

千葉工業大学  
学位論文

スマートホーム環境における  
発話者の位置・姿勢推定を用いた  
ロボットの音声操作

Voice operated robot using estimation of position  
and posture of speaking person in the Smart home environment

令和2年3月

所属学科 : 未来ロボティクス学科  
学生番号 : 16C1115  
氏名 : 松永達人  
指導教員 : 王志東教授

# 目 次

<b>1 序論</b>	<b>3</b>
1.1 研究背景 . . . . .	3
1.2 関連研究 . . . . .	4
1.3 研究目的 . . . . .	4
1.3.1 スマートホーム環境の利用 . . . . .	4
1.3.2 発話者の位置・姿勢推定 . . . . .	4
1.4 本論文の構成 . . . . .	5
<b>2 システム概要</b>	<b>6</b>
2.1 システム図 . . . . .	6
2.2 センシング部 . . . . .	7
2.3 制御部 . . . . .	7
2.4 ソフトウェア部 . . . . .	8
2.4.1 Robot Operating System (ROS) . . . . .	8
2.4.2 slackbot . . . . .	8
2.4.3 OpenPose . . . . .	9
2.4.4 AprilTag . . . . .	9
<b>3 ゴミ箱ロボット「TrashBot」</b>	<b>10</b>
3.1 制御部 . . . . .	11
3.2 電子回路 . . . . .	11
3.2.1 電源回路基板 . . . . .	12
3.2.2 Arduino シールド基板 . . . . .	12
<b>4 スマートスピーカーを用いたロボットの音声操作</b>	<b>13</b>
4.1 スマートスピーカーと Slack の通信 . . . . .	13
4.1.1 IFTTT を経由した通信 . . . . .	13
4.1.2 Slack を用いたテキストの取得 . . . . .	14
4.2 Slackbot で参照するテキスト . . . . .	14
<b>5 OpenPose を用いた発話者の位置・姿勢推定</b>	<b>15</b>
5.1 各キーポイントの座標推定 . . . . .	16
5.1.1 カメラ座標からフィールド座標を推定 . . . . .	16
5.1.2 カメラ座標系からロボット座標系に座標変換 . . . . .	17
5.2 ロボットがアプローチする半円領域 . . . . .	18
5.2.1 半円領域の定義 . . . . .	18
5.2.2 人間の中心座標の定義 . . . . .	19
5.2.3 人間の方向の定義 . . . . .	20
5.2.4 半円領域の半径の実測 . . . . .	21
5.3 手の動作に応じた領域選択 . . . . .	22

<b>6 結論</b>	<b>23</b>
6.1 まとめ	23
6.2 今後の課題	23
<b>7 参考文献</b>	<b>24</b>
<b>8 謝辞</b>	<b>24</b>

# 1 序論

## 1.1 研究背景

近年、IoT や AI の技術を用いた多様なデバイスの組み合わせにより、住居の利便性を提供するスマートホームという考えが進められている。また、ロボット家具や料理ロボットなど、人間の生活環境内で活動する自律型ロボットが登場している。これらのシステムを相互に利用することで、生活環境内の自律型ロボットと人間との協調活動の幅が広がると考えられる。本研究では、スマートホームの環境を利用して、発話者の位置・姿勢推定を用いた、ロボットの音声操作手法の提案を行う。



Fig. 1: Smart Home



Fig. 2: Autonomous robot in living environment

## 1.2 関連研究

マイクロホンアレイ(複数のマイクロホン)を用設置することで、発話者の位置を推定する研究がある。マイクロホンを複数設置する必要がありコストが高いという点や、発話者の位置は認識できるが姿勢は認識できないという点が課題として挙げられる。

## 1.3 研究目的

スマートホーム環境内における発話者の位置・姿勢推定を用いたロボットの音声操作手法の提案

### 1.3.1 スマートホーム環境の利用

本研究では、発話者の位置・姿勢推定とロボットの音声操作にスマートホーム環境を利用する。想定するスマートホーム環境は、ホームオートメーションで用いられるスマートスピーカーと、見守りシステムで用いられる屋内用のカメラのある環境とする。

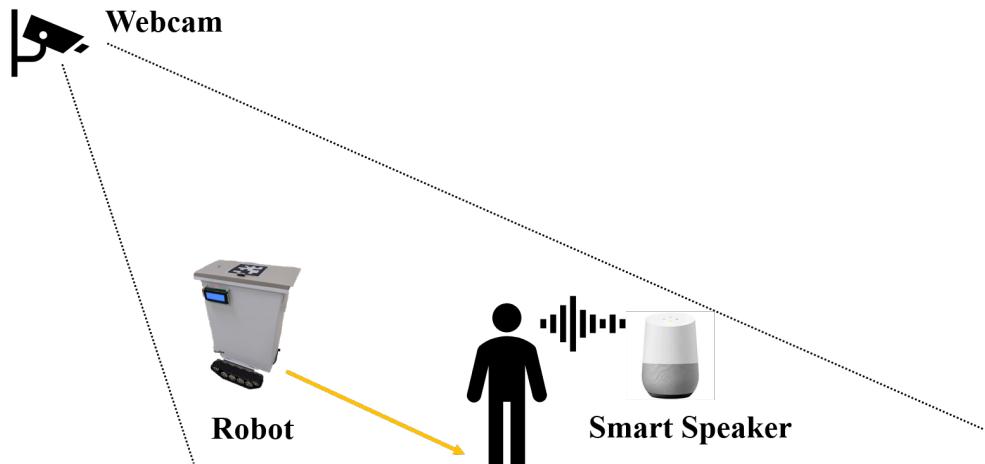


Fig. 3: Assumed smart home environment

### 1.3.2 発話者の位置・姿勢推定

発話者の位置・姿勢を認識することで、「こっち来て」など曖昧な言語表現を含む音声でロボットを操作。

## 1.4 本論文の構成

本論文の構成を以下に示す。第1章である本章では研究背景と目的について述べた。第2章では、本研究のシステム図を示し、システムの概要に関して、センシング部、制御部、ソフトウェア部の3つに分けて述べる。第3章では、第2章で述べたシステムを実装するにあたり製作した、ゴミ箱ロボット「Trash-Bot」の概要について述べる。第4章では、スマートスピーカーを用いたロボットの音声操作について述べる。第5章では、OpenPoseを用いた発話者の位置・姿勢推定について述べる。第6章では、本論文の結論として、本研究のまとめと今後の課題について述べる。

## 2 システム概要

### 2.1 システム図

スマートスピーカーから得た音声をテキストに変換し、そのテキストに応じてカメラ側からロボット側に指令を送信する。このとき、指令によって必要な人間とロボットの位置・姿勢の情報を送信する。通信にはROSを利用し、slackbotやOpenPose、AprilTagの各機能をノードとしてソフトウェアを組み、独立性の高い疎結合なシステムとした。本研究のシステムを図4に示す。

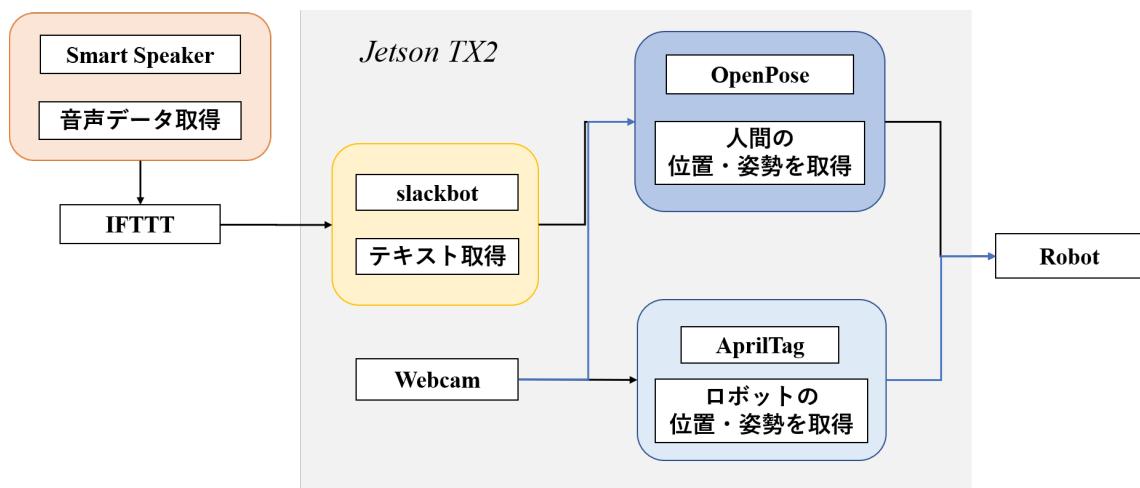


Fig. 4: System overview

## 2.2 センシング部

音声データの取得には、スマートスピーカー (GoogleHome) を使用。発話者とロボットの位置推定には、Web カメラ (Logicool c920r) を使用。



[1] GoogleHome



[2] Web カメラ

Fig. 5: Sensing devices

## 2.3 制御部

本システムの制御には Jetson TX2 (NVIDIA) を使用した。Web カメラと接続し、各プログラムを実行。ロボット側の制御ボードである RaspberryPi と SSH 接続で通信。



Fig. 6: Jetson TX2

## 2.4 ソフトウェア部

### 2.4.1 Robot Operating System (ROS)

ROS は Willow Garage 社によって開発されたオープンソースのミドルウェアである。各機能をノードとしてソフトウェアを組み、分散処理することでアルゴリズムごとの再利用性が増す。開発された各機能はオープンソースのパッケージとして提供されるため、柔軟に機能の追加や改良が可能である。本研究では、ROS のバージョンとして kinetic を利用している。



Fig. 7: ROS

### 2.4.2 slackbot

Slack (チームコミュニケーションツール) から特定のメッセージを取得し、処理を行う Python のライブラリである。Slack のワークスペースに Bot(自動返信システム)を作成し、Bot の API キーを slackbot とリンクして通信する。



Fig. 8: slack

### 2.4.3 OpenPose

OpenPose はカーネギーメロン大学 (CMU) で開発された、深層学習を用いて人物の骨格情報を可視化する手法である。Keypoint と呼ばれる人体における特徴点 (Pose Output Format BODY\_25 では 25箇所の特徴点) を認識し、それらの 2 次元座標を取得できる。

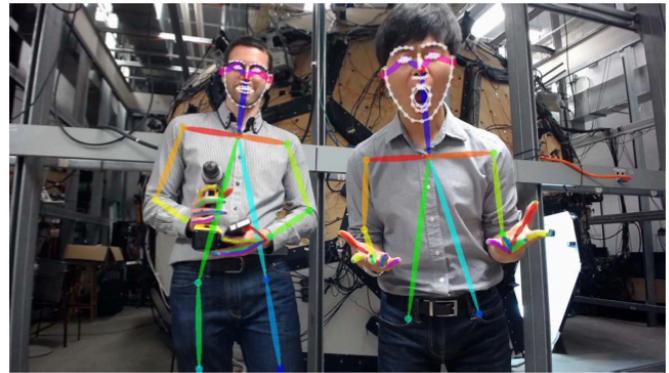
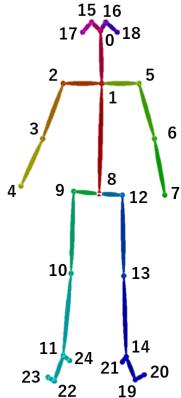


Fig. 9: OpenPose

### 2.4.4 AprilTag

AprilTag はミシガン大学 (UM) の AprilTags Visual Fiducial System プロジェクトで開発された、図 10 のような AR マーカーを用いてカメラに対する位置・姿勢を推定できるシステムである。



Figure 1: AprilTag 36h11 ID = 12



Fig. 10: AprilTag

### 3 ゴミ箱ロボット「TrashBot」

本研究のシステムを実装するにあたり、ゴミ箱ロボット「TrashBot」を製作した。蓋の開閉にはスライダクランク機構を用いた。走行部は、機体の安定性を考慮してクローラを用いた。蓋に設置した AprilTag を認識し、自己位置推定を行う。また、内部に設置した重量センサを用いて、ゴミの検知や容量の通知を行う。

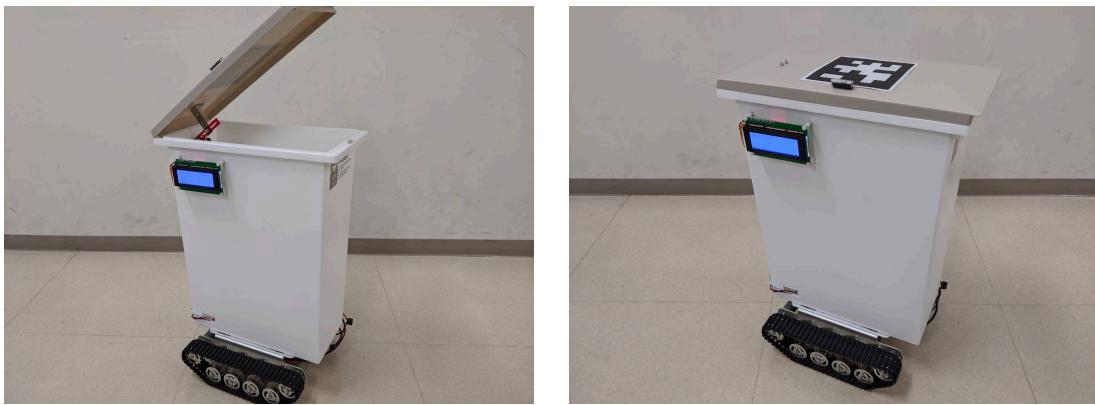


Fig. 11: TrashBot

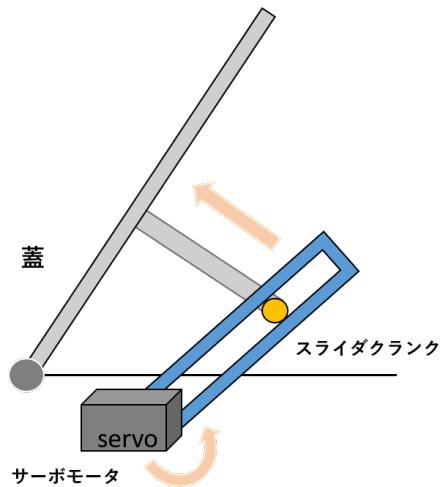


Table. 1: Robot size

Width	260 [mm]
Depth	365 [mm]
Height	530 [mm]
Capacity	20 [L]

Fig. 12: Slider crank mechanism

### 3.1 制御部

RaspberryPi 3B (OS: Ubuntu MATE), Arduino MEGA(互換ボード)を使用。  
rosserial を用いて接続し、ロボット側の RaspberryPi とカメラ側の Jetson は SSH で接続。



[1] RaspberryPi 3 Model B      [2] Arduino MEGA (compatible)

Fig. 13: Control board

### 3.2 電子回路

TrashBot の回路基板として、図 14 の電源回路基板(右)と Arduino のシールド基板(左)に分けて製作した。

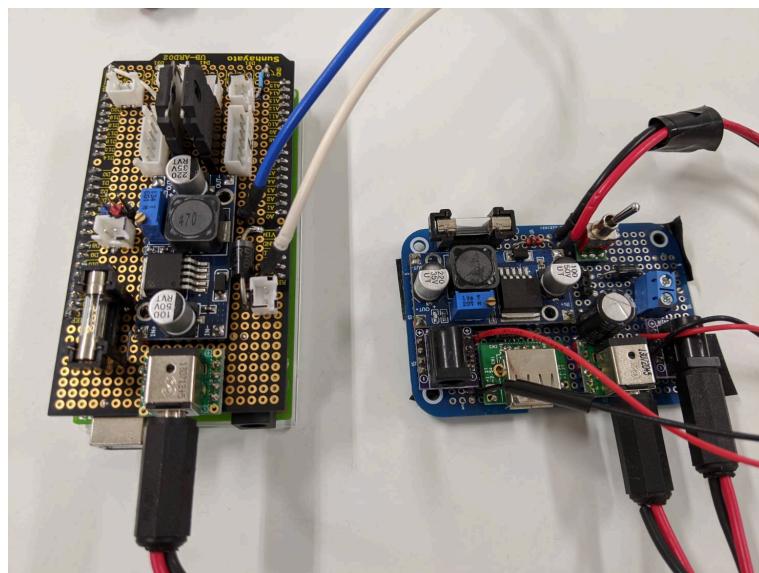


Fig. 14: Circuit board

### 3.2.1 電源回路基板

LiPo バッテリ (11.1V) を電源とし, Arduino と DC モータに電力を供給.  
また, DCDC コンバータで 5.0V に降圧し, RaspberryPi に電力供給.

### 3.2.2 Arduino シールド基板

走行部の DC モータは, モータドライバ (TB6643KQ) を用いて PWM 出力により駆動.  
ロードセルは, AD コンバータ (HX711) に接続して使用.  
LCD は, LiquidCrystal\_I2C.h (Arduino ライブラリ) を使用し, I2C 通信で接続.  
蓋の開閉用のサーボモータは, Servo.h (Arduino ライブラリ) を使用し, PWM で角度制御.  
各部品に接続した Arduino ピンは表 2 に記載した.

Table. 2: Connected Arduino pins

[1] Servo motor			[2] Load cell			[3] LCD		
赤	Vcc	7.4V	赤	Vcc	5V	赤	Vcc	5V
黒	GND	GND	黒	GND	GND	茶	GND	GND
白	Signal	D5(PWM)	茶	SCK	D36	黄	SCL	D21
			白	DT	D34	橙	SDA	D20

[4] DC motor with encoder

		left motor	right motor
赤	In1	D11(PWM)	D6(PWM)
黒	In2	D12(PWM)	D7(PWM)
緑	enA	D18(割込み)	D2(割込み)
橙	enB	D19(割込み)	D3(割込み)
白	Vcc	5V	
黄	GND	GND	

## 4 スマートスピーカーを用いたロボットの音声操作

スマートスピーカー (GoogleHome) とロボットの通信に Slack を利用する。スマートスピーカーに音声を入力する代わりに、Slack でテキストを打ち込んでも、同じ操作が可能。

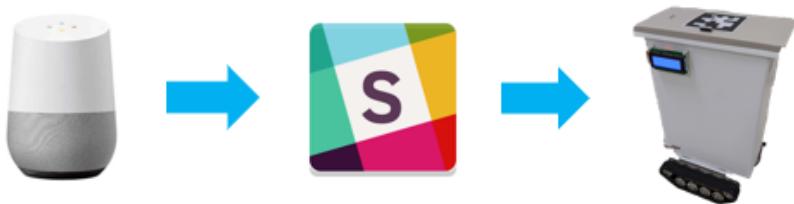


Fig. 15: Voice operated robot

### 4.1 スマートスピーカーと Slack の通信

#### 4.1.1 IFTTT を経由した通信

IFTTT (If This Then That) という、Web サービス同士を連携できる Web サービスを利用し、GoogleHome(Google Assistant) と Slack を通信。

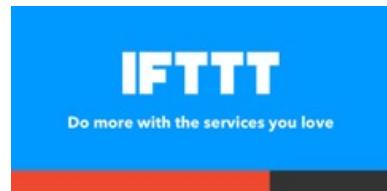
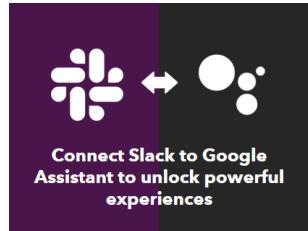


Fig. 16: IFTTT

#### 4.1.2 Slack を用いたテキストの取得

「ゴミ箱」というキーワードを含む音声をスマートスピーカーが取得すると、Slackにその音声のテキストを送信するように、IFTTTで設定する。



[1] Google Assistant to Slack



[2] Keyword

Fig. 17: IFTTT connection

#### 4.2 Slackbot で参照するテキスト

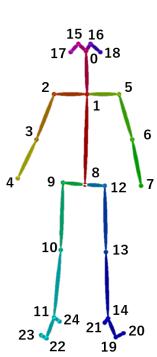
Slackで受け取ったテキストを、Web APIを利用してslackbotに送信する。slackbotで取得したテキストに応じて、表3のテキスト参照を行い、ROSの通信を利用してロボットに指令を送信する。

Table. 3: Reference text

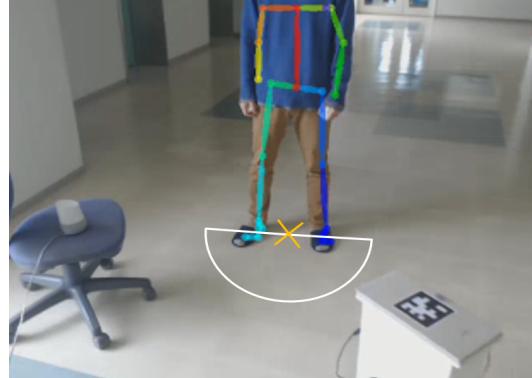
音声指令の例	参照テキスト	ゴミ箱ロボットの作動
ゴミ箱 <u>こ</u> っち來 <u>て</u>	來 <u>て</u> , <u>こ</u> っち	ゴミを入れる距離に近づく
ゴミ箱投 <u>げ</u> 入れモード	投 <u>げ</u>	ゴミを投げ入れる距離に近づく
ゴミ箱戻 <u>つ</u> って	戻 <u>つ</u> , 帰, 定位置	定位置に戻る
ゴミ箱センサモード	センサ, 重	センサが反応するまで蓋を開けて待機
ゴミ箱タイマモード	タイマ, 時間	一定時間が経過するまで蓋を開けて待機
ゴミ箱開 <u>け</u> て	開 <u>け</u> , 蓋開	蓋を開ける
ゴミ箱閉 <u>じ</u> て	閉 <u>じ</u> , 蓋閉	蓋を閉じる
ゴミ箱前 <u>行</u> って	前, フロント	前方移動
ゴミ箱後 <u>ろ</u> 行 <u>う</u> って	後, バック	後方移動
ゴミ箱右	右, ライト	右旋回
ゴミ箱左	左, レフト	左旋回

## 5 OpenPose を用いた発話者の位置・姿勢推定

「ゴミ箱こっちに来て」という音声指令を受け取ると、OpenPose を利用し取得する人体のキーポイント（各部 25箇所）より推定した、発話者を中心とする半円領域の円弧上を「こっち」として、ロボットがアプローチする。

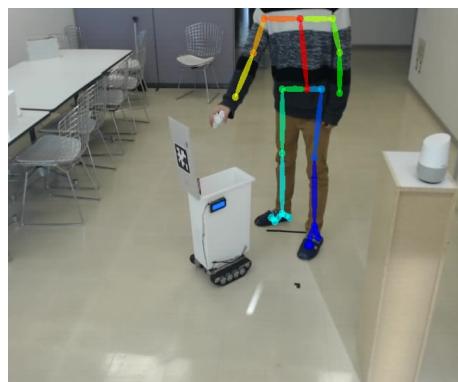


[1] Keypoint



[2] Semicircle area

Fig. 18: Keypoint and semicircle area



[1] Handing Mode



[2] Throwing Mode

Fig. 19: Approach

## 5.1 各キーポイントの座標推定

### 5.1.1 カメラ座標からフィールド座標を推定

画面上のキーポイント座標  $(x_i, y_i)$  から、カメラの焦点距離  $f$  を用いて、フィールド上のキーポイント座標  $(x_t, y_t)$  を推定。

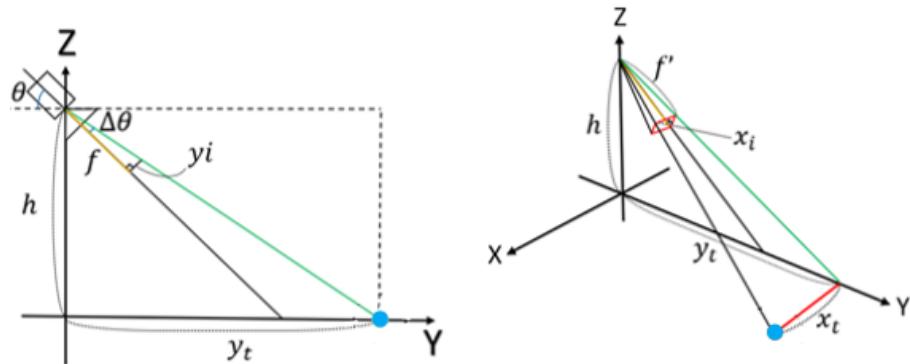


Fig. 20: Position estimation

$$\begin{aligned}
 \tan \Delta\theta &= \frac{y_i}{f} \\
 \Delta\theta &= \tan^{-1}\left(\frac{y_i}{f}\right) \\
 \frac{h}{y_t} &= \tan(\theta - \Delta\theta) \\
 f' &= \sqrt{f^2 + y_i^2}
 \end{aligned} \tag{1}$$

$$\begin{aligned}
 y_t &= \frac{h}{\tan(\theta - \tan^{-1}\left(\frac{y_i}{f}\right))} \\
 x_t &= \frac{x_i \sqrt{y_t^2 + h^2}}{f'}
 \end{aligned} \tag{2}$$

### 5.1.2 カメラ座標系からロボット座標系に座標変換

カメラ座標系におけるキーポイント座標  $(x_{key}, y_{key})$  とロボット座標  $(x_{robot}, y_{robot})$  から、式 4 の同次変換行列より、ロボット座標系におけるキーポイント座標  $(u_{key}, v_{key})$  を推定。

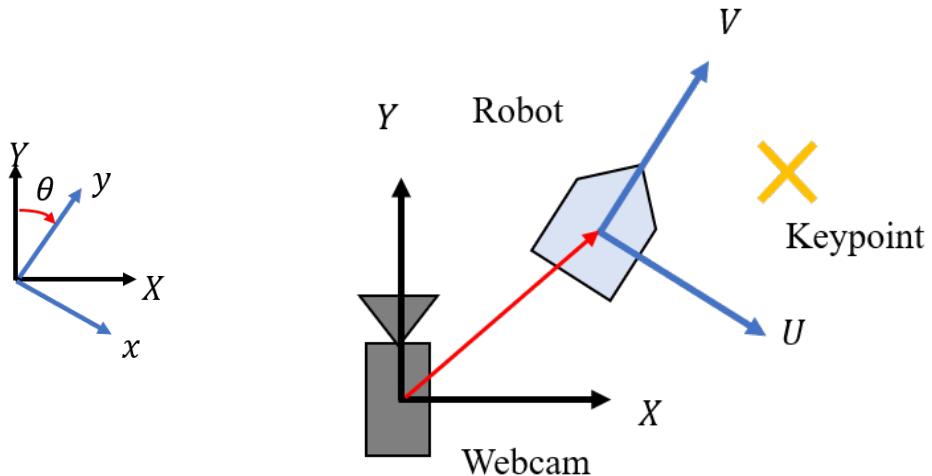


Fig. 21: Coordinate transformation

$${}^R\mathbf{p} = {}^R\mathbf{q}_C + {}^R\mathbf{R}_C(\theta) {}^C\mathbf{p} \quad (3)$$

$$\begin{bmatrix} u_{key} \\ v_{key} \\ 1 \end{bmatrix} = \begin{bmatrix} \cos \theta & \sin \theta & x_{robot} \\ -\sin \theta & \cos \theta & y_{robot} \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} x_{key} \\ y_{key} \\ 1 \end{bmatrix} \quad (4)$$

## 5.2 ロボットがアプローチする半円領域

### 5.2.1 半円領域の定義

発話者を中心とした、半径  $r$ 、式 5 の半円領域に、ロボットがアプローチする。

$$y = \sqrt{x^2 + r^2} \quad (5)$$

また、Handing Line と Throwing Line を以下のように定義した。

**Handing Line** ゴミを手渡しで入れる時の距離を半径とする円弧

**Throwing Line** ゴミを投げ入れる時の距離を半径とする円弧

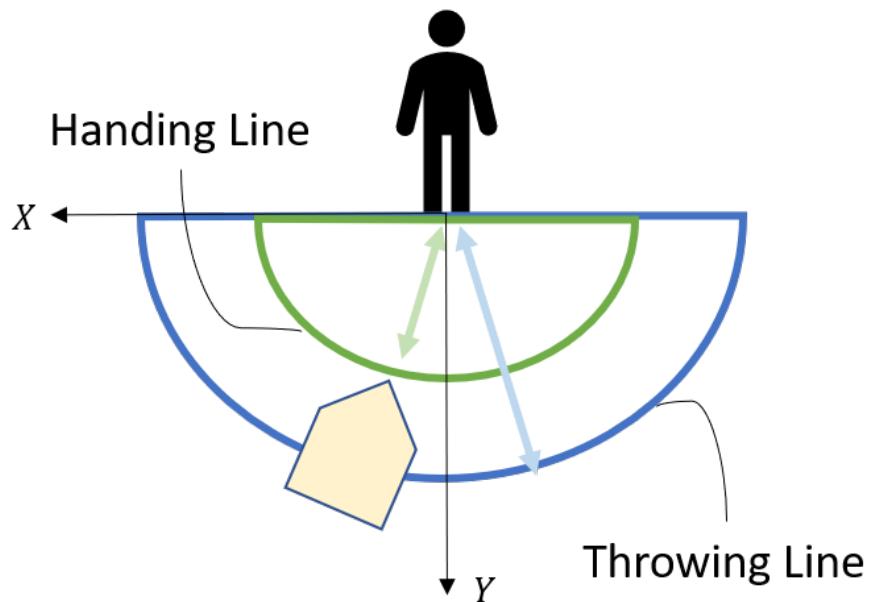


Fig. 22: Semicircle area

### 5.2.2 人間の中心座標の定義

人間の中心座標について、x 座標は腰の中心点の x 座標とし、y 座標は両足のつま先、かかとの 4 点の y 座標の平均値と定義した。z 方向の個人差を考慮し、z 成分が 0 となる座標を人間の中心座標に選択した。

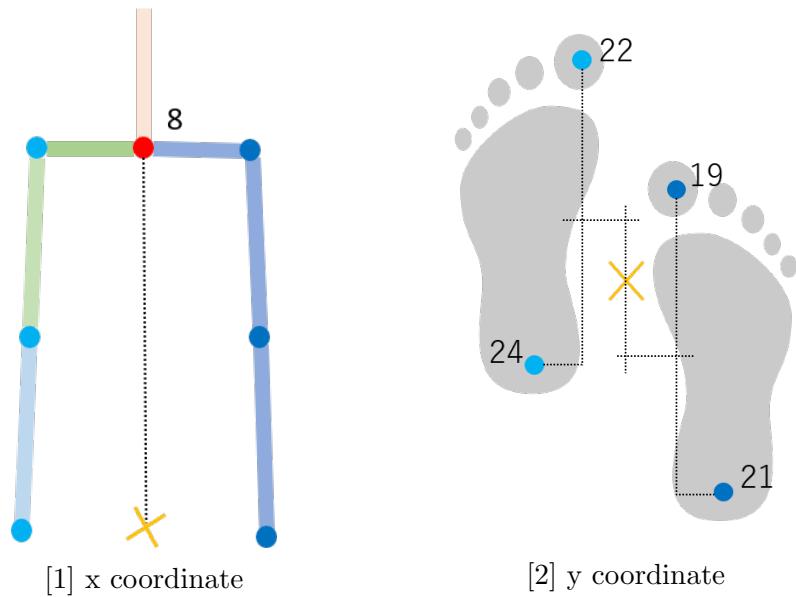


Fig. 23: Center coordinate of human

### 5.2.3 人間の方向の定義

人間の方向について、両かかとを結ぶ線分の中点を始点、両つま先を結ぶ線分の中点を終点としたベクトル  $\mathbf{h}$  と定義した。カメラ座標系の直行座標におけるベクトル  $\mathbf{h}$  の偏角  $\theta$  は、傾きの逆正接 (Python の atan2 関数) より、 $-\pi \leq \theta \leq \pi$  の範囲で求め、半円領域を偏角  $\theta$  の分、回転移動する。

$$\begin{aligned}\mathbf{h} &= (h_x, h_y) \\ \theta &= \tan^{-1}\left(\frac{h_y}{h_x}\right)\end{aligned}\tag{6}$$

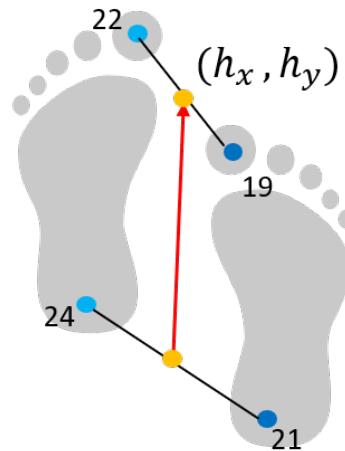


Fig. 24: Direction of human

#### 5.2.4 半円領域の半径の実測

ゴミを捨てる際の、人間とロボットの距離を測定し、半円領域の半径の実測実験を行った。ゴミを手渡しで捨てる場合と、ゴミを投げ入れて捨てる場合で、各10回ずつ測定し、その平均値を半円領域の半径とした。図25は実験の様子で、オレンジ色の点(OpenCVで出力)は人間の中心座標を示す。人間の中心座標が正しく認識できていない時のデータは除外した。



Fig. 25: Measurement Experiment

Table. 4: Distance between robot and human

[1] Handing Line		
回数	距離 [mm]	正誤
1	448.7	○
2	330.1	○
3	568.7	○
4	306.6	○
5	551.9	○
6	493.2	○
7	337.2	○
8	543.8	○
9	517.1	○
10	590.1	○
平均値	468.8 [mm]	

[2] Throwing Line		
回数	距離 [mm]	正誤
1	1796.7	×
2	1451.3	○
3	1533.7	○
4	1474.1	○
5	1428.2	○
6	1425.6	○
7	1473.3	○
8	1518.1	○
9	1569.0	○
10	1797.1	×
平均値	1546.7 [mm]	

### 5.3 手の動作に応じた領域選択

ゴミを持つ手を肘より上に持ち上げた場合を判断基準とし、肘関節の角度が  $90^\circ$  より小さいという条件(式7)を満たすとき、ロボットのアプローチする半円領域を、持ち手の四分円の領域に絞る。

$$\cos \phi = \frac{\mathbf{a} \cdot \mathbf{b}}{|\mathbf{a}||\mathbf{b}|} \geq 0 \quad (7)$$

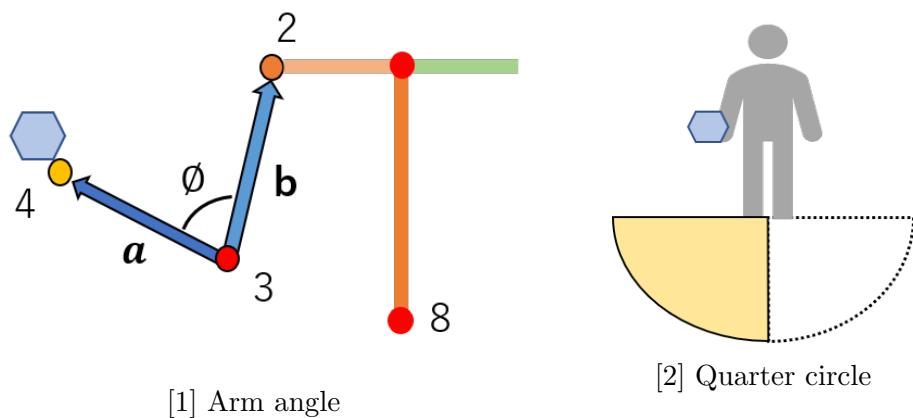


Fig. 26: Area limited by arm angle

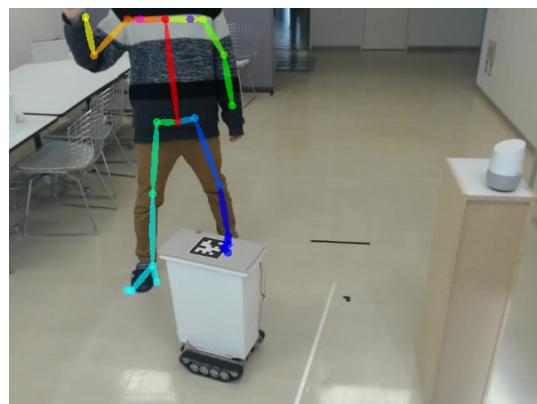


Fig. 27: Raise right hand

## 6 結論

### 6.1 まとめ

本研究では、製作したゴミ箱ロボットを使用し、スマートホーム環境における発話者の位置・姿勢推定を用いたロボットの音声操作を行った。

### 6.2 今後の課題

今後の課題としては、人間とロボットの双方向の応答や、複数人数の場合の発話者の推定が挙げられる。

## 7 参考文献

- [1] Zhe Cao, Gines Hidalgo, Tomas Simon, Shih-En Wei, and Yaser Sheikh, “OpenPose: Realtime Multi-Person 2D Pose Estimation using Part Affinity Fields ” , 2017
- [2] Edwin Olson, ”AprilTag: A robust and flexible visual fiducial system”, IEEE International Conference on Robotics and Automation, 2011
- [3] Carlos T.Ishi, Jani Even, and Norihiro Hagita, ”Evaluation of speech interval detection and face orientation estimation based on sound directions by multiple microphone arrays and human positions” , 2016

## 8 謝辞

本研究は、千葉工業大学未来ロボティクス学科において、王志東教授の御指導の元に行われたものです。常に懇切な御指導と御鞭撻を賜った王志東教授に対し、謹んで感謝の意を表します。大学生活4年間でロボット工学の基礎から様々な応用技術まで様々な講義を行って頂いた千葉工業大学未来ロボティクス学科の教員の皆様に対し謹んで感謝の意を表します。