卒業論文

ブレイン-オブ-ロボットの開発―
インタラクションによる人の嗜好の自動認識

指導教員	タン ジュークイ 教授
所　　属	九州工業大学工学部 機械知能工学科知能制御工学コース
学生番号	181A2093
氏　　名	中島　　万貴人
提出日	２０２２年2月１４日

摘要

近年，人工知能を搭載したコミュニケーションロボットをペットの代わりに家に置き，心の癒しやユーザーの健康管理として活用する人が増加している．また，国内のコミュニケーションロボットの市場動向として，2030 年には 900 万台普及と予測され，今後のロボット市場は成長していく見込みである．

そこで，本論文では AI コミュニケーションロボットの新しい機能として，ユーザーが日々の生活で食べる物を対象とし，物体認識および物体に対する味覚，好き嫌いなどの嗜好を判別するインタラクションシステムを提案する．本研究ではロボット上の音声アシスタントを作成し，会話の音声データからユーザーの嗜好と物体の情報を獲得して，ユーザー専用のデータベースを作成する．また，物体の識別器は，色情報と形状を特徴量とする Random Forest を用いて構築する．

本研究では，多クラス分類の精度を評価することにより，提案システムの有効性を確認した．

目 次

第 1 章　序論

シード・プランニングの調査によると，国内のコミュニケーションロボットの市場動向として，2030 年には 900 万台の普及が予測されている．また，家庭用で 10 万円以下の低価格帯機種は，450 万台程度の普及が予定されている．現在は約 30 万台の普及となっているが，コミュニケーションロボットの市場は今後さらに成長していく見込みである[1]．

現在，人工知能（AI）を搭載したコミュニケーションロボットの例として，富士ソフトのパルロが挙げられる．このロボットは，ユーザーとの会話の中でユーザーの行動や趣向などをデータとして記録を取り，データベースに蓄積していく．会話をすればするほどユーザーの情報を蓄え，ユーザーへの理解を深めていく．今後，AI コミュニケーションロボットが人に寄り添う思考を持つことができるようになり，ロボットと人が共に暮らせるという新しいライフスタイルが可能ではないだろうか．それを実現するため，本研究では，AI コミュニケーションロボットの新しい機能と考えられる，インタラクションによる人の嗜好認識および物体認識のシステムを提案する．

AI による物体認識は，事前に知識データベースを作成することが前提として考えられる．しかし，大量の物体データをデータベース化するのは手間を要する．従って，ユーザーが頻繁に用いる物体データのみを初めにデータベースに蓄積し，新しい知識を漸次加えて，ユーザー専用のデータベースを自動的に構築する手法を提案する．

本研究では，ユーザーが日々の生活でよく食べている果物を対象として，物体認識および物体に対する味覚，好き嫌いなどの嗜好を判別するシステムを開発する．本研究ではヴイストン株式会社のロボビーゼット（Robovie-z）を用いて，Raspberry Pi システム上の音声アシスタントのシステムを作成し，ロボットのカメラから得られる物体画像とマイクから得られる会話の音声データを獲得する．そして，物体画像の色情報を特徴量として，機械学習の一つである Random Forest を用いて物体認識を行う手法を提案する．また，得られた会話の音声データに対してテキスト変換や形態素解析などを行い，食べ物に対するユーザーの嗜好を分析し，嗜好を判別する手法を提案する．

本論文の構成について述べる．第２章では提案法の概要を説明する．第３章では会話解析のシステム，第４章では物体認識の手法，第５章では音声アシスタントについて述べる．第６章では本研究の実験の方法と結果，および実験結果に対する評価を記述し，第７章で考察を述べ，最後に第８章で結論を述べる．

第２章　研究の概要

本章では，本研究の全体の処理の流れと，使用するロボットの概要について述べる．

2.1　処理の流れ

本研究では，インタラクションによる画像認識と嗜好判別を行うことにより，人に寄り添う思考を持つロボットシステムを提案する[2]．図 2.1 に本研究で提案するシステムの流れを示す．まず，音声アシスタントのように，ユーザーはロボットに向けて発話する．発話した言葉をテキスト化し，キーワードを取得する．ロボットは取得されたキーワードに応じたタスクを実行する．

「記録」というタスクとは，ユーザーの回答を理解し，選別物の画像と選別物に対する嗜好を記録することである．ここで，物体識別のためには, Random Forest を用いて選別物の識別器の構築を行う．また，「認識」というタスクとは，選別物を識別し，識別結果選別物に対する嗜好を出力することである．最後，「味覚＋食べたい」というタスクとは，ユーザーの要求に応じて，味覚に該当する物を推奨することである．

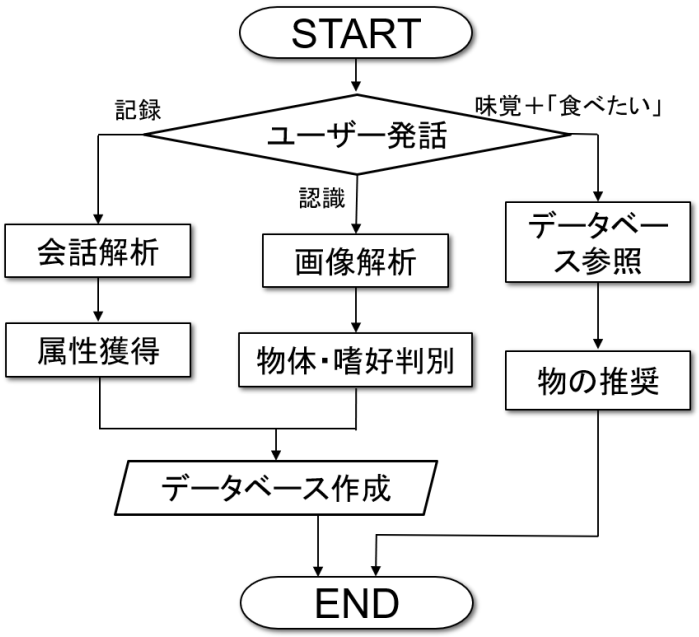


図2.1　提案システムの処理の流れ

2.2　ロボットの仕様

本研究で使用するロボットの外観を図 2.1 で示し，仕様を表 2.1 で示す．このロボットはヴイストン社が開発した Robovie-Z Raspberry Pi 版（以後 Robovie-Z）である[3]．Robovie-Z はサーボーモーター，ロボット制御基板，フレーム構造を採用し，メイン基板として Raspberry Pi 4 Model B を搭載した二足歩行ロボットのことである．また，ロボット内の基板については以降で説明する．

図2.2　Robovie-Z

表2.1 Robovie-Zの仕様

サイズ	(W)164×(D)110×(H)315[mm]
重量	約1020g
自由度	20軸（脚部：6軸×2 / 腕部：3軸×2 / 頭部：2軸）
搭載サーボモーター	脚部：VS-S055×12　その他：VS-S055C×8
電源	LiPo 7.4V1600mAh

2.2.1　Raspberry Pi 4

ロボットのCPUはラズベリーパイ財団のRaspberry Pi 4 Model B基板を用いる．外観を図2.3に示し，その仕様を表2.2で示す[4]．これはARMプロセッサを搭載したシングルボードコンピュータであり，ロボット本体内で画像処理やネットワーク連携などの高度な演算処理を行うことができる．

2.2.2 制御基板

ロボットにはサーボモーターを直接制御するロボット制御基板として，ヴイストン社のVS-RC026を用いる．これには3軸ジャイロセンサーや3軸加速度センサーが搭載されており，二足歩行ロボットとしての基本的な動作と制御を完給できるように設計されている．CPUから各モーションに対するコマンドを送信することによりロボットの制御が可能である．ここで，CPUと制御基板間の通信にはヴイストン社のRaspberry Pi 4専用拡張基板VS-RC019を使用する．

図2.3　Raspberry Pi 4 Model B基板

表2.2 Raspberry Pi 4 Model Bの仕様

OS	Raspbian OS
CPU	1.5GHz クアッドコア Cortex-A72
RAM	4.00GB

第３章　会話解析

本章では，ロボットとユーザーとの会話を解析し，ユーザーの嗜好をデータベースに変換するシステムについて述べる．処理の流れを図3.1で示し，会話内容の一例として図3.2で示す．

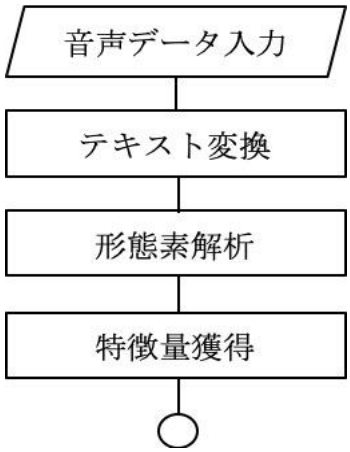


図3.1　会話解析の流れ

図3.2　ロボットとユーザーの会話内容の例

3.1 音声認識システム

ユーザーとロボットが会話を行う時，ロボットのマイクから得られるユーザーの声が音声データとして入力される．音声をテキストに変換するために，GoogleのCloud Speech-to-Text APIを用いる．このAPIを使用すると，Googleの音声認識技術をデベロッパーのアプリケーションに簡単に統合することが可能になり，音声をSpeech-to-Text APIサービスに送信し，文字変換されたテキストを受け取ることができる．また，ユーザーの回答を特徴量として獲得するために，文章の単語情報を認識する必要がある．本研究では工藤が開発したMecab[5]を用いて単語の分かち書きを行う．Mecabは与えられた日本語の文章に対応する全ての単語の組み合わせを単語辞書から抽出し，最小コスト法を用いて，もっともらしい組み合わせを出力する形態素解析エンジンである[6]．これを用いることにより入力された文章に含まれている単語とその品詞を調べることができる．

また，ロボットがユーザーに嗜好を尋ねるために，質問集を用意する必要がある．質問を以下に述べる．

（1）何を食べようとしていますか？
（2）どんな味がしますか？
（3）好きですか？

ユーザーは質問に対する答えを発話し，入力された音声はテキストに変換され，テキスト内の単語はMecabによって品詞が特定される．そして，質問に対して必要な品詞だけが特徴量として保存される．質問ごとに特定される品詞は以下の通りである．

（1）名詞
（2）形容詞
（3）名詞・感動詞

例を挙げると，質問（1）に対して抽出される品詞は名詞と特定され，ユーザーが「バナナを食べようとしている」と答えると「バナナ」が抽出される．また，質問（2）に対しては形容詞が特定され，ユーザーが「甘いです」と答えると「甘い」が抽出される．

3.2　嗜好情報の抽出

前節で述べた提案法により，ユーザーの食べ物に対する嗜好の回答は単語ごとに保存される．ロボットとユーザーが会話をすればするほど，回答の情報が多くなる．蓄積された回答を整理・分析するために，データの統計量を求める必要がある．そのために，辞書を作成する．質問２に対する味覚とそれに対応する名義

ラベル, 質問３に対する好き嫌いとそれに対応する順序ラベルをそれぞれ表3.1および表3.2に示す.

表3.1 味覚を表す辞書

味覚	名義ラベル
甘い	A
酸っぱい	B
苦い	C
渋い	D
水っぽい	E
甘酸っぱい	F
無味	G
ジューシー	H

表3.2　好き嫌いを表す辞書

好き嫌い	順序ラベル
大好き	4
好き,はい	3
普通	2
いいえ，嫌い	1
大嫌い	0

各味覚に対する食べ物の出現回数の統計を取り，出現回数が最も多い食べ物がその味覚に対応するお勧めのアイテムとして反映される．例を挙げると，A(甘い)で表現されたデータはバナナ５回，りんご２回，みかん０回の場合，バナナの回数が最も多いので，ユーザーが「甘いものが食べたい」と発話すると，答えがバナナになる．例の処理の詳細は第５章で述べる．

また，各食べ物に対する好き嫌いVを表すために，順序ラベルを用いる．好き嫌いの結果は次式で与えられる．

$$V \approx \frac{\sum_{j=0}^{4} j \times n_j}{\sum_{j=0}^{4} n_j} \tag{3.1}$$

ここで，jは順序ラベルの数値，n_jは$j(j = 0,1,2,3,4)$が現れた回数を表す.

第４章　画像認識

本章では，ユーザーの嗜好をよく理解するために，提案システムの一つの機能として，物体を画像から認識する手法を提案する．

ロボットが搭載しているカメラを用いて，物体の写真と物体が置かれていない状態の背景写真を撮影し，背景差分により物体領域を抽出した後，Random Forestを用いて物体の識別を行う[7]．特徴量として円形度，色相と彩度の情報を抽出する．

4.1　物体領域の抽出

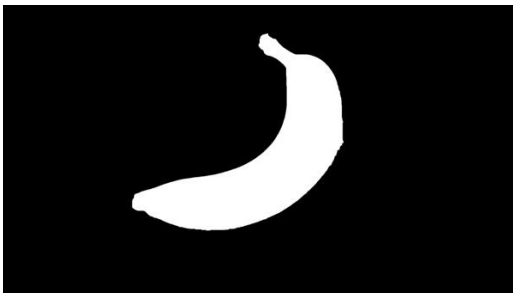
まず，ロボットのカメラから得られた背景画像と物体画像を入力する．物体画像の例を図4.1(a)に示す．物体の形状や色情報を抽出するために，背景領域の削除を行う．ここで，物体の影と光の明るさによる誤抽出を防ぐため，入力画像であるRGBカラー画像をRGB空間からHSV空間にチャンネル変換し，入力画像の色相と彩度と明度からなる３枚のグレースケール画像を生成する．生成された３枚の画像で背景差分を行い，それぞれ得られた差分画像I_dをチャンネル合成して，図4.1(b)に示すように物体のマスク画像を求める．次式により背景差分を行う．HSV変換の式は4. 2節で記述する．

$$I_d(i,j) = \begin{cases} 1 & |I_a(i,j) - I_b(i,j)| \leq T \\ 0 & |I_a(i,j) - I_b(i,j)| > T \end{cases} \tag{4.1}$$

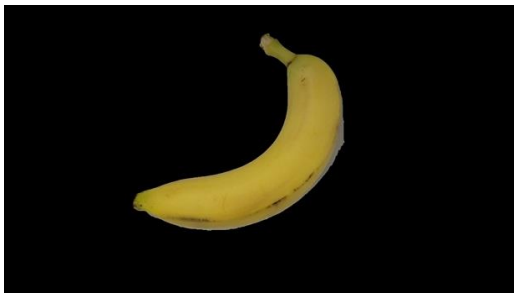
ここで，I_aは入力画像，I_bは背景画像であり，閾値Tは実験的に決定される値である．また，マスク画像に対して，8近傍のメディアンフィルタを用いて，膨張収縮処理を行うことにより，図4.1(c)に示すような物体のマスク画像が得られる．図4.1(d)は物体領域抽出画像である．

(a)

(b)



(c)



(d)

図4.1 物体領域の抽出
(a)物体画像，(b)マスク画像，(c)膨張収縮後画像，(d)物体領域抽出画像

4.2　特徴量抽出

物体の形状と色情報から特徴量を抽出するために，円形度の計算とHSV変換が必要である．本節では物体認識で用いられる特徴量について記述する．

4.2.1　円形度

円形度Rは，図形がどれだけ円に近いかを表す尺度で，面積をA，周囲長をPとして，次式で計算される．

$$R(A,P) = \frac{4\pi A}{P^2} \tag{4.2}$$

ここで，面積Aは図4.1に示すようにマスク画像の連結成分を構成する画素の数であり，周囲長Pはマスク画像の連結成分を輪郭追跡し一周する移動量である．

4. 2. 2　HSV 変換

色空間のRGBモデルは, 赤(Red), 緑(Green), 青(Blue)の３原色をそれぞれ0～255の範囲で表し，色を表現する．それに対して，HSVモデルは，色相(Hue)，彩度(Saturation)，明度(Value)という3つの値で色を表現する．Hは0～360の範囲で表現され，SとVはそれぞれ0～255の範囲で表現される．HSV変換を視覚化したモデルを図4.2に示す．変換式は次式で与えられる．

$$H = \begin{cases} 60 \times \dfrac{G - B}{MAX - MIN} & (MAX = R\text{のとき}) \\ 60 \times \dfrac{B - R}{MAX - MIN} + 120 & (MAX = G\text{のとき}) \\ 60 \times \dfrac{R - G}{MAX - MIN} + 240 & (MAX = B\text{のとき}) \end{cases} \tag{4.3}$$

$$S = MAX - MIN \tag{4.4}$$

$$V = MAX \tag{4.5}$$

ただし，RGB空間における赤，緑，青成分のうち最も大きい値をMAX，最も小さい値をMINとする．また，$H < 0$のときは，Hに360を加算して0～360の範囲に収める．MAX=0のときは$S = 0$，$H =$ 不定とする．

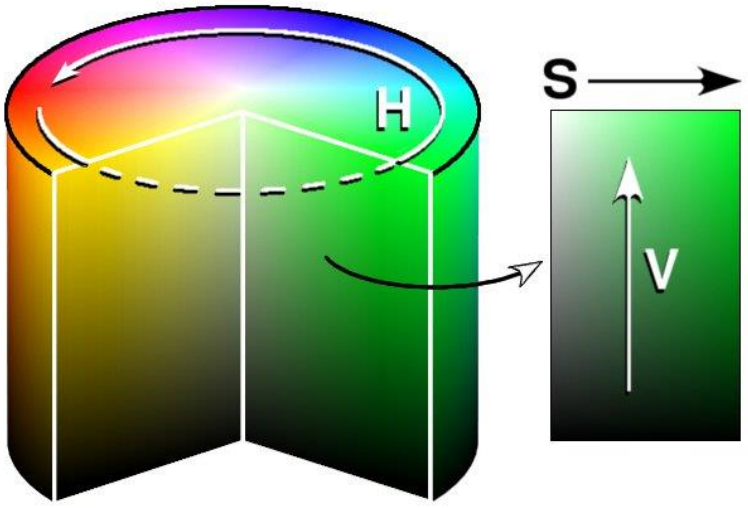


図 **4.2** HSV 空間の円柱モデル

4. 2. 3　色特徴量の抽出

物体の色特徴量を抽出するために，HSVモデルを利用する．明度(V)は照明の明るさに影響されるので，本研究では色相(H)と彩度(S)の2つの成分で処理を行う．

物体を撮影するときに，光の影響で物体に影が写ることがあるので，誤検出を防ぐために，特徴量は物体全体の色相と彩度の最頻値で表す．また，物体は多色の場合が多いと考えられるので，物体の色相ヒストグラムを作成し，各色成分の割合を特徴量として用いる．ただし，色相の値の範囲は本来0~255であるが，高速化のため0~31の32階調とする．図4.4に示すように，各度数の出現割合を特徴量として使用する．

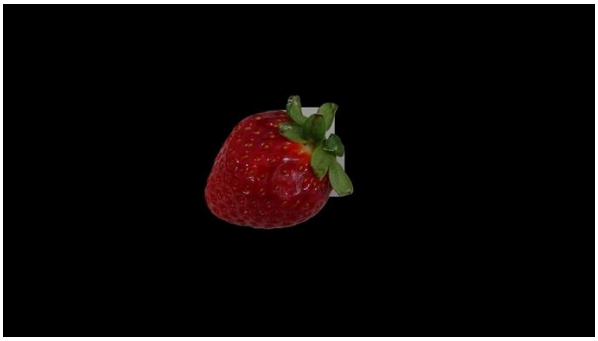


図4.3 いちごの物体領域抽出画像

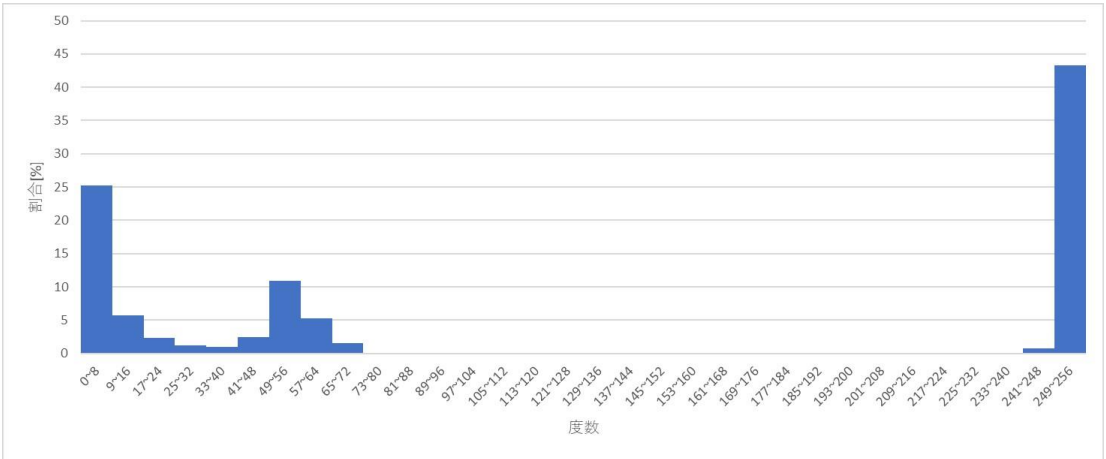


図4.4 図4.3の色相累積ヒストグラム

4.3　Random Forest

本研究では，識別器として多クラスの識別が可能であるRandom Forestを使用する．Random Forestとは図4.5に示すように複数の決定木でアンサンブル学習を行う手法であり，複数の学習器で学習することによって高精度な学習器を構築することができる[8]．また，学習データのランダム選択による影響を抑制することや特徴量の正規化や標準化が必要ではないことがこの手法の特徴である．

4.3.1　サブセットの作成

まず，学習サンプルからランダムにサンプルを選択して，T 個のサブセットを作成する．

次に，前節で述べた円形度と色情報の特徴量を用いて特徴ベクトルを $\boldsymbol{v_i}$ とし，教師信号 c_j を付与したサンプル集合を用意する．そのサンプルをランダムに選択して T 個のサブセット t を作成する．サブセットは，学習サンプルから重複を許してランダムに選択する．

その後，図4.5のように各サブセットから決定木を構築する．サブセット１つにつき，１つの決定木が作成されるので，T個の決定木が得られる．

各決定木は，分岐ノードと末端ノードによって構成され，分岐ノードを反復作成して，一定基準により分岐が不可能になったときに，末端ノードが構築される．

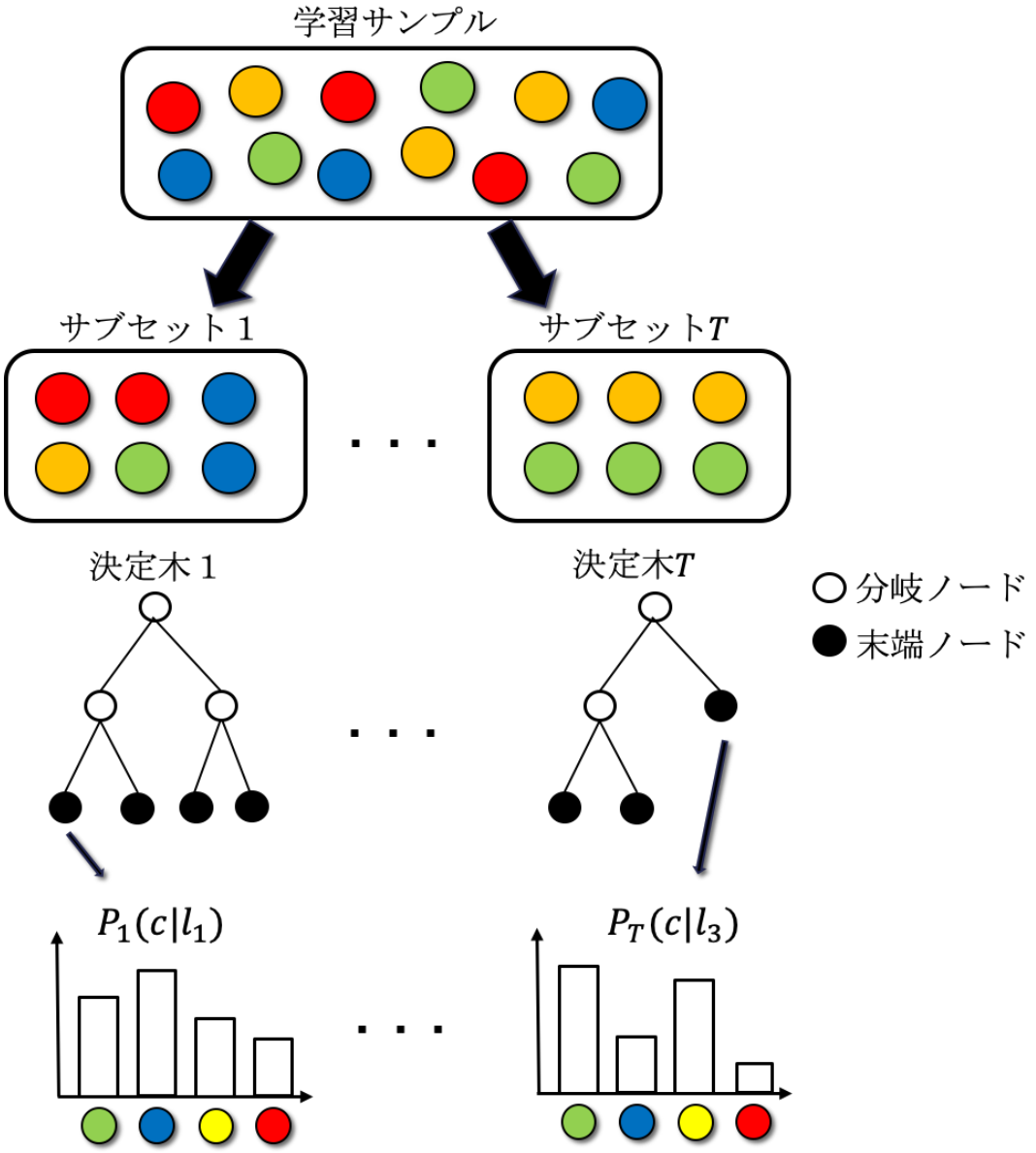


図4.5 Random Forestの仕組み

4.3.2　分岐ノードの作成

分岐ノード n に存在するサンプル集合を Q_n とすれば，分岐ノードで右の子ノードに分岐するサンプル集合 Q_r，左の子ノードに分岐するサンプル集合 Q_l はそれぞれ以下の式で表される．

$$Q_l = \{i | f(v_i) < th, \quad i \in Q_n\} \tag{4.6}$$

$$Q_r = Q_n \setminus Q_l \tag{4.7}$$

ここで，v_i は i 番目の学習サンプルの特徴量，f は分岐関数，th は閾値を表す．$Q_n \setminus Q_l$ は集合 Q_n と Q_l の差集合を表す．次に，(4.5)，(4.6)式で求めたQ_l，Q_rより，情報利得 ΔE を算出し，分岐の評価を行う．情報利得とは，分岐関数によりどの程度情報量が減少したかを表す量で，式(4.7)で示される．

$$\Delta E = E(Q_n) - \frac{|Q_l|}{|Q_n|} E(Q_l) - \frac{|Q_r|}{|Q_n|} E(Q_r) \tag{4.8}$$

ここで，$|Q|$ はサンプルの集合の大きさを表す．
関数 $E(Q)$ は現在のノードの情報エントロピーを表し，分岐後のサンプルの分布の偏りを示す情報エントロピーは次式で表される．

$$E(Q) = -\sum_{j=1}^{n} P(c_j) log P(c_j) \tag{4.9}$$

ここで，$P(c_j)(j = 1,2,\cdots,n)$はクラス c_j の発生確率であり，学習サンプルの教師信号の出現頻度により求められる．
これら以上の処理を繰り返して分岐ノードを作成し，情報利得 ΔE が0になるか，指定した決定木の深さに到達するまで行う．

4.3.3　末端ノードの作成

末端ノードは情報利得 ΔE が 0 になった場合や決定木の深さが指定した深さに達した場合に作成する．末端ノードは到達したサンプル集合から確率分布を得るノードである．末端ノード l において，そこにたどり着いた学習サンプルに付与された教師信号 c_j を投票することにより，各クラスの出現確率$P_t(c_j|l_n)$ $(t = 1,2,\cdots,T)$が次式のように得られる．

$$P_t(c_j|l_n) = \frac{\left|I_{c_j}\right|}{|I|} \quad (j = 1,2,\cdots,n) \tag{4.10}$$

ここで，$|I|$は末端ノード l の全サンプル数，$\left|I_{c_j}\right|$は末端ノード l 内のクラス c_j のサンプル数，nは末端ノード数である．

4. 3. 4　識別

作成した決定木にテストサンプル $\boldsymbol{v}$ を入力する．各末端ノードには各クラスの出現確率が保存されているため，入力したテストサンプルがたどり着いた末端ノードにおける各クラスの出現確率$P(c_j|l)$を出力する．各決定木から出力された出現確率$(P_1(c_j|l), P_2(c_j|l), \cdots, P_T(c_j|l))$を用いてテストサンプル$v$の事後確率$P(c_j|\boldsymbol{v})$を次式より算出する．

$$P(c_j|\boldsymbol{v}) = \frac{1}{T}\sum_{t=1}^{T} P_t(c_j|l) \quad (j = 1,2,\cdots,n) \tag{4.11}$$

ここで，$c_j(j = 1,2,\cdots,n)$は事後確率$P(c_j|\boldsymbol{v})$が最大のクラスをテストサンプルの推定クラスとする．

第５章　音声アシスタント

音声アシスタントとは，情報を提供したり，特定のタスクを実行したりするデバイス上の音声で反応するシステムである．例を挙げると，Apple社のSiriやAmazonのAlexaなどがある．本研究ではインタラクションによってロボットの機能を開発するために，音声アシスタントのシステムを提案する[9]．システムの処理の流れを図5.1に示す．

5.1　提案システムの概要

３章で述べたように，ロボットはユーザーの発話を音声認識し，変換されたテキストから単語情報を取得することにより，それぞれのキーワードに該当するタスクを実行する．タスクの概要は図5.1に示すように4つに分かれている．以下にそれらの機能を述べる．

（1）タスク１：ロボットと人間との会話による画像の獲得と嗜好の回答を保存する．

（2）タスク２：撮影された物体を識別し，識別結果とその結果に対する好みの程度を出力する．

（3）タスク３:現在食べたい味覚の物に応じて,おすすめの食べ物を推奨する．

（4）タスク４：発話していない場合，または「バイバイ」を発話した場合，システムは終了し，ロボットの電源がオフになる．

5.2　データベースの構築について

本研究では事前にデータを収集して決定木を訓練するのではなく，ユーザーと会話しながら，データを蓄積していくという手法となっている．従って，ロボットはタスクを実行し会話を行うことで，データが自動的データベースに追加される．ここで，物体認識や嗜好判別が可能となるために，複数回の会話が既に進められていることが前提となっている．会話した回数に対する物体認識の精度は第６章で述べる．

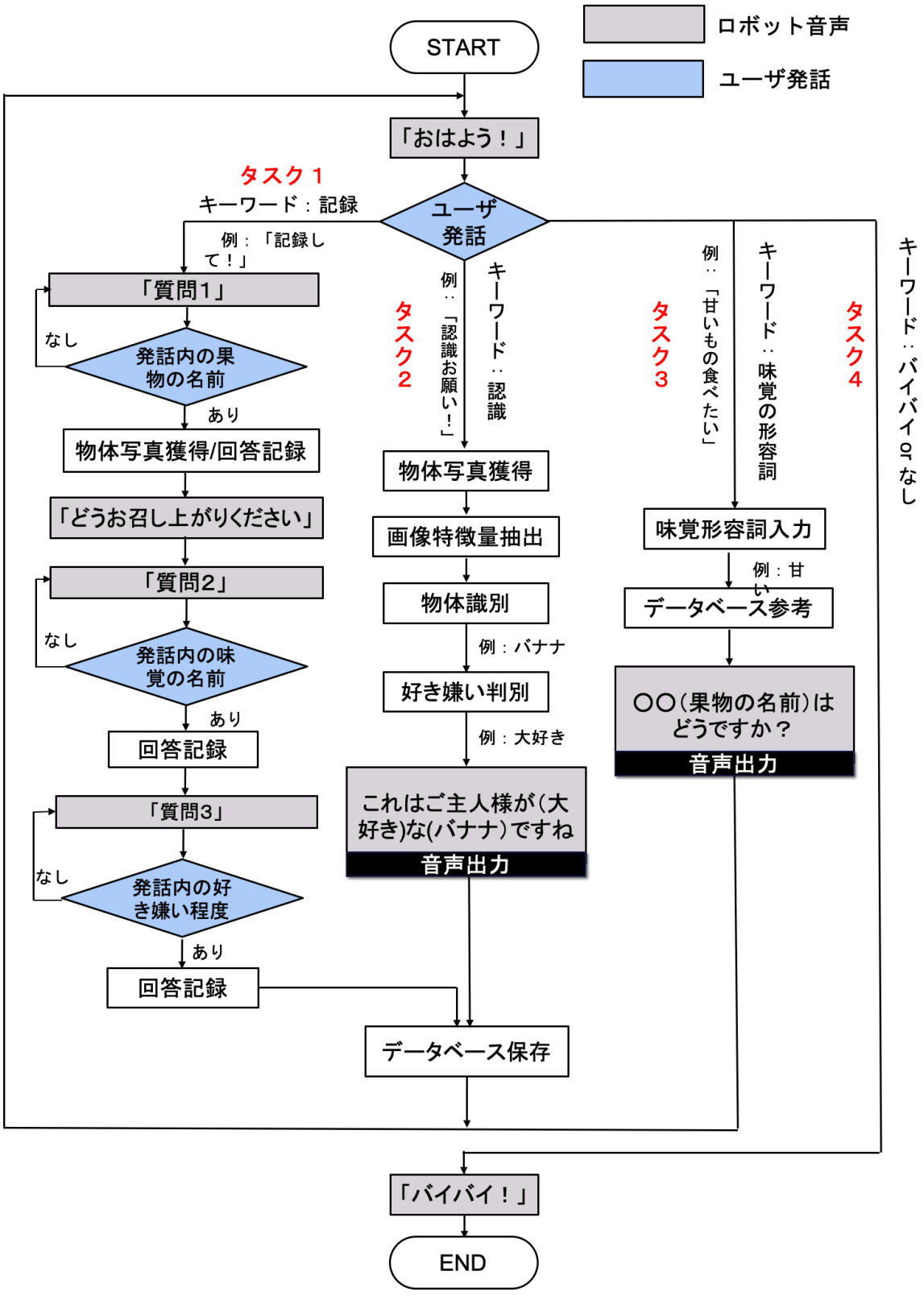


図5.1 音声アシスタントの処理の流れ

第６章　実験

本章では，実験とその結果について述べる．実験１では，音声アシスタントのタスク１とタスク３の機能を実証する．実験２はタスク２の機能実証と物体の画像認識の精度評価を行う．

6.1 実験１

6.1.1 実験環境

屋内環境において，ロボットとユーザーの会話を行う．本実験の撮影環境を図6.1に示す．使用したPCは第２章で述べたRaspberry Pi 4である．

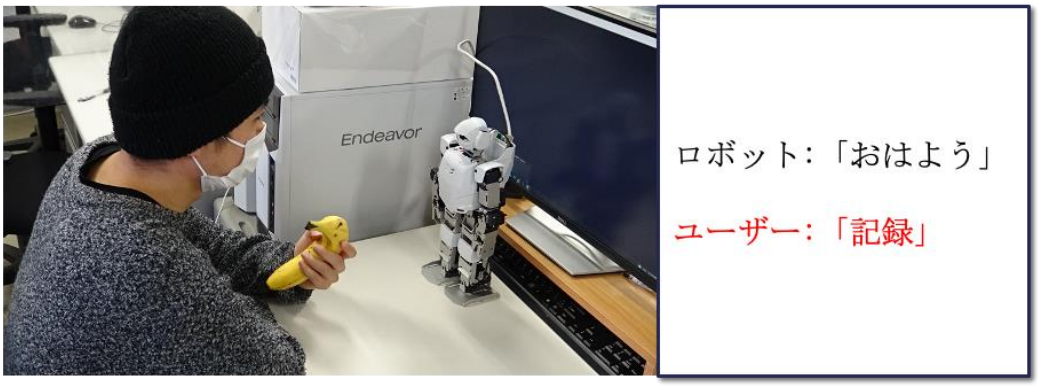
6.1.2　実験方法

タスク１の実験では，ユーザーが食べ物を食べようとしている場合を想定し，食べ物の写真とそれに対するユーザーの嗜好を記録して，機能の有効性を検証する．ここで，タスク１の実験対象物はバナナとする．また，タスク３の実験では，ユーザーの質問に対して，ロボットは正確に答えられるかを検証する．ここで，選別物１個につき，4回会話を行い，計28回の会話で構築されたデータベースをタスク３で使用する．

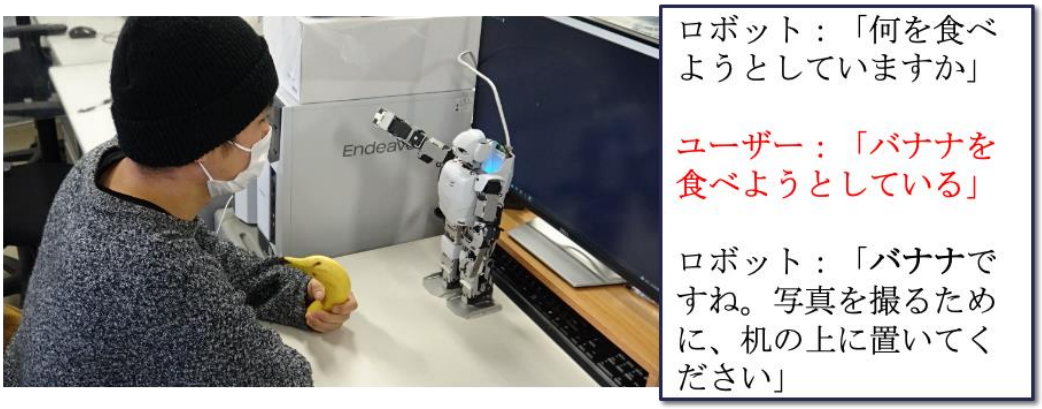
6.1.3　実験結果

タスク１の実行動画を図 6.2 に示し，音声はテキストで表す．図 6.3 は PC の実行画面である．また，タスク３の実行動画を図 6.4 に示し，判定した実験結果の一例を表 6.1 に示す．

図 6.1 実験環境

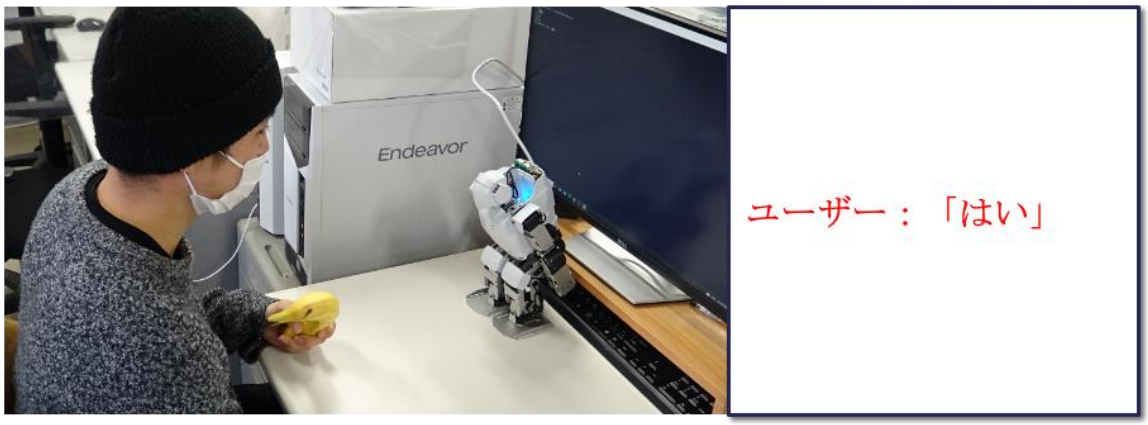


(a)



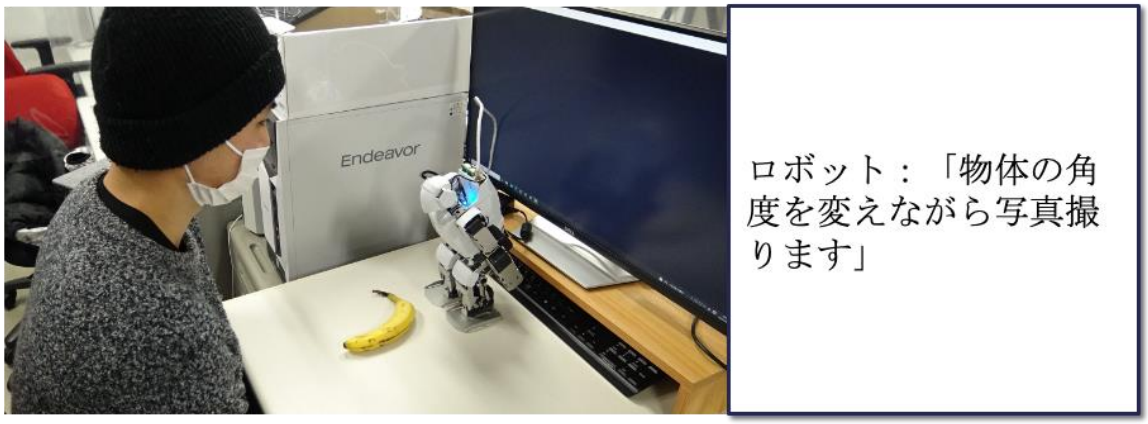
(b)

図 6.2 タスク 1 の実行動画(続く)



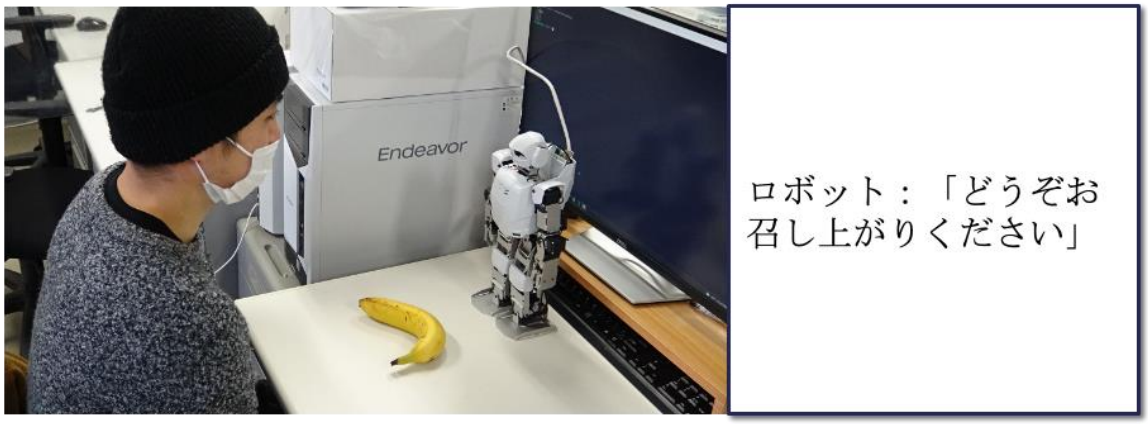
(c)

(d)

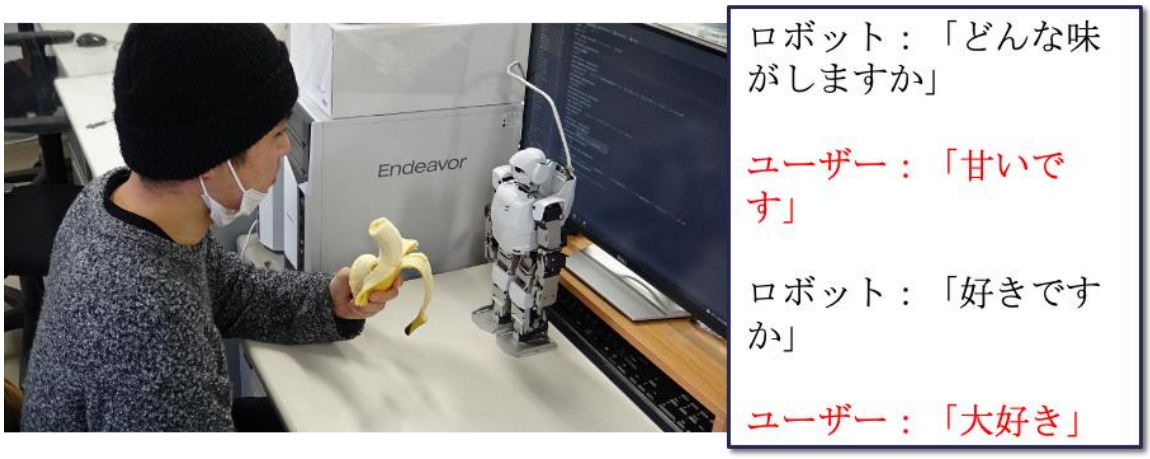


(e)

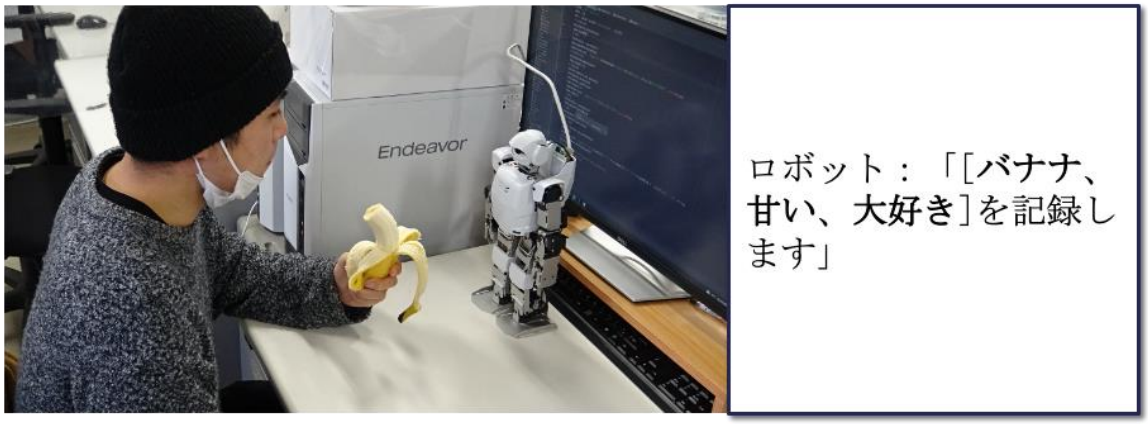
図 6.2 タスク 1 の実行動画（続く）



(f)



(g)



(h)

図 6.2　タスク 1 の実行動画

```
JackShmReadWritePtr::~JackShmReadWritePtr - Init not done for -1, skipping unloc
JackShmReadWritePtr::~JackShmReadWritePtr - Init not done for -1, skipping unloc
記録
w 0048 01
サーボ電源オン
w 09c0 0c
w 09c0 00
w 0048 00
何を食べようとしていますか
jack server is not running or cannot be started
JackShmReadWritePtr::~JackShmReadWritePtr - Init not done for -1, skipping unloc
JackShmReadWritePtr::~JackShmReadWritePtr - Init not done for -1, skipping unloc
バナナです
バナナ   名詞,一般,*,*,*,*,バナナ,バナナ,バナナ
です     助動詞,*,*,*,特殊・デス,基本形,です,デス,デス
EOS

バナナですね。写真を撮るために、机の上に置いてください
w 0048 01
サーボ電源オン
w 09c0 0d
w 09c0 00
w 0048 00
<VideoCapture 0xa736f040>
<VideoCapture 0xa736f030>
w 0048 01
サーボ電源オン
w 09c0 00
どうぞお召し上がりください
どんな味がしますか
甘いです
好きですか
はい
バナナ   名詞,一般,*,*,*,*,バナナ,バナナ,バナナ
EOS

甘い     形容詞,自立,*,*,形容詞・アウオ段,基本形,甘い,アマイ,アマイ
です     助動詞,*,*,*,特殊・デス,基本形,です,デス,デス
EOS

はい     感動詞,*,*,*,*,*,はい,ハイ,ハイ
EOS

['バナナ', '甘い', 'はい']
バナナ,甘い,はい,を記録します
['バナナ', 1, 2]
バイバイ
会話終了
```

図 **6.3** タスク 1 の PC 実行画面

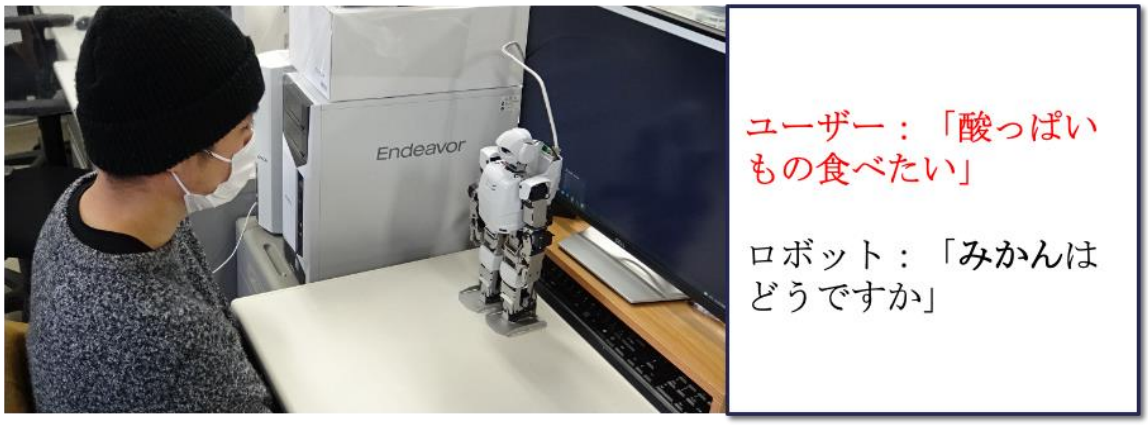


図 **6.4** タスク 3 の実行動画

表 **6.1**　タスク 3 の判定結果

ユーザー発話	ロボット回答	判定結果
甘いもの食べたいな	バナナはどうですか	○
酸っぱいもの食べたいです	みかんはどうですか	○
渋いもの食べたい	柿はどうですか	○
苦いもの食べたい	まだわかりません	○

6.2　実験２

6.2.1　実験環境

実験２ではタスク２の機能を検証し，Random Forest を用いて，ロボットカメラから撮られた果物のデータベース画像に対して物体識別を行う．実験に用いる果物画像のクラスを図 6.5 に示す．また，図 6.6 に示すようにロボットカメラは小型カメラ Raspberry Pi Camera V2 を利用し，画角は 640×480[pixel]である．

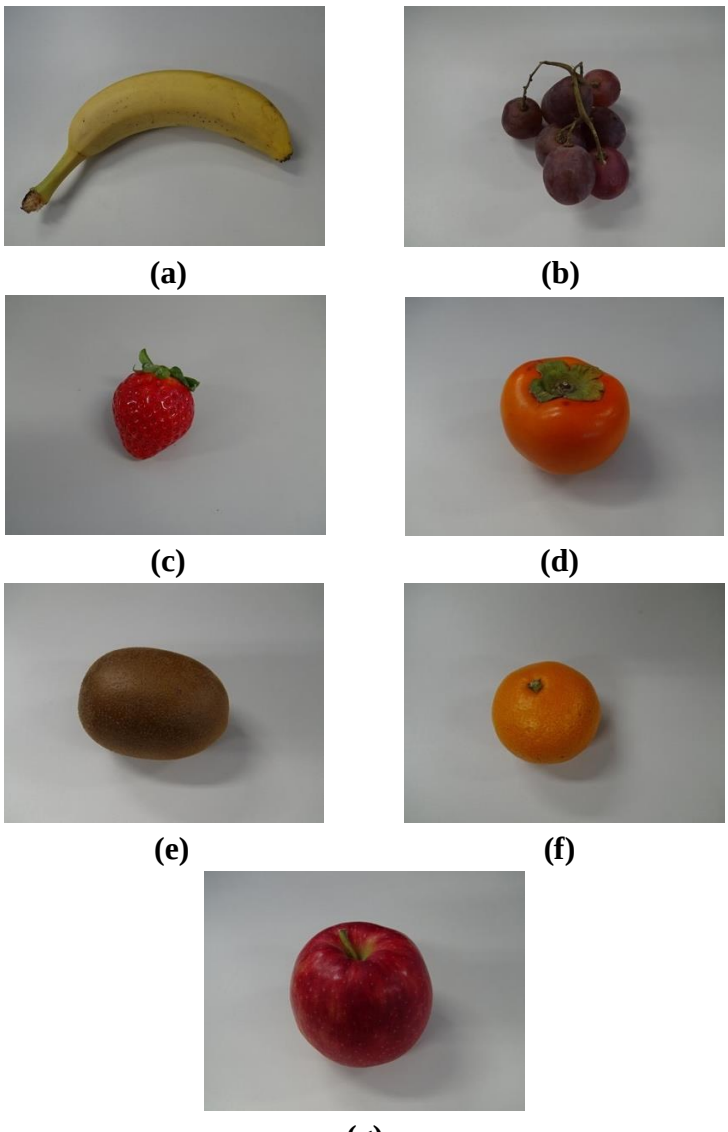
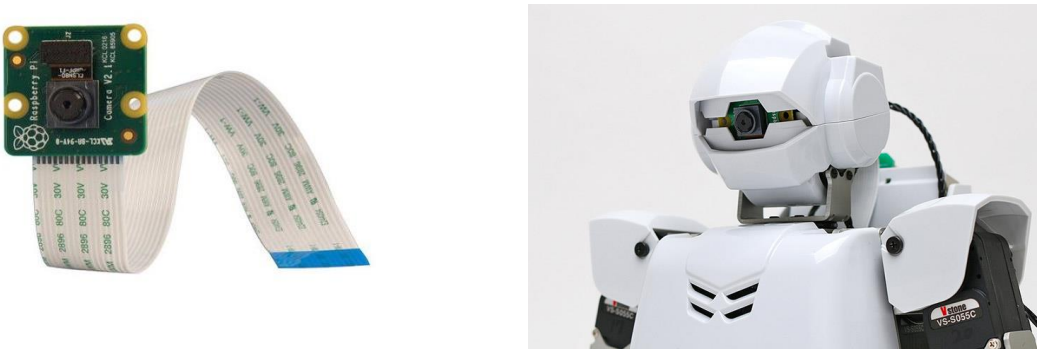


図 6.5　果物画像：(a)バナナ, (b)葡萄, (c)いちご, (d)柿
(e)キウイ, (f)みかん, (g)りんご



(a) **(b)**

図 6.6　CPU とロボットカメラ：(a) Raspberry Pi Camera V2,
(b) カメラを搭載したロボットの外観

6.2.2　実験方法

実験では，各クラスの物体識別精度を$Recall, Precision, F-measure$で評価し，それぞれの評価式は式 6.1，式 6.2，式 6.3 で定義される．また，タスク 2 で使用されるデータベースでは，事前に数回会話を行い，画像が保存されていることが前提となっている．ここで，会話 1 回につき画像 5 枚を獲得するとし，会話の回数による物体識別の精度の変化を考察する．実験では，データベースが 140 枚（7 種類×20 枚），175 枚（7 種類×25 枚），210 枚（7 種類×30 枚），280 枚（7 種類×40 枚），350 枚（7 種類×50 枚）に分けて，それぞれの精度評価と$accuracy$を求める．

$$Recall = \frac{TP}{TP + FN} \times 100[\%] \tag{6.1}$$

$$Precision = \frac{TP}{TP + FP} \times 100[\%] \tag{6.2}$$

$$F - measure = \frac{2 \times Recall \times Precision}{Recall + Precision}[\%] \tag{6.3}$$

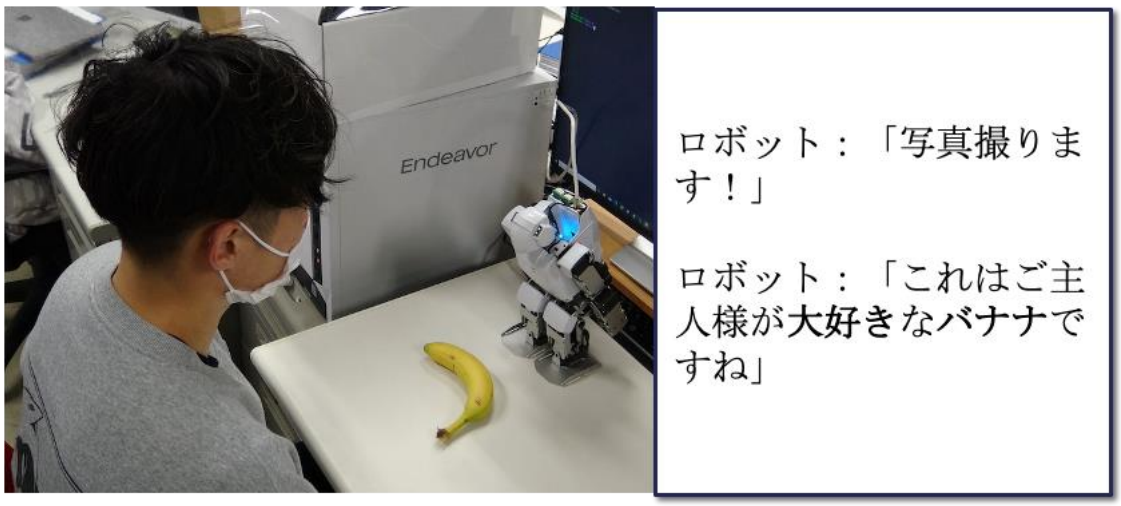
$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \times 100[\%] \tag{6.4}$$

ここで, TPは真の画像を真と識別した枚数. FPは偽の画像を真と識別した枚数. TNは偽の画像を偽と識別した枚数. FNは真の画像を偽と識別した枚数である.

6.2.3　実験結果

タスク 2 の実行動画を図 6.7 に示し，音声はテキストで表す．データベースの画像枚数による物体識別の精度評価を表 6.1~表 6.5 に示す.

(a)



(b)

図 6.7 タスク２の実行動画

表 **6.1** データベース 140 枚の物体識別結果

	$\boldsymbol{Recall}$[%]	$\boldsymbol{Precision}$[%]	$\boldsymbol{F-measure}$[%]
バナナ	100.00	100.00	100.00
葡萄	76.92	100.00	86.96
いちご	100.00	100.00	100.00
柿	87.50	70.00	77.78
キウイ	81.82	90.00	85.71
みかん	90.91	100.00	95.24
リンゴ	100.00	70.00	82.35
$\boldsymbol{Accuracy}$[%]	90.00		

表 **6.2** データベース 175 枚の物体識別結果

	$\boldsymbol{Recall}$[%]	$\boldsymbol{Precision}$[%]	$\boldsymbol{F-measure}$[%]
バナナ	100.00	100.00	100.00
葡萄	100.00	100.00	100.00
いちご	100.00	100.00	100.00
柿	100.00	60.00	75.00
キウイ	100.00	100.00	100.00
みかん	71.43	100.00	83.33
リンゴ	100.00	100.00	100.00
$\boldsymbol{Accuracy}$[%]	94.29		

表 **6.3** データベース 210 枚の物体識別結果

	$\boldsymbol{Recall}$[%]	$\boldsymbol{Precision}$[%]	$\boldsymbol{F-measure}$[%]
バナナ	100.00	100.00	100.00
葡萄	100.00	100.00	100.00
いちご	100.00	100.00	100.00
柿	100.00	80.00	88.89
キウイ	100.00	100.00	100.00
みかん	83.33	100.00	90.91
リンゴ	100.00	100.00	100.00
$\boldsymbol{Accuracy}$[%]	97.14		

表 **6.4** データベース 280 枚の物体識別結果

	$\boldsymbol{Recall}$[%]	$\boldsymbol{Precision}$[%]	$\boldsymbol{F-measure}$[%]
バナナ	100.00	100.00	100.00
葡萄	100.00	100.00	100.00
いちご	100.00	100.00	100.00
柿	100.00	90.00	94.74
キウイ	100.00	100.00	100.00
みかん	90.91	100.00	95.24
リンゴ	100.00	100.00	100.00
$\boldsymbol{Accuracy}$[%]	98.57		

表 **6.5** データベース 350 枚の物体識別結果

	$\boldsymbol{Recall}$[%]	$\boldsymbol{Precision}$[%]	$\boldsymbol{F-measure}$[%]
バナナ	100.00	100.00	100.00
葡萄	100.00	100.00	100.00
いちご	100.00	100.00	100.00
柿	100.00	100.00	100.00
キウイ	100.00	100.00	100.00
みかん	100.00	100.00	100.00
リンゴ	100.00	100.00	100.00
$\boldsymbol{Accuracy}$[%]	100.00		

第７章　考察

本章では，第６章で行った実験の結果ならびに評価について考察を行う．

タスク２の実験において，ロボットが物体を識別するまでの所要時間は約9.7秒であった．システムを実用的に構築するためには，物体識別の処理速度を上げる必要がある．また，表6.3と表6.4のように，みかんと柿の精度が低い．その理由として，2つの物体の色情報と形状が類似し，誤検出が生じることである．改善策として，カメラから物体までの距離を一定にし，物体の大きさを特徴量として追加することが挙げられる．

タスク２のデータベース実験において，140枚，175枚，210枚，280枚，350枚のそれぞれの*Accuracy*は90%，94.29%，97.14%，98.57%と100.00%である．枚数を増やすことにより，*Accuracy*が上がることが検証された．従って，物体１個につき会話を４回行うと物体識別の正解率は90%になる．また，10回まで行うと，*Accuracy*は100%になった．よって，会話はすればするほど，物体の識別率が上がることが証明された．

第８章　結論

本研究では，音声アシスタントを用いたインタラクションによる，人の嗜好判別と物体識別を行うロボットシステムを提案した．

提案法は，ユーザーと会話を行うことにより，選別物の画像と選別物に対する嗜好を記録することができた．また，Random Forest を用いて物体識別を行い，ユーザーの要求に応じた物体を推奨することができた．実験では，物体識別の所要時間は約 9.7 秒，$Accuracy$は 90%以上となった．

今後の課題として，質問を追加しユーザーの嗜好を深く理解できるようにシステムを改善する必要がある．また，物体識別において，撮影距離を一定にして物体の大きさを検出することにより識別精度を向上させることが挙げられる．

参考文献

[1] コミュニケーションロボットの最新動向とAI対応状況2018，株式会社シードプランニング，2018.

[2] 石井 雅樹，石島樹，藤野慎也: “人間とのインタラクション検出に基づく物体認識に関する基礎検討”，知能と情報，30巻，5号，pp.675-681，2018.

[3] Robovie-Z Raspberry Pi版，https://www.vstone.co.jp/products/robovie_z/index.html，(参照2022-2-1)

[4] P. Vashistha，J. P. Singh，P. Jain，J. Kumar: “Raspberry Pi based voice-operated personal assistant(Neobot)”，*International Conference on Electronics Communication and Aerospace Technology*，pp.974-978，2019.

[5] Mecab：Yet Another Part-of-Speech and Morphological Analyzer，http://taku910.github.io/mecab/，(参照2022-1-11)

[6] 本山 和也: “言語の可視化に基づくコミュニケーション・システムの開発”，平成27年度九州工業大学卒業論文,2016.

[7] H. M. Zawbaa，M. Hazman，M. Abbass，A. E. Hassanien: “Automatic fruit classification using random forest algorithm”，*International Conference on Hybrid Intelligent Systems*，pp.164-168，2014.

[8] P. He, H. Li, H. Wang: “Detection of fake images via the ensemble of deep representations from multi color spaces”，*IEEE International Conference on Image Processing*，pp.2299-2303，2019.

[9] S. Subhash, P. N. Srivatsa, S. Siddesh, A. Ullas, B. Santhosh: “Artificial intelligence-based voice assistant”，*World Conference on Smart Trends in Systems，Security and Sustainability*，pp.594-596，2020 .

謝辞

本研究を行うにあたり，指導教員のタンジュークイ教授から，ご多忙の中多大なご指導をいただき，深く感謝申し上げます.

また，数々の助言をいただいた研究室の皆様に心から感謝いたします．この場を借りて厚く御礼申し上げます.

付録

表１　タスク３の実験データベース

	味覚	好き嫌い
バナナ	A	4
キウイ	F	2
葡萄	A	1
葡萄	G	1
りんご	G	2
みかん	B	3
バナナ	A	4
柿	D	1
柿	D	1
みかん	B	3
りんご	A	2
キウイ	A	3
いちご	A	4
バナナ	A	3
キウイ	B	1
いちご	A	2
りんご	G	1
柿	D	0
みかん	F	4
りんご	A	2
バナナ	A	4
柿	D	0
葡萄	F	1
キウイ	F	4
葡萄	G	2
いちご	A	3
いちご	A	4
みかん	F	4