

1 . はじめに

マルチメディアの時代における人間と音響・音場インタ - フェースではこれまでのような電話機のハンドセットやヘッドホンに拘束されない自由なインターフェースが必要となる。例えばマイクロホンとスピーカを用いて通話を行う通信会議システムなどがあげられる。その通信会議においてあたかも同一空間にいるかのように会話ができることや、マイクロホンやスピーカの位置を意識しないで会話ができるなど距離や場所を意識しないシームレスな音響環境の実現が課題となっている。しかしその為には、物体の影響を考慮した送話・受話の音響設計や音響フィードバック（エコー）の抑圧が必要である⁽¹⁾。本論文では後者である音響フィードバック（エコー）の抑圧の目的として使われている音響エコーキャンセラの構成要素である、適応フィルタの基本特性を評価する。具体的には、matlab によるプログラムを作成し、ハンドフリー電話機やデスクトップ会議、テレビ会議装置等、拡声通話系を有する通信機器の異なる使用形態や設置音響環境においてのシュミレーションや実測をもとに、適応フィルタの基本特性の評価を行う。

2 . 基本事項

2 . 1 音の大きさ⁽²⁾

まず初めに、音の大きさをどのように数値表現するか説明する。音の大きさを数量化するとき、正弦波であれば、振幅でも比較できるが、一般波形の場合、振幅が不規則であり、比較することはできない。そこで、通常はパワーで定量化する。パワー P は音圧 $p(t)$ の二乗平均値である。

$$P = \frac{1}{T} \int_0^T p(t)^2 dt \quad (1)$$

また、感覚に対応した単位として利用されるのが、[dB]である。[dB]はパワーと基準量を比較した相対値を計算し、常用対数をとった値 $10 \log_{10}$ を 10 倍した値である。10 倍する理由は、そのままでは、小数点以下が多く、使いにくいいためである。人は音のパワーが 10 倍となったのと、100 倍となったのでは、同じような変化を感じる。[dB]では、0[dB] 10[dB]の変化と、10[dB] 20[dB]という変化を表す。

以下、音の大きさを表す単位として、[dB]を用いていく。

2 . 2 インパルス応答⁽²⁾

次に、インパルス応答について説明する。アナログ系のインパルス応答は、幅が 0 で高さが 1、積分値が 1 のパルス（インパルス信号 $\delta(t)$ ）を入力した時の系の出力である。また、デジタル系では、時間 0 で値 1、その他の点では値 0 の信号（単位サンプル信号 $\delta(k)$ ）に対する応答 $g(k)$ をインパルス応答という。デジタル系の単位サンプル信号 $\delta(k)$ はアナログ系のインパルス信号 $\delta(t)$ と等価である。インパルス応答は、線形系の全ての情報をふくんでおり、インパルス応答がわかれば、あらゆる入力に対する出力がわかる。また、インパルス応答のフーリエ変換は、その線形系の周波数応答特性となる。

2 . 3 畳み込み⁽²⁾

次に畳み込み演算の説明をする。パルス信号 $\delta(t)$ を使うと任意の $x(t)$ は次のように表現できる。

$$x(t) = \int_{-\infty}^{\infty} x(\tau) \delta(t - \tau) d\tau \quad (2)$$

この表現を使い、線形系 T に対する入力 $x(t)$ の出力を $T[x(t)]$ と表すと、その

関係は次のようになる。

$$\begin{aligned}
 y(t) &= T[x(t)] = T\left[\int_{-\infty}^{\infty} x(\tau)\delta(t-\tau)d\tau\right] \\
 &= \int_{-\infty}^{\infty} T[x(\tau)\delta(t-\tau)]d\tau = \int_{-\infty}^{\infty} x(\tau)T[\delta(t-\tau)]d\tau \\
 &= \int_{-\infty}^{\infty} x(\tau)h(t-\tau)d\tau \quad (3)
 \end{aligned}$$

ここで $h(t) = T[\delta(t)]$ であり、シフト不変性を仮定すれば、

$$h(t) = T[\delta(t)] \quad (4)$$

上式の定義からわかるように、 $h(t)$ はこの系のインパルス応答である。

$t \rightarrow -t$ と変換すれば、

$$\int_{-\infty}^{\infty} x(\tau)h(t-\tau)d\tau = \int_{\infty}^{-\infty} x(t-\theta)h(\theta)(-d\theta) = \int_{-\infty}^{\infty} x(t-\theta)h(\theta)d\theta \quad (5)$$

上式から、線形系の入出力はその系のインパルス応答によって完全に表現できることがわかる。

系の入力 $x(t)$ と出力 $y(t)$ の関係が (3) 式または (5) 式によって表される
とき、出力 y を、入力 x と系のインパルス応答 h との畳み込みといい、次の
ように表す。

$$y = x * h \quad \text{または} \quad y = h * x \quad (\text{ベクトル演算では、} y = x^T h) \quad (6)$$

畳み込み表現においては畳み込まれる関数の順序は重要ではない。

3. 音響エコーキャンセラ

図1にモデル化した拡声通話と音響エコーキャンセラの関係を示す。

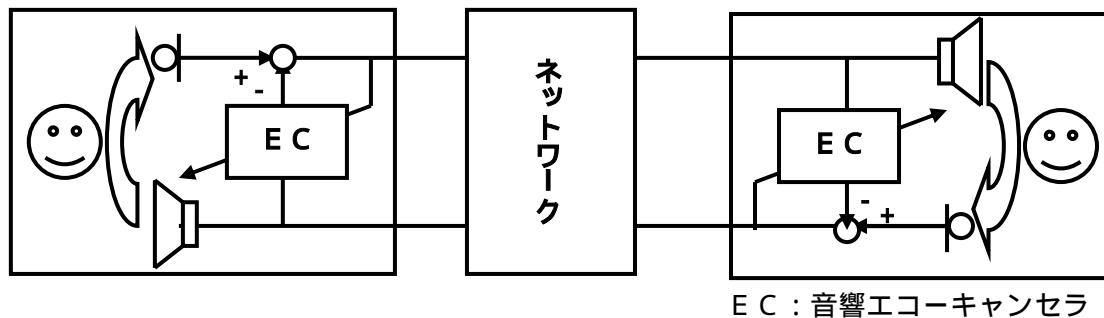


図1：拡声通話のモデルと音響エコーキャンセラ

拡声通話をする場合、マイクロホンから入った自分の声はネットワークを経由して相手のスピーカから出ることになる。いったん相手側に出た声は相手のマイクロホンに拾われ、ネットワークを経由して再びこちら側のスピーカから出る。つまり、自分がしゃべる声に戻ってくることによるエコーが生じる。スピーカから出た自分の声は、再び自分のマイクロホンで拾われるので拡声通話系の利得設定によってはエコーだけでなく、ハウリングを生じる場合もある。音響エコーキャンセラは、このようなスピーカからマイクロホンに回り込む音を消去する装置である⁽¹⁾。

音響エコーキャンセラの原理は図2のように示される。スピーカに入力される受話信号を x 、マイクで受音されるエコー信号を d とするとこれらは室内のインパルス応答 g を用いて $d = x * g$ ($*$ は畳み込み演算) と関係づけられる。そこで音響エコーキャンセラはこの g の推定値 h を求め、これにより推定エコー信号 y を作り、それをマイクの受音信号から差し引くことで音響帰還を防止する⁽³⁾。

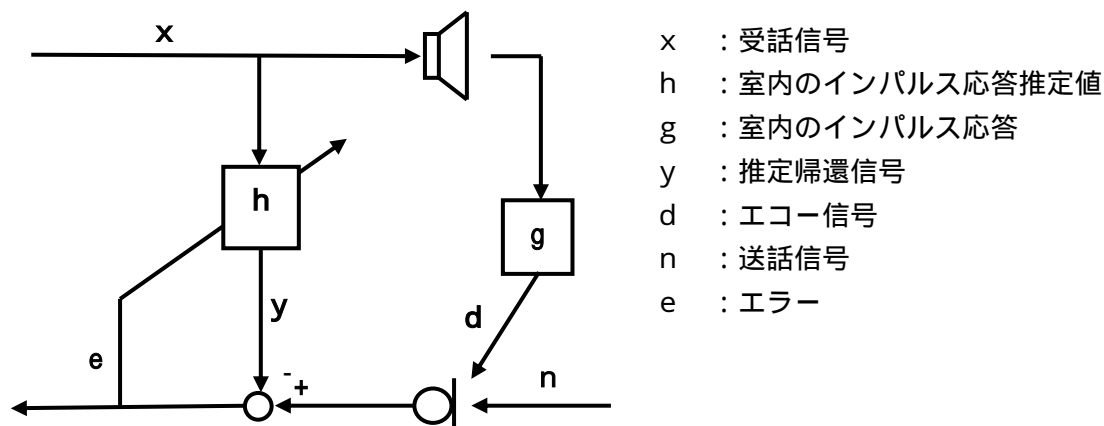


図2：音響エコーキャンセラ - の原理

このような動作は制御理論におけるシステム同定そのものであるが、次のような点が音響エコーキャンセラの技術を難しくしている。まず、第一は系の次数が高いことである。後で詳しく説明は書くが、 h は通常、FIR フィルタにより実現される。例えば、通常の会議において 7kHz 帯域で 40dB 程度のエコー消去量を得ようとする、6000 次以上ものフィルタが必要となる。その結果、大きな規模のハードウェアが必要となる。第二は、 g を推定するための入力信号 x が、高い自己相関を持つ音声信号であるという点である。このことはフィルタ次数が高いことも関連して、 g の推定速度を遅いものとしている。その他にも送話音声 n が推定の妨害要因となること(同時通話の問題)など種々の問題点がある。

4 . 適応フィルタ⁽⁴⁾

室内のインパルス応答は、人の移動や扉の開け閉めなどによって時間的変動を伴う為、その推定に適応フィルタが用いられる。 k を離散的時間を表わす変数とした時 $e(k)$ は、

$$e(k) = d(k) - y(k) \quad (7)$$

と表わされ、適応フィルタでは、図 2 の e のパワー (誤差パワー) を最小にするように推定値 h を逐次的に修正する。

また、フィルタ特性の修正手順は、適応アルゴリズムと呼ばれ、さまざまな手法がある。

適応フィルタの内部は、図 3 に示すような入力信号 $d(k)$ をフィルタリングして出力信号 $y(k)$ を生成するフィルタ部と、そのフィルタ特性を適応アルゴリズムに基づいて修正する適応部から構成される。フィルタ部には一般的に FIR フィルタが使われていて、適応部で求められた係数を $h(k)$ とすると、

$$y(k) = x(k)^T h(k) \quad T: \text{転置} \quad (8)$$

と表わすことができる。また、適応部では $e(k)$ のパワーが小さくなる様に適応アルゴリズムによって $h(k)$ が、計算される。

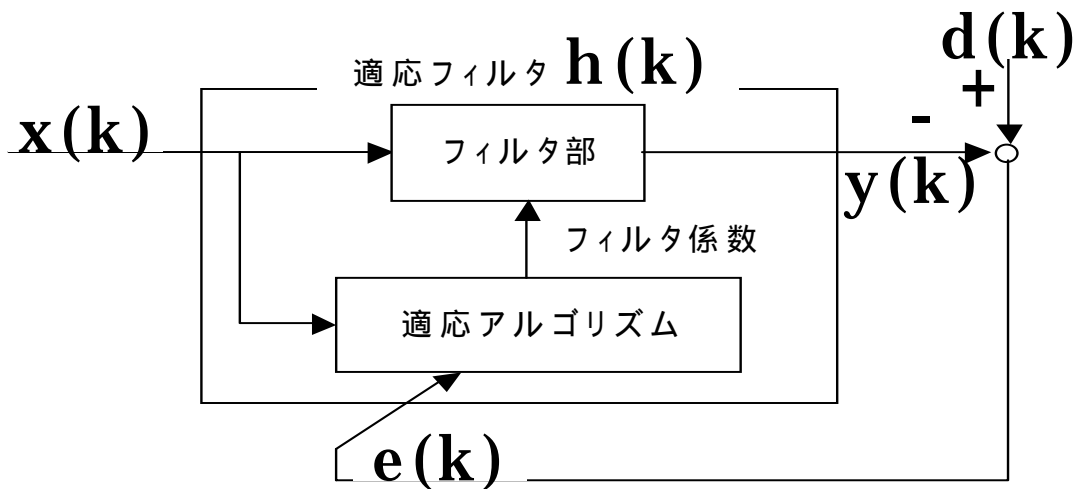


図 3 : 適応フィルタの内部構成

5 . 適応アルゴリズム⁽⁴⁾

適応アルゴリズムとは、各時間において観測される入力 $x(k)$ と誤差 $e(k)$ に基づいて、係数 $h(k)$ を最適係数 h_0 に近づけていく修正手順である。適応フィルタの動作開始時刻 $k=0$ において、係数には適当な初期値（例えば、 $h(0)=0$ ）が与えられる。適応アルゴリズムは各時刻 k において、式(9)に示すような修正式で係数 $h(k)$ を $h(k+1)$ に修正し、少しずつ最適係数に近づけていくのである。

$$h(k+1) = h(k) + \delta h(k) \quad (9)$$

ただし $h(k)$ は、 L 次のベクトルである。また、アルゴリズムによっては、修正量の大きさを制御するためのスカラー量 α を導入して、

$$h(k+1) = h(k) + \alpha \cdot \delta h(k) \quad (10)$$

と表わす場合もある。この α は時間の変化によらない一定値で、ステップサイズ (step size) と呼ばれている。(9) 式、(10) 式の係数修正は、時刻 k において観測される入力 $x(k)$ と誤差 $e(k)$ に基づいて、またアルゴリズムによっては、それまでに観測された入力 $x(k-1), x(k-2), \dots$ と誤差 $e(k-1), e(k-2), \dots$ にも基づいて行われる。そして、計算された $h(k+1)$ と $x(k+1)$ および $d(k+1)$ に基づいて、時刻 $k+1$ の誤差 $e(k+1)$ が次式の様決定される。

$$e(k+1) = d(k+1) - x(k+1)^T h(k+1) \quad (11)$$

(9) または (10) 式の修正は、厳密に言えば k と $k+1$ の間において行われ

るものであるが、本論文では「時刻 k における修正」と呼ぶことにする。
以上より適応アルゴリズムの一般的な手順は以下のようなものとなる。

時刻 $k=0$ として、初期値 $h(0)$ を設定する。

時刻 k における誤差 $e(k)$ を次式により計算する

$$e(k) = d(k) - x(k+1)^T h(k) \quad (12)$$

修正ベクトル $h(k)$ を計算し、(9) または、(10) 式に基づいて $h(k+1)$ を計算する。

k の値を 1 つ増やして上記の 手順を繰り返す。

このうち 手順については、適応フィルタは、 $x(k)^T h(k) = y(k)$ だけを計算し、 $e(k)$ は、外部の系によって計算される場合も多い。しかしシミュレーションの場合やシステムの構成によって $d(k)$ のみが与えられる場合には $e(k)$ を計算によって求める必要がある。また上記の手順のうち、 については各アルゴリズム共通の手順である。このあとに書かれる学習同定法の説明は、上記 手順のみを説明する。適応アルゴリズムは修正ベクトル $h(k)$ を計算する際に、どの程度さかのぼって過去の入力および誤差の情報を利用するかという点から分類することができる。修正直前の入力 $x(k)$ と誤差 $e(k)$ のみを用いる手法としては、LMS アルゴリズム、学習同定法などが知られている。またタップ数を L とし、 L 個以下の過去の信号を利用する方法として射影アルゴリズムが知られ、過去すべての信号に重みをつけて利用する方法の RLS アルゴリズムなどが知られる。

6 . 学習同定法⁽⁴⁾

このアルゴリズムは LMS アルゴリズムと比較した時、演算量は多少増加するが、入力信号の大きさによって収束特性が変化するという問題点がなく、扱いやすいアルゴリズムとなっている。そのため入力信号の大きさが時間的に変動するような場合には、主としてこのアルゴリズムが使用される。

では、学習同定法の動作原理を説明し、アルゴリズムを導いてみよう。図4において、雑音 $n(k)$ が十分に小さいものと仮定する。この時 $d(k)$ は、最適フィルタに $x(k)$ を入力した時の出力とに考えることができる。

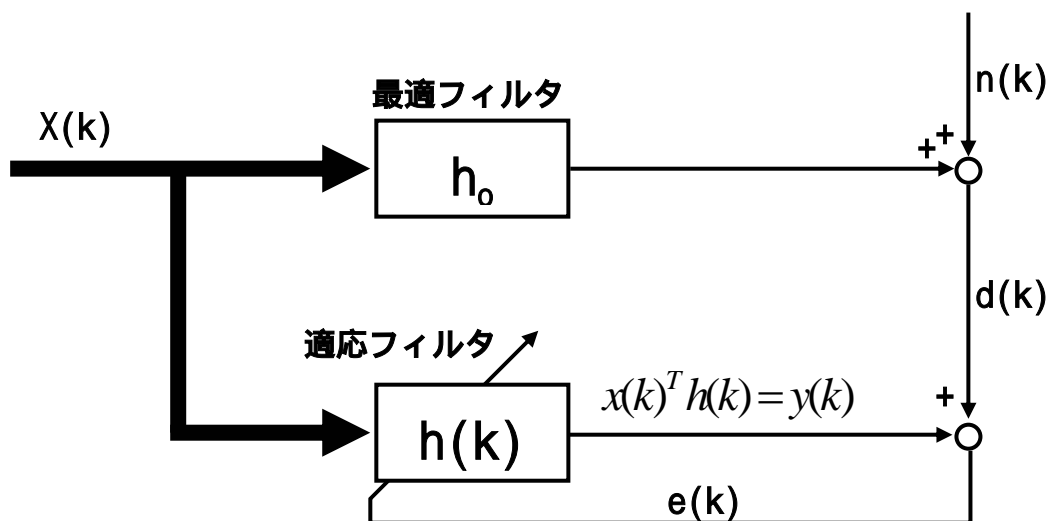


図4：適応フィルタの同定モデル

そこで、最適フィルタも $x(k)$ を入力した時の出力が $d(k)$ となるように係数を修正する、というのが学習同定法の基本的な考え方である。すなわち、修正後の係数 $h(k+1)$ が次式の関係を満たすように係数を修正する。

$$x(k)^T h(k+1) = d(k) \quad (13)$$

時刻 k における係数の修正式が、式(9)と表わせるものとし、式(9)を式(13)に代入して整理すると、

$$x(k)^T \delta h(k) = d(k) - x(k)^T h(k) = e(k) \quad (14)$$

が得られる。

これより、学習同定法の修正ベクトル $h(k)$ は、(14)式を満たすように決定される。しかし、(14)式の解は未知数が L 個あるのに対して、満たすべき方程式の数は1個であるため、その解は無数に存在する。そこで、学習同定法では、(14)式を満たす解の中で、修正量の大きさ $\|h(k)\|^2$ が最小となるものを選び出すと、

$$\delta h(k) = \frac{e(k)x(k)}{\|x(k)\|^2} \quad (15)$$

ここで修正量の大きさを制御するステップサイズ α を導入し、また $x(k) = 0$ の時、式 (1 5) の分母項が 0 になって、アルゴリズムが発散する事を防ぐために十分小さい値のパラメータ ϵ を導入して、係数の修正式は、

$$h(k+1) = h(k) + \frac{\alpha \cdot e(k)x(k)}{\epsilon + \|x(k)\|^2} \quad (1 6)$$

となる。

7. シミュレーションによる適応フィルタの基本特性評価

1) 目的

適応フィルタは、収束速度の速さと、どれだけエコーを消せるか (ERLE) が求められる。学習同定法において、最適な適応フィルタを構成するためには、ステップサイズ とフィルタの長さ L を変化させると、それらの条件がどのように変化するかを理解する必要がある。そのために、シミュレーションを用いて、適応フィルタの性能を評価する。

2) 概要

まず初めに、シミュレーションによる適応フィルタの性能評価を行う。下にそのブロック図を示す。

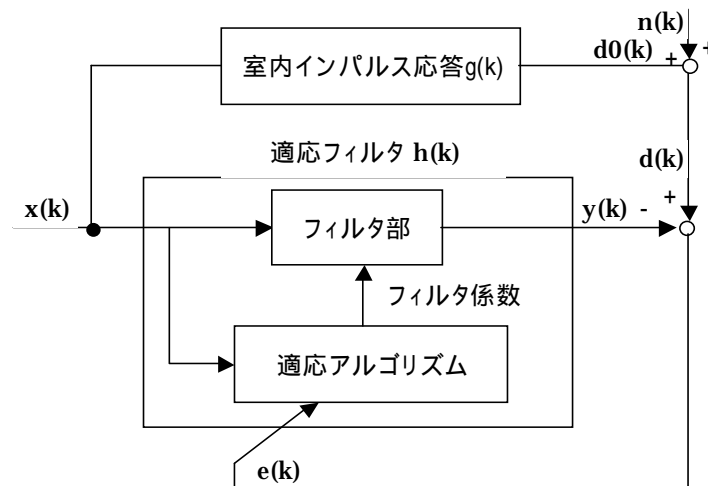


図5：シミュレーションのブロック図

室内のインパルス応答 g を測定し、ある信号 x と畳み込んだものに雑音 $n(k)$ を加算したものを目的信号 d とする。ある信号 x を適応フィルタに通し、推定した y を差し引き、その誤差 e のパワーを測定し、ステップサイズ、フィルタの長さ L を変化させた場合における、誤差 e のパワー (ERLE) の変化について検討する。また、インパルス応答 g の残響時間を換え、その時の比較を行う。ただし ERLE (Echo Return Loss Enhancement : エコー低減量) とは、 $e(k) - y(k)$ と $d0(k)$ のパワー比である。 $e(k) - n(k)$ は、(17) 式のように表される。

$$\begin{aligned} e(k) - n(k) &= d(k) - n(k) - y(k) \\ &= d0(k) - y(k) \end{aligned} \quad (17)$$

以下に、使用したプログラムを示す。

```

% xx の合成
Fs=8000;
% サンプリング周波数
k=4*Fs;
% データ xx の長さ
xx=randn(k,1);
% 信号 xx (白色雑音) の合成

% ----- 未知系のインパルス応答 gg -----
% gg (部屋のインパルス応答)を読み込む
[gg0,Fs]=wavread('D:\¥imp¥imp11.wav');
Lg=500;
% インパルス応答の有効長
gg=gg0(291:290+Lg);
% インパルス応答を短縮 (演算量軽減)

% ----- 未知系の出力 dd -----
% 実測した雑音重畳音声 dd の読み込み
% xx と gg を畳み込んで dd を合成

dd=conv(xx,gg);
% xx と gg のたたみこみ
Pd=dd'*dd/length(dd);
% dd のパワー (二乗平均値) の計算
SNR=60;
% dd の SN 比
nnn=randn(length(dd),1)*10^(-SNR/20)*sqrt(Pd);
% SN 比 SNR で雑音を付加
dd=dd+nnn;
% SN 比 SNR で雑音を付加

% ----- 適応フィルタ hh の設定 -----
L=500;
% フィルタの長さを決定
hh=zeros(L,1);
% 適応フィルタ係数の初期化 (全ての要素が 0 の L*1 の行列を作成)

% ----- パラメータの初期化、など -----
% 全ての要素が 0 の 1*k の行列を作成

yy=zeros(k,1);
% 全ての要素が 0 の k*1 の行列を作成
ee=zeros(k,1);
% 全ての要素が 0 の k*1 の行列を作成
ee1=zeros(k,1);
% 全ての要素が 0 の k*1 の行列を作成
alp=1;
% ステップサイズの設定

% ----- 適応フィルタの動作開始 -----
Nrep=length(xx)-L;
% 適応フィルタの動作時間長
for ip=1:Nrep;

```

```

% ip が 1 ~ Nrep まで繰り返す
xxx=xx(ip:ip+L-1);
% L 次の入力ベクトル xxx
yy(ip)=hh'*xxx;
% 信号 xx を適応フィルタ hh にかけて、出力 y を計算する
ee(ip)=dd(ip+L-1)-yy(ip);
% 誤差 ee の計算
ee1(ip)=ee(ip)-nnn(ip+L-1);
hh=hh+alp*ee(ip)*xxx/(xxx'*xxx);
% 誤差をもとに、学習同定法で適応フィルタを修正する。
end

% 誤差の減少のをグラフにして確認

figure(1)
plot(gg)
hold on
plot(flipud(hh), 'r')
% 推定された未知系のインパルス応答の図示
title('推定した未知系のインパルス応答')

Pee=ee.*ee/Pd;
% 誤差 ee の二乗と正規化
Pee1=ee1.*ee1/Pd;
Pee2=conv(Pee,ones(100,1))/100;
% 誤差を平均する (平坦化)
Pee3=conv(Pee1,ones(100,1))/100;

figure(2)
plot(10*log10(Pee2+10^(-10)));
% 誤差パワーをプロット
axis([0,31000,-140,20])
% グラフの範囲を設定
grid
% 目盛りをつける
ylabel('誤差パワー [dB]')
title('誤差パワーの減少')
hold on

figure(3)
plot(10*log10(Pee3+10^(-10)), 'b');
% 誤差パワーをプロット
axis([0,31000,-70,5])
% グラフの範囲を設定
grid
% 目盛りをつける
ylabel('誤差パワー [dB]')
title('誤差パワーの減少')
hold on

return

```

3) 条件

室内のインパルス応答 g の残響時間は、
imp11 が、0.05[s]、imp12 が、0.1[s]、imp13 が、0.15[s]の三種類のものを用いた。

インパルス応答の波形は、以下のようなになる。

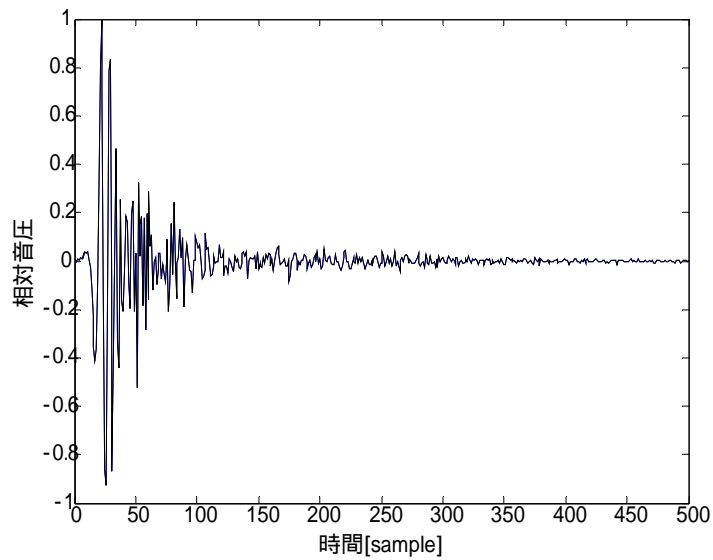


図 6 : imp1 1 (残響時間 0.05[s])

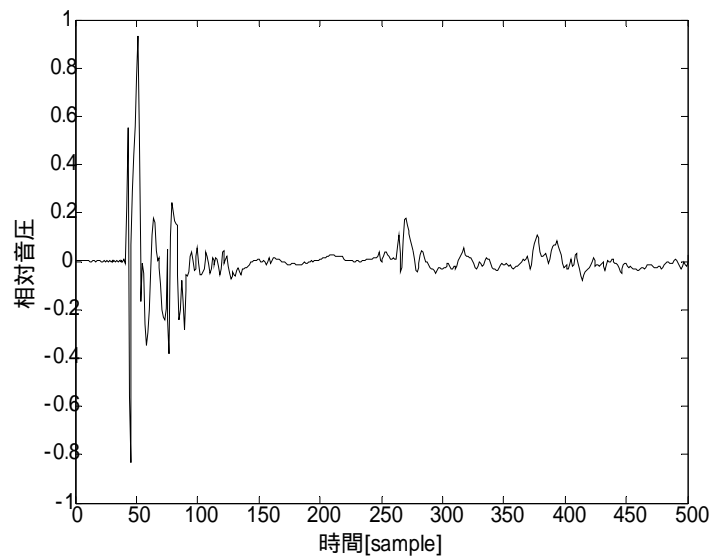


図 7 : imp12 (残響時間 0.1[s])

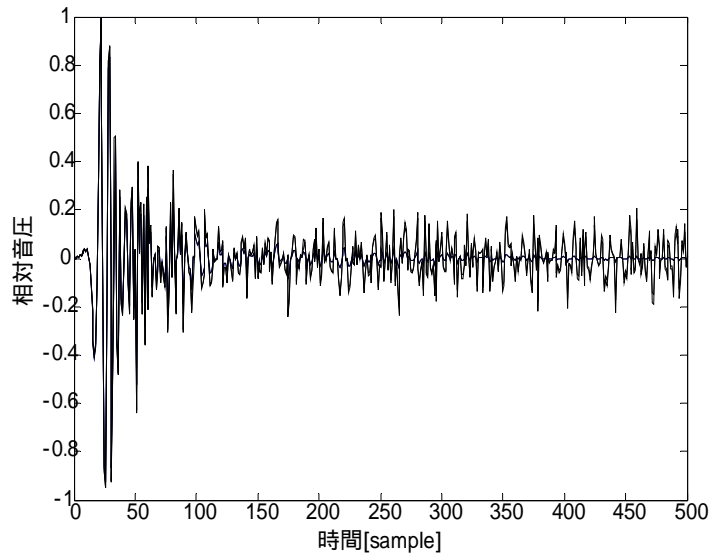


図 8 : `imp13` (残響時間 0.15[s])

また、入力信号 $x(k)$ は白色雑音とし、サンプリング周波数は 8000[Hz]、図 5 の $d_0(k)$ と $n(k)$ のパワー比である SN 比は、40[dB] とした。

4) 結果、検討

4.1 ステップサイズ を変化させた時の結果 ($L=500, \text{imp11}$)

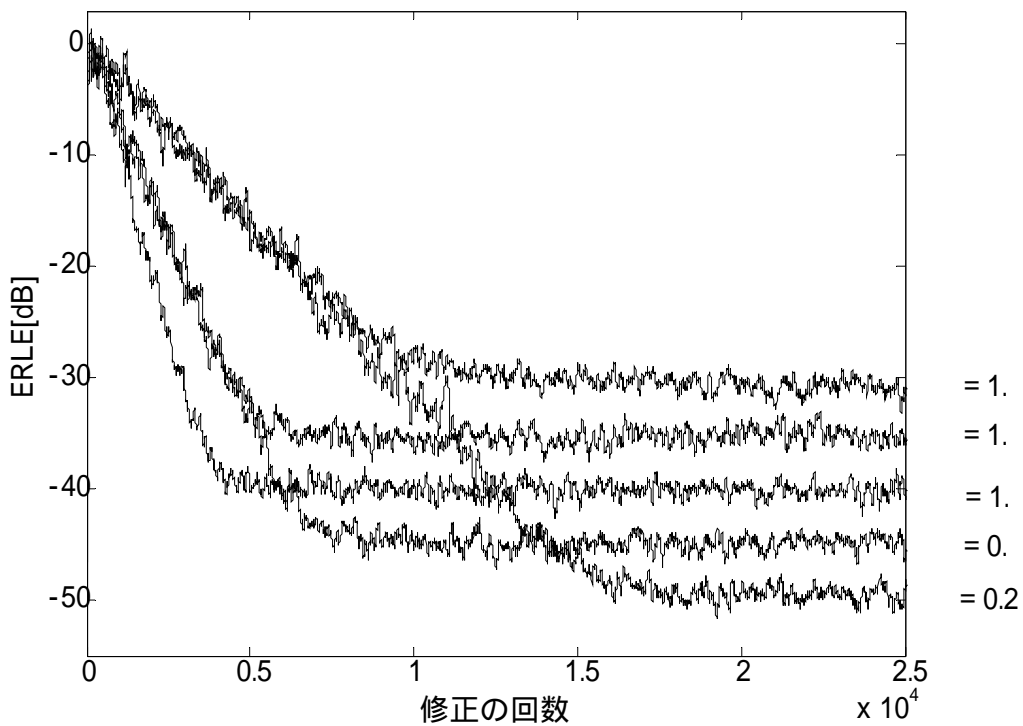


図 9 : を変化させたときの ERLE の変化 ($L=500$)

図9は、ステップサイズを変化させたとき、ERLEの減少経過の変化を表したものである。縦軸はERLE、横軸は経過時間として、修正の回数を表しており、修正の回数8000回が経過時間1秒に相当する。

図9から読み取れることを以下にあげる。まず、ステップサイズを大きくすると、ERLEが大きくなっていることがわかる。また、変化の様子に注目すると、0.2と1.8、0.5と1.5それぞれにおいて、収束特性の傾きが同じように読み取れる。これは、ステップサイズの規則性として、あるステップサイズ ($0 < \alpha < 2$) を考えた場合、 α と $2-\alpha$ の傾きが同じになると言える。

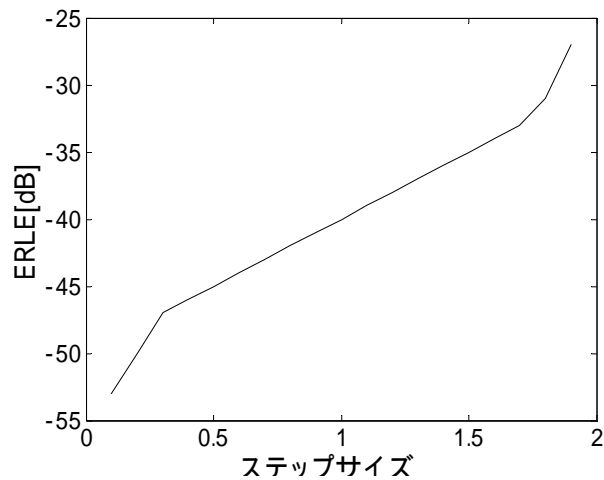


図10：ERLEとステップサイズ の関係

図10はERLEとステップサイズ の関係を表している。縦軸はERLE、横軸はステップサイズ の値を表している。

図10から読み取れることをあげると、ステップサイズ が大きくなるにつれ、ERLEは大きくなっていることがわかる。

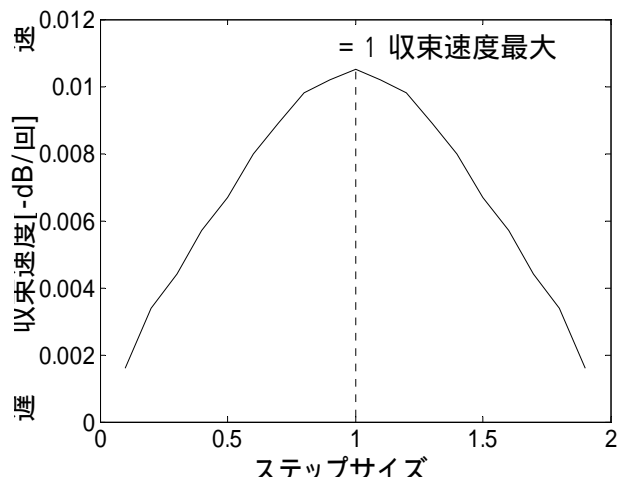


図11：ステップサイズ と収束速度の関係

図 1 1 は収束速度とステップサイズ の関係を表している。縦軸は収束速度、横軸はステップサイズ の値を表している。

図 1 1 から読み取れることをあげると、ステップサイズ が $0 < \alpha < 1$ の範囲にあるとき、収束速度は次第に速くなっていくが、 $1 < \alpha < 2$ の範囲にあるときには収束速度は次第に遅くなっていく。また、図 9 に見られた規則性を検討してみると、図 1 1 からその規則性(あるステップサイズ $a(0 < a < 2)$ を考えた場合、 a と $2-a$ の傾きが同じになる)を読み取ることができる。その変化の仕方は、二次関数的と読み取ることができる。

4. 2 フィルタの長さ L を変化させた時の結果 ($\alpha = 1, \text{imp11}$)

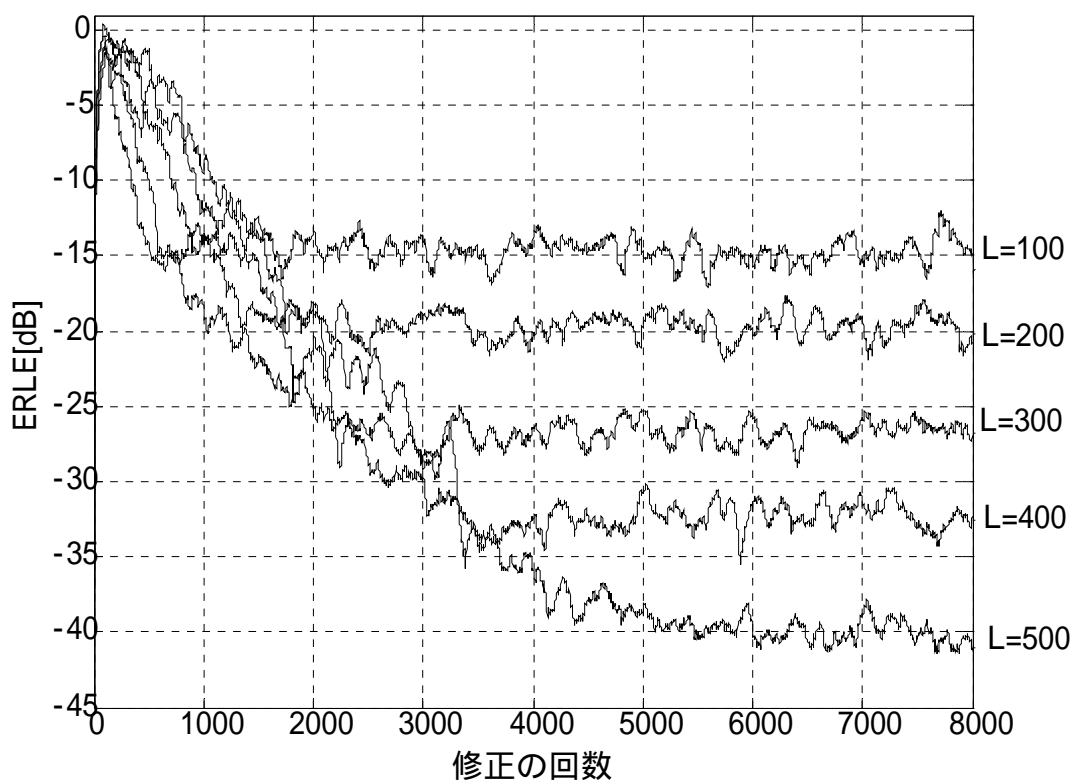


図 1 2 : L を変化させた時の ERLE の変化

図 1 2 はフィルタの長さ L を変化させたとき、ERLE の変化を表したものである。縦軸は ERLE、横軸は経過時間として、修正の回数を表しており、修正の回数 8000 回が経過時間 1 秒に相当する。

図 1 2 から読み取れることとして、フィルタの長さ L を長くすると、ERLE は次第に小さくなっていることが読み取れる。また、フィルタの長さ L を長くすると、収束速度は次第に遅くなっていることが読み取れる。

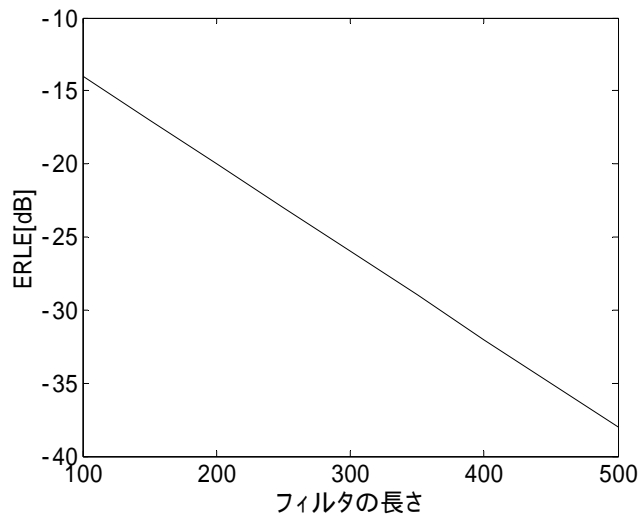


図 1 3 : ERLE とフィルタの長さの関係

図 1 3 は ERLE とフィルタの長さ L の関係を表している。縦軸は ERLE、横軸はフィルタの長さ L の値を表している。

図 1 3 から読み取れることをあげると、フィルタの長さ L が長くなるにつれ、ERLE は小さくなっていることがわかる。その変化の仕方は、比例的と読み取ることができる。

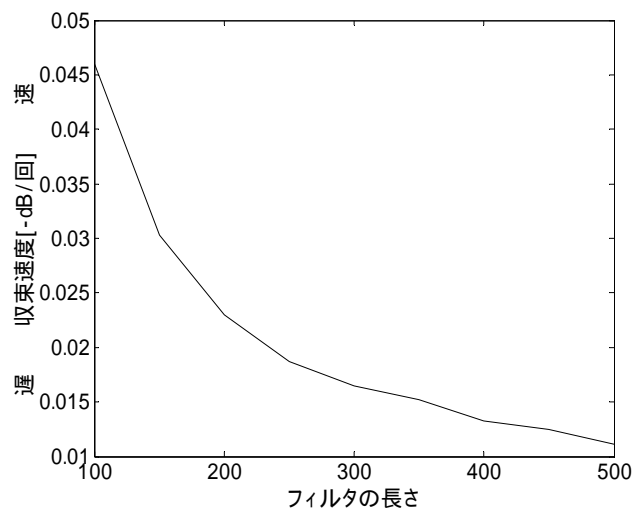


図 1 4 : フィルタの L の長さ と収束速度の関係

図 1 4 は、収束速度とフィルタの長さ L の関係を表している。縦軸は収束速度、横軸はフィルタの長さ L の値を表している。

図 1 4 から読み取れることをあげると、フィルタの長さ L を長くすると、収束速度は次第に、遅くなることが読み取れる。その変化の仕方は、反比例的と読み取れる。

ここで、なぜLを大きくするとERLEが小さくなるのか考えてみた。
 図6のようなインパルス応答がある時、図15のようにパラメータをあたえる。

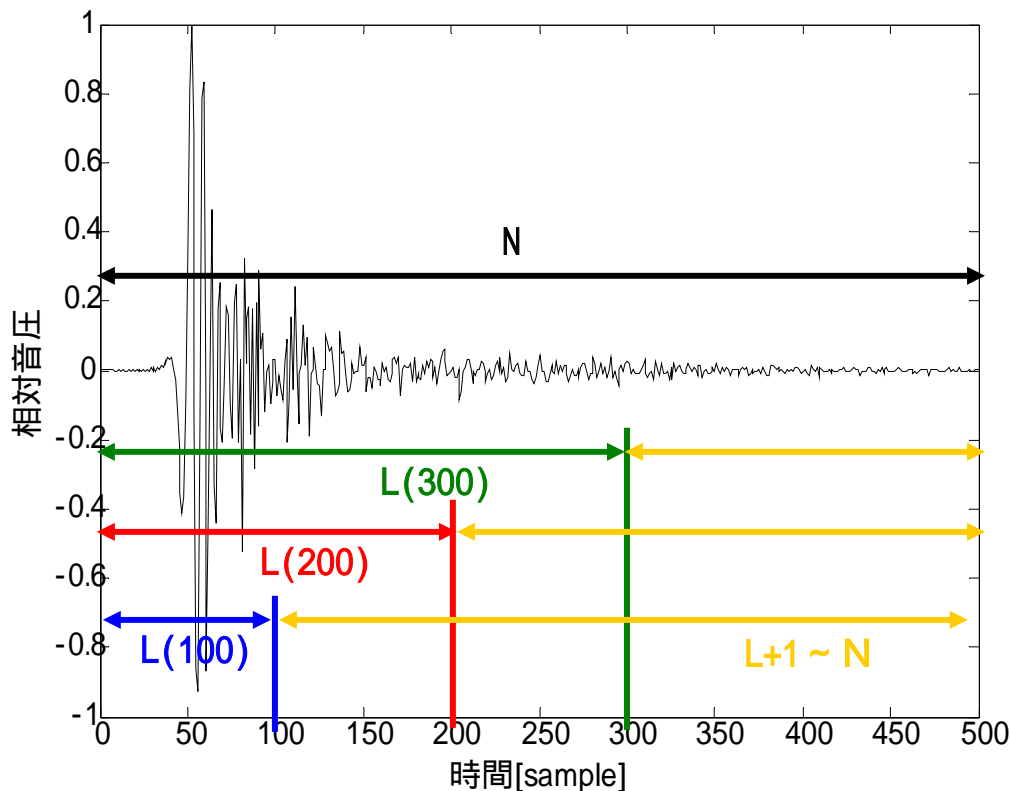


図15：インパルス応答 (imp11) のパラメータ

インパルス応答全体の長さがN(この場合500)、フィルタの長さがL(k)、黄色の範囲はNからLを引いた、Lの範囲によって切り捨てられた部分である。ただし、kは時間[sample]を表す。この図からも分かるように、L(100)、L(200)、L(300)とLを長くするにつれて全体のインパルス応答に対してLの割合が大きくなっている。つまりLによって切すてられる部分が少なくなっている。ここでインパルス応答全体のパワー P_0 とLによって切り捨てられる部分のパワー P_L を(17)、(18)式に表す。

$$P_0 = \sum_{k=1}^N h^2(k) \quad (17)$$

$$P_L = \sum_{k=L+1}^N h^2(k) \quad (18)$$

以上よりERLEの理論値Eは、

$$E = \frac{\sum_{k=L+1}^N h^2(k)}{\sum_{k=1}^N h^2(k)} \quad (19)$$

と表すことができる。

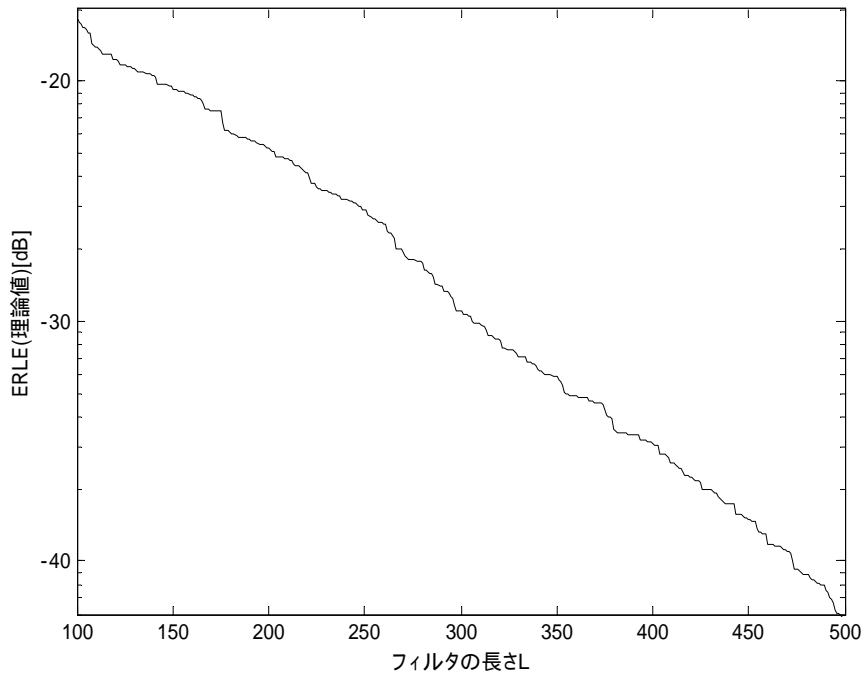


図 1 6 : フィルタの長さ L に対する ERLE(理論値)の変化

図 1 6 は、フィルタの長さ L に対する ERLE(理論値)の変化のグラフである。縦軸は(19)式によって求めた E を図 1 3 と対応するように dB 変換してのもので、横軸はフィルタの長さ L を表す。ここでも、フィルタの長さを長くすると ERLE が減少するのが、覗える。また、図 1 3 と比較してみると L が 100 ~ 500 の間で同じような直線となっている。これは、測定した L に対する ERLE の変化が正しいと考えることができる。

4.3 インパルス応答の残響時間を変化させた時の結果 ($\alpha=0.5, L=500$)

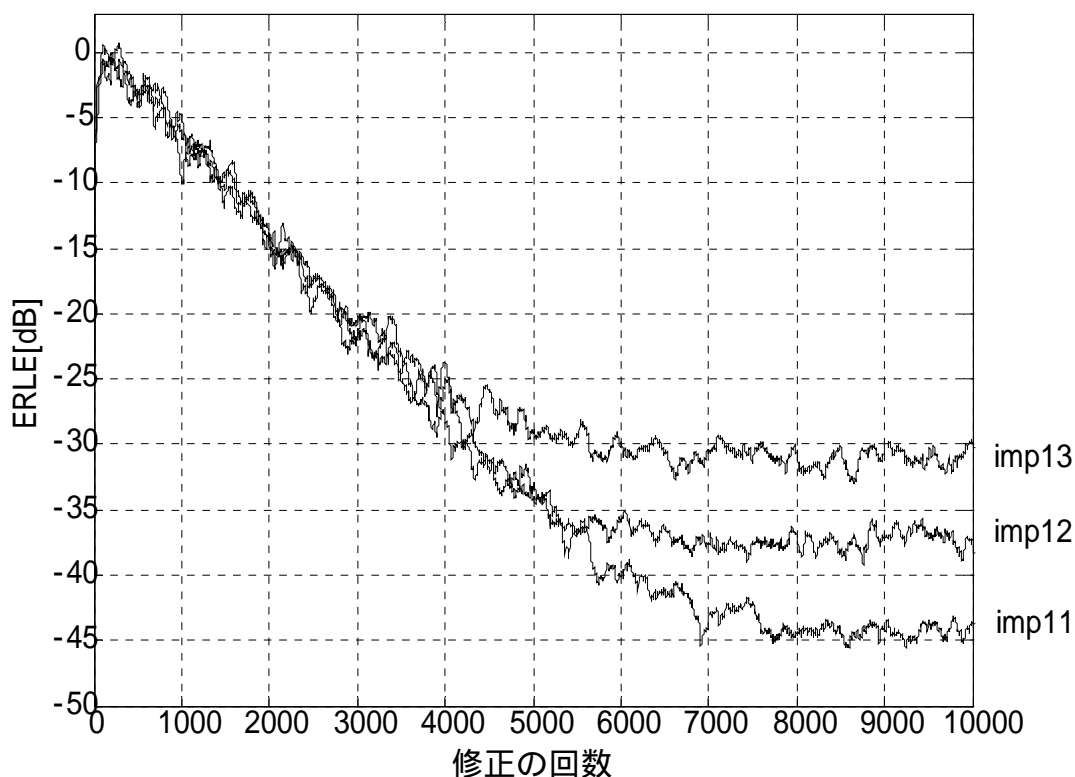


図17：インパルス応答の残響時間を変化させた時の ERLE の変化

図17は、インパルス応答の残響時間を変化させた時の、ERLEの減少経過の変化を表したものである。縦軸はERLE、横軸は経過時間として、修正の回数を表示しており、修正の回数8000回が経過時間1秒に相当する。

図17から読み取れることを以下にあげる。残響時間が変わってもERLEの収束速度（傾き）は変わらないが、残響時間が長いほどERLEの減少は早いうちに頭打ちされる。つまり、残響時間が長いほどERLEは大きくなる。

5) まとめ

収束速度とERLEの減少はトレードオフの関係がある。この関係は、 α とLを変化させることで制御できるが、残響の多い部屋ではERLEの減少はあまり期待できないというのがわかった。また実用的な音響エコーキャンセラのERLEは、30[dB]あれば十分なので、今回の条件（imp11）では、 $\alpha=1, L=400$ 、ぐらいが適当である。

Lに対するERLEの変化は(19)式で表すことができる。

8 . 実測による適応フィルタの基本特性評価

1) 目的

適応フィルタの性能としては、収束速度の速さと、どれだけエコーを消せるか (ERLE) が求められる。学習同定法において、最適な適応フィルタを構成するためには、ステップサイズ とフィルタの長さ L を変化させると、それらの条件がどのように変化するかを7章でわかった。今度は実際にマイクロフォンで雑音と声を録音し、どれくらい雑音を消去出来るかを評価してみた。音源からまたマイクロフォンまでの距離を変化させた時の ERLE の減少の変化も見た。

2) 概要



図 1 8 : 測定風景

実測として、室内のスピーカから、ある距離にマイクロホンを置き、1分間ノイズを流しながら録音する。30秒を過ぎた時点から、人の音声を加える(今回は、ある本を音読したもの)。録音したものを目的信号 d とし、ノイズを適応フィルタに通し、推定した y を差し引き、ノイズを消去し、音声だけを取り出す。その、ノイズの消去率と、音声の音質について検討する。

また、今回はフィルタの効果を見るのが目的であるため、雑音だけと音声+雑音の時間を見つけ出し、雑音 1 秒・音声+雑音 2 秒で h の測定と雑音の消去を行っている。理論的には部屋の実験室はほとんど変化していないため h が測定できた時点 (1 秒) で h の更新を止め、その h を使い、残り 2 秒にフィルタをかければ雑音だけ軽減させることができるのだが、それではうまくいかなく h を更新させたまま (3 秒間 h を更新する) だと、雑音の軽減ができたのでこちらの方法で測定を行っている。以下に作成したプログラムを示す。

```

Clf
    %今までの作業をクリアする

%パラメータの設定
Fs=8000;
    %サンプリング周波数
k=60*Fs;
    %データの長さ
tt=1;
    %フィルタにかける時間を設定
oo=30*Fs;
    %データの中心点を設定
ip0=1+oo-(tt+1)*Fs;
ip1=1+oo-tt*Fs;
    %データのどこからフィルタにかけるかを設定
ip2=1+oo+(tt+3)*Fs;
    %データのどこまでフィルタにかけるかを設定

%データを読み込む
[xx0,Fs]=wavread('D:\MATLAB\WORK\SOUND\xx01.wav');
[dd0,Fs]=wavread('D:\MATLAB\WORK\SOUND\kyoribetu\ddsp25.wav');
xx=xx0(1:k);    %xx0のうち、1~kまでのデータをxxとする
dsp1=dd0(38585:k);
    %dd0のうち、38501~kまでのデータをdsp1とする
dsp2=dsp1(ip1:ip2);
    %dsp1のうち、ip1~ip2までのデータをdsp2とする
dsp3=dsp1(ip0+L:ip2+L);
    %dsp1のうち、ip0+L~ip2+Lまでのデータをdsp3とする

%フィルタの設定
L=150;
    %フィルタの長さを決定
hh=zeros(1,L)';
    %全ての要素が0の1×Lの行列を作成
ee=zeros(ip2,1);
    %全ての要素が0のip2×1の行列を作成
alp=0.2;
    %誤差の重さを設定

%データを適応フィルタにかける
for ip=ip0:ip2;
    %ipがip1~ip2まで繰り返す
    if mod(ip,1000)==0, disp(num2str(ip)), end
    %進行状況を1000ごとに把握する
    yy=hh'*xx(ip:ip+L-1);
    %信号を適応フィルタにかける
    ee(ip)=dsp1(ip+L-1)-yy;
    %ddとyyの誤差を計算
    if ip>ip1
        hh=hh+alp*ee(ip)*xx(ip:ip+L-1)/(xx(ip:ip+L-1)'*xx(ip:ip+L-1));
    end
end

```

```

        %誤差をもとに、フィルタ係数を修正する。
    end
end

%フィルタにかけた結果をグラフにして確認
Pd=dsp1'*dsp1/length(dsp1);
    %dsp1 を二乗平均する
ee1=ee(ip0:ip2);
Pee=ee1.*ee1/Pd;
    %誤差を二乗平均する
Pee2=conv(Pee,boxcar(100))/100;
    %誤差を平均する
figure(1)
plot(10*log10(Pee2));
    %誤差パワーをプロット
axis([0,14000,-80,20])
    %グラフの範囲を設定
grid
    %目盛りをつける

figure(2)
plot(flipud(hh));

figure(3)
plot(dsp3);
hold on
xlabel('修正の回数');
ylabel('相对音圧')
plot(ee1,'r');
hold off
    %ee1 のグラフを書く

%フィルタにかけた結果を音として確認
dmy=input('?');
    %データ確認のため、一呼吸おく
sound(dsp2,Fs);
    %dp2 を再生
dmy=input('?');
    %作業が同時に実行されないように一呼吸おく
sound(ee1,Fs);
    %ee1 を再生

return

```

3) 条件

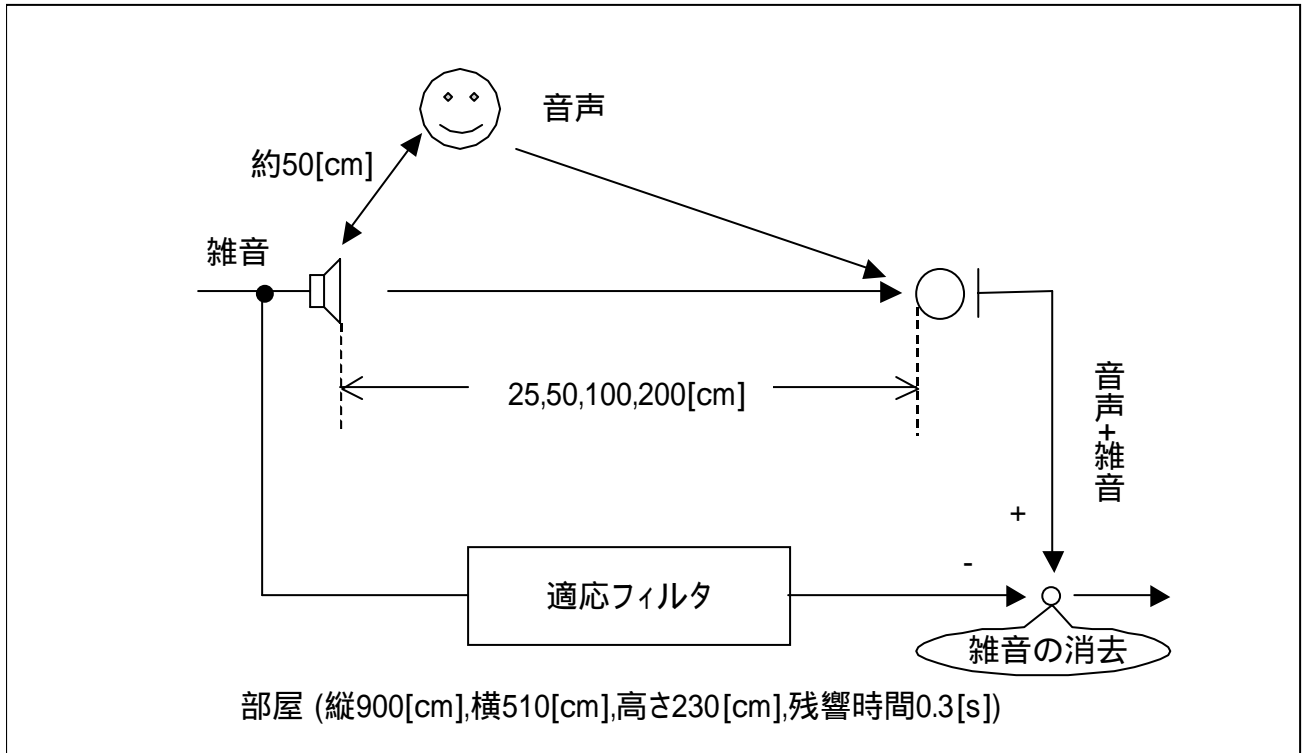


図 1 9 : 配置図

縦 900[cm]、横 510[cm]、高さ 230[cm]、残響時間 0.3[s]の室内で行った。実験のための配置図は図 1 6 に示すように、スピーカと人の位置は固定したまま、マイクロホンの位置を 25、50、100、200[cm]と変化させた。サンプリング周波数は 8000[Hz]とする。

4) 結果、検討

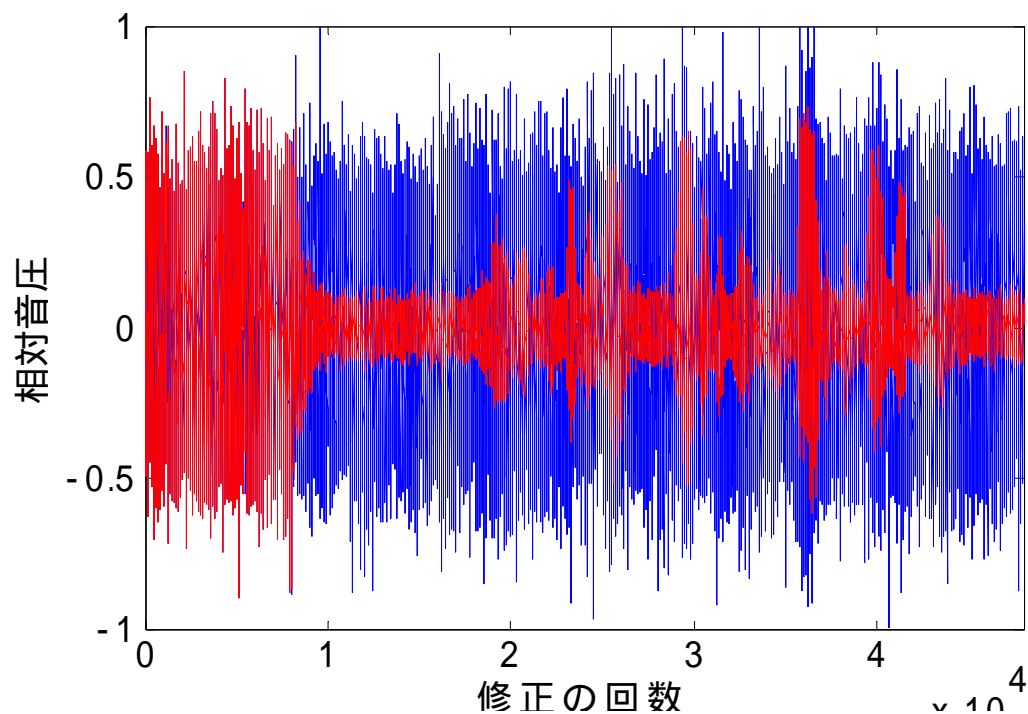


図 2 0 : フィルタにかける前と後の波形の比較(25[cm])

図 2 0 はスピーカから、マイクロホンまでの距離が 25[cm]の時のものである。縦軸は相対音圧、横軸は経過時間として、修正の回数を表しており、修正の回数 8000 回が経過時間 1 秒に相当する。青い波形がフィルタにかける前、赤い波形がフィルタにかけた後の波形を表している。

図 2 0 を見てみると、フィルタに通した後は、ノイズに隠れていた音声信号がはっきりと見て取れるようになっている。ノイズ信号も 0.5 ほどあったものが 0.1 ほどまで消去できている。

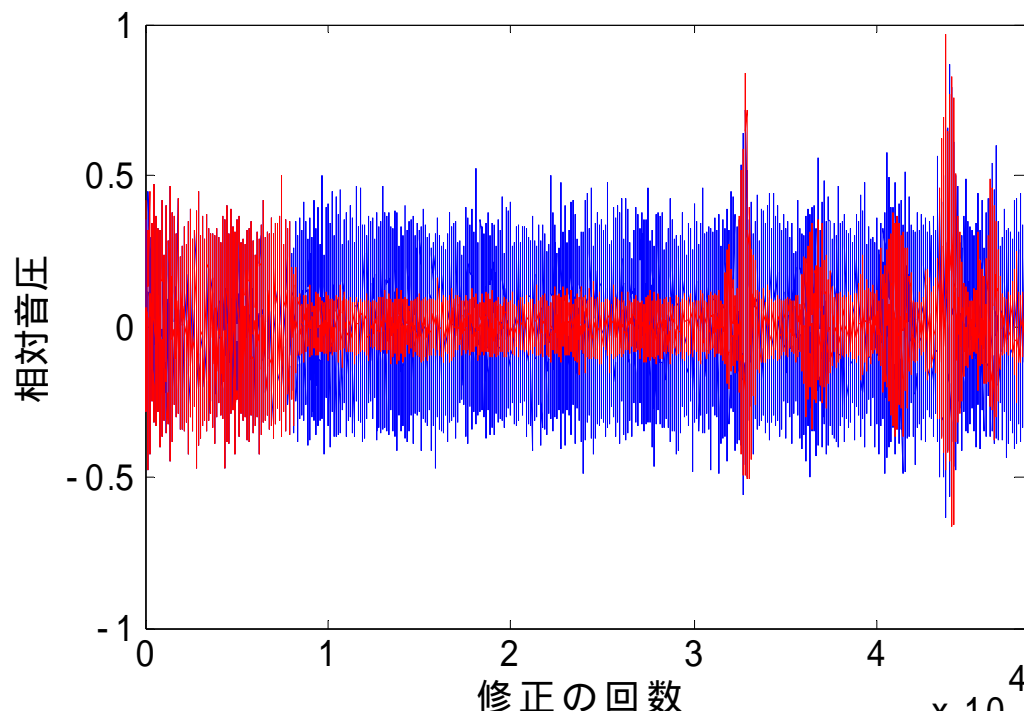


図 2 1 : フィルタにかける前と後の波形の比較(50[cm])

図 2 1 はスピーカから、マイクロホンまでの距離が 50[cm]の時のものである。縦軸は相対音圧、横軸は経過時間として、修正の回数を表しており、修正の回数 8000 回が経過時間 1 秒に相当する。青い波形がフィルタにかける前、赤い波形がフィルタにかけた後の波形を表している。

図 2 1 を見てみると、フィルタに通した後は、ノイズに隠れていた音声信号がはっきりと見て取れるようになっている。ノイズ信号も 0.3 ほどあったものが 0.1 ほどまで消去できている。しかし、25[cm]の場合に比べ、響きがあるように感じ取れる。

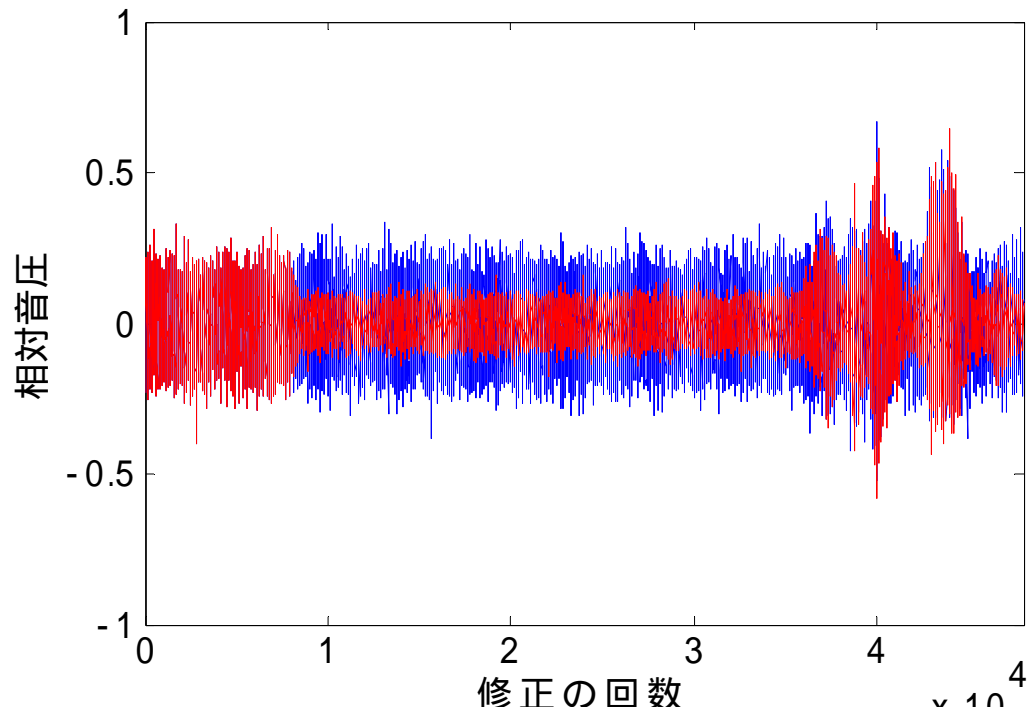


図 2 2 : フィルタにかける前と後の波形の比較(100[cm])

図 2 2 はスピーカから、マイクロホンまでの距離が 100[cm]の時のものである。縦軸は相対音圧、横軸は経過時間として、修正の回数を表しており、修正の回数 8000 回が経過時間 1 秒に相当する。青い波形がフィルタにかける前、赤い波形がフィルタにかけた後の波形を表している。

図 2 2 を見てみると、フィルタに通した後は、ノイズが 0.2 から 0.1 ほどに減少しているのが見て取れる。ノイズに比べ、音声信号がしっかりとわかるので、ノイズが減少したというのは感じにくい。けれども、音声信号に響きのようなものが見て取れるようになった。

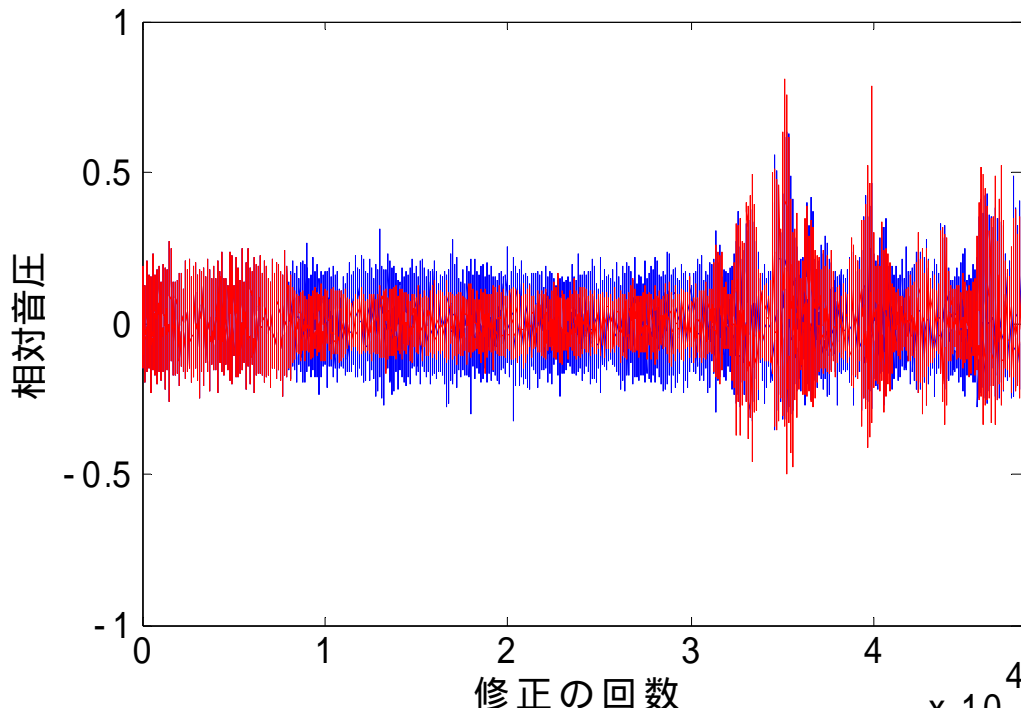


図 2 3 : フィルタにかける前と後の波形の比較(200[cm])

図 2 3 はスピーカから、マイクロホンまでの距離が 200[cm]の時のものである。縦軸は相対音圧、横軸は経過時間として、修正の回数を表示しており、修正の回数 8000 回が経過時間 1 秒に相当する。青い波形がフィルタにかける前、赤い波形がフィルタにかけた後の波形を表している。

図 2 3 を見てみると、フィルタに通した後は、ノイズが 0.15 から 0.1 ほどに減少しているのが見て取れる。ノイズに比べ、音声信号がしっかりとわかるので、ノイズが減少したというのは感じにくい。けれども、音声信号に響きのようなものがはっきりと見て取れるようになった。

図 2 0 ~ 2 3 から、マイクロホンからスピーカまでの距離が遠くなるほど、ノイズの消去率は悪くなる。この原因は、響きのある音がマイクロホンに入力されているからである。距離が短ければ、それだけ直接にスピーカの音（響きの無い音）が、マイクロホンに入力されるが、長くなればなるほど、部屋の特徴（残響時間 0.3[s]）がかかった音が入力される。これはフィルタを通した後の音声を知るとわかりやすい。そして 7 章でも述べたように、響きがあればあるほど、ノイズ消去率は悪くなるのでこのような結果が出た。また、部屋の特徴は刻々と変化していくので、予測した特性とずれてしまう。したがって、消去率が落ちてしまうとも考えられる。

実際に、距離によって、ノイズの消去率はどのように変化するかを次の図に示す。比較しやすいようにフィルタの長さを統一した。

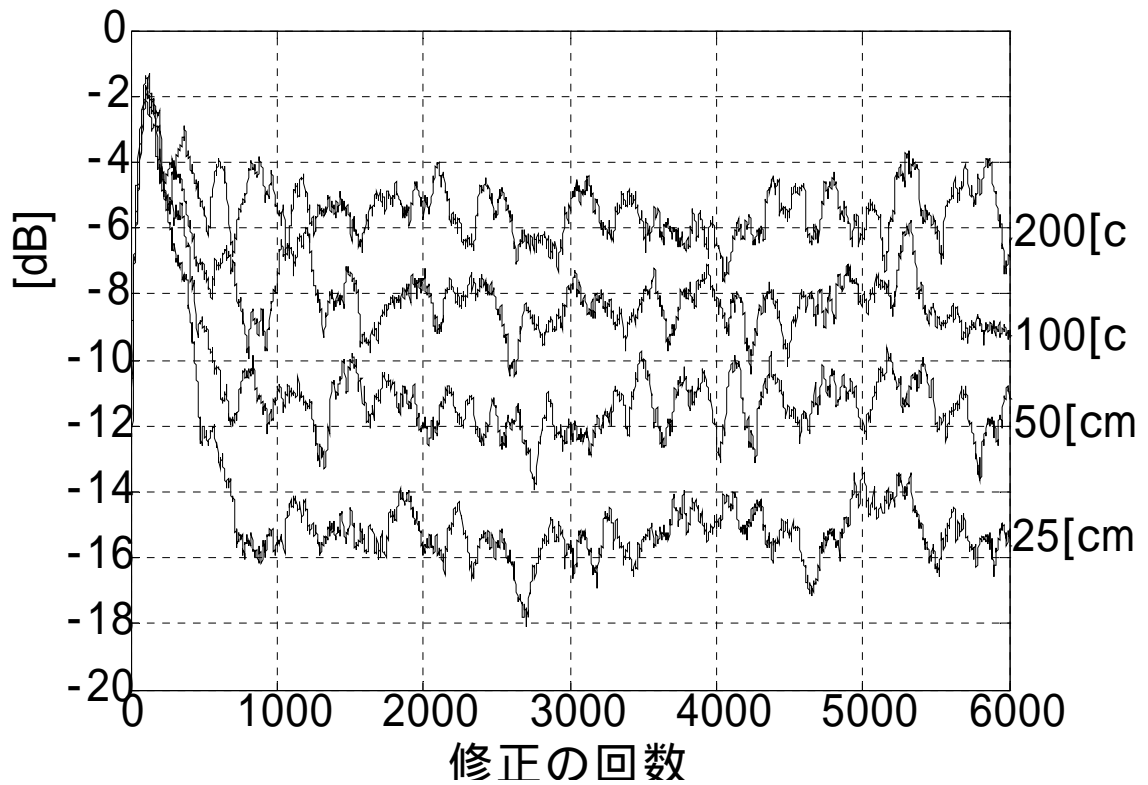


図 2 4 : 距離による ERLE の違い ($L=150$)

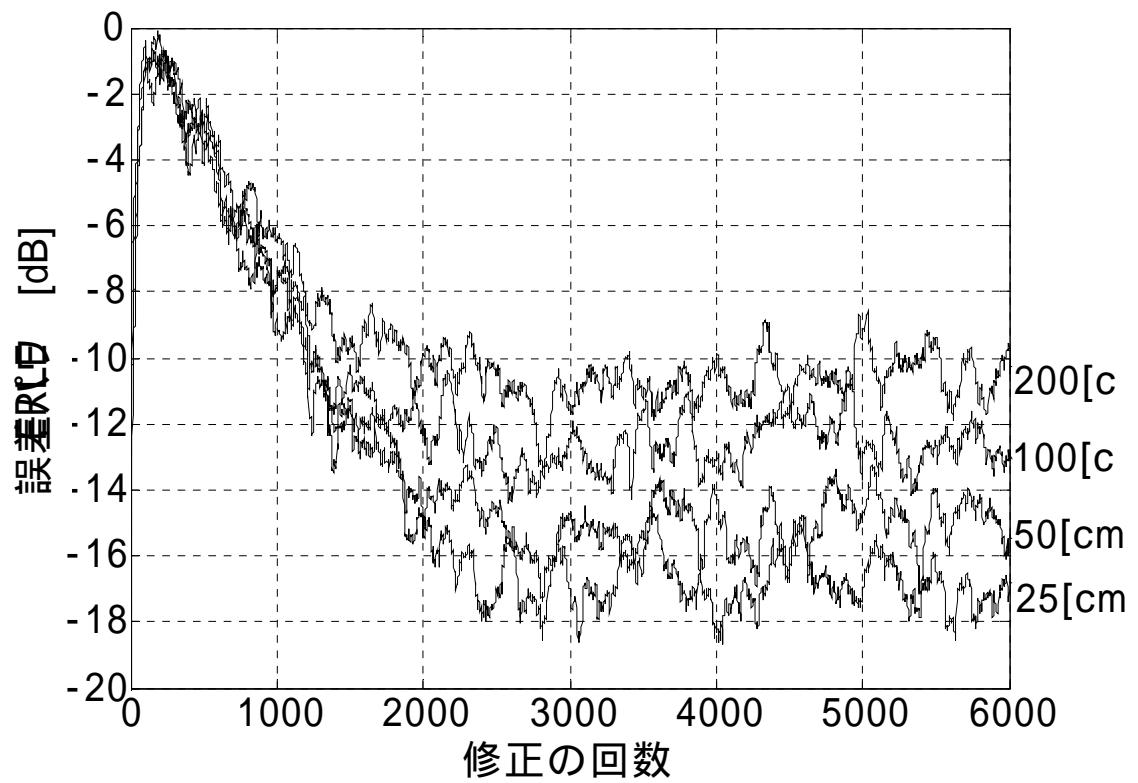


図 2 5 : 距離による ERLE の減少の違い ($L=500$)

図 2 4、図 2 5 は距離による ERLE の変化を表している。縦軸は誤差パワー、横軸は経過時間として、修正の回数を表しており、修正の回数 8000 回が経過時間 1 秒に相当する。図 2 4 はフィルタの長さ L を $L=150$ 、図 2 5 はフィルタの長さ L を $L=500$ に設定した。

ここから、スピーカからマイクロホンまでの距離を長くすると、ERLE の減少が少なくなることが読み取れる。これは、部屋の反響などにより、特性の変化が敏感に関係してくるからと考えられる。また、前述にもあるように、フィルタの長さ L を長くすると、ERLE は小さくなるが、収束速度は遅くなることが、ここでも読み取れる。

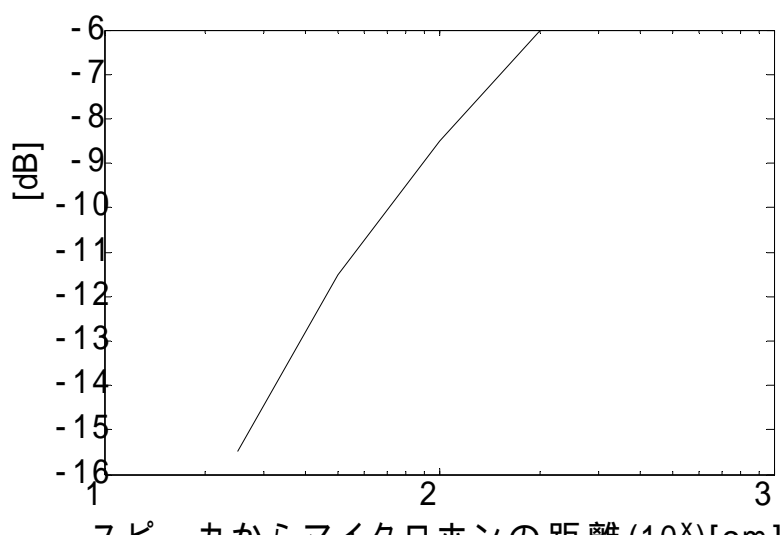


図 2 6 : 距離による ERLE ($L=150$)

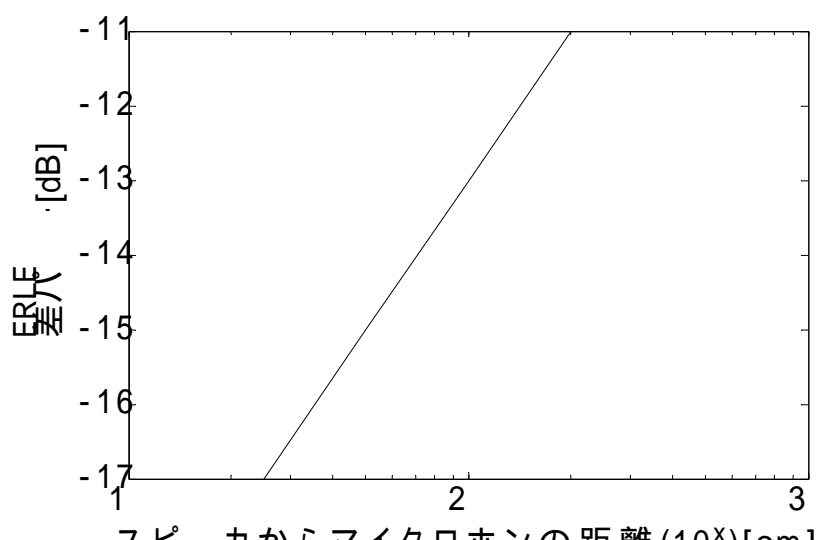


図 2 7 : 距離による ERLE の違い ($L=500$)

図 2 6、図 2 7 は距離による ERLE の変化を表している。縦軸は誤差パワー、横軸はスピーカとマイクロホンとの距離を対数軸で表したものである。図 2 6 はフィルタの長さ L を $L=150$ 、図 2 7 はフィルタの長さ L を $L=500$ に設定した。

図 2 6、図 2 7 から読み取れることとして、ここでもスピーカからマイクロホンまでの距離を長くすると、ERLE が大きくなることが読み取れる。また、スピーカからマイクロホンまでの距離を長くすると、ERLE は対数的に大きくなっていることが読み取れる。これは、部屋の反響などにより、特性の変化が敏感に関係してくるからと考えられる。また、前述にもあるように、フィルタの長さ L を長くすると、ERLE は少なくなることがここでも読み取れる。

5) まとめ

実測実験から以下のことがわかった。

図 2 0 ~ 2 3 からは、

スピーカからマイクロホンまでの距離を長くすると、次第に ERLE の減少が少なくなる。

スピーカからマイクロホンまでの距離を長くすると、次第に音声に響きがあるように感じられるようになる。

より、今回の実験では 100[cm]、200[cm] は、フィルタをかけないほうが音声を聞き取りやすい気がした。

図 2 4 ~ 2 7 からは

スピーカからマイクロホンまでの距離を長くすると、ERLE は大きくなっていることが読み取れる。その変化は対数的に変化する。

7 章の 4) 4 . 2 で述べたようにフィルタの長さ L を大きくするほど、ERLE は少なくなっている。

9. まとめ

本研究ではシミュレーションおよび、実測によって、適応フィルタの処理パラメータと、性能との関係について調べた。

まず、残響時間の違う3つの室内インパルス応答を用いたシミュレーションから以下のことがわかった。

修正の大きさを制御するステップサイズ α は、 $0 < \alpha < 1$ の範囲で大きくするほど収束速度は速くなるが、ERLEの減少は少ない。

ステップサイズ α が1を超えた場合、 α が大きくなるにつれて、収束速度は遅くなり、ERLEも大きくなる。したがって、 α を1より大きくすると、その成果は期待できない。

あるステップサイズ α ($0 < \alpha < 2$) を考えた場合、 α と $2 - \alpha$ の傾きが同じになる。

ステップサイズ α を変化させたときの収束速度の変化は二次関数的に変化し $\alpha = 1$ の時に最大となる。

フィルタの長さ L は、大きくするほどERLEは比例的に小さくなり、収束速度は反比例的に遅くなる。

また、 L に対するERLEの変化は(19)式で表すことができる。

残響時間が変わってもERLEの収束速度(傾き)は変わらないが、残響時間が長いほどERLEの減少は早いうちに頭打ちされる。つまり、残響時間が長いほどERLEは大きくなる。

次に、実際の実測として室内において、音声と雑音の混じった受信信号から雑音を軽減するという実験を行い、以下のことがわかった。

スピーカからマイクロホンまでの距離を長くすると、次第にERLEは大きくなっていく。またその変化は対数的である。

スピーカからマイクロホンまでの距離を長くすると、次第に音声に響きがあるように感じられるようになる。原因として考えられるのは、マイクロホンまでの距離が長くなったことにより微妙に変化している部屋の特徴が音声にかかったためだと思われる。

で述べたように、実測においてもフィルタの長さ L を大きくするほど、ERLEは小さくなっている。

で述べたように、実測においても録音した音に大きな響きを感じる時(残響時間が長い)ERLEは大きくなっている。

これらのことより、最適な適応フィルタというのは、処理速度と消去率をどれほど求めるかによって変わってくることがわかる。また、使用する音響環境条件によって、その性能も変わってきてしまう。主に音響エコーキャンセラの使用される音響環境条件は以下のようになっている。ハンドフリー電話機は個人が家庭、またはオフィスで使い、デスクトップ会議はオフィス、テレビ会議装

置は特定の会議室で多人数で使う場合が多いと考えられる。ではその設置環境を考えてみよう。ハンドフリー電話が送・受話距離 50[cm]～1[m]くらいで普通の部屋で使われるのに対し、デスクトップ会議では送受話とも 1～2[m]、テレビ会議では送話距離 50[cm]～1[m]くらい、受話距離で 2～5[m]くらいでカーペット、天井の吸音ボードなどにより多少の吸音処理がなされている部屋で使う場合が多い。さらに、エコーの発生状況に密接に関係のあるスピーカ、マイクロホンの距離については、電話機で 5～20[cm]、デスクトップ会議で 50[cm]～1[m]、テレビ会議で、2～5[m]と考えられる。

これらの違いをエコーキャンセラのフィルタ長 (8[kHz] サンプリングの場合) で考えてみるとおおむね図 2 8 のように書くことができる。⁽⁵⁾

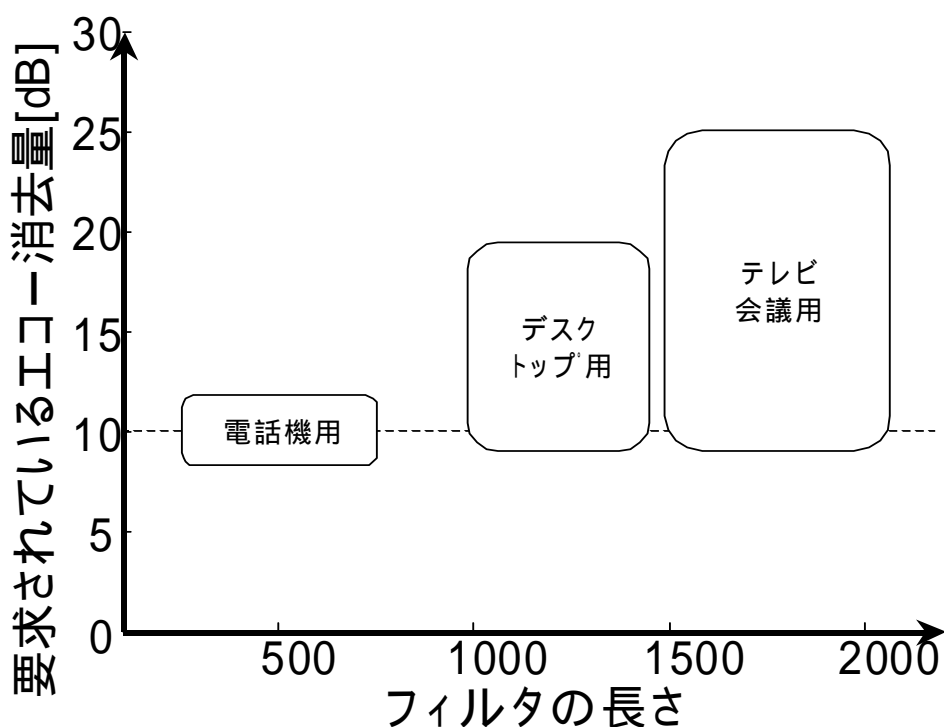


図 2 8 : エコーキャンセラの規模と適応域⁽⁵⁾

図の縦軸は要求されるエコー消去量であり、横軸はフィルタの長さをサンプリング数で表している。⁽⁵⁾

本研究の結果を応用して、図からわかるように用途によって、要求されるエコー消去量とフィルタの長さは異なっており、これに基づいて、ステップサイズパラメータやフィルタの長さを定めて、最適な適応フィルタ特性をシステム化することができるようになる。

参考文献

- (1) 北脇信彦・小島順治：” シームレスな音響環境の実現を目指して ” ,NTT R&D vol.44 No.1, pp.33~35,1995年.
- (2) 金田豊：デジタル信号処理の基礎,日本音響学会第 67 回技術講習会資料, pp.22~27,1999年11月
- (3) 金田豊：音響信号処理技術の現状と展望、pp.3~4.
- (4) 大賀寿郎・山崎芳男・金田豊：音響システムとデジタル処理,電子情報通信学会, pp.136~142,1995年3月.
- (5) 小島順治・牧野昭二・羽田陽一・島内末廣：” 高性能音響エコーキャンセルの開発 ” ,NTT R&D vol.44 No.1,pp.41,1995年.

謝辞

本研究を進めるに当たり、提案、指導をしていただいた、情報通信工学科金田豊教授、音響信号処理研究室の卒研究生の皆様には厚くお礼申し上げます。深く感謝します。