二次の射影法とスペクトルサブトラクションを用いた音響エコー抑圧*

◎大久保俊也　滝口 哲也　有木 康雄（神戸大・工）

1　はじめに

近年，カーナビゲーションシステムなど，音声を用いた対話型の電子機器が増加している．このようなシステムは，システムのスピーカとマイクロフォンの音響結合による音響エコーの発生により，システムの発する音とユーザの発話とが区別できず，誤動作を起こすという問題がある．

これらの問題を解決するためには，音響エコーキャンセラが用いられる．音響エコーキャンセラの適応アルゴリズムには，最も簡易な学習同定法がよく用いられるが，二次の射影法の方が適応フィルタの収束速度が速く，音響エコー抑圧の性能が良いとされる．

また，音響エコーキャンセラは音響エコー以外の雑音を除去できないため，自動車内での走行ノイズなど，音響エコー以外の雑音のある環境下では，音声認識性能が劣化する．そのため，システムの音声認識性能を高める為には，そのような音響エコー以外の雑音も除去する必要がある．

本稿では，定常的な雑音環境下にて音響エコーが発生した場合を想定し，二次の射影法を適応アルゴリズムに用いた音響エコーキャンセラにスペクトルサブトラクションを組み合わせることにより，音響エコーと定常雑音を共に抑圧した結果について報告する．

2　音響エコーキャンセラ

2.1　音響エコーキャンセラとは

音響エコーキャンセラは，図 1 のように，室内伝達系のインパルス応答 $\mathbf{h}(t)$ を, 適応フィルタ $\mathbf{w}(t)$ により推定し，これを音源信号 $\mathbf{x}(t)$ と畳み込むことにより擬似的な音響エコー成分を作成し，観測信号 $\mathbf{y}(t)$ から差し引くことにより音響エコーを抑圧する．ここで，$\mathbf{w}(t)$ と $\mathbf{x}(t)$ は，時刻 t において，それぞれ以下の要素をもつ，長さ L(タップ長) のベクトルである．

$$\mathbf{w}(t) = [w_0(t), w_1(t), \cdots, w_{L-1}(t)]^T \quad (1)$$

$$\mathbf{x}(t) = [x(t), x(t-1), \cdots, x(t-L+1)]^T \quad (2)$$

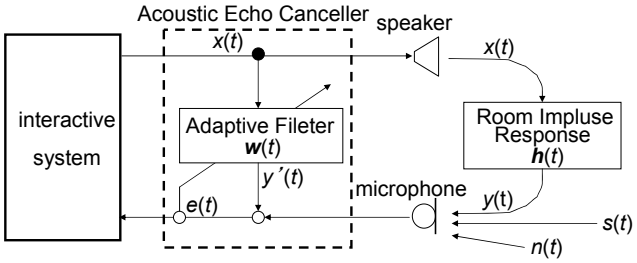


図 1: 音響エコーキャンセラ

この適応フィルタの更新方法には，いくつかのアルゴリズムがあるが，基本的には次式の誤差信号 $e(t)$ が最小となるように更新を行う．

$$e(t) = y(t) - \mathbf{w}(t)^T\mathbf{x}(t) \quad (3)$$

以降では，最もよく用いられる学習同定法と，二次の射影法について概説する．

2.2　学習同定法

学習同定法 (NLMS：Normalized Least-Mean Square) は，過去の 1 つの入出力関係を満足するように適応フィルタの更新を行うアルゴリズムである．具体的には (4) 式のようにして更新する [1].

$$\mathbf{w}(t+1) = \mathbf{w}(t) + \frac{\alpha}{\mathbf{x}(t)^T\mathbf{x}(t) + \beta} e(t)\mathbf{x}(t) \quad (4)$$

ここで，α はフィルタ係数の更新量を制御するステップサイズであり，β は分母が 0 にならないための小さな定数である．

学習同定法は，関係式を 1 つ満たすだけで良いので計算量が少なく，簡易に用いることができる一方，フィルタ係数の収束に時間を要し，即応性に欠けるという問題がある．

2.3　二次の射影法

射影法は，過去の複数の入出力関係を満足するように適応フィルタの更新を行うアルゴリズムである．ここでは，過去の 2 つの関係を満たす二次の射影法について述べる．二次の射影法では，(5) 式のようにしてフィルタを更新する [1].

$$\mathbf{w}(t+1) = \mathbf{w}(t) + \begin{bmatrix} \mathbf{x}(t) & \mathbf{x}(t-1) \end{bmatrix} \cdot \begin{bmatrix} \mathbf{x}(t)^T\mathbf{x}(t) & \mathbf{x}(t)^T\mathbf{x}(t-1) \\ \mathbf{x}(t-1)^T\mathbf{x}(t) & \mathbf{x}(t-1)^T\mathbf{x}(t-1) \end{bmatrix}^{-1} \cdot \begin{bmatrix} e(t) \\ (1-\mu)e(t-1) \end{bmatrix} \quad (5)$$

ここで，μ は次式で表わされる定数である．

$$\mu = \frac{\alpha}{\mathbf{x}(t)^T\mathbf{x}(t) + \mathbf{x}(t-1)^T\mathbf{x}(t-1)} \quad (6)$$

二次の射影法は，満たすべき関係式が 2 つと制約が強く，フィルタ係数の収束が速いため，音響エコー抑圧性能が高い．表 1 は，音響エコーのみが存在する場合での学習同定法と二次の射影法によるエコーキャンセラの性能を，ERLE(Echo Return Loss Enhancement) により比較したものである．表 1 からも，二次の射影法を用いた方がよりエコーを抑圧できることがわかる．なお，ERLE は次式により求められる．

$$ERLE = 10\log_{10}\frac{\Sigma y(t)^2}{\Sigma(y(t) - \mathbf{w}(t)^T\mathbf{x}(t))^2} \quad (7)$$

表 1: ERLE による比較

	学習同定法	二次の射影法
ERLE [dB]	20.84	31.60

*An acoustic echo reduction using a 2-order projection algorithm for filter adaptation and a spectral subtraction.
by Toshiya Ohkubo, Tetsuya Takiguchi and Yasuo Ariki (Kobe Univ.)

2.4 二次の射影法とスペクトルサブトラクションの連結

音響エコーキャンセラは，音響エコーの抑圧には有効であるが，音響エコー以外の雑音を除去することはできない．しかしながら，自動車内などでの実環境下においては，走行音などといった比較的定常的な雑音が存在する．そのため，音響エコーキャンセラに定常的な雑音を抑圧する処理を組み合わせることで，より目的音声を強調できると考えられる．

本稿では，最も簡易に用いることができる定常雑音抑圧処理であるスペクトルサブトラクション (SS：Spectral Subtraction)[2] を組み合わせることにより，目的音声がどの程度強調されるかについて検討する．

3 シミュレーション実験

3.1 実験概要

実験で使用する音響エコー信号は，3つの音楽データに対し，RWCP 実環境音声・音響データベース内の残響時間の異なる 3 種類の実環境インパルス応答（無響室，残響時間 120ms，300ms）を畳み込んで作成した．また，目的信号である音声発話にも同様のインパルス応答を畳み込んだ．定常雑音にはホワイトノイズを用い，目的信号，音響エコーおよび定常雑音信号を足し合わせて観測信号を作成した．エコーキャンセラの分析パラメータは，表 2 のとおりである．

本実験では，エコーキャンセラによる音響エコー抑圧処理後の信号に対してスペクトルサブトラクションを行った．定常雑音成分の推定は，音響エコー抑圧処理後の信号の非音声区間約 2 秒を用いて行った．

3.2 実験結果

雑音抑圧効果を SN 比により評価した結果を表 3 に示す．表 3 より，どの残響環境下においても，二次の射影法とスペクトルサブトラクションの組み合わせたが最も良好な SN 比を得ていることがわかる．

また，残響時間 300ms の場合において，各手法により音響エコーおよび定常雑音を抑圧した波形を図 2 に示す．図 2(c) の学習同定法を用いた場合には音響エコーの消し残しが見られるが，図 2(d) の二次の射影法を用いた場合，音響エコー成分が見られない，よって，二次の射影法は定常雑音環境下においても音響エコー抑圧の効果がより大きいといえる．

ところで，図 2(d) において，目的音声が減衰しているが，これは，ダブルトーク制御を行っていないため，二次の射影法の収束性の速さにより，発話区間において，即座に，音声も抑圧するようフィルタを更新したためであると考えられる．

表 2: 分析パラメータ

サンプリング周波数	16000Hz
L（タップ長）	256
α（ステップサイズ）	1.0
β	0.1

4 おわりに

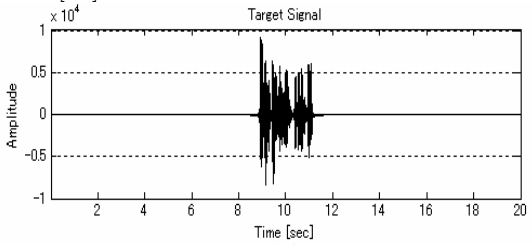
本稿では，二次の射影法を用いた音響エコーキャンセラとスペクトルサブトラクションを用いて，音響エコーと定常雑音を抑圧する手法について検討した．定常雑音が存在する場合であっても，フィルタ係数の収束速度の速い二次の射影法を用いた方が，学習同定方を用いた場合より，音響エコー抑圧の効果が高いことが確認できた．

今後は，周波数上において音響エコーキャンセラを実装し，フィルタ適応と同時に音響エコー以外の雑音を抑圧する [3][4] などを行い，より効果的に音響エコーならびに雑音の抑圧を行いたい．

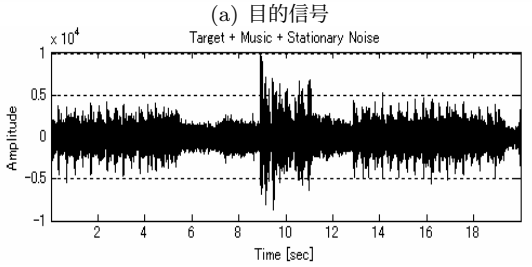
表 3: SN 比による雑音抑圧効果の比較

残響時間	無処理	学習同定法+SS	二次射影法+SS
無響	3.09	27.24	27.92
120ms	3.53	22.71	27.95
300ms	0.63	13.98	16.92

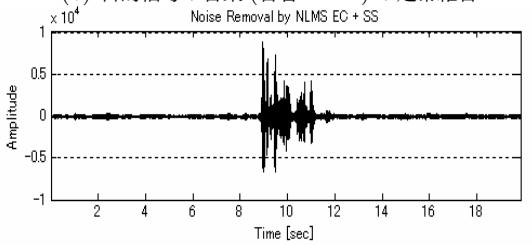
∗ 単位は [db].



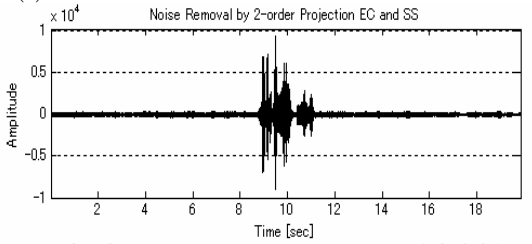
(a) 目的信号



(b) 目的信号＋音楽 (音響エコー) ＋定常雑音



(c) 学習同定法を用いた EC と SS による雑音除去結果



(d) 二次の射影法を用いた EC と SS による雑音除去結果

図 2: 波形による雑音抑圧効果の比較

参考文献

[1] 大賀寿郎, 山崎芳男, 金田豊, ”音響システムとディジタル処理”, 電子情報通信学会, 1995.

[2] Steven F.Boll, ”Suppression of Acoustic Noise in Speech Using Spectral Subtraction”, *IEEE Trans. Acoust., Speech and Signal Process.*, Vol.ASSP-27, No.2, pp.113-120, 1979.

[3] 阪内澄宇, 羽田陽一, 田中雅史, 佐々木潤子, 片岡章俊, ”雑音抑圧及びエコー抑圧昨日を備えた音響エコーキャンセラ”, 信学論 (A), Vol.J87-A, No.4, pp.448-457, April, 2004.

[4] O.Ichikawa, M.Nishimura, ”Simultaneous Adaptation of Echo Cancellation and Spectral Subtraction for In-Car Speech Recognition”, *IEEE Trans. Fundamentals*, Vol.E88-A, No.7, pp.1732-1738, JULY, 2005.