

修士論文

反射音の周波数特性を用いた 透明・鏡平面の方位角推定手法

片山 寛基

奈良先端科学技術大学院大学

先端科学技術研究科

知能社会創生科学プログラム

主指導教員: 小笠原 司 教授

ロボティクス研究室 (情報科学領域)

令和2年 3月 13日提出

本論文は奈良先端科学技術大学院大学先端科学技術研究科に
修士(工学) 授与の要件として提出した修士論文である。

片山 寛基

審査委員：

小笠原 司 教授	(主指導教員, 情報科学領域)
中村 哲 教授	(副指導教員, 情報科学領域)
高松 淳 准教授	(副指導教員, 情報科学領域)
Garcia Ricardez Gustavo Alfonso 助教授	(副指導教員, 情報科学領域)

反射音の周波数特性を用いた 透明・鏡平面の方位角推定手法*

片山 寛基

内容梗概

ロボットにおいて、暗闇での把持や、透明物体の把持するにはまだ、問題がある。解決策として、音波を用いての対象平面の方位角方向推定手法を提案する。Photometric Stereo に代表される陰影解析を参考にして、波の強弱に着目することで、これまでは無い手法で平面の方位角の推定を行うことを目的とする。

本論文では、二通りの手法を提案する。1つ目は光学における物体面での波の反射モデルと音響学における音の周波数における反射特性の変化の2点に着目することで、モデルを作成し、モデルベースで平面の方位角方向を推定する。2つ目は実際に対象平面、スピーカ、マイクを用いて実験を行って取得したデータに対して、周波数領域での特徴について、サポートベクター回帰を用いて、モデルを作成することで、モデルフリー学習ベースで平面の方位角を推定する手法である。

モデルベースによる結果、平面の方位角推定はできなかった。原因は2つ考えられ、1つ目は仮定として、考えていた音響分野のモデルが壁面などのある程度の大きさを持った物体でなければならない点で、今回対象とした平面には合わなかった点が考えられる。2つ目は光学における反射モデルを考える際に、ある1点での反射のモデルを立てたが、音声分離がうまくいかず、欲しい点の音の情報を取得することができなかった点が考えられる。

モデルフリー学習ベースによる結果、 $-45 \sim 45$ の範囲で、誤差 $5[\text{deg}]$ 以内で方位角を推定できた。また、この手法は無響室だけでなく、実環境でも認識可能であった。

*奈良先端科学技術大学院大学 先端科学技術研究科 修士論文, 令和2年3月13日.

さらに，モデルフリー学習ベースの手法に関して，まず，必要な実験機器等の実験条件を検証するために，次の 1 ～ 3 の検証を行った．1. 使用した音の種類 (TSP 信号, トーン音), 2. 使用した周波数の範囲, 3. 使用したスピーカ (平面波, 球面波)．次に，対象物体の条件を検証するために，次の 4 ～ 6 の検証を行った．4. 対象面の材質 (ガラス, ダンボール), 5. 対象面の大きさ (大, 中, 小), 6. 対象面の数 (1 面, 2 面)．

キーワード

ロボット聴覚, 物体形状認識, 透明物体認識, サポートベクター回帰, 平面波

Estimating Azimuth Angle of Transparent and Mirror Plates Based on Reflected Sound Property in Frequency Domain*

Hiroki Katayama

Abstract

In a robot, there are still problems in grasping in the dark or grasping a transparent object. As a solution, we propose an azimuth angle estimation method of the target plane using sound waves. Aiming at estimating the azimuth of a plane by a method that has never been seen before, focusing on the strength of waves with reference to shading analysis represented by PhotometricStereo.

In this paper, we propose two methods. The first is to create a model by focusing on two points, a reflection model of waves on the object plane in optics and a change in reflection characteristics at the frequency of sound in acoustics, and estimate the azimuth angle of the plane on a model basis. The second is model-free learning by creating a model using support vector regression for features in the frequency domain for data acquired by conducting experiments using the target plane, speakers, and microphone. This is a method of estimating the azimuth of a plane on the base.

As a result of the model base, the azimuth of the plane could not be estimated. There are two possible causes. The first is assuming that the model in the acoustic

*Master's Thesis, Graduate School of Science and Technology, Nara Institute of Science and Technology, March 13, 2020.

field that I was thinking was not an object with a certain size such as a wall, and the second was when thinking about a reflection model in optics. We made a model of reflection at the point, but sound separation is perfect. This is a point that did not fit the plane targeted this time. It is conceivable that the information of the sound at the desired point could not be obtained.

As a result of the model-free learning base, the azimuth could be estimated within an error of 5 [deg] within the range of $-45 \sim 45$. This method could be recognized not only in an anechoic room, but also in a real environment.

Furthermore, regarding the model-free learning-based method, first, the following 1 $\sim$ 3 was verified in order to verify experimental conditions such as necessary experimental equipment. 1. Type of sound used (TSP signal, Tone sound), 2. Range of frequencies used, 3. Speakers used (Plane wave, Spherical wave). Next, in order to verify the condition of the target object, the following 4 $\sim$ 6 was verified. Material of target surface (Glass, Cardboard), 5. Size of target surface (Large, Medium, Small), 6. Number of target surfaces (Solo, Multiple).

Keywords:

Robot Audition, Object Shape Recognition, Transparent Object Recognition, Support Vector Regression, Plane Wave

目 次

1. 緒言	1
1.1 研究背景	1
1.2 関連研究	1
1.2.1 カメラを用いた透明物体認識	2
1.2.2 触覚センサを用いた透明物体認識	3
1.2.3 音を用いた物体認識	3
1.2.4 法線方向推定手法	4
1.2.5 センサ	5
1.3 研究目的	6
1.4 本論文の構成	6
2. 提案手法	7
2.1 提案手法の概要	7
2.2 モデルベース手法による方位角推定	7
2.2.1 音波の反射特性	7
2.2.2 反射音のモデル化	10
2.2.3 鏡面反射での反射音の振る舞い	11
2.2.4 拡散反射での反射音の振る舞い	12
2.2.5 推定モデルの考察	13
2.3 モデルフリー学習ベース (SVR) 手法による方位角推定	15
2.3.1 データセット作成方法	15
2.3.2 サポートベクター回帰	17
3. 方位角推定の実験	19
3.1 実験のセッティング	19
3.1.1 実験の目的	19
3.1.2 実験環境	20
3.1.3 実験器具	20

3.1.4	実験方法	23
3.2	モデルベース推定の結果	23
3.2.1	周波数に対するパワースペクトルの変化	24
3.2.2	方位角に対するパワースペクトルの変化	24
3.2.3	方位角に対する2周波数でのパワースペクトルの比の変化	25
3.3	モデル作成に関する考察	27
3.4	マイク, スピーカ, 環境による影響の考察	28
3.5	モデルフリー学習 (SVR) の結果	29
3.5.1	使う音の評価	31
3.5.2	使う周波数の範囲の評価	33
3.5.3	使うスピーカの評価	36
3.6	SVR を用いた場合の考察	37
3.6.1	サポートベクターの解析	38
3.6.2	汎化性能についての考察	40
4.	材料, 形状の違い, 複数面への応用実験	43
4.1	実験目的	43
4.2	実験セッティング	45
4.2.1	材料, 大きさ実験	45
4.2.2	複数面での実験	46
4.3	材料, 大きさ実験の結果	47
4.3.1	材料の違いによる変化の考察	49
4.3.2	大きさの違いによる変化の考察	51
4.4	複数面での実験の結果と考察	54
5.	結言	56
5.1	まとめ	56
5.2	今後の課題と展望	57
	謝辞	58

目 次

1	The overview of the proposed method	8
2	The energy acquired by the microphone changes as the frequency of the sound and the angle of the surface's azimuth change.	9
3	Model environment	10
4	Random incident diffusion coefficient of one-dimensional sinusoidal wall[Messured], Calculated value: Directed correlation method [Cal 1], Sample rotation method [Cal 2], Infinite wall[Infinite](Source: <i>Sakuma et al.</i> [1] Fig.9)	11
5	Relationship between energy (E) change with respect to frequency (f) and azimuth direction (N) when the sound source direction (S) and microphone direction (V) are fixed in the model	14
6	Relation between energy ratio $D(\mathbf{N} f_m, f_h)$ and azimuth angle $\mathbf{N}$ between mid and high frequencies at a point Q on the plane($f_m =$ 1000, $f_h = 2000$)	15
7	Flowchart to create dataset	16
8	Comparison between Zero cross and Short-time fourier transform	17
9	Experiment for checking model	19
10	Experiment environment	20
11	Experiment equipments	22
12	Position and parametor for experiment checking model	23
13	elationship between energy (E) change with respect to frequency (f) and azimuth direction (N) when the sound source direction (S) and microphone direction (V)	25
14	Relation between energy ratio $D(\mathbf{N} f_m, f_h)$ and azimuth angle $\mathbf{N}$ between mid and high frequencies at a point Q on the plane($f_m =$ 1000, $f_h = 2000$)	26

15	Relation between energy ratio $D(\mathbf{N} f_m, f_h)$ and azimuth angle $\mathbf{N}$ between mid and high frequencies at a point Q on the plane($f_m = 1000, f_h = 2000$)	26
16	Comparison of change of amplitude in frequency domain	28
17	Result of Experiment based on SVR	30
18	Comparison of result about sound kinds	32
19	Comparison of frequency range based on case that I create models individually	35
20	Experiment for speaker speculat speaker	36
21	SVR result using beamforming data	37
22	Result of PCA	38
23	Parametric eigen space method	39
24	Comparison of result about the difference of microphon	41
25	Exsample of experiment environment for checking the changes in characteristics of material, size and plate number	44
26	Experiment Materials size: left is 200 [mm] $\times$ 400 [mm], middle is 100 [mm] $\times$ 200 [mm], right is 50 [mm] $\times$ 100[mm]	46
27	Comparison of the changes in characteristics of material and size .	48
28	Comparison of the changes in characteristics of material	50
29	Comparison of the changes in characteristics of size by the error .	53
30	SVR result of the data that target object is 2 cardboard plate . .	54
31	SVR result of the data that target 1 plate from 2 cardboard plate data	55

表 目 次

1	Characteristic of each sensor	5
2	Response of reflection in various frequencies.	9
3	Environment Equipment	21

4	Parameter for experiment checking model	24
5	Result of SVR	29
6	Comparison of R^2 , RMSE by change of freequency range	34
7	Contribution rate	39
8	Comparison of result about the difference of microphon at same environment	42
9	Three types of sizes to verify changes due to size	44
10	Parameter for experiment checking the changes in characteristics of material, size and plate number	45
11	Comparison of R^2 , RMSE by change of material and size	47
12	Comparison of ϵ by changing the target object sie	52

1. 緒言

1.1 研究背景

家庭内には、ロボットにおいて把持が難しい物品が数多く存在する．例としては、コップやペットボトルといった透明な物体，鏡などの光が鏡面反射する物体がある．更に、夜間などの光がない環境でも認識は難しい．そのため、夜間は人間と同じで明かりがないと動くことができない．このようなロボットが不得手とする物体の位置と姿勢を認識する需要は増加している．特に、透明物体の把持については、コップやペットボトルといった身近なものにも存在するため、需要は大きい．

このような透明な物体”検出”手法に関して研究は多数存在する．使用センサとしては、光学式距離センサ [2]，カメラ [3]，超音波センサ [4] 等，その種類も様々である．このような研究により、どの位置に透明な物体があるのかを認識することは可能である．

しかし、透明な物体”把持”について考えると、別の問題が出てくる．それは物体把持をする際に、位置というのは非常に重要であるが、位置だけでなく、同時に面の法線方向も重要ということである．[5]

このような、透明物体の形状や法線方向の認識はロボットにおける透明物体の把持において重要である．

1.2 関連研究

ロボットにおける透明物体認識の研究を紹介する．1つ目はカメラを用いる手法である．これは非接触で認識可能であるため、一般的であり、第 1.2.1 節にて紹介する．2つ目は触覚センサを用いる手法である．これは、カメラとともに使われることも多く、第 1.2.2 節で紹介する．3つ目は音波を用いた手法である．これは、音波が非接触で、かつ物体の色の影響を受けずに測定可能である性質を利用できる．そこで、対象が透明物体だけでなく、音波を用いての物体形状認識手法を第 1.2.3 節で紹介する．

次に、物体把持において、重要な物体の法線方向を取得する手法について、これまでに提案された手法は、主に2つの方法が提案されてきた。1つ目は、距離センサを用いる場合である。物体面上の任意の3点の位置を計測することで、平面をあてはめることで法線方向を推定する手法がある。2つ目はカメラを用いる場合である。Photometric Stereoに代表される陰影を用いることで法線方向を直接推定する手法がある。これら2つの手法に関する研究を第1.2.4節

これらの研究に加えて、最後に、これまで紹介した研究で用いられている一般的なセンシングデバイス(画像、音、触覚)についてまとめ、第1.2.5節で紹介し、物体の形状認識についてそれぞれの利点や欠点などの特徴について述べる

1.2.1 カメラを用いた透明物体認識

透明物体ではないが、3DCADモデルを元にして、カメラを使い認識を行う研究がある。Jeffreyら[6]によって作成されたDex-Net:2.0では物体の3DCADモデルを用いることによって、物体の把持計画を立てる。このように、現在物体の3DCADモデルがあれば、カメラによって様々な物体を把持することができる。

しかし、ここで問題がある。それは、透明物体の場合、形状を3Dスキャンする技術は確立されておらず、まずどのようにして透明物体のCADモデルを作成するのか?といった問題がある[7],[8]。そこで、はじめに思いつくのは手動でCADモデルを作ること、透明物体の認識を行う手法である。しかし、この手法を用いる場合、3DCADモデルを作成するのに、時間がかかり、効率的ではないという問題点があった。そのため、Phillipsら[9]は、インターネットから似ている3DCADモデルを探してくることによって、より簡単に認識を行った。しかし、似た形状の3DCADモデルがない場合、自分で作成する必要がある。Osadchyら[10]は透明なものにパウダーやスプレーを行うことによって、物体表面を不透明にし、従来の3Dスキャン技術を用いて3Dモデルを作成した。

このような、透明物体の3DCADモデル作成手法をロボットの把持への応用した例として、Ilyaら[11]は、Microsoft社のKinectを用いて透明な物体の認識を行った。このとき、透明な物体の表面をスプレーで不透明にすることで、3DCADモデルを作成し、そのデータとRGB-Dカメラの距離情報を合わせることで、透

明物体の物体姿勢の認識をリアルタイムに行い、透明物体の把持を行った。

しかし、これらの手法は多くの場合、3DCAD モデルを使用する必要があるため、既知の物体のみに対応しており、未知の物体の認識は難しい。

1.2.2 触覚センサを用いた透明物体認識

Kim ら [12] は、磁気を利用した触覚センサを開発した。彼らはこの触覚センサーを指ごとに3つ使用した3指グリッパーを作成した。センサ感度は 0.016mV/kPa で検出範囲は 350kPa であった。そのため、この触覚フィードバックを使用することで、様々な形状の対象物体 (例えば、造花、ワイングラスや紙コップ) を損傷なく把持に成功した。

Shaoxiong ら [13] は、触覚を用いた形状認識は全体形状を認識する効率が低い問題点に対して、カメラによって、物体の大体の形状を事前知識より認識後、触覚センサによって未知の物体でも正確な 3D モデルをより効率的に作成した。これにより、カメラのみでは認識が難しい透明なペットボトルの 3D モデルを作成することに成功した。

このように、マルチモーダルなセンシングにより、視覚の弱点である 2D-3D の曖昧さ、触覚の弱点である形状全体を再構築するための効率の低さについて、お互いの強みによって補う合うことによって、より汎用性の高いセンシングシステムを作成することができる。

1.2.3 音を用いた物体認識

音波を用いての物体形状認識手法としてもっとも用いられる手法として、音を発してから物体表面で反射し、返ってくるまでの時間を用いた距離計測により物体形状を認識する手法が研究されている。Olaya ら [14] は、水中における、ソナーを使用した環境認識を行った。水中の場合は空気中に比べて音速が早く、減衰も少ないため理想的な環境ではある。しかし、空気中では音の減衰の影響を考慮する必要がある、同様の手法は用いることができない。David ら [15] は、死角にある物体の形状をスピーカから出した音が反射して帰ってきたときの時間差を見るこ

とによって、見えない位置にある物体の形状を認識した。しかし、物体が 40 [cm] × 40 [cm] 以上のものを対象としており、空間分解能は 1.5 [cm] である。

Amit ら [16] は、人間の手のジェスチャーを音を使って識別する手法を開発した。Ivan ら [17] が作成した超音波マイクアレイを用いて、スピーカから出た音が反射してマイクに入る音の変化によって、ジェスチャーを分類した。このときに、到達時間差だけでなく、音波の強さの情報を用いることによって、精度向上だけでなく、到達時間差だけのときと比べて、比較的に小さい物体の形状認識を成功した。

Ohtani ら [18] は、音波を用いて 5 種類の透明なペットボトルの分類を行った。超音波マイクロホンアレイを用いる事によって、スピーカから出た音が物体に反射して帰って来た際にマイクに入る音の強さの変化に対して、Deep learning を使用することによって、ペットボトルのような軽い物体の分類に成功した。

しかし、これらの手法の多くの場合、物体姿勢や形状の分類はできているが、法線方向の数値での推定までできていない。

1.2.4 法線方向推定手法

距離センサを用いる方法として、Yamaguchi ら [19] は、光学式の距離センサアレイの開発を行い、対象物体反射して受光器で取得した光の到達時間差だけでなく、反射光の強さも測定するセンサをエンドエフェクタの指先に敷き詰めることによって、掴む部分の物体形状を認識し、物体把持を行う手法を提案した。さらに、反射光の波の強さから布と紙パックの識別をすることで、パックの中に入っている対象物体の取り出しが可能となった。しかし、欠点としては透明物体の把持は考慮されていない。

2 つ目のカメラを用いる方法として、Photometric Stereo が挙げられる。この手法は光源の方向が移動することによって、カメラにうつる物体の陰影は変化することを利用している。この陰影の変化について光の反射モデルを当てはめることによって、写真の 1 ピクセルの明るさとその点の法線方向に関係があるというランバート反射モデル [20] を利用した手法である。Guanying ら [21] は、Photometric Stereo の精度を上げるべく、Deep learning 手法を付け加えることによって、物

体の色や環境光に対して，これまでの手法よりもロバストな手法を提案した．

1つ目の距離センサを用いる場合の手法に関しては，透明な物体に対してロバストな音を使った手法もある．しかし，2つ目の中でも特にこの Photometric Stereo に代表される法線方向を取得する手法は透明物体や鏡に適応されることは少なく，さらに，特に光の陰影を用いる手法は，同じ波である音波を用いて法線方向を推定する手法は我々の知る限り存在しない．

1.2.5 センサ

これまで，透明物体の認識について，光 (距離センサ，カメラ)，触覚，音の三種類のセンサの視点から関連研究を紹介した．Table 1 では，それぞれのセンサについて，特徴をまとめている．

Table 1 より，カメラと触覚センサの組み合わせについては，第 1.2.2 節で述べたが，有用である．しかし，カメラ，触覚センサのマルチモーダルセンシングシステムの弱点として，物体近傍のセンシングの弱さがある．カメラでは，物体近傍では，オクルージョンが発生することが多く，触覚センサは接触するまでは認識ができない．このような近接覚の認識手法として，小山ら [22] は光学式の距離センサアレイをエンドエフェクタの指先につけ，物体表面の形状を素早く認識することによって，移動している物体を掴むことに成功している．しかし，光学式の距離センサを用いている点で透明物体等を把持することは考慮されていない．

Table 1 Characteristic of each sensor

Sensor	Strong point	Weak point
Optical	Measurement range is large	Object color like transparent
Tactile	Mesurement the shape accurately	Inefficient to measure entire shape
Auditional	Robustness against object color	Spacial resolution is low

1.3 研究目的

センサによりそれぞれ利点、欠点がある。この中でも、私が着目したのは音である。現在、最も頻繁に用いられているカメラと触覚センサのマルチモーダルなセンシングは、透明物体での近接での非接触センシングが必要な場合、弱点となることが多い。そこで、さらに、マイクを用いて物体にあたって反射した音を用いることができれば、より、ロバストに透明物体の検出をすることができると考えられる。

そのため、本研究では音に着目する。しかし、現手法でよく用いられている音の到達時間差を用いる方法は空間分解能が低いため使用が難しい。そこで、光学においてより分解能が必要な場合に用いられる Photometric Stereo のような直接的に法線方向を取得する手法を参考にする。これにより、音波の時間差ではなく、波の強さに着目した、新しい音波を用いた法線方向の分類ではなく数値の推定手法の開発を目指す。

1.4 本論文の構成

本論文は全 6 章から構成される。以下に各章の概要を述べる。

- 第 1 章である本章は本研究の背景、関連研究と研究目的について述べた。
- 第 2 章では、最も単純な形状である平板の方位角の方向を取得する手法について 2 通り述べる。
- 第 3 章では、提案手法の有効性を示すための実験方法と実験結果、考察について述べる。
- 第 4 章では、さらに、第 3 章において見つかった疑問点について追加の検証実験について、方法と結果、考察を述べる。
- 第 5 章では本研究の結論と今後の展望について述べる。

2. 提案手法

2.1 提案手法の概要

本研究ではスピーカから出た音が対象平面に反射してマイクにて検出された音を使用し、発する音の周波数と反射音の強度の関係をを使うことによって、対象平面の方位角方向を推定する手法を提案する。この提案手法の概要について、Fig. 1にて示す。

推定手法としては、二通りの方法によって、対象面の方位角方向とマイクで取得した音の関係をモデル化を行った。

一つ目は音響、物理の音の反射モデルを参考にすることで、方位角度と反射音の関係をモデルベースで解いた手法である。詳細は第 2.2 節で示す。

二つ目は無響室、実環境二つの環境で集めたデータに対して、サポートベクター回帰を用いて、モデルフリー学習で解いた手法である。詳細は第 2.3 節で示す。

2.2 モデルベース手法による方位角推定

2.2.1 音波の反射特性

光の可視範囲の波長が 780[nm] ~ 380[nm] なのに対して、音の可聴範囲での波長は 17[m] (20[Hz]) ~ 17[mm] (20[kHz]) で長さの差が 1000 倍もある。そのため、光に比べて、音波はその周波数によって反射特性に変化がでる [23]。

周波数の違いによる反射特性の変化を Table 2 に示す。音波の反射は low(低域)では、小さな傾きや凹凸には関係なく鏡面反射する。Middle(中域)では、鏡面反射成分が強い。しかし、周波数が高くなるにつれて、拡散反射成分が強くなるため、High(高域)では、拡散反射の影響が大きくなる。本論文では、中域から高域に渡って、反射特性が変化することに着目する。低域を無視した理由としては、低域では、マイクとスピーカの入力音圧レベルと出力音圧レベルの周波数特性が著しく小さくなることが多いという点と低域は生活雑音が多い点の 2 点の問題が挙げられる。それぞれの反射のエネルギーに着目すると、鏡面反射では、反射音は入射角に応じて、ある定まった方向にのみ反射する。それに対し、拡散反射で

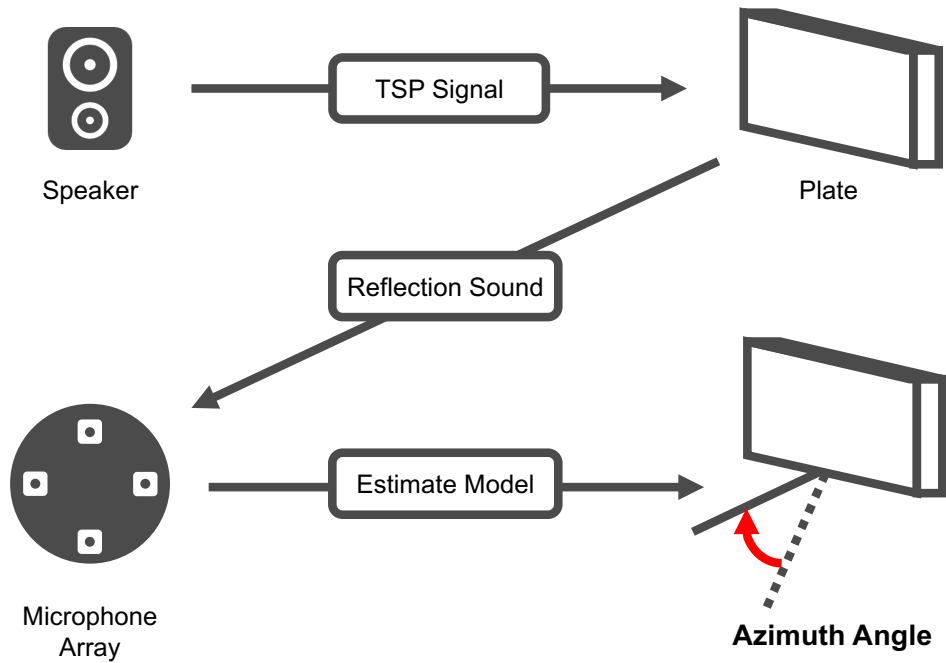


Fig. 1 The overview of the proposed method

は、入射角により反射する音のエネルギーは影響されるものの、すべての方向に
一様に反射する。

Fig. 2(a) のように、鏡面反射した音がマイクに入りやすい面の角度の場合、中
域では鏡面反射した音波を主に観測し、高域では拡散反射した音波を観測する。
鏡面反射で観測されるエネルギーの方が大きいため、周波数の変化により全体と
して観測されるエネルギーが減少していくと考えられる。

Fig. 2(b) のように、鏡面反射した音がマイクに入りにくい面の角度の場合、中
域では鏡面反射した音波を観測することは難しく、高域で拡散反射した音波が増
えるに従い観測される音波のエネルギーも増加していくと考えられる。

このように、中域 (鏡面反射) ・高域 (拡散反射) のエネルギーの差と物体面の方
位角にはある関係があると考えられる。

Table 2 Response of reflection in various frequencies.

	Frequency	Characteristic
Low	20[Hz]-100[Hz]	Perfect specular
Middle	100[Hz]-1[kHz]	Specular
High	1[kHz]-20[kHz]	Diffuse

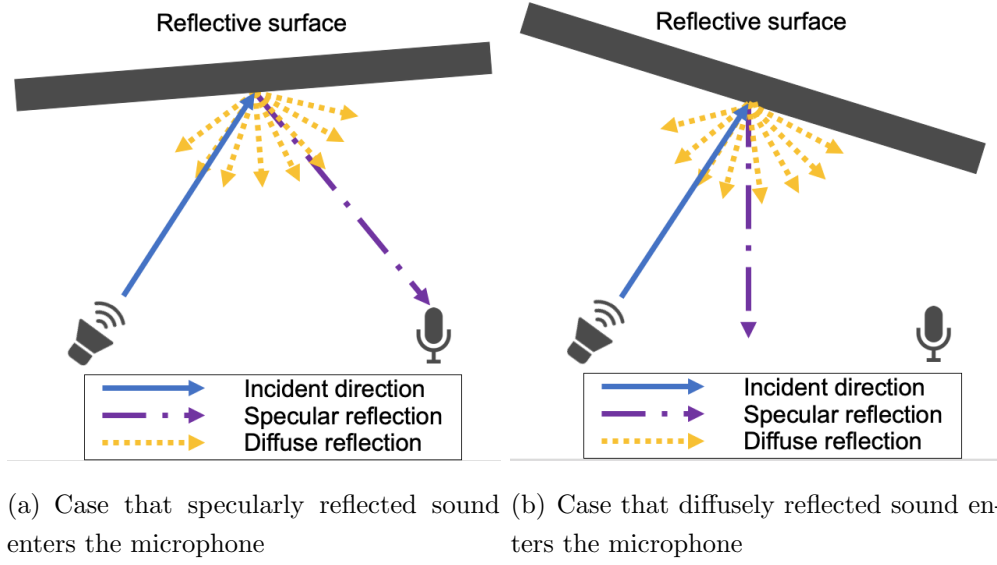


Fig. 2 The energy acquired by the microphone changes as the frequency of the sound and the angle of the surface's azimuth change.

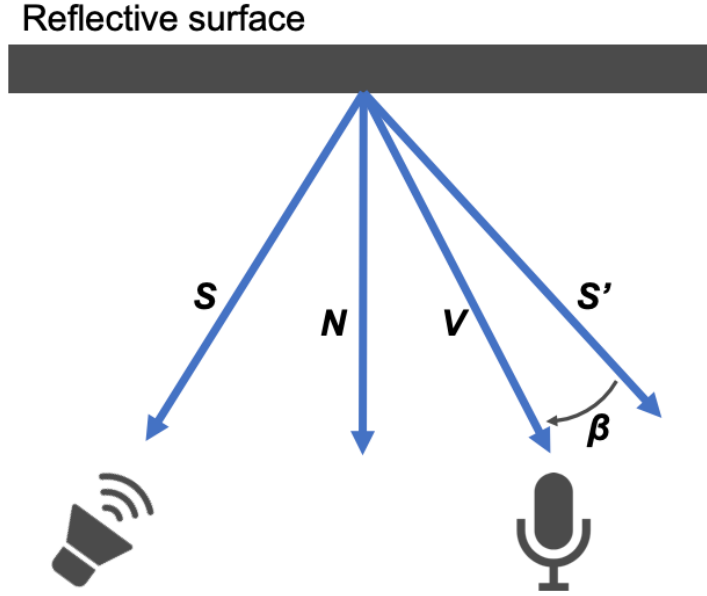


Fig. 3 Model environment

2.2.2 反射音のモデル化

光学における物体面での反射モデルを参考にすると、反射音は鏡面反射と拡散反射の和でマイクに入ってくると考える。光と違い、音は周波数に応じて、鏡面反射、拡散反射の割合が変化していくものと考えられる。

Fig. 3 のような環境において、レイトレーシングにより、仮想的な音波の線を用いて反射について考える。音源が S から入射した際に平面上のある点 Q 点での反射したとする。ここで、面の方向を N 、音波が鏡面反射した際の反射方向を S' 、観測する方向を V とした時、 S' と V の角度の差を β とおく。ある点 Q で反射してマイクで取得できる音のエネルギー E は以下の用に表す事ができる。

$$E(f, S, N, V) = \gamma_{\text{spe}}(f)e_{\text{spe}}(S, N, V) + \gamma_{\text{dif}}(f)e_{\text{dif}}(S, N). \quad (1)$$

ただし、 $e_{\text{spe}}, e_{\text{dif}}$ は、ある基準の音源における、マイクが受け取る鏡面反射、拡散反射音のエネルギー、 $\gamma_{\text{spe}}, \gamma_{\text{dif}}$ は、ある基準の音源における、鏡面反射率、拡散

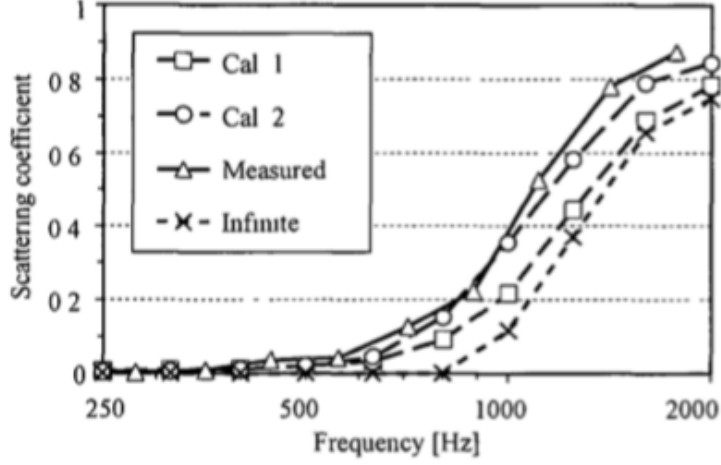


Fig. 4 Random incident diffusion coefficient of one-dimensional sinusoidal wall[Messured], Calculated value: Directed correlation method [Cal 1], Sample rotation method [Cal 2], Infinite wall[Infinite](Source: *Sakuma et al.* [1] Fig.9)

反射率のをそれぞれ示す．また， e_{spe}, e_{dif} について詳細はそれぞれ以下にて説明している．

このとき，今回反射以外でのエネルギー損失をないと仮定し， $\gamma_{spe} + \gamma_{dif} = 1$ と考える．反射音における拡散反射率の関数 γ_{dif} は佐久間らが実測によって求めた周波数が変化した際の拡散反射率の変化の結果から引用した図 Fig. 4 [1] を参考にして，シグモイド関数を用いて以下の式 (2) のように設定した．

$$\begin{cases} \gamma_{spe}(f) = 1 - (\gamma_{dif}(f)) \\ \gamma_{dif}(f) = 1/\{1 - \exp(-af + b)\}. \end{cases} \quad (2)$$

このとき，定数 a, b は Fig. 4 を参考にして $a = 1/150, b = -25/3$ とした．

2.2.3 鏡面反射での反射音の振る舞い

鏡面反射における観測点でのエネルギー e_{spe} について考える．式の定義には，光学における Phong のモデル [24] を参考にした．Phong のモデルは経験に基づ

く古典的なモデルであり，次式のように照明方向の正反射方向と観測方向のなす角 β_{spe} の余弦のべき乗として近似するモデルである．このとき，反射波の強さ i は

$$i = \rho_{\text{spe}} \cos^n(\beta_{\text{spe}}), \quad (3)$$

と表せる．

このとき， ρ_{spe} は鏡面反射率であり，鏡面反射の強さを表す．係数 n は，表面の粗さを表すパラメータである．このモデルは，あくまで経験に基づくものであり，エネルギー保存の法則を満たす保証がないなどの性質は気をつける必要はあるが，簡単に計算できるため，今回は使用した．

以上の Phong のモデルを参考にして，鏡面反射音 e_{spe} のエネルギーを式 (4) のように定義した．

$$e_{\text{spe}}(\mathbf{S}, \mathbf{N}, \mathbf{V}, n) = \cos^n \beta(\mathbf{S}, \mathbf{N}, \mathbf{V}). \quad (4)$$

このとき， $\beta(\mathbf{S}, \mathbf{N}, \mathbf{V})$ について反射の法則と余弦定理を用いることで，求めていく．まず， $\mathbf{S}'$ は反射の法則より $\mathbf{S}, \mathbf{N}$ を用いて以下のように表すことができる．

$$\mathbf{S}' = \mathbf{S} + 2(-\mathbf{S}^T \mathbf{N}) \mathbf{N}. \quad (5)$$

次に， $\mathbf{V}$ と $\mathbf{S}'$ のなす角 β は余弦定理より以下の用に表す事ができる

$$\cos \beta = \frac{\mathbf{S}'^T \mathbf{V}}{|\mathbf{S}'| |\mathbf{V}|}. \quad (6)$$

これより，

$$\cos \beta(\mathbf{S}, \mathbf{N}, \mathbf{V}) = \frac{(\mathbf{S} + 2(-\mathbf{S}^T \mathbf{N}) \mathbf{N})^T \mathbf{V}}{|\mathbf{S} + 2(-\mathbf{S}^T \mathbf{N}) \mathbf{N}| |\mathbf{V}|}, \quad (7)$$

となる．

2.2.4 拡散反射での反射音の振る舞い

拡散反射における観測点でのエネルギー e_{dif} について考える．拡散反射が入射光が表面層内部で乱反射することで生じる成分であり，観測方向に依存せず，あらゆる方向に均一の強度で観測される．この式の定義には光学における Lambert

のモデル [20] を参考にした．Lambert のモデルは，式 (8) のように照明方向と法線方向のなす角の余弦に比例すると仮定する．反射波の強さ i は

$$i = \rho_{\text{dif}} \mathbf{S}^T \mathbf{N}, \quad (8)$$

となる．

拡散反射音のエネルギーを式 (9) のように定義した．

$$e_{\text{dif}} = \kappa \mathbf{S}^T \mathbf{N}. \quad (9)$$

このとき， κ は拡散反射の中でもマイクに入ってくる音の割合を示す

2.2.5 推定モデルの考察

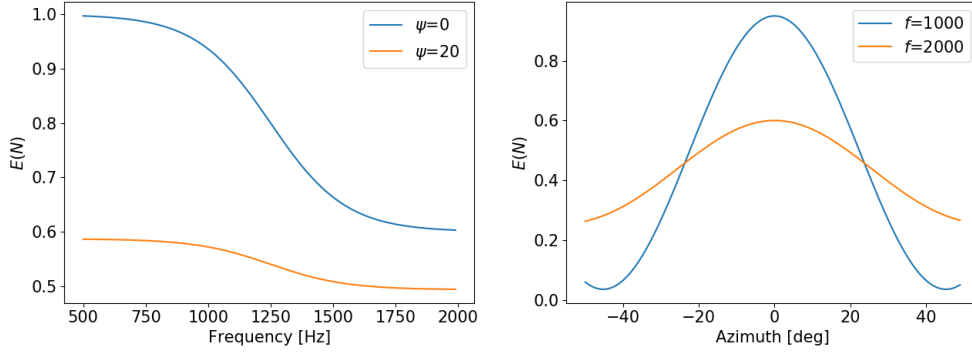
平面上のある点 Q における式 (1) に式 (2)，式 (4)，式 (9) を当てはめる．音源方向 $\mathbf{S}$ とマイク方向 $\mathbf{V}$ は既知であるため，この式は音の周波数の高さ f と面の方位角方向 $\mathbf{N}$ の 2 つが変数となる．音源方向 $\mathbf{S}$ とマイク方向 $\mathbf{V}$ を固定した場合にこの 2 つの変数において，変化させてみたものを Fig. 5 に示す．ただし， $\mathbf{V} = (0, -1)$ ， $\mathbf{S} = (0, -1)$ とおいた．そのため，今回の平面とマイク，スピーカの配置では $\mathbf{N} = (-\cos \psi, -\sin \psi)$ とした時， $\psi = 0[\text{deg}]$ で最大のエネルギーになる

Fig. 5(a) のように，横軸を周波数でとると，エネルギーの周波数応答を見ることができる．よって，鏡面反射成分と拡散反射成分では鏡面反射のほうが強いいため前半のほうが大きく，周波数が高くなるにつれて小さくなっていることがわかる．方位角の例として， $\psi = 0, \psi = 20$ の時を比べてみると，確かに $\psi = 0$ のほうが大きくなっている

つぎに Fig. 5(b) のように，横軸を方位角を取ると，エネルギーと角度の関係がわかりやすい．また，周波数で比べてみても鏡面反射成分が強い $f = 1000$ のほうが大きく変化していることがわかる．

最後に，物体面の方位角と中域 ($f_m[\text{Hz}]$)，高域 ($f_h[\text{Hz}]$) での比 D との関係について考えると， D は $\mathbf{N}$ のみを変数となり，以下のような式 (10) になる

$$D(\mathbf{N}) = \frac{E(\mathbf{N}|f_m)}{E(\mathbf{N}|f_h)}. \quad (10)$$



(a) Frequency response of energy $E(f)$ with constant azimuthal direction N at a point Q on the plane surface
 (b) Relation between energy $E(N)$ and azimuth N when sound frequency f is constant at a certain point Q on the plane surface

Fig. 5 Relationship between energy (E) change with respect to frequency (f) and azimuth direction (N) when the sound source direction (S) and microphone direction (V) are fixed in the model

また、これを図に表してみると、Fig. 6 のようになる。

$-50 < \psi < 50$ の範囲で方位角に対して、エネルギーの比 D が関係するモデルを作成することができた。このとき、音源の入射方向 $S = (-\cos \phi, -\sin \phi)$ ($0[\text{deg}] \leq \phi \leq 180[\text{deg}]$) とする。 D と N の関係の図より、物体の方位角を式 (10) から求める場合、二通りの解き方が考えられる。一つは方位角を制限する。 $\phi = 90[\text{deg}]$ の場合 $0 < \psi < 50$ の範囲に制限することで、単調減少するため一意に求める事ができる。もう一つは複数方向からの音源を用意することである。 Fig. 6 でもあるように $\phi = 90[\text{deg}]$ だけでは一意には求めることができないが、複数方向からの音源例えば、 $\phi = 45[\text{deg}]$ があれば、連立方程式より求めることが可能になる。

これより、この D を N に関して方程式を解く事によって、方位角方向 N を推定することができる。

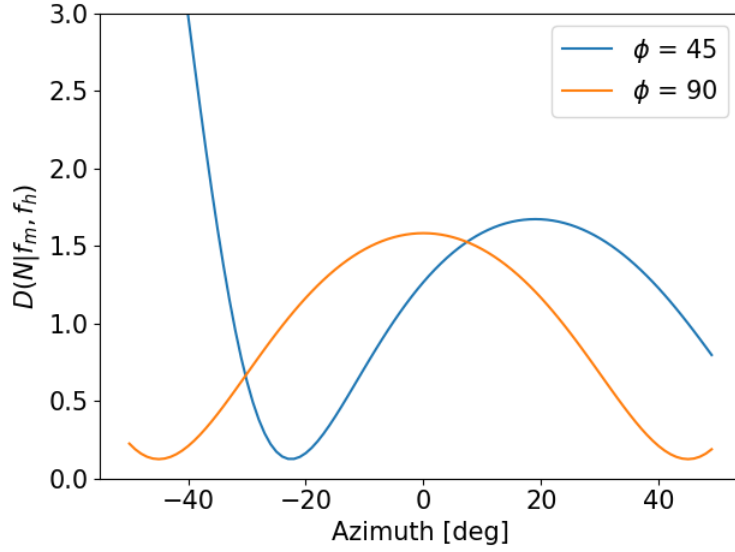


Fig. 6 Relation between energy ratio $D(N|f_m, f_h)$ and azimuth angle N between mid and high frequencies at a point Q on the plane ($f_m = 1000$, $f_h = 2000$)

2.3 モデルフリー学習ベース (SVR) 手法による方位角推定

2.3.1 データセット作成方法

音のデータを作成するフローについて、Fig. 7に示す。フローについて、順番に示す。

まず、実験データから使用する音を切り取る。このとき、ゼロクロス検出を用いる。ゼロクロス検出とは、ある区間において正から負、もしくは負から正に値が移動した回数を数えることで信号の周波数等変化を認識が可能である。これにより、周波数の変化をフーリエ変換なしに見ることができる。今回使用している音波は徐々に周波数が大きくなる音を使用しているため、最初の音の起こりは認識しづらいが音の終わりはわかりやすいため、今回使用した音のゼロクロスの結果と周波数の結果の比較の図を Fig. 8に示す。

このように、ゼロクロスの値が音が終わるさいに急激に落ちる。このときの値を音が終わる時間として、それ以前の約1秒のデータを今回の実験データとして

使用する。

次に周波数領域になおす。音のデータセットを作成する際には時間領域のデータでは、雑音の影響を受ける。そのため、フーリエ変換を用いる。このときのフーリエ変換では、高速フーリエ変換を用いる。このとき、フーリエ変換のパラメータとして、窓関数は周波数分解能が良いハミング窓、サンプリング周波数は44100 [Hz]、データ数は45120 個 (データ時間は約1 秒) を用いた。

最後にデータを整理する。今回は第 2.2 節でも示したが、1000 [Hz] から 2000 [Hz] の変化に着目するため、まず、1000 [Hz] から 2000[Hz] のデータのみで切り取る。次に、変化を見るために、平均0 分散1 で正規化を行う。最後に、ノイズを減らすために平滑化を行う。今回は周波数領域において、移動平均法を用いた。このとき、平滑化の幅は512 個のデータごとで行っている。

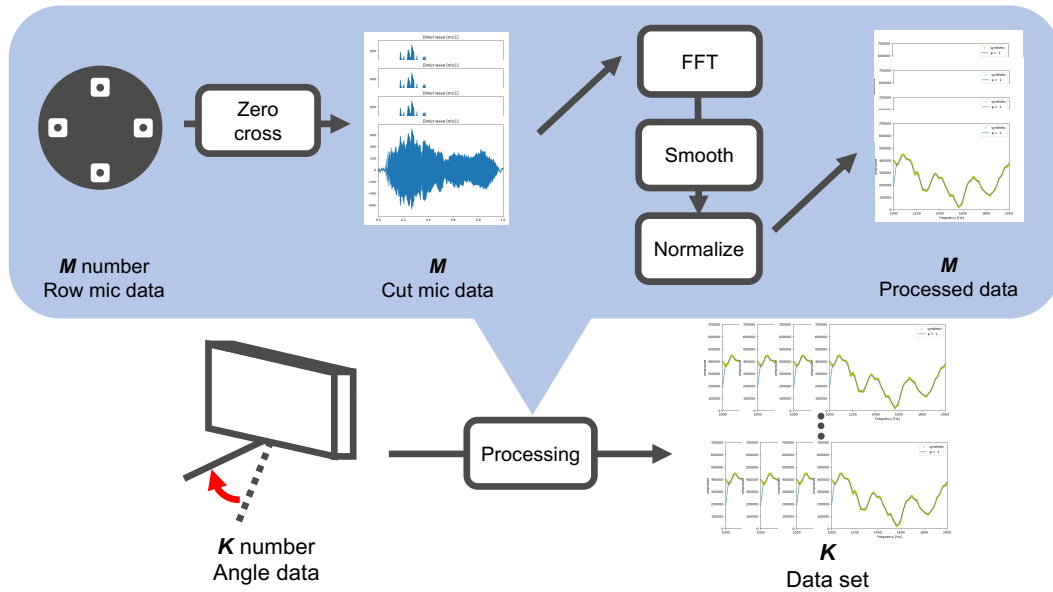


Fig. 7 Flowchart to create dataset

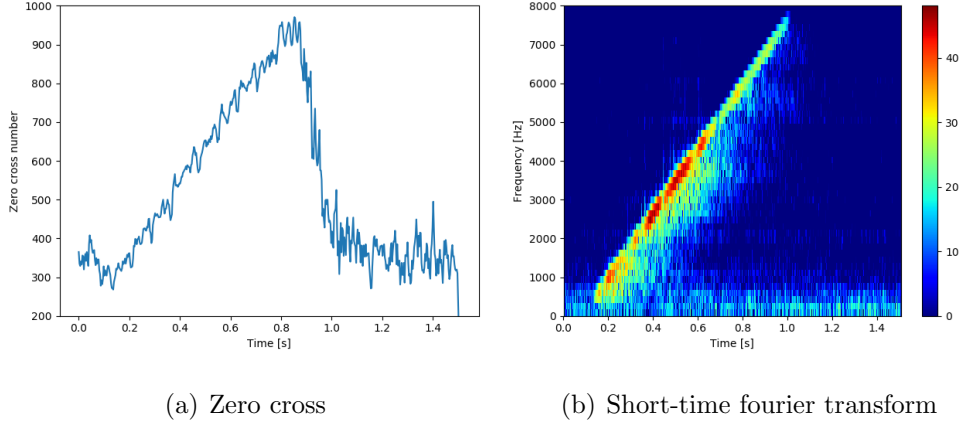


Fig. 8 Comparison between Zero cross and Short-time fourier transform

2.3.2 サポートベクター回帰

本研究では，平面の方位角を推定するのに，サポートベクター回帰 (Support Vector Regression: 以降 SVR [25]) を用いる．SVR の特徴として，誤差に不感帯を設けることで，ノイズの影響を受けにくく，さらに，基本的には線形な回帰分析手法であるが，カーネルトリックを使うことによって，非線形の回帰モデルになる．本研究では，ガウシアンカーネルの非線形の回帰モデルを用いることで，対象平面の方位角推定を行う．

以下に SVR の簡単な理論について説明する．

非線形 SVR では，次を最小化する係数を求めます．

$$L(\alpha) = \frac{1}{2} \sum_{i=1}^N \sum_{j=1}^N (\alpha_i - \alpha_i^*)(\alpha_j - \alpha_j^*) G(x_i, x_j) + \epsilon \sum_{i=1}^N (\alpha_i + \alpha_i^*) - \sum_{i=1}^N y_i (\alpha_i - \alpha_i^*) \quad (11)$$

これには，次の条件が適用される．

$$\begin{aligned} \sum_{n=1}^N (\alpha_n - \alpha_n^*) &= 0 \\ \forall n : 0 &\leq \alpha_n \leq C \\ \forall n : 0 &\leq \alpha_n^* \leq C. \end{aligned} \quad (12)$$

条件より，推定に使用される関数は以下に等しくなる．

$$f(x) = \sum_{i=1}^n (\alpha_i - \alpha_i^*) G(x_n, x) + b \quad (13)$$

このとき、 G は非線形カーネル関数で以下のように考えられる。

$$G(x_n, x) = \exp(-\gamma \|x_n - x\|^2) \quad (14)$$

α_n, α_n^* が誤差 ϵ の範囲内にない場合に $\alpha_n - \alpha_n^* = 0$ となり、SVR の回帰式を作成するサンプルとなる。 $\epsilon = 0$ のときが最もノイズに強いモデルとなる。

このとき、 C, γ, ϵ の3つがハイパーパラメータとして存在する。しかし、 C は ϵ から導くこともできるため、 C, γ の二つのパラメータを求める。この2つのハイパーパラメータを求めるためにグリッドサーチ交差検証を行う。グリッドサーチ交差検証とは、モデルの精度を向上させるために用いられる手法であり、全てのパラメータの組み合わせを試すグリッドサーチに加えて、さらにテストデータと訓練データの組み合わせを試す交差検証をすることで、最も評価精度が高いパラメータを探索する手法である。今回、この評価指標として、決定係数 (R^2 Score) を採用した。

3. 方位角推定の実験

3.1 実験のセッティング

3.1.1 実験の目的

モデルベースで解いた場合，モデルフリー学習 (SVR) で解いた場合の2通りについて，方位角の推定精度を検証するため．

本実験の様子を Fig. 9 にて示す．これは無響室での実験の様子である．

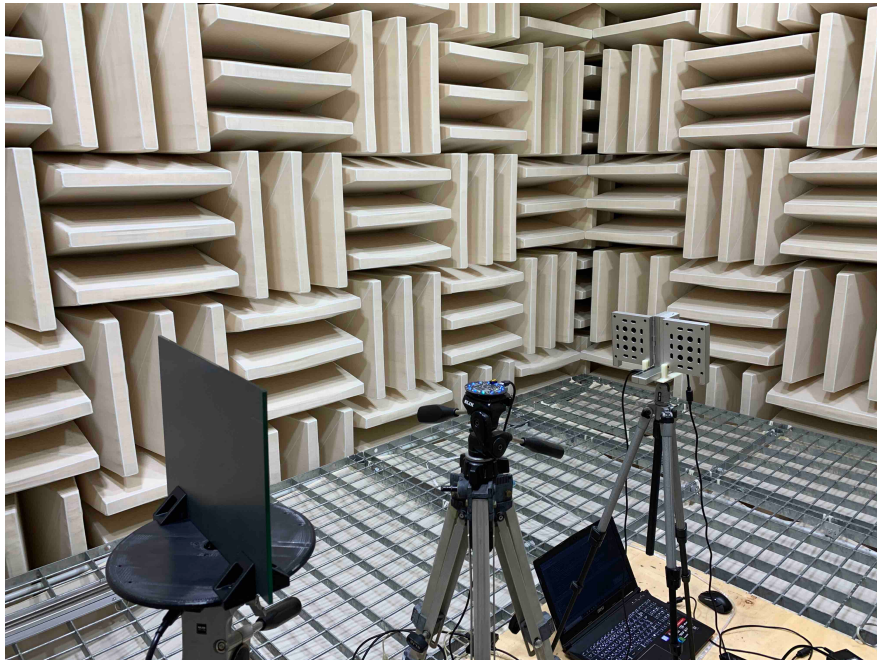


Fig. 9 Experiment for checking model



(a) Anechoic chamber (ORIST Izumi center [26]) (b) Real environment (NAIST S1 room)

Fig. 10 Experiment environment

3.1.2 実験環境

本実験は無響室と実環境の2通りで実験を行った。

無響室は、大阪産業技術研究所和泉センター [26] の無響室を使用した。実験環境の図を Fig. 10(a) に示す。

また、実環境として、奈良先端科学技術大学院大学情報科学棟 S1 教室を使用した。このとき、他の物体からの反射音を減らすために机や椅子といったものはすべて除外している。このときの環境の様子を図を Fig. 10(b) に示す。

3.1.3 実験器具

実験に用いた道具について Table 3 に示す。それぞれの実験器具について、詳細を以下で説明し、写真を Fig. 11 に示す。

第一に、マイクについて、4つのマイクロホンを搭載したマイクロホンアレイを用いた。理由は2点ある。1点目は Beamforming 技術を用いることで音源の到来方向を制限するためである。これはモデルベースで解く際にある1方向の音のデータを取る際に使用する。この Beamforming におけるマイクの個数の評価に

Table 3 Environment Equipment

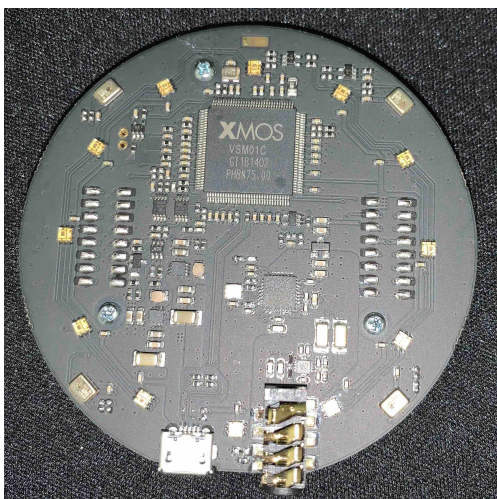
	Name (Company or Size)
Microphone	Respeaker Microphone Array V2 (Seed Inc.)
Speaker	Compact audio speaker SW5500-SS (FPS Inc.)
Turn table	OSMS-60YAW (SIGMAKOKI Inc.)
Target Object	Mirror (hight:23 [cm], length:32[cm])

については、後ほど述べる．2点目は1度に多くのデータを集めることができるためである．これはモデルフリー学習で解く際に、より効率的にデータを集めるために使用する．

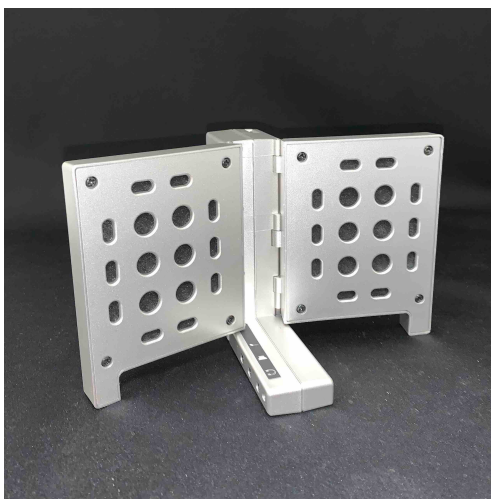
第二に、スピーカについて、平面波スピーカを使用している．一般的に、音は近傍界、遠方界によって波の形状が変化する．近傍界では球面波、遠方界では平面波となる．モデルベースで解く手法を用いる場合、音源の入射方向が一意に決まる必要があるため、平面波を用いることが必要である．

第三に、回転台について、平面の方位角との収録音の関係を取得するために使用する．本実験では、1[deg] ごとに変化させた実験データ収録する．今回使用した回転台は0.0001[deg] の位置決め精度で回転させることができる．そのため、今回欲しい平面の方位角を1[deg] ごとに-50[deg] ～ 50[deg] を取るには十分だと考えている．また、停止時の振動と移動時の雑音も少ないため、今回のような音を使った実験を行うのに最適だと考えた．さらに、本実験ではバックラッシュ誤差が少しでも出ないように1[deg] を連続で変化させるのではなく、10[deg] ずつ変化させ、データを集めた．

最後に、対象平面について、ガラス板の平面を方位角方向推定の対象とした．理由は2点ある．1点目は、ガラスは表面が非常になめらかであるため、拡散反射率がより理論値に近づくと考えた点である．2点目はカメラでの認識が難しく音波を使うことの有用性が考えられる点である．



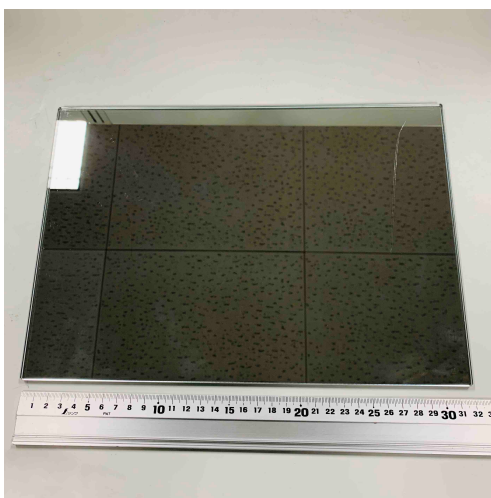
(a) Respeaker Microphone Array V2.0



(b) Compact audio speaker SW5500-SS



(c) Automatic rotating stage OSMS-60YAW



(d) Mirror plate (height:23cm, length:32cm)

Fig. 11 Experiment equipments

3.1.4 実験方法

スピーカから Time-Stretched Pulse(TSP 信号) を出力し，対象平面に反射してきた音をマイクロホンアレイにより取得する．今回，TSP 信号を使用した理由は2点ある．1点目は全周波数の情報を持っているためである．2点目はインパルス応答を見ることができ，伝達関数を作成可能という点である．この伝達関数を参考にするこことで，モデルベースで解く際に伝達関数と比較することで，評価ができると考えた．さらに，その中でも Up-TSP 信号を採用した理由は，周波数特性に対して外乱が少ないためである．

実験の簡略図を Fig. 12, 実験をした際のパラメータを Table 4, 実験の様子を Fig. 9 に示す．

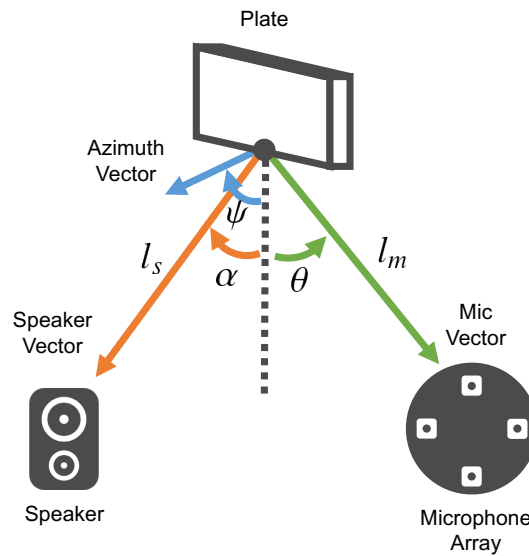


Fig. 12 Position and parameter for experiment checking model

3.2 モデルベース推定の結果

モデルベースでの方位角の推定はできなかった．

Table 4 Parameter for experiment checking model

Parameter	Value
$\alpha[\text{deg}]$	0
$\psi[\text{deg}]$	$-50 \sim 50$ (per 1 [deg])
$\theta[\text{deg}]$	0
$l_m[\text{m}]$	0.2
$l_s[\text{m}]$	0.4

この結果についての原因を明らかにするために、以下の3つ変化について解析を行う。

- 周波数に対するパワースペクトルの変化
- 方位角に対するパワースペクトルの変化
- 方位角に対する2周波数でのパワースペクトルの比の変化

3.2.1 周波数に対するパワースペクトルの変化

Fig. 13 はマイクで取得した音のフーリエ変換後の 1000 [Hz] から 2000 [Hz] のデータをプロットしたものである。全時間をフーリエ変換した。この結果の中でも $\psi = 0, \psi = 20$ の値を描画したものである。Fig. 5(a) と比較しても、周波数領域において、考えたモデルのような結果を見ることはできなかった。

3.2.2 方位角に対するパワースペクトルの変化

Fig. 14 は、ある周波数についての方位角と振幅の関係を見たものである。この結果より、モデル作成時の仮定としておいていた拡散反射と鏡面反射の割合によって、変化が見えるという予測に対して、1000[Hz], 2000[Hz] の比較と、Fig. 13 の結果も踏まえて考えると、周波数におけるモデルのときに考えた拡散反射、鏡面反射の割合の変化については考えることが難しいのでは無いかと考えた。しかし、一方で、ある周波数における方位角に対しての振幅の変化は連続的であり、

かつモデルで考えていた形に似ている形をしている．そのため，この点より，方位角の角度と反射音にはある一定の関係性があることがわかる．

3.2.3 方位角に対する 2 周波数でのパワースペクトルの比の変化

式 (10) に当てはめ，1000 [Hz], 2000[Hz] の値で割った結果を Fig. 15 に示す．予測モデルとは違った変化をする．この原因は，(1) の仮定について，これまでの結果より現状のデータでは当てはまらないことが大きいと考えられる．

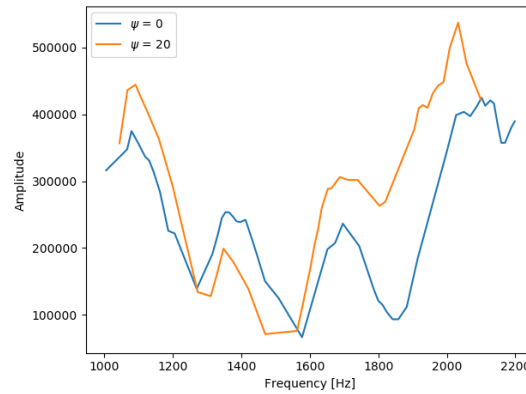


Fig. 13 relationship between energy (E) change with respect to frequency (f) and azimuth direction (N) when the sound source direction (S) and microphone direction (V)

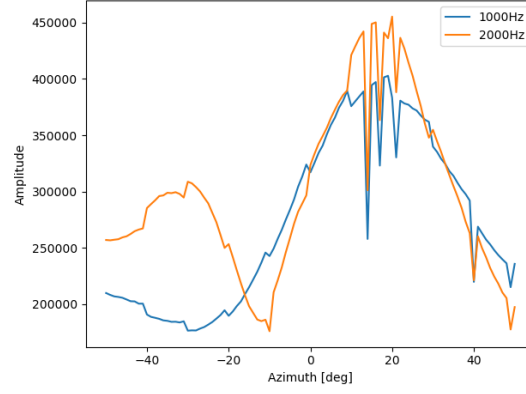


Fig. 14 Relation between energy ratio $D(\mathbf{N}|f_m, f_h)$ and azimuth angle $\mathbf{N}$ between mid and high frequencies at a point Q on the plane($f_m = 1000$, $f_h = 2000$)

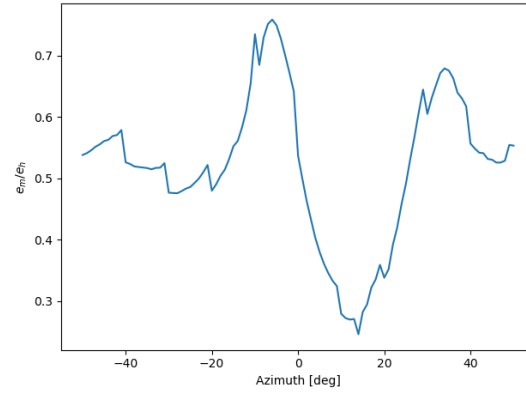


Fig. 15 Relation between energy ratio $D(\mathbf{N}|f_m, f_h)$ and azimuth angle $\mathbf{N}$ between mid and high frequencies at a point Q on the plane($f_m = 1000$, $f_h = 2000$)

3.3 モデル作成に関する考察

結果として、周波数領域における振幅の関係がなかった。しかし、方位角と振幅の関係は似たような関係を見ることができた。そのため、モデル自体を使うことはできないが、方位角と反射音には何かしらの連続的な関係があることがわかった。

まずは、理論通りの結果にならない原因を考える。原因の一つは直接音の影響であると考え。今取得している音波は完全に反射音のみを取得しているのではなく、マイクには直接音と反射音の合成音を取得されている。さらに、もう一つが反射音の干渉である。この2点について、理論で考えたモデルでは、直接音は考えておらず、反射音はある1点の反射音のみを考えている。そのため、直接音、他の点での反射音が干渉した音をマイクで取得していることになる。この原因より、周波数領域での変化というのは干渉の影響を非常に受けやすいため、この結果になったと考えている。

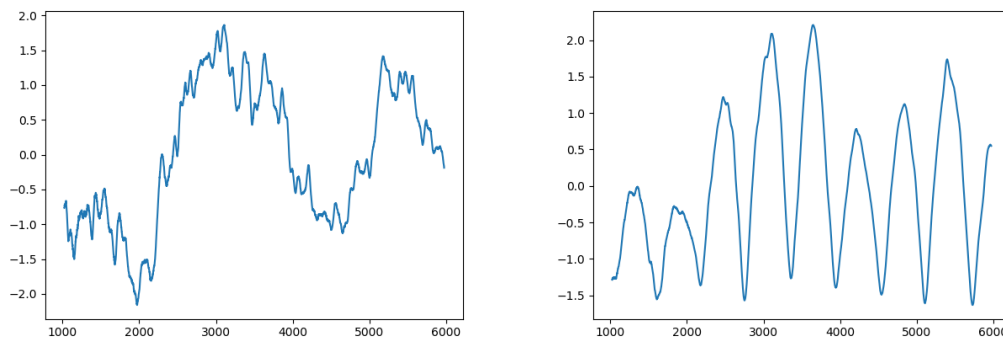
反射音の取得に関する対策はデータの改善である。方法は音声分離手法が考えられる。音声分離手法の中でも位置関係等が与えられるため、Beamformingを用いることが考えられるが、後の第3.5節で述べるが、結果としては、現在使用している4つのマイクロホンアレイでは空間分解能の精度が出ないため、難しいことがわかった。現状はDelay-Sum Beamformingを用いているが、MVDR Beamformingに代表される適応Beamformingについて今後検討をしていきたいと考えている。

また、モデルの改善についても考察していく。今回使用したモデルのなかでも特に、不安があるのが、音響学における鏡面反射と拡散反射の周波数の変化による割合の変化についてのモデルである。Fig. 13にも示したが、干渉の影響ももちろん考えられるが、同時に参考にしたモデルが壁といった十分に大きい物体に対しての反射率であった。そのため、今回のような大きさが限られているものに対しては適応が難しかったとも考えられる。

3.4 マイク，スピーカ，環境による影響の考察

本実験において，使用したスピーカ，マイク，無響室でのデータにおいて，対象物体がない場合で，スピーカとマイクのための環境で音を録音することで，本実験データについて，マイクとスピーカの特性における誤差の有無について検討を行う．以下に，スピーカ，マイクのための音源とスピーカ，マイク，対象物体の場合のそれぞれの周波数に対する振幅スペクトラムの変化を示す．

Fig. 16 より，スピーカマイクのための場合で，スピーカとマイクの周波数による感度の違いにより，周波数において，変化する．これが，反射面がある場合は，反射音の影響がはいるため，波形が変化することがわかった．しかし，この時，マイクで取得している波は直接音と反射音の合成音であると考えられる．これより，前節の考察のように，モデルベースで考える際は干渉の影響を考える必要があることの信用度が上がった．



(a) Amplitude change in frequency domain at environment no wall (b) Amplitude change in frequency domain

Fig. 16 Comparison of change of amplitude in frequency domain

3.5 モデルフリー学習 (SVR) の結果

第 3.2 節より，方位角と反射音について関係があることがわかった．まず，SVR を用いての結果を Fig. 17 に示す．このとき，横軸は実際の対象面の角度，縦軸は推定した対象面の角度である．つまり，正解の値はオレンジの線のような直線がかかることがわかる．そのため，求める結果としては，このオレンジの線にどれだけ沿っているのかを評価する．

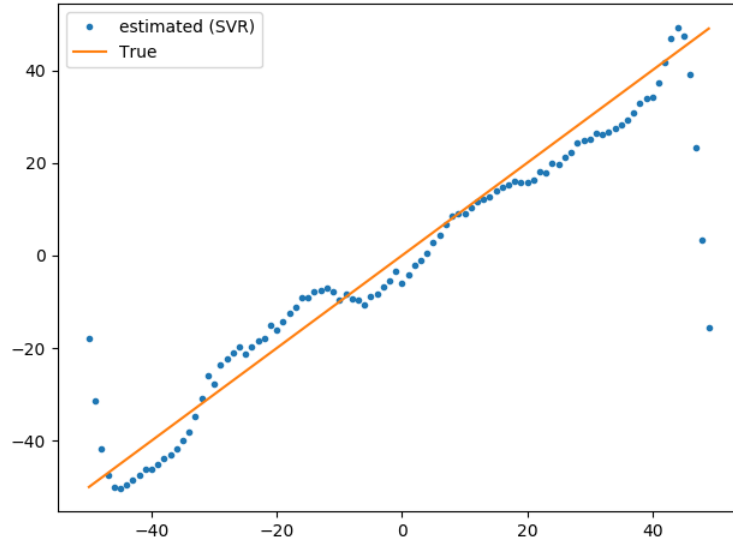
このときのそれぞれの決定係数 (Coefficient of determination, R^2) と二乗平均平方誤差 (Root Mean Squared Error, RMSE) で評価した結果を Table. 5 に示す．

以上の結果より，SVR を用いることによって，5 [deg] 以下の誤差で推定することができることがわかった．ここで，モデルベースでの結果について，考察をするために，現状の実験器具についてそれぞれ検証を行う．

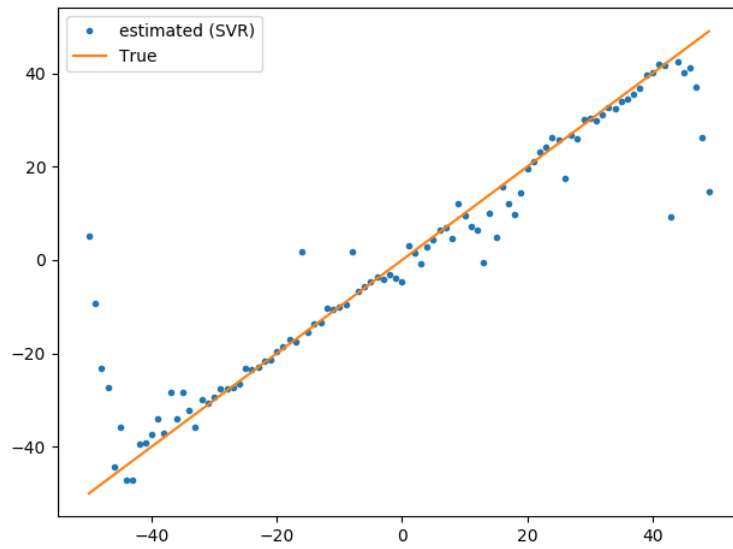
使用した音の種類，周波数の範囲，スピーカについて，この組み合わせが正しかったのかを考察すべく，第 3.5.1 節 ～ 第 3.5.3 節でそれぞれについて評価していく．

Table 5 Result of SVR

Place	R^2	RMSE [deg]
Anechoic	0.8	3.5
Real	0.7	4.0



(a) Result using SVR anechoic data



(b) Result using SVR real data

Fig. 17 Result of Experiment based on SVR

3.5.1 使う音の評価

今回、音波として、TSP 信号を用いている。このような、周波数成分を多く持ったデータを使うことが有効であるかどうかを確認する。

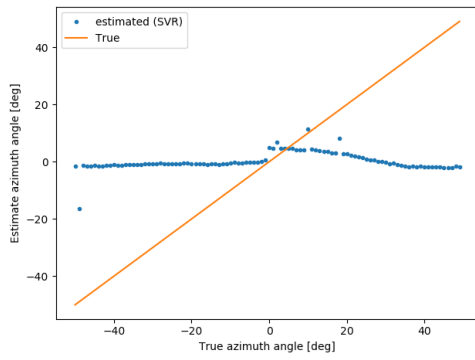
方法は、単一の高さの音を流した音の場合でも同様にデータを作成し、SVR をつかってモデル化したモデルを用いて、推定を行い、その二乗平均平方誤差を比較する。この結果を Fig. 18 に示す。

Fig. 18(a) ~ Fig. 18(c) より、単一の音を用いた場合は、SVR を用いても推定をすることはできていない。これは Fig. 14 からわかるが、単一の音を切り取った場合その音には線形な関係を見ることができない。そのため、角度を推定することができなかったと考えられる。

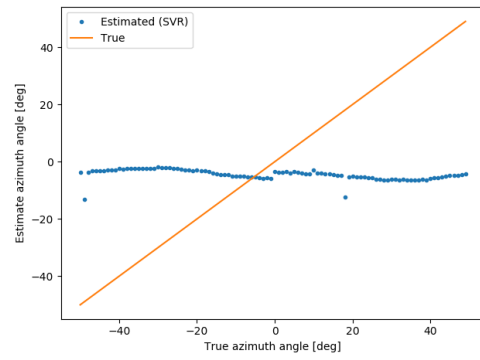
そこで、800, 1000, 2000 [Hz] での 3 つのデータを連結することで、3 つの周波数の情報をもつデータセットを作成した。そのデータセットを用いて同様に SVR のモデル作成し、テストデータで検証した結果を Fig. 18(d) に示す。

Fig. 18(d) のように、3 つのデータを連結したデータを用いても認識はできなかった。この結果より、周波数領域のデータの重要性が理解できる。

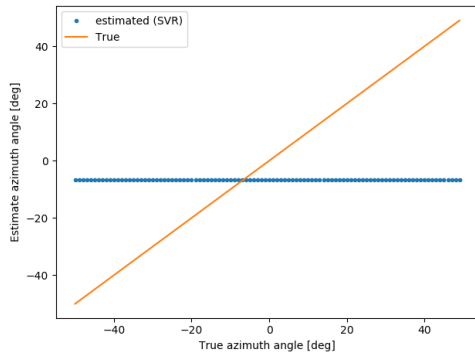
そのため、TSP 信号のように、周波数領域に関して、多くのデータを集めることができる信号を使うことはモデルベースで解く際にも有効であると考えられる。



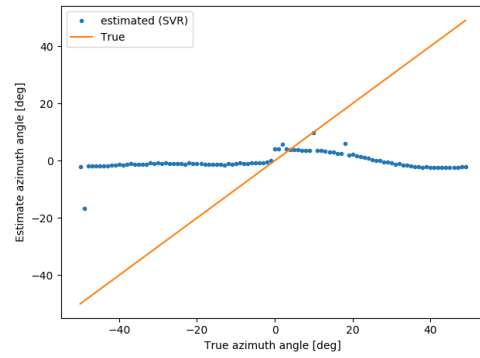
(a) 800Hz



(b) 1000Hz



(c) 2000Hz



(d) Result using torn signal 800, 1000, 2000Hz

Fig. 18 Comparison of result about sound kinds

3.5.2 使う周波数の範囲の評価

今回、使用する音の周波数として、音の反射の特性の変化を用いるために 1000 [Hz] から 2000[Hz] の音波を使用した。今回、このような音を用いることが有効であるかどうかを確認する。

方法として、1000[Hz] から 1000[Hz] おきにこれまでと同様の方式でデータセットを作成し、それぞれの決定係数、二乗平均平方誤差を比較する。結果をそれぞれ Fig. 19, Table 6 に示す。

Table 6, Fig. 19 より、0 ～ 1000 [Hz] と 6000 [Hz] 以上の音のデータに関して、著しく結果が悪いことがわかる。この原因としては、2 点考えられる。1 点目はスピーカの出力の周波数応答、2 点目はマイクの感度の周波数応答である。スピーカの周波数応答に関して、使用しているスピーカの周波数応答は 0 ～ 1000 [Hz] において、出力が低くなっている。この原因により、0 ～ 1000 [Hz] において、音が安定しないため、0 ～ 1000 [Hz] において、SVR の推定結果が悪くなったと考えられる。

次に、マイクの感度に関して、5000[Hz] 以上においてマイクの感度が上がっている。そのため、スピーカのとくと同様に、5000[Hz] で徐々に精度が下がり、6000[Hz] を超えると RMSE 10 [deg] になり、Fig. 19 で見た際のずれも明らかに大きくなっていると考えられる。

最後に、2000[Hz] ～ 4000[Hz] の音が使いやすく、その中でも Fig. 19 では、2000 ～ 3000 [Hz] がきれいに見え、Table 6 では、RMSE が一番小さいのは 3000 ～ 4000 [Hz] になる理由はこれらのスピーカの出力とマイクの感度の最大出力と最低出力のちょうど中間程度であるため、安定しているためであると考えられる。

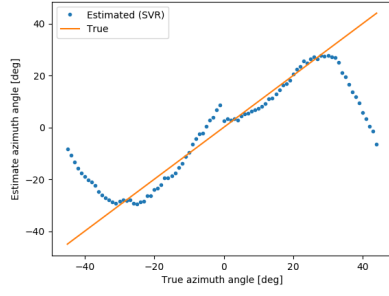
これらの結果より、今回検証したかった、1000 ～ 2000 [Hz] における反射特性の変化が重要であるという仮定について、本実験では検証できなかった。本実験では、マイクとスピーカの感度が安定する点であれば、モデルを作成することができることがわかった。

今後、モデルベースで解く際にはスピーカから出す音に関しては、スピーカとマイクの許容範囲ギリギリではなく、余裕を持った音の高さを用いることが有効であることがわかった。しかし、音の反射特性の変化を用いることが有効である

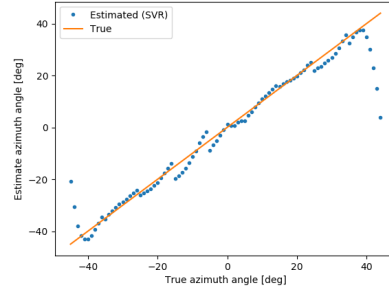
か否かについては本実験では，検証することができなかった．今後，違った方法で検証していこうと考えている．

Table 6 Comparison of R^2 , RMSE by change of frequency range

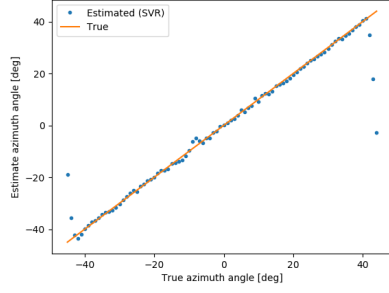
Frequency Range [Hz]	R^2	RMSE [deg]
0 ~ 1000	0.692	14.218
1000 ~ 2000	0.934	6.482
2000 ~ 3000	0.939	6.437
3000 ~ 4000	0.969	4.469
4000 ~ 5000	0.955	5.826
5000 ~ 6000	0.915	9.780
6000 ~ 7000	0.786	11.067
7000 ~ 7000	0.624	15.971



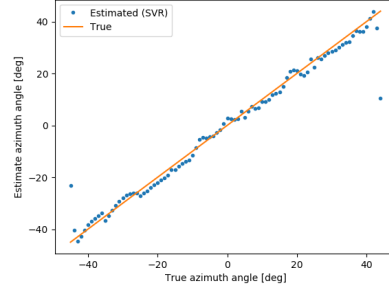
(a) 0-1000Hz



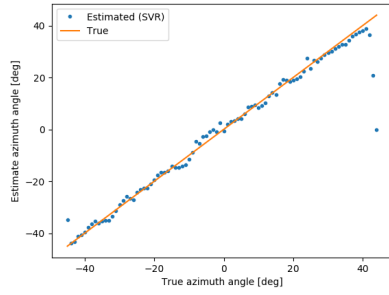
(b) 1000-2000Hz



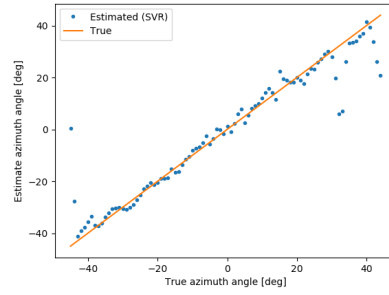
(c) 2000-3000Hz



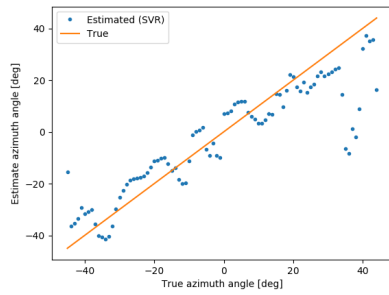
(d) 3000-4000Hz



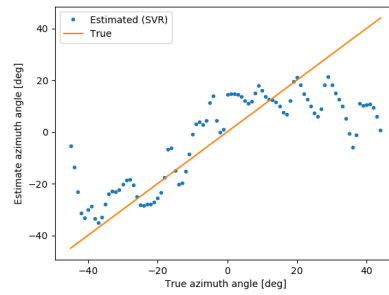
(e) 4000-5000Hz



(f) 5000-6000Hz



(g) 6000-7000Hz



(h) 7000-8000Hz

Fig. 19 Comparison of frequency range based on case that I create models individually

3.5.3 使うスピーカの評価

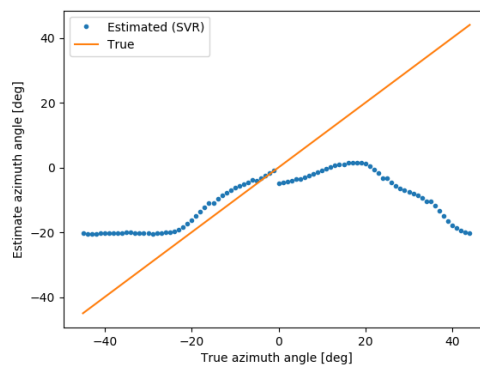
今回の実験では、平面波のスピーカを用いている。しかし、これが球面波スピーカの場合ではどうなるのかを確認する。

方法は球面波スピーカで発した音のデータを同様に作成したものでSVRをもちいてモデルを作成した。このときのスピーカはONKYO社のパワードスピーカシステム GX-D90を用いている。GX-D90について、使用した物の写真をFig. 20(a)に示す。今回、本来はステレオスピーカとして、2対でつかうものであるが、Fig. 20(a)のように、片方のみでしようした。理由としては、二つの方向から同時に音になることによる干渉を避けるためである。この実験における結果をFig. 20(b)に示す

この結果はFig. 20(b)となった。このときのRMSEは24.961 [deg]であり、球面波スピーカを用いた場合、これまでの平面波スピーカを使用した場合と比べて、SVRでの推定ができなくなっていることがわかる。以上の結果より、スピーカとして、平面波スピーカを用いたことはモデルベースで解く場合にも有効な手段であると考えられる。



(a) GX-D90 ONKYO Inc.



(b) Result of the data by spherical wave

Fig. 20 Experiment for speaker speculat speaker

3.6 SVR を用いた場合の考察

SVR を用いた場合、物体平面の面の方向を誤差 5 [deg] 以下で推定することができた。このように推定できた理由としては、物体平面の方位角とある周波数でのパワースペクトルの様子には連続的な関係があった。この連続的な関係については、Fig. 14 にてすでに示している。このような変化の特徴を使うことで SVR では推定モデルを作成することができたと考えている。

しかし、第 3.3 節でも述べたが、このような変化はモデルベースで解く際には考えていた変化とは違う。この原因として、直接音や他の点の反射音の影響が考えられ、対策として、音声分離手法の Beamforming を考えた。そこで、Delay-Sum Beamforming を用いて、対象物体の中心から反射してきた音を分離した。この分離音を用いて、SVR のモデルを作成し、推定した結果を Fig. 21 で示す。

Fig. 21 より、精度は Beamforming 前よりも低下した。この理由としては、2 点あると考えている。1 点目は Beamforming の精度が悪い。より、4 個のマイクアレイではほとんど分離が不可能であることがわかる。2 点目は、物体形状の影響が大きいという点である。物体形状の情報をわざと消している Beamforming を用いることでその情報が落ち、精度が落ちたと考えられる。

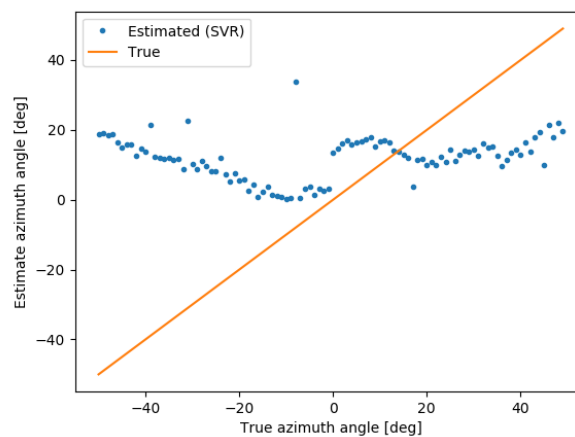


Fig. 21 SVR result using beamforming data

また、SVRにおいて、周波数に関係があるのか、どのような関係が見えるかを調べるために、主成分分析とパラメトリック固有空間法を用いて、検証を行った。その検証について、第 3.6.1 節で行う。

3.6.1 サポートベクターの解析

まず、主成分分析を用いて、第一、第二、第三主成分の周波数における固有値の変化を無響室でのデータ、現実環境でのデータそれぞれで比較する。結果を Fig. 22 に示す。

このように、両者は似たような変化をしていることがわかる。この変化にサポートベクターは着目しているように考えられる。

次に、パラメトリック固有空間法をもちいて解析を行う。この手法を用いるために、第一主成分から第三主成分までの主成分を 3 軸として 3 次元プロットを行う。このような方法は、実際のデータに今回のような物理的な変化がある際に用いることで、3 次元プロットした際にある規則的な動きをすることがあり、そこからモデル化の道筋を立てることができる。今回の場合、角度が -50 から 50 で回転しているため、回転のような連続的な動きが見れることが望ましい。無響室データ、実環境データそれぞれの結果を Fig. 23 に示し、このときの、主成分分析の寄与率を Table. 7 に示す。

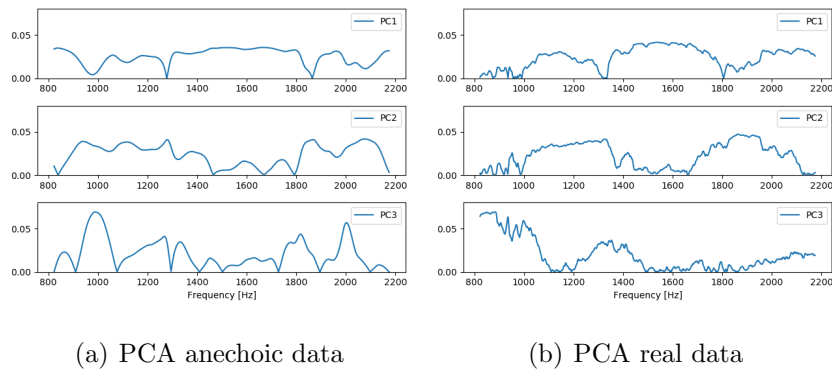
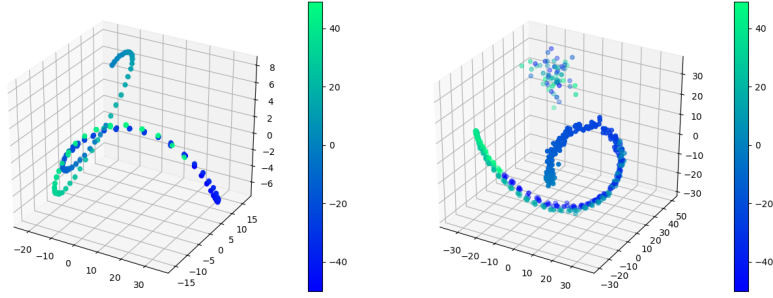


Fig. 22 Result of PCA



(a) Parametric eigen space method of anechoic data (b) Parametric eigen space method of real data

Fig. 23 Parametric eigen space method

Table 7 Contribution rate

PC	Anechoic	Real
PC1	0.73	0.52
PC2	0.18	0.21
PC3	0.03	0.09
SUM	0.91	0.82

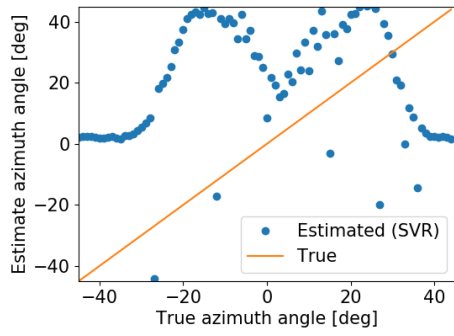
以上の結果より，取得した音データの中には方位角による連続的な関係があることはわかった．さらに，動きもなめらかに円を描くような結果を出力している．このような点より，方位角と反射音の周波数領域の大きさには連続的な関係が見ることはできる．このため，サポートベクターによる回帰が可能であったと考えられる．

3.6.2 汎化性能についての考察

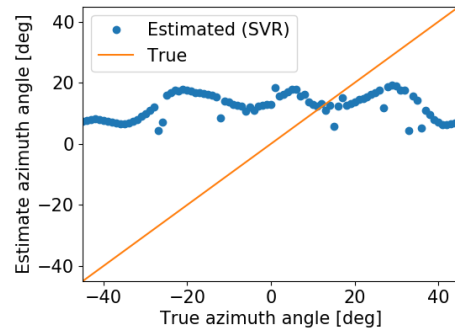
違う環境，違う位置でのデータを集めることによって，これまでデータを集めた．特に，違うスピーカ，マイク，対象平面の位置関係でのデータを統合したデータセットを作成し，学習を行うことで，未知の位置での方位角の推定に使用できるのかの検証を行った．

今回，使用するデータとしては，学習データとして，無響室でのデータを使用する．マイクと対象平面の距離 0.5m でのデータを学習データとして，使用することで，学習を行い，テストデータとして，マイクと対象平面の距離 0.2 [m] でのデータの推定を行った．この時，マイクロホンアレイを 1 のみつ使用 (mic1), 2 つのみ使用 (mic1, mic2), 3 つ使用 (mic1, mic2, mic3), 4 つ使用 (mic1, mic2, mic3, mic4) の 4 種類のモデルを作成する．それぞれのモデルを用いて，テストデータの方位角の推定を行った．

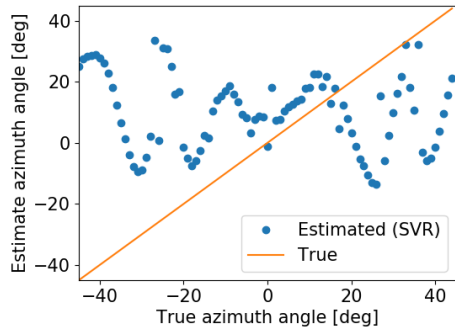
このときの，推定結果を Fig. 24 に示す．



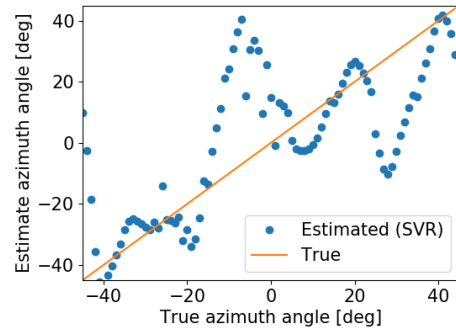
(a) Number of microphone is one



(b) Number of microphone is two



(c) Number of microphone is three



(d) Number of microphone is four

Fig. 24 Comparison of result about the difference of microphone

Table 8 Comparison of result about the difference of microphone at same environment

	RMSEs [deg]			
Mic number	Mic1	Mic2	Mic3	Mic4
One	3.2	20.4	31.6	31.6
Two	3.4	4.3	28.1	30.0
Three	3.4	4.4	3.5	19.4

同じ環境下で、距離を変えた場合の推定は精度が下がることがわかった。しかし、たしかに、マイクの数が増えるほどに正解の値により沿うようになっていように見える。この点より、たしかにマイクロホンアレイを用いることで、汎化性能を高めることができるのではないかと考えられる。

このマイクの位置を増やすことが汎化性能を高めるために有用であるかをより詳細に調べるために、マイクロホンアレイの4つの違う位置にあるマイクを用いて、同一環境のデータを用いて評価を行う。まず、先程同様にマイクロホンアレイを1のみつ使用 (mic1), 2つのみ使用 (mic1, mic2), 3つ使用 (mic1, mic2, mic3) の3種類のモデルを作成する。次に、それらのモデルに対して、学習データとして使用していない mic1, mic2m, mic3, mic4 のデータをテストデータとして、それぞれのモデルを用いて推定した際のそれぞれの RMSE をまとめた表を Table 8 に示す。

それぞれ、未知のマイクの位置の推定精度は非常に悪いことが表より分かる。しかし、mic4 という一切学習していない位置でのデータに対して、マイクの数が増えるごとに RMSE の誤差は小さくなっていることがわかる。

以上の結果より、今後マイクの数を増やすことや違う位置でのデータを増やすことによって、より対象物体、スピーカの位置関係にロバストなモデルを作成することができると思われる。

4. 材料，形状の違い，複数面への応用実験

4.1 実験目的

モデルフリー学習 (SVR) で解いた場合の2通りについて，方位角の推定精度について，以下の3点の実験条件の変化があった際にどのように変化するかを検証するため．

- 対象物体の材料
- 対象物体の大きさ
- 対象物体が複数

それぞれについて，調べる目的について，記述する．まず，材料について，音の反射率は対象物体の吸音率，拡散反射率は対象物体の表面の粗さにそれぞれ影響を受けると考えられる．そのため，2つの目的がある．1点目は材料が代わることによる方位角推定の可否の検証のため，2点目は他の材料で作った SVR モデルの方位角推定の可否を検証のためである．

次に，大きさについて，反射音の大きさは対象物体の幅，高さ，厚さに影響を受ける．特に，方位角推定の場合幅に関して，著しく関係性があると考えられる．そのため，3つの目的がある．1点目は大きさが代わることに寄る方位角推定の可否の検証のため，2点目は対象物体幅と方位角推定における関係性の検証のため，3点目は他の大きさに SVR モデルによる方位角推定の可否の検証のためである．

最後に，物体面の数が複数の場合について，これまで最も簡単な形状である面の推定はについて述べてきた．物体形状が明らかに違うためこれまでの SVR モデルによつての推定は難しいと考え，トレーニングデータから学習をした結果を使つての SVR による方位角推定による可否について検証を行うことを目的とする．

実験の様子について，Fig. 25 にて示す．



(a) Material is glass, Size is $200\text{mm} \times 400\text{mm}$ (b) Multiple plate (Material is cardboard, Size is $200\text{mm} \times 400\text{mm} \times 2\text{plate}$)

Fig. 25 Exsample of experiment environment for checking the changes in characteristics of material, size and plate number

Table 9 Three types of sizes to verify changes due to size

Size Name	Height [mm]	Length [mm]
Small	50	100
Middle	100	200
Middle	200	400

Table 10 Parameter for experiment checking the changes in characteristics of material, size and plate number

Parameter	Value
$\alpha[\text{deg}]$	0
$\psi[\text{deg}]$	$-45 \sim 45$ (per 1 [deg])
$\theta[\text{deg}]$	0
$l_m[\text{m}]$	0.2
$l_s[\text{m}]$	0.4

4.2 実験セッティング

実験環境は，奈良先端科学技術大学院大学情報科学領域 S1 教室において実験を行った．無響室ではなく，実環境で行う理由としては，第 3 章において，実環境でも結果がでることがわかったためである．

それぞれの実験について詳細を第 4.2.1 節，第 4.2.2 節で示す．

本実験の環境については，Table 10 のような環境で行った．Table 3 での変更点として， ψ を $-45 [\text{deg}] \sim 45 [\text{deg}]$ に， l_m を $0.1 [\text{m}]$ とした．まず，対象物体の方位角の変化の範囲 ψ の変更理由は，第 3 章にて， $-45 [\text{deg}]$ 以上の推定精度の誤差の増加について，反射音が返ってこないからだと考えた．本実験でも $-45 [\text{deg}]$ 以下， $45 [\text{deg}]$ 以上では反射音の取得が困難になると考え，省略した．次に，マイクと対象物体との距離 l_m の変更理由は，本実験で使用する平面が小さい場合に，距離 l_m 遠いとマイクから見た際の方位角の変化量が小さくなってしまい，方位角推定が難しいと考えたためである．

4.2.1 材料，大きさ実験

厚さを均一にしたガラス，ダンボールの 2 種類の材料に対して，それぞれ，Table 9 の 3 種類のサイズの平面板を用意した．厚さに関しては 3mm を選択した．また，今回使用したガラス，ダンボールの写真を Fig. 26 に示す．

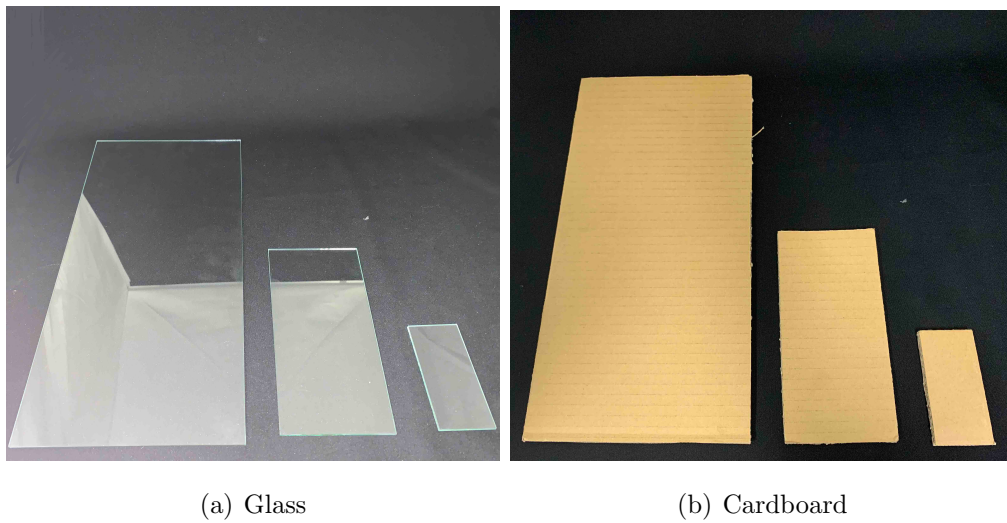


Fig. 26 Experiment Materials size: left is 200 [mm] $\times$ 400 [mm], middle is 100 [mm] $\times$ 200 [mm], right is 50 [mm] $\times$ 100[mm]

4.2.2 複数面での実験

対象物体としては，ダンボールの 200 [mm] $\times$ 400 [mm] を 2 枚用意し，実験を行った．この時，2 枚の平面の成す角が直角になるように対象物体を重ね合わせた．この 2 枚を重ね合わせた際にできる辺がマイクがある方向を正面とした際に，正面を向いている時，方位角 0 [deg] と考えた．

4.3 材料，大きさ実験の結果

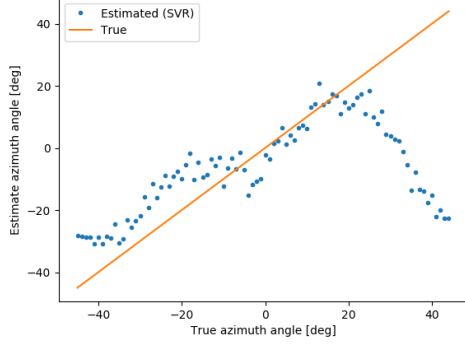
8回の実験データのうち，6回分のデータをもちいて，SVRのモデルをそれぞれ作成し，残り2回分のデータに対して，方位角の推定を試みた結果を Fig. 27 に示す．また，このときの決定係数 R^2 と二乗平均平方誤差 RMSE については Table 11 に示す．

まず，Table 11 より，特に RMSE に関して $-45 \leq \psi \leq 45$ の範囲で，目標としていた 5 [deg] 以下の精度を出すことができなかった．原因としては，35 [deg] 以上のところの誤差が非常に大きくなっている．この原因は実験時の回転台の角度のズレや，スピーカの方向のズレなどが考えられる．特に，スピーカのズレに関して，平面波を使用しているためズレによる影響が顕著に出たのではないかと考えうる．このズレをなくしていくためにもスピーカとマイクの固定を考える必要がある．

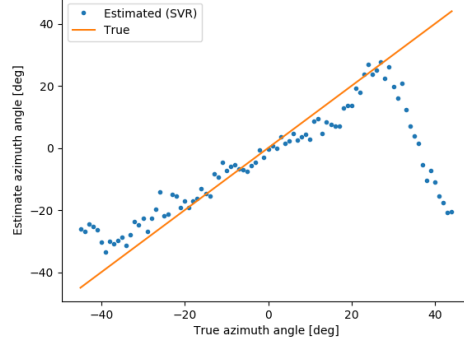
材料，形状に関する考察については，それぞれ，第 4.3.1 節，第 4.3.2 節にて述べる．

Table 11 Comparison of R^2 , RMSE by change of material and size

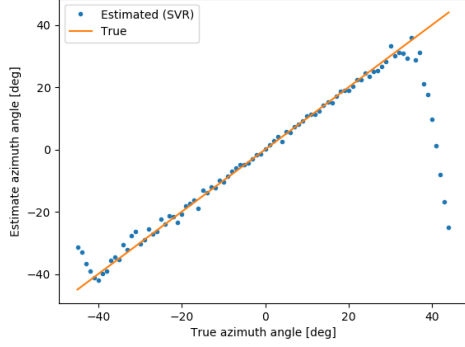
Material, Size	R^2	RMSE [deg]
Glass, 50mm × 100mm	0.234	22.737
Glass, 100mm × 200mm	0.486	18.616
Glass, 200mm × 400mm	0.740	13.236
Cardboard, 50mm × 100mm	0.431	19.604
Cardboard, 100mm × 200mm	0.594	16.560
Cardboard, 200mm × 4000mm	0.517	18.054



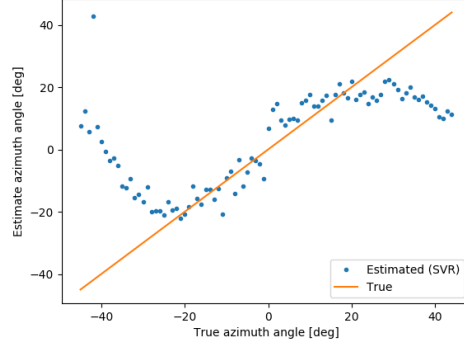
(a) Glass 50mm $\times$ 100mm



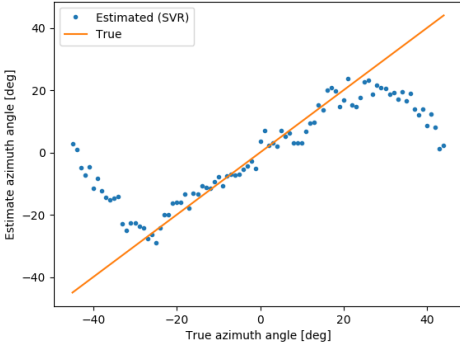
(b) Glass 100mm $\times$ 200mm



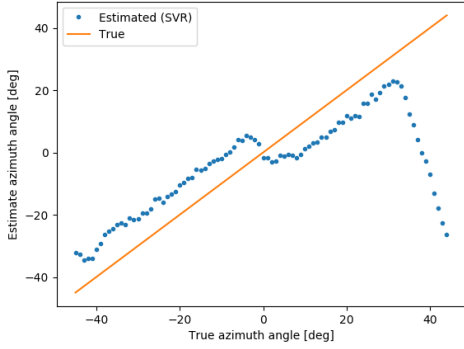
(c) Glass 200mm $\times$ 400mm



(d) Cardboard 50mm $\times$ 100mm



(e) Cardboard 100mm $\times$ 200mm



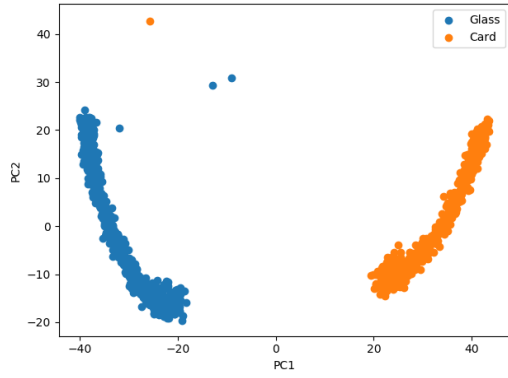
(f) Cardboard 200mm $\times$ 400mm

Fig. 27 Comparison of the changes in characteristics of material and size

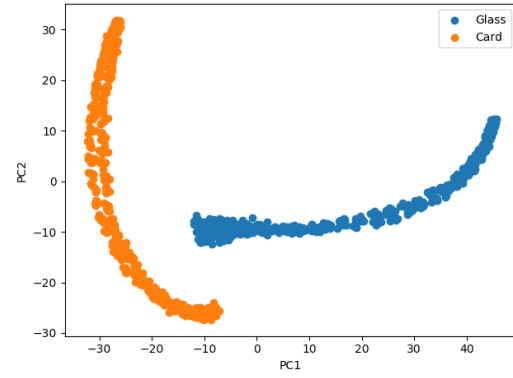
4.3.1 材料の違いによる変化の考察

ガラスとダンボールとを比べた際に、ガラスのほうが推定精度は高かった。これは、ガラスの表面が滑らかであるため、表面での音の反射が複雑にならないため、推定しやすかったと考えられる。反面、ダンボールのような表面が波を打つような凹凸があるような場合では、反射が複雑になってしまい、推定精度が落ちたと考えられる。また、このように、ガラスとダンボールでは異なった性質を持っているため、ガラスとダンボールの分類はできるのかを検証してみた。以下はそれぞれの大きさにおいて、ガラスとダンボールのデータを主成分分析した際の結果を Fig. 28 に示す。このように、材質による違いを認識することができる。そのため、このような周波数領域の音の大きさの変化のみを用いる場合は材質によって特性は変化するため、材質それぞれに対して、モデルを作成する必要がある。

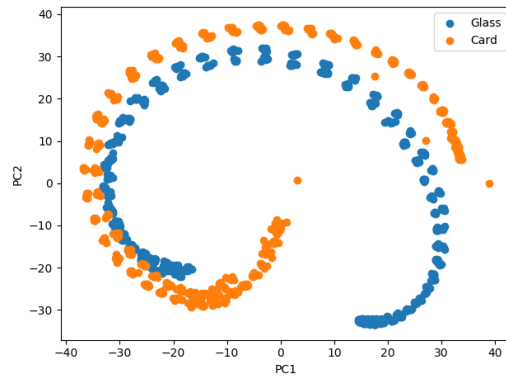
この問題点に対して、Photometric Stereo を参考にすることができるのではないかと考えている。Photometric Stereo では、光源方向が違えば、データを最低 3 方向使うことにより、法線ベクトルに加えて、その点の反射率を取得することができる。このような、手法を用いることによって、音を使って、材質の特性も含めた上で物体法線方向を推定することができると考えている。



(a) 50mm $\times$ 100mm



(b) 100mm $\times$ 200mm



(c) 200mm $\times$ 400mm

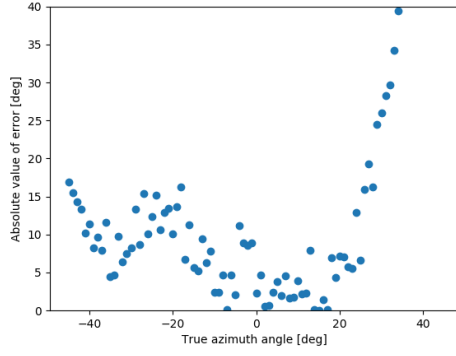
Fig. 28 Comparison of the changes in characteristics of material

4.3.2 大きさの違いによる変化の考察

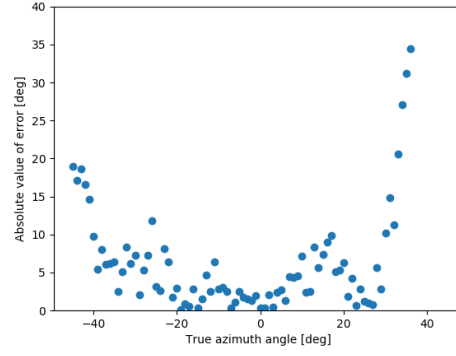
大きさについて、ガラス、ダンボールともに、大きさが大きくなるにつれて、推定精度が高くなることがわかる。これは第 3.6 節でも述べたが、スピーカからでた音が平面全体に対して反射した音をマイクで取得している。そのため、物体の大きさの影響を大きく受ける。さらに、この物体の大きさに関して、ある規則性があると考えられる。その検証のために、全体の誤差ではなく、 $-45 \leq \psi \leq 45$ の範囲の点それぞれの誤差を出力したものが Fig. 29 になる。このように、全体の誤差では単に精度が下がったように見えるが実際は認識できない角度が増加したと考えられる。これは、SVR でのモデル推定の際に使用している ϵ 誤差が関係していると考えられる。この SVR のモデルを作成した際の ϵ はグリッドサーチにて、最適な値を選んでいる。そのため、この ϵ について、それぞれのモデル作成の際の結果を Table 12 に示す。このように、小さくなるに連れて、 ϵ の値が小さくなる理由はマイクからの方位角方向の変化の幅が物体が小さいほど狭くなることが考えられる。例えば、0.1 [m] 離れた位置にある幅 0.4m の物体が $-45 \sim 45[\text{deg}]$ 回転した場合に反射音を取得することができる範囲を考える。平面からの反射音が最大するとき ($\psi = 0$ の時)、マイクから見た時第一反射音を取得できる範囲は $\tan^{-1}(0.2/0.1) = 63.4 [\text{deg}]$ の 2 倍で $126.9 [\text{deg}]$ である。また、これが反射音が最小の時 ($-45, 45[\text{deg}]$ の時) は $\tan^{-1}(0.1\sqrt{2}/0.2) = 35.3[\text{deg}]$ の 2 倍で $70.6 [\text{deg}]$ である。これに対して、幅が 0.05m の物体になると最大で、 $\tan^{-1}(0.05/0.1) = 26.6[\text{deg}]$ 、最小で、 $\tan^{-1}(0.05/0.1\sqrt{2}) = 19.5[\text{deg}]$ の 2 倍でそれぞれ、 $53.2[\text{deg}]$, $38.9[\text{deg}]$ となる。このように、幅、50 [mm] の物体の場合、幅、400 [mm] のものに比べて、取得できる反射音が著しく少なくなる。この時、対象物体の幅も重要ではあるが、同様にマイク、物体間の距離も重要になる。もし、マイク、物体間の距離を近づけることができれば、それだけ取得できる範囲は広がる。このように、対象物体距離に対しての依存関係があるため、もし、実際に同じモデルを使用する際は物体、マイクの距離は同じ条件にする必要がある。このように、物体の大きさも重要ではあるが、物体距離に関して、検証を増やしていきたいと考えている。

Table 12 Comparison of ϵ by changing the target object size

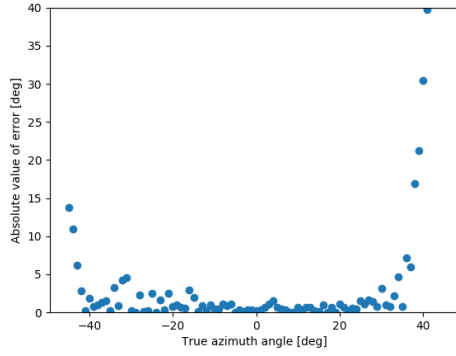
Material	Size (height,Length)	ϵ
Glass	100,200	10.0
	100,200	7.85
	200,400	0.10
Cardboard	100,200	0.33
	100,200	0.37
	200,400	10.0



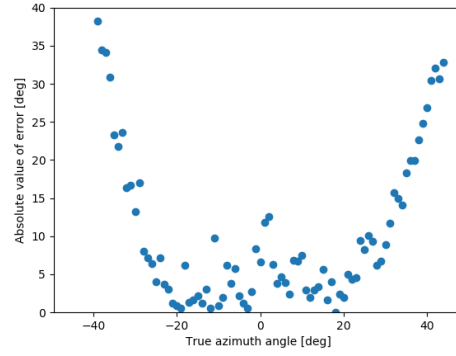
(a) Glass 50mm $\times$ 100mm



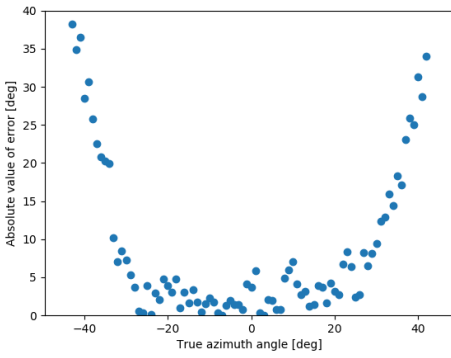
(b) Glass 100mm $\times$ 200mm



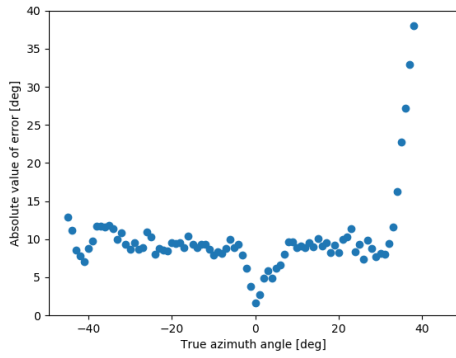
(c) Glass 200mm $\times$ 400mm



(d) Cardboard 50mm $\times$ 100mm



(e) Cardboard 100mm $\times$ 200mm



(f) Cardboard 200mm $\times$ 400mm

Fig. 29 Comparison of the changes in characteristics of size by the error

4.4 複数面での実験の結果と考察

8回の実験データのうち、6回分のデータをもちいて、SVRのモデルを作成し、残り2回分のデータに対して、方位角の推定を試みた結果をFig. 30に示す。また、このときの決定係数 R^2 は0.834、二乗平均平方誤差 RMSE は20.35 [deg] であった。

このように、複数平面であっても方位角の推定ができることがわかった。しかし、今回2枚の平面の要素となっている200 [mm] × 400 [mm] のダンボールで作成したSVRのモデルを用いて、2枚の平面のうちの1面の方位角推定を行った図をFig. 31に示す。Fig. 31に示すように、方位角推定は難しかった。しかし、傾きに関しては似たような傾きをしているようにも見える。難しい原因は、実験時のマイクとスピーカと対象物体の位置関係が違うことが考えられる。

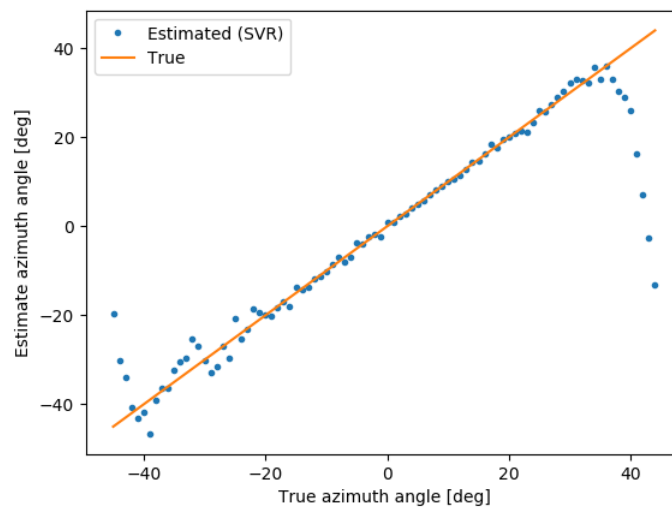


Fig. 30 SVR result of the data that target object is 2 cardboard plate

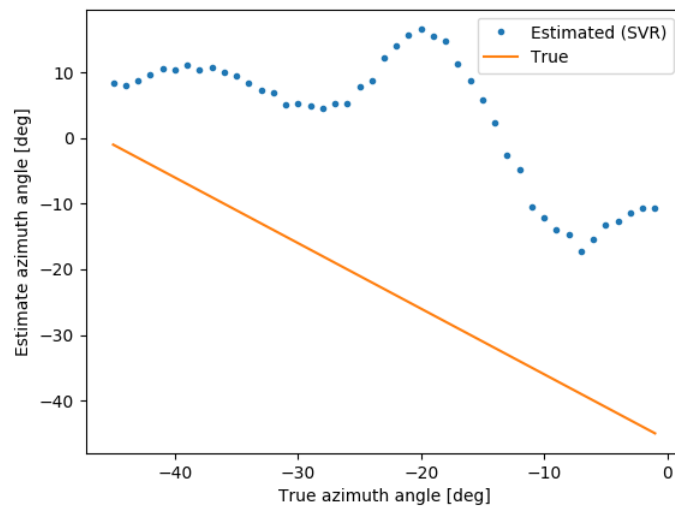


Fig. 31 SVR result of the data that target 1 plate from 2 cardboard plate data

5. 結言

5.1 まとめ

本論文では、ロボットでの把持が難しい透明物体や鏡といったカメラでは認識しにくく物体に対して、音波を用いることによって、形状を認識することで解決できるのではないかと考えた。しかし、これまでの時間領域の情報では空間分解能に限界があり、形状認識が難しい。そこで、光学における Photometric Stereo を参考にし、音の周波数領域の音の大きさに着目することで、法線方向を推定できるのではないかと考えた。

そこで、2通りの方法によって、まずは一番簡単な形状である平板の方位角の推定を行った。一つ目はモデルベースで考える手法、二つ目はサポートベクター回帰を用いる手法である。

結果として、モデルベースで解く手法は音の干渉の影響により推定自体できなかった。次にサポートベクター回帰を用いた場合は、反射音をマイクで取得できる範囲 (今回は方位角 -45 [deg] ~ 45 [deg]) の間で、誤差 5 [deg] 以内で推定ができた。この誤差 5 [deg] はロボットがものをつかむ際には許容できる範囲であると考えている。

また、本論文では、モデルベースで解けていない原因を考えるために SVR ではどういった部分に着目しているのか。どこまで SVR で認識が可能なのかを検証実験を行った。

結果として、平面波スピーカ、TSP 信号のような周波数領域に多くの特徴を持つ信号の2点を使うことの有用性について説明した。さらに、物体の大きさに対して、平面であれば物体の大きさと測定可能範囲に関して、関係性があることがわかった。これらより、ある環境における物体マイクの距離がわかった際に、その環境における物体のモデルがあれば、対象物体の方位角の推定ができることができる。

しかし、ある環境における対象物体のモデルのデータは必要となるため、このデータを集めることが必要になる問題がある。

5.2 今後の課題と展望

到達時間差を用いて距離を求める超音波距離センサでも面の方位角推定は可能である。モデルベースで解くことで初めて、実際の Photometric Stereo のように空間分解能を改善することができると考えている。今後はモデルの定式化に向けて、2点ほど解決策を考えている。

一点目は超音波を使用することである。今回は反射の法則が 1000 [Hz] - 2000 [Hz] で変わることを用いるために可聴域を使用した。しかし、可聴域よりも超音波のほうが指向性が高く、より光の波の性質に近い。超音波を用いて研究を行いたいと考えている。

二点目は超音波でもピンク TSP 信号を用いることである。これはこうもりが環境を認識するために用いている音だと知られている。この信号を用いることによって、周波数領域の情報も取得できる。

以上の2点の方法について、より今後知識を深め、Photometric Stereo で使用されている Lambert モデルといった光学の反射モデルのようなモデルをまずは音でも作成していきたいと考えています。

謝辞

本論文は著者が奈良先端科学技術大学院大学先端科学技術研究科先端科学技術専攻博士前期在学中にロボティクス研究室で行った研究をまとめたものである。

本論文の執筆にあたり、貴重なご指導とご助言、ご助力を賜りました小笠原 司教授に深く感謝いたします。論文執筆にあたり、丁寧なご検討、ご教示を頂きました副指導教員である本研究科の中村 哲 教授に御礼申し上げます。

論文執筆にあたり、丁寧なご検討、ご教示を頂きました副指導教員である本研究科の浦岡 行治 教授に御礼申し上げます。

本研究の遂行、方針決定、論文執筆などにおいて多くのご助言、ご指導いただきました本研究科の高松 淳 准教授に深く感謝いたします。

論文執筆、研究会での多くのご指導、ご助言を頂きました名古屋大学の丁 明客員准教授に深く感謝いたします。

論文執筆、研究会での多くの点においてご指導、ご助言を頂きました本研究科の GARCIA RICARDEZ Gustavo Alfonso 助教に深く感謝いたします。

本研究の遂行にあたり、学生生活や事務手続き等でサポートをいただきました本研究室秘書の大脇 美千代 氏、池田美輝 氏に深く感謝いたします。

本研究の進行にあたり、数多くのご助言、ご相談を頂きました本研究の博士後期課程学生の 袴田 有哉 氏、山崎 亘 氏、趙 崇貴 氏、湯口 彰重 氏、清川 拓哉 氏、白倉 尚貴 氏、URIGUEN ELJURI Pedro Miguel 氏に深く感謝いたします。

特に清川先輩には、研究の基礎となる方針や国内学会の論文執筆等で多大なご助言、ご相談をさせていただきました。この場を借りて深く御礼申し上げます。

本研究の遂行にあたり、実装面において多くのご相談やご助言を頂きました、同輩の永谷 智貴 氏に深く感謝いたします。

本研究室の先輩及び同輩の皆様、後輩達には研究を進める上で様々なご協力を頂きました。ここに深く御礼申し上げます。

最後に、学生生活を支えていただいた両親、家族、および友人達に深く感謝いたします。

参考文献

- [1] 佐久間哲哉: “壁面の拡散性の解析及び評価法 (〈小特集〉室内音響における最近の話題)”, 日本音響学会誌, vol. 61, no. 1, pp. 39–44, 2004.
- [2] “Omron 透明体検出光電センサー e3s-db”, <https://www.fa.omron.co.jp/products/family/3455/>.
- [3] Yichao Xu, Hajime Nagahara, Atsushi Shimada, and Rin-ichiro Taniguchi: “Transcut: Transparent object segmentation from a light-field image,” In IEEE International Conference on Computer Vision and Pattern Recognition (CVPR2016), pp. 3442–3450, Oct 2016.
- [4] T. Ugajin, K. Ohtani, and M. Baba: “Evaluation of recognition algorithms in ultrasonic object recognition,” In 2012 Proceedings of SICE Annual Conference (SICE), pp. 1252–1255, Aug 2012.
- [5] Wan Weiwei, Harada Kensuke, and Nagata Kazuyuki: “Assembly sequence planning for motion planning,” Assembly Automation, vol. 38, no. 2, pp. 195–206, 2018.
- [6] Mahler Jeffrey, Liang Jacky, Niyaz Sherdil, Laskey Michael, Doan Richard, Liu Xinyu, Aparicio Ojea † Juan, Goldberg Ken: “Dex-net 2.0: Deep learning to plan robust grasps with synthetic point clouds and analytic grasp metrics,” arXiv preprint arXiv, 2017.
- [7] Meriaudeau Fabrice, Rantson Rindra, Fofi David, and Stolz Christophe: “Review and comparison of non-conventional imaging systems for three-dimensional digitization of transparent objects,” Journal of Electronic Imaging, vol. 21, no. 2, pp. 1 – 6, 2012.
- [8] Ihrke Ivo, N. Kutulakos Kiriakos, P. A. Lensch Hendrik, Magnor Marcus, and Heidrich Wolfgang: “State of the art in transparent and specular object reconstruction,” 2008.

- [9] C. J. Phillips, K. G. Derpanis, and K. Daniilidis: “A novel stereoscopic cue for figure-ground segregation of semi-transparent objects,” In 2011 IEEE International Conference on Computer Vision Workshops (ICCV Workshops), pp. 1100–1107, Nov 2011.
- [10] Osadchy, Jacobs, and Ramamoorthi: “Using specularities for recognition,” In Proceedings Ninth IEEE International Conference on Computer Vision, pp. 1512–1519 vol.2, Oct 2003.
- [11] N. Roy, P. Newman, and S. Srinivasa. Recognition and Pose Estimation of Rigid Transparent Objects with a Kinect Sensor, pp. 273–280. MITP, 2013.
- [12] S. Kim, S. Oh, K. B. Kim, Y. Jung, H. Lim, and K. Cho: “Design of a bioinspired robotic hand: Magnetic synapse sensor integration for a robust remote tactile sensing,” In IEEE Robotics and Automation Letters, vol. 3, pp. 3545–3552, Oct 2018.
- [13] S. Wang, J. Wu, X. Sun, W. Yuan, W. T. Freeman, J. B. Tenenbaum, and E. H. Adelson: “3d shape perception from monocular vision, touch, and shape priors,” In IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS2018), pp. 1606–1613, Oct 2018.
- [14] O. Álvarez-Tuñón, Á. Rodríguez, A. Jardón, and C. Balaguer: “Underwater robot navigation for maintenance and inspection of flooded mine shafts,” In IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS2018), pp. 1482–1487, Oct 2018.
- [15] B. Lindell David, Wetzstein Gordon, and Koltun Vladlen: “Acoustic non-line-of-sight imaging,” In IEEE International Conference on Computer Vision and Pattern Recognition (CVPR2019), pp. 6780–6789, Oct 2019.
- [16] A. Das, I. Tashev, and S. Mohammed: “Ultrasound based gesture recognition,” In IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP2017), pp. 406–410, March 2017.

- [17] I. Dokmanić and I. Tashev: “Hardware and algorithms for ultrasonic depth imaging,” In IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP2014), pp. 6702–6706, May 2014.
- [18] Kozo Ohtani and Mitsuru Baba: “Shape recognition for transparent objects using ultrasonic sensor array,” In SICE Annual Conference 2007, pp. 1813–1818, Sep. 2007.
- [19] N. Yamaguchi, S. Hasegawa, K. Okada, and M. Inaba: “A gripper for object search and grasp through proximity sensing,” In 2018 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), pp. 1–9, Oct 2018.
- [20] Michael Oren and Shree K. Nayar: “Generalization of lambert’s reflectance model,” In SIGGRAPH 94, pp. 239–246. 1994.
- [21] Guanying Chen, Kai Han, Boxin Shi, Yasuyuki Matsushita, and Kwan-Yee K. Wong: “Self-calibrating deep photometric stereo networks,” In IEEE International Conference on Computer Vision and Pattern Recognition (CVPR2019), pp. 1813–1818, Sep. 2019.
- [22] 小山佳祐, 鈴木陽介, 明愛国, 下条誠: “指先に近接覚センサを備えたロボットハンドによるプリグラス制御”, 日本ロボット学会誌, vol. 33, no. 9, pp. 712–722, 2015.
- [23] 藤原恭司: “音場の拡散性と壁面の拡散反射性 (< 小特集); 室内音響設計の現状と課題”, 日本音響学会誌, vol. 53, no. 4, pp. 301–305, 1997.
- [24] Bui Phong: “Illumination for computer generated pictures,” Communications of The ACM-CACM, vol. 18, pp. 311–317, 1975.
- [25] Christopher M. Bishop: “Pattern Recognition and Machine Learning”. Information Science and Statistics, 2011.

- [26] “大阪産業技術研究所 和泉センター”, <http://www.omtri.or.jp/guide/equipment>.