

# 認知症者の介護を支援するシステムの予備検討

大島 千佳<sup>1,a)</sup> 田中 竜二<sup>2</sup> 浜崎 亮大<sup>2</sup> 中山 功一<sup>1</sup>

**概要:** 認知症者が穏やかに過ごせる環境が提供できれば、介護者の負担を減らすことができると考える。我々はもの忘れ支援として、物体が置かれるだけで物体を検知・認識するシステム、認知症者が徘徊などの行動を起こしそうな状態を検知するシステム、そしてこのような認知症者の状態を多忙な介護者に伝える手法の研究を行っている。本稿では、それぞれのシステム・手法について予備検討を行った。システムは事前に、多数の様々な眼鏡の画像を撮影してカテゴリ化した学習をせずに、無地の背景に置かれた眼鏡を検知・認識できた。次に、行動を起こしそうな状態を検知するシステムを使って、健常の大学生が二者の会話に割り込もうとしている動作の検知を試みた。正解率は高くなかったが、検知できる可能性が示唆された。最後にオープンキャンパスで来客者への対応に追われている大学院生1名に、ロボットから来客者の状態を伝える実験を行った。ロボットの配置や呼びかけ方によって、大学院生のロボットへの感情や自分の情緒が変わることがわかった。

## Preliminary Examinations of Support Systems for Dementia Care

### 1. はじめに

アルツハイマー病による認知症やその手前の軽度認知障害(MCI)により、怒りやすくなる人たちがいる。認知症の中核症状である「もの忘れ」を自覚し不安になることや、もの忘れによる失敗や誤りを他人から指摘されることで怒りへと向かう[1]。不安や怒りが高まった不穏な状態は、徘徊、暴言・暴力といった行動・心理症状(BPSD: Behavioral and Psychological Symptoms of Dementia)の言動に発展することも多い。一方で、このような言動が始まる前に、他者から適切に声をかけると、穏やかな状態に戻ることもある。

我々は、物体が置かれるだけで物体を検知・認識するシステム、認知症者が徘徊などのBPSDを発症しそうな状態を検知するシステム、そしてこのような認知症者の状態を多忙な介護者に伝える手法の研究を行っている。本稿では、各システム・手法について、予備検討を行った結果を報告する。

### 2. 事前学習の要らない物体検出

本章は、文献[2]で発表した内容である。

#### 2.1 背景

もの忘れの1つに、持ち物の紛失が挙げられる。日常的に使う物をどこに置いたか忘れてしまう。また保管している物を突然思い出すことがあっても、保管場所を思い出せない。物にあらかじめタグをつけておくことで、紛失した際にスマートフォンのアプリケーションを使って場所を突き止める方法は商品化されている。しかし、すべての物体にタグをつけることは現実的ではない。従来、機械学習で物体認識を行うには、1物体につき1000枚程度の画像を準備して特徴を抽出し、Support Vector Machine(SVM)などの学習器を使ってその物体を識別するための識別器を作成するという大掛かりな手順が必要であった。我々は、物体が出現(誰かが部屋に置いた)や消滅(誰かが持ち去った)するたびに、その物体が何であるか記録するシステムを開発した。

#### 2.2 システム

図1は、1台のカメラにより物体を検知して認識する過程を示している。システムはまず、OpenCV(Open Source

<sup>1</sup> 佐賀大学理工学部  
Faculty of Science and Engineering, Saga University, Japan  
<sup>2</sup> 佐賀大学大学院理工学研究科  
Graduate School of Science and Engineering, Saga University, Japan  
a) karin27@sa3.so-net.ne.jp

Computer Vision Library) で提供されている背景差分のアルゴリズム (BackgroundSubtractorMOG [3]) を使って、物体を検出する。図 2 に示すように、単純に説明すると、何も置かれていない背景の画像と、その背景に物体が置かれた画像の両方を OpenCV に入力すると、OpenCV は物体を検知してそのまわりをバウンディングボックス (矩形) で囲う。その 4 つの角の座標を出力する (返す)。次に、図 3 に示すように、物体が置かれた画像を X 軸、Y 軸でそれぞれ反転し、さらに 180 度回転させた画像を作成する。この方法により、1 枚の画像から 4 倍の枚数に増やすことができる。この過程を  $n$  回繰り返した後に、図 4 に示すように、画像のファイル名とその画像での物体の位置を示す座標データ (正解) を、教師データとして YOLO[4] に渡す。YOLO はリアルタイムに物体を検知・認識するためのディープラーニングを使ったアルゴリズムである。処理速度が速く、正解率が高いことで知られているため、本研究では YOLO を採用した。システムは YOLO を使って、教師データを学習した。

開発言語は Python で、ライブラリは OpenCV を使用した。また YOLO を使うために、ライブラリの Keras を使用した。

### 2.3 検証実験

システムの学習モデルを検証した。図 5 に示すように、1 台のカメラで眼鏡を置いた画像を 40 枚撮影した。軸で回転させることにより、4 倍の 160 枚になった。今回はもう 1 台カメラを設置して撮影したため、320 枚の画像セット (データセット) を取得した。320 枚のデータセットのうち、教師データとして 224 枚を使用した。残りの 96 枚を検証に使った。これらとは別にテスト用に 20 枚準備した。テスト用データは教師データと同様に OpenCV を用いて作成した。学習した回数は 100 エポックだった。

判定は適合率、再現率、F 値を使用した。本検証では図 6 に示すように、正解のバウンディングボックスとの重なり具合の率を示した。

$$Precision = \frac{TruePositive}{TruePositive + FalsePositive} \quad (1)$$

$$Recall = \frac{TruePositive}{TruePositive + FalseNegative} \quad (2)$$

$$F - measure = \frac{2 * Precision * Recall}{Precision + Recall} \quad (3)$$

### 2.4 結果と考察

OpneCV を用いたテスト用データの本システムの適合率、再現率、F 値はそれぞれ、0.91, 0.90, 0.91 であった。比較のために筆者はテスト用データを OpenCV で作成する方法とは別に、手で検知物をバウンディングボックス (矩形) で囲った “マニュアル” のテスト用データを作成した。判定結果はそれぞれ、0.70, 0.98, 0.82 であった。

これらの結果から、システムはかなり精密に物体を検知しているといえる。本検証では枚数を増やすためにカメラを 2 台使用したが、1 台分の 160 枚のデータセットのみを学習しても、十分に物体を検知できた可能性は高い。今後は、小さい物体でも検知できるようにアルゴリズムを改良していく。

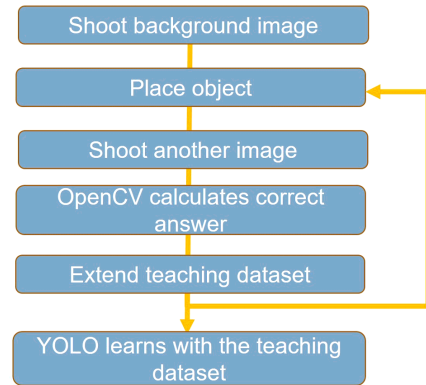


図 1 物体を検知・認識する過程

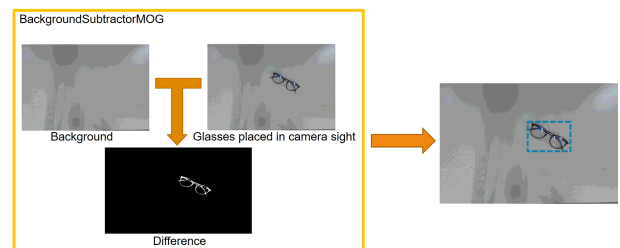


図 2 背景差分による物体検知

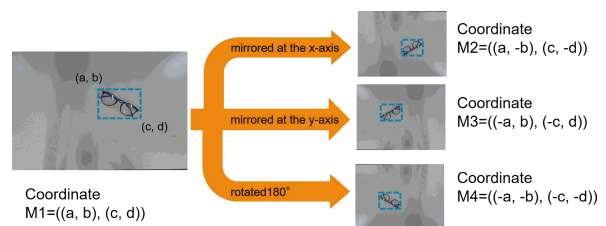


図 3 枚数を増やす方法

Path	Coordinates
Sample_image1.jpg	((a1, b1), (c1, d1))
Sample_image2.jpg	((a2, b2), (c2, d2))
Sample_image3.jpg	((a3, b3), (c3, d3))
Sample_image4.jpg	((a4, b4), (c4, d4))
⋮	⋮

図 4 YOLO に渡すリストのイメージ

## 3. 予備動作の検出

本章は、文献 [5] で発表した内容である。

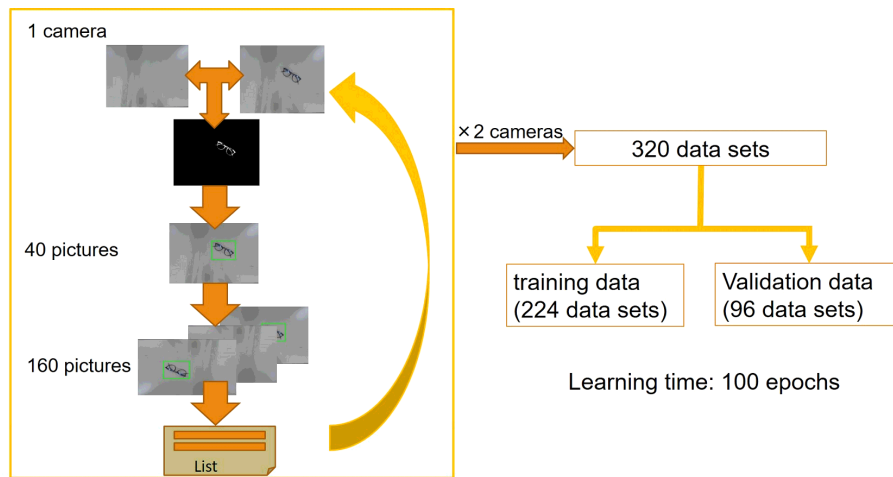


図 5 検証の方法

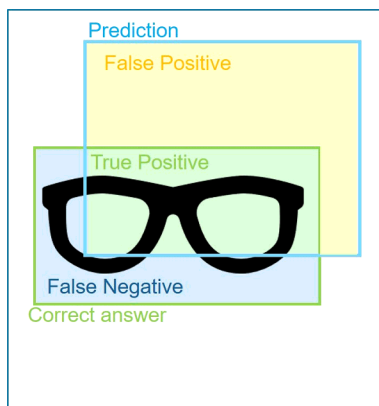


図 6 適合率, 再現率, F 値の出し方

### 3.1 背景

認知症者が不穏な状態になり、BPSD を発症する前には、そわそわした状態やじっと黙り込む状態など、個々の患者によって何らかの動作があると考えられる。いつも近くに居る介護者ならば察知できる可能性があるが、あまり一緒に居たことがない介護者や、大勢と一緒に過ごす介護施設の介護者などは、個々の患者の BPSD 発症前の動作に気づくことは難しい。個々の認知症者の BPSD 発症前の動作推定が可能であれば、BPSD 発症前に介護者に通知でき、介護者がその認知症者に、声かけをすることで、徘徊や暴力といった BPSD を回避できる。

本稿では、本研究の第一段階として、健常者を対象に、二者の会話に割り込もうとする人の予備動作を判定するシステムを構築できないか検証した。

### 3.2 予備動作のデータ収集

図 7 に示すように、二者が会話をしている状況に、別の人(被験者)が会話に割り込もうとする動作を収録した。以下、割り込もうとする人を「インタラプタ」と呼ぶ。インタラプタの座っている姿をビデオカメラ (Sony Handycam HDR-CX480) で収録する。

9 人の男性の大学生が実験に協力した。3 人ずつ 3 グループに分かれ、9 人がそれぞれのグループで順番にインタラプタを務めた。会話時間は 1 ペアにつき 20 分間である。あらかじめ設定したテーマについて片方からもう片方の話者に質問をする形で会話を始める。

2 人の話者のそばには、20 種類の小物を置いた。インタラプタの前にはスマートフォンが置かれており、その画面に、「話者 A さんから話者 B さんにペンを渡す」というように、20 種類の小物を話者同士で受け渡しする指示が不定期に表示される。インタラプタは、指示が表示されると、二者の会話に割り込み、「話者 A さんは、話者 B さんにペンを渡してください」とお願いする。

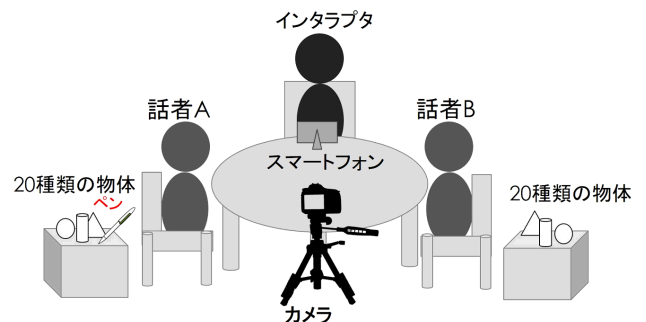


図 7 会話をする二者と割り込むインタラプタ

### 3.3 倫理的配慮

収録前には各参加者に、会話を録画することと、個人を特定できる情報は開示しないことを説明し、署名により承諾を得た。収録後に、本研究の主旨を説明し、再度同意を得た。

### 3.4 データの処理と学習

まず第二筆者が、インタラプタの画像から、二者の会話に割り込もうとしている予備動作と、それ以外の部分に分

けた。次に録画データを OpenPose [6] に入力し、骨格データを作る。

OpenPose は、顔の詳細も含めた全身に“キーポイント (keypoint)”があるが、今回はそのうち 18 のキーポイントでの 1 秒間 30 フレームのデータを使った。図 8 に示すように、教師データとして学習モデルに入れる前に、データの編集をした。1 回の予備動作は 3 秒間とした。1 秒間に 30 フレームの骨格データがある。1 秒ごとの最初の 1 フレームを取り出した 3 フレーム分を、1 セットとした。同様のことを次の 1 フレームに行い、繰り返すことで、30 セットが作成された。この方法により、連続した 3 フレームよりも、1 秒間隔で集めた 3 フレームの方が、より多い動作の情報が含まれると考えたからである。

図 9 に示すように、20 分間の会話のうち、17 分間分の骨格データの 9 割は教師データ、1 割は検証データに使った。残りの 3 分間分の骨格データはテストデータに使った。20 分間のうちの 17 分間と 3 分間はランダムに選ばれる。17 分間のデータの中で、会話に割り込む予備動作の骨格データと、それ以外の動作の骨格データは、それぞれ 1230 セットずつになった。学習回数は 2000 エポック行った。その後この学習モデルの評価を、テストデータを使って行った。

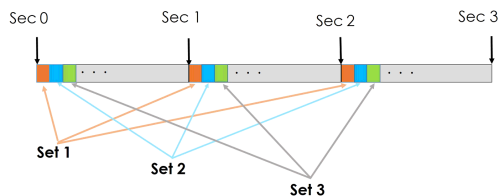


図 8 データの編集

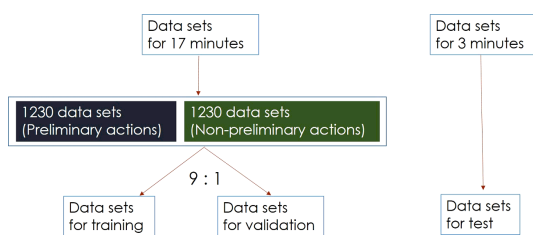


図 9 教師データ、検証データ、テストデータの分け方

### 3.5 結果と考察

表 1 は、結果の一例として、6 分から 9 分をテストデータに使ったときの時間と正解の関係を示す。適合率、再現率、F 値はそれぞれ、0.71, 0.56, 0.63 であった。

このようにテストに使った 3 分間ごとの、適合率、再現率、F 値を表 2 に示す。F 値がテストデータに使った区間によってばらついたことがわかる。

これらの結果から、作成したモデルが 3 秒間のスケルト

表 1 6-9 分をテストデータに使った結果

実際の予備動作 開始時刻	推定された予備動作 開始時刻	正解
	6'03"	
6'25"		
6'32"		
6'50"	6'50"	正解
7'17"	7'17"	正解
7'39"		
	7'42"	
7'48"	7'48"	正解
8'24"	8'24"	正解
8'29"	8'29"	正解
8'40"		

表 2 テストに使った 3 分間ごとの結果

テストデータ (分)	適合率	再現率	F 値
0-3	0.43	0.50	0.46
3-6	0.44	0.57	0.50
6-9	0.71	0.56	0.63
9-12	0.43	0.43	0.43
12-15	0.50	0.50	0.50
15-18	0.29	0.25	0.27
17-20	0.36	0.57	0.44
Average	0.45	0.48	0.46

ンデータ (予備動作) により、会話に割り込もうとしている予備動作かどうかを識別可能であるとは言い難い。時間による予備動作の様子を検証したところ、「座りなおす」動作は識別しやすいと考えられた。一方で、視線を動かす動作 (目が泳ぐ) は、今回 OpenPose で認識するように設定していなかったため、識別できなかった。また、会話が盛り上がっていないときには、割り込みやすく、予備動作なしで指示をしていたため、識別できなかった。

## 4. ロボットを通じた状態の伝達

### 4.1 背景

介護施設の需要が増大し、介護スタッフの離職率が高い近年、少ないスタッフで多くの高齢者の介護をしなければならない。ある認知症者が不穏であることをシステムが判断して、スタッフに直接伝えても、その認知症者の近くにいないければ、他のスタッフを当てにするかもしれない。また排泄の付き添いなど、急ぎの用事があればそれらを優先するのは当然であろう。

そこで本研究では、図 10 に示すように、ロボットから認知症者へ話しかけることにより、近くにいるスタッフがその内容から認知症者の状態を知り、その認知症者に声をかけるように促す手法を提案している [7]。この手法により、介護者がスマートフォンなどでいちいち通知を受け取る必要はなく、近くにいる介護者が対象の高齢者の様子を直に見ることもできる。

本稿ではこの手法を使って、健常の大学院生を対象にした実験を行う。大学のオープンキャンパスで来場者への対応に忙しい大学院生1名が被験者になり、ロボットから来場者への声かけと、その大学院生へ直接声かけする方法の2条件を行う。ロボットから声をかけられた大学院生がどのように感じるか明らかにする。

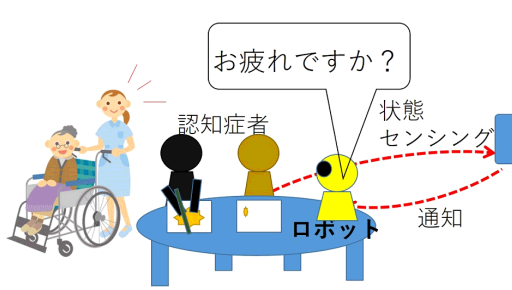


図 10 ロボットが認知症者に話しかけることでスタッフが状態に気が付く

#### 4.2 倫理的配慮

実験前に、被験者となる学生に、研究のために、その学生が担当するブースにロボットを置いて、ビデオカメラで撮影することを説明し、署名にて同意を得た。実験の目的の都合上、実験前には詳細な内容を説明することはできなかった。そのため、実験終了後に実験の主旨を説明し、データを研究に使うことについて再度同意を得た。またそのブースには、人目につきやすい位置にビデオカメラを設置し、研究のために録画をしていることを文書で明示した。そのブースでゲームの体験に参加する高校生には、研究のために録画していることと、個人を特定するような会話や映像は外部に発表しないことを説明し、口頭で同意を得た。

#### 4.3 実験方法

被験者となる学生 A は B 大学大学院修士課程の学生である。B 大学で開催された高校生を対象にしたオープンキャンパスでは、学生らが作ったゲームを高校生に披露して、体験してもらうブースを担当した。同じブースには、ロボット“CATARO[7]”も展示して、来場者から説明を求められたときには、学生 A が説明した。要するに、学生 A は来客者への対応で非常に多忙であった。

2 種類の条件を設定した。どちらの条件も、展示用とは別の CATARO を使った。高校生のゲームの状況に応じて発話する。発話のパターンはあらかじめ録音したファイルが準備されており、学生 A ではない他の学生がノートパソコン上で再生ボタンを押すことで、選択された音声ファイルが再生された。スピーカは CATARO の胴体の中に入れることができるが、今回は来場者が多く、音量が足りなかったため CATARO の背後に置いた。時間は 1 条件 1 時間であった。

ゲームを体験する高校生が、学生 A に質問したくなる状況を作るために、ゲームの展開の中で、操作がわからない部分を設定した。

条件 1 では、図 11 に示すように、高校生がゲームを体験するテーブルを 2 台準備し、テーブルごとに CATARO を配置した。つまり展示用の CATARO のほかに、発話をする CATARO を 2 台使用した。テーブルの上に置かれた CATARO はゲームを体験する高校生に向かって発話をした。発話の種類を下記に示す。

- (1) あれあれ？やり方がわからないかな？ぼくにはわからないから、係の人に聞いてね。
- (2) あれあれ？何か困ってるのかな？ぼくにはわからないから、係の人に聞いてね。
- (3) ねえねえ、あきちゃった？ゲームおもしろくないのかな？
- (4) ねえねえ、つかれてるんじゃない？
- (5) ねえねえ、終わった？終わったら、係の人に聞いてね。
- (6) はじまったね。いっしょに楽しもうね。
- (7) 調子いいね。がんばれー。
- (8) どんどん進んでいるね。

条件 2 では、図 12 に示すように、学生 A のそばに CATARO を 1 台配置した。CATARO は学生 A に向かって発話をした。発話の種類を下記に示す。

- (1) A 君、1 番のお客さんが何か困ってますよ。見に行ってください。
- (2) A 君、1 番のお客さんが飽きてますよ。見に行ってください。
- (3) A 君、1 番のお客さんが終わったようですよ。見に行ってください。
- (4) A 君、2 番のお客さんが何か困ってますよ。見に行ってください。
- (5) A 君、2 番のお客さんが飽きてますよ。見に行ってください。
- (6) A 君、2 番のお客さんが終わったようですよ。見に行ってください。
- (7) A 君、CATARO の説明を待っているお客さんがいますよ。戻ってきてください。

実験後、学生 A に質問票により、文献 [8] を参考にして、各条件でのロボットに対する各感情を、1 (全く感じない) ~ 5 (とても感じた) の 5 段階で評価してもらった。さらに、各条件での自分の情緒について、同様に 5 段階で評価してもらった。その後第一筆者がインタビューを行った。

#### 4.4 結果と考察

表 3 に、ロボットに対する感情の結果を示す。条件間で 3 以上の評価の差がついた感情は、「屈辱感 (条件 1 の評価が 1, 条件 2 の評価が 4)」と「けむたさ (条件 1 の評価が 1, 条件 2 の評価が 4)」であった。

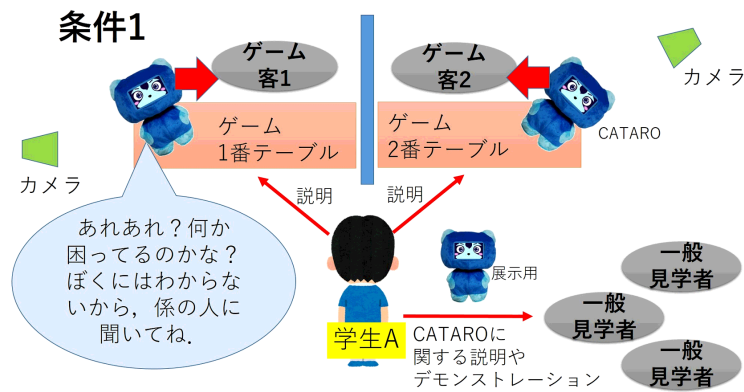


図 11 条件 1：各テーブルに CATARO を配置する

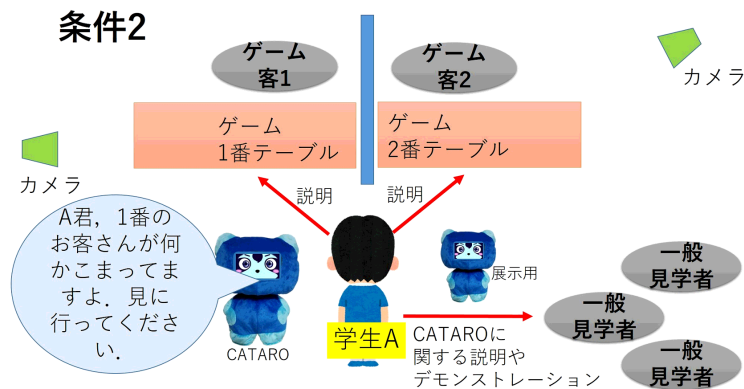


図 12 条件 2：学生 A のそばに CATARO を配置する

表 4 に、学生 A 自身の情緒の結果を示す。条件間で 3 以上の評価の差がついた感情は、「さびしさ（条件 1 の評価が 5，条件 2 の評価が 1）」「不安（条件 1 の評価が 1，条件 2 の評価が 4）」「恥ずかしさ（条件 1 の評価が 2，条件 2 の評価が 5）」であった。

学生 A にインタビューして得られた結果の要約を下記に示す。

#### 条件 1 について

- (CATARO から) 名前を呼ばれないと、言っている意味を聞こうとしなかった。
- 人 (他の学生) が発話を選んで出力していることがわかっていたので、「やめてくれ」と思った。

#### 条件 2 について

- (ロボットに自分の) 名前を呼ばれたことは、人がたくさんいる状況だったので、迷子のお知らせをされているような気持ちになった。
- 操作する人が大人 (後輩) だとむかっとくるかも。ロボットから仕事に関する内容を話しかけられることについて
- もし CATARO が高度な人工知能で、自分で判断し、発話を出力していたら、「しょうがないいな、つきあってやるか」「犬だしな (と同様の)」という気持ちにな

るかもしれない。

- 自分と CATARO の間に関係が築けていたら、親近感もわくと思う。

条件 1 より条件 2 の方が、ロボットに対して屈辱感やけむたさがあり、不安や恥ずかしさがあったことがわかる。その理由として、条件 2 の CATARO は、常に学生 A の名前から始める呼びかけをしていたため、大勢の高校生などが出入りする会場で、恥ずかしい思いをしたという点が挙げられる。さらに、CATARO の発話の操作を学生 A の後輩が行っていたことがわかってしまったため、余計にロボットに対して屈辱感が生まれたと考えられる。

条件 1 が、本来提案したい手法であるが、学生 A は CATARO の言っている意味をほとんど聞いていなかったことがわかった。学生 A は次々と外部から来ている高校生やその保護者といった客への対応に追われていた。まさに、高齢者への対応に追われている介護スタッフのような状況であった。人間と接しているということは、会話を行っていることが多く、たとえ CATARO の声は聞こえていても、内容を理解することは難しかった。また、多忙な状態のときに、CATARO から執拗に声を掛けられることはけむたさにもつながるようだった。

他者からの依頼を受理する理由は、他者に対して何らか

表 3 ロボットに対する感情

感情	条件 1	条件 2	差
好感	4	4	0
軽蔑	1	2	1
愛情	3	4	1
嫌悪	1	3	2
かわいい	4	4	0
憎悪	1	1	0
なじみ	4	3	1
しっと	2	1	1
敬意	2	3	1
屈辱感	1	4	3
感謝	4	4	0
劣等感	1	1	0
甘え	2	2	0
怒り	1	1	0
優越感	4	2	2
恐怖	1	1	0
気安さ	4	2	2
気味悪さ	1	1	0
かわいそう	1	1	0
けむたさ	1	4	3
気がね	4	3	1
うらやましき	3	1	2
失望	1	1	0

表 4 学生 A の情緒

情緒	条件 1	条件 2	差
さびしさ	5	1	4
喜び	3	4	1
悲しみ	2	1	1
うれしさ	4	4	0
苦悩	1	1	0
安心	3	3	0
絶望	1	1	0
満足感	4	3	1
イライラ	1	1	0
誇り	3	3	0
不安	1	4	3
驚き	1	1	0
恥ずかしさ	2	5	3
不満	1	1	0
くやしき	1	3	2

の権威を持っているからだ [9] とされる。人とロボットとの関係（インタラクション）に関する研究が行われている。小野ら [10] は、携帯端末に表示したエージェントが物理的なロボットに移動するシステムを開発した。ユーザは端末上のエージェントとインタラクトすることで、エージェントに信頼感や親近感を持つようになり、ロボットへ移動した後も、その関係が継承されることを示した。本研究も、ロボットと学生、ロボットと介護スタッフとの間に信頼感や親近感が生まれる仕組みが必要だと考えられる。

## 5. おわりに

本稿では、背景差分の技術を応用して、背景に物体が置かれるとシステムが学習するための画像を自動的に生成し、YOLO に渡すことで物体の検知・認識ができる手法を検証した。背景に眼鏡を置いたところ、非常に高い精度で眼鏡を検知・認識することができた。また、二者の会話に割り込もうとしている人の動作の検知を試みた。精度はあまり高くなかったが、目や足の動きも判断材料にすると、もっと高い精度で検知ができる可能性が示唆された。最後に、大学のオープンキャンパスで、来客者への対応で多忙な学生に対して、ゲームをしている来客者の状態を、ロボットから伝える実験を行った。ロボットの配置により、ロボットに対する感情や自分の情緒が異なることがわかった。今後はロボットとの信頼関係を構築する手法も取り入れることで、ロボットからの伝達に耳を傾けられるようにする。

謝辞 本研究は JSPS 科研費基盤研究 B (17H01950) の助成を受けたものである。

## 参考文献

- [1] 高橋智. (2011). 認知症の BPSD. 日本老年医学会雑誌, 48(3), 195-204.
- [2] Hamasaki, R. and Nakayama, K.: A Deep Learning System that Learns a Discriminative Model Autonomously Using Difference Images, *Proceedings of the Genetic and Evolutionary Computation Conference Companion*, pp. 1683-1685, ACM, 2019.
- [3] KaewTraKulPong, P. and Bowden, R.: An improved adaptive background mixture model for real-time tracking with shadow detection, *Video-based surveillance systems*, Springer, Boston, MA, pp. 135-144, 2002.
- [4] Redmon, J., Divvala, S., Girshick, R., and Farhadi, A.: You only look once: Unified, real-time object detection. *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 779-788, 2016.
- [5] Tanaka, R., Oshima, C., and Nakayama, K.: Intention Inference from 2D Poses of Preliminary Action Using OpenPose, *Proceedings of the Genetic and Evolutionary Computation Conference Companion*, pp.1697-1700, ACM, 2019.
- [6] Cao, Z., Simon, T., Wei, S,E, and Sheikh, Y.: Realtime multi-person 2d pose estimation using part affinity fields. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 7291-7299, 2017.
- [7] 大島 千佳, Patrick Hock, 中山 功一: CATARO: 介護者に被介護者の状態を気づかせるロボット, 情報処理学会アクセシビリティ研究会, 2018-AAC-7(3), 2018.
- [8] 齋藤勇: 対人感情と情緒の人間関係的アプローチ, 心理学研究, 56(4), pp. 222-228, 1985.
- [9] 村上直隆, 片上大輔, 山田誠二: 権威付けによるヒューマンロボットインタラクション, 人工知能学会全国大会論文集, 第 17 回全国大会, pp. 133-133, 2003.
- [10] 小野哲雄, 今井倫太, 江谷為之, 中津良平: ヒューマンロボットインタラクションにおける関係性の創出, 情報処理学会論文誌, 41(1), pp. 158-166, 2000.