■2 群（画像・音・言語）－ 6 編（音響信号処理）

5 章 音響エコーキャンセラ

■2 群 － 6 編 － 5 章

5-1 通信における音響エコー

（執筆者：羽田陽一） ［2011 年 11 月受領］

通信相手がスピーカとマイクロホンを用いた拡声通話を行っている場合に，自分の話した声が時間遅延を伴って自分の耳に戻ってくることがある．この現象は，山に登ったときにほかの山から反射してくる自分の声を聞く，「やまびこ」と同じように聞こえることから「エコー」と呼ばれている．通信におけるエコーは，その発生要因から大きく二つに分類され，それぞれ回線エコーと音響エコーと呼ばれている．

回線エコーは，通信網と電話端末との間の 2 線と 4 線を変換する箇所において，インピーダンスのミスマッチによって生じる．通信網が 2 線というのは，アナログ電話網が経済性を理由に 2 本の銅線で上りと下りの通信を行っていることを指す．一方，電話端末では，音を拾うためのマイクロホンに 2 本の線（＋と－），音を再生するためのスピーカに 2 本の線（＋と－）を利用するため，合計 4 線を必要とする．回線エコーは，この 2 線を 4 線に変換する 2 線 4 線変換器で発生する．

音響エコーは，回線エコーと同様に通信時に発生するエコーではあるが，ハンドセットの代わりにスピーカとマイクロホンを用いて通信を行うハンズフリー拡声通話時に発生する．ハンズフリー拡声通信は，参加者が複数人となるような通信会議（音声会議，あるいはテレビ会議）などで主に用いられ，通信会議参加者がハンドセットをもつ必要がない，通常の対面での会議と同じように資料の閲覧やパソコンの操作ができるなど，利便性が高いため，近年，一般に利用されるようになってきている．

音響エコーの発生要因はスピーカとマイクロホンの間の音響的な結合にあり，**図 5･1** に示すように地点 A の話者 A の声が，地点 B のスピーカ B で再生された後にマイクロホン B で収音され再び地点 A に戻り，スピーカ A で再生されることで発生する．このとき，地点 A と地点 B の両地点ともスピーカとマイクロホンを用いた拡声通話である場合には，地点 A に戻ってスピーカ A で再生された声が，更にマイクロホン A で収音されて再度地点 B に戻るときがある．このような状態で通信系全体に一巡ループができあがると，音がぐるぐると回るハウリングと呼ばれる現象が発生する．

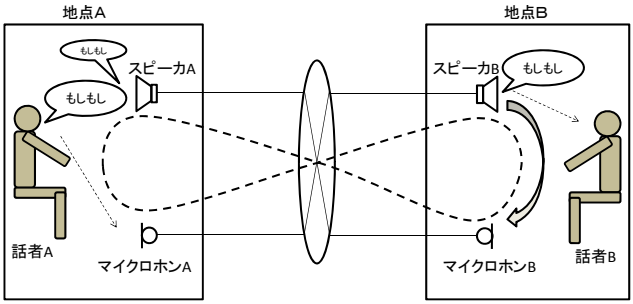


図 5･1　音響エコーとハウリング

さて，戻ってきた自分の声をエコーとして知覚するということには，エコーの大きさやエ

コーが戻ってくるまでの時間が関係している [1]．エコーの大きさが小さい，あるいは耳に戻ってくるまでの時間が短い場合には，エコーとして知覚されないが，エコーが大きい場合や，エコーが戻ってくるまでの時間が長い場合にはエコーとして知覚されやすくなる．したがって，衛星などを用いた長距離通信時や，音声を IP パケット化して送受信する VoIP 通信時など，通信における遅延が増大する場合や，スピーカとマイクロホンの距離が近い場合にエコーが問題となることが多い．

このようなエコーを防止する装置は一般にエコーキャンセラ [2] と呼ばれており，適応フィルタ [3)-5)]，音声スイッチ・エコーサプレサ [6]，エコーリダクション [7),8)] などの技術から構成されている．主に適応フィルタを用いてエコーを消去する技術 [9] または装置をエコーキャンセラと呼ぶが，広義の意味ではエコーを防止する技術全体をエコーキャンセラと呼ぶことも多い．音響エコーを消去するエコーキャンセラと回線エコーを消去するエコーキャンセラは，エコーの発生要因となるエコー経路の性質が異なるため，区別されて開発されており，前者は音響エコーキャンセラ，後者は回線エコーキャンセラと呼ばれている．以下では，対象を音響エコーキャンセラとして話を進める．

■2 群 － 6 編 － 5 章

5-2 音声を入力とする音響エコーにおける適応フィルタ

（執筆者：羽田陽一） ［2011 年 11 月受領］

音響エコーを防止あるいは消去する手法の中で，マイクロホンで収音された音の中から音響エコーのみを消去し，それ以外の伝えたい声などはそのまま送信可能とする技術として適応フィルタ技術がある．

図 5･2 に時間領域で動作する適応フィルタを用いたエコーキャンセラの原理を示す．図において，通信相手から受信してスピーカから再生する受話信号を $x(k)$，スピーカ・マイクロホン間のエコー経路のインパルス応答信号を h_n とすると，マイクロホンに回り込むエコー信号 $z(k)$は，

$$z(k)=\sum_{n=0}^{N-1} h_n x(k-n)=\mathbf{h}^T\mathbf{x}(k)$$

と表される．ここで，k, n は離散時間，N はインパルス応答長，$\mathbf{h}=[h_0, h_1, \ldots, h_{N-1}]^T$, $\mathbf{x}(k)=[x(k), x(k-1), x(k-2), \ldots, x(k-N+1)]^T$ を表す．適応フィルタを論じるときにはしばしばベクトルで表記する方が簡単な場合があるため，ここでも必要に応じてベクトル表記を併記する．適応フィルタは，エコーを消去するために擬似インパルス応答 $w_n(k)$を用意し，この $w_n(k)$と $x(k)$を畳み込むことによって擬似エコー $y(k)$を生成する．

$$y(k)=\sum_{n=0}^{L-1} w_n(k)x(k-n)=\mathbf{w}(k)^T\mathbf{x}(k)$$

ここで，$\mathbf{w}(k)=[w_0(k), w_1(k), \ldots, w_{L-1}(k)]^T$ ，L は擬似インパルス応答長であり，タップ長とも呼ばれる．擬似エコー $y(k)$をマイクロホン入力信号から差し引くことによりエコーを消去することが，適応フィルタによるエコー消去の基本である．ただし，前提として，短時間内においてインパルス応答は線形時不変であることを仮定している．

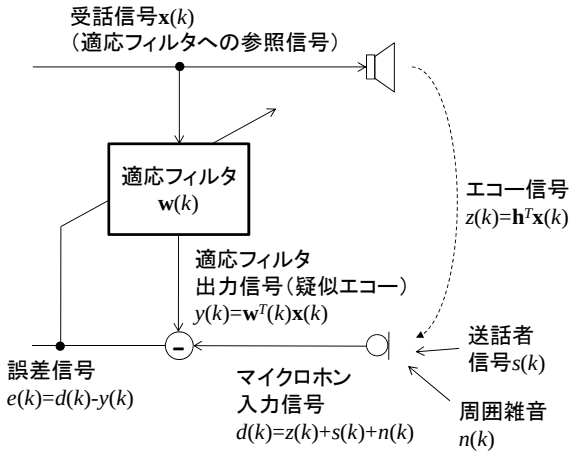


図 5･2　適応フィルタの構成例

さて，適応フィルタにおいては，$s(k) = 0$, $n(k) = 0$ であったとしても初期状態 $k = 0$ では $\mathbf{w}(0) \neq \mathbf{h}$ であるため，マイクロホン入力信号から擬似エコー信号を差し引いても誤差 $e(k)$は 0 にはならない．適応フィルタは，この誤差があらかじめ決めた規範の中で最小になるように自らフィルタ係数 $\mathbf{w}(k)$を逐次修正する．逐次修正とは，時刻 k でのフィルタ係数 $\mathbf{w}(k)$に，誤差 $e(k)$が小さくなるように算出された修正ベクトル$\Delta\mathbf{w}(k)$を加えて，新たに $\mathbf{w}(k+1)$を生成する（アップデートする）ことを指す．

$$\mathbf{w}(k+1) = \mathbf{w}(k) + \Delta\mathbf{w}(k)$$

この修正量$\Delta\mathbf{w}(k)$を計算する方法は適応アルゴリズムと呼ばれ，「学習同定法（NLMS 法）」[10]，「アフィン射影法」[11]，「RLS 法」[12]など複数の方法が知られている．これらのアルゴリズムは，一般的には誤差 $e(k) = d(k) - y(k)$ の 2 乗期待値を最小とするように動作する．ここで，$d(k)$はマイクロホン入力信号 $d(k) = z(k) + n(k) + s(k)$であり，エコー信号 $z(n)$以外に周囲雑音などの信号 $n(k)$を含む（以降，説明を簡単化するため，「**ダブルトーク問題**」の節まで $s(k)$は 0 とする）．この雑音が混入した観測信号 $d(k)$からフィルタ係数 $\mathbf{w}(k)$を推定するために，一般には以下の式で表される誤差の 2 乗期待値ε を最小化する最小自乗の規範が用いられる．

$$\varepsilon = E[e^2(k)] = E\left[\left(d(k) - \sum_{n=0}^{L-1} w_n(k)x(k-n)\right)^2\right] = E\left[\left(d(k) - \mathbf{w}^T(k)\mathbf{x}(k)\right)^2\right]$$

ここで，$E[\cdot]$ は期待値を表す．適応フィルタは期待値である集合平均（複数の試行に対する平均）を最小化することが目標であるが，実際の動作としては，ある一定の時間内で観測される誤差を最小（つまり時間的な平均を最小）にするように動作する．このとき，どれくらいの時間範囲内で誤差を最小にするかによって，代表的な三つの適応アルゴリズムを分類することができる．

5-2-1 学習同定法(NLMS)（1 時刻の誤差を最小化）[10]

修正した後のフィルタ係数 $\mathbf{w}(k+1)$によって現時刻 k の誤差を最小，つまり“0”にすることを規範にしたアルゴリズムであり，下記の式を満たす修正ベクトル$\Delta\mathbf{w}(k)$により $\mathbf{w}(k)$をアップデートする．

$$d(k) - \mathbf{w}^T(k+1)\mathbf{x}(k) = d(k) - \left(\mathbf{w}(k) + \Delta\mathbf{w}(k)\right)^T\mathbf{x}(k) = 0$$

この式を満たすΔw (k)をノルム最小の規範で求めると，

$$\Delta\mathbf{w}(k) = \frac{e(k)}{\mathbf{x}(k)^T\mathbf{x}(k)}\mathbf{x}(k)$$

となる．ここで，$e(k) = d(k) - \mathbf{w}^T(k)\,\mathbf{x}(k)$ の関係を利用した．この$\Delta\mathbf{w}(k)$を用いて，次の時刻における擬似エコーを生成するためのフィルタ係数は，

$$\mathbf{w}(k+1) = \mathbf{w}(k) + \mu\Delta\mathbf{w}(k) = \mathbf{w}(k) + \mu\frac{e(k)}{\mathbf{x}(k)^T\mathbf{x}(k) + \beta}\mathbf{x}(k)$$

として計算する．μ はステップサイズと呼ばれるもので，修正量をコントロールする値であり，0 から 1 までの正の数であり，一般的には 1 が基本である．μが小さいほど収束速度は

遅くなるが，外乱（周囲の雑音）などがあっても安定して $\mathbf{w}(k)$が求められるというメリットがある．β は $\mathbf{x}(k)$のノルムが小さく分母が 0 になることを防ぐために導入される正の数である．

5-2-2 アフィン射影法（2 時刻以上 L 時刻以下の誤差を最小）[11)]

アフィン射影法は，修正した後のフィルタ係数 $\mathbf{w}(k+1)$によって，その時刻 k から，$p-1$ 時刻まで遡って誤差 $e(k-i+1)$ $(i=1\sim p)$ を同時に 0 とすることを規範としたアルゴリズムである．つまり，

$$\begin{aligned} d(k)-\mathbf{w}^T(k+1)\mathbf{x}(k)&=0\\ d(k-1)-\mathbf{w}^T(k+1)\mathbf{x}(k-1)&=0\\ &\vdots\\ d(k-p+1)-\mathbf{w}^T(k+1)\mathbf{x}(k-p+1)&=0 \end{aligned}$$

を同時に満たす $\mathbf{w}(k+1)$を求めるアルゴリズムである．アルゴリズムの詳細は述べないが $p=1$ では学習同定法と同じであり，p は一般に射影次数と呼ばれる．

5-2-3 RLS 法（過去の観測データすべての誤差を最小化）[12)]

RLS 法は，過去の観測データすべてに遡って誤差を最小化することを理想とするが，ある程度過去のデータは忘却するためにパラメータ λ を利用し，下記の式の関係で表される 2 乗誤差和を最小化するように動作するアルゴリズムである．

$$\varepsilon=\sum_{i=0}^{k}\lambda_i\left(d(k-i)-\mathbf{w}^T(k+1)\mathbf{x}(k-i)\right)^2$$

ここでも，アルゴリズムの詳細は述べないが，RLS 法は，入力信号 $x(k)$の相関行列の逆行列を求める必要があるが，この逆行列を逐次計算する手法として知られている．

5-2-4 適応アルゴリズムの比較方法

音響エコーキャンセラで利用される適応フィルタとして三つのアルゴリズムを紹介したが，これらを含むアルゴリズムは，大きくは収束特性と演算量とによって比較される．収束特性は，一般には，ERLE（Echo Return Loss Enhancement）[13)] という指標の時間変化をグラフ化することで，定常消去量と収束速度によって評価される．ERLE は，

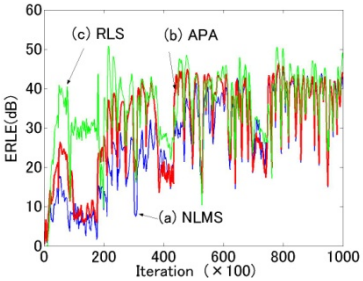
$$\mathbf{ERLE}(k)=10\log_{10}\frac{E[z^2(k)]}{E[(z(k)-y(k))^2]}\ [\mathbf{dB}]$$

として定義されるもので，マイクロホン収音信号 $d(k)$ではなく，エコーのみの信号 $z(k)$が用いられていることに注意する必要がある．つまり，この指標は，純粋にエコーがどれだけ減ったかを評価するものであるが，一方で，エコーのみを分離して測定可能な状態でないと利用できないため，通常は計算機シミュレーションの段階でのみ用いられる．また，期待値 $E[\cdot]$ は，入力信号列 $x(k)$を変えるなどして，試行を繰り返し行って平均をとることを意味するが，多くの時間を要するため，通常は，試行平均と時間平均を併用する．

$$\mathrm{ERLE}_{\mathrm{avg}}(k)=10\log_{10}\frac{\frac{1}{M}\sum_{m=1}^{M}\left\{\frac{1}{T+1}\sum_{i=-T}^{T}z_m^2(k-i)\right\}}{\frac{1}{M}\sum_{m=1}^{M}\left\{\frac{1}{T+1}\sum_{i=-T}^{T}\left(z_m(k-i)-y_m(k-i)\right)^2\right\}}\ (\mathrm{dB})$$

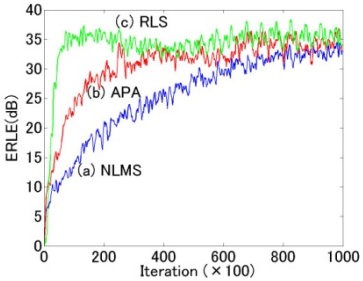
ここで，M は試行の回数を表し，T は時間平均の区間を表す．音声信号は，常にパワーが変動するため，試験に使う音声信号を変える，あるいは同じ音声信号を用いる場合には試行を開始する時刻を変えるなどして，試行平均を繰り返す必要がある．

図 5･3 に 8 kHz サンプリングの音声信号を参照信号 $x(k)$として用いて，エコー経路のインパルス応答 **h** を 512 タップの FIR フィルタとし，マイクロホン収音時にエコー信号に対して－35 dB の白色雑音を周囲雑音 $n(k)$として付加した場合の，学習同定法（NLMS），2 次のアフィン射影アルゴリズム，RLS アルゴリズムの収束特性のシミュレーション結果を示す．図において，縦軸は ERLE (dB)， 横軸は修正回数（Iteration）である．図 5･3 の ERLE は，時間平均区間を 512 サンプルとし， 1 回の試行のみで求めた結果であるため，収束速度や定常消去量の比較が正確にできない様子が分かる．次に，**図 5･4** に同じ条件で音声信号を用いて 20 回の試行を行い，時間平均区間を 512 サンプルとした場合の結果を示す．20 回の試行を行った場合には，平均化されるため，収束する様子や定常消去量が分かりやすくなり，NLMS から RLS に従って収束速度が速くなっている様子が分かる．



(a) NLMS，(b) 2 次射影法(APA)，(c) RLS

図 5・3　音声信号を用いたアルゴリズムの収束特性の比較（試行回数 1 回）



(a) NLMS，(b) 2 次射影法(APA)，(c) RLS

図 5・4　音声信号を用いたアルゴリズムの収束特性の比較（試行回数 20 回）

さて，収束速度は，横軸の修正回数に対して，どれくらいの速度で消去量が増えるかを表す量であり，エコー経路が変わった場合にどのくらいのスピードでエコーが消去されるかの目安になる．定常消去量は，エコー経路が変動しないときに，これ以上修正を繰り返しても ERLE の値が上昇しないで一定値に収束した状態のエコー消去量を指す．周囲に雑音がある場合には，雑音の影響で定常消去量が劣化する場合や，アルゴリズムによっては，収束速度は速いが定常消去量が悪い場合などがあるため，この指標が重要となる．

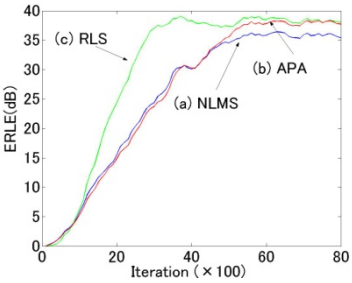
以下，複数のアルゴリズムの性能を計算機シミュレーションにて比較する際の注意事項についてまとめる．

(1) 定常消去量を一致させる

通常，異なるアルゴリズムを比較する場合には定常消去量を合わせた状態で収束速度を比較するようにしなければ厳密な比較とはならない．例えば，NLMS ではステップサイズの大きさにより定常消去量と収束速度がトレードオフの関係にある．また，エコー経路を模擬する FIR フィルタのタップ数も定常消去量と収束速度に影響する．

(2) 白色雑音を入力とした比較は音声信号入力の結果と異なる

適応フィルタへの入力参照信号 $x(k)$として，白色雑音を用いた場合の収束速度の比較を**図 5･5** に示す．白色雑音を用いた場合には，シミュレーションの試行が 1 回でも見やすい収束特性を得ることができる一方，音声信号を用いた場合とは異なる結果となってしまうことが分かる．このため，音声入力を前提とする音響エコーキャンセラの応用を目指したアルゴリズムの検討では，音声信号を用いて収束特性の評価を行う必要がある．



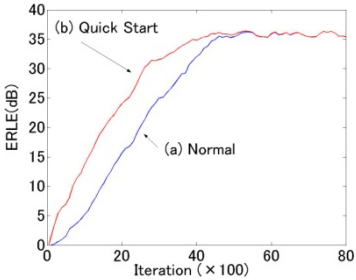
(a) NLMS，(b) 2 次射影法(APA)，(c) RLS

図 5・5　白色雑音入力に対するアルゴリズムの収束特性の比較（試行回数 1 回）

(3) 適応フィルタの動作開始時刻

収束速度を比較する場合，参照信号 $x(k)$に対して，エコーキャンセラの動作がいつ始まったかに注意する必要がある．つまり，参照信号 $x(k)$の値が，$k < 0$ では $x(k) = 0$ であるとして，$\mathbf{y}(k) = \mathbf{h}^T\mathbf{x}(k)$を k=0 から計算をはじめてシミュレーションを開始すると，例えば，エコー消去の途中でエコー経路が変動した場合の収束速度よりも速くなる傾向にある(**図 5･6**)．これを避けるためには，いったんエコー消去を実施し，収束した状態からエコー経路を変動させ（$\mathbf{h}(k)$を例えばそれとは無相関な $\mathbf{h}'(k)$に置き換える），その時点から再度収束するまでの様子から収束速度を比較する方法がある．または，シミュレーションを行うときに，エコーキャンセラのフィルタタップ数以上になるまで，$\mathbf{w}(k)$のアップデートを停止

し，アルゴリズムを空回しさせ，$x(k)$に十分値がたまってからシミュレーションを開始するという方法もある（例えば，$k = 2L$ からシミュレーションを開始する）．



(a) ある程度入力信号がたまってから開始，(b) 0 時刻以前のデータが 0 として開始

図 5・6　適応動作の開始点の違いによる収束特性の違い

(4) ERLE と MSE

収束特性の比較に ERLE ではなく，MSE（平均自乗誤差）の時間変化を比較する場合もあるが，MSE は誤差信号 $e(k)$そのものの自乗平均値をプロットするため，マイクロホンに混入した雑音 $n(k)$以下にはならない．つまり，$n(k)$よりも小さいレベルでエコーが消えているか否かを判断することができない．このため，音響エコーキャンセラとして適応フィルタを比較する場合には，MSE で比較すべきではないと考える．

5-2-5 演算量の比較

ところで，各アルゴリズムの収束速度を見てみると，NLMS が一番遅く，RLS が一番早いことが分かる．これは，過去の信号 $x(k)$をどこまで利用しているかの差となっている．一方で，フィルタタップ数 L を基準として演算量を比較してみると，NLMS を $3L$ とすると，2 次のアフィン射影法は約 $4L$，RLS は L^2 といわれている [14]．これらは主に 1 回の係数修正に行われる積和演算の量であるが，アフィン射影法や RLS では高速演算手法が開発されており，実際にはこれほど演算量は大きくならない．しかし，エコーキャンセラ装置を実現する場合には，コストを考えると利用する CPU やディジタルシグナルプロセッサー（DSP）の能力に限界があるため，それに従ってアルゴリズムが決定されることも多い．

5-2-6 収束特性の向上

さて，音声信号入力に対しては，三つのアルゴリズムの収束特性が異なる理由は何であろうか？　おおざっぱに言えば，収束速度の違いは，参照信号 $x(k)$の自己相関が原因である．自己相関の影響は，簡単には**図 5・7** に示すように白色信号と音声信号の振幅周波数特性の違いとして現れる．この自己相関が取り除かれ，白色信号に近くなれば収束速度が速くなることが知られている [15]．最も簡単な方法の一つとしてはエンファシス・ディエンファシスにより，参照信号の音声信号の周波数特性が高域下がりである特性を平坦に近づけるようなフィルタリングをしてから参照信号とするような方法がある．また，信号を周波数分割することで，各周波数帯域内では音声信号が平坦にみなせることから収束速度の向上を図るサブバン

ド方式[16]も知られている．また，フーリエ変換を用いて信号を周波数領域にしてエコーを消去するアルゴリズムも検討されている．周波数領域の適応アルゴリズムは時間領域信号を周波数領域に変換するために，時間信号をためてフレーム処理するため，時間遅延が生じるというデメリットがあるが，畳込み演算を掛け算で実行できるため，演算量を大幅に削減できるというメリットもある．一方で演算量を削減するためだけに周波数領域変換が行われる場合もあり，収束速度が時間領域と同一の周波数領域アルゴリズムもある[17),18)]．

また，収束特性を向上するためには，エコー経路を推定するための目的信号であるエコー信号がマイクロホンに対して SN 比良く収音されている場合にのみフィルタ係数のアップデートを行う方法が有効である．簡単には，参照信号のパワーがある程度以上のときにのみ，修正を実行する，SN 比に合わせてステップサイズの大きさをコントロールするなどの方法が考案されている．

その他，入力信号ではなく，同定すべきエコー経路の性質に着目したアルゴリズムとして Exponential Weighted Step-size 法が提案されており，大幅な収束速度の向上を達成している[19)]．

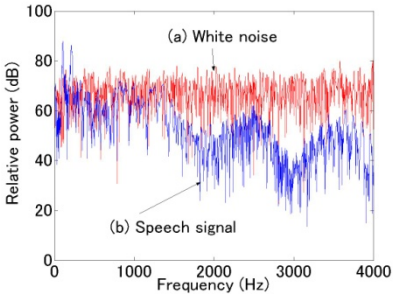


図 5・7　(a)白色雑音と(b)音声信号の振幅周波数特性の違い

5-2-7 ダブルトーク問題

収束速度を向上させる手段の一つとして，エコー信号に対する雑音信号の SN 比が高いときのみ学習することが有効と述べたが，最も SN 比が悪くなる状態として，ダブルトーク問題が知られている．ダブルトークとは，送話者信号 $s(k)$がエコー信号 $z(k)$とともにマイクロホンに混入した状態であり，この状態で学習を行うと，フィルタ係数が発散する可能性が高い．これは，適応アルゴリズムが誤差信号 $e(k)$を小さくしようとして $\mathbf{w}(k)$を修正するが，送話者信号分はどうしても小さくならないために，あらぬ方向に $\mathbf{w}(k)$を導いてしまうために起こる問題である．この問題を解決するために，適応アルゴリズムは，ダブルトークを検出する機能[20)]とともに用いられることが一般的である．また，FG/BG（Foreground/Background）構成[21)]といって，バックグラウンドではダブルトークか否かに関わらず常に学習を行い，フィルタ係数が安定してエコーを消去しているときの係数をフォアグラウンド側にコピーし，実際のエコー消去はフォアグラウンド側で行うという方法が知られている．また，SN 比に合わせてステップサイズコントロールを行う手法も有効な方法[22)]として知られている．

5-2-8 タップ数の選択

エコー経路を模擬するための FIR フィルタのタップ数は，エコー消去性能と収束速度に関連する重要なパラメータである．エコーはスピーカ・マイクロホン間のインパルス応答で発生するため，このインパルス応答をどこまで模擬するかによってエコー消去性能は決まる．例えば，インパルス応答のエネルギーが指数減衰するとみなせる場合，残響時間 T_R = 400 ms の部屋で，エコー経路の模擬を T_L = 100 ms 分の FIR フィルタで行う場合には，エコーは 15 dB しか消えない．これは，T_L = 100 ms 分のインパルス応答を完全に模擬できたとしても，ERLE が，

$$\mathrm{ERLE} = 10\log_{10}\frac{\sum_{n=T_L/T_S}^{\infty} h_n^{\,2}}{\sum_{n=0}^{\infty} h_n^{\,2}}\ [\mathrm{dB}]$$

であること（ただし，T_s はサンプリング周期），及び，シュレーダーのインパルス積分法 [23] による残響時間の計算式が，

$$-60[\mathrm{dB}] = 10\log_{10}\frac{\sum_{n=T_R/T_S}^{\infty} h_n^{\,2}}{\sum_{n=0}^{\infty} h_n^{\,2}}\ [\mathrm{dB}]$$

であることの比較から計算することができる．

一方で，適応フィルタのタップ数は，推定すべき未知数の数そのものであるため，タップ数が少ないほど，収束するまでの時間が短くなる．つまり，収束速度はタップ数に反比例するため，音響エコーキャンセラにおけるタップ数は，エコー消去量と収束までの時間を考慮しながら決定する必要があることになる．

一方，通信相手に戻るエコーの大きさは，スピーカとマイクロホンの間の音響的な結合の大きさに依存するため，同じエコー消去量でも，例えばスピーカの音量が大きければ，相手に戻るエコーは大きくなってしまう．また，スピーカとマイクロホン間の距離が近ければ，直接回り込む音が大きくなり，消去しなくてはならないエコー消去量が増えるが，一方で，インパルス応答の直接音成分のエコーを消去するだけでエコー消去量が大きくなったように見えるが，実際には後半の残響成分のエコーが通信相手に聞こえる場合がある．このように，エコー消去量は，タップ数を決める目安ではあるが，実際のシステムに照らし合わせて決定する必要がある．

■2 群 － 6 編 － 5 章

5-3 ステレオ音響エコーキャンセラ

（執筆者：羽田陽一） ［2011 年 11 月受領］

臨場感の高いテレビ会議などを提供することを目的に，近年，二つ以上のマイクロホンを用いてステレオ収音し，通信相手側でステレオスピーカから再生するというステレオ通信会議システムが商用化されつつある．このようなシステムでは，二つのスピーカと二つのマイクロホンに対応したステレオ音響エコーキャンセラが必要となり，その結果，モノラル通信とは異なる課題が生じる．最も大きな課題は，左右のステレオ参照信号に相関がある場合にエコー経路を正しく推定できないという問題である [24].

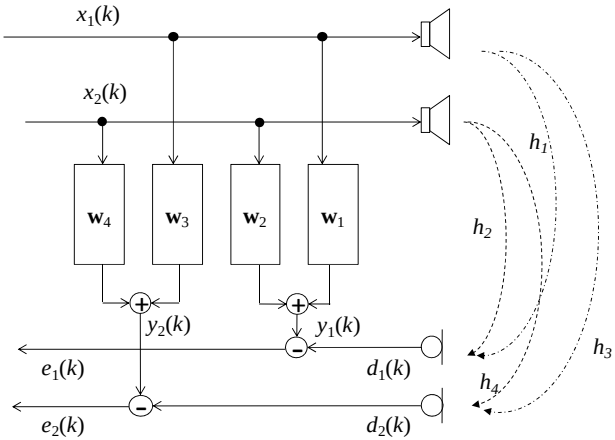


図 5・8　ステレオエコーキャンセラの構成例

いま，**図 5･8** に示すようなステレオ再生・ステレオ収音を用いたステレオ音声通信システムを考える．このとき，スピーカとマイクロホンの間の音響結合経路，すなわちエコー経路は四つ（$\mathbf{h}_1, \mathbf{h}_2, \mathbf{h}_3, \mathbf{h}_4$）存在するため，ステレオ通信用のエコーキャンセラはこの四つのエコー経路を推定し，疑似エコーを生成して，エコーを消去する．ここで，左右のマイクロホンに対するエコーキャンセラの構成は同じであるため，片方のマイクロホンにのみ着目して問題を考える．このとき，マイクロホンへ回り込むエコー信号は，d_1 側では，

$$d_1(k) = \mathbf{h}_1^T \mathbf{x}_1(k) + \mathbf{h}_2^T \mathbf{x}_2(k)$$

と表される．一方，疑似エコーを生成するための疑似エコー経路を $w_1(k)$と $w_2(k)$のフィルタ係数で表すと，疑似エコーは

$$y_1(k) = \mathbf{w}_1(k)^T \mathbf{x}_1(k) + \mathbf{w}_2(k)^T \mathbf{x}_2(k)$$

となる．いま，$d_1(k)$と $y_1(k)$が同じになれば，誤差信号 $e_1(k)$が 0 になり，エコー消去が達成されるが，ここで，ステレオ参照信号の相関が高い極端な例として，ステレオ参照信号が $x_1(k)$

$= a_1 s(k)$，$x_2 = a_2 s(k)$という場合について考える．この場合，$\mathbf{w}_1(k)$と $\mathbf{w}_2(k)$は，

$$\mathbf{w}_1(k) = \mathbf{h}_1 + a_2\mathbf{g}$$
$$\mathbf{w}_2(k) = \mathbf{h}_2 - a_1\mathbf{g}$$

であれば，任意の $\mathbf{g}$ に対して $e_1(k)$が 0 となり，エコー消去が達成できる．ただし，次の瞬間，通信相手の話者が交代するなどして，ステレオ参照信号が $x_1(k) = b_1 s(k)$，$x_2 = b_2 s(k)$に変化した場合を考える．すると，$e_1(k)$は，

$$e_1(k) = \mathbf{h}_1{}^T b_1 \mathbf{s}(k) + \mathbf{h}_2{}^T b_2 \mathbf{s}(k) - \left\{ (\mathbf{h}_1 + a_2\mathbf{g})^T b_1 \mathbf{s}(k) + (\mathbf{h}_2 - a_1\mathbf{g})^T b_2 \mathbf{s}(k) \right\}$$
$$= -a_2 b_1 \mathbf{g}^T \mathbf{s}(k) + a_1 b_2 \mathbf{g}^T \mathbf{s}(k)$$

となり，いったん誤差が 0 となり収束したように見えてはいたが，次の瞬間にはエコーが返ってくることになる．このように，ステレオ参照信号 $\mathbf{x}_1$ と $\mathbf{x}_2$ に相関がある場合には，$\mathbf{w}_1, \mathbf{w}_2$ が真のインパルス応答である $\mathbf{h}_1$, $\mathbf{h}_2$ に収束せず，参照信号の相関変動によってエコーが返ることになる．この問題は，ステレオ参照信号の相関がある程度の時間一定であり，その次の段階で相関が変化する場合に起こるが，その一方で，相関が変化するたびに，$\mathbf{w}_1, \mathbf{w}_2$ は $\mathbf{h}_1, \mathbf{h}_2$ に近づくことが知られている．このため，真のインパルス応答に収束するようにするために，ステレオ参照信号間の相互相関の変化を強調することで，収束速度を速める手法が検討されている [25]．また，積極的に人間の耳には気にならない程度に相互相関を変動させる [26] ことで，やはり収束速度の向上を図ることが研究されており，実用的な面を考えると有効な手段といえる．

また，このほかの問題としては，送信チャネル数と受信チャネル数を掛け合わせた数だけの疑似エコー経路を推定することになり，演算量も増えるうえに，収束速度が遅くなるため，高速なアルゴリズムが必要になる，チャネル間の音量の差によっても収束速度が遅くなるなどの問題 [27] がある．

■2群 － 6編 － 5章

5-4 システム化技術

（執筆者：羽田陽一）［2011年11月受領］

音響エコーを消去するための方法として適応フィルタを用いたエコーキャンセラについて説明してきたが，実際に通信会議を行う場合には，適応フィルタのみでは，十分に性能を発揮することができない場合がある．例えば，電源を入れた直後などは，フィルタ係数は収束していないので，そのままではエコーが返る場合がある．また，急にマイクロホンを移動したような場合においても，適応フィルタの収束に時間がかかるためエコーが発生したり，最悪の場合，ハウリングなどが生じる．また，適応フィルタは前述したように，スピーカ・マイクロホン間のインパルス応答を模擬することで疑似エコーを生成するが，模擬するインパルス応答の長さ，つまり準備できる適応フィルタのタップ数の制限により完璧にはエコーを消去することができず，残留エコーが発生する．更に，多くの適応アルゴリズムの場合，雑音に埋もれたエコーを消去するには，収束速度を犠牲にする必要があり，十分にエコーが消えるまで多くの時間がかかるなどの問題が生じる．これらの課題を克服するためには，適応フィルタ以外に音声スイッチやエコーリダクションを用いてエコーとハウリングを抑止することが必須となる．

5-4-1 音声スイッチ

音声スイッチ[7]は，**図5･9**に示すように受話信号と送話信号のレベルを比較し，どちらか大きい方のみの通話路を通すことで，エコーの発生を防止する機能であり，原理が単純でアナログ回路でも実現可能なため，安価なエコー・ハウリング防止装置として広く用いられている．しかし，送受話信号のどちらか一方しか通らないため，ダブルトークができない，送話信号や受話信号の検出の遅れにより，話頭や話尾が途切れるなど，通話品質に問題がある．

この問題を解決するために，どちらか一方のみを通すのではなく，ハウリングを防止する程度の損失を適応的に計算し，片方の通話路に挿入することで，話頭・話尾の欠損を極力少なくする方法や，周波数領域に分割して損失を挿入する方法などが検討されている[28]．

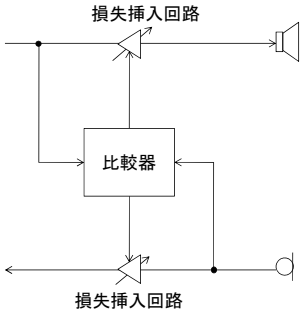


図5・9　音声スイッチの構成例

適応的に挿入損失を与える手法は，適応フィルタとともに用いられ，適応フィルタが収束していない段階では，大きな損失でエコー・ハウリングを防止し，適応フィルタが収束した

段階では，損失の挿入を停止することで，トータルとしての通話品質を保持することが可能となる．

5-4-2 エコーリダクション

エコーリダクション[9]は，適応フィルタの誤差出力信号に含まれる残留エコーを抑圧するために用いられる機能であり，適応フィルタの後段で使用されることからポストフィルタ[8]とも呼ばれている．エコーリダクションの構成例を**図 5･10** に示す．エコーリダクションの原理は，1 入力系の雑音抑圧手法であるスペクトルサブトラクション[29]やウィナーフィルタ法と同じ短時間スペクトラム振幅（Short Time Spectrum Amplitude : STSA）推定であり，ウィナーフィルタの場合，

$$G_{\mathrm{wiener}}(i,\omega)=\frac{E^2(i,\omega)-A^2(i,\omega)X^2(i,\omega)}{E^2(i,\omega)}$$

で表されるエコー抑圧ゲインを，以下の式のように周波数分析後の信号にかけ合わせることでエコーを抑圧する．

$$\hat{S}(i,\omega)=G_{\mathrm{wiener}}(i,\omega)E(i,\omega)$$

ここで，$A(i,\omega)$ は周波数ごとに推定した音響結合量（適応フィルタによってエコーが抑圧された後の音響結合量）であり，エコー信号のみの区間や，あるいはコヒーレンス関数を用いて計算する．$E(i,\omega)$ および $X(i,\omega)$ は，それぞれ適応フィルタの誤差信号 $e(k)$と参照信号 $x(k)$ の短時間スペクトルであり，i は短時間スペクトルの時刻インデックス，ω は離散周波数を表す．ここで，エコーリダクションは，適応フィルタのみでは完全にはエコーを消去することができないことを仮定しており，$E(i,\omega)$ の平均エネルギーが 0 であることは考慮していない．また，短時間スペクトルは時間信号をフレームで切り出して FFT によって求め，次のフレームはフレームサイズの 1/2 程度シフトした時点から切り出す．$\hat{S}(i,\omega)$ はエコーが抑圧された後の信号であり，周波数合成により送信信号 $\hat{s}(k)$ となるが，$e(k)$に雑音成分 $n(k)$がなく，送話者信号 $s(k)$が混在している場合には，$\hat{s}(k)\approx s(k)$，$s(k)$がない場合には，$\hat{s}(k)\approx 0$ となる．

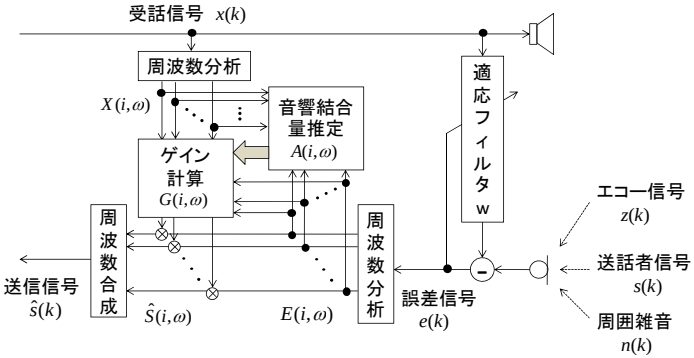


図 5・10　エコーリダクションの構成例

STSA に基づくエコー抑圧方法は，エコー信号の振幅スペクトルのみを推定するため，位相も含めて推定する適応フィルタに比べてインパルス応答の変動に対し，エコー抑圧性能があまり変化しないという利点をもつ．一方で，振幅のみの推定であるため，出力される信号の位相は残留エコーが混在していたときの位相を用いており，主通話信号が歪むといった欠点もある．更に，音響結合量の推定誤差の影響でも信号が歪みやすい．このため，エコーリダクションは単体の機能として用いられることは少なく，適応フィルタの後段である程度小さくなった残留エコーを更に抑圧するために用いられることが多い．

■参考文献

1) H. Yasukawa, M. Ogawa, M. Nishino, “Echo return loss required for acoustic echo controller based on subjective assessment,” IEICE Trans. on Fundamentals of Electronics, Communications and Computer Sciences, vol.E74-A, no.4, pp.692-705, 1991.
2) E. Hänsler, G. Schmitdt, “Acoustic echo and noise control,” Wiley-IEEE Press, 2004.
3) S, Haykin, “Adaptive filter theory,” Prentice-Hall Prentice Hall, 2001.
4) シモン ヘイキン(著)，武部 幹(訳), “適応フィルタ入門”, 現代工学社, 1987.
5) B. Widrow, S. Stearns, “Adaptive signal processing,” Prentice-Hall, 1985.
6) ITU-T Recommendation G.165.
7) S. Gustafsson, R. Martin, P. Jax, P. Vary, “A psychoacoustic approach to combined acoustic echo cancellation and noise reduction,” IEEE Trans. Speech Audio Processing, vol.10, pp.245–256, 2002.
8) 阪内澄宇, 羽田陽一, 田中雅史, 佐々木順子, 片岡章俊, “雑音抑圧及びエコー抑圧機能を備えた音響エコーキャンセラ,” IEICE vol.J87-A, no.4, pp.448-457, 2004.
9) M. M. Sondhi, “An adaptive echo canceller,” Bell System Technical J. 46 (March), pp. 497-511, 1967.
10) 野田淳彦, 南雲仁一, “システムの学習的同定法,” 計測と制御, vol.7, no.9, pp.597-605, 1968.
11) 尾関和彦, 梅田哲夫, “アフィン部分空間への直交射影を用いた適応フィルタ・アルゴリズムとその諸性質”, 電子情報通信学会論文誌, vol.J67-A, no.2, pp.126-132, 1984.
12) R. L. Plackett, “Some theorems in least squares,” Biometrika, 37, pp.149-157, 1950.
13) ITU-T Recommendation G.168.
14) 大賀寿郎, 金田 豊, 山崎芳男, “音響システムとディジタル処理,” 電子情報通信学会, 1995.
15) S. Yamamoto, S. Kitayama, J. Tamura, H. Ishigami, “An adaptive echo canceller with linear predictor,” IEICE vol.E62-E, no.12, pp.851-857, 1979.
16) H. Yasukawa, S. Shimada, “An acoustic echo canceller using subband sampling and de-correlation method,” IEEE Trans. Signal Processing, vol.41, no.2, pp.926-930, 1993.
17) J. Benesty, P. Duhamel, “A fast exact least mean square adaptive algorithm,” IEEE Trans. Signal Processing, vol.40, no.12, pp.2904-2920, 1992.
18) M. Tanaka, S. Makino, Y. Kojima, “A block exact fast affine projection algorithm,” IEEE Trans. Speech Audio Processing, vol.7, pp.79-86, 1999.
19) S. Makino, Y. Kaneda, N. Koizumi, “Exponentially weighted step-size NLMS adaptive filter based on the statistics of a room impulse response,” IEEE Trans. Speech Audio Processing, vol.1, no.1, pp.101-108, 1993.
20) D. L. Duttweiler, “A twelve-channel digital echo canceler,” IEEE Trans. Comm., vol.26, no.5, 1978.
21) K. Ochiai, T. Araseki, T. Ogihara, “Echo canceler with two echo path models,” IEEE Trans. Communications, vol.COM-25, no.6, pp.589-595, 1977.
22) S. Yamamoto, S. Kitayama, “An adaptive echo canceller with variable step gain method,” IEICE, vol.E65-E, no.1, pp.1-8, 1982.
23) M. R. Schroeder, “A new method of measuring reverberation time,” J. Acoust. Soc. Am., vol.37, pp. 409-412, 1965.
24) M. Sondhi, D. Morgan, J. Hall, “Stereophonic acoustic echo cancellation —An overview of the fundamental problem,” IEEE Signal Processing Lett., vol.2, pp.148-151, 1995.
25) J. Benesty, F. Amand, A. Gilloire, Y. Grenier, “Adaptive filtering algorithms for stereophonic acoustic echo

cancellation,” in Proc. IEEE Int. Conf. Acoustics, Speech, Signal Processing, pp.3099-3101, 1995.
26) S. Shimauchi, S. Makino, “Stereo projection echo canceller with true echo path estimation,” in Proc. IEEE Int. Conf. Acoustics, Speech, Signal Processing, pp.3059-3062, 1995.
27) 中川　朗, 羽田陽一, “ステレオ信号間のパワー差を考慮したステレオエコーキャンセラに関する一検討,” IEICE, vol.J86-A, no.10, pp.989-997, 2003.
28) 一ノ瀬裕, “周波数帯域分割による音声スイッチの等価挿入損失の軽減について,” 日本音響学会誌, vol.52, no.12, 1996.
29) S. F. Boll, “Suppression of acoustic noise in speech using spectral subtraction,” IEEE Trans. Acoust., Speech, Signal Process., vol.27, pp.113-120, 1979.