

時間差に基づく音源方向探査の研究

指導教員 金田 豊 教授

97k c 6 1 5 山口 智識

目次

1 . はじめに	(pp.3)
2 . 時間差に基づく方向推定の理論	(pp.4)
2 - 1 . 音の到来方向を推定する	(pp.4)
2 - 2 . 自己相関関数と相互相関関数	(pp.5)
3 . RWCP 実環境音声・音響データベース	(pp.6)
3 - 1 . マイクロホンアレーデータベースについて	(pp.6)
3 - 2 . データ収録の方針	(pp.6)
3 - 3 . マイクロホンアレー 収録の概要	(pp.6)
3 - 4 . OPTOTRAK (実時間3次元位置計測装置)	(pp.8)
3 - 5 . 収録風景	(pp.9)
3 - 6 . 収録状況	(pp.10)
3 - 7 . データの種類	(pp.11)
3 - 8 . それぞれのデータの見方	(pp.14)
3 - 9 . プログラム入力データ	(pp.16)
3 - 10 . データの一覧表	(pp.17)
4 . 時間差に基づく方向推定のプログラム	(pp.19)
4 - 1 . インパルス応答の方向推定	(pp.19)
4-1-1 . インパルス応答の方向推定プログラムの手順	(pp.19)
4-1-2 . プログラム使用上の注意	(pp.20)
4-1-3 . インパルス応答の方向推定プログラム	(pp.21)
4-1-4 . インパルス応答の方向推定プログラムの実行結果	(pp.24)
4 - 2 . 移動音声の方向推定	(pp.26)
4-2-1 . 移動音声の方向推定プログラムの手順	(pp.26)
4-2-2 . 移動音声の方向推定プログラム	(pp.28)
4-2-3 . 移動音声の方向推定プログラムの実行結果	(pp.35)
5 . まとめ	(pp.40)
参考文献	(pp.41)
謝辞	

1 . はじめに

目的とする音がどの方向から到来しているかを推定する音源方向推定の技術は、利用範囲の広い技術である。ロボットの耳機能や通信会議システム等、様々な分野で利用されている。

この技術の問題点として、音源が離れた位置にある場合、直接音に加えて部屋の残響音が影響してくることがあげられる。例えば、マイクロホンとスピーカーを使用する、ハンズフリー技術には、複数の話者、複数の雑音源、部屋の残響が存在する環境の中から対象の話者を同定し、その音だけを取り出す技術が不可欠となる。しかし、現存する技術は残響がある部屋では良好に動作しない。これらの技術を達成するためには、マイクロホンアレー（複数のマイクロホン素子から構成される受音器）を用いた信号処理が有効であることが言われている。

人間は騒音の中から注目した音を聞き分け、また音の方向を見いだす機能を有している。これらの機能は、ふたつの耳で捉えた音の空間的情報を利用した信号処理（両耳機能）であると考えられている。このことから、音の空間情報を効果的に捉えることのできるマイクロホンアレーを用いた信号処理（多点受音信号処理）は、方向推定に有望な技術であると考えられる。

マイクロホンアレーを用いた研究の共通のデータベースとして、近年「RWCP実環境音声・音響データベース マイクロホンアレーデータベース」が構築された。このデータベースには、まだ十分な評価がなされていないという問題がある。

音源方向推定の手法には様々なものがあるが、本研究では音源方向を推定するプログラムを時間差に基づく方法で作成し、マイクロホンアレーデータベースから様々な環境でのデータを読み取り、プログラムとデータベースそれぞれについて評価した。また、残響が方向推定に与える影響を評価して解析した。

2 . 時間差に基づく方向推定の理論

2 - 1 . 音の到来方向の推定

s 方向から到来する平面波を距離 d だけ離れて設置された二つのマイクロホンで受信することを考える。この様子を図 1 に示す。

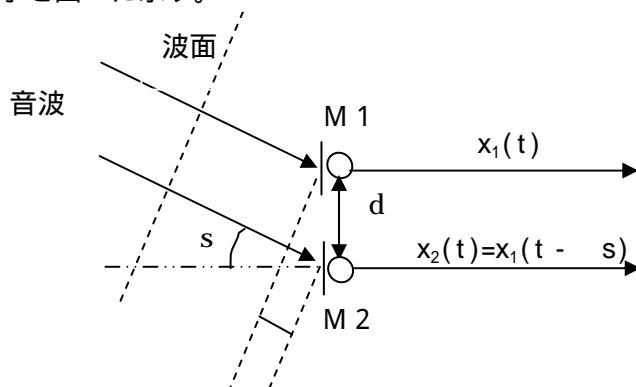


図 2.1.1 : 音の到来方向推定モデル

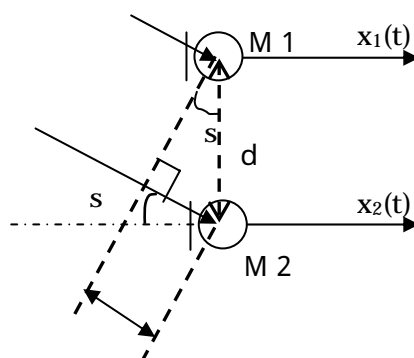


図 2.1.2 : マイクロホンの部分の拡大図

図 2.1.1 の s 方向から到来した音波は、まずマイク M1 において受信される。次に、音波は図 2.1.1 に示した距離 $d \sin s$ だけ進んでマイク M2 に到達する。図 2.1.2 より、この距離 $d \sin s$ は、

$$d \sin s \quad \dots (1)$$

と表される。したがって、M2 での受信信号 $x_2(t)$ は M1 での受信信号 $x_1(t)$ と比べて音波が距離 $d \sin s$ だけ進行するのに要する時間 s だけ遅れた信号となっている。即ち

$$x_2(t) = x_1(t - s) \quad \dots (2)$$

$$s = (d \sin s) / c \quad \dots (3) \quad c : \text{音速}$$

の関係が成立する。したがって逆に、信号 $x_1(t)$ と $x_2(t)$ の間の時間差 s がわかれば、音波の到来方向 s は次式から求められる。

$$s = \sin^{-1} (c s / d) \quad \dots (4)$$

時間差 s は、 $x_1(t)$ 、 $x_2(t)$ との相互相関関数 $r_{12}(\tau)$ から求めることができる。

$$\begin{aligned} r_{12}(\tau) &= (1/N) \{x_1(t) \cdot x_2(t+\tau)\} \\ &= (1/N) \{x_1(t) \cdot x_1(t+\tau - s)\} \dots (5) \end{aligned}$$

ただし、信号 $x_1(t)$ 、 $x_2(t)$ はそれぞれ標本化されたものとする。また、 t は離散時間を表し、 τ はN個の標本点 ($\tau=1,2,\dots,N$)の総和を表す。 $r_{12}(\tau)$ の最大値を与える τ を求めれば s が得られ ($\tau = s$ で $r_{12}(\tau)$ は最大値をとる。(2-2 参照)) 式(4)に代入して、音波の到来方向 s が推定できる。

2 - 2 . 自己相関関数と相互相関関数

自己相関関数と相互相関関数について述べる。

自己相関関数とは一つの信号波形に対して相関処理を行い、一つの信号波形中に繰り返し(周期性)があるかどうかを調べることである。

信号波形 $x_1(t)$ の自己相関関数は

$$r_{11}(\tau) = (1/N) \{x_1(t) \cdot x_1(t+\tau)\} \dots (6)$$

で表される。

この式は、N個のサンプリング値からなる一つの信号波形 $x_1(t)$ と、それを遅れ時間 τ だけずらした信号波形 $x_1(t+\tau)$ に対して、それぞれを掛け合わせて累積し、平均化するという処理を行っている。この場合、遅れ時間がない ($\tau = 0$) ときには、 $x_1(t)$ と $x_1(t+\tau)$ とは同じ波形になるのでぴったりと重なり、自己相関関数 $r_{11}(\tau)$ は最大値となる。

相互相関関数とは異なる二つの信号波形に対して相関処理を行い、二つの類似度を調べることである。

信号波形 $x_1(t)$ と信号波形 $x_2(t)$ の相互相関関数は

$$r_{12}(\tau) = (1/N) \{x_1(t) \cdot x_2(t+\tau)\} \dots (7)$$

で表される。基本的には、自己相関処理と同じように信号波形 $x_1(t)$ と遅れ時間 τ だけずらした信号波形 $x_2(t+\tau)$ において、それぞれ対応する部分を掛け合わせて累積し、平均化する処理を行う。今回の時間差に基づく方向推定では一つの音源をマイクロホン M1、M2 で受信するものなので、上に記したように、M2 での受信信号 $x_2(t)$ は M1 での受信信号 $x_1(t)$ と比べて音波が距離 s だけ進行するのに要する時間 s だけ遅れた信号となっている。よって、式(7)は式(5)となり、さらに自己相関関数 $r_{11}(\tau)$ で表すと

$$\begin{aligned} r_{12}(\tau) &= (1/N) \{x_1(t) \cdot x_1(t+\tau - s)\} \\ &= r_{11}(\tau - s) \dots (8) \end{aligned}$$

となる。式(8)からも分かるように、相互相関関数 $r_{12}(\tau)$ は自己相関関数 $r_{11}(\tau)$ を s ずらしたものであることが分かる。さらに、自己相関関数 $r_{11}(\tau)$ では $\tau = 0$ で最大値をとることが分かっている。したがって、相互相関関数 $r_{12}(\tau)$ では、 $\tau = 0$ より s ずらした

$\tau = s$ で最大値をとる。 [1] , [2]

3 . R W C P 実環境音声・音響データベース

3 - 1 . マイクロホンアレーデータベースについて

音源方向探査の研究において、共通なデータベースが存在しないため、各研究者がいろいろな手法を別個の室内で検討しているため、手法の優劣の比較がしにくいことが問題となっている。このことを解決する為、手法の効果を比較できるような共通のデータベースとして、近年「R W C P 実環境音声・音響データベース マイクロホンアレーデータベース」が構築された。この章では、研究で使用したデータベースについて詳しく説明する。このデータベースについては、オフィシャルのホームページ（参照：<http://www.slt.atr.co.jp/~tnishi/DB/micarray/index.htm>）があるのだが、必要なデータが少々探しにくい為、この章で分かりやすくまとめた。特に、方向推定のプログラムに使用するデータ一覧を 3 - 10 でまとめているので参照するとよい。

3 - 2 . データ収録の方針

ドライソース（色々な音を無響室で録音したもの）と種々の部屋のインパルス応答を畳み込むことで、多種類の音環境をシミュレートすることができる。このことを目的として、このデータベースは種類の室内のインパルス応答を測定したものが中心となっている。また、畳み込みでシミュレートできない音源である、移動音源などは実音場で収録が行われている。このデータベースには、9種類の部屋でのインパルス応答と5種類の部屋での実収録した音声、および5種類の部屋での実収録した移動音源などが存在する。

3 - 3 . マイクロホンアレー 収録の概要

収録では以下の2種類のマイクロホンアレー（受音器）が用いられている。

circle: 円形マイクロホンアレー(16channel)

circleline: 円形+直線マイクロホンアレー(30channel)

図 3.3.1 にマイクロホン素子の配置図を示す。素子番号 1 が正面方向である。円形アレーの直径は 30cm でその円周上に等間隔に 16 素子（#1～#16）が配置してある。また直線型アレーは素子間隔 2.83cm で 14 素子（#17～#30）配置されている。

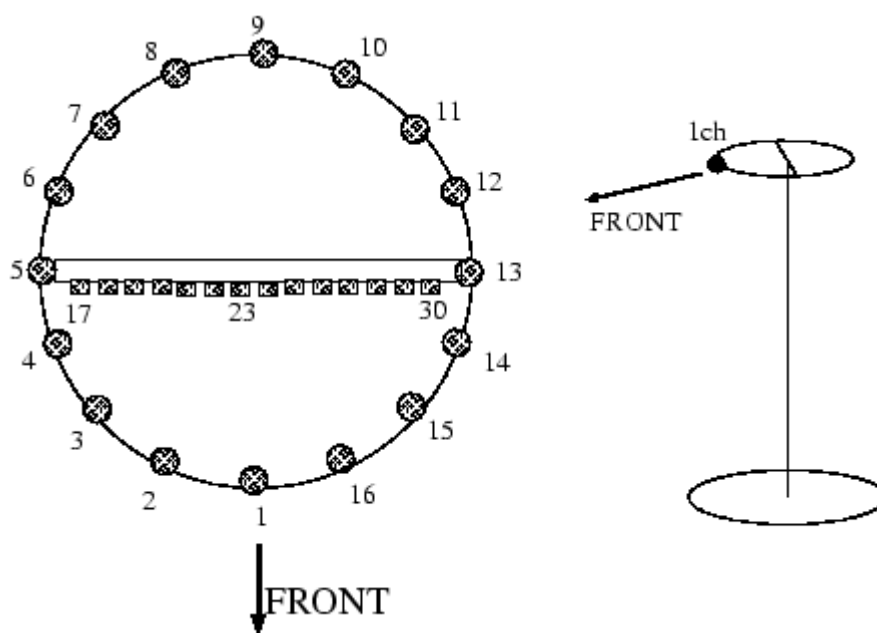


図 3.3.1 : マイクロホンアレーの素子配置

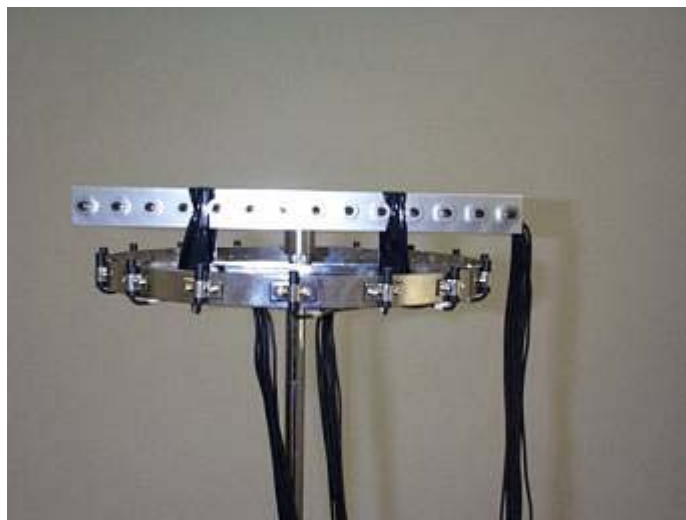


図 3.3.2 : 円形 + 直線マイクロホンアレー

3 - 4 . OPTOTRAK (実時間 3 次元位置計測装置)

実時間 3 次元位置計測装置 OPTOTRAK について説明する。OPTOTRAK は、移動する計測対象物に取り付けられた赤外線発光ダイオード (赤外線マーカ) を正確に追跡し、リアルタイムに 3 次元座標を計測する装置である。ロボット工学・身体運動研究・医療技術など多様な分野で利用されている。データ収録に際し、音源の移動に対して、各時刻における音源の位置を正確に計測するために OPTOTRAK が使用されている。OPTOTRAK の計測手順は次の通り。

- 1 . 計測対象物に取り付けた赤外線マーカを発光させる。
- 2 . 発光された赤外線を位置センサが計測する。
- 3 . 計測されたデータは、システム制御装置で計算され、リアルタイムに 3D データが得られる。
- 4 . 計測結果は、PC の画面に表示される。

収録に使用された OPTOTRAK 3020 の主な特徴

- 1 . 最大で 256 個の赤外線ダイオードを実時間で追跡
- 2 . 3 次元位置データと剛体位置データのリアルタイム表示が可能
- 3 . 最大 0.01mm の分解能、0.1mm の RMS 精度
- 4 . 最大 400Hz のデータ収録が可能
- 5 . ± 60 度まで計測可能



図 3.4.1:OPTOTRAK

参照

<http://www.asco.co.jp/asco/products/optotrak.html>

3 - 5 . 各室内における収録風景



図 3.5.1 : 残響可変室



図 3.5.2 : 会議室

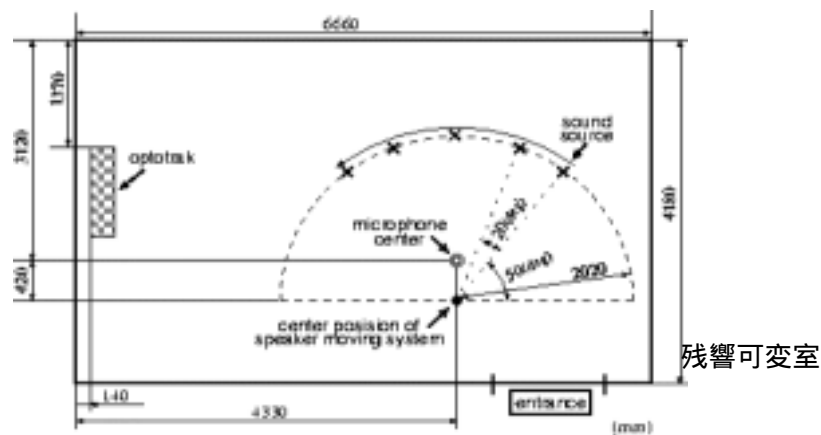
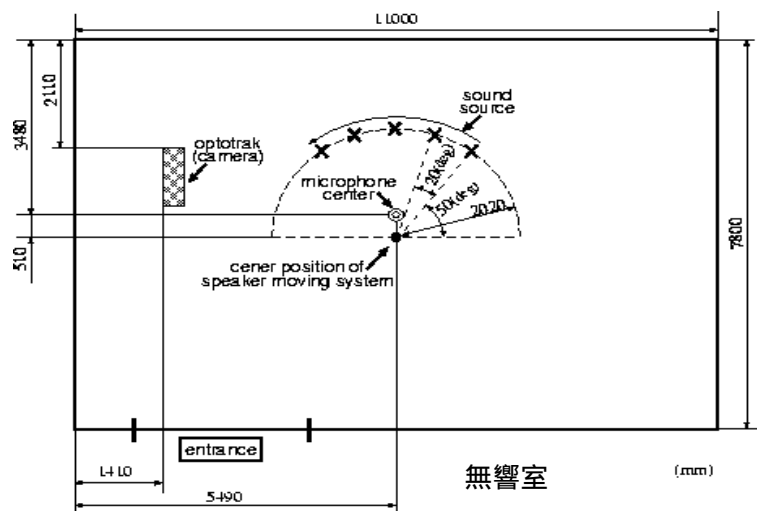
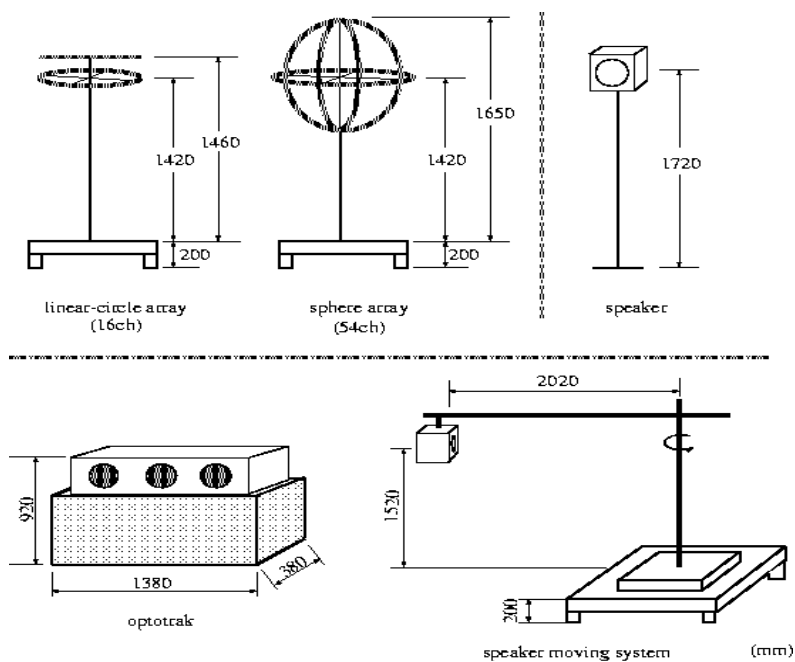


図 3.5.3 : 畳部屋 (大)



図 3.5.4 : 畳部屋 (小)

3 - 6 . 測定器材と収録状況の寸法図



3 - 7 . データの種類

本データベースでは以下に示す5項目が収録されている。データはインパルス応答が float 型(4バイト)、音声データが short 型(2バイト)で、サンプリング周波数は 48KHz である。

1 . 固定音源

- (ア) 固定インパルス応答 ... マイクロホンアレーを中心に2mの距離にスピーカを配置し、スピーカの方向が異なった複数のインパルス応答を測定
- (イ) 固定音声 ... マイクロホンアレーを中心に正面2mの距離にスピーカを配置して収録

2 . 移動音源

- (ア) 固定インパルス応答 ... 専用移動音源装置を用いてスピーカを軌道上の数箇所に固定してインパルス応答を測定
- (イ) 移動音声 ... 専用移動音源装置を用いて音源を 50 度から 130 度付近まで手動で音源を移動(各時刻の音源位置は OPTOTRAK により記録)

3 . 拡散音場

- (ア) インパルス応答 ... スピーカを壁に向けて配置して拡散音場を再現(マイクロホンに直接波が到来しない状況)

4 . 背景雑音

- (ア) 背景雑音 ... スピーカから音が出ていない状態の背景雑音(暗騒音、空調音など) (より忠実に実音場を再現するためにドライソースとインパルス応答を畳み込んだ後に背景雑音を加算)

5 . ピストンホン

- (ア) 素子補正用 ... マイクロホンの素子は1つずつの固体差がある。そこでピストンホンにより素子補正を行い個体差を補正

インパルス応答

図にサンプルのインパルス応答を示す。

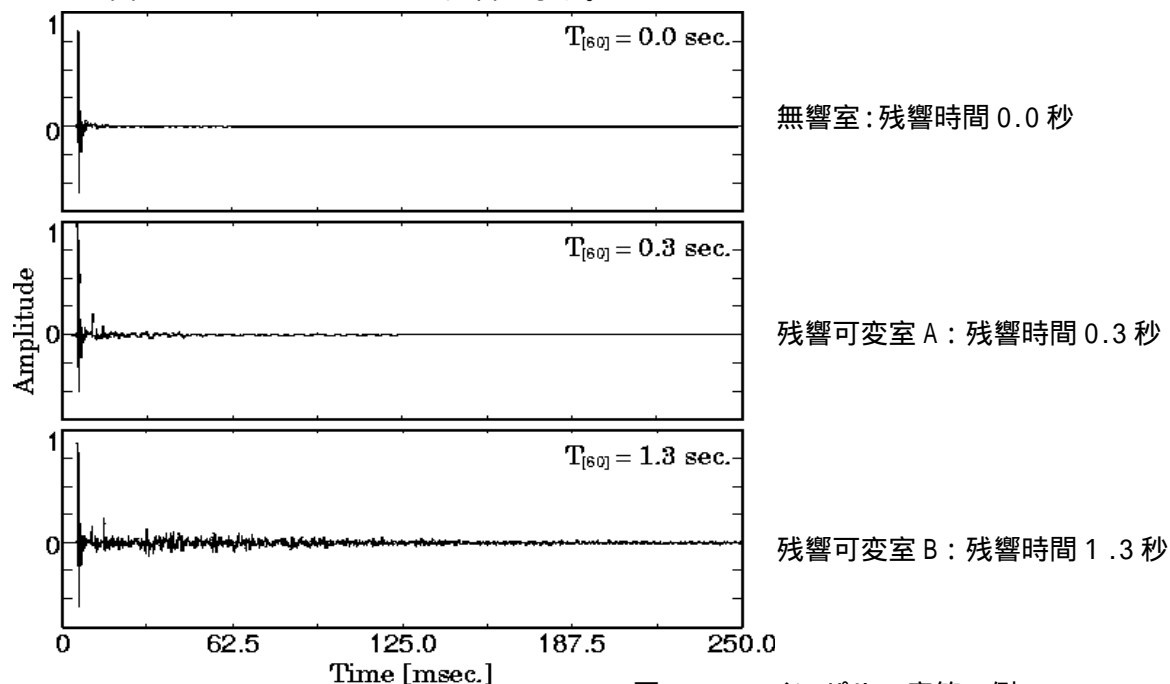


図 3.7.1：インパルス応答の例

音声データ

音声データは ATR 音韻バランス 50 文（日本人男性 1 名）と TIMIT (The DARPA TIMIT Acoustic-Phonetic Continuous Speech Corpus) データベース中に含まれる 10 文 × 4 名（外国人男性 2 名、外国人女性 2 名）のあわせて 90 文を収録した。音声データは全て、各音声ごとに ZIP 形式で圧縮されている。ファイルを解凍すると、各音声ごとにディレクトリ構造になっており、その中に各マイク素子に対応するファイル（例 .mmysda01.1 ~ .mmysda01.30）が格納されている。無響室と残響可変室にて収録した音声波形の一例を図 3.7.2 に示す。無響室では振幅が小さいのに対し、残響可変室では残響によって振幅が大きくなっている。

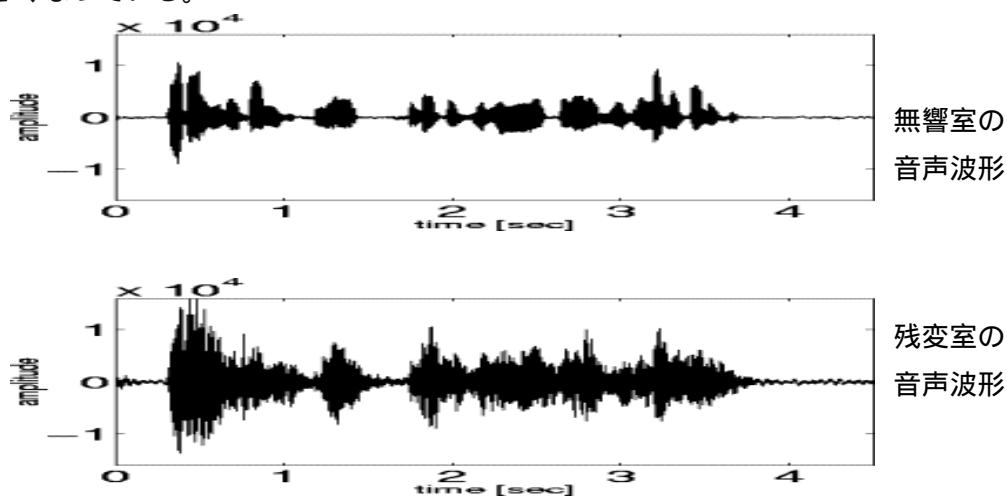


図 3.7.2：音声波形の例

移動音源

移動音源は図 3.7.3 に示す移動音源装置を用いて収録された。棒の先にスピーカを固定して、支柱を中心としてこの棒を回転させる。マイクロホンアレーは、この支柱の根元付近（3-6 節の図参照）に配置してある。3-7 節の移動音源に対する固定のインパルス応答もこの軌道上の数点において収録されたものである。移動音声に対しては人力により棒を動かすことによりデータ収録が行われ、50 度から 130 度付近まで手動で音源を移動している。そこでの各時刻においての音源位置は OPTOTRAK を用いて測定され、収録音声と同期させてデータベース内に保存されている。音源位置データは float 型（4 バイト）、周波数は 100Hz で x, y, z 座標が file.x, file.y, file.z となっている。座標は音源装置の中心からの距離で単位は [mm] である。図は移動音源装置と OPTOTRAK を用いた音源移動の軌跡を示す。

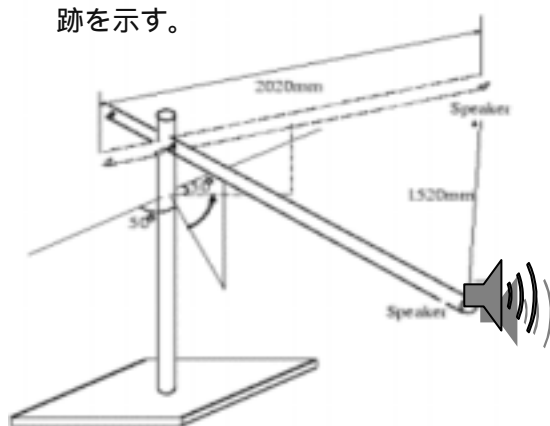


図 3.7.3：移動音源装置

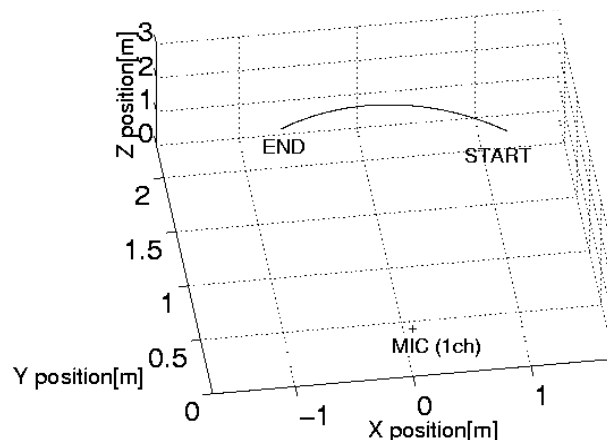


図 3.7.4：移動音源の軌跡

拡散音場

拡散音場については、図 3.7.6 に示すようにスピーカを壁側に向けてインパルス応答の測定を行った。

背景雑音

会議室において背景雑音を 10 秒間 2 セット収録した。

ピストンホン ピストンホンは素子校正用のキャリブレーション用のデータとして収録。ピストンホン ONSOKU TYPE2126 を使用した。



図 3.7.5：拡散音場の収録風景（1）



図 3.7.6：拡散音場の収録風景（2）

3 - 8 . それぞれのデータの見方

DVD のデータのフォルダを開くと、
micarray/MICARRAY/data1/ circle/ANE/imp010/imp010.1 というような構造になっている。
それぞれのフォルダの意味を次に示す。

data1 : インパルス応答	circle : 円形マイクロホンアレー
data2 : 音声	cirline : 円形 +
data3 : 移動インパルス応答	直線マイクロホンアレー
data4 : 移動音声	
data5 : 拡散音源	ANE : 無響室
data6 : 背景雑音	E1A : 残変室シリンダ A
data7 : ピストンホン	E1B : 残変室シリンダ B
	E1C : 残変室シリンダ C
imp : インパルス応答	E2A : 残変室パネル A
mmysda : 音声 (日本人男性)	E2B : 残変室パネル B
faks : 音声 (外国人女性 1)	OFC : 会議室
felc : 音声 (外国人女性 2)	JR1 : 畳部屋大
mdeb : 音声 (外国人男性 1)	JR2 : 畳部屋小
mreb : 音声 (外国人男性 2)	

インパルス応答

インパルス応答は float 型 (4 バイト) で、データベース中のフォルダ構造は、例えば

data1/circle/ANE/imp010/imp010.1

のようになっている。この場合は

インパルス応答 / 円形マイクロホンアレー / 無響室 / 10° / 素子番号 1 のデータを意味している。

また、data1 (インパルス応答) だけ、円形アレーと円形 + 直線アレーの 2 種類で収録されたデータがあるが、残るデータ (data2 ~ data7) は全て円形 + 直線アレーのデータのみである。

音声

音声データは short 型 (2 バイト) で、50 文 × 日本人 1 名と 10 文 × 外国人 4 名のあわせて 90 文が収録されている。データファイルは各音声ごとに ZIP 形式で圧縮されている。(mmysda01.zip ~ mre90.zip) 解凍すると、各音声ごとにディレクトリ構造になっており、その中に各マイク素子に対応するファイル (例 .mmysda01.1 ~ mmysda01.30) が格納されている。

音声のデータベース中のフォルダ構造は、例えば

data2/cirline/E2A/ mmysda40.zip

で、解凍後は

data2/cirline/E2A/ mmysda40/mmysda40.30

のようになっている。この場合は

音声 / 円形 + 直線マイクロホンアレー / 残変室パネル A

/ 日本人男性音声データの 40 個目 / 素子番号 30 のデータ

を意味している。

また、cirline のデータは、円形 + 直線の 30 素子分のデータであって、直線のみではない。

移動音声

移動音声のデータも、音声と同様に圧縮されている。ファイルを解凍すると、音声データの他に、OPTOTRAK により記録された各時刻の音源位置データ、file.x, file.y, file.z が追加されている。それぞれ x 座標, y 座標, z 座標を表している。座標は移動音源装置の中心を原点としたもので、単位は [mm]、0.01mm 単位で記録されている。(注: ホームページ上には「受音器の中心からの距離」と書かれているが、検討した結果こちらの方が正しいと思われる。)

移動音声のフォルダ構造は音声と同じ構造である。

音源位置データのフォルダ構造は、例えば

data3/cirline/0FC/ faks60/ faks60.x

となっており、この場合は

移動音声 / 円形 + 直線マイクロホンアレー / 会議室

/ 外国人女性 1 の音声データの 60 番目 / x 座標の音源位置データ

を意味している。

3 - 9 . プログラム入力データ一覧

方向推定に必要なデータをここにまとめた。

円形アレー : 直径 30cm、円周上に等間隔に 16 素子配置

直線アレー : 素子間隔 0.0283[m]、14 素子配置

インパルス応答 : float 型 (4 バイト)

音声データ : short 型 (2 バイト)

音源位置データ : float 型 (4 バイト)

サンプリング周波数 : 48[kHz]

移動音源装置の中心軸からマイクロホンアレーの中心までの距離

無響室 : 510[mm]

残響可変室 : 420[mm]

3 - 10 . 収録データの一覧表

表記の注釈

RT:	室内の残響時間
CH:	マイクロホンの素子数
Ang:	収録データの方向数
Len:	1素子あたりの長さ（音声では全体の文数：sent.=センテンス）
Byte:	ファイルのバイトオーダー（全てリトルインディアン表記）
Total:	全容量

	部屋	RT	CH	Ang	Len	Byte	Total
/MICARRAY/data1 ... インパルス応答 (DVD-ROM VOL.1)							
+--/circle							
----	ANE/imp000 無響室	0.00	16	36	5000	4	11.2MB
----	E1A/imp000 残変室(シリング)A	0.12	16	1	12000	4	753kB
----	E1B/imp000 残変室(シリング)B	0.31	16	19	30000	4	34.9MB
----	E1C/imp000 残変室(シリング)C	0.38	16	19	36000	4	41.8MB
+--/cirline							
----	E2A/imp000 残変室(パネル)A	0.30	30	9	30000	4	31.0MB
----	E2B/imp000 残変室(パネル)B	1.30	30	9	125000	4	128.8MB
----	OFC/imp000 会議室	0.78	30	9	75000	4	77.3MB
----	JR1/imp000 畳部屋(大)	0.60	30	9	60000	4	61.9MB
----	JR2/imp000 畳部屋(小)	0.47	30	7	45000	4	36.1MB
/MICARRAY/data2 ... 音声 (E2A...DVD-ROM VOL.1 , E2B,OFC,JR1,JR2...DVD-ROM VOL.2)							
+--/cirline/							
----	E2A/mmysda01 残変室(パネル)A	0.30	30	1	90sent.	2	1.2GB
----	E2B/mmysda01 残変室(パネル)B	1.30	30	1	90sent.	2	1.2GB
----	OFC/mmysda01 会議室	0.78	30	1	90sent.	2	1.2GB
----	JR1/mmysda01 畳部屋(大)	0.60	30	1	90sent.	2	1.2GB
----	JR2/mmysda01 畳部屋(小)	0.47	30	1	90sent.	2	1.2GB

	部屋	RT	CH	Ang	Len	Byte	Total
/MICARRAY/data3 ...移動インパルス応答 (DVD-ROM VOL.1)							
+-- /cirline							
+---	ANE/imp000	無響室	0.00	30	5	5000	4 2.9MB
+---	E2A/imp000	残変室(パネル)A	0.30	30	5	30000	4 17.2MB
+---	E2B/imp000	残変室(パネル)B	1.30	30	5	125000	4 71.6MB
+---	OFC/imp000	会議室	0.78	30	5	78000	4 43.0MB
+---	JR1/imp000	畳部屋(大)	0.60	30	5	60000	4 34.4MB
/MICARRAY/data4 ...移動音声 (DVD-ROM VOL.3)							
+-- /cirline							
+---	ANE/mmysda01	無響室	0.00	30	1	90sent.	2 1.2GB
+---	E2A/mmysda01	残変室(パネル)A	0.30	30	1	90sent.	2 1.2GB
+---	E2B/mmysda01	残変室(パネル)B	1.30	30	1	90sent.	2 1.2GB
+---	OFC/mmysda01	会議室	0.78	30	1	90sent.	2 1.2GB
+---	JR1/mmysda01	畳部屋(大)	0.60	30	1	90sent.	2 1.2GB
/MICARRAY/data5 ...拡散音源 (DVD-ROM VOL.1)							
+-- /cirline							
+---	OFC/imp_rev	会議室	0.78	30	1	75000	4 8.6MB
/MICARRAY/data6 ...背景雑音 (DVD-ROM VOL.1)							
+-- /cirline							
+---	OFC/ambient1	会議室	0.78	30	1	480000	2 27.5MB
+---	OFC/ambient2	会議室	0.78	30	1	480000	2 27.5MB
/MICARRAY/data7 ...ピストンホン (DVD-ROM VOL.1)							
+-- /cirline							
+---	Equalize	-	-	30	1	144000	2 8.3MB

4. 時間差に基づく方向推定のプログラム

実際にマイクロホンアレーデータベースから、インパルス応答、移動音声のデータを読み取り、方向推定のシミュレートを行う。

4-1. インパルス応答の方向推定

マイクロホンアレーデータベースから、インパルス応答のデータを2チャンネル分読み込み、プログラムにより方向推定を行う。

4-1-1. インパルス応答の方向推定プログラムの手順

1). パラメータの設定

```
c=340[m/s]           % 音速
d=0.0283*10[m]      % 2素子間距離
Fs=48000[Hz]        % サンプリング周波数
```

2). データの読み込み

```
読み込むデータ micarray/MICARRAY/data1/cirline/Ane/imp070/imp070.18
micarray/MICARRAY/data1/cirline/Ane/imp070/imp070.28
```

上記の2つのファイルを data1 ~ data2 として読み込む。上から順に、インパルス応答(無響室)の18素子のデータ、インパルス応答(無響室)の28素子のデータである。

3). 読み込んだ波形の表示

```
figure(1)に読み込んだ2つのデータ波形を色違いで表示
```

4). 相関関数の計算

まず相関の計算長を決める。次に時間差 τ の範囲を決める。マイク M 1 での受信信号 $x_1(t)$ と マイク M 2 での受信信号 $x_2(t)$ の間の時間差 s の最大値 (\max) は音波が図 4.1.1 のように、直角に到来してきた場合である。

$$\max = d / c \quad (c : \text{音速}) \cdots (9)$$

式(9)は、単位が[秒]であるので、[サンプル]にするには、サンプリング周波数 F_s をかければよい。式(5)の相互相関関数 $r_{12}(\tau)$ は、 s で最大値となることがわかっているので、 $r_{12}(\tau)$ が最大になる値は、 \max より大きくなることはない。よって τ に \max を含むような範囲をとると、 $r_{12}(\tau)$ が最大になる値を含むため、正常な結果を得ることができる。

(pp. 5 参照)

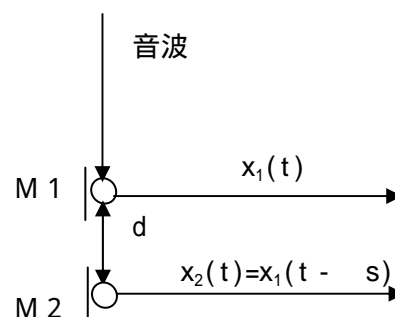


図 4.1.1 : 音の到来方向推定モデル
(到来方向直角の場合)

5). 相関関数を表示する

figure(2)に相互相関関数のグラフを表示。このグラフの最大値が、時間差 となる。MATLAB のコマンドウィンドウより、saidaiti が に相当する。

6). 時間差により角度を求める

時間差より角度を計算し、コマンドウィンドウに表示する

4-1-2. プログラム使用上の注意

このプログラムを使用する場合、読み込むフォルダの構造をを次のようにしなければならない。

D:¥array¥ ¥ ¥ ¥

- : imp,onsei ,idouonsei ,... (データの種類)
- : Ane,E2a,E2b,... (部屋の種類)
- : imp070,mmysda03,... (ファイルの種類)
- : imp070.18,mmysda03.28,... (ファイルの素子番号)

また、本研究では cirline のデータのみを扱った。

以下に使用したプログラムを示す。

4-1-3. インパルス応答の方向推定プログラム

```

% imp

% % マイクロホンアレーのインパルス応答データの方向検出プログラム
% 2002.2.8 YAMAGUCHI

% マイクロホンアレーデータベースからのインパルス応答の2個のファイル
% を読み取り、その後相関処理を行い、個々のデータの波形とその相関波形
% を表示し、角度を求めて音波の到来方向を推定する

clf % 前の図を消す

% 初期設定

c=340 % 音速[m/s]
d=0.0283*10 % 素子間距離[m]
Fs=48000 % サンプリング周波数[Hz]

% データの読み込み

rn='Ane'; % 部屋の指定（無響室）
fn='070'; % ファイル指定（方向）

fname1=['d:%array%imp%' rn '%imp' fn '%imp' fn '.18']; % インパルス応答の18素子
% 目のデータ

fid=fopen(fname1,'r'); % 指定されたファイルを開く。'r'は読み込み
data1=fread(fid,'float'); % バイナリデータ（文字形式以外のデータ）を読み込む

fname2=['d:%array%imp%' rn '%imp' fn '%imp' fn '.28']; % インパルス応答の28素子
% 目のデータ

fid=fopen(fname2,'r');
data2=fread(fid,'float');

```

```
fclose(fid); % ファイルを閉じる

% 波形の表示

figure(1) % 読み込んだインパルス応答データの波形 図 4.1.2

plot(data1)

hold on % 前の図に次の図を重ね書きする

plot(data2, 'r')

xlabel('points')
ylabel('amplitude')
axis([320 600 -0.5 0.5 ]);
grid

%hold off

% 相関の計算

np=4800; % 相関の計算長 N

taumax=fix(d/c*Fs); % 信号の時間差の最大値
% fix は()内を 0 方向へ丸める (ex. 24.5 24)

Ctsum=zeros(2*taumax+1,1); % 相関の計算で使用する分だけの行列を作る

for tau=-taumax:taumax

    xx=data1(taumax+1+tau:taumax+1+np+tau);
    yy=data2(taumax+1:taumax+1+np);
```

```

        Ctsum(tau+taumax+1)=xx'*yy/np;
end

% 相互相関関数の表示

figure(2) % 相関関数波形 図 4.1.3

plot(-taumax:taumax, Ctsum(1:taumax*2+1));

xlabel('tau')
ylabel('相互相関関数 ')

[yjiku xjiku]=max(Ctsum) % 表示されたグラフの最大値を示す
                        % x軸の値を xjiku, y軸の値を yjiku とする

saidaiti=xjiku-(taumax+1) % グラフではx軸の中心0が、左端から taumax+1
                        % ずれた部分にあるため

% 角度の計算

hoko=asin(c*saidaiti/Fs/d); % 時間差から角度を求める

kaku=hoko/pi*180 % 音の方向の単位を[ラジアン]から[°]に変換

disp(['角度=' num2str(kaku)]) % 角度の数値を文字に変えて表示

```

4-1-4. インパルス応答の方向推定プログラムの実行結果

プログラムを実行すると、次の2つの図が表示される。

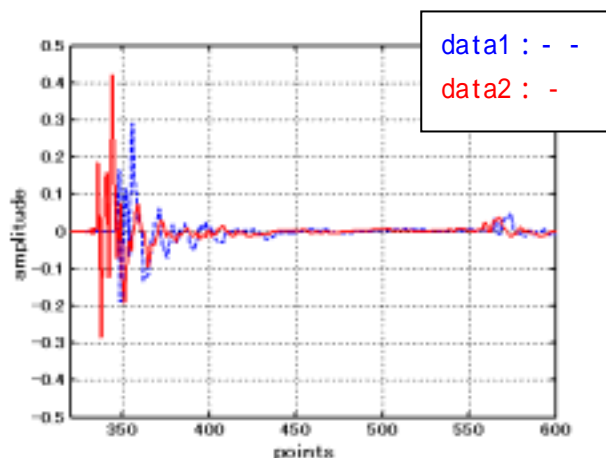


図 4.1.2: 読み込んだ2つの
インパルス応答の波形

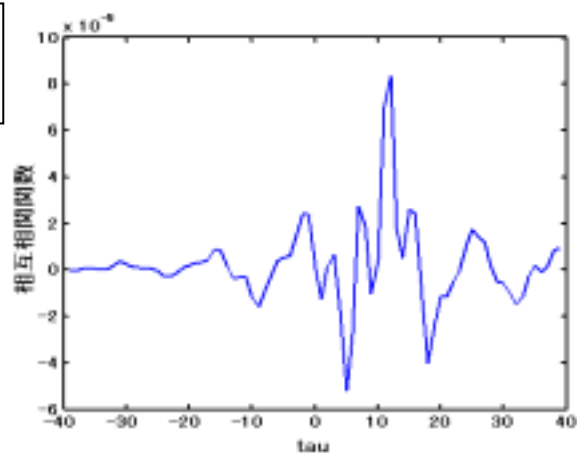


図 4.1.3: 相互相関関数波形

図 4.1.2 はプログラムにより読み込んだ、無響室の2つのインパルス応答波形である。両波形にはマイクロホンアレー10素子分の時間差があるため、グラフにずれが生じている。また振幅に若干の差が出ているのが分かる。

図 4.13 は相関処理を行って、その結果を出力した波形である。MATLAB のコマンドウィンドウより、saidaiti=tau=12 のところで最大となっている。これより角度を求めると、 $17.4788[^\circ]$ となる。

読み込んだファイルは「----/imp070.18」なので、マイクロホンアレー正面を0度とした場合、 $-20[^\circ]$ になれば良い。計算結果は $-17.4788[^\circ]$ なので、約 2.5° の誤差があるが、方向推定はできていると言える。

残る無響室のデータも同様に方向推定し、その結果を表にまとめた。

(a) データベースのフ ァイル No.	(b) データベースの公 称角度 [°]	(c) プログラム実行結 果 [°]
imp010	-80	-58.3
imp030	-60	-48.7
imp050	-40	-33.4
imp070	-20	-17.5
imp090	0	0
imp110	+20	17.5
imp130	+40	35.1
imp150	+60	50.9
imp170	+80	58.3

表 . インパルス応答の方向推定結果

表より、角度が0°から双方に大きくなるにつれ、方向推定誤差が大きくなっているのが分かる。これについては、何らかの形で補正をするのが望ましいと思われる。

音声データについては、「3-10. 収録データの一覧表」より、収録データの方向数は1つしかない。全て0°のデータなので、インパルス応答と同様にデータを読み込んで方向検出をしても結果は0°となる。

4 - 2 . 移動音声の方向推定

時間差に基づく方法で移動音声の方向推定を行う。

4-2-1 . 移動音声の方向推定プログラムの手順

1). パラメータの設定

```
c=340          % 音速[m/s]
d=0.0283*10    % 素子間距離[m]
Fs=48000       % サンプリング周波数[Hz]
```

2). データの読み込み

```
micarray/MICARRAY/data4/cirline/Ane/mmysda03/mmysda03.x  data1
micarray/MICARRAY/data4/cirline/Ane/mmysda03/mmysda03.y  data2
micarray/MICARRAY/data4/cirline/Ane/mmysda03/mmysda03.28 data3
micarray/MICARRAY/data4/cirline/Ane/mmysda03/mmysda03.18 data4
```

上記の4つのファイルをdata1～data4として読み込む。上から順に、OPTOTRAKのx座標データ、OPTOTRAKのy座標データ、移動音声（無響室）の28素子のデータ、移動音声（無響室）の18素子のデータである。

z座標の音源位置データは今回方向推定に使用しなかった。移動音源装置は基本的にz座標上を平行移動し、始点から終点まで10cm程度しか変化しないので、方向推定に与える影響も少ないと判断した為である。参考までに、図4.2.1にx座標とz座標の位置データ波形を示す。

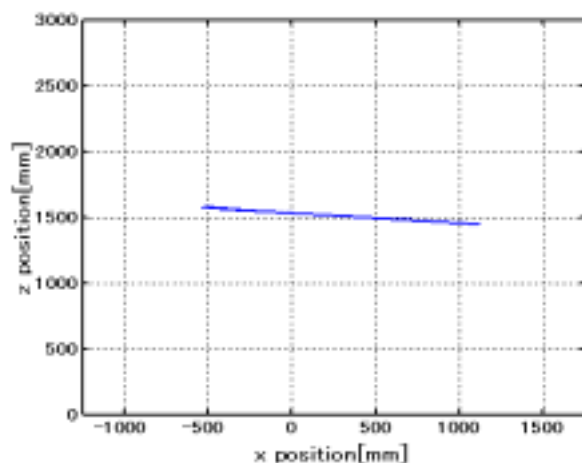


図 4.2.1 : OPTOTRAK による位置データ (x,z)

3).OPTOTRAK による位置データと、音声データの波形を出力

figure(1)に OPTOTRAK による位置データを表示する。figure(2)には読み込んだ 2 つの音声波形を色違いで重ねて表示する。

4).OPTOTRAK による位置データから、角度を算出

マイクロホンアレーの正面を 0° としているので、角度を求める式は
 $\text{atan}(x \text{ 座標}/y \text{ 座標})$ (図 4.2.2 参照)
という形になる。

5).相関関数の計算

figure(3)に、読み込んだデータの np (相関の計算長) 分の波形が連続で表示される。読み込んだデータを np ずつ区切って、それを np/2 ずつずらしながら相関処理をしていく。np が 4800 の場合は、1 ~ 4800, 2401 ~ 7200, 4801 ~ 9600, . . . というふうにずらしていく。

6).相関関数を表示する

figure(4)に相互相関関数の波形を表示する。これも連続表示されるので、プログラム終了後は読み込みデータの一番最後の部分の相関波形になる。

7).時間差により角度を求める

相関関数のグラフより得た時間差の値から角度を求める。次に音声データの短時間パワーを求め、その小さい部分は方向推定できないので、プログラムより求めた角度を 200 に置き換える。こうすることで、角度のグラフには誤ったデータがプロットされることはなくなる。

8).OPTOTRAK から求めた角度と時間差から求めた角度を 1 つの図に表示

figure(5)に OPTOTRAK から求めた角度を、figure(6)に時間差から求めた角度をそれぞれ表示する。

以下に使用したプログラムを示す。

4-2-2. 移動音声の方向推定プログラム

```

%idouopta

% マイクロホンアレーの移動音声データの方向検出プログラム
% 2002.2.8 YAMAGUCHI

% マイクロホンアレーデータベースから、移動音声の2個のファイルと、OPTOTRAKに
% よる音源位置データファイルを読み込む。その後相関処理を行い、個々のデータの波形と
% その相関波形を表示し、角度を求めて、OPTOTRAKによる音源位置データから算出した角度と、
% プログラムから算出した角度を表示する

clf          % 前の図を消す

% 初期設定

c=340        % 音速[m/s]
d=0.0283*10  % 素子間距離[m]
Fs=48000     % サンプリング周波数[Hz]

% データの読み込み

rn='Ane';    % 部屋の指定（無響室）
fn='03';     % 音声ファイル指定（音声データ(人と文章)の種類）

fname1=['d:%array%idouonseï%' rn '%mmysda' fn '%mmysda' fn '.x']; % OPTOTRAKのx座標
                                           % データの読み込み
fid=fopen(fname1,'r'); % 指定されたファイルを開く。'r'は読み込み
data1=fread(fid,'float'); % バイナリデータ（文字形式以外のデータ）を読み込む

fname2=['d:%array%idouonseï%' rn '%mmysda' fn '%mmysda' fn '.y']; % OPTOTRAKのy座標
                                           % データの読み込み
fid=fopen(fname2,'r');
data2=fread(fid,'float');

```

```

fname3=['d:¥array¥idouonsei¥' rn '¥mmysda' fn '¥mmysda' fn '.28'];
                                     % 移動音声の28素子目データの読み込み
fid=fopen(fname3,'r');
data3=fread(fid,'short');

fname4=['d:¥array¥idouonsei¥' rn '¥mmysda' fn '¥mmysda' fn '.18'];
                                     % 移動音声の28素子目データの読み込み
fid=fopen(fname4,'r');
data4=fread(fid,'short');

fclose(fid); % ファイルを閉じる

% 波形の表示

figure(1) % OPTOTRAKの位置データ波形 図4.2.2

plot(data1,data2) % data1をx軸、data2をy軸に表示

axis([-1250 1750 0 3000]);

xlabel('x position[mm]')
ylabel('y position[mm]')

grid % plot図に格子を引く

figure(2) % 音声データの入力波形 図4.2.3

plot(data3)

hold on % 前の図に次の図を重ね書きする

```

```

plot(data4, 'r') % 赤色で表示

xlabel('points')
ylabel('amplitude')
axis([10000 170000 -1.5e4 1.5e4]);
grid

% OPTOTRACK の角度の計算

kakudo1=zeros(1000,2); % 1000 行 2 列の 0 行列を作る

m=510; % 移動音源装置の根元からマイクロホンアレーの中心までの距離

i=1;

for ip2=1:5:length(data1)-5+1 % 1 から 5 間隔で data1 のデータ長まで繰り返す

    ip2 % ip2 のの値を表示

    data11=data1(ip2);
    data22=data2(ip2);

    data11
    data22

    hoko1=atan(data11/(data22-m)); % OPTOTRACK による音源位置データから角度を算出
    % する。
    % 無響室の場合、510[mm]あるので、y 軸のデータ
    % 値から差し引く

    kakudo=hoko1/pi*180 % 音の方向の単位を[ラジアン]から[°]に変換

    kakudo1(i,1)=0.05*i; % OPTOTRACK は 0.01 秒毎にデータを取っている。
    % ip2 を 5 個ずつずらしているの、0.05 秒毎に
    % 角度を出す

```

```

kakudo1(i,2)=kakudo;          % 時間と角度を行列に格納

i=i+1;

disp(['角度=' num2str(kakudo)]) % 変数「kakudo」を文字列に変換

end

% 相関関数の計算

np=4800;    % 相関の計算長 N

taumax=fix(d/c*Fs); % fix は()内を0方向へ丸める (ex. 24.5 24)

Ctsum=zeros(2*taumax+1,1); % 相関の計算で使用する分だけの行列を作る

kakudo2=zeros(1000,2);

ikaku=1;

PP=zeros(1000,1);

xx2=zeros(1000,1);

for ip1=1:np/2:length(data3)-np+1-taumax*2-1

    ip1

    data33=data3(ip1:ip1+np-1+taumax*2+1);
    data44=data4(ip1:ip1+np-1+taumax*2+1);

```

figure(3) % 音声データを np ずつ区切った波形 図 4.2.4

```

plot(data33)

axis([0 np -10000 15000]);

xlabel('points')
ylabel('amplitude')

drawnow % 強制的にスクリーンを更新する。この場合、for 文によって
        % 変化する figure3 を連続で表示する形になる

for tau=-taumax:taumax % 相関計算

    xx=data33(taumax+1+tau:taumax+1+np+tau);
    yy=data44(taumax+1:taumax+1+np);

    Ctsun(tau+taumax+1)=xx'*yy/np;
end

kakudo2(ikaku,1)=np/2/Fs*ikaku;

```

% 相関関数の表示

figure(4) % 相関関数波形 図 4.2.5

```

plot(-taumax:taumax, Ctsun(1:taumax*2+1));

xlabel('tau')
ylabel('相互相関関数 ')

[yjiku xjiku]=max(Ctsun)

```



```

saidaiti=xjiku-(taumax+1)

% 角度の計算

hoko=asin(c*saidaiti/Fs/d); % 時間差から角度を求める

kaku=hoko/pi*180

PP(ikaku)=(data33'*data33)/length(data33); % data33の短時間パワーの計算

    PP(ikaku)

if PP(ikaku) < 0.3e6 % 読み込んだ音声の短時間パワーが低い部分は、角度を
    kaku= 200; % 表示しないようにする為に200を代入
end

kakudo2(ikaku,2)=kaku;

ikaku=ikaku+1;

disp(['角度=' num2str(kaku)]) % 角度の数値を文字に変えて表示

end

figure(5) % 音声データの短時間パワー波形 図4.2.6

plot(kakudo1(1:i-1,1),PP(1:ikaku+1))
xlabel('時間[s]')
ylabel('Power ')

```

```
figure(6) % OPTOTRAK から求めた角度と時間差から求めた角度 図 4.2.7
```

```
clf
```

```
plot(kakudo2(1:i-1,1),kakudo2(1:i-1,2),'xb'); % kakudo2 のデータが入っている  
% 部分のみ表示
```

```
hold on
```

```
plot(kakudo1(1:i-1,1),kakudo1(1:i-1,2),'r');
```

```
axis([0 4.5 -50 50]);
```

```
xlabel('時間[s]')
```

```
ylabel('角度 [°]')
```

```
grid
```

4-2-3. 移動音声の方向推定プログラムの実行結果

このプログラムを実行すると、次の6つのグラフが表示される。

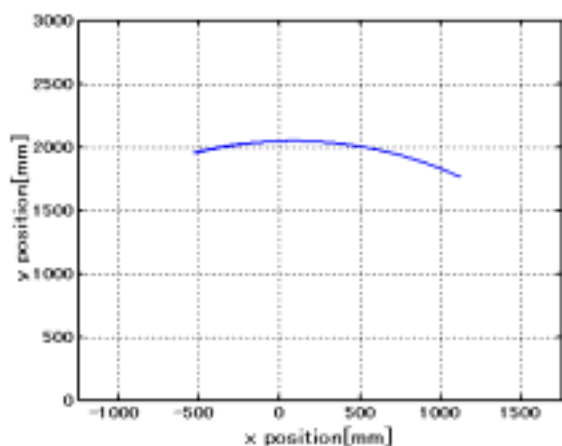


図 4.2.2: OPTOTRAK による位置データ(x,y)

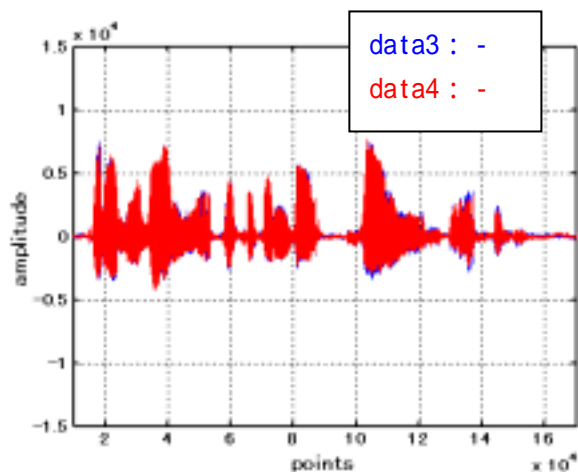


図 4.2.3: 読み込んだ2つの音声データ波形

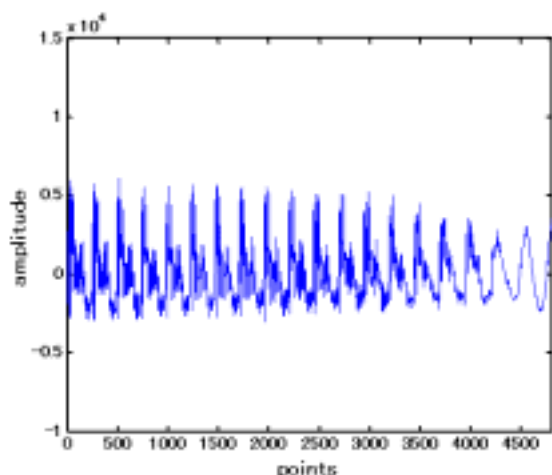


図 4.2.4: 相関計算長で区切った音声の波形

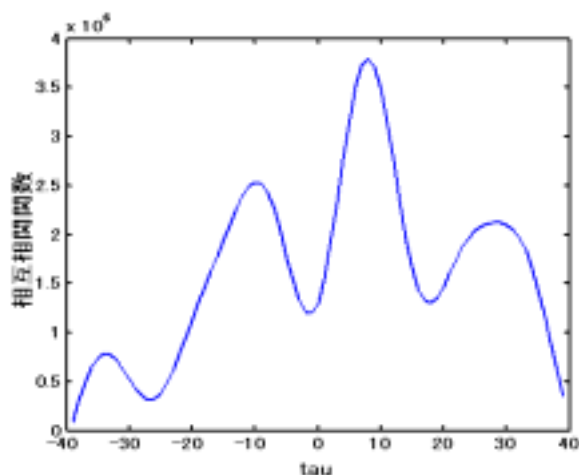


図 4.2.5: 相互相関関数波形

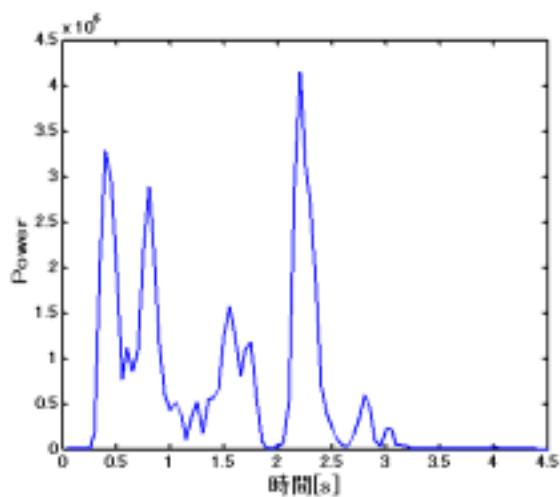


図 4.2.6: 音声の短時間パワー波形

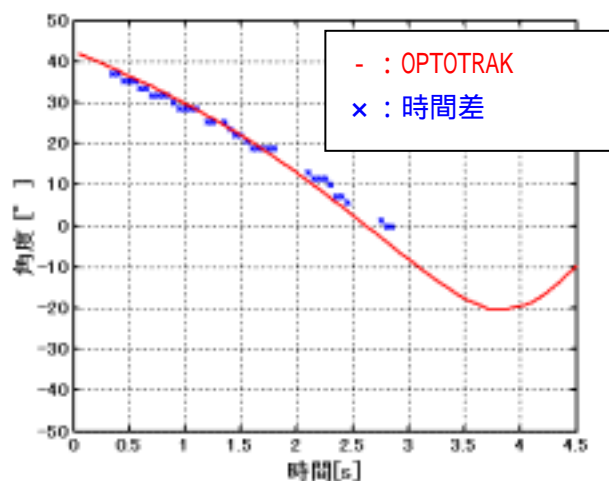


図 4.2.7: OPTOTRAK による音源角度と時間差による推定角度

図 4.2.2 は OPTOTRAK による音源位置データの x 座標と y 座標を表示したものである。図から見て分かるように、移動音源の x 座標の軌跡は、中心が 0 になっていない(図の中心は 300point 付近)。手動で装置を動かしている為、収録する毎に多少のズレが生じるようである。

図 4.2.3 は読み込んだ 2 つの音声データの波形である。赤と青で色を分けて表示してある。

図 4.2.4 は相関の計算長で読み込んだ音声データを区切ったものである。この場合は相関長が 4800point , ip1=103201point なので、音声データの 103201 ~ 108000point の部分の波形となる。

図 4.2.5 は図 4.2.4 の部分の相互相関関数波形である。図から見てとれるように、 $\tau = 8$ の部分でグラフは最大となり、この時間差から算出された角度は $11.5507[^\circ]$ である。

図 4.2.6 は読み込んだ音声データの短時間パワー波形である。出力の極端に低い部分は音がよく聴き取れないので、その部分で方向推定を行っても、無音部分の方向推定をすることと同じなので無意味である。ここでは 0.3×10^6 乗未満の部分、算出された角度の値の代わりに 200 を代入して、グラフ上にプロットできないようにしている。

図 4.2.7 は OPTOTRAK による位置データから算出した角度(真の音源角度)と、相関処理を行って時間差により求めた角度を同時に表示したものである。OPTOTRAK の位置データは正確なので、時間差より求めた角度がこの波形と一致すれば、このプログラムは正常であると言える。この場合は多少のズレはあるものの、良好に推定が行われている。

次に相関の計算長を変えて同じように方向推定を行った。角度のグラフのみ表示する。

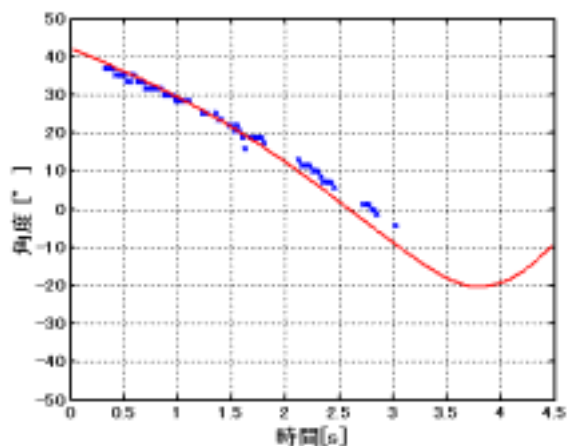


図4.2.8:角度のグラフ(相関長:2400)

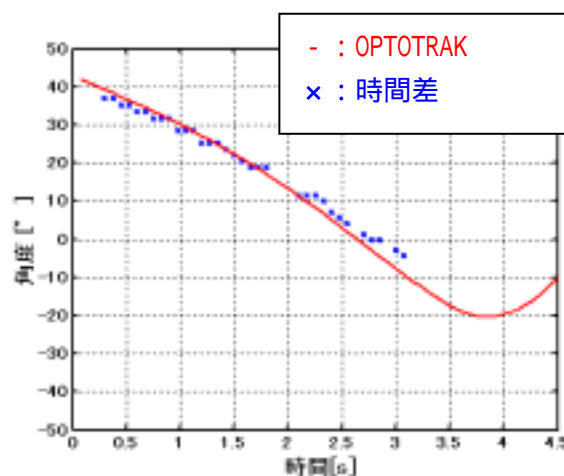


図4.2.9:角度のグラフ(相関長:7200)

上の両図を比較すると、図4.2.8の方の1.7秒あたりに1つだけズレて出力されているもの以外はどれもほぼ同じである。残響のない無響室のデータでは正常に方向推定が行われているのが分かる。

次に残響可変室のデータで同じように方向推定を行った。先のプログラム上でデータを変更する部分は、

```
rn='Ane'; % 部屋の指定
```

```
m=510; % 移動音源装置の根元からマイクロホンアレーの中心までの距離
```

```
np=4800; % 相関の計算長
```

の3箇所である。部屋は残響室 A の場合には $rn='E2a'$ 、残響室 B の場合は $rn='E2b'$ に、また m は $420[\text{mm}]$ に、 np を $2400, 4800, 7200$ (サンプリング周波数 $F_s = 48\text{kHz}$ であるので、それぞれ $0.05, 0.10, 0.15$ 秒に相当する) に変えてプログラムを実行する。それぞれの部屋で、相関長を変えて方向推定し、角度のグラフのみを次にまとめた。グラフのタイトルの () 内は、(部屋, 相関計算長) を意味している。なお、残響室 A の残響時間は 0.3 秒、残響室 B の残響時間は 1.3 秒である。

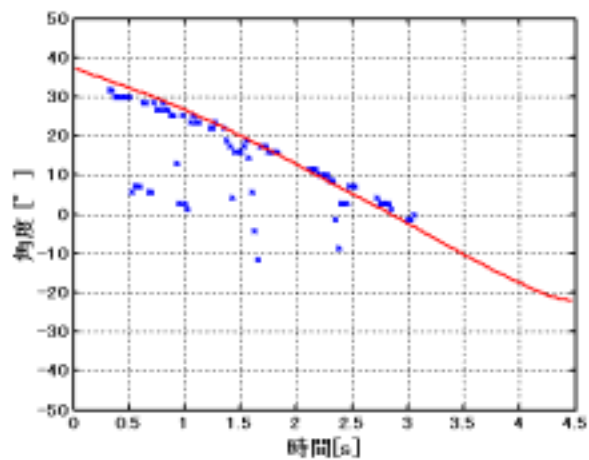


図 4.2.10 : 角度のグラフ (E2a , 2400)

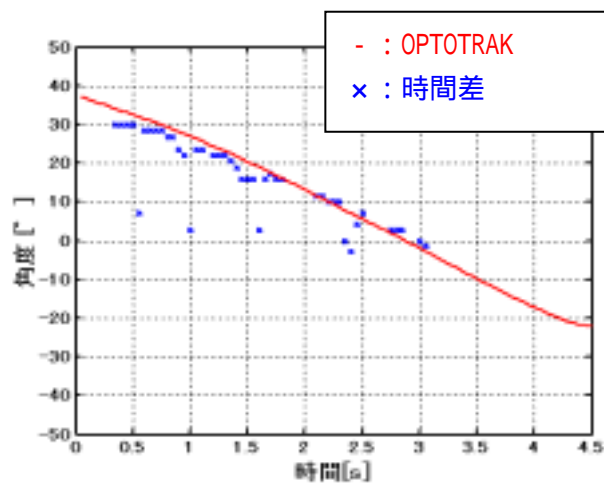


図 4.2.11 : 角度のグラフ (E2a , 4800)

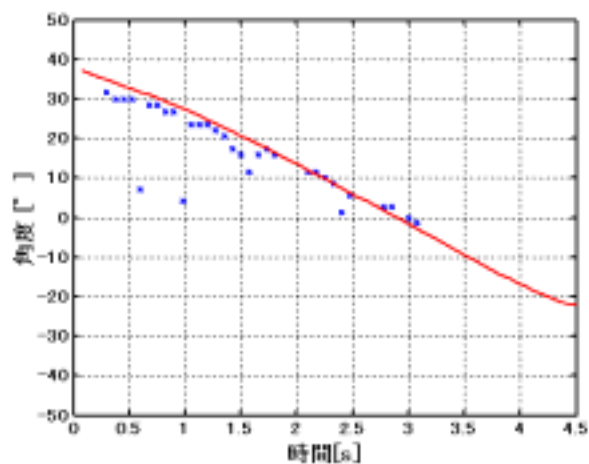


図 4.2.12 : 角度のグラフ (E2a , 7200)

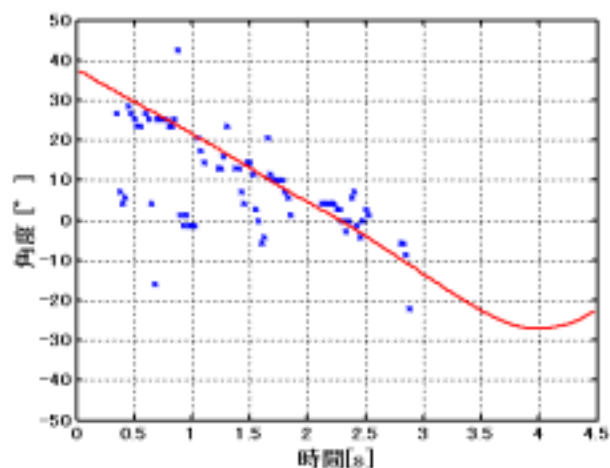


図 4.2.13 : 角度のグラフ (E2b , 2400)

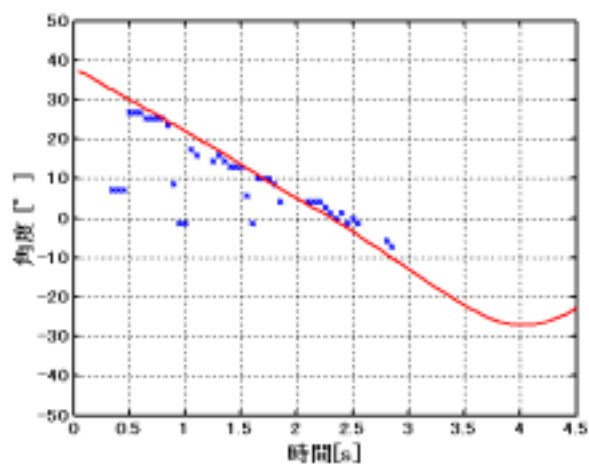


図 4.2.14 : 角度のグラフ (E2b , 4800)

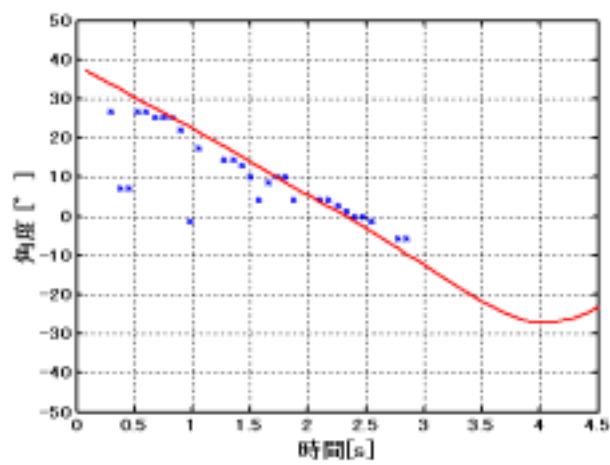


図 4.2.15 : 角度のグラフ (E2b , 7200)

まず残変室 A を見てみる。図 4.2.9 では、時間差により求めた角度が OPTOTRAK のラインと重なる部分以外に、飛び散ってプロットされている部分がある。図 4.2.11 を見ると、図 4.2.10 よりばらつきが少ない。図 4.2.12 ではそれほどばらつきが目立たなくなっている。

次に残変室 B の図 4.2.13 を見ると、残変室 A よりも更にグラフが飛び散っている。しかしこれも同様に、相関長を 4800 , 7200 と変えていくにつれ、飛び散る部分が減っている。

無響室では正しく方向推定できていたが、残変室 A では方向推定の結果にバラつきが出ている。残変室 B では更にバラつきが目立つようになる。このことから、残響が方向推定に影響を及ぼすことが分かる。

また、残変室でも相関の計算長が長くなるにつれ、方向推定の結果にばらつきがなくなっている。このことから、相関長が方向推定の精度に影響を与え、相関長を長くすると方向推定の精度が上がる事が分かる。

相関長を上げていったとき、相関長が 7200 になるとほぼ実際の角度と一致しているので、1.3 秒ほどの残響がある部屋でも、約 0.15 秒単位で方向推定することが可能であると言える。

まとめ

本研究では、まず時間差に基づく方向推定の理論を示し、音源方向を推定するプログラムをその理論より作成した。次にマイクロホンアレーデータベースから様々な環境でのデータを読み取り、プログラムとデータベースの評価をし、また残響が方向推定に与える影響を評価して解析した。

インパルス応答の解析では、マイクロホンアレーの正面 0° から $\pm 80^\circ$ まで角度が大きくなっていくにつれ、プログラムにおける方向推定の誤差が大きくなった。これは正面の音の方がより正確に方向推定ができているということである。人間の耳もそのような性質があることから、マイクロホンアレーを使用した方向推定においても、鈍角から聞こえてくる音の解析は困難であることが分かった。

移動音源の解析では、残響時間0秒の無響室では良好に推定が行われており、プログラムが正常に動作していることが分かった。その上で、残響可変室で同様に方向推定したところ、残響時間が0.3秒、1.3秒と長くなるにつれ、出力される角度のグラフがばらつき、方向推定誤差が大きくなった。このことより、残響が方向推定に影響を及ぼしていることが分かった。また、残響がある部屋でも、相関の計算長を2400,4800,7200と長くしていく毎に、グラフのばらつきも減っていき、誤差が少なくなった。これより、相関の計算長が方向推定に影響を与え、相関の計算長を長くとると方向推定の精度があがることが分かった。更に、残響可変室Bで相関長を上げていったとき、相関長が7200になると実際の角度とほぼ一致していることから、1.3秒ほどの残響がある部屋でも、約0.15秒単位で方向推定することが可能であると言える。

マイクロホンアレーデータベースに関して、ホームページでは、「OPTOTRAKの位置データ、x,y,z座標は受音器の中心からの距離」となっているが、3-6の無響室および残響可変室の図にある、移動音源装置からマイクロホンアレーの中心までの距離をy座標の値から引いてやらないと、正常に方向推定ができないことが分かった。よって、これは「移動音源装置からの中心」が正しいと思われる。またホームページには、移動音源装置からマイクロホンアレーの中心までの距離は、無響室と残響可変室のデータしか書いていない。会議室や畳部屋で方向推定をする為には、このことを検討しなければならない。

今後は、残響可変室のデータで、相関長を長くしても推定誤差を生じている部分を分析する。また、OPTOTRAKの読み取りを、z座標も加えた3次元のものにすれば、OPTOTRAKにより求める角度はより正確なものになる。それらをふまえた上で残響に強い方向推定方法を開発すること、更にデータベースから様々なデータを読み取り、データベースの評価を続けることが今後の課題である。

参考文献

- [1]坂巻 佳美,「見てわかる デジタル信号処理」, 工業調査会
pp.160 ~ 164,173 ~ 175 (1998 年)
- [2]大賀、山崎、金田 :「音響システムとデジタル処理」電子情報通信学会,
pp.176 ~ 181,197 ~ 199 (1995 年)
- [3]RWC P 実環境音声・音響データベース
<http://www.slt.atr.co.jp/~tnishi/DB/micarray/index.htm>
- [4]OPTOTRAK3020 3次元運動計測システム
<http://www.asco.co.jp/asco/products/optotrak.html>

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、ならびに音響信号処理研究室の研究生の皆様に厚くお礼申し上げますとともに、深く感謝します。