

アクションユニットレベルでの表情模倣と記号レベルでの表情模倣における共感性の比較

Comparison of empathy between action unit level and symbolic level in facial expression mimicry

山崎真由香¹ 港隆史² 飯尾 尊優^{1*}
Mayuka Yamasaki¹ Takashi Minato² Takamasa Iio¹

¹ 同志社大学

¹ Doshisha University

² 理化学研究所ガーディアンロボットプロジェクト

² Guardian Robot Project, RIKEN

Abstract: Humans tend to engage in “conforming behavior” in which they unconsciously imitate the facial expressions and gestures of others. Because attuned behavior gives the other person a favorable impression, the effect of robots recognizing and attuning to the other person’s emotions has been studied. However, prior research has shown that it is difficult for robots to recognize and reproduce complex emotions and small nuances. The process of imitation is also different from that of human communication. In this research, we propose a system that imitates action units and verify empathy. We implement the action unit imitation system and the symbolic level imitation system on an android robot, and conduct an experiment in which the robot responds to a participant’s narration using only facial expressions. The data obtained from the experiment will be used to compare impressions of the robot. The results showed that there was no significant difference between the two conditions, but a certain number of supports were obtained in both conditions, indicating that the robot is expected to be used effectively in different situations.

1 はじめに

ヒトは対人相互作用場面において無意識に相手の表情や仕草を真似する傾向にある。この行動は相手に安心感など好感を与え [1]、この模倣行動はミラーリングや同調行動と呼ばれる [2]。ロボットと人間の円滑なコミュニケーションを実現するため、ロボットも同調行動をとることで得られる効果について多く研究されている [3][4]。近年ではヒトの感情を認識し同調する方法についても研究が進められているが、ロボットが認識する感情は基本6感情などカテゴリーが大きい [5]。そのため微笑や苦笑といった小さなニュアンスを、ロボットが認識し再現することは難しい。また、人間がとる同調行動は無意識に起こるものであり、相手の感情を把握しその真似をするメカニズムとは異なる。以上の問題点を踏まえ、本研究では次のリサーチクエスチョンに取り組む。『記号化された感情の模倣よりもア

クションユニットそのものを模倣する方が、ロボットの共感性を高めるのではないか。』このリサーチクエスチョンを解明することで、ロボットの感情的なサポートを更に提供することができ、人間とより自然なコミュニケーションできるロボットの開発につながる。

2 関連研究

2.1 同調動作に関するロボット研究

山野 (2008) ではロボット KAMIN を用いて、ヒトの情動に同調させる場合と同調させない場合での効果を明らかにするために、1名の実験参加者が1台のコミュニケーションロボットに話しかける形式で実験を行っている。その結果、情動に同調させる方がインタラクティブ時間が長く、ポジティブな印象を持つことがわかった。一方で行動や発話のタイミング、速度の同調手法も課題として挙げられている [4]。

米津 (2017) では遠隔コミュニケーションの課題であ

*連絡先：同志社大学文化情報学部
(京田辺市多々羅都谷 1-3)
E-mail: tiio@mail.doshisha.ac.jp

る遅延やリアクションの遅れを解決するために、通話間の仲介役としてロボットが一方の感情を認識し模倣する実験を行った。その結果、より共感性や安心感を高める存在になり得ることがわかった [6]。

2.2 ロボットの開発

これまでに様々なロボットが開発されているが、近年では人間とそっくりのアンドロイドロボットの開発が進んでいる。石黒 (2019) では自律対話アンドロイド「ERICA」を開発し、人間の曖昧な指示も解釈できるシステムを構築した [7]。人間とロボットが自然に対話をするには、日常対話にあるような他者理解と共感を得ることが必要である。そのため、ロボットが一方的に人間の要求に応えるのではなく、ロボットも行動の意図や欲求を持ち互いに発信することで共生につながると石黒は述べている [8]。

このようなアンドロイドロボットは高齢者などの支援、労働力不足を解消できると期待されている。河本 (2022) ではアンドロイドロボット「I」を旅行代理店のカウンターセールスとして旅行を考えている体験者と対話する実験を行った [9]。その結果、アンドロイドロボットの顔つきや姿勢、笑顔などが好印象につながっていることを確認した一方で、不自然な動作パターンも見受けられ、改善点が明らかとなった。

2.3 本研究の位置づけ

このようにコミュニケーションにおいて表情は非常に重要であり、表情に関する同調効果をロボットで検証する研究は多く行われてきた。しかしこれまでの研究では、基本6感情では表せない感情を認識し再現することは難しい点や、模倣のプロセスが人間同士のコミュニケーションとは異なる点が問題として挙げられる。本研究では表情表出レベルが高いアンドロイドロボット「Nikola」を使用し、アクションユニットを模倣するシステムを構築し共感性を測る。

3 ロボットシステム

本研究ではアクションユニットを模倣するシステムを構築した。本システムではカメラで読み取ったヒトの表情をアンドロイドロボット Nikola にリアルタイムで模倣させることができる。ロボットが発話をしていないときに口が動いているのは不自然であるため、口の動きを除いた。

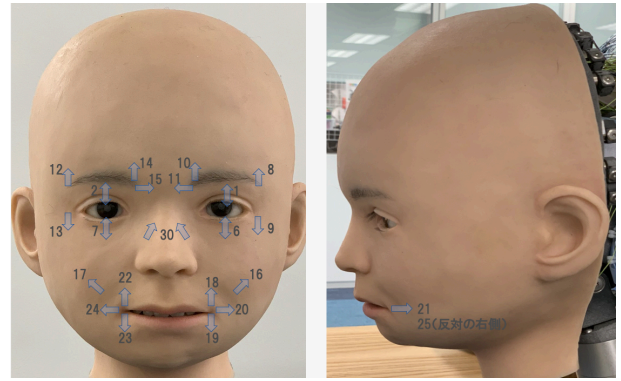


図 1: 本システムで使うアクチュエーターの位置

3.1 使用ロボット

本研究では理化学研究所が開発したヒト型ロボット「Nikola」を使用する。Nikola はヒトの表情筋の動きを緻密に再現することができるアンドロイドロボットである。35 個のアクチュエーターを持ち、空気圧でそれぞれ動かすことでヒトのような表情を作る。本システムでは図 1 の箇所のアクチュエータを使用する。各アクチュエータは 0 から 255 の力で制御する。既に Nikola は図 2 のように、基本感情やウインクなどが表出可能で 14 種類のアクションユニットが再現できる。

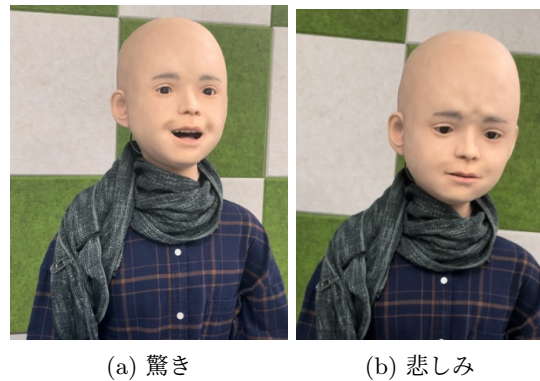


図 2: Nikola の表情の例

3.2 システム構成

図 3 は本システムの全体図である。まず、画像を取得し FaceMesh パッケージを搭載した PC1 に送る。FaceMesh パッケージによって顔を検出し、獲得した座標データから、指定した点のデータのみ抽出する。その後、事前に取得した真顔の画像と比較し、その差を Nikola に送信する値に変換し、実装する。FaceMesh パッケージは MediaPipe[10] が提供しており、2次元顔

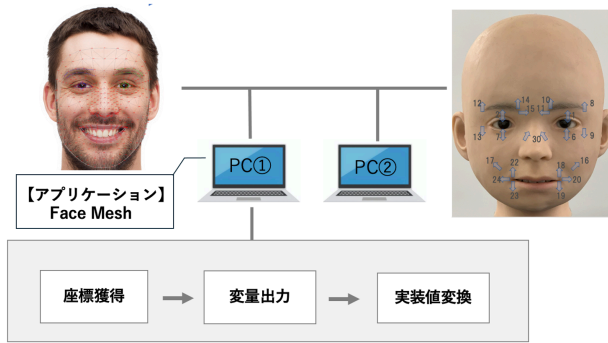


図 3: 本システムの構成図

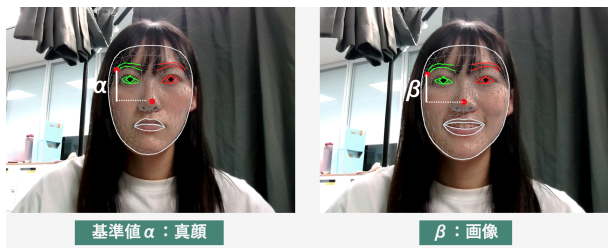


図 4: 変量出力のイメージ図

画像から 3次元ランドマークを推定するプログラムである。画像の左上を原点とし、右方向に x 座標正、下方に y 座標正、奥方向に z 座標正とした座標を 478 箇所取り込む [11]。このパッケージを提供する MediaPipe は Google 社が提供するオープンソースの画像処理 ML ライブラリである。非常に速く動作するため、リアルタイムに実装させたい今回の目的に適している。

3.3 指定したランドマーク

図 4 のように本システムでは真顔の状態と比較し、取得した画像との差 (α と β の差) を送信値に変換する。本節では、図 5 のように各顔の部位を制御するために参照したランドマークについて説明する。

・目

1 番と 2 番は顔の縦幅 (10 と 152 の距離) と目の縦幅 (159 と 145 / 386 と 374 の距離) の比率を用いて目の開口量から決定した。

・眉

顔の中で鼻先が最も表情の動きに影響されないと考えたため、8 番、10 番、12 番、14 番は眉間と鼻先の距離 (107 と 4 / 336 と 4 / 55 と 4 / 285 と 4) を用いて決定する。

・頬

16,17 番は発話における動きの影響をなくすため、顎先と鼻先、頬の 3 点の角度 (152 と 4 と 57 / 152 と 4

と 287) を用いて決定する。

また、6 番、7 番、9 番、13 番、18 番、19 番、20 番、22 番、23 番、24 番は外眉 (70,300)、目 (23,253)、口 (57,287) と鼻先 (4) との距離、21 番と 25 番は口 (57,287) の z 座標を用いて決定する。

・鼻

30 番は顔と鼻の距離 (10 と 217 / 10 と 437) を出力し、その平均から決定する。

3.4 ロボットへの対応づけ

次に前章で指定した距離や角度を、取り込んだ画像と真顔の時の値で比較し、その差をロボットに適応させていく。今回使用するロボット Nikola は各アクチュエーターを 0 から 255 の力で制御するため、各ランドマーク間の最大変形量が 255 になるように変換する。指定したランドマーク間の最大変形量は産業技術総合研究所の表情データベース 8 人分 [12] を用いて定めた。この表情データベースは喜びや悲しみといった基本感情を表す表情が、約 3 秒の動画形式で格納されているが、感情の表現度が非常に高いため、最大変形量を定めるのに適している。8 人の最大変形量の平均値を、本実験におけるランドマーク間の最大変形量と定める。そしてこのランドマーク間の最大変形量を Nikola の 255 の力に対応させる。

4 評価実験

4.1 実験設計

本研究は模倣の方法 (アクションユニット条件・記号レベル条件) を独立変数とする 1 要因参加者内計画で行う。本研究ではアクションユニットを模倣するシステムを開発した。そこでこのシステムの有効性を検証する。そのため、読み取った感情の表情をとる記号レベル条件と比較する 1 要因参加者内計画とした。

4.2 実験参加者

本実験は対面形式で、実験参加者は 19 歳から 22 歳までの 31 名 (男性 18 名・女性 13 名) で行なった。しかし実験初日に行った 8 名は実験条件が整っていなかったため、分析には 23 名 (男性 14 名・女性 9 名) のデータを使用した。実験場所は国立研究開発法人理化学研究所けいはんな地区の一室である。各参加者には人材派遣会社から報酬が支払われた。

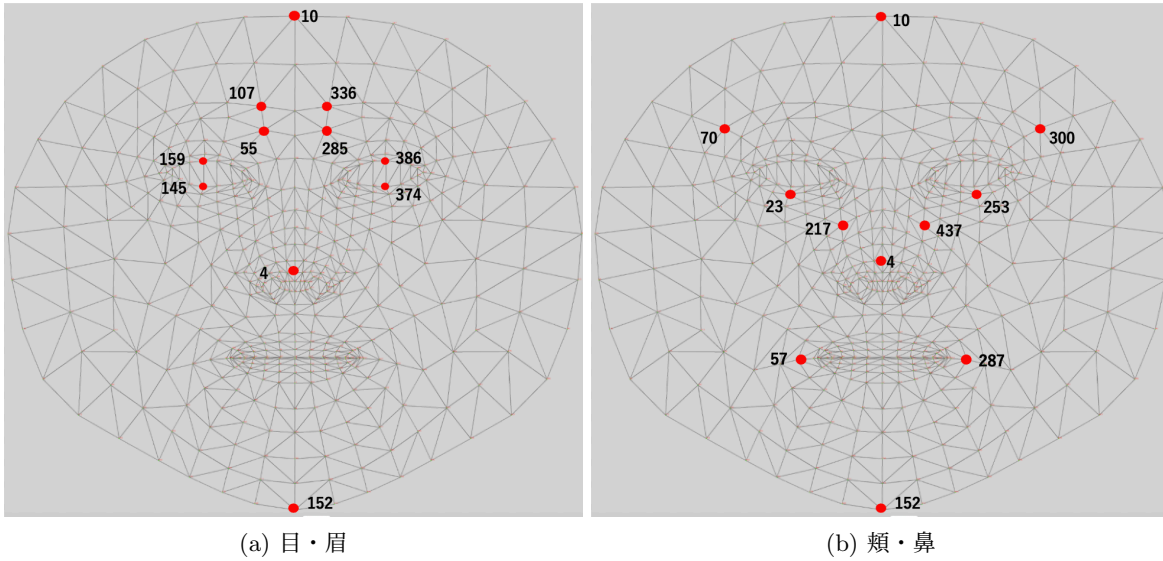


図 5: 指定したランドマーク

4.3 実験環境

図 6 のように実験参加者と Nikola を向き合うようにする。Nikola は元々椅子に座っているため、実験参加者も椅子に座ってもらい、顔が同じ高さになるようにセッティングする。実験参加者と Nikola の間にカメラ 1 を置き、実験参加者の顔を読み取り PC1 と PC2 を経由し Nikola に制御値を送信していく。この時、カメラ 1 は実験参加者目線で Nikola の顔が全て見えるような位置にする。また、プログラム通りに実装されたかを確認するため、被験者の右後ろにカメラ 2、Nikola の左横にカメラ 3 を設置する。実験者は実験参加者に見えないよう別室で待機する。

4.4 実験刺激

本実験では図 7,8 のようにアクションユニット条件と記号レベル条件で比較する。記号レベル条件では Py-feat[12] を使用し実験参加者の感情判定を行う。Py-feat とはアクションユニットや感情、顔のランドマークを検出できる Python のライブラリである。検出する感情は anger・disgust・fear・happiness・sad・surprise・neutral で、その感情の強さを 0 から 1 の値で示す。本実験ではカメラから取得した画像を Py-feat によって解析し、判定結果から最も強いと検出された感情を Nikola の表情に表す。図 8 のように最大感情の強さによって Nikola の表情を変化させた。この時の表情は既存のアクションユニットを参考とし、各アクチュエーターの制御値を確定させた。また、感情条件でも口や首、頭が動かないように統一している。



図 6: 実験環境

4.5 実験手順

実験参加者はまず実験環境に入り、実験の説明を受ける。この時、2 分間 Nikola に話す実験を 2 回行うことを伝え、カメラ撮影の同意も得る。そして実験参加者は Nikola に話す内容を決める。実験参加者が話しやすいように、トピックを 7 個 (「失敗談」「恥ずかしかった話」「悔しかった話」「怒った話」「最近うまくいかないこと」「後悔していること」「滑った話」) 用意した。実験参加者はそのトピックから話を考えてもらう。苦笑いなどなるべく複雑な感情や感情が変化するようなトピックを用意した。Nikola のようなヒト型ア



図 7: アクションユニット条件の様子

ンドロイドロボットに対して、不気味さや不自然さを感じないようにするため、実験参加者は怒りの表情やウインクなど Nikola の 6 種類の顔の動きを 30 秒間見る。次に基本情報や顔ロボットの使用経験などを問う GoogleForm の事前アンケートに回答する。条件の順序はカウンターバランスをとり、実験参加者は条件終了ごとに GoogleForm で作成したアンケートに回答する。両条件のアンケートに回答後、「実験を踏まえてどちらが共感してくれて良いと感じたか。またその理由は何か。」「アクションユニット条件の実験中、自分の顔を模倣されていると気づいたか。」を口頭質問し、実験終了となる。

4.6 測定項目

ロボットとのインタラクションに関する心理状態を測定する項目には 7 段階のリッカート尺度を用いる。また、評価項目は表 1 のように L.ChARRIER(2019) で提案されている RoPE 尺度を参考にして作成した [13]。RoPE 尺度はロボットを対象とした共感を測る尺度であり、ロボットの理解度、反応、フィラー項目の 3 つに分けて質問項目を作成している。(-) がついている項目には -1 をかけ、カテゴリごとに項目の点数を足し合わせたものを評価点数とする。尺度のとりうる最大値はロボットの理解度が 56、反応が 49 で、最小値はロボットの理解度が 8、反応が 7 となる。値が大きくな

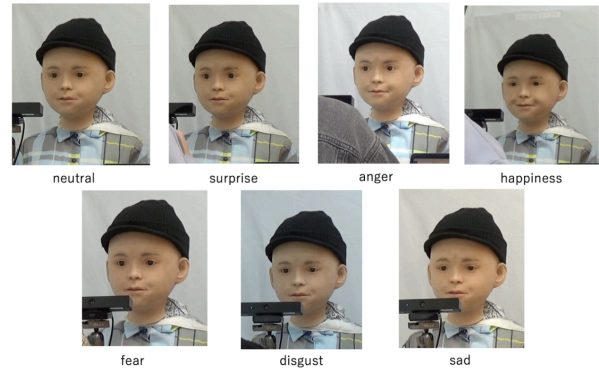


図 8: 記号レベル条件で表出する表情

るほど、ロボットが持つ共感性が高いことを意味する。

4.7 分析方法

本実験では、アクションユニットレベルの模倣は記号化された感情の模倣よりも共感性が高まるかを明らかにするため、アクションユニット条件と記号レベル条件を比較する。そこで印象評価アンケートを有意水準 5% とした対応ありの t 検定にかける。また、口頭質問の回答に関しては実験中の Nikola の動作と照らし合わせて、考察する。

5 結果

5.1 印象評価アンケート

図 9, 図 10 は得られたデータを箱髷図に示したものである。まずロボットの理解度を測る項目では、アクションユニット条件は $\text{mean}=8.78, \text{SD}=6.83$ 、記号レベル条件は $\text{mean}=11.35, \text{SD}=6.87$ となり、 $t=-1.29, p=0.21$ であった。ロボットの反応を測る項目では、アクションユニット条件 $\text{mean}=12.74, \text{SD}=3.66$ 、記号レベル条件は $\text{mean}=12.04, \text{SD}=4.52$ となり、 $t=0.56, p=0.58$ であった。そのため、両項目ともアクションユニット条件と記号レベル条件で有意差はないという結果になった。

5.2 口頭質問

「実験を踏まえてどちらが共感してくれて良いと感じたか」という質問に対して 23 名中 7 名がアクションユニット条件、16 名が記号レベル条件と回答した。その回答について以下のような理由が挙げられた。

アクションユニット条件と回答した人

- ・記号レベル条件の方が話を聞いてくれているのか不信感があったから。

表 1: ロボットの印象についてのアンケート項目

ロボットの理解度
あなた自身が話に対して抱いている感情をロボットは理解している。 ロボットはあなた自身が必要としていることを理解している。 ロボットはあなたの感情を気にかけている。 ロボットはあなたを理解していない。(-) ロボットはあなたの個性を受け入れている。 ロボットはあなたの話を理解している。 ロボットはあなたの話には反応するが、感情を理解していない。(-) あなたが悲しい時や失望している時に、ロボットも悲しく失望しているようである。
ロボットの反応
ロボットはあなたの考えや感情に影響されず、行動に変化がない。(-) ロボットはあなたの話に影響されず、行動に変化がない。(-) ロボットはあなたを励ましてくれる。 ロボットはあなたが必要な時に助けてくれる。 ロボットはあなたが話すタイミングを察してくれる。 ロボットの反応が固定化されているため、あなたはロボットに通じない。(-)
フィルター項目
ロボットの行動は自然に感じられる。 ロボットは自分(ロボット)は何をしているのか知っている。 ロボットは自分(ロボット)の行動に責任を持っている。 ロボットと対話する時に不安を感じる。 ロボットに楽しく話せた。

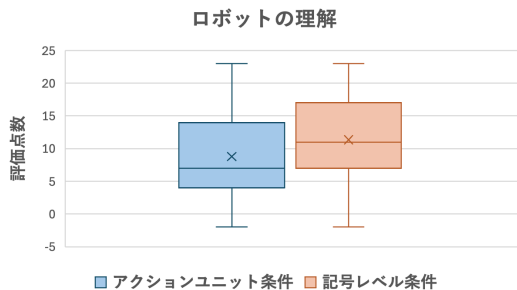


図 9: ロボットの理解度に関するデータ

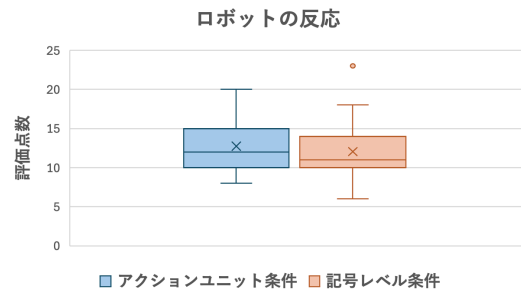


図 10: ロボットの反応に関するデータ

- ・記号レベル条件に比べて表情の変化があったから。
- ・話を聞いてくれている感じがしたから。

記号レベル条件と回答した人

- ・表情が分かりやすいから。
- ・アクションユニット条件は常に動いているが、記号レベル条件は表情の変化があって応答してくれている感じがしたから。

・暗い話をしている時に笑ってくれるのが良かったから。

「アクションユニット条件の実験中、自分の顔を模倣されていると気づいたか」という質問に対しては23名

中6名が「気づいた」、1名が「なんとなく気づいた」、16名が「気づかなかった」と回答した。

6 考察

今回の実験結果から、アクションユニットレベルの模倣と記号化された感情の模倣の共感性に数値的な差は見られなかった。しかし「どちらが共感してくれて良いと感じたか」という質問に関していずれの手法も一定数の支持を得られた。

実験結果がこのように分かれた理由は、記号レベル条件における表情変化の回数だと考えている。記号レベル条件を支持した人は、表情変化の回数が平均 17.2 回と多く、表情のわかりやすさが評価の高さにつながった。しかし表情変化の回数が 20 回以上と非常に多くなると、話の内容と異なる表情をとることがありロボットに不信感を抱いたと考える。一方でアクションユニット条件を支持した人は表情変化の回数が 7 回以下と少なく、アクションユニット条件の豊かな表情が評価の高さにつながったと考える。

従って、アクションユニット条件のメリットは常に表情の変化があることであり、デメリットは時折表情が不自然になることが挙げられる。一方で記号レベル条件のメリットは表情の変化がわかりやすいことであり、デメリットは表情が話の内容と違う時があることや分析に時間を要することが挙げられる。そのため、場面に応じた使い分けが期待できる。

7 結論

本研究では、アクションユニットレベルの模倣が記号化された感情の模倣よりも共感性が高まるかを明らかにするため、アクションユニットを模倣するロボットシステムを構築し、そのシステムの有効性を検証した。具体的には、構築したシステムと Py-feat の判定結果をロボットの表情に表出させるシステムの比較を行った。今回の実験を踏まえ、アクションユニットレベルの模倣と記号化された感情の模倣の共感性に数値的な差は見られなかった。しかし両条件とも一定数の支持は得られたことから、場面に応じた効果的な使い分けが有効であることが示唆される。しかし今回の実験設定では、ロボットの取る表情よりも表情の変化から印象を形作られた場面もあった。そこで今後はシステムの滑らかさや実験設定を更に改善した上で、具体的な使用場面や効果を明らかにしていく必要がある。

謝辞

本研究は JSPS 科研費 24K21327 の支援を受けたものです。

参考文献

- [1] 齋藤 千夏・高橋 英之・岡田浩之. 共同注視と模倣によるコミュニケーションロボットとの関係性の変容. 人工知能学会全国大会論文集第 25 回, 2011.
- [2] 鈴木啓太・横山正典・吉田成朗・望月崇由・布引純史・鳴海拓志・谷川智洋・廣瀬通孝. 同調的な表情変形技術を用いた遠隔コミュニケーションの拡張. 情報学会論文誌, Vol. 59, No. 1, pp. 52–60, 2018.
- [3] 齋藤千夏・高橋英之・岡田浩之・金岡利知・渡辺一郎・玉川大学・富士通研究所. 無意識的同調がロボットへの親和性に与える影響. 2012 年度人工知能学会全国大会論文集, 2012.
- [4] 山野美咲・薄井達也・橋本稔. 音声情動認識による人間とロボットのコミュニケーション. 自動制御連合講演会講演論文集第 51 回自動制御連合講演会, 2008.
- [5] 吉田伶司・黒野侑哉・棚橋優・松本幸大・菅谷みどり. 感情を理解するロボットの実現へ向けて. 組込みシステムシンポジウム 2018 論文集, 2018.
- [6] 米津壮二・大澤博隆. テレプレゼンスロボットにおける同調動作の効果. 人工知能学会全国大会論文集第 31 回, 2017.
- [7] 石黒浩・港隆史・小山虎. 意図欲求を持つ自律対話アンドロイドの研究開発. 日本ロボット学会誌, Vol. 37, No. 4, pp. 312–317, 2019.
- [8] 石黒共生ヒューマンロボットインタラクションプロジェクト国立研究開発法人科学技術振興機構.
- [9] 河本真琴・河窪 大介・杉山 弘晃・酒造 正樹・前田英作・東京電機大学・NTT コミュニケーション科学基礎研究所. 共生社会におけるアンドロイドロボットのためのマルチモーダル対話戦略の構築. 人工知能学会全国大会論文集第 36 回, 2022.
- [10] J. Nash H. Grundmann M. Lugaresi. Mediapipe: A cross-platform framework for building multimodal applied machine learning pipelines. github. 2019.
- [11] 問世田秀・赤松茂. Mediapipe で生成した顔点群に対する conditionalvae を用いた表情操作. 法政大学大学院紀要 理工学研究科編, Vol. 64, , 2023.
- [12] H. Fujimura, T. Umemura. Development and validation of a facial expression database based on the dimensional and categorical model of emotions. *Cognition Emotion*, Vol. 32, No. 4, pp. 1663–1670, 2018.
- [13] Laurianne Charrier・Alisa Rieger・Alexandre Galdeano・Amélie Cordier・Mathieu Lefort・Salima-Hassas. The rope scale: a measure of how empathic a robot is perceived. *14th ACM/IEEE International Conference on Human-Robot Interaction*, 2019.