

長岡技術科学大学大学院
工学研究科修士論文

題 目

可聴音による
物体位置の推定に関する研究

指導教員

准教授 杉田 泰則

著 者

電気電子情報工学専攻
信号処理応用研究室
21317184 東 響樹

令和 7 年 2 月 7 日

目次

第 1 章	はじめに	1
1.1	研究背景	1
1.1.1	位置情報	1
1.1.2	従来手法	2
1.2	研究目的	3
1.3	本論文の構成	3
第 2 章	提案手法	4
2.1	提案手法の概要	4
2.2	仮想的な受信信号	5
2.3	プリアンファシスフィルタ	7
2.4	標準化	7
2.5	等間隔フィルタバンク	8
2.6	ユークリッド距離	9
第 3 章	位置推定	10
3.1	インパルス応答測定	10
3.2	シミュレーション	12
3.2.1	条件	12
3.2.2	結果	14
3.3	実験	15
3.3.1	条件	15
3.3.2	結果	15
3.4	切り出すフレーム幅変更	17
3.4.1	条件	17
3.4.2	結果	17

第 4 章	おわりに	19
4.1	まとめ	19
4.2	今後の展望	20
謝辞		21
付録 A	各フレーム幅の位置推定精度	22
参考文献		25

第 1 章

はじめに

1.1 研究背景

1.1.1 位置情報

位置情報は、人の動線情報が得られることから、地図アプリなどのルートナビゲーションや現在地や移動履歴による広告や近隣店舗のセール情報の提供に活用されている。また、ヘルスケアモニタリングにおける高齢者や患者の検知や在室状態の検出にも活用されている。屋外における位置情報の取得は一般的に人工衛星からの電波をカーナビやスマホなどの受信機で受信し位置を特定する GPS(Global Positioning System) 測位や携帯電話が接続している複数の基地局情報を基に位置を特定する携帯基地局測位が普及している。しかし、屋内においては GPS 信号や基地局からの信号が届かないことから、GPS 測位や携帯基地局測位による位置情報取得は困難である。

そこで、近年、屋内での位置情報取得の手法に関する研究が行われており、ユーザがデバイスを携帯する測位法として、Wi-Fi [1] や BLE(Bluetooth Low Energy) [2] 等の電波強度に基づく手法や慣性センサを用いた手法 [3] が提案されている。しかし、全てのユーザがデバイスを携帯しているとは限らないことから、屋内における位置情報取得の手法はユーザがデバイスを携帯しないデバイスフリー測位法 (Device-free Localization : DFL) が求められている。デバイスフリーな屋内測位手法として、画像や赤外線、超音波などを用いた手法が提案されている [4] [5] [6]。画像を用いた測位では時系列順の画像から差分をとることで推定できるが、暗い場所での不向きやプライバシー侵害といった問題がある。赤外線や超音波は、物によってセンサの反応が異なる、広い空間での計測に適していないなどの課題がある [7] [8]。

1.1.2 従来手法

上記の課題を解決するために、[8] では可聴音を用いて物体 (人体) 検知する手法を提案している。可聴音を用いる理由としては、プライバシーの問題や広範囲の計測に向いていないことを解決できるうえに、スピーカーやマイクなどの一般的なデバイスだけで事足りるためである。提案している手法は、可聴音の透過法計測を用いて、物体なしと物体ありの受信信号の最大値の比と到達時間変動量の値から物体の検出を行うというものである。透過法計測とは、可聴音における光や音波などの波動を物体や媒体に透過させ、その波動の挙動や変化を利用して測定を行う方法であり、物質の成分分析や品質検査などで用いられている。可聴音を用いた物体検出の従来手法の流れは以下に示す。

1. 予め、物体がない場合のスピーカから発せられた可聴音信号 (チャープ信号) の透過波をマイクにより受信 (図 1.1 の左図)
2. 物体がある時も同様に可聴音信号の透過波をマイクにより受信 (図 1.1 の右図)
3. 物体なしの信号と物体ありの受信信号の最大値の比と到達時間変動量の値から物体検出

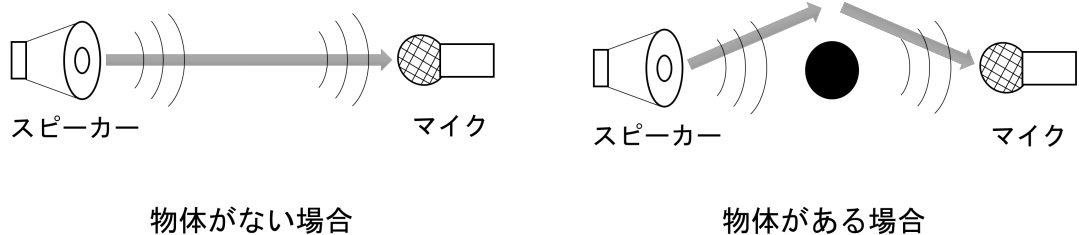


図 1.1 透過法計測の例

[8] では、図 1.2 の条件で物体を様々な位置に配置し、上記の手法で物体検出が可能であるかどうかシミュレーションを行った。その結果、物体なしの信号と物体ありの信号の最大値の比と到達時間差の値が物体の位置によって異なり、物体検出が可能であることが示唆されていた。しかし、実機実験での未検証や送信信号がチャープ信号という特殊な信号 (固定信号) であることが課題に挙げられる。

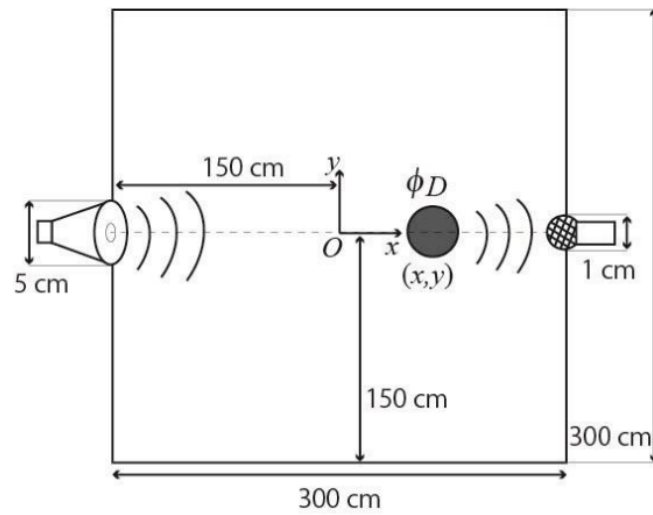


図 1.2 シミュレーション条件

1.2 研究目的

そこで、本論文では、インパルス応答および可聴音 (音声や音楽など) による物体位置の推定を目的とする。

1.3 本論文の構成

第1章では研究背景および目的を述べた。第2章では提案手法について述べる。第3章では実験について述べる。最後に第4章では本論文のまとめ及び今後の展望について述べる。

第 2 章

提案手法

2.1 提案手法の概要

とある場所に物体を置いた時の受信信号は送信信号とインパルス応答を畳み込んだ信号で仮想的に再現できる。この特性を用いて、物体をある位置に置いた時の受信信号と学習データのインパルス応答から生成した各位置の仮想的な受信信号を比較し、物体の位置推定する手法を提案する。提案手法のフロチャートを図 2.1 に示す。提案手法の具体的なプロセスの説明は以下に示す。

1. 物体がどこかにある時、スピーカから送信された信号をマイクで受信する (推定対象)。
2. 事前測定した各位置に物体がある時のインパルス応答 (学習データ) に 1. でのスピーカからの送信信号を畳み込み、各位置に物体がある時の仮想的な受信信号を生成する。
3. 推定対象と仮想的な受信信号をプリアンファシスフィルタ処理する。
4. プリアンファシスフィルタ処理後は、標準化する。
5. 標準化した後、フーリエ変換で周波数領域に変換し振幅スペクトルを求める。
6. 振幅スペクトルに等間隔フィルタバンク処理を行う。
7. 推定対象と仮想的な受信信号の振幅スペクトルのユークリッド距離により位置推定する。

次小節から図 2.1 で使用されている処理および各用語について説明を行う。

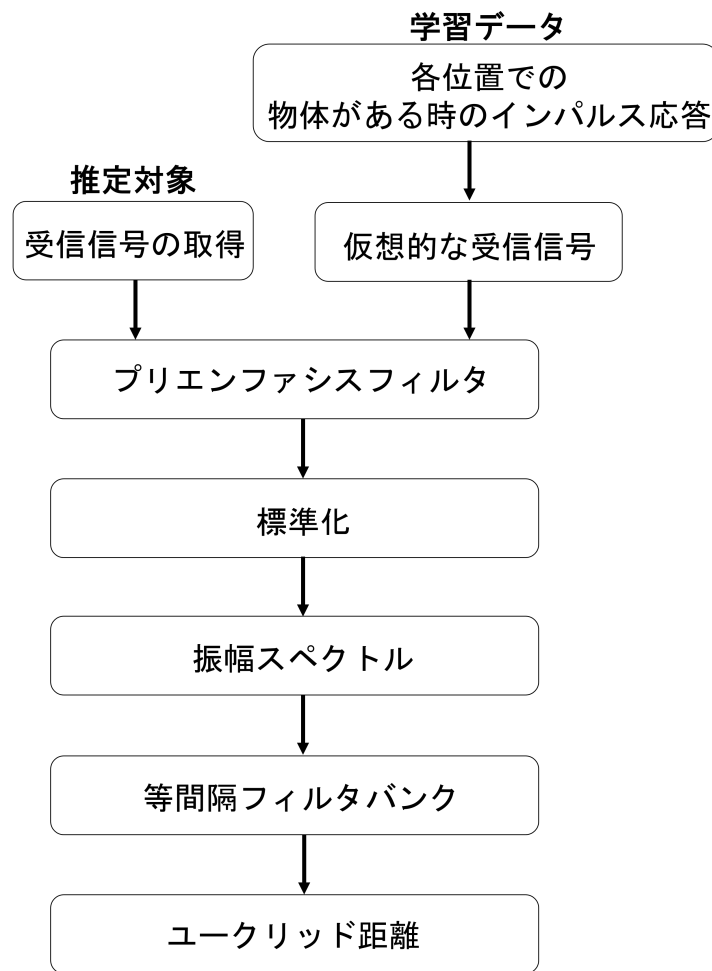


図 2.1 提案手法のフローチャート

2.2 仮想的な受信信号

推定対象と比較するための仮想的な受信信号は、事前に測定した学習データのインパルス応答から生成する。

インパルス応答とは、あるシステムにインパルス信号 $\delta(n)$ (非常に短いパルス状の信号) を入力した時の出力信号 $h(n)$ のことである (図 2.2)。ここでのシステムは、部屋やコンサートホール、電気回路などであり、信号処理や制御でよく扱われる。提案手法において、スピーカからマイクまでの空間のことである。また、図 2.2 でのインパルス信号 $\delta(n)$ は式 (2.1) に示した性質である。

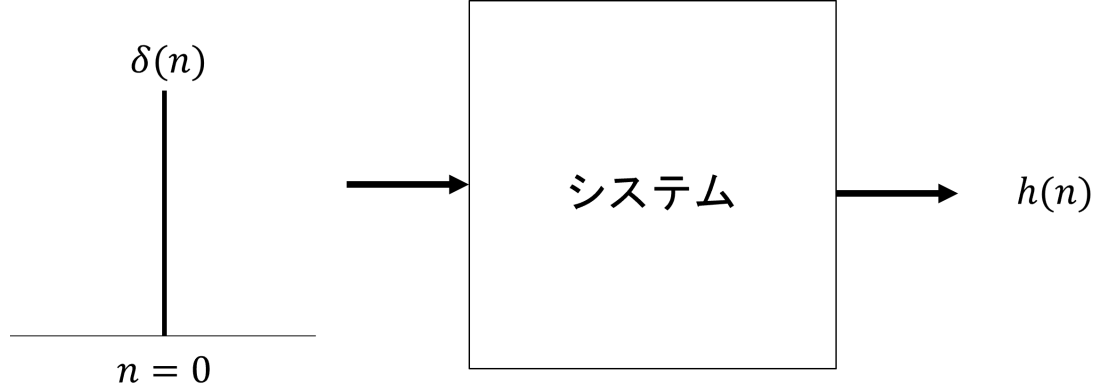


図 2.2 インパルス応答

$$\delta(n) \begin{cases} = \infty (n = 0) \\ = 0 (n \neq 0) \end{cases} \quad (2.1)$$

図 2.2 でのシステムを T とした時、入出力関係は

$$h(n) = T[\delta(n)] \quad (2.2)$$

と表される。この時、 $h(n)$ がシステム T でのインパルス応答である。また、線形時不変システムでは、任意の入力 $z(n)$ とそれに対応する出力 $y(n)$ の関係を

$$y(n) = \sum_{k=-\infty}^{\infty} z(k)h(n-k) \quad (2.3)$$

と記述することができる。ここでは、任意の入力 $z(n)$ に対する出力 $y(n)$ がインパルス応答 $h(n)$ のみで計算できることを意味している。式 (2.3) の関係を $z(n)$ と $h(n)$ の畳み込みと呼び、

$$y(n) = z(n) * h(n) \quad (2.4)$$

と略記される [9] [10] [11]。

このことから、提案手法では、各位置に物体がある時のインパルス応答 $h_i(n)$ ($i = 1, 2, \dots, N$) と任意の可聴音 $x(n)$ から、仮想的な受信信号 $y_{train,i}(n)$ を生成する (式 (2.5))。 N は学習データの数を示す。

$$y_{train,i}(n) = x(n) * h_i(n) \quad (2.5)$$

2.3 プリエンファシスフィルタ

音声信号などは一般に高周波成分が弱く、低周波成分が強い傾向があるため、推定対象の受信信号と各位置に物体がある時の仮想的な受信信号にプリエンファシスフィルタ処理を行う。プリエンファシスフィルタは、信号の高周波成分を強調し低周波成分を相対的に抑制するためのフィルタである。音声信号処理や音声認識の分野で広く使用され、一般的に人間の聴覚特性に基づいた特徴抽出手法の一つである MFCC(Mel Frequency Cepstrum Coefficient) の取得前に用いられる。プリエンファシスフィルタの式は

$$y_{pre}(n) = s(n) - \alpha \cdot s(n-1) \quad (2.6)$$

と表される。 $y_{pre}(n)$ はプリエンファシスフィルタ後の信号、 $s(n)$ はフィルタ前の元信号、 α はプリエンファシスフィルタの強度係数である。また、一般的に α は 0.9 から 1.0 の範囲の値が使われる (今回は 0.97 に設定) [12] [13] [14]。例として、ある音声にプリエンファシスフィルタ処理を行わない時と行った時の振幅スペクトルを図 2.3、図 2.4 に示す。2つの図より、プリエンファシスフィルタ処理をすることによって高周波成分が強調されていることが分かる。

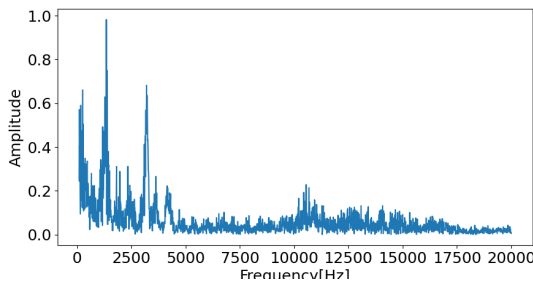


図 2.3 プリエンファシスフィルタなし

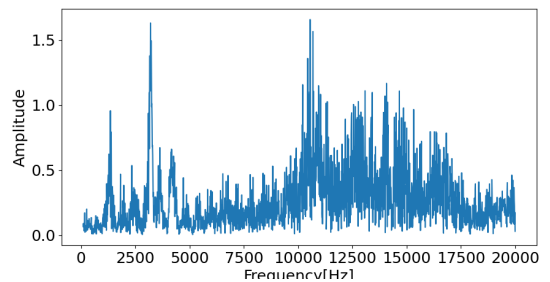


図 2.4 プリエンファシスフィルタあり

2.4 標準化

振幅スペクトルに変換する前に、プリエンファシスフィルタ後の推定対象と仮想的な信号の比較ができるよう、それぞれを平均 0、分散 1 となるように標準化する。

2.5 等間隔フィルタバンク

信号をフーリエ変換した振幅スペクトルと等間隔フィルタバンクの内積をとり、データ圧縮、ノイズ除去および特徴抽出を行う。等間隔フィルタバンクの例を図 2.5 に示す。等間隔フィルタバンクは、周波数軸上で等間隔に配置された複数のバンドパスフィルタの集合体である。この等間隔フィルタバンクと振幅スペクトルの内積をとることで、各バンドパスフィルタにおける周波数成分を抽出する。バンドパスフィルタの数 (三角形の頂点の数) はチャンネル数と呼ばれ、図 2.5 は 30 チャンネルの等間隔フィルタバンクであるといえる。また、一般的に下限周波数が 0、上限周波数がサンプリング周波数の半分になり、帯域間が 50% の重なりを持つ [12]。

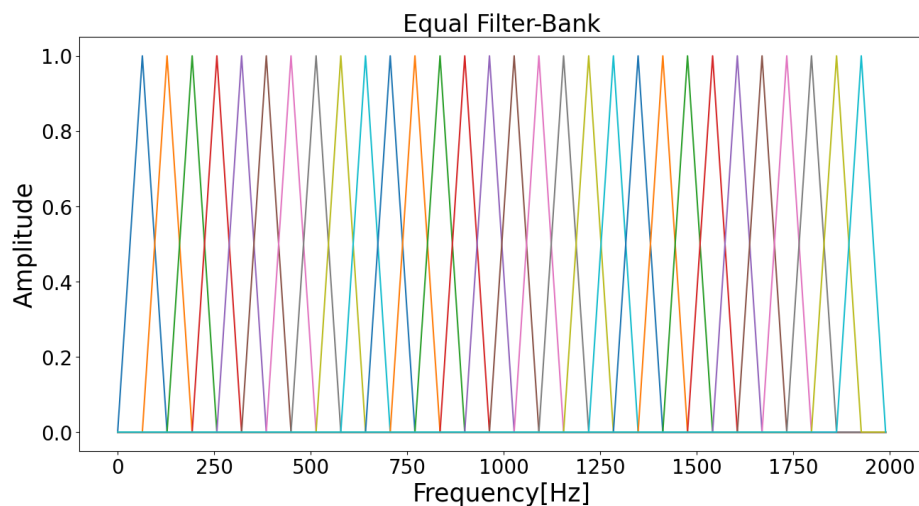


図 2.5 30 チャンネルの等間隔フィルタバンク

また、実際、図 2.4 の周波数成分に図 2.5 の等間隔フィルタバンク処理を行うと図 2.6 のような振幅スペクトルになる。図 2.5 のチャンネル数分の 30 個のデータが出力されているのが分かる。

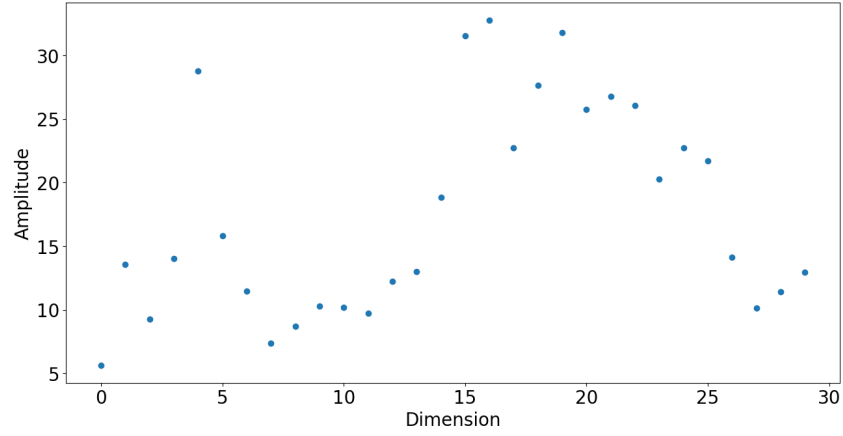


図 2.6 等間隔フィルタバンク処理した結果

2.6 ユークリッド距離

最後に、フィルタバンク処理後である推定対象と仮想的な受信信号の振幅スペクトルのユークリッド距離により位置推定する。ユークリッド距離は一般的に教師なしのクラスタリングにおけるデータ間の距離尺度として用いられるものである [15]。推定対象の振幅スペクトル Y_{test} と仮想的な受信信号の振幅スペクトル $Y_{train,i}$ のユークリッド距離 d_i は

$$d_i = \sqrt{\sum_{j=1}^{num} (Y_{test}(j) - Y_{train,i}(j))^2} \quad (2.7)$$

と求める。ここでの、 num は振幅スペクトルのデータ数 (次元数) である。そして、 d_i の中で最小値であった i に対応する物体位置を推定結果とする。

第 3 章

位置推定

今回の位置推定実験は、スピーカからマイクまでのインパルス応答からスピーカとマイクの直線上 (x 軸上 9 点) にある物体の位置を推定することを目的として実験を行う。また、今回は 1 秒ごとの信号という短い信号データから位置推定可能であるか検証する。

3.1 インパルス応答測定

はじめに、各位置に物体を置いた時のインパルス応答を測定する。今回は TSP 信号を用いてインパルス応答を測定した。TSP 信号とは時間が経過するにつれ、低い周波数から高い周波数へ変化する信号のことである。測定場所は無響室で、物体 (ダミーヘッド) を図 3.1 に示す x 軸上 9 点の位置に物体を置いた時のインパルス応答 $h_i(n)$ ($N = 9$) を測定する。以下の手順でインパルス応答を測定した。

1. 物体を置き、スピーカから TSP 信号を再生し、マイクで音響応答を録音
2. TSP 信号の逆特性の逆 TSP 信号を生成
3. 録音した信号に逆 TSP 信号を畳み込むことでインパルス応答を取得

また、今回は S/N を高めるために、1. では TSP 信号を複数回繰り返し再生し、加算平均したものを録音した信号とした。各位置に対してインパルス応答を 110 回測定する。ただし、100 回分のデータは学習データ、残り 10 回分のデータはテストデータとして扱う。スピーカおよびマイクの高さは約 100cm、ダミーヘッドの高さは 120cm、ダミーヘッドの大きさは縦 30cm、横 15cm、奥行 20cm である。図 3.2 に実際の測定環境を示す。また、測定に用いた機器は表 3.1 に示す。

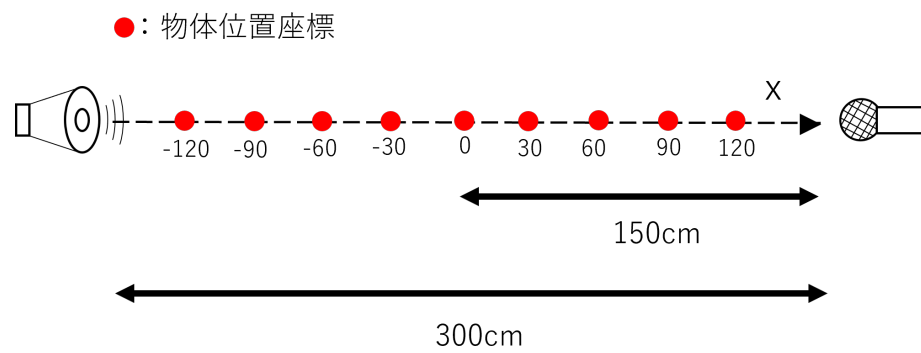


図 3.1 インパルス応答測定環境

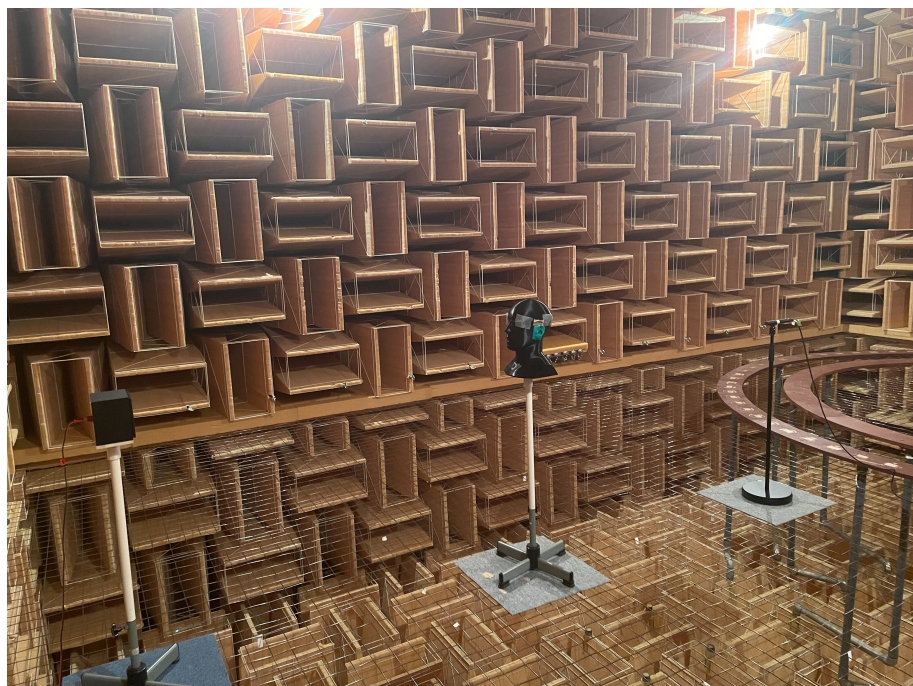


図 3.2 実際のインパルス応答測定環境

表 3.1 測定機器

Device Name	Model	Maker
Condenser Microphone	AT4022	Audio-Technica
Audio Interface	OCTA-CAPTURE	Roland
Power Amplifier	XM4080	YAMAHA

3.2 シミュレーション

まず、推定対象は実測受信信号ではなく、各位置のインパルス応答のテストデータ $h_{i-test}(n)$ (10 回分) に可聴音 $x(n)$ を畳み込みした仮定の推定対象を用いて、位置推定可能であるか検証する。

3.2.1 条件

仮定の推定対象 $y'_{test}(n)$ と各位置の仮想的な受信信号 $y_{train,i}(n)$ は

$$y'_{test}(n) = x(n) * h_{i-test}(n) = \sum_{k=0}^{L-1} x(k) h_{i-test}(n-k) \quad (3.1)$$

$$y_{train,i}(n) = x(n) * h_i(n) = \sum_{k=0}^{L-1} x(k) h_i(n-k) \quad (3.2)$$

で生成する。上式での L はインパルス応答の長さで、本論文では 2048 サンプルとした。 $x(n)$ は音声 5 種類、音楽 5 種類の全 10 種類で全て 10 秒の信号である [16] [17] [18]。音声は 5 種類の内、女性の声が 2 種類、男性の声が 3 種類である。位置推定行う際は、全時間データを使うのではなく、整数秒 (1 秒、2 秒...) から 0.1 秒分の信号を切り出したものを入力データとして順に位置推定する。入力のデータのイメージ図を図 3.3 に示す。0 秒からの信号を使わない理由は、無音区間が存在するためである。等間隔フィルタバンクのチャンネル数は 200 (図 3.4)、学習データのインパルス応答 h_i は 100 回分を平均したものを用いる。また、同じ位置の物体を置いた時のインパルス応答の周波数成分が測定ごと変動している部分 (100Hz 以下、20kHz 以上) があったため、本論文での位置推定では安定した結果を得る目的で振幅スペクトルを算出する際、100Hz ~ 20kHz に帯域制限する。

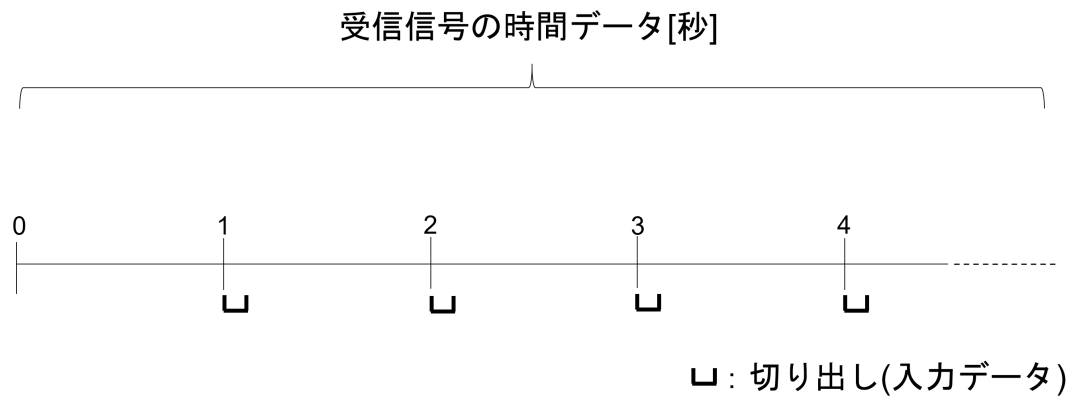


図 3.3 入力データのイメージ

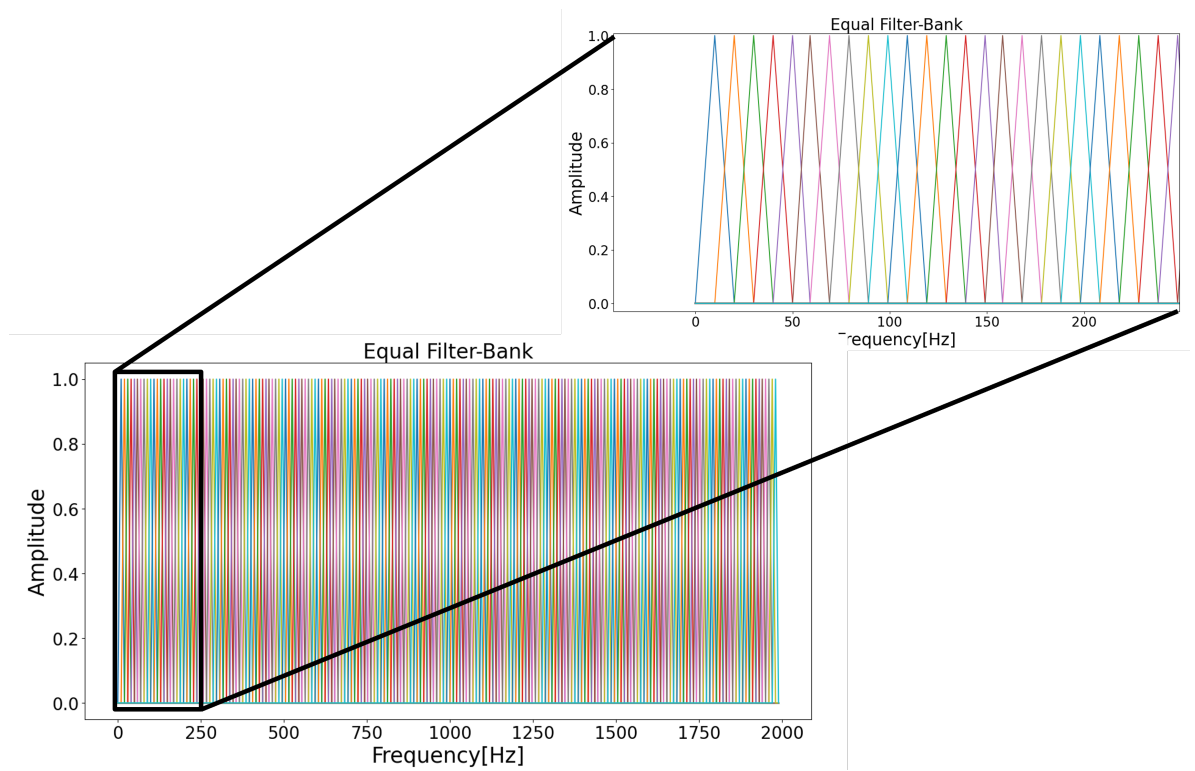


図 3.4 200 チャンネルの等間隔フィルタバンク

3.2.2 結果

シミュレーションでの位置推定精度の混同行列を図 3.5 に示す。各位置の入力データ数は 1000 個である ($10 \times 10 \times 10$ (信号の長さ \times 可聴音の種類 \times テストデータのインパルス応答数))。図 3.5 より、位置-120、-30、0、60、90、120 の信号は 100% もしくは 100% に近い精度で位置推定できていることが分かる。また、それ以外の位置-90、-60、30 の信号に関しても、他の位置の信号と比べ、少し精度が下がるが位置推定できていることが分かる。このことから、インパルス応答および可聴音を用いて物体の位置推定が可能であることが分かった。

		推定結果								
		-120	-90	-60	-30	0	30	60	90	120
入力	-120	100	0	0	0	0	0	0	0	0
	-90	0	78	12	0	0	0	8.3	1.1	0
	-60	0	0	73	4.1	0	2	21	0	0
	-30	0	0	0	100	0.2	0	0	0	0
	0	0	0	0	0	100	0	0	0	0
	30	0	0	0	35	0	65	0	0	0
	60	0	0	0	2	0	0	98	0	0
	90	0	3.1	1.6	0	0	0	0	95	0
	120	0	0	0	0	0	0	0	0	100

図 3.5 位置推定精度の混同行列 (入力：テストデータのインパルス応答による仮定の推定対象)

3.3 実験

次に、推定対象が実測した受信信号 $y_{test}(n)$ の位置推定を行う。

3.3.1 条件

$y_{test}(n)$ は、図 3.1 の各位置での、各可聴音をスピーカから流した時の受信信号である。各位置の仮想的な受信信号 $y_{train,i}(n)$ は 3.2 と同様に式 (3.2) を用いて求めた。可聴音、切り出したフレーム幅、フィルタバンクのチャンネル数、学習データのインパルス応答、インパルス応答の長さといったパラメータや条件については 3.2.1 と同様である。

3.3.2 結果

実測の受信信号による位置推定精度の混合行列を図 3.6 に示す。各位置の入力データ数は 100 個である (10×10 (信号の長さ \times 可聴音の種類))。図 3.6 より、位置 0 の信号は 88% と高い精度で、位置-30 の信号は 61% と十分といえないが 50% 以上の精度で位置推定できていることが分かる。しかし、他の位置の信号に関しては 50% に満たない精度で、位置推定できているとはいえない精度であり、3.2.2 の結果と比べ、大きく精度が落ちていることが分かる。また、位置推定精度の混合行列での正解精度の平均(以下、正答率)は 36% であった。3.2.2 の結果と比べ、精度が低くなった原因としては、3.2 のシミュレーションではインパルス応答を基にした各位置の仮想的な受信信号と仮想的な推定対象の比較であったため、差異が顕著に現れたが、今回の実験では、インパルス応答を基にした各位置の仮想的な受信信号と実測信号の比較で、実測によるノイズの影響で全体的に信号成分が異なってしまう、明確な差異が現れず位置推定ができなかったと考えられる。

		推定結果								
		-120	-90	-60	-30	0	30	60	90	120
入力	-120	27	11	2	13	12	4	9	4	18
	-90	5	29	9	10	17	5	15	6	4
	-60	4	0	32	10	16	4	33	0	1
	-30	0	1	4	61	25	2	2	0	5
	0	1	0	0	6	88	1	0	0	4
	30	0	1	3	63	22	6	0	0	5
	60	2	0	16	14	16	4	47	0	3
	90	4	21	10	6	17	1	20	19	2
	120	4	39	3	5	16	1	8	8	16

図 3.6 位置推定精度の混同行列 (入力：各位置での実測受信信号)

3.4 切り出すフレーム幅変更

3.3.2 で実測の受信信号による位置推定の精度が低かったため、切り出すフレーム幅を変更することで、精度向上を試みる。

3.4.1 条件

1 秒ごとの信号という短い信号データからフレーム幅を変更させるということで、今回は 0.2、0.4、0.6、0.8、1.0 秒の 5 種類のフレーム幅で実験を行う。切り出すフレーム幅以外のパラメータや条件に関しては 3.3.1 と同様である。

3.4.2 結果

フレーム幅を変更した際の正答率を表 3.2 に、フレーム幅が 1.0 秒の時の位置推定精度の混合行列を図 3.7 に示す。表 3.2 より、フレーム幅を大きくしていくにつれ、位置推定精度が向上していることが分かる。向上した理由としては、フレーム幅を大きくしたことにより、扱うデータ数が多くなることで、周波数軸の分解能が高くなり、位置によって異なる周波数成分の差が大きく現れ、位置推定の精度向上につながったと考えられる。ただし、図 3.7 より、1 秒ごとの信号のうち、1 秒全てのデータを入力データとしても、推定できている位置は少ないことが分かる。これは、実測信号でのノイズや 1 秒という短い信号では周波数軸の分解能が低くスペクトルが滑らかで、物体の位置を判別する際に重要な差異が現れない、といったことが考えられ、1 秒ごとの信号データから位置推定は困難であると分かった。ただ、各可聴音ごとの 10 秒分の精度結果で最も多く推定された位置を推定結果とすると、位置推定できる場所があったため、1 秒ごとに位置推定するのではなく、数秒ごともしくは数秒分の推定結果で位置推定を行えば、精度が向上すると考えられる。

表 3.2 正答率

フレーム幅	0.1s	0.2s	0.4s	0.6s	0.8s	1.0s
正答率	36%	39%	44%	47%	49%	51%

		推定結果								
		-120	-90	-60	-30	0	30	60	90	120
入力	-120	39	8	2	7	6	6	8	8	16
	-90	2	46	12	3	3	4	19	8	3
	-60	0	3	51	3	1	2	38	1	1
	-30	0	3	1	87	4	3	0	1	1
	0	0	1	0	2	95	0	0	1	1
	30	0	3	0	82	2	11	0	1	1
	60	0	3	24	3	1	2	65	1	1
	90	0	36	7	1	1	1	17	36	1
	120	0	50	3	0	1	1	4	12	29

図 3.7 位置推定精度の混同行列 (フレーム幅：1.0 秒)

第 4 章

おわりに

4.1 まとめ

本論文では、インパルス応答および可聴音 (音声や音楽など) による物体位置を推定することを目的として、シミュレーションと実測信号による位置推定を行った。今回の位置推定では、スピーカとマイクの直線上にある物体の位置を推定した。

シミュレーションでは、物体位置の推定を行う受信信号を実測ではなく、各物体位置のインパルス応答のテストデータ (10 回分) に可聴音を畳み込みした仮想の推定対象の受信信号を用いて、位置推定可能であるか検証した。その結果、多くの物体位置の信号が 100% もしくは 100% に近い精度で位置推定できていることが分かり、インパルス応答および可聴音を用いて物体の位置推定が可能であることが分かった。

実測信号による位置推定では、実測した受信信号を推定対象として位置推定を行った。その結果、精度が高い結果もあったが、多くが 50% に満たない精度で、位置推定できているとはいえない精度であった。シミュレーションの結果と比べ、精度が低くなった原因としては、シミュレーションではインパルス応答を基にした各位置の仮想的な受信信号と仮想的な推定対象の比較であったため、差異が顕著に現れたが、実測信号による位置推定では、インパルス応答を基にした各位置の仮想的な受信信号と実測信号の比較で全体的に信号成分が異なってしまう、物体位置の判別に必要な差異が現れず位置推定ができなかったと考えられる。また、切り出すフレーム幅を変更し、精度向上を試みた。その結果、フレーム幅を大きくしていくにつれ、位置推定精度が向上していることが分かった。向上した理由としては、フレーム幅を大きくしたことにより、扱うデータ数が多くなることで、周波数軸の分解能が高くなり、位置によって異なる周波数成分の差が現れ、位置推定の精度向上につながったと考えられる。しかし、精度が最も良いフレーム幅 1.0 でも推定できている結果は少なかった。このことから、物体位置の判別で重要とされている顕著な周波数成分が入力データ (1 秒分の

データ)に十分に含まれておらず誤推定が多くなったと考えられ、1秒ごとの信号データから位置推定は困難であると分かった。ただ、各可聴音ごとの10秒分の推定結果を見ると、最も多く推定された位置を推定結果とした場合、推定できる物体位置があったため、数秒ごともしくは数秒分の推定結果で位置推定を行えば、精度が向上すると考えられる。

4.2 今後の展望

今回の条件での位置推定では実測の受信信号での位置推定の精度が低いことが分かった。そのため、今後は位置推定精度の向上がまず必要である。精度向上するにあたって、数秒ごとの信号もしくは数秒分の推定結果から位置推定を行ったり、1秒という短い信号を使うにしても、どの周波数成分が推定に寄与しているか、逆に寄与していないか周波数成分の解析を行い、その結果をもとに位置推定を行うアルゴリズムを設計する必要があると考えられる。また、位相といった振幅スペクトル以外の情報や複数マイクによる受信信号から位置推定を行うことでも位置推定の精度が向上するのではないかと考えられる。環境に関しては、測定場所は無響室、物体の位置はスピーカーとマイクの直線上のx軸のみ、と限定的な環境であったため、今後は反射を考慮した位置推定やx軸上以外の物体位置の推定を行う必要がある。

謝辞

本研究を進めるにあたり、多大なご指導とご助言を賜りました杉田泰則准教授に心から感謝申し上げます。また、本研究の審査において適切なご指摘をいただいた圓道知博教授、ならびに南部功夫准教授に厚く御礼申し上げます。さらに、本研究に関して的確な意見や日々の議論をしていただいた信号処理応用研究室の皆様に感謝の意を表します。最後に、これまでの学業を支えていただきました家族に深く感謝いたします。

付録 A

各フレーム幅の位置推定精度

		推定結果								
		-120	-90	-60	-30	0	30	60	90	120
入力	-120	31	8	4	10	13	3	9	3	19
	-90	7	36	10	8	10	4	18	4	3
	-60	3	0	36	10	10	3	34	0	4
	-30	0	1	2	67	19	3	1	0	7
	0	1	0	0	7	86	0	0	0	6
	30	0	1	3	69	14	6	0	0	7
	60	2	0	16	9	10	3	55	0	5
	90	6	27	8	6	11	1	21	18	2
	120	6	40	4	4	10	0	8	10	18

図 A.1 位置推定精度の混同行列 (フレーム幅：0.2 秒)

		推定結果								
		-120	-90	-60	-30	0	30	60	90	120
入力	-120	28	6	3	11	12	4	6	8	22
	-90	1	37	21	6	8	1	17	6	3
	-60	0	2	42	7	7	1	38	2	1
	-30	0	2	2	78	11	4	0	2	1
	0	0	0	0	3	91	2	0	3	1
	30	0	2	1	76	10	8	0	2	1
	60	0	3	24	8	7	1	56	1	1
	90	0	27	6	4	8	0	18	35	2
	120	0	50	1	2	7	0	7	12	21

図 A.2 位置推定精度の混同行列 (フレーム幅：0.4 秒)

		推定結果								
		-120	-90	-60	-30	0	30	60	90	120
入力	-120	31	9	2	10	9	2	7	10	20
	-90	1	43	16	6	7	1	14	9	3
	-60	0	2	48	12	6	2	26	3	1
	-30	0	2	3	80	7	4	0	3	1
	0	0	1	1	4	91	0	0	3	0
	30	0	2	1	77	7	10	0	2	1
	60	0	2	24	8	5	2	55	3	1
	90	0	35	6	5	6	0	10	37	1
	120	2	50	2	2	5	0	5	10	24

図 A.3 位置推定精度の混同行列 (フレーム幅：0.6 秒)

		推定結果								
		-120	-90	-60	-30	0	30	60	90	120
入力	-120	36	10	0	7	10	3	9	7	18
	-90	1	43	12	6	4	2	20	8	4
	-60	0	3	50	5	2	2	34	3	1
	-30	0	3	2	85	4	3	0	2	1
	0	0	1	1	4	91	0	0	2	1
	30	0	2	3	80	3	10	0	1	1
	60	0	2	24	5	2	2	61	3	1
	90	0	35	5	2	3	1	16	37	1
	120	0	51	3	1	2	1	4	13	25

図 A.4 位置推定精度の混同行列 (フレーム幅：0.8 秒)

参考文献

- [1] S. Gao, S. Prasad, "Employing spatial analysis in indoor positioning and tracking using wi-fi access points", Proceedings of the Eighth ACM SIGSPATIAL International Workshop on Indoor Spatial Awareness pp.27-34, 2016 年
- [2] Yu-Chi Pu, Pei-Chun You, "Indoor positioning system based on BLE location fingerprinting with classification approach", Applid Mathematical Modelling, Volume 62, pp.654-663, October, 2018 年
- [3] 安齋恵一, 岡島匠吾, 坪川宏, "スマートフォンを用いた屋内位置の推定と歩行ナビゲーションシステム", 「マルチメディア, 分散, 協調とモバイル (DICOMO2011) シンポジウム」, 7 月, 2011 年
- [4] François Fleuret, Jérôme Berclaz, Richard Lengagne, Pascal Fua, Senior Member, "Multicamera People Tracking witha Probabilistic Occupancy Map", IEEE TRANSACTIONS ON PATTERN ANALYSIS AND MACHINE INTELLIGENCE, VOL.30, NO.2, FEBRUARY, 2008 年
- [5] 山藤浩明, 前川卓也, 松下康之, "赤外線カメラと再帰性反射材を用いたデバイスフリー屋内位置推定に関する検討", 情報処理学会研究報告 IPSJ SIG Technical Report, 11 月, 2015 年
- [6] Nissanka B. Priyantha, Anit Chakraborty, Hari Balakrishnan, "The Cricket Location-Support System ", Proceedings of the 6th annual international conference on Mobile computing and networking, pp.32-43, 2000 年

- [7] 杉野恭兵, 大園忠親, 新谷虎松, ”環境変化を考慮した Bluetooth ビーコンに基づくデバイスフリー屋内測位機構”, The 30th Annual Conference of the Japanese Society for Artificial Intelligence, pp.1-4, 2016 年
- [8] 宮本隆典, 堀江和正, 北川博之, ”可聴音を用いた透過法計測による屋内での人体検知”, 第 12 回データ工学と情報マネジメントに関するフォーラム, J6-5, 2020 年
- [9] 島田正治, 安川博, 伊藤良生, 田口亮, 張熙, 岩橋政宏, ”デジタル信号処理の基礎”, コロナ社, 第 1 版, pp.16-19, 6 月, 2006 年
- [10] 河原英紀, ”インパルス応答の基本概念”, 日本音響学会誌, 76 巻 3 号, pp.148-149, 2020 年
- [11] 金田豊, ”知識ベース”, 電子情報通信学会, 2 群-6 編 p4, 6 月 2012 年
- [12] 西谷洋哉, 杉田泰則, ”励起特徴と声道特徴を用いた音声感情認識の精度向上に関する研究”, 長岡技術科学大学大学院工学研究科修士論文, pp.19-20, 2024 年
- [13] ZRAR KH. ABDUL, ABDULBASIT K. AL-TALABANI, ”Mel Frequency Cepstral Coefficient and Its Applications”, IEEE Access, p.122137, November 23, 2022 年
- [14] J. Ancilin, A. Milton, ”Improved speech emotion recognition with Mel frequency magnitude coefficient”, ScienceDirect, p.3, March 9, 2021 年
- [15] 尼子就都, 増山直輝, 能島祐介, 石渕久生, ”クラスタリング手法における距離尺度の影響調査”, 第 35 回ファジィシステムシンポジウム 講演論文集 (FSS2019 大阪大学), no.TC3-3, pp.121-122 2019 年
- [16] <https://bgmer.net/>
- [17] 話速変換コーパス: SpeedSpeech-JA-2022, 先進的音声技術研究室, https://ast-astrec.nict.go.jp/release/speedspeech_ja_2022/download.html 2022 年

- [18] Takuma Okamoto and Yoshinori Shiga and Hisashi Kawai "Hi-Fi-CAPTAIN: High-fidelity and high-capacity conversational speech synthesis corpus developed by NICT", <https://ast-astrec.nict.go.jp/en/release/hi-fi-captain/>, 2023 年