令和元年度

学士学位論文

録音位置の違いによる音声信号のずれ検知

Detection of Time Shift of Audio Signal Due to Difference in Decording Position

1200304　　尾辻明里

指導教員　　福本昌弘

2020 年 3 月 17 日

高知工科大学 情報学群

要 旨

録音位置の違いによる音声信号のずれ検知

尾辻明里

近年，音声認識技術は急速に発展を遂げ，会議での議事録作成や字幕付与などに貢献し，必要性が高まってきている．実際に使用されている音声認識システムの性能は高く，IBM 社と Microsoft 社によると，電話会話の音声認識で 95%の認識率を記録した，といった報告がなされている．しかし，その認識率は雑音が少ない環境や，少人数の対話であることなどの条件を満たさなければ急激に低くなってしまう．また，会議などの実環境では音源数が複数であるため，所望している音声以外を取り除き，所望している音声のみを抽出する技術が必要である．

本研究では， 録音された音から所望音以外の音を取り除くことで所望音を抽出する方法を提案する．原理的には，録音された音から所望音以外の音を減算できればよい．録音された音から所望音以外の音を減算するためには，録音された音と所望音以外の音の同期が求められる．信号のずれは，振幅の減衰や時刻のずれなど様々な要因によって生じており，それらのずれを検知することで，信号の同期がとれる．今回は，信号がマイクに到達する時刻のずれを検知した．2 本のマイクから所望音を抽出するために，音源とマイクの距離の差から生じる音声信号の入力時刻のずれは，各マイクに入力された観測信号のどちらかを 1 サンプルずつずらし相互相関をとることで検知している．このときの相互相関係数は 1 とはならず，相互相関を用いたずれの検知だけでは信号の一致はできなかった．そこで適応フィルタを用い，マイクに到達する時刻以外のずれを合わせ，信号を一致させた．この提案方法により，録音位置の違う音声信号が一致し，所望音以外の音を取り除けることを確認している．

キーワード　適応信号処理, 相互相関，適応フィルタ, 信号の同期，音声抽出

Abstract

Detection of Time Shift of Audio Signal Due to Difference in Decording Position

Akari OTSUJI

In recent years, speech recognition technology has been developing rapidly, contributing to the creation of conference minutes and the addition of video captions. The performance of the speech recognition system actually used is high, and according to IBM and Microsoft, it has been reported that a 95% recognition rate has been recorded in speech recognition of telephone conversations. However, the recognition rate drops sharply unless conditions such as low noise environment and small number of conversations are satisfied.

In this research, we propose a method of extracting desired sounds by removing sounds other than the desired sounds. In order to extract the desired sound from the mixed signal, it is only necessary to be able to subtract sounds other than the desired sound. In order to subtract and eliminate sounds other than the desired sound, perfect synchronization of the observed signal is required. The signal shift is caused by various factors such as the attenuation of the amplitude and the time shift. By detecting these shifts, the signal can be synchronized. This time, we detect the shift of the input time of the signal. In order to extract the desired sound from two microphones, the difference in the input time of the audio signal caused by the distance between the microphones is detected by cross-correlating the observation signals input to each microphone. In this case, the cross-correlation coefficient did not become 1, and signal matching could not

be achieved only by detecting the shift using cross-correlation. Therefore, processing using an adaptive filter was proposed. Errors were minimized by shifting either signal until the cross-correlation coefficient was maximized and updating the filter coefficients on a sample-by-sample basis. It has been confirmed that the proposed method matches audio signals recorded at different recording positions and removes sounds other than the desired sound.

key words adaptive signal processing, cross-correlation, adaptive filter, signal synchronization, desired sound extraction

目次

目次

図目次

図目次

表目次

第 1 章

序論

1.1 本研究の背景と目的

近年，音声認識技術は急速に発展を遂げ，会議での議事録作成や映像の字幕付与などに貢献し，必要性が高まってきている．実際に，スマートフォンやカーナビなどで動作しているものの性能は極めて高く，IBM 社と Microsoft 社によると，電話会話の音声認識で 95%の認識率を記録した，といった報告がなされている [1]．しかし，この認識率は雑音が少ない環境であること，少人数での対話であること，認識させる音声の音量が大きく発音が明瞭であることなどの理想的な条件を満たさなければ低くなってしまう [2]．音声認識のシステムが用いられる会議などの実環境では，音源数が複数であるため，個々の発話の認識率を高める必要がある．したがって，録音された音声の中から所望している音声以外を取り除き，所望音のみを抽出する技術が必要である．

本研究では，録音された音声の中から所望音以外の音を取り除くことで所望音を抽出する方法を提案する．

1.2 本論文の構成

本論文は 5 章より構成されている．以下に各章の概要を述べる．

2 章は, 音声の抽出について所望音が強く入力されている観測信号の選択に必要となる音源方向推定技術や，所望音の抽出方法について，2 本のマイクを用いた簡単なモデルで説明し，音声がマイクに到達する時刻のずれについて述べる．3 章では，予測される録音位置の

1.2　本論文の構成

違いによるマイクに到達する時刻のずれに誤差を実際のずれを測定し調べ，その誤差が生まれた要因について述べる．4 章は，相互相関と適応フィルタを用いた所望音抽出システムについて述べる．5 章では，本研究のまとめと今後の課題について述べる．

第 2 章

音声の抽出

近年，音声認識技術が急速に発展しており，スマートフォンにおける音声入力型情報検索システムなどを始め，様々なサービスが広く利用されている．マイクの近くで 1 人の話者が発話した音声に対して高精度な音声認識が可能であるが，複数の話者が同時に発話した音声に対する認識率は低いことがわかっている [2]．複数人の音声に対して高精度な音声認識を実現するためには，認識させたい話者の音声を録音された音から完全に分離し，抽出できれば良い．そのため，録音された音の中から所望音のみを抽出する技術が必要である．所望音のみを抽出するには，マイクに入力された観測信号から所望音以外を取り除かなければならない．

本研究では，2 本のマイクを用いたモデルで，所望音と所望音以外の音声が混ざった観測信号から所望音以外を取り除き，音声を抽出する方法を提案する．2 本のマイクには，それぞれ所望音と所望音以外の音が観測信号として入力されている．しかし，マイクが複数の場合，所望音がどの観測信号に最も強く入力されているかこのままでは分からない．そこで，所望音が強く入力されている観測信号の選出については，複数のマイクを用いることで必要な音の方向を推定できる音源方向推定技術を用いて選出したものとする [4].

本章では，音源方向推定技術と 2 本のマイクを用いて所望音と所望音以外の音声が混ざった観測信号から所望音以外を取り除く方法と問題について述べる．

2.1　音源方向推定技術

音源方向推定技術は，音の位置情報を用いた音源分離のほかに，テレビ会議システムでの話者方向の特定，ロボットによる救助システム，監視システムなど幅広い分野で求められ，研究が行われている [4][5]．音源方向推定技術には，相関関数，遅延和アレー，高分解能法などがある [5]．ここでは，マイクに信号が届く時間差を用いた手法について説明する．図 2.1 は録音位置の違いによるマイクへ到達する時間の差をモデル化したものである．図 2.1 のように，距離 d 離れた 2 つのマイクに音速 c でそれぞれ音声が入力されたとき，マイク 1 とマイク 2 における時間差 Δt は，

$$\Delta t = \frac{dcos(\theta)}{c} \tag{2.1}$$

である．観測信号 x1 と x2(t) の間の時間差 Δt がわかれば、音の到来方向 θ_s は，

$$\theta_s = cos^{-1}(\frac{c \times \Delta t}{d}) \tag{2.2}$$

より求まる．

2.2　所望音の抽出方法

所望音を抽出するには，所望音と所望音以外の音声が混ざった観測信号から所望音以外の音を取り除ければよい．これは，録音された音声信号から所望音以外の音を減算することで所望音が得られると考えられる [3]．ここでは 2 つの音声信号に対して，マイク 2 本に入力された観測信号から所望音を抽出する方法を簡単なモデルで説明する．

図 2.2 は録音位置の違う 2 本のマイクを使った所望音の抽出モデルである．図 2.2 のように，2 つの音声信号がそれぞれ 2 本のマイクに入力されたとき，あるマイクに入力された観測信号からもう一方の観測信号を用いて，所望音だけを抽出する方法を述べる．2 本のマイクは，それぞれ音声 1 と音声 2 の信号をそれぞれ受信している．音声 1 を所望音とすると，マイク 1 はマイク 2 より音声 1 に近いため，このモデルにおいて所望音が強く入力されている観測信号は観測信号 1 だとすることができる．よって，観測信号 1 から音声 1 を抽出する

ために，観測信号 1 から音声 2 の信号を減算することで音声 1 を抽出できる．しかし，音声 2 からマイク 1 とマイク 2 の距離は異なるため，マイクに到達する時刻にずれが生じてしまい，位相が揃っていないためこのままでは減算することができない．したがって，録音位置の違いによるマイクに到達する時刻のずれを検知する必要がある．

本研究では，録音位置の違いによって生じるマイクに到達する時刻のずれを検知し，観測信号 1 と観測信号 2 を一致させ，所望音を抽出する．

2.2　所望音の抽出方法

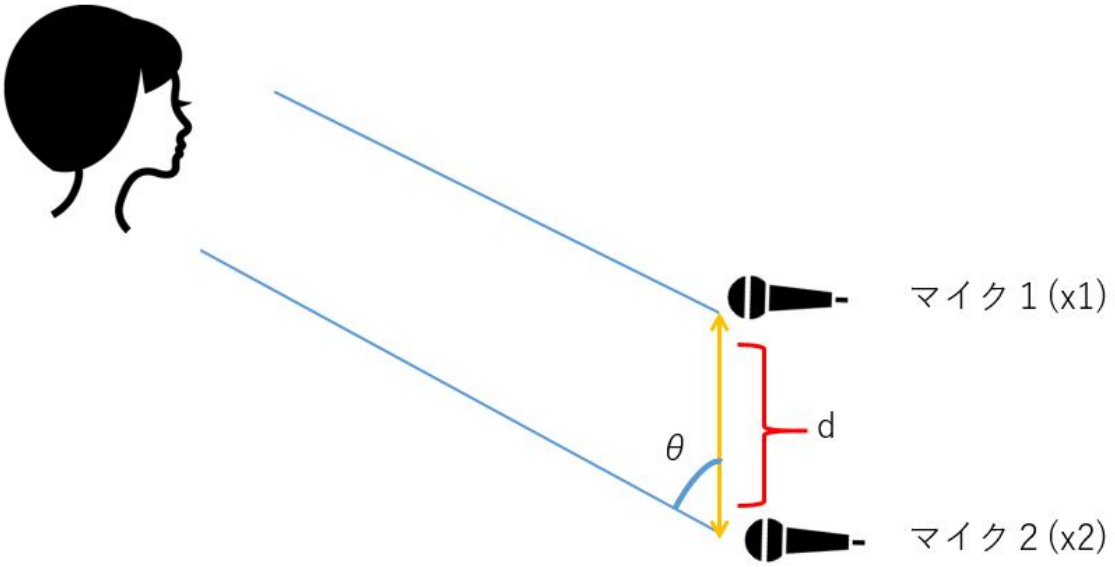

図 2.1　マイクに信号が届く時間差モデル

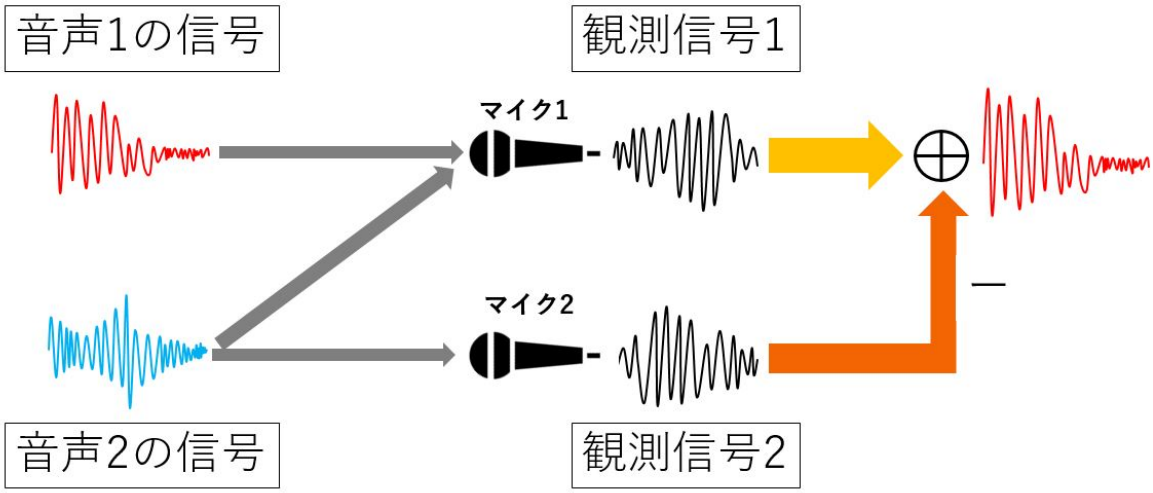

図 2.2　所望音抽出モデル

第 3 章

録音位置の違いによるマイクに到達する時刻のずれ

所望音を抽出するために，録音位置の違いによってマイクに到達する時間のずれを検知し観測信号を一致させ，減算できるようにする．今回，スピーカーから出力された音声を，録音位置の違うマイク 2 本で同時に録音し，観測信号をそれぞれ記録した．このとき，スピーカーに近いマイクの観測信号がどちらであるかはすでに分かっているものとする．

本章では，音源とマイクの距離の差から生じる観測信号のマイクに到達する時刻のずれを検知する．

3.1　予測される入力時刻のずれ

図 3.1 は録音位置の違いによるマイクに到達する時刻のずれを予測するためのモデルである．図 3.1 における録音位置の違いによる音声信号の入力時刻のずれを予測する．マイクに到達する時刻のずれ s は，

$$s = fs \times \frac{d}{c}[\text{サンプル}] \tag{3.1}$$

と予測されるはずである．fs はサンプリング周波数，c[m/s] は音速，d[m] はマイク間の距離とする．また，音速 c は $c = 331.5 + 0.6t$[m/s] とし，ここで t[℃] は摂氏温度である．このとき，マイク間の距離と計測時の温度によって 1 サンプルのずれがどのくらいで発生するのか述べておく．

3.1.1　距離による 1 サンプルのずれ

気温 0 ℃のとき，マイクに到達する時刻のずれが 1 サンプルであるときのマイク間の距離 d を求める．このとき，$fs = 44100$，$c = 331.5$[m/s] とする．マイク間の距離 d は，

$$d = 1 \times \frac{331.5}{44100} = 0.007517006802721[m] \tag{3.2}$$

となる．したがって，式 (3.2) より気温 0 ℃のとき，マイクに到達する時刻のずれが 1 サンプルであるときのマイク間の距離は約 7.517mm である．

3.1.2　温度による 1 サンプルのずれ

マイク間の距離が 20cm のとき，マイクに到達する時刻のずれが 1 サンプルであるときの温度差を求める．このとき，$fs = 44100$，$c = 331.5 + 0.6t$[m/s] とする．

1. 気温 0 ℃のとき 2 つの信号のずれ S_1 は，

$$S_1 = 44100 \times \frac{0.2}{331.5}[\text{サンプル}] \tag{3.3}$$

   で求まる．

2. 気温 1 ℃のとき 2 つの信号のずれ S_2 は，

$$S_2 = 44100 \times \frac{0.2}{331.5 + 0.6}[\text{サンプル}] \tag{3.4}$$

   で求まる．

式 (3.3)(3.4) より，温度 1 ℃あたりのずれ S は，

$$S = S_2 - S_1 = 0.048069258973133 \text{ サンプル} \tag{3.5}$$

である．したがって，マイク間の距離が 20cm のとき 1 サンプルずれる温度差は $\frac{1}{S} = 20.803311632653049$ より，約 20.8 ℃である．表 3.1 は，マイク間の距離ごとに 1 サンプルずれる温度差を示している．

3.2　相互相関関数

相互関数は，ある信号波形を一定時間ずらしたとき，ずらした信号波形とずらす前のもとの信号波形にどのくらい関連があるのか，あるいはどのくらい類似しているのかを図る尺度である．その中でも，2 つの信号の時間的なずれを知るために相互相関関数が利用される[8]．ある音源から出力された音は，録音位置の違いによって図 3.2 のように位相がずれている類似した信号であることがわかる．そこで，信号 $y(t) = \{y_1, y_2, \cdots, y_N\}$ をどれだけずらせば信号 $x(t) = \{x_1, x_2, \cdots, x_N\}$ と類似しているのかを示す尺度として相互相関関数を用いる．信号 $y(t)$ は信号 $x(t)$ より時間 τs 遅れてマイクに到達したとすると，信号 $x(t)$ と信号 $y(t)$ の相互相関関数は $\phi_{xy}(\tau)$ は，

$$\phi_{xy}(\tau) = \sum_{t=1}^{N} x(t)y(t+\tau) \tag{3.6}$$

と定義され，式 (3.6) は，

$$\begin{aligned}\phi_{xy}(\tau) &= \sum_{t=1}^{N} x(t)y(t+\tau) \\ &= \sum_{t=1}^{N} x(t)x(t+\tau-\tau s) \\ &= \phi_{xx}(\tau-\tau s)\end{aligned} \tag{3.7}$$

と変形できる．式 (3.7) は，自身との相互相関関数である自己相関関数である．自己相関関数とは，1 つの信号波形にに対して相関をとり，周期性があるかどうかを調べることができる．この自己相関関数 $\phi_{xx}(\tau-\tau s)$ は $\tau = \tau s$ のとき，遅延時間がなく，信号が一致し，最大値をとる．したがって，相互相関関数 $\phi_{xy}(\tau)$ は自己相関関数 $\phi_{xx}(\tau-\tau s)$ を τs ずらしたものであることから，相互相関関数 $\phi_{xy}(\tau)$ は $\tau = \tau s$ で最大値をとる。

3.3　録音位置の違いによるマイクに到達する時刻のずれ測定

この予測されるずれが実際のずれと一致しているか測定し，検証する．図 3.3 は測定環境，図 3.4 は測定のモデル図である．

3.3　録音位置の違いによるマイクに到達する時刻のずれ測定

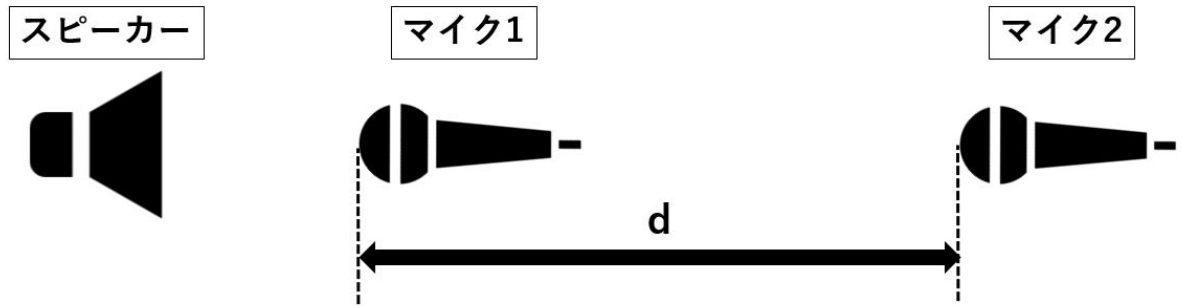

図 3.1　マイク間の距離差による音声信号の入力時刻のずれ

表 3.1　距離毎の 1 サンプルずれる温度差

d[cm]	20	21	22	23	24	25
t[℃]	20.8	19.8	18.9	18.1	17.3	16.6

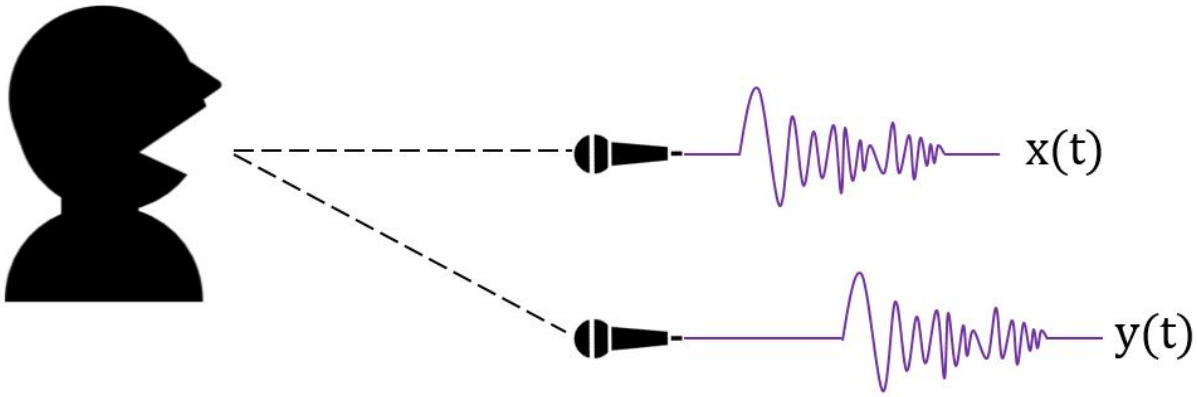

図 3.2　録音位置の違う受音された 2 つの信号

表 3.2　測定時に使用した機材

PC	VAIO VPCZ23AJ
マイクロホン	BEHRINGER C-2
オーディオインタフェース	Roland UA-1010
スピーカー	Roland MA-7A

3.3　録音位置の違いによるマイクに到達する時刻のずれ測定

図 3.3　実験環境

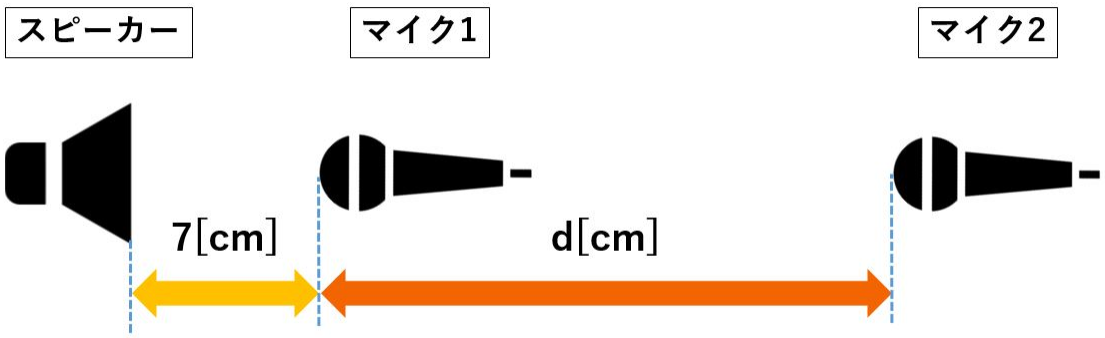

図 3.4　実験の測定モデル

図 3.4 のように，スピーカーと 2 本のマイクを一直線上に並べ，吸音材で覆った．音声は図 3.5 のような手を 1 度たたいた破裂音と，約 1 秒程度の人間の声を含む 8 秒間の録音音声を使用した．縦軸は振幅，横軸は時間 [s] を表している．測定には表 3.2 の機材を使用した．このときサンプリング周波数は 44.1kHz，気温 21.2 ℃であった。

3.3.1　測定方法

図 3.4 のようにスピーカーとマイク 1 の距離を 7cm に固定し，マイク 1 とマイク 2 の距離 d を 20 から 25 まで 1cm ずつ変化させ，マイク 1 とマイク 2 の観測信号を記録した．図 3.6，図 3.7，図 3.8，図 3.9，図 3.10，図 3.11 はマイク間の距離ごとにマイク 1 の観測信号，マイク 2 の観測信号を示し，縦軸は振幅，横軸は時間 [s] を表している．

3.3.2　検証方法

マイクに到達した時間のずれを推定するために相互相関関数を利用する．予測したずれが実際のずれと一致しているかを検証するために，マイク 2 の観測信号を 1 サンプルずつずらし，マイク 1 の観測信号と相互相関を取る [7]．マイク 1 の観測信号を $\mathbf{x} = \{x_1, x_2, \cdots, x_N\}$，マイク 2 の観測信号を $\mathbf{y} = \{y_1, y_2, \cdots, y_N\}$ とするとき，相互相関係数 $\tilde{R}_n^{(xy)}$ は，

$$\tilde{R}_n^{(xy)} = \frac{\frac{1}{N}\sum_{i=1}^{N} x_i y_{i+n}}{\sqrt{\frac{1}{N}\sum_{i=1}^{N} x_i^2}\sqrt{\frac{1}{N}\sum_{i=1}^{N}(y_{i+n})^2}} \tag{3.8}$$

と定義される．マイク 2 の観測信号はマイク 1 の観測信号より遅れて入力されている．式 (3.8) は，$\mathbf{x}$ の時間軸はずらさず，$\mathbf{y}$ を左へ n サンプルシフトした信号，つまり，

$$\mathbf{y}^{(n)} = \{y_{1+n}, y_{2+n}, \cdots, y_{N+1}\} \tag{3.9}$$

表 3.3　距離毎のずれの比較

d[cm]	20	21	22	23	24	25
s[サンプル]	25	26	28	29	30	32
n[サンプル]	23	24	26	26	28	29
$\tilde{R}_n^{(xy)}$	0.9007	0.8968	0.915	0.9237	0.9199	0.917

との内積を計算している．内積は 2 つの信号間の類似性を評価する尺度のため，この類似性を時間ずれ n サンプルを変数に持った係数として表される．また，式 (3.9) は通常，

$$y_i = y_{i+N} \tag{3.10}$$

と N サンプルごとの周期性を持っていることを考慮すると，式 (3.8) は，

$$\tilde{R}_n^{(xy)} = \frac{\sum_{i=1}^{N} x_i y_{i+n}}{\sqrt{\sum_{i=1}^{N} x_i^2}\sqrt{\sum_{i=1}^{N} y_i^2}} \tag{3.11}$$

と変形することができる．$\tilde{R}_n^{(xy)}$ は $-1 \leq \tilde{R}_n^{(xy)} \leq 1$ となり，正規化された相関値となる．$|\tilde{R}_n^{(xy)}|$ の値が大きいほど相関があるといえる．これは $|\tilde{R}_n^{(xy)}|$ が大きいほど $\mathbf{x}$ と $\mathbf{y}^{(n)}$ は類似性が高いといえ，このときの n を実際のずれとみなせる．また，$\tilde{R}_n^{(xy)} = 1$ のとき，マイク 1 の観測信号とマイク 2 の観測信号は一致している．

3.3.3　実験結果

表 3.3 に実験の結果を示す．各 d の距離に対して，s が予測されるずれ，n が実際のずれ，$\tilde{R}_n^{(xy)}$ がそのときの相互相関係数を示している．また，図 3.12，図 3.13，図 3.14，図 3.15，図 3.16，図 3.17 はマイク間の距離ごとに相互相関係数を示し，縦軸は相関係数，横軸はサンプル数を表している．この相互相関係数が最も高いところが実際のずれである．表 3.3 より，s と n は一致せず，2，3 サンプルの差であった．

3.3.4　誤差

予測したずれに誤差が生じたのは，予測時に用いたマイク間の距離または気温の計測ミスによって誤差が生まれたと考えられる．式 (3.2) より本実験における 1 サンプルずれるのに要した距離は 7.8mm である．また，表 3.1 より本実験における 1 サンプルずれるのに要した温度は最大で 20.8 ℃であることから，温度の計測ミスによる誤差の可能性はかなり低く，距離の計測ミスによる誤差であると考えられる．

また，実験で実際のずれを検知するときに用いた相関係数 $\tilde{R}_n^{(xy)}$ は最大で 0.924 であった．相関は極めて強いが，$\tilde{R}_n^{(xy)} = 1$ でないため，信号は完全に一致していない．図 3.18，図 3.19 はマイク 1 とマイク 2 の観測信号の相互相関によってマイクに到達する時刻のずれを検知し，そのずれの分だけマイク 1 をずらしたものとマイク 2 の観測信号の誤差を，マイク間の距離毎に示したものである．図 3.18，図 3.19 からも，マイク 1 とマイク 2 の観測信号が一致していないことがわかる．これは，信号がマイクに到達するまでに減衰や残響，歪みなどの影響でもとの信号と異なっているものを観測しているためだと考えられる．そのため，位相のずれだけだけを補正しても信号は一致しない．そこで，相互相関をとって位相のずれを補正した信号と実際に観測された信号との誤差を，適応フィルタを用いることで誤差を最小化し，信号の一致を目指す．

3.3　録音位置の違いによるマイクに到達する時刻のずれ測定

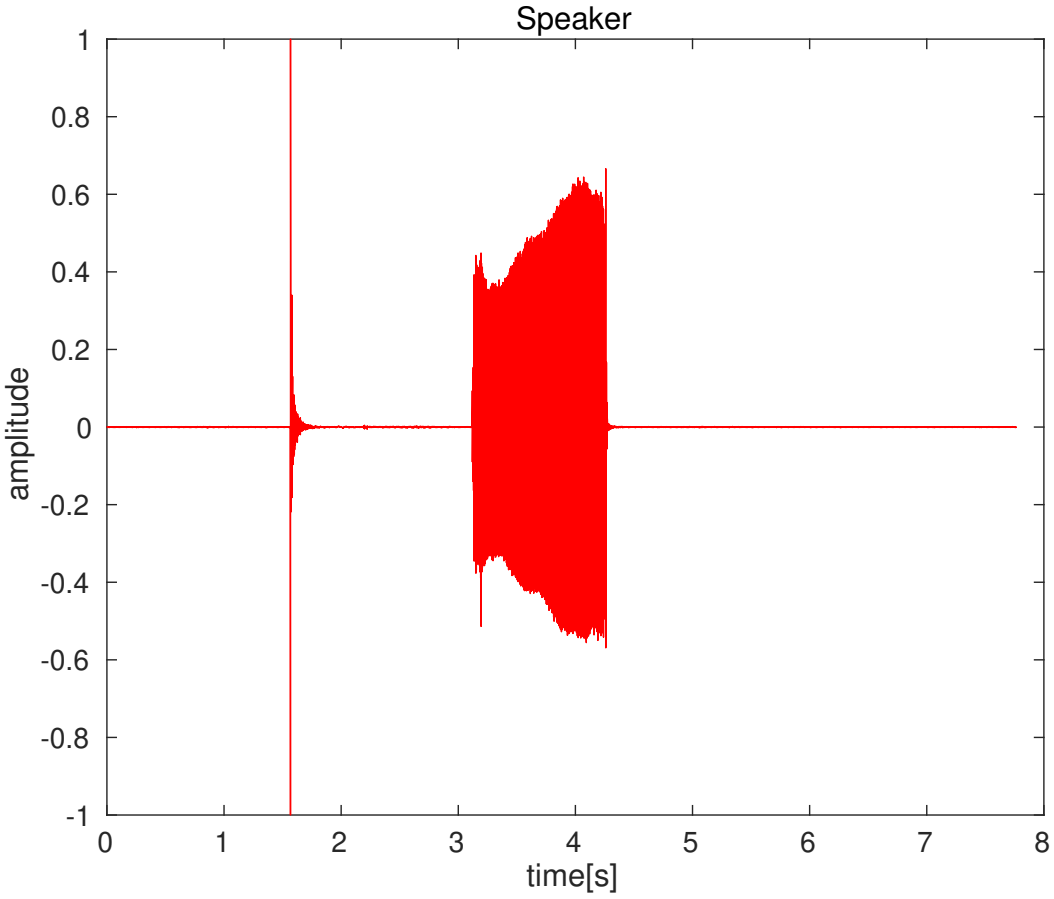

図 3.5　使用した録音音声

3.3　録音位置の違いによるマイクに到達する時刻のずれ測定

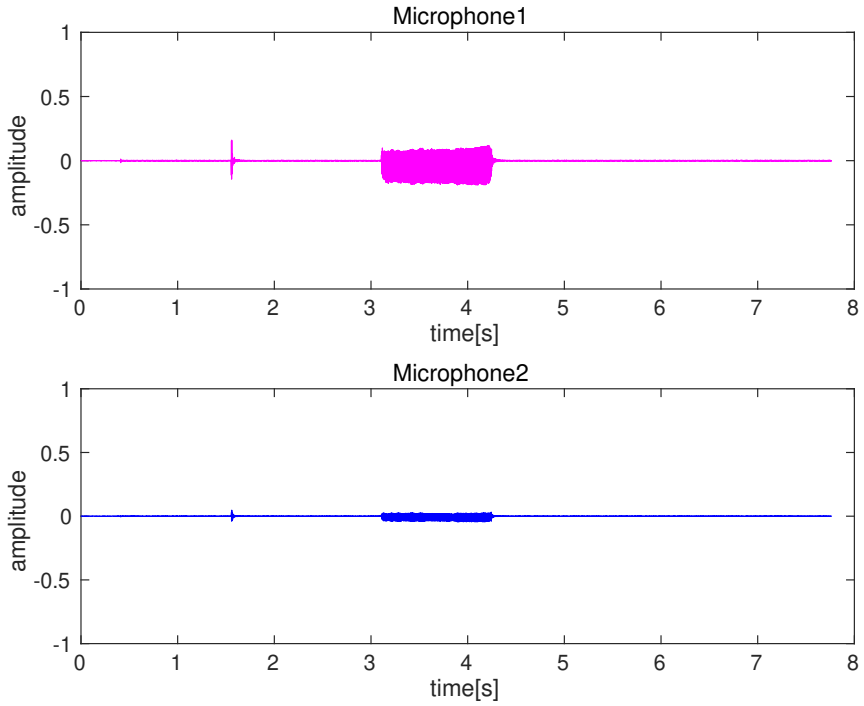

図 3.6　マイク間の距離が 20cm のときの観測信号

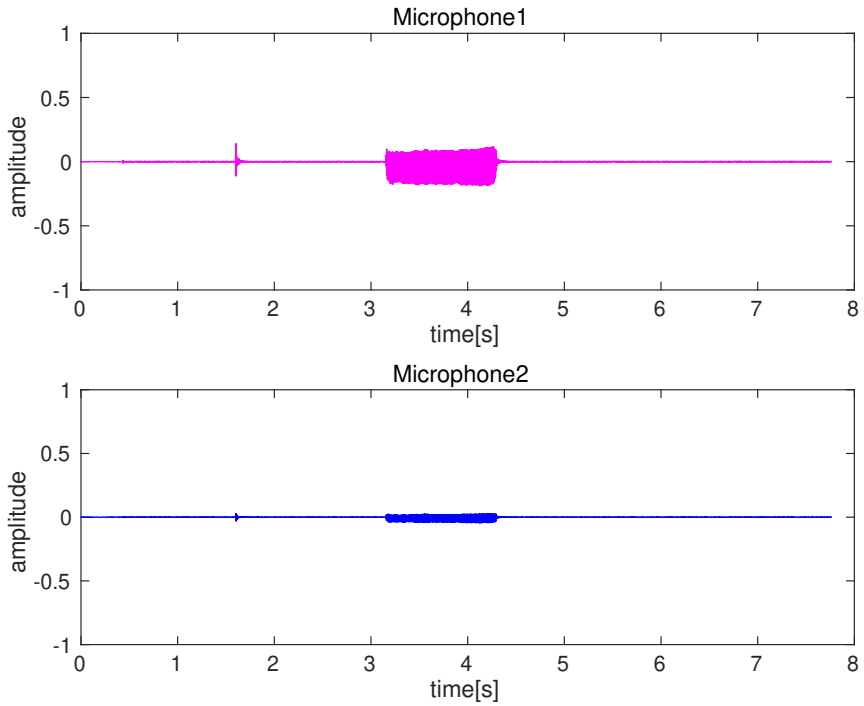

図 3.7　マイク間の距離が 21cm のときの観測信号

3.3　録音位置の違いによるマイクに到達する時刻のずれ測定

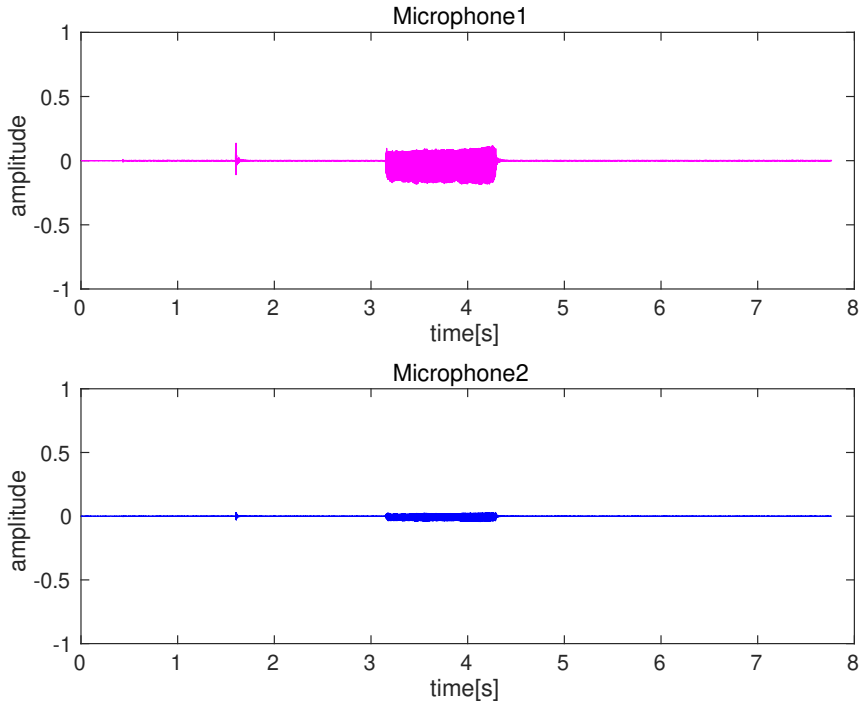

図 3.8　マイク間の距離が 22cm のときの観測信号

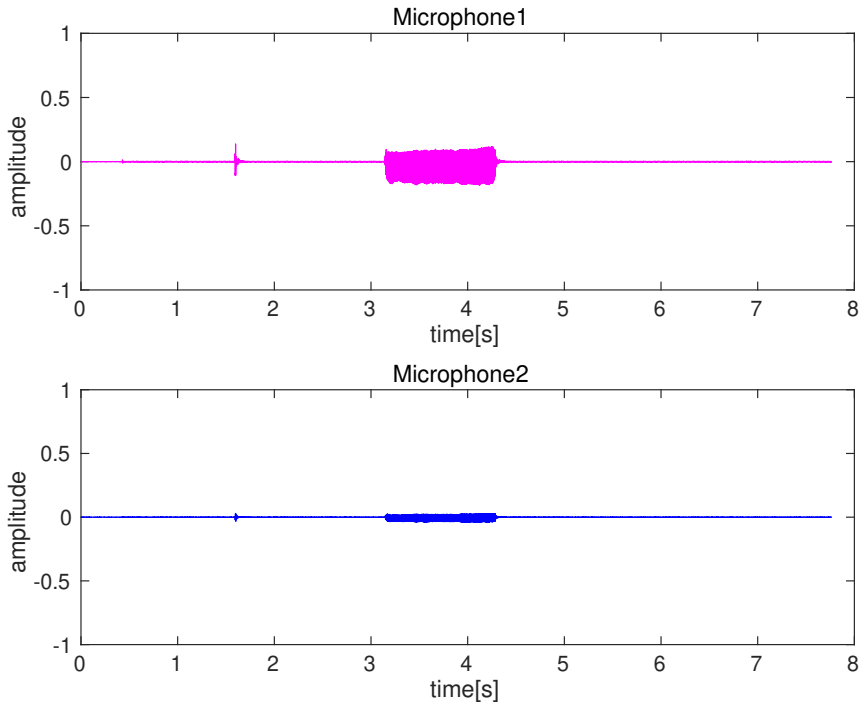

図 3.9　マイク間の距離が 23cm のときの観測信号

3.3　録音位置の違いによるマイクに到達する時刻のずれ測定

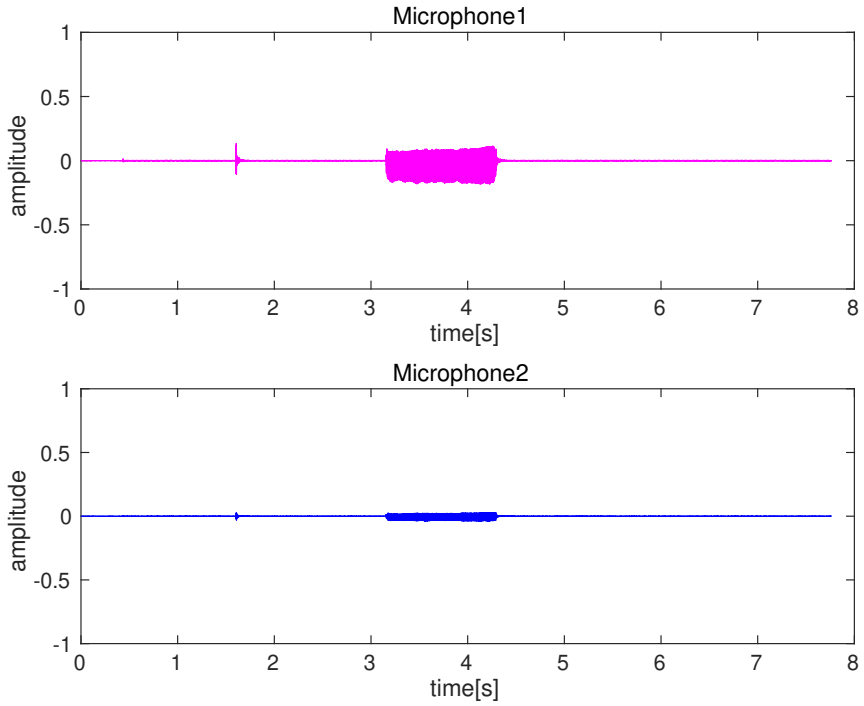

図 3.10　マイク間の距離が 24cm のときの観測信号

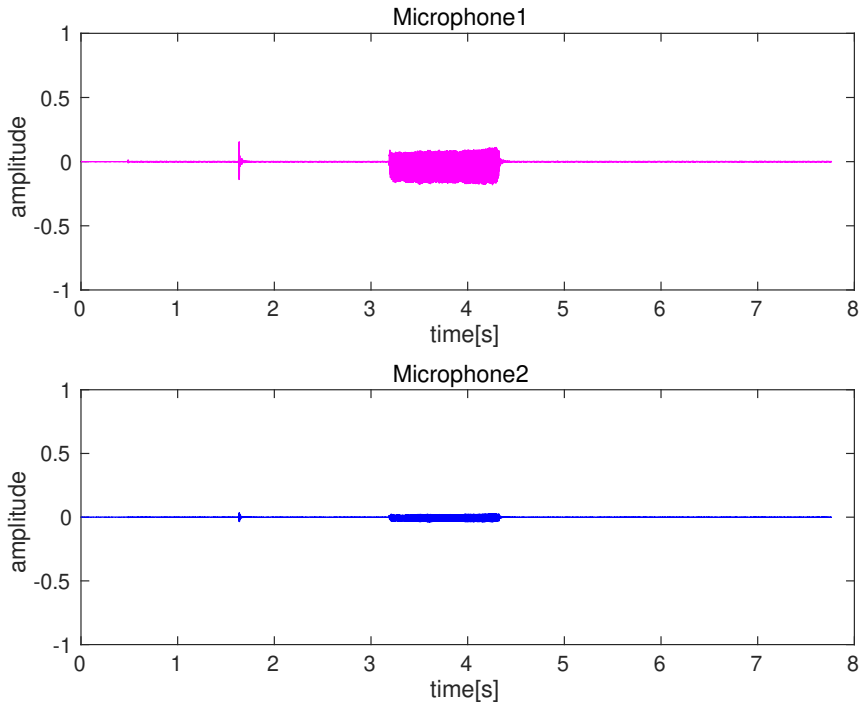

図 3.11　マイク間の距離が 25cm のときの観測信号

3.3　録音位置の違いによるマイクに到達する時刻のずれ測定

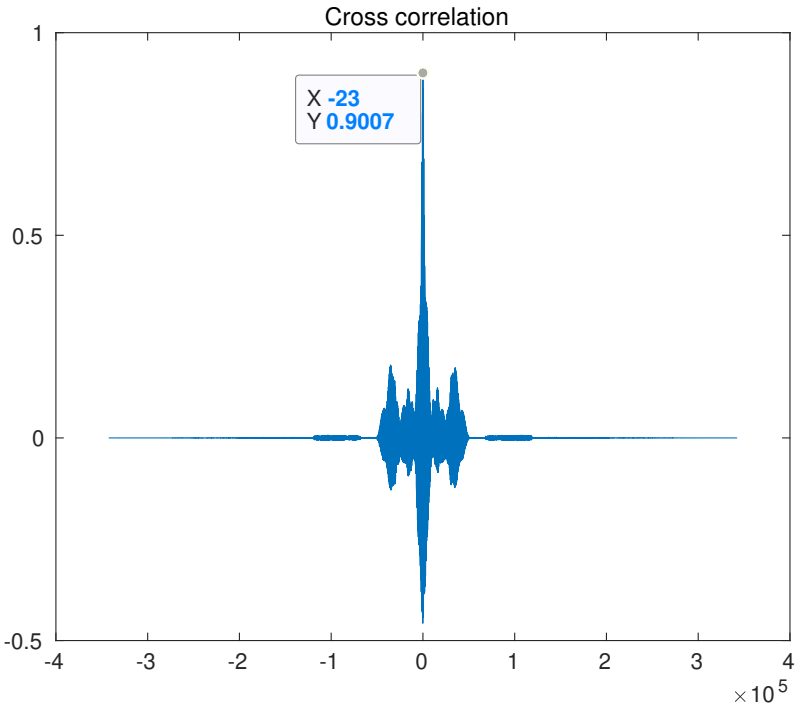

図 3.12　マイク間の距離が 20cm のときの相互相関

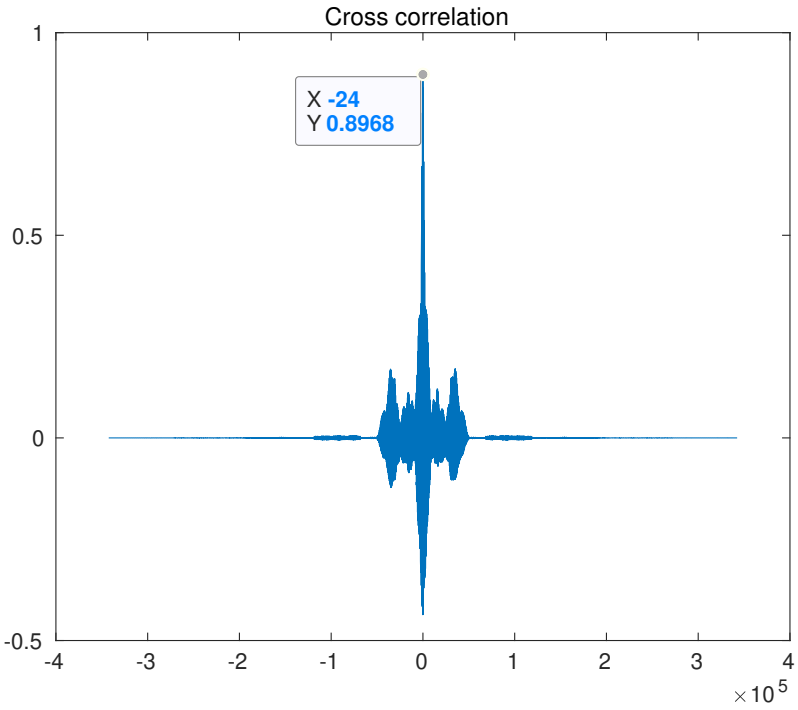

図 3.13　マイク間の距離が 21cm のときの相互相関

3.3　録音位置の違いによるマイクに到達する時刻のずれ測定

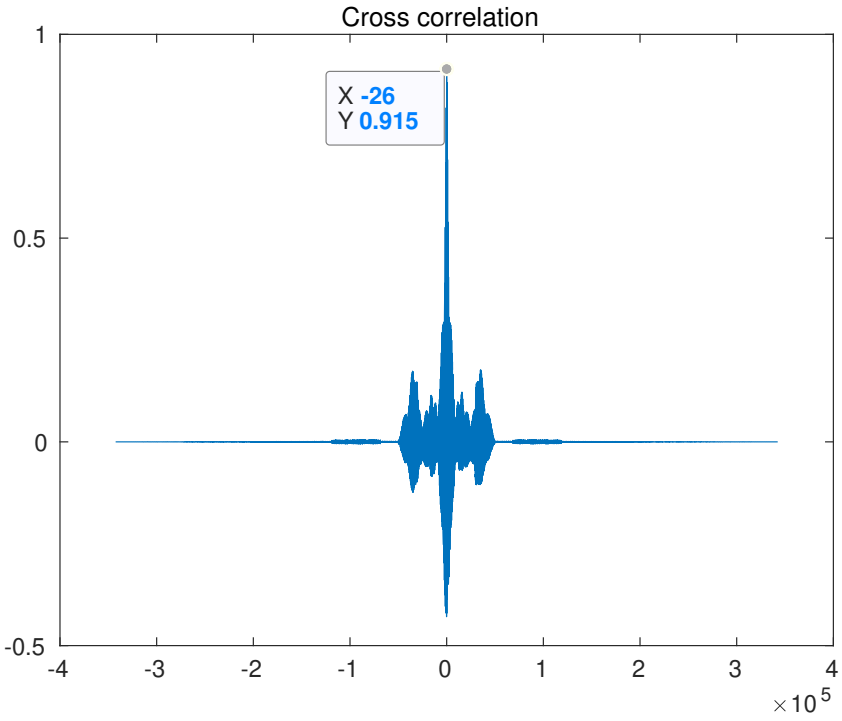

図 3.14　マイク間の距離が 22cm のときの相互相関

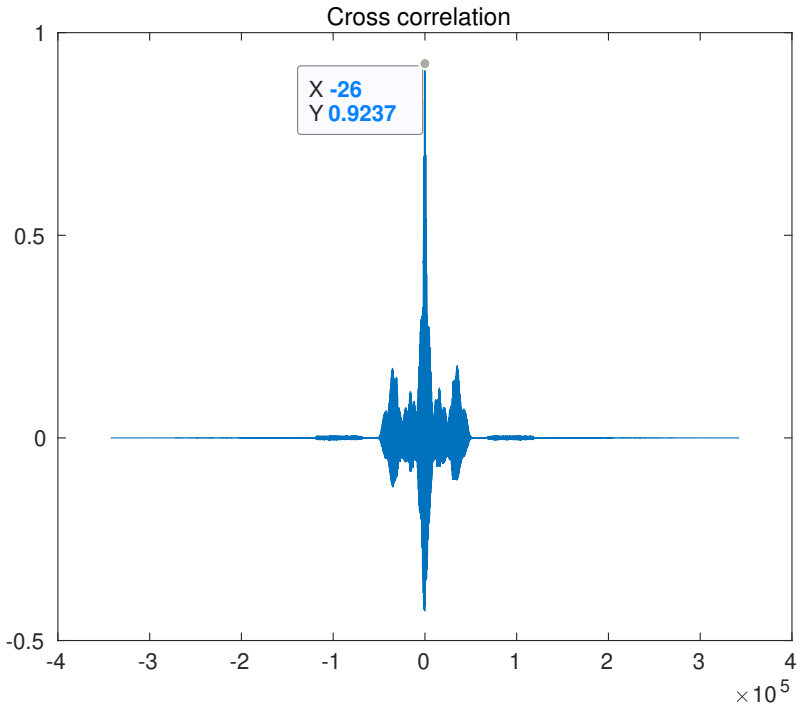

図 3.15　マイク間の距離が 23cm のときの相互相関

3.3　録音位置の違いによるマイクに到達する時刻のずれ測定

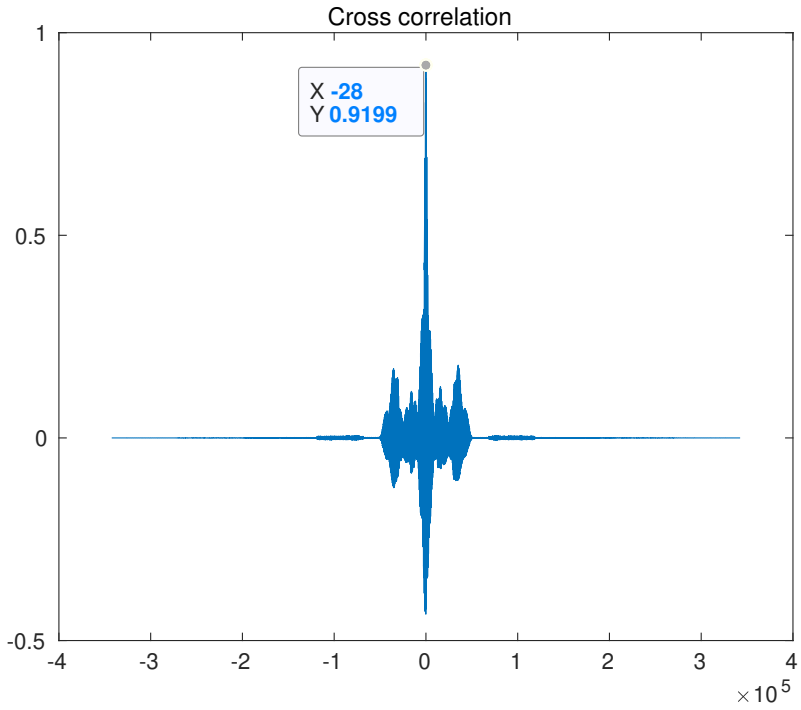

図 3.16　マイク間の距離が 24cm のときの相互相関

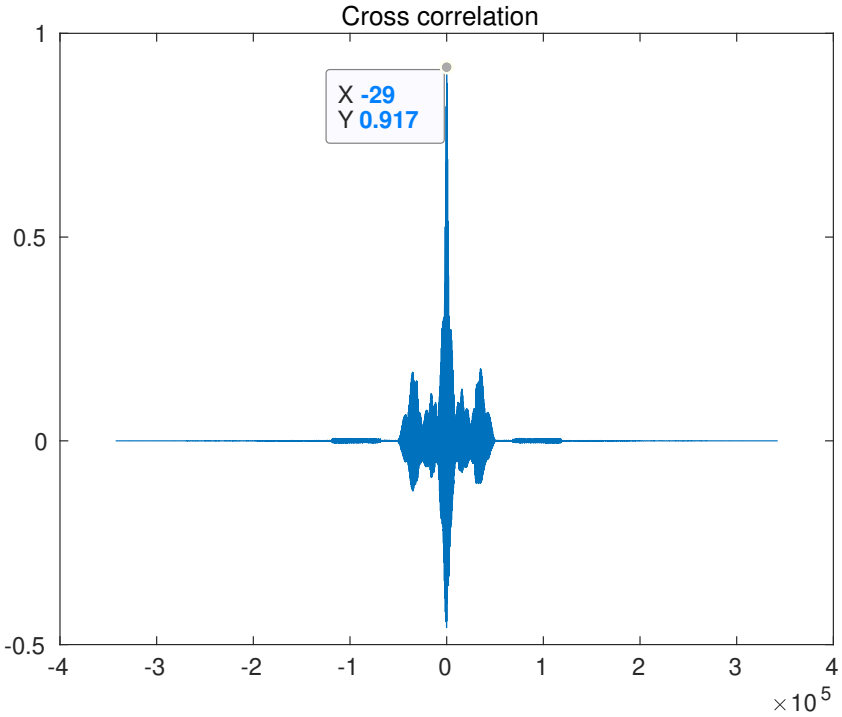

図 3.17　マイク間の距離が 25cm のときの相互相関

3.3　録音位置の違いによるマイクに到達する時刻のずれ測定

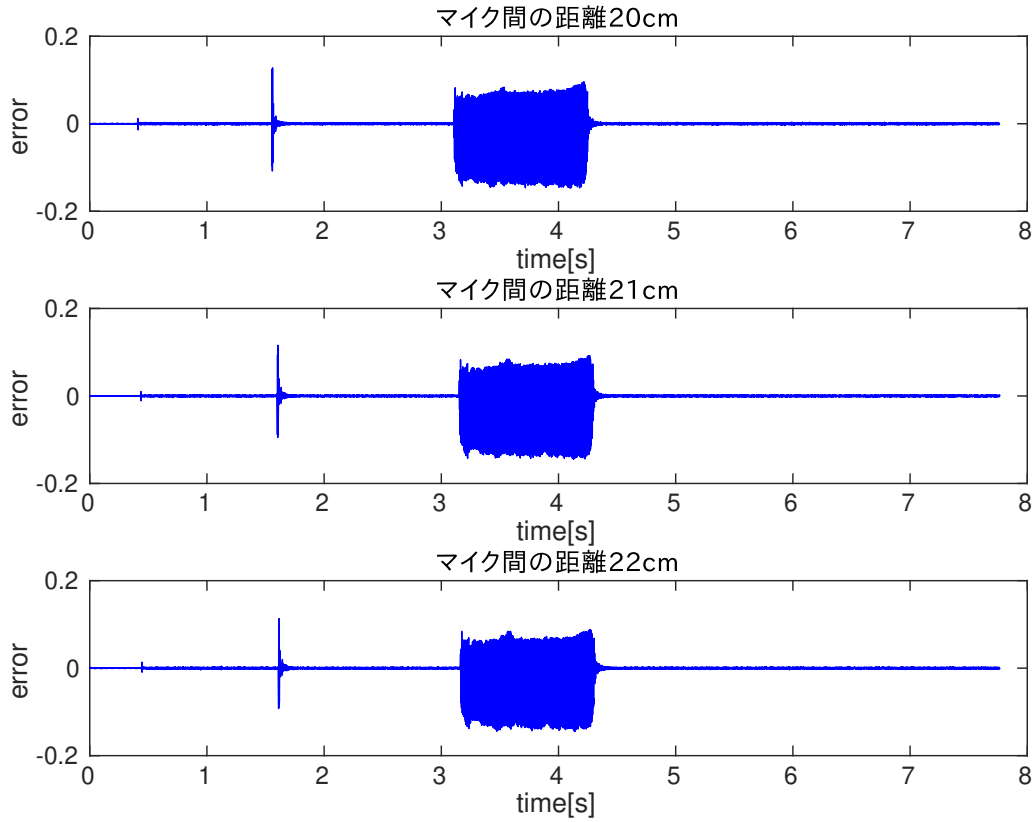

図 3.18　マイク間の距離毎の誤差

3.3　録音位置の違いによるマイクに到達する時刻のずれ測定

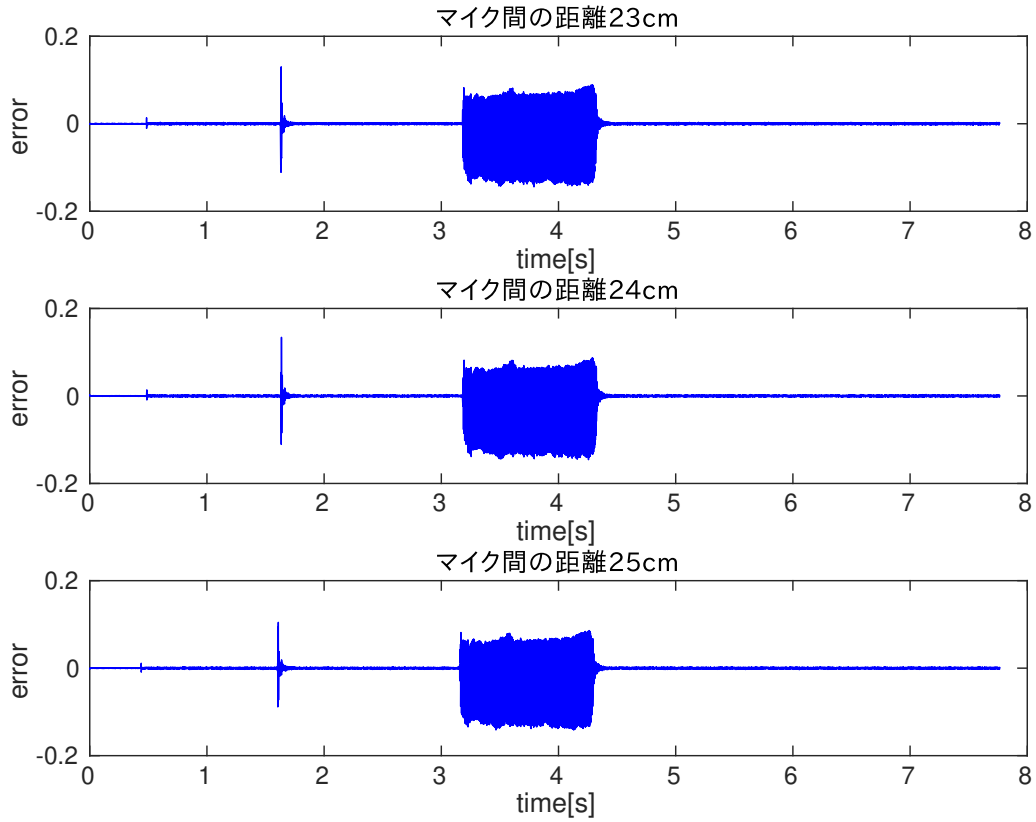

図 3.19　マイク間の距離毎の誤差

第 4 章

相互相関と適応フィルタを用いた所望音抽出システム

観測信号の相互相関をとることで，録音位置の違いから生じる観測信号の入力時刻のずれを検知したが，相関係数は 1 にならず観測信号の完全一致はできなかった．そこで，適応フィルタを用いることで，相互相関だけでは検知できなかったずれを補正し，誤差を最小化することで，信号の一致を目指す．

本章では，相互相関と適応フィルタを用いた所望音抽出システムについて述べる．

4.1 適応フィルタ

適応フィルタとは，システムのパラメータを逐次的に推定していくフィルタである．パラメータが固定されているフィルタは，特性や性質を予め知っておく必要があるため，時間とともにこれらの特性や性質が変化する信号に対しては，変化に応じてフィルタを設計し直す必要がある．しかし，その変化に応じて自動的にパラメータを変化させる適応フィルタであれば，正確に処理を行うことができる．適応フィルタの代表例として，エコーキャンセラ，自動等化器，アクティブノイズコントロール，ノイズキャンセラなどが挙げられる [9]．スピーカーからそれぞれのマイクに到達する時刻は，信号の相互相関をとることで知ることができる．しかし，それぞれのマイクに到達する音声信号は空間を通るため，歪みや雑音など周りの環境による影響を受け，相関係数は 1 未満になってしまう．この影響は，壁からの反射音なども含め，予測することができない．そのため，信号の相互相関をとるだけでは信号

の一致はできないが，適応フィルタを用いることは，未知である周りの環境の影響に対し，効果的なものであるといえる．

図 4.1 は FIR 形の適応フィルタの構成例である．この図 4.1 をもとに，適応フィルタのパラメータ推定について説明する [9][10]．入力信号 x_n，y_n はそれぞれ時刻 $t = nT$ におけるフィルタの入力信号と出力信号である．ここで，T はサンプリング周波数である．また，フィルタ係数 $a_i(n)$ は i 番目のフィルタ係数，N はフィルタ係数の数である．フィルタ係数 $a_i(n)$ はインパルス応答 a_τ が時変であることを $a_\tau(t)$ と表す．このとき，$\tau = iT$，$t = nT$ に対応し，出力 $y(n)$ は，

$$y(n) = \sum_{i=0}^{N} a_i(n)x(n-i) \tag{4.1}$$

となる．これは，

$$y(t) = \int_0^{NT} a_\tau(t)x(t-\tau)d_\tau \tag{4.2}$$

に対応する線形時変フィルタである．この時変フィルタにおける伝達関数の z 変換は，n の関数であることより，

$$a(z,n) = \sum_{i=0}^{N} a_i(n)z^{-1} \tag{4.3}$$

となる．基準信号 $d(n)$ は，出力 $y(t)$ を近づけたい理想的な信号 $d(t)$ の標本値に対応し，$d(n)$ と $y(n)$ の差である誤差信号

$$\varepsilon(n) = d(n) - y(n) \tag{4.4}$$

から，フィルタ係数 $a_i(n)$ は適応制御される．

4.2　パラメータ推定問題

パラメータ推定の問題は適当に定められた評価量に関する最小化の問題として定式化できる．図 4.2 より，評価量 J は，

$$J = E[\varepsilon^2(n)] = E[d(n) - y(n)^2] \tag{4.5}$$

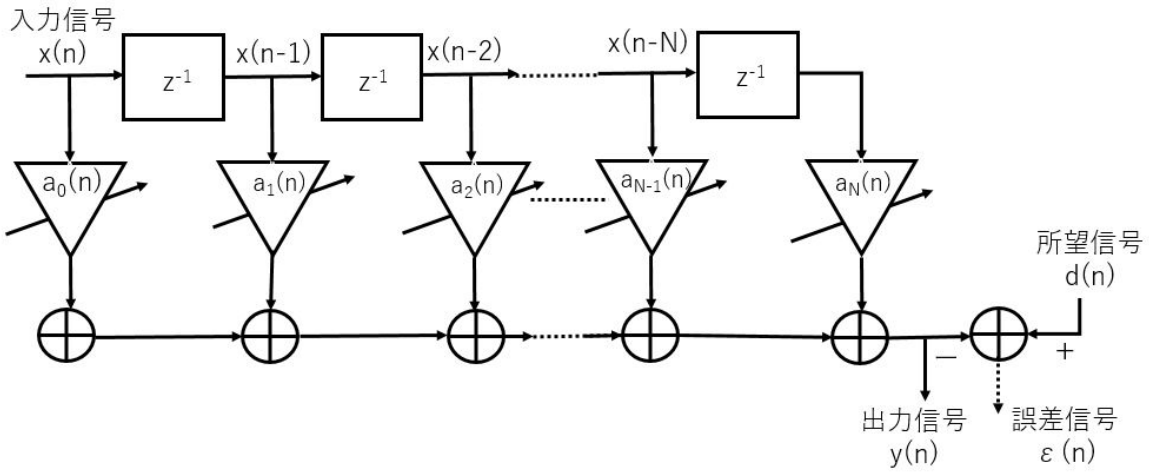

図 4.1　適応フィルタの構成

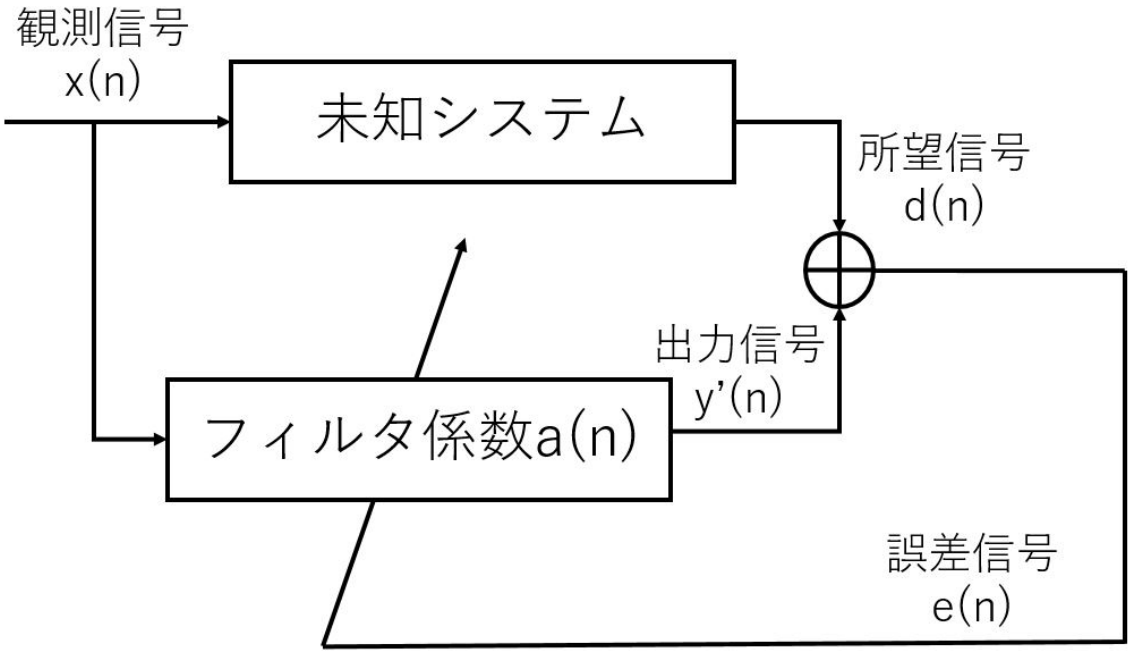

図 4.2　パラメータ推定の表現

で与えられる．式 (4.5) に，適応フィルタの入出力関係

$$y(k) = \sum_{i=0}^{N-1} a(i)x(n-i) \tag{4.6}$$

を代入し，行列表現すると a_N に関する 2 次関数

$$J = a_N^T A_{N,N}(n) a_N - 2a_N^T v_N(n) + E[d^2(n)] \tag{4.7}$$

が得られる．ここで，

$$\begin{aligned}
A_{N,N}(n) &= E[x_N(n)x_N^T(n)] \\
v_N(n) &= E[x_N(n)d(n)] \\
d(n) &= W_M^T x_M(n) \\
x_N(n) &= (x(n), x(n-1), \cdots, x(n-N+1))^T \\
x_M(n) &= (x(n), x(n-1), \cdots, x(n-M+1))^T \\
a_N &= (a(0), a(1), \cdots, a(N-1))^T \\
W_M &= (w(0), w(1), \cdots, w(M-1))^T
\end{aligned}$$

とする．式 (4.7) より J を最小化する問題は無制約最適化問題となる．$A_{N,N}(n)$ は正定値であることが知られており．2 乗平均誤差 $E[\varepsilon^2(n)]$ は h_N に関する代表的な凸関数で，唯一の最小値を持つことが示されている．時刻 n において J を最小にする推定ベクトル a_N を $a_N(opt, n)$ と表すと，$a_N(opt, n)$ は両辺を a_N で偏微分して 0 とすると得られる．式 (4.7) の両辺を a_N で偏微分すると，

$$\frac{\partial J}{\partial a_N} = 2A_{N,N}(n)a_N - 2v_N(n) \tag{4.8}$$

となり，

$$a_N(opt, n) = A_{N,N}^{-1}(n)v_N(n) \tag{4.9}$$

が得られる．式 (4.9) は Wiener-Hoff のかいと呼ばれている．ゆえに，$x_M(n)$ の自己相関ベクトルと，入力信号ベクトル $x_M(n)$ と所望信号 $d(n)$ との相互相関ベクトルが定まれば，フィルタ係数ベクトル a_N の最適値 $a_N(opt, n)$ が求まる．

4.3　所望音抽出システム

所望音の抽出方法について図 3.4 のスピーカーから出力された音を 2 つの観測信号から消音することで確認する．

まず，マイク 1 の観測信号とマイク 2 の観測信号の相互相関を取り，位相の同期を取る．位相以外のずれは適応フィルタで補正し，観測信号の完全な同期を図る．マイク 1 の観測信号を $x(n)$，マイク 2 の観測信号を $y(n)$ とする．今回，FIR 形の適応フィルタを用いて所望信号 $d(n)$ と出力信号 $y'(n)$ の誤差を最小にする．このとき，$x(n)$ と $y(n)$ は s サンプルずれていることが相互相関を取ることでわかったとすると，出力信号は $a(n)x(n+s)$ となる．今回，所望信号は $y(n)$ であることから，この所望信号 $y(n)$ と出力信号 $a(n)x(n+s)$ の誤差信号 $e(n)$ が最小になれば良い．この誤差信号の 2 乗平均値 $E[e^2(n)]$ が最小になるようにフィルタ係数 $a(n)$ を更新していく．

この適応フィルタを用いた結果，マイク 1 とマイク 2 の観測信号の誤差は図 4.3，図 4.4 のように 1×10^{-17} 未満であり，これは消音できたとみなせる．

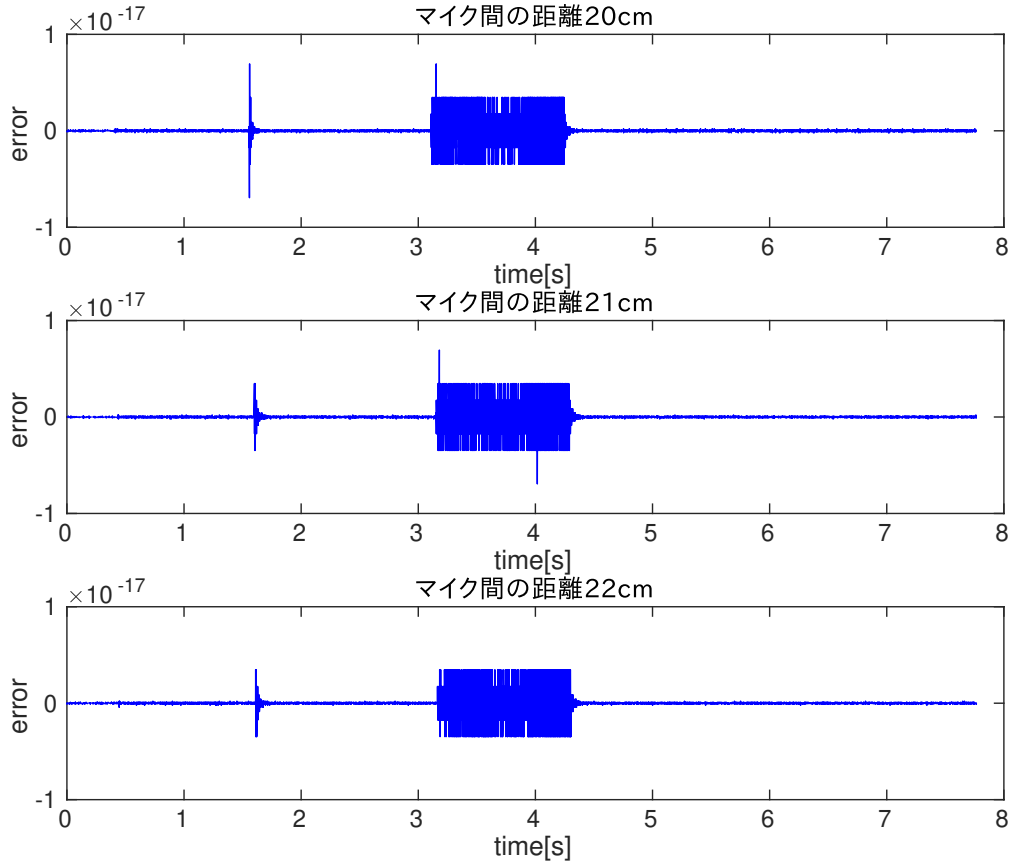

図 4.3　マイク間の距離毎の適応フィルタ処理後の誤差

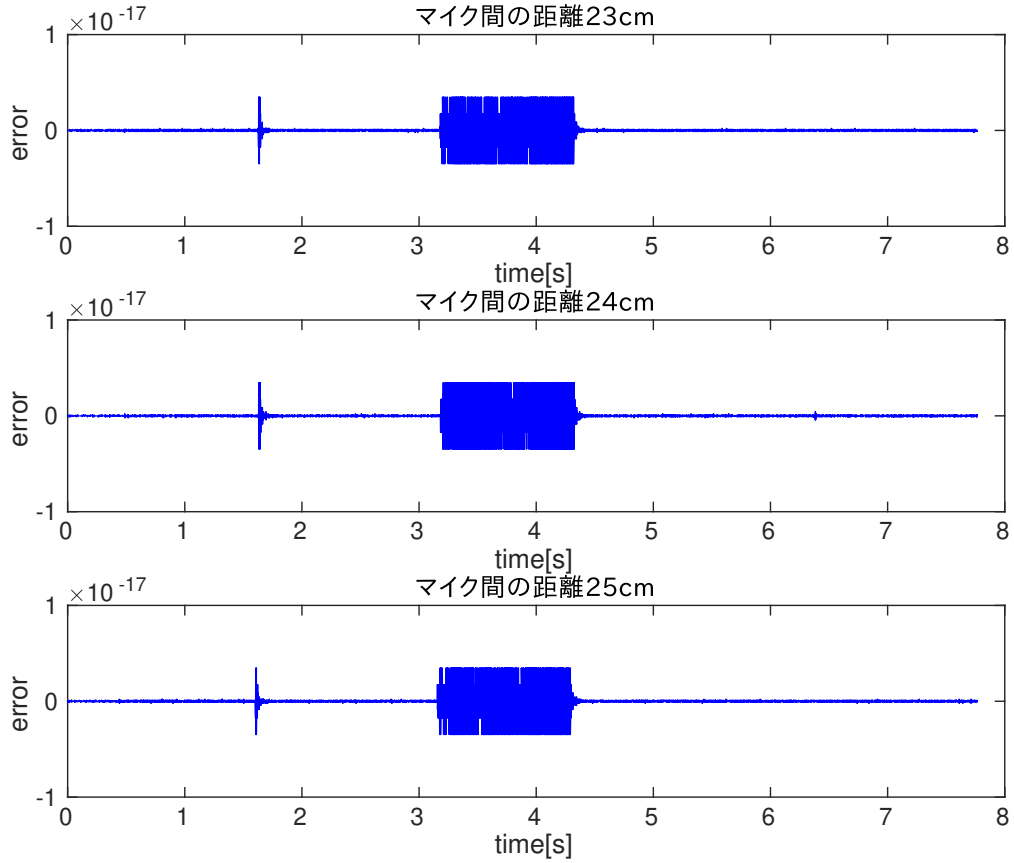

図 4.4　マイク間の距離毎の適応フィルタ処理後の誤差

第 5 章

結論

5.1 本研究のまとめ

本研究では，所望音以外の音を取り除くことで所望音を抽出する方法を提案した．録音された音声から所望音を抽出するために，観測信号の相互相関をとることでマイクに到達する時刻のずれを検知した．また，相互相関で時刻のずれを検知し，位相を合わせただけでは信号の一致はできなかったため，適応フィルタを用いて時刻以外のずれを合わせた．適応フィルタを用いたことで，マイクに到達する時刻以外のずれを合わせ，消音できたことを確認した．この提案方法により，録音位置の違う音声信号が一致し，所望音以外を取り除き，所望音を抽出できることを確認した．

5.2 今後の課題

本研究で用いた適応フィルタはあくまで理想的な適応フィルタが存在するという仮定のもとで，所望音の抽出方法を提案した．実際に適応フィルタを作成し，それを用いて所望音が抽出可能であるか確認する必要があると考える．

また，今回スピーカーとマイク 2 本を一直線上に配置した場合でのみ提案方法を確認した．今後は，スピーカーとマイクの配置が一直線上でない場合にも提案方法が可能であることを確認する必要がある．

謝辞

本研究を行うにあたり，ご指導していただいた福本昌弘教授に謹んで感謝致します．鹿児島まで研究授業を見に来てくださったときは驚きましたが，とても嬉しかったです．また，週次報告をほとんど出さず，ギリギリに慌てて実験したり，勘違いしたまま研究をしていた私を最後まで面倒見てくださり本当にありがとうございました．本研究の副査をしていただいた浜村昌則教授，植田和憲講師のお二人にも謹んで感謝致します．

また，NOC の福冨英次氏にも謹んで感謝致します．仕事終わりや休日に夜遅くまで研究のアドバイスをしていただいたり，お食事に連れて行ってもらったりと大変お世話になりました．効率のよい食事の作り方も学んだので，今後の生活に活かしていきたいと思います．

原田崇司助教にも謹んで感謝いたします．中間発表の合宿で初めてお話をし，犬嫌いとして分かち合い，ちょくちょく情報提供していただきありがたかったです．また，ご飯に連れて行ってもらったり，イベントに参加していただいたりとありがとうございました．

Bandhit.Suksiri 氏にも心より感謝致します．研究のお話や分からないところを卒業後も電話などで教えてくださり，とても感謝しています．タイの料理や外国の歌についても知ることができてよかったです．

そして，福本研究室修士 1 年の中村巴氏にも感謝いたします．先輩なのに舐めた態度でいてすみませんでした．巴氏には研究だけでなく，色んな話を聞いてくださったり，実験のために工作を手伝ってもらったりと大変お世話になりました．研究室が楽しかったのは巴氏のおかげだったと思います．

研究室の同期たちにも心より感謝致します．３年生のときはそこまででしたが，４年生になりたくさん話すようになってなんだかんだ楽しかったです．

福本研究室 3 年生の皆さん，あまり関わることはありませんでしたが，イベントの企画などありがとうございました．これからも頑張ってください．

最後に，高知工科大学で過ごした 4 年間を支えていただいたすべての方に感謝致します．

参考文献

[1] 河原達也，“音声認識技術の変遷と最先端-深層学習による-End-to-End モデル-，” 日本音響学会誌 74 巻 7 号，pp.381-386，2018.

[2] 安藤厚志，丹羽健太，北岡教英，武田一哉，“特徴量領域音源分離のためのクロススペクトル抑圧，” 電子情報通信学会，Vol.112No.369，pp.107112, Dec.2012.

[3] 鹿野清宏，中村哲，伊勢史郎，“音声・音情報のディジタル信号処理，” 昭晃堂，1997.

[4] Bandhit Suksiri，“Advanced Direction-of-Arrival Estimation for Acoustic Signal Processing and its Applications，” 令和元年度高知工科大学博士学位論文，2019.

[5] 森山佳奈，“遅延和アレーに基づく音源方向推定の研究，”http://www.asp.c.dendai.ac.jp/thesis/H13_moriyama.pdf，2020 年 2 月 12 日閲覧.

[6] 関口航平，中村圭佑，坂東宜昭，糸山克寿，吉井和佳，中臺一博，“音源到来方向・時間差を用いた非同期複数マイクロホンアレイ位置のオンライン推定，”http://sap.ist.i.kyoto-u.ac.jp/members/yoshii/papers/ipsjnc-2016-sekiguchi.pdf，第 78 回情報処理学会全国大会講演論文集，情報処理学会，2016，pp.483-484，2020 年 2 月 12 日閲覧.

[7] 三谷政昭，“やり直しのための信号数学，” CQ 出版社，2005.

[8] 澳本 拓郎，“相互相関法による音源位置推定，” 平成 22 年度高知工科大学学士学位論文，2011.

[9] 辻井重男，“適応信号処理，” 昭晃堂，1995.

[10] 羽鳥光俊，“適応フィルタの最近の動向，”https://www.jstage.jst.go.jp/article/sicejl1962/25/12/25_12_1082/_pdf，計測と制御 Vol.25No.12，2020 年 2 月 13 日閲覧.