この問題を解決する為に次項のような方法を提案する。

$$\begin{aligned} \mu &= \frac{K_1}{K_2|\Gamma|^\xi + \delta} \mu_0 \\ &(K_1, K_2, \xi, \delta, \mu_0 \text{は定数}) \end{aligned} \quad (6)$$

5 短時間平均と長時間平均の利用

前項の問題を解決する為に、式(5)で求めた相互相関 Γ の短時間平均と長時間平均を組み合わせた方法を提案する。この方法では、長時間平均の値を基本として用いるが、ある時点で短時間の平均値が長時間の平均値を上回った場合は、短時間平均の値をその時点での相関値とする。これにより、安定、且つ変化に素早く追従する相互相関の値を得ることができる。図5にこの方法の例を示す。提案する方法により、51250サンプル付近における音声間の相互相関の変化に追従できていることが確認できる。

求めたこの相関値に対して反比例するようにステップサイズを可変することで、相互相関が急激に増加した場合でも瞬時に学習を抑制し、また、多少の変化では影響を受けない安定した学習が可能となる。

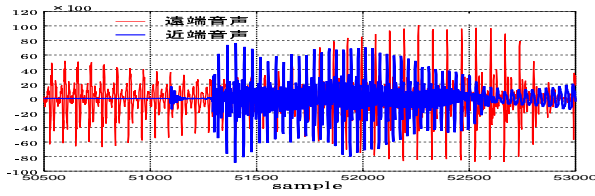


図 4-a: 入力音声

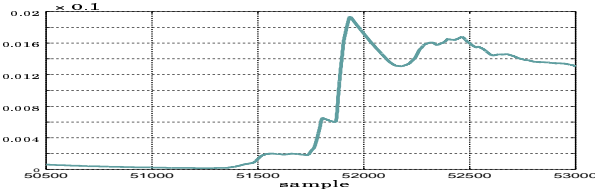


図 4-b: 相互相関

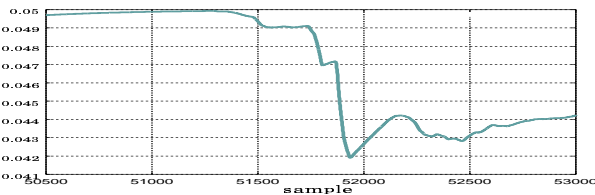


図 4-c: 可変ステップサイズ

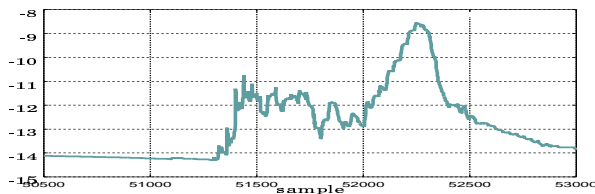


図 4-d: 係数誤差の学習曲線

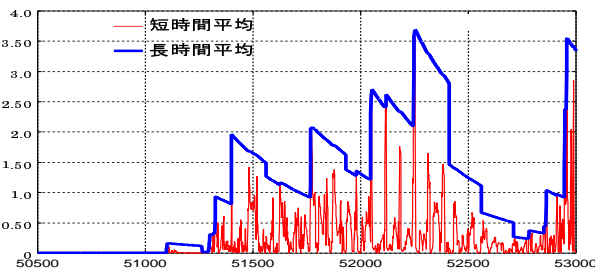


図 5: 短時間、長時間平均を組み合わせた相互相関

6 計算機シミュレーション

6.1 可変ステップサイズの導出

本稿では、上記で求めた相互相関に基づいて、3通りの方法で可変ステップサイズを導出した。まず方式1として閾値を設け、この閾値に対して相互相関の値が大きい場合はステップサイズを小さく、逆に相互相関の値が小さい場合はステップサイズを大きくする。

ステップサイズは大小2つの値のみとする。図6にこの方法の例を示す。

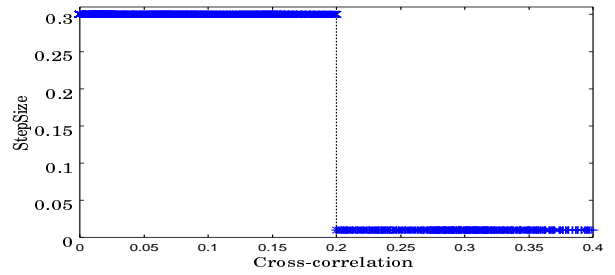


図 6: 方式1のステップサイズ

次に方式2として、相互相関の値に対してステップサイズが直線的に変化する方法を用いた。その様子を図7に示す。方式1のような2値のみの離散的な変化に対して、方式2は連続的に変化する。

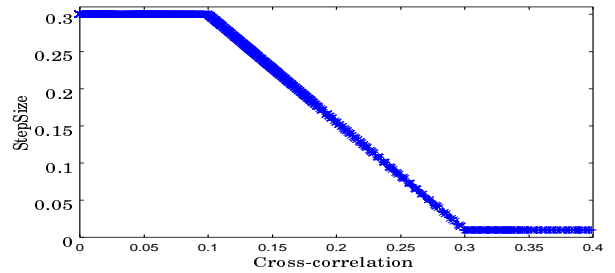


図 7: 方式2のステップサイズ

方式3として、相互相関の値に対してステップサイズが2次曲線的に変化する方法を用いた。その様子を図8に示す。

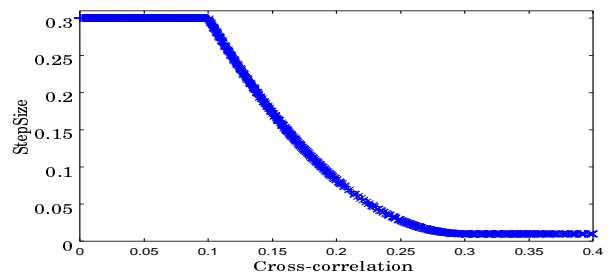


図 8: 方式3のステップサイズ

6.2 シミュレーション

提案する3通りの可変ステップサイズと固定ステップサイズ、及び従来法の場合に対して計算機シミュレーションを行い比較した。

遠端音声、近端音声には図9のような音声を用い、エコーパスにはバターワースフィルタ、適応フィルタにFIRフィルタを用い、そのタップ数を64とした。相互相関をとる範囲は、短時間の平均幅を約10サン

プル、長時間の平均幅を約 500 サンプルとした。また、可変ステップサイズの最大値を 0.3, 最小値を 0.01 とし、比較対象の固定ステップサイズの値は経験的に求めた最良の収束特性となる 0.03 とした。係数学習には NLMS を用いた。この条件の基で、エコーパスのインパルス応答 h と適応フィルタのフィルタ係数 w を比較し、式 (7) で示されるような係数誤差をノルムを用いて dB 評価した。但し、係数更新において、遠端音声の平均パワーが 1000 以上で学習を行うこととした。結果を図 10 に示す。

$$D = 10 \log_{10} \frac{\|h - w\|^2}{\|h\|^2} \quad (7)$$

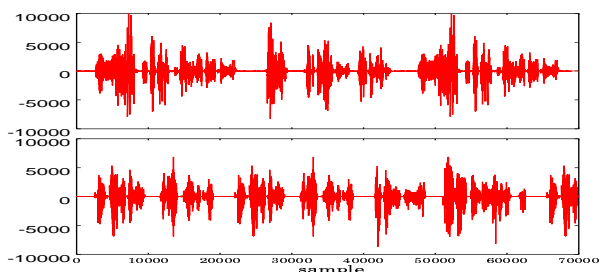


図 9: 遠端音声 (上), 近端音声 (下)

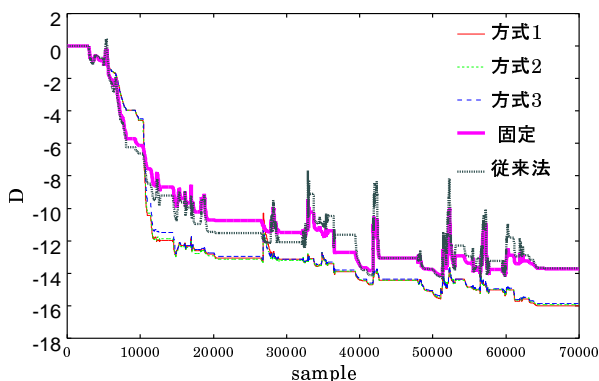


図 10: 係数誤差の学習曲線

固定ステップサイズ、及び従来法と比較して、3つの提案法はいずれも収束特性が改善されている。51000 サンプル付近において、従来法では相互相関の急激な変化に対応しきれず収束特性が悪化しているのに対し、提案法では素早く追従し悪化を最小限に抑えていることが確認できる。

また、3つの提案法の間では大きな違いは見られなかったが、閾値やパラメータによる変化を検討する余地がある。

7 むすび

音響エコーキャンセラにおいて、双方向通話時における収束特性の改善を目的として、遠端音声のエコーと近端音声の相互相関を近似的に導出する方法 [8] を基に、相互相関の新たな利用法を提案した。この相互相関を用いて 3通りの可変パターンをもつステップサイズによりシミュレーションを行った。その結果、短時間平均と長時間平均の相関値を組み合わせることにより、相関の変化に素早く追従し、安定した学習が可能となり、固定ステップサイズ、及び従来法と比べ良好な収束特性が得られることが確かめられた。

参考文献

- [1] Simon Haykin, "Adaptive Filter Theory" PRENTICE HALL, 1996
- [2] 佐藤 明, 田村潤三, 来山征士: "相関を利用した重畳通話検出方式", 信学技報, CS75-136(1975).
- [3] 来山征士, 田村潤三, 山本誠一, 石上彦一: "共通の適応制御部をもつ多重エコーキャンセラ", 信学技報, CS78-23(1978).
- [4] 古川博基: "適応フィルタの実用技術 (3)", 音響誌, 48, 7, pp.513-516(1992).
- [5] Yamamoto S. and Kitayama S.: "An Adaptive Echo Canceler with Variable Step Gain Method", Trans.IECE Japan, E65, 1, pp.1-8(1982).
- [6] Akihiko Sugiyama: "An Interference-Robust Stochastic Gradient Algorithm with a Gradient-Adaptive Step-Size", Proc. of ICASSP '93, vol.3, pp.539-542(1993).
- [7] Hector Perez Meana, et. al.: "A Time Varying Step Size Normalized LMS Echo Canceler Algorithm", Proc. of ICASSP '94, vol.2, pp.249-252(1994).
- [8] 牛丸 晋一郎, 中山謙二, 平野晃宏: "A Variable Step size Acoustic Echo Canceler: A Step Size Control Method Based on Cross-correlation between Echo and Near-end Signal"