

2次ランダム可変前処理を用いたステレオエコーキャンセラ

A Stereo Echo Canceler Using Second-Order Randomly Time-Varying All-Pass Filter

武部 歌津衣 平野 晃宏 中山 謙二
Kazue Takebe 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: takabe@leo.ec.t.kanazawa-u.ac.jp

あらまし

2次のAPFを用いた前処理を行なったステレオエコーキャンセラ(SEC)を提案する。SECでは入力信号の相互相関のため、適応フィルタはエコーパスを正しく同定することができない。この問題を解決するために入力信号にランダムに時間変化する1次のオールパスフィルタ(APF)を用いて前処理を行なう方法が試みられている[5]。そこで、この方法の収束特性および音声品質について解析し、その結果から従来の前処理を用いた方法より実環境に適した付加回路を提案し、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 proposes a stereophonic echo canceler (SEC) with pre-processing using second-order randomly time-varying all-pass filter (APF). In a stereo echo canceler, adaptive filters can't identify the correct echo paths due to cross-correlated stereo signals. Pre-processing technique based on time-varying first-order all-pass filter have been proposed to solve this problem [5]. This paper analyzes convergence characteristics and quality of pre-processed signals and proposes pre-processing circuit that is more suitable for actual environment. Computer simulation results confirm the analyses.

1 まえがき

TV会議システムでは、音響エコーを除去するためにエコーキャンセラが用いられている。現在、TV会議システムは音声伝達に関してシングルチャンネルシステムがほとんどであるが、より空間的なリアリズムの追求のためにはマルチチャンネルシステム、少なくともステレオチャンネルシステムが必要となる。そこでステレオシステムのためのSECが必要となる。

SECでは、入力音響信号の両チャンネル間の強い相互相関の影響が大きな問題となる[1-3]。これにより適応フィルタ係数は最適値に収束せず、また、収束速度も遅くなってしまう。即ち、エコーパス特性を正しく同定することができない。そこでこの問題を解決するためにSECの2つの入力信号に前処理を行なう方法が試みられている[4,5]。その一つとして、1次のAPFを用い、係数をランダムに変化させる方法が提案されている[5]。しかし、収束速度と音声品質の両面からすると十分な性能は得られていない。今回は、APFを2次に拡張し、収束特性と音声品質の両面で良好な特性を実現するための付加回路の特性について検討を行なった。

2 ステレオエコーキャンセラ

2.1 基本構成

SECの基本構成を図1に示す。SECは4つの適応フィルタから成る。各適応フィルタは対応するエコーパス特性を同定し、エコーを除去する。

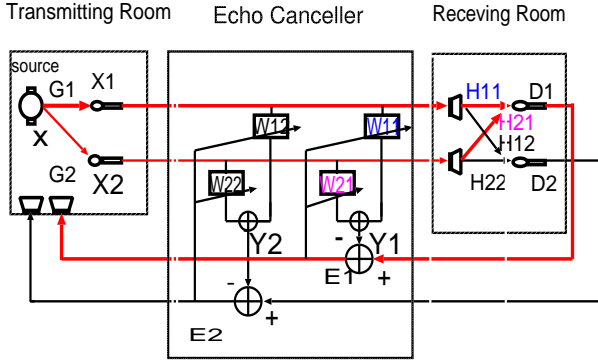


図 1: ステレオエコーキャンセラ・システムの構成

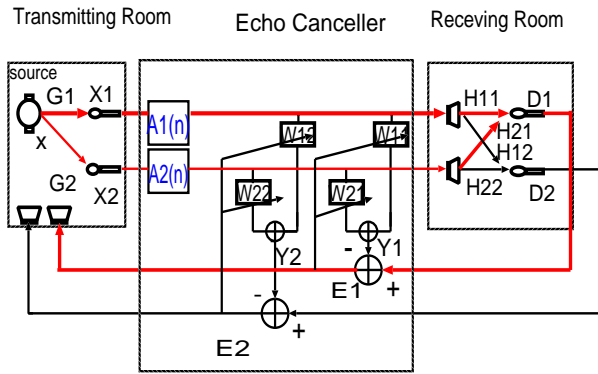


図 2: 付加回路を用いたステレオエコーキャンセラ・システムの構成

2.2 係数不確定性の問題

ここでは一つのマイク信号のエコーキャンセラだけを考慮するが(図 1 太線), 同じ議論がもう一方の経路に対してもそのまま成り立つ。

図 1 において, 送信側の部屋の伝達特性を $G_1(z), G_2(z)$ 入力信号の周波数特性を $X(z)$, エコーパスの伝達関数を $H_{11}(z), H_{21}(z)$, 適応フィルタの伝達関数を $W_{11}(z), W_{21}(z)$ とする。このとき, $E_1(z)$ は

$$E_1(z) = \sum_{j=1}^2 (H_{j1}(z) - W_{j1}(z)) G_j(z) X(z) \quad (1)$$

となり, $E_1(z) = 0$ を満たすフィルタ係数は, $W_{j1}(z) = H_{j1}(z)$ で一意に決まるとは限らない。そのため, エコーは消えても必ずしもフィルタ係数がエコーパス特性を同定できるとは限らない。これが SEC における係数不確定性の問題である。

3 付加回路を用いた SEC

前述した問題を解決するため入力信号に前処理を行なう方法が試みられている [4,5]。ここでは, 時間変化する APF $A(z, n)$ を用いて入力信号を前処理する。入力信号に前処理を行なう付加回路を用いた SEC の構成を図 2 に示す。対称性より, 片チャンネルのみを考え, 同時に 1 人以上が発言しないこととする。この場合, 誤差信号は, 次のようになる。

$$E_1(z, n) = \sum_{j=1}^2 (H_{j1}(z) - W_{j1}(z)) A_j(z, n) G_j(z) X(z) \quad (2)$$

$A_j(z, n)$ を時間変化させるとすると, 多くの異なる $A_j(z, n)$ に対して $E_1(z, n) = 0$ を満たす必要がある。よって, 適応フィルタのフィルタ係数は $W_{j1}(z) = H_{j1}(z)$ で一意に決定することができる。

4 1 次の APF を用いた付加回路

4.1 従来法

現在提案されている付加回路として, ランダムに時間変化する 1 次の APF を用いる方法があり [5], APF $A(z, n)$ の伝達関数は,

$$A(z, n) = \frac{z^{-1} - b(n)}{1 - b(n)z^{-1}} \quad (3)$$

で表され, 変数 $b(n)$ は以下のように変更する。

$$b(n+1) = b(n) + \text{rand}(n) \quad (4)$$

$$b(n+1) = b_{max} \text{ if } b(n+1) > b_{max} \quad (5)$$

$$b(n+1) = b_{min} \text{ if } b(n+1) < b_{min} \quad (6)$$

$\text{rand}(n)$ は, $-a$ から a までの範囲のランダムな数とし, a は定数とする。

4.2 従来法の解析

1 次の APF で入力信号を前処理することによって収束にどのように影響があるか解析する。1 次の APF の群遅延特性を図 3 ($0 \leq b(n) \leq 0.9$), 図 4 ($-0.9 \leq b(n) \leq 0$) に示す。1 次の APF の群遅延特性は, $0 \leq b(n)$ のとき低域で大きくなり, 反対に $b(n) \leq 0$ のとき高域で大きくなる。

TV 会議システムなどの実環境では部屋の音響特性がローパス型で入力信号が音声であることを考慮し, 部屋の音響特性である $G_i(z), H_{ij}(z)$ には 20 次のバターワースフィルタを用いローパスフィルタで構成

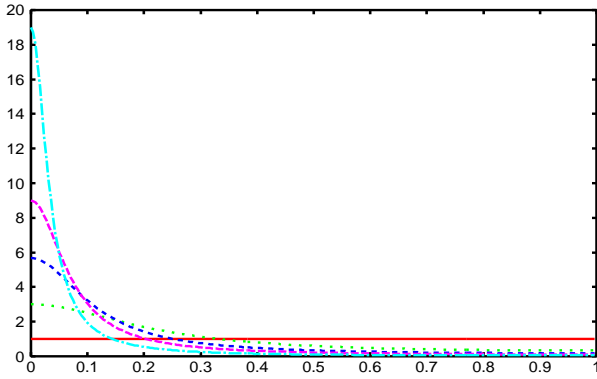


図 3: 群遅延特性
(下から $b = 0, 0.5, 0.7, 0.8, 0.9$)

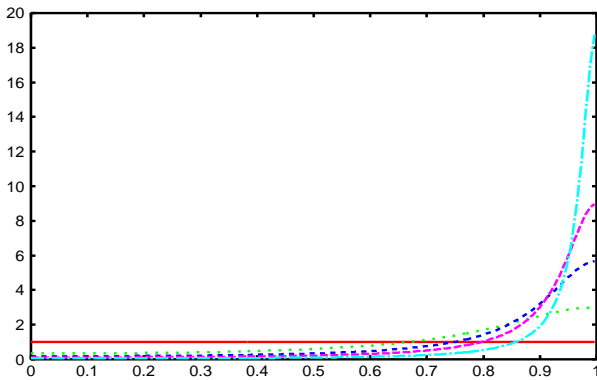


図 4: 群遅延特性
(下から $b = 0, -0.5, -0.7, -0.8, -0.9$)

した。よって信号成分は低域に多くある。そのため、 $0 \leq b(n)$ のとき $b(n) \leq 0$ のときよりも信号変化が大きくなる。したがって $0 \leq b(n)$ のとき、その群遅延特性から (2) 式の独立性が増すため収束が速くなると考えられる。入力に白色雑音を用い、適応フィルタを 64 タップの FIR フィルタで NLMS アルゴリズムを用いて、従来法で適応フィルタの係数の収束特性の解析を行なった。ステップサイズは $\mu = 0.5$ とした。 $b(n)$ を $-0.9 \leq b(n) \leq 0$ と $0 \leq b(n) \leq 0.9$ の範囲で時間変化させたときのシミュレーション結果を図 5 に示す。

シミュレーション結果より、 $0 \leq b(n)$ のときは $b(n) \leq 0$ のときよりも収束が速くなった。付加回路は収束を速くする一方で、元の音声を劣化させてはならない。エコーキャンセラで用いられる入力信号は音声であり、音声信号は基本的には低域にほとんどの成分がある。そのため、群遅延特性が低域で大きい場合には、音声品質が劣化すると考えられる。入力信号

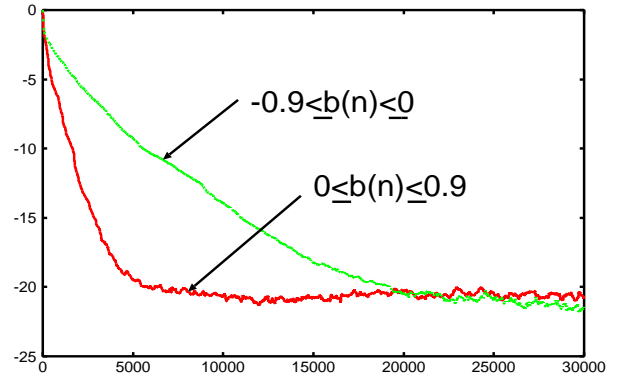


図 5: 従来法の収束特性 (フィルタ係数誤差 [dB])

を音声にしたとき、付加回路通過後の音声は $0 \leq b(n)$ では激しく劣化した。

5 2 次 の APF を用いた付加回路の提案

1 次 の APF を用いた付加回路の解析結果より、付加回路の群遅延特性が低域で大きい方が収束特性が良いことがわかった。また、音声品質に関しては群遅延が低域で小さい方が良いことがわかった。そこで、これらのトレードオフを最適化する必要がある。この 2 つの条件を両方とも満たすような最適付加回路を 2 次 の APF で構成する。2 次 の APF は、次式で表される。

$$A(z) = \frac{b_2(n) - b_1(n)z^{-1} + z^{-2}}{1 - b_1(n)z^{-1} + b_2(n)z^{-2}} \quad (7)$$

$$= \frac{(r(n))^2 - 2r(n) \cos \theta(n)z^{-1} + z^{-2}}{1 - 2r(n) \cos \theta(n)z^{-1} + (r(n))^2 z^{-2}} \quad (8)$$

$b_1(n) = 2r(n) \cos \theta(n)$, $b_2(n) = (r(n))^2$ の更新を次の 3 通りの方法で行なった。

- θ は固定とし、 r をランダム変化させる。
- r は固定とし、 θ をランダムに変化させる。
- θ が小さいと r も小さくし、 θ が大きいと r も大きくなるように、 r と θ をランダムに変化させる。

まず、 θ は固定とし、 r をランダムに変化させたときの r の更新式は、

$$r(n+1) = r(n) + \text{rand}(n) \quad (9)$$

$$r(n+1) = r_{max} \text{ if } r(n+1) > r_{max} \quad (10)$$

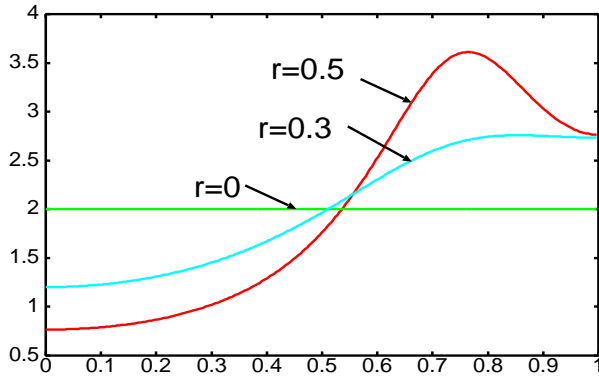


図 6: r を変化させたときの群遅延特性

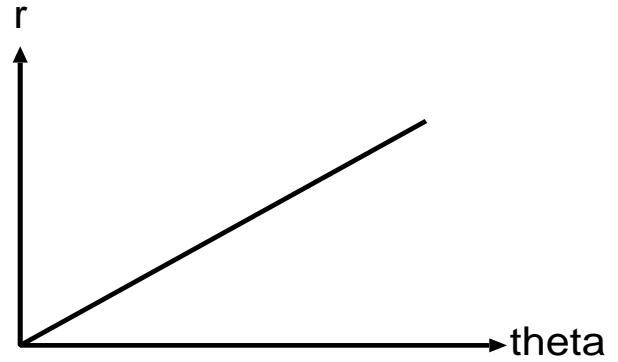


図 8: r と θ の関係

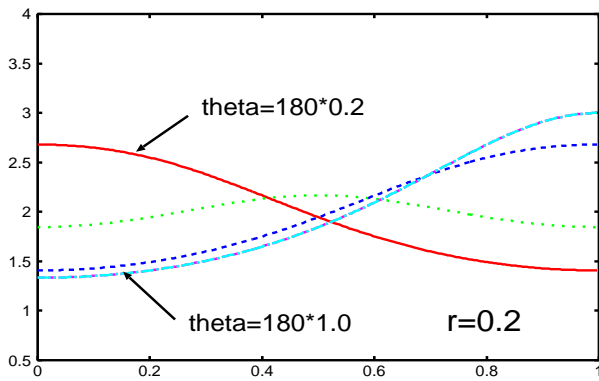


図 7: θ を変化させたときの群遅延特性
($\theta = 180 * 0.2, 180 * 0.5, 180 * 0.8, 180 * 1.0$)

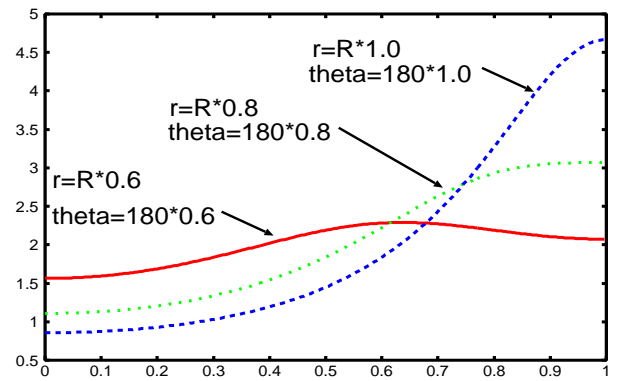


図 9: r, θ を時間変化させたときの群遅延特性

$$r(n+1) = r_{min} \text{ if } r(n+1) < r_{min} \quad (11)$$

となる。ただし、 $rand(n)$ はランダム変数である。

2 次の APF で $r = 0, 0.3, 0.5, \theta = 135^\circ$ と変化させたときの群遅延特性を図 6 に示す。

次に、 r は固定とし、 θ をランダムに時間変化させたときの更新式は、

$$\theta(n) = 180 * Z(n) \quad (12)$$

$$Z(n+1) = Z(n) + rand(n) \quad (13)$$

$$Z(n+1) = Z_{max} \text{ if } Z(n+1) > Z_{max} \quad (14)$$

$$Z(n+1) = Z_{min} \text{ if } Z(n+1) < Z_{min} \quad (15)$$

θ を変化させたときの群遅延特性を図 7 に示す。

次に、 r と θ を両方をランダムに時間変化させた。1 次の APF でシミュレーションしたときに、群遅延が低域にあると音声品質が悪くなるという結果から、 θ が小さいとき r も小さくして、群遅延の変化を小さく

し音声品質の劣化を防ぎ、 θ が大きい時は r も大きくなるようにして収束速度が速くなるように設計した。 r と θ の関係は図 8 のようにした。

r と θ の更新式は次のようになる。

$$r = R * Z(n) \quad (16)$$

$$\theta = 180 * Z(n) \quad (17)$$

$$Z(n+1) = Z(n) + rand(n) \quad (18)$$

$$Z(n+1) = Z_{max} \text{ if } Z(n+1) > Z_{max} \quad (19)$$

$$Z(n+1) = Z_{min} \text{ if } Z(n+1) < Z_{min} \quad (20)$$

R は定数とする。

r と θ を変化させたときの群遅延特性をを図 9 に示す。

図 6, 図 7, 図 9 より、 r を変化させたときと θ を変化させたときは、ほぼ全帯域において群遅延の変化があるので収束速度は速くなると考えられる。また、 r と θ を両方時間変化させたときは低域で群遅延の変化が小さいので、音質は良くなると考えられる。

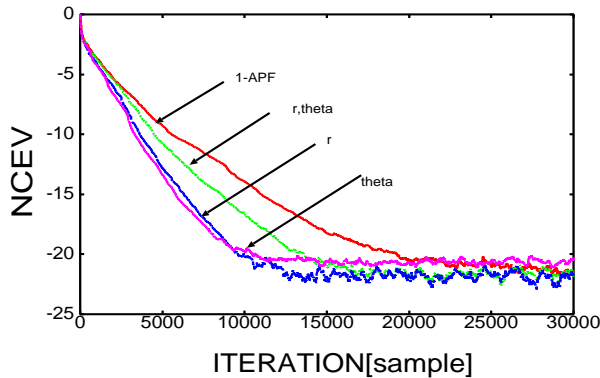


図 10: 収束特性 (フィルタ係数誤差 [dB])

6 計算機実験

付加回路に2次のAPFを用いたときの計算機実験を行なった。シミュレーション条件は、付加回路に1次のAPFを用いたときと同じ条件とした。収束速度と音声品質のトレードオフがあるため、それぞれのパラメータは、 r を時間変化させるときは $0 \leq r(n) \leq 0.55$, $\theta = 135^\circ$ をとり、 θ 時間変化させるときは $0.2 \leq Z(n) \leq 1.0$, $r = 0.2$ をとり、 r と θ を時間変化させるときは $0.6 \leq Z(n) \leq 1.0$, $R = 0.4$ とした。このときのシミュレーション結果を図10に示す。付加回路に2次のAPFを用いた場合、3通りとも収束速度は1次のAPFで $b(n) \leq 0$ のときより速くなった。収束速度においては、 r をランダムに時間変化させたときと θ をランダムに時間変化させたとき収束が速くなった。音声品質においては、 r と θ を両方ランダムに時間変化させたとき他の2つに比べて、若干良好であった。

7 むすび

ステレオエコーキャンセラの付加回路と収束特性の関係について解析した。ランダム可変する1次のAPFを付加回路に用いた場合、付加回路と収束特性、音声品質の関係には群遅延が深くかかわっていることがわかった。付加回路の群遅延が低域で大きい場合、収束が速くなるが、音声品質は劣化し、群遅延が高域で大きい場合、収束は遅いが音声品質は良くなる。そこで、1次のAPFでは、その群遅延特性から収束速度、音声品質の両面における向上に限度があるので、付加回路に2次のAPFを用いて、収束速度が速く音質の良いSECを設計した。そして、計算機実験により、この解析の妥当性を確認した。群遅延の変化と収束特性の関係及び音声品質の関係を定量的に解析することが今後の課題である。

参考文献

- [1] A. Hirano and A. Sugiyama, "Convergence Characteristics of a Multi-Channel Echo Canceller with Strongly Cross-correlated Input Signals - Analytical Result -," Proc. of 6th DSP Symposium, pp. 144-149, November 1991.
- [2] 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, August 1995
- [3] A. Hirano and S. Koike, "Convergence Analysis of a Stereophonic Acoustic Echo Canceller Part I: Convergence Characteristics of Tap Weights," Proc. of 11-th DSP Symposium, pp. 569-574, November 1996.
- [4] Y. Joncour and A. Sugiyama, "A Stereo Echo Canceller with Pre-Processing for Correct Echo-Path Identification," ICASSP, pp.3677-3680, 1998.
- [5] M. Ali, "Stereophonic Acoustic Echo Cancellation System Using Time-varying All-pass Filtering for Signal Decorrelation," ICASSP, pp.3689-3692, 1988.
- [6] 渡辺和伸, 中山謙二, 平野晃宏: "付加回路をもちいるステレオエコーキャンセラの収束性解析" 第13回デジタル信号処理シンポジウム, pp177-181, 1998.