

エコーキャンセラにおける 適応アルゴリズムとダブルトーク検出の関係

Relationship between Adaptive Algorithms and Double Talk Detection in Echo Cancellers

○箱石 直士, 勝俣 友紀, 西山 清, 佐藤 克昌[†], 三澤 栄治[†]

○Naohito Hakoishi, Tomonori Katsumata, Kiyoshi Nisiyama,
Katsuaki Satou[†], Eiji Misawa[†]

岩手大学, Iwate University

[†](株)エー・アール・アイ, Amenity Research Institute

キーワード : ハイパー H_{∞} フィルタ (hyper H_{∞} filter), 高速 H_{∞} フィルタ (fast H_{∞} filter)
エコーキャンセラ (echo canceller), システム同定 (system identification)

連絡先 : 〒020-8551 盛岡市上田4-3-5 岩手大学 工学部 情報システム工学科 西山研究室
西山清, Tel.: (019)621-6475, Fax.: (019)621-6475, E-mail: nisiyama@cis.iwate-u.ac.jp

1. はじめに

エコーキャンセラは国際通信や携帯電話、ハンズフリー通話システムなどで広く用いられており、ダブルトーク検出はエコー除去において通話品質を維持するために重要である。エコーキャンセラにおいてはエコーパスを推定するために適応フィルタが用いられ、残留エコーを用いるダブルトーク検出法は適応アルゴリズムの性能の影響を受けるものと考えられる。本研究では、残留エコーを用いてダブルトーク検出を行うエコーキャンセラにおいて、適応アルゴリズムがダブルトーク検出性能へ及ぼす影響をシミュレーションにより明らかにする。

2. エコーキャンセラ

2.1 エコーキャンセラの原理

国際電話など長距離電話回線では、信号増幅などの理由から4線式回線が用いられている。一方、加入者回線は比較的短距離なので、2線式回線が使用されている。2線式回線と4線式回線の接続部にはハイブリッドトランスが導入され、インピーダンス整合が行われている。このインピーダンス整合が完全であれば、話者Bからの信号(音声)は話者Aに一度だけ到達する。しかし、一般に整合は完全にできず、受信信号の一部は4線式回線に漏れ、増幅された後、再び受信者(話者A)に戻るといった現象が起こる。これがエコー(echo)である。エコーは伝送距離が長くなるにつれて(遅延時間が長くなるにつれて)影響が大きくなり、著しく通話

の品質を劣化させる(パルス伝送においては近距離であってもエコーによる通話品質の劣化は問題となる)。そこで、図1のようにエコーキャンセラを導入し、直接観測可能な受信信号とエコーを用いてエコーパスのインパルス応答を逐次推定し、それを利用して得た疑似エコーによって、実際のエコーを打ち消し、その除去を図っている。

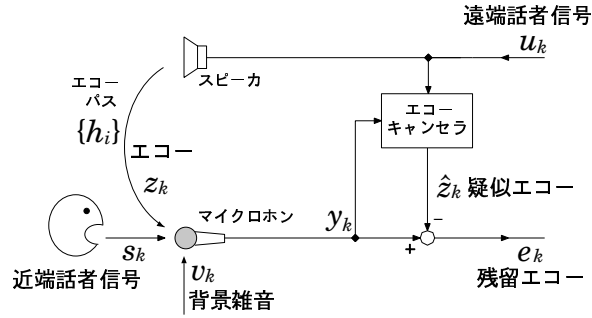


Fig. 1 エコーキャンセラの原理

図1にエコーキャンセラの原理を示す。

エコーパスのインパルス応答の推定は、残留エコー e_k の平均2乗誤差が最小になるように行われる。このとき、エコーパスの推定を妨害する要素は、背景雑音と近端話者からの信号(音声)である。一般に、話者2人が同時に話し始めた(ダブルトーク)ときはインパルス応答の推定を中断する。

次に、遠端話者信号 $\{u_k\}$ がエコーパスへの入力信号となることを考慮すれば、エコーパスのインパルス応答 $\{h_i\}$ により、エコーの観測値 $\{y_k\}$ は次式で表される。

$$y_k = z_k + v_k = \sum_{i=0}^N h_i u_{k-i} + v_k, \quad k \geq 0 \quad (1)$$

ただし、近端話者信号 s_k は無視し、 N はタップ数(インパルス応答長)とする。このとき、エコーパスのインパルス応答の推定値 $\{\hat{h}_i\}$ が得られれば、これより疑似エコーが次のように得られる。

$$\hat{z}_k = \sum_{i=0}^N \hat{h}_i u_{k-i}, \quad k = 0, 1, 2, \dots \quad (2)$$

この疑似エコーを送信信号から差し引くことによ

てエコーをキャンセルすることができる。ただし、 $k-i < 0$ のとき $u_{k-i} = 0$ とする。

以上より、エコーキャンセラは直接観測可能な遠端話者信号 $\{u_k\}$ とエコーの観測値 $\{y_k\}$ から適応フィルタを用いてインパルス応答を推定することによって実現できることがわかる。

この推定したインパルス応答と遠端話者信号 $\{u_k\}$ から、疑似エコー \hat{z}_k を次のように求め、送信信号から差し引けばエコーキャンセラが実現できる。

$$\hat{z}_k = \hat{\mathbf{h}}_k^T \mathbf{u}_k = \mathbf{u}_k^T \hat{\mathbf{h}}_k \quad (3)$$

ただし、

$$\begin{aligned} \hat{\mathbf{h}}_k &= [\hat{h}_0[k], \hat{h}_1[k], \dots, \hat{h}_N[k]]^T, \\ \mathbf{u}_k &= [u_k, u_{k-1}, \dots, u_{k-N}]^T \end{aligned} \quad (4)$$

2.2 適応フィルタ

ある初期値からフィルタ係数を適応的に少しずつ自己調整し、最終的に最適フィルタを実現するフィルタが考案されている。このようにフィルタ係数の自己調整機能(適応アルゴリズム)をもったデジタルフィルタを適応フィルタ(adaptive filter)と呼ぶ。本研究では適応フィルタとして正規化LMSアルゴリズム(NLMS)と高速 H_∞ フィルタ(FHF)を用いる。

高速 H_∞ フィルタはアルゴリズムの計算量が $O(N)$ であり、 H_∞ ノルムの上界 γ_f が評価関数である最大エネルギーゲインに含まれている。フィルタの存在条件を満たす最小の γ_f を求めることで、時変システムへの優れた追従特性を付与することができる。最良の追従特性を与えるシステム雑音(あるいは忘却係数)を H_∞ の意味で(準)最適に決定できる。計算量 $O(N)$ でアルゴリズムの正当性(フィルタの存在性)を検証できる。

2.3 ダブルトークとその対策

エコーキャンセラにおいて、ダブルトークは適応フィルタの係数を大きく乱すため、エコー消去量が減少し、通話品質が大幅に低下する危険性がある。その対策として音響エコーキャンセラでは通常、ダブルトークの発生を検出する回路を設け、その発生時に適応フィルタの係数の更新を休止している。またダブルトーク状態でなくなった場合は、エコーパスの変動に追従するために適応フィルタ係数の更新を再開する必要がある。

ダブルトーク検出法は、その検出に利用される信号がエコー相殺前の信号であるか、相殺後の信号であるかによって大きく二つに分かれる。後者を代表する方法では残留エコーのパワーからダブルトークを検出している。この方法はエコーに埋もれた近端話者音声の混入も比較的容易に検出できる。図2に本研究で用いたダブルトーク検出法のアルゴリズムを示す。ここで、 u_k は入力(遠端話者音声)、 y_k はエコーの観測値、 e_k は残留エコーを表し、 λ_u 、 λ_y 、 λ_e はそれぞれ1未満とする。

残留エコーはエコーパスのインパルス応答の推定値と遠端話者信号より得られる疑似エコーをエコーの観測値から引いたものであるため、用いる適応アルゴリズムによって異なる。そのため、残留エコーのパワーを用いてダブルトーク検出を行う場合は、用いる適応アルゴリズムによって検出性能が異なってくるものと考えられる。

3. シミュレーション

3.1 シミュレーション条件

エコーキャンセラにおいて、ダブルトークの検出を行い、適応フィルタの係数更新を制御するシミュレーションを行った。ダブルトーク検出のために分散 $E\{u_k^2\}$ 、 $E\{y_k^2\}$ 、 $E\{e_k^2\}$ を用い、その近似 $\sigma_{u|k}^2$ 、 $\sigma_{y|k}^2$ 、 $\sigma_{e|k}^2$

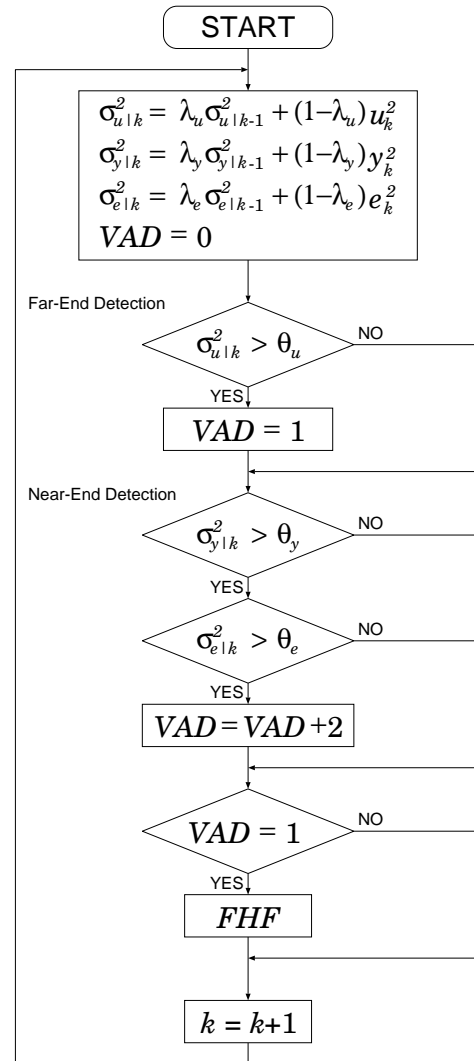


Fig. 2 閾値に基づくダブルトーク検出法

は以下のように計算される。

$$\sigma_{u|k}^2 = \lambda_u \sigma_{u|k-1}^2 + (1 - \lambda_u) u_k^2 \quad (5)$$

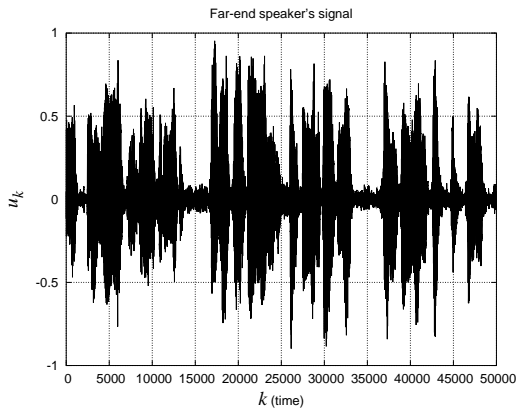
$$\sigma_{y|k}^2 = \lambda_y \sigma_{y|k-1}^2 + (1 - \lambda_y) y_k^2 \quad (6)$$

$$\sigma_{e|k}^2 = \lambda_e \sigma_{e|k-1}^2 + (1 - \lambda_e) e_k^2 \quad (7)$$

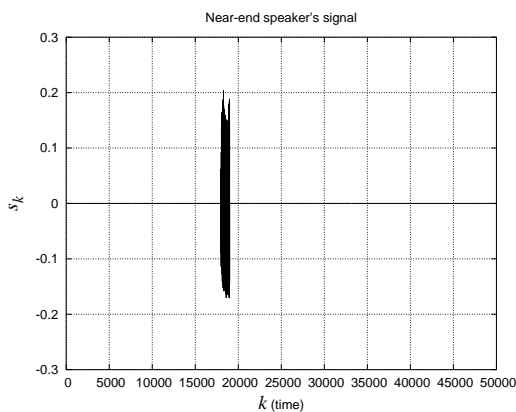
これらより図2に示したアルゴリズムを用いてダブルトークの検出を行い、ダブルトークが検出されている間は適応フィルタの係数更新を休止する。但し、最初の5000ステップはダブルトーク検出の結果によらず更新を継続する。適応アルゴリズムとしてタップ数48の正規化LMSアルゴリズム(NLMS)($\alpha = 1.0, \beta = 0.01$)及び高速 H_∞ フィルタ(FHF)($\gamma_f = 42$)を用いた場合のそれぞれについてダブルトーク検出性能を評価し比較した。

シミュレーション時に用いたパラメータは $\lambda_u = 0.99, \lambda_y = 0.99, \lambda_e = 0.99$ とした。図2に示した検出法における閾値はそれぞれ $\theta_u = 1.0 \times 10^{-5}, \theta_y = 1.0 \times 10^{-6}, \theta_e = 2.0 \times 10^{-6}$ とした。また、背景雑音 $v_k = 0$ とした。

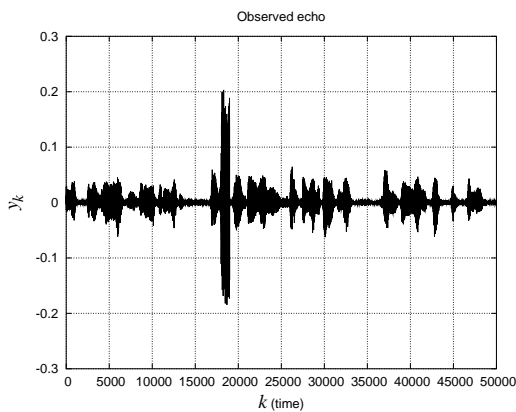
図3(a)に遠端話者信号、図3(b)に近端話者信号としてシミュレーションに用いた50000サンプルの音声データを示す。この例では、ダブルトークの発生位置は $k = 18001 \sim 19000$ となる。このときのエコーの観測値を図3(c)に示す。また、このときのエコーの観測値 $\{y_k\}$ のパワー $\{\sigma_{y|k}^2\}$ を図4に示す。



(a) 遠端話者信号 $\{u_k\}$



(b) 近端話者信号 $\{s_k\}$



(c) エコーの観測値 $\{y_k\}$

Fig. 3 シミュレーションに用いた信号

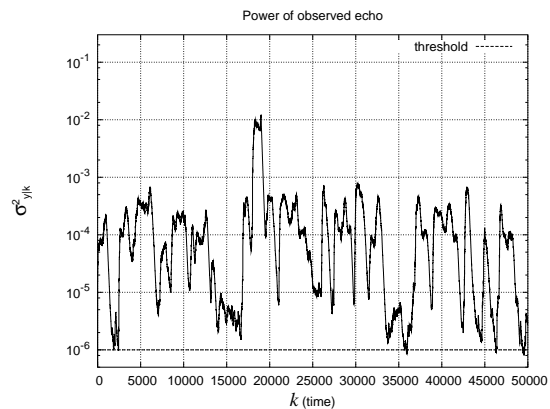
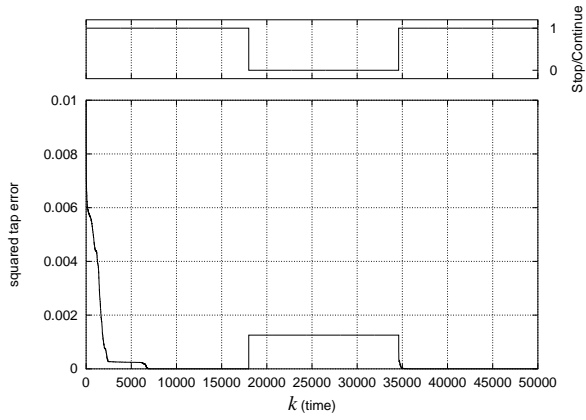
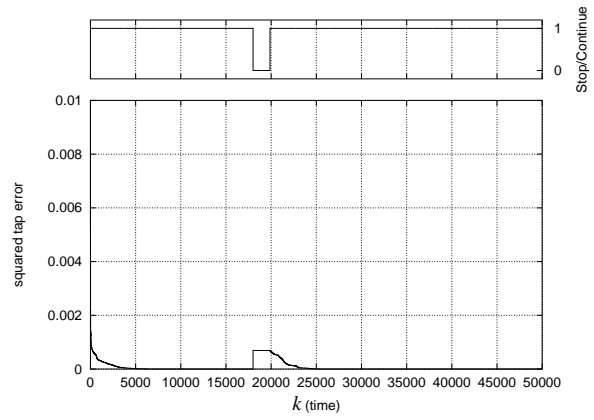


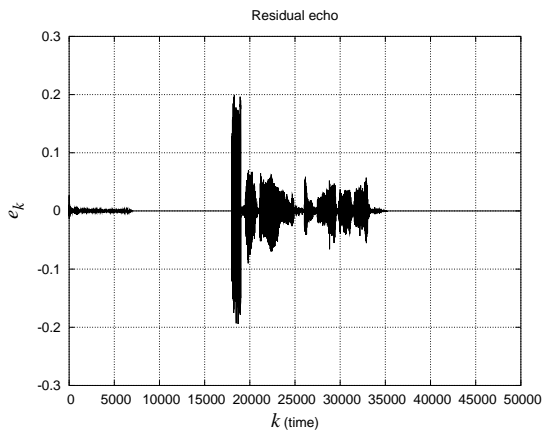
Fig. 4 エコーの観測値のパワー $\{\sigma_{y|k}^2\}$



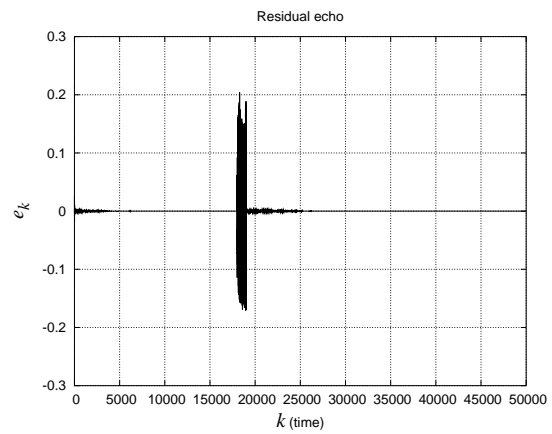
(a) ダブルトーク検出結果とタップ誤差



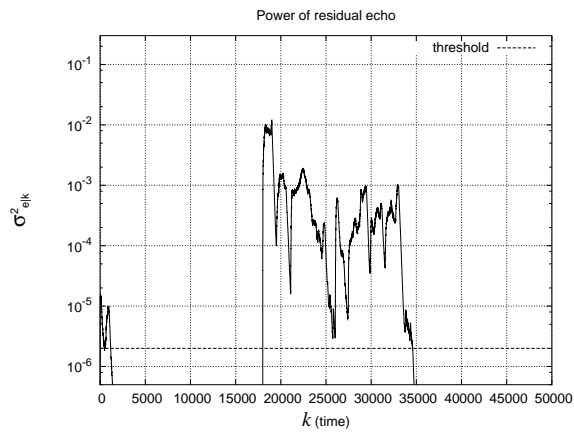
(a) ダブルトーク検出結果とタップ誤差



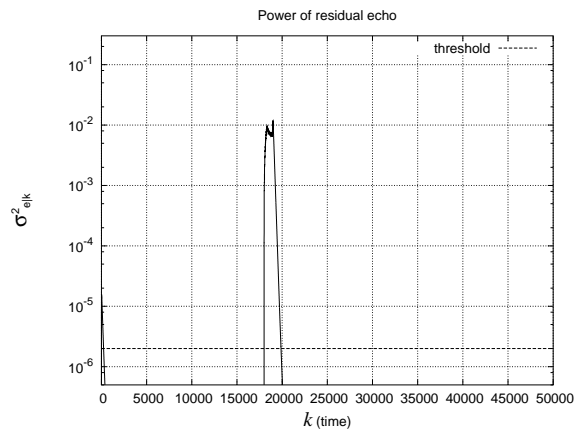
(b) 残留エコー $\{e_k\}$



(b) 残留エコー $\{e_k\}$



(c) 残留エコーのパワー $\{\sigma^2_{e|k}\}$



(c) 残留エコーのパワー $\{\sigma^2_{e|k}\}$

Fig. 5 NLMSを用いたエコーキャンセラでの制御結果

Fig. 6 FHFを用いたエコーキャンセラでの制御結果

3.2 シミュレーション結果

図5にNLMSを用いた場合、図6にFHFを用いた場合のダブルトーク検出結果、タップ誤差 $\{E_k^2\}$ 、残留エコー $\{e_k\}$ 、残留エコーのパワー $\{\sigma_{e|k}^2\}$ をそれぞれ示す。ここで、

$$E_k^2 = \|\hat{\mathbf{h}}_k - \mathbf{h}_k\|^2 \quad (8)$$

である。また、表1にダブルトーク検出された区間を示す。

Table 1 ダブルトーク検出結果

適応アルゴリズム	開始	終了
NLMS	$k = 18002$	$k = 34596$
FHF	$k = 18002$	$k = 19877$
(実際のダブルトーク)	$(k = 18001)$	$(k = 19000)$

ダブルトーク後の各アルゴリズムにおける残留エコーを図7、エコーパスのインパルス応答の真値、ダブルトーク直前とダブルトークを検出し更新を休止した時点での適応フィルタによる推定値を図8に示す。

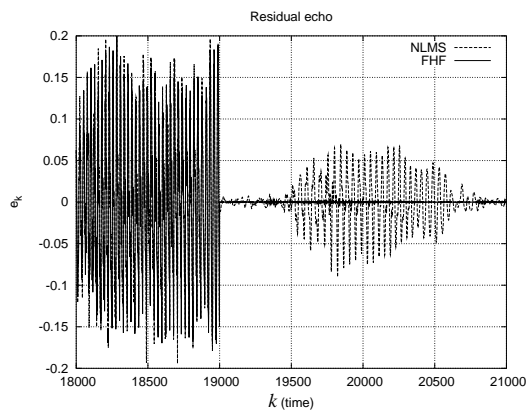
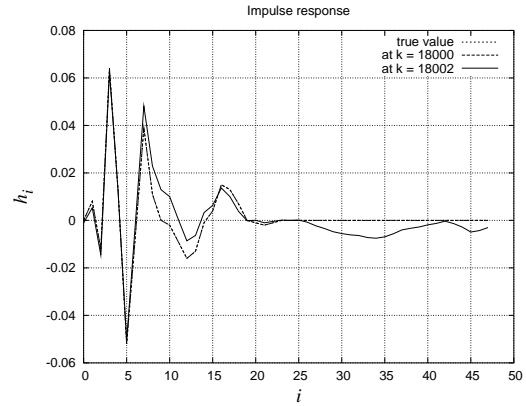
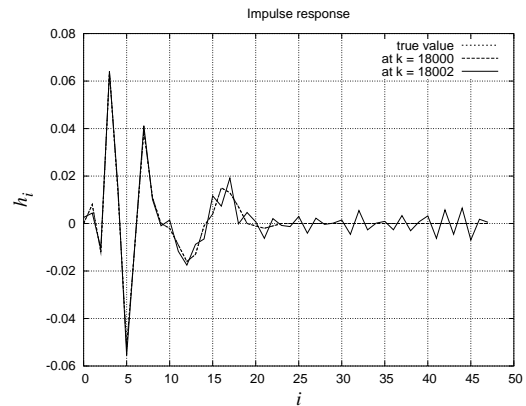


Fig. 7 ダブルトーク後の残留エコー(NLMS,FHF)

NLMSを用いた場合とFHFを用いた場合を比較すると、双方ともダブルトーク状態となってから1ステップ遅れで検出され、適応フィルタの更新を休止できている。しかし、近端話者信号がなくな



(a) NLMS



(b) FHF

Fig. 8 エコーパスのインパルス応答; 真値及びダブルトーク直前($k = 18000$)と後($k = 18002$)での適応フィルタによる推定値

りダブルトーク状態でなくなったときにNLMSを用いた場合はダブルトーク検出状態が長く続き、本来ならば適応フィルタの更新を再開しなければならないところで更新を休止し続けている。一方、FHFを用いた場合はNLMSの場合に比べ、ダブルトーク終了後の更新再開が即座に行われている。

図7を見ると、NLMSを用いた場合とFHFを用いた場合ではダブルトーク終了直後の残留エコーに大きく差が生じていることがわかる。また、図8を見ると、NLMSではダブルトーク直前と直後においてインパルス応答の推定値が大きくずれていることがわかる。これらより、NLMSを用いた場合に更新再開の遅延が生じる原因としては、ダブルトーク発生初期の検出遅延によりフィルタ係数の更新を休止できずに係数が大きく乱され、エコーパスのインパルス応答の推定値が真値から大きく離れることで、ダブルトーク終了後に本来ならばキャンセルされるべきであるエコー成分が十分にキャンセルされずに残留エコーが生じ、図2に示したアルゴリズムにより更新休止されてしまうためと考えられる。

FHFを用いた場合では、ダブルトーク検出遅延はNLMSの場合と同じであるが、観測信号への近端話者信号の混入によるタップ誤差の増加がNLMSの場合に比べて小さい。図8を見るとダブルトーク検出後におけるインパルス応答の推定値のずれが小さいことがわかる。このため、ダブルトーク状態の終了後に十分なエコー消去量が確保されており、ダブルトーク終了を素早く検出でき、フィルタ更新の更新を再開できるものと考えられる。

NLMSを用いる場合、閾値を大きめに設定することで更新再開を早めることができると考えられるが、その場合ダブルトーク検出による更新休止が遅れると予想される。これはフィルタ係数の大きな乱れを招く。

4. おわりに

エコーキャンセラにおいて、適応アルゴリズムの違いによるダブルトークの検出性能への影響をシミュレーションにより考察した。その結果、近端信号(外乱)に対して、FHFがNLMSよりロバストであり、近端信号直後の残留エコーを大幅に軽減できることを明らかにした。

参考文献

- 1) 西山 清：最適フィルタリング，培風館，2001.
- 2) 藤井 健作，大賀 寿郎：“エコー経路変動検出を併用するダブルトーク検出法”，電子情報通信学会論文誌，Vol.J78-A，No.3，pp.314-322，1995.