

◎島内 末廣 牧野 昭二 (NTT ヒューマンインタフェース研究所)

1 はじめに

ステレオ音声による拡声通信システムは、臨場感の高い快適な通話手段として期待されている。例えば、テレビ会議システムにおいて、複数の相手同士との通話時に、容易に話者の識別ができる等、会議進行の円滑化に役立つ。しかし、ステレオ拡声通話実現に不可欠であるステレオ音響エコーキャンセラには、受話信号の相互相関により音響エコー経路の推定を誤る問題がある [1][2][3][4]。

本報告では、実環境におけるステレオ信号の相互相関の変化に着目したステレオ射影アルゴリズムを提案し、本手法により真の音響エコー経路を推定可能であることを示す。

2 ステレオ音響エコーキャンセラの構成

ステレオ拡声通話用の音響エコーキャンセラは、図1のように2つの独立な線形結合型エコーキャンセラにより構成できる。1つの線形結合型エコーキャンセラは、入力ベクトル $x_1(k), x_2(k)$ と疑似音響エコー経路ベクトル $\hat{h}_1(k), \hat{h}_2(k)$ をそれぞれ $x(k) = [x_1^T(k), x_2^T(k)]^T$, $\hat{h}(k) = [\hat{h}_1^T(k), \hat{h}_2^T(k)]^T$ とまとめて扱う。但し、 $*^T$ はベクトルの転置を表す。1つの音響エコー経路を L タップの FIR フィルタで模擬する場合、 $2L$ タップの1チャンネルエコーキャンセラと同様な構成となる。

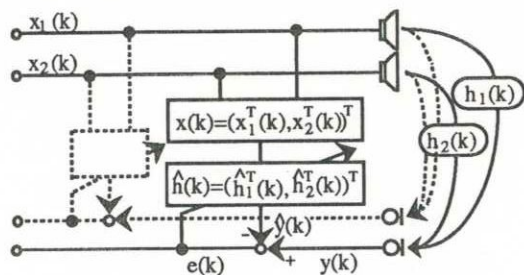


図1: ステレオエコーキャンセラの構成

3 ステレオ音響エコーキャンセラの問題点

ステレオ信号が強い相互相関を持っているために生じる問題点を述べる。入力ベクトル $x_1(k)$ と $x_2(k)$ とが、ある一定な相互相関を持つ時間区間に

*Stereo projection echo canceller based on the variations in the cross-correlation between stereo signals.
By Suehiro Shimauchi and Shoji Makino (NTT Human Interface Laboratories)

において、真の音響エコー経路 $h_1(k), h_2(k)$ を経て得られるエコー $y(k) = h_1^T(k)x_1(k) + h_2^T(k)x_2(k)$ と疑似エコー $\hat{y}(k) = \hat{h}_1^T(k)x_1(k) + \hat{h}_2^T(k)x_2(k)$ とを一致させる結合ベクトル $[\hat{h}_1^T(k), \hat{h}_2^T(k)]$ は無数に存在し、 $x_1(k)$ と $x_2(k)$ との相互相関に固有な部分空間 H_x を形成する。このため、NLMS等の逐次誤差最小化アルゴリズムを用いた場合に、 $[\hat{h}_1^T(k), \hat{h}_2^T(k)]$ は初期値から部分空間 H_x までの距離が最小となる点に収束し、一般に真値 $[h_1^T(k), h_2^T(k)]$ には収束しない。

簡単な例として、一定なスカラー値 α_1, α_2 と原信号 $s(k)$ により、入力ベクトルが、

$$x_1(k) = \alpha_1 s(k), \quad x_2(k) = \alpha_2 s(k) \quad (1)$$

と表せる場合を考える。 $[\hat{h}_1^T(k), \hat{h}_2^T(k)]$ が存在し得る部分空間 H_x は、図2の直線のように見なすことができ、初期値 o から適応を開始した場合、収束点 $[\hat{h}_{1\alpha}^T(k), \hat{h}_{2\alpha}^T(k)]$ は、

$$\hat{h}_{1\alpha}(k) = \frac{\alpha_1^2}{\alpha_1^2 + \alpha_2^2} (h_1(k) + \frac{\alpha_2}{\alpha_1} h_2(k)) \neq h_1(k) \quad (2)$$

$$\hat{h}_{2\alpha}(k) = \frac{\alpha_2^2}{\alpha_1^2 + \alpha_2^2} (\frac{\alpha_1}{\alpha_2} h_1(k) + h_2(k)) \neq h_2(k) \quad (3)$$

のように得られる。

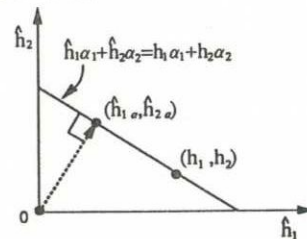


図2: 相互相関のある入力に対する適応動作

4 ステレオ信号の相互相関の変化の影響

入力ベクトル $x_1(k), x_2(k)$ の相互相関が変化した場合として、式(1)の α_1, α_2 が β_1, β_2 に変化した例を考える。まず、 $[\hat{h}_1^T(k), \hat{h}_2^T(k)]$ は、式(2),(3)の $[\hat{h}_{1\alpha}^T(k), \hat{h}_{2\alpha}^T(k)]$ に収束する。次に、 $[\hat{h}_{1\alpha}^T(k), \hat{h}_{2\alpha}^T(k)]$ を“初期値”として、そこから最短距離にある $[\hat{h}_{1\beta}^T(k), \hat{h}_{2\beta}^T(k)]$ に収束する(図3)。ここで、フィルタ係数誤差ベクトルの大きさは、一般に $\|e_\alpha\| > \|e_\beta\|$ であることが自明であるから、入力信号の相互相関の変化の度に、フィルタ係数誤差が小さくなることが分かる。

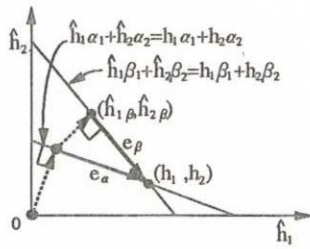


図 3: 入力の相互相関の変化の影響

5 ステレオ射影アルゴリズム

適応フィルタの真の音響エコー経路への収束を速めるには、入力信号の相互相関の変化に着目することが有効であると考えられる。射影アルゴリズム [5] は、現時点での入力ベクトル $x(k)$ 以外に過去の入力ベクトルの情報を用いる。そのため、各時刻毎の入力ベクトルの相互相関の変化成分を抽出できることが期待できる。そこで、以下のように射影アルゴリズムを線形結合型エコーキャンセラに適用したステレオ射影アルゴリズムを提案する。

$$\hat{h}(k+1) = \hat{h}(k) + \frac{e(k)}{u^T(k)u(k)}u(k) \quad (4)$$

但し、

$$u(k) = x(k) - \frac{x^T(k-1)x(k)}{x^T(k-1)x(k-1)}x(k-1) \quad (5)$$

$$e(k) = y(k) - \hat{y}(k) \quad (6)$$

6 計算機シミュレーション

スピーカ・マイクロホン間のインパルス応答を残響時間 150ms の部屋で測定し、これを模擬する $\hat{h}_1(k)$, $\hat{h}_2(k)$ を各 500 タップで構成した。標準化周波数は 8kHz である。

6.1 相互相関固定のステレオ音声信号入力時

入力信号は、モノラルの音声信号を 2 つの異なるインパルス応答と計算機上で畳み込むことにより作成した。NLMS, 射影アルゴリズムともに、30 dB 以上の ERLE を得ることができる (図 4)。一方、フィルタ係数誤差はともに -3 dB 程度で停滞している (図 5)。これは、適応フィルタが真の音響エコー経路以外の局所解に収束したことを意味している。

6.2 実測したステレオ音声信号入力時

入力信号は、2 本のマイクロホンによりステレオ集音した。集音時に話者は頭や体を動かしていないが、フィルタ係数誤差は、NLMS を用いた場合でも、僅かずつ減少し続けている (図 6)。また、

射影アルゴリズムを適用した場合は、誤差の減少はさらに加速され真値へ近づく。

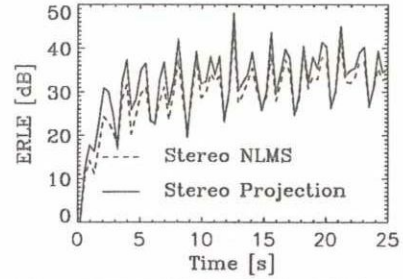


図 4: ERLE(相互相関固定信号)

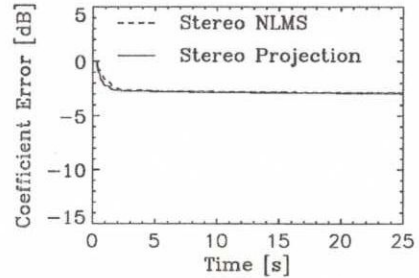


図 5: フィルタ係数誤差(相互相関固定信号)

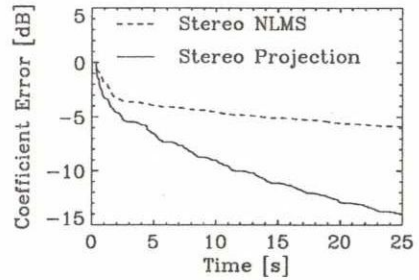


図 6: フィルタ係数誤差(実測信号)

7 まとめ

本報告では、ステレオ射影アルゴリズムを提案し、ステレオ信号の相互相関が僅かに変動する実環境を想定した計算機シミュレーションにより真値を推定できることを確認した。

謝辞 本研究を進めるにあたり、御指導・御討論頂いた音声情報研究部北脇部長、小島リーダを始め、部員諸氏に感謝する。

参考文献

- [1] M. M. Sondhi and D. R. Morgan, *1991 IEEE Workshop on Applications of Signal Processing to Audio and Acoustics* (1991).
- [2] Y. Mahieux, A. Gilloire, and F. Khalil, *EU-SIPCO 93* (1993).
- [3] A. Hirano, A. Sugiyama, Y. Arasawa, and N. Kawayachi, *ICASSP94*, pp. 245-248 (1994).
- [4] 田川, 古川 他, 音講論集, 1-9-4 (1994).
- [5] S. Makino and Y. Kaneda, *IEICE Trans.*, Vol. E75-A, No. 11, pp. 1500-1508, (Nov. 1992).