

多人数対話音声の感情情報を考慮する対話システムの開発

Development of a dialogue system that takes into account emotional information

in multi-people dialogue speech

原田 和麻¹ 片上 大輔¹

Kazuma Harada¹, Daisuke Katagami¹

¹ 東京工芸大学工学部

¹ Faculty of Engineering, Tokyo Polytechnic University

Abstract: 本研究では、多人数が参加する対話場面に対して適切な介入を行う対話エージェントを作ることを目指し、対話参加者の声に現れる感情を考慮して動作する多人数対話システムを開発した。個々の対話参加者の識別情報、声から推定される感情を組み込んだ対話履歴を保存する仕組みを作り、対話履歴をもとに文脈と感情を考慮して動作する対話エージェントを構築した。また、提案システムが感情を考慮することの効果や、対話エージェントが対話中における立場の違いが他の対話参加者に与える印象の変化を調査した。

1. はじめに

従来、人とコンピュータが自然言語を用いて対話を行うことは絵空事だと思われてきた。しかし、音声認識技術やディープラーニングなどといった近年の目覚ましい技術発展により、音声を通じた対話、自然言語による連続した対話を実装したエージェントは現実のものとなっている。

音声を通じた対話エージェントのサービスとして有名なものでは、Amazon Alexa (図 1 左) や Google Home (図 1 右) をはじめとするスマートスピーカが挙げられる。スマートスピーカの動作は基本的に一問一答形式であり、連続した対話こそ難しいものの、言葉だけで IoT 機器の操作や情報の検索を行うことができる。また、自然言語による連続した対話を行うことができるサービスとしては、ChatGPT (図 2) が存在する。スマートスピーカとは異なり IoT 機器の操作こそできないが、対話経緯を考慮し連続した対話を行うことができる。

しかし、現在サービスとして提供される対話エージェントの多くに共通した問題点として、3 名以上による日常対話に介入することを考えていない点が挙げられる (表 1)。例として tender[1] というサービスがある。tender は会議全体の進行を自動で行ったり、議論の場で発言者を指名したりと、議事進行として様々なふるまいをすることができる。しかし動作のためには、あらかじめ進行やテーマに関するルールを定義しておく必要があり、同じ話し合いでも



図 1 スマートスピーカ

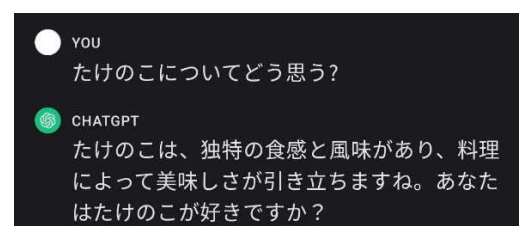


図 2 ChatGPT との対話

家庭内の所謂家族会議に利用するには少しハードルが高い。また、ユーザに対する自発的な呼びかけや一対一の日常対話などをこなすことができる、スマートスピーカの Romi (図 3) [2] というものがある。しかし Romi は、Romi 以外に向けられる発話の存在を対話経緯の一部として利用することを想定していないつくりとなっており、多人数対話への十分な対応がなされているとは言えない。

そこで本研究では、音声から推定される感情情報



図 3 Romi[2]

を組み込んだ多人数対話ログを大規模言語モデルに受け渡すことで、多人数対話場面で感情を考慮し発話による介入を行う対話システムを提案する。

なお、2名以上のユーザが対話している場面においてエージェントが発話を行う際、ユーザの発話が終わったことを確認してからエージェントが発話する「介入」、ユーザの発話が終わることを待たずに発話を割り込ませる「割り込み」の2つが存在するとされる[3]。本研究におけるエージェントの発話では、このうち介入のみを扱うものとしている。

2. 関連研究

2.1 対多人数対話システム

対話エージェントサービスの多くは、エージェント以外に向けられた発話を受け取らないように設計されている。例として、Alexa や Siri などに代表される音声アシスタントの場合、ウェイクワードと呼ばれる特別な言葉を命令の前につけ、音声アシスタントに対する呼びかけであることを明示しなければ動作しない。また、ChatGPT をはじめとした各種チャットボットも、ユーザの入力のすべてをボットに対する呼びかけとして解釈し動作する。

これらの特徴を持つエージェントは、連続しない命令や質問の投げかけや、エージェント以外への発話が発生しえない一対一の対話場面においては問題なく動作する。しかし、多人数対話場面においては、対話の経緯として重要な意味を持つ発話のすべてがエージェントに向けられたものであるとは限らない。そのため、エージェント以外に向けられた質問に勝手に答えてしまう、対話経緯をうまく把握できないなどの問題が生じる。

黄ら[3]は、ガイドエージェントによる対話への介入が必要な場面の定義、システムによる自動検出の2つを目的として、観光案内をはじめとした人-エー

表 1 対話エージェントサービスの現状

サービス名	日常生活への導入	多人数対応	非タスク志向
Amazon Alexa	○	△	×
Google Home	○	△	×
ChatGPT	×	×	○
tender	×	○	×
Romi	○	△	○

※ 可：○ 不可：× 運用次第、不十分：△

ジェント間の対話コーパスの収集と分析、自動検出する仕組みの開発を行った。黄らの研究では、対話中の言語情報、ユーザ間の向き合い方をはじめとするカメラから取得される各種特徴量を利用した分類を検証している。検証の結果として、どのような種類の介入をするかを判断する場面においては言語情報の有効性が高く、いつ介入するかというタイミングについては非言語情報の有効性が高いという結果が得られている。

2.2 感情情報を考慮する対話システム

従来、対話システムの多くはユーザから受け取った言語情報のみを考慮して発話判断を行うシステムが多かった。しかし、実際の人間同士の対話においては、言語情報だけではなく非言語情報を考慮したうえでお互いの発話や意思疎通がなされることが大半である。近年では大規模言語モデルの発展により言語情報のみに基づいた高度な返答が可能となっている背景もあり、非言語情報を考慮することのできる対話エージェントの研究が進みつつある。

Tang らの研究[4]では、音声の音響特徴量からの感情推定と音声認識によって振る舞いを決定する図 4 に示すような構造のシステムを製作し、動作に対する第三者評価実験を行っている。実験の結果、言語情報だけで振る舞いを決定する表 2 に示すような振る舞いと比べ、感情推定の機能を利用して振る舞いを調整した表 3 に示すような振る舞いのほうが、より人間に近く、より快に寄った印象を与えていると評価された。

2.3 大規模言語モデル (LLM)

人が普段用いている自然言語をコンピュータが理解できる形に落とし込もうとする研究は、遅くとも 2000 年代の初めにはすでに試みられていた[5]。しかし近年になって、コンピュータの高性能化や深層学習をはじめとする新しい技術の登場により、世間一般にその存在が普及するほどの急速な発展を遂げている。

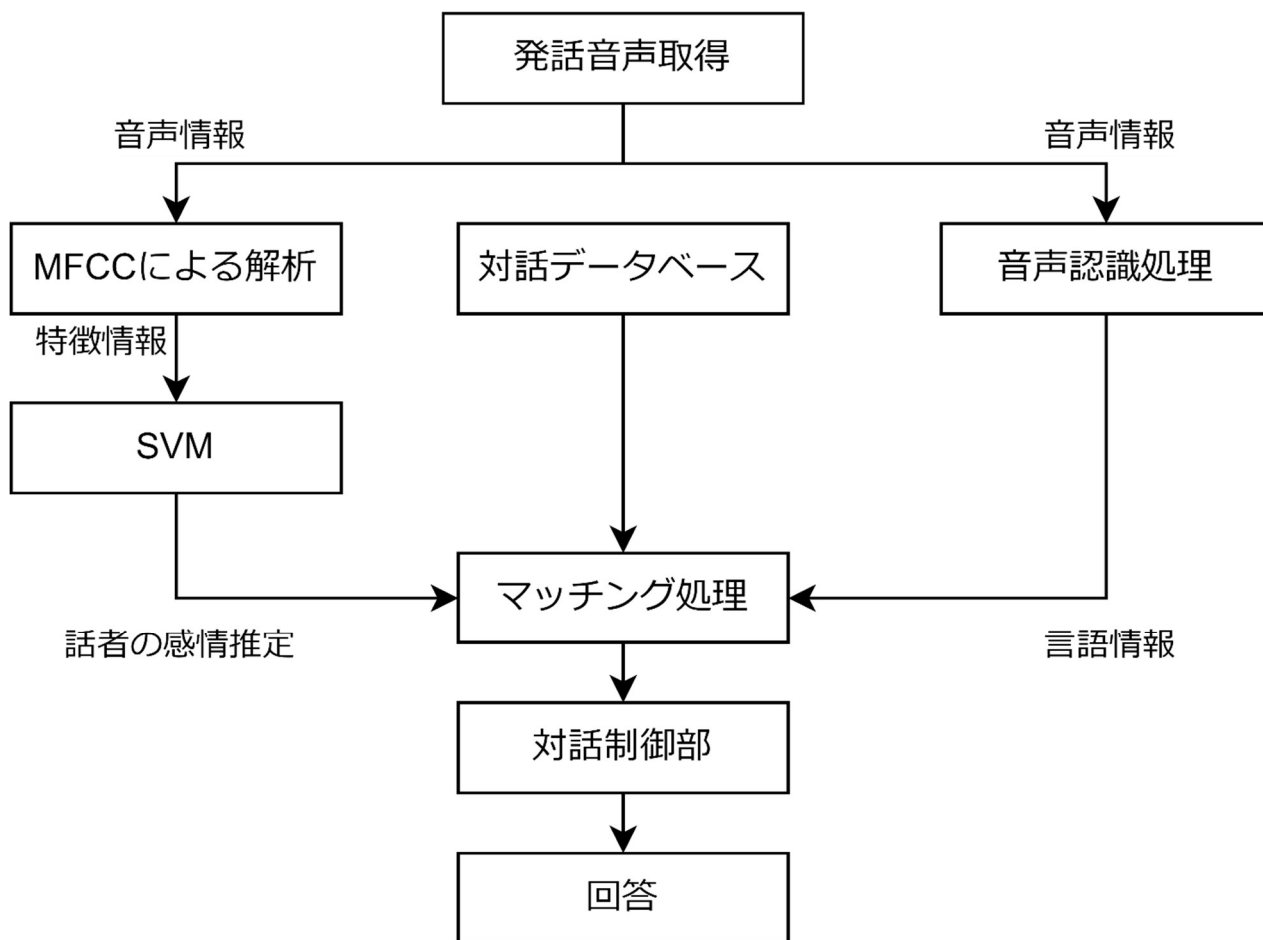


図 4 感情を考慮する音声対話システムの概要図

表 2 感情を考慮しない応答例

発話者	発話内容
システム	こんにちは
ユーザ	こんにちは
システム	元気ですか？
ユーザ	元気だよ（元気がなく）
システム	それはよかった。今日も一日頑張りましょう。

表 3 感情を考慮する応答例

発話者	発話内容
システム	こんにちは
ユーザ	こんにちは
システム	元気ですか？
ユーザ	元気だよ（元気がなく）
システム	本当ですか？声に元気がないですよ
ユーザ	では、あなたの好きな曲をおかけしましょう

表 4 IoT 機器の機能

機器	機能
ライト	点灯 色の変更 状態の取得
温湿度計	温度の計測 湿度の計測
ロボット掃除機	掃除の開始・終了
ファンヒータ	暖房 送風

ChatGPT は、Open AI 社の所有するさまざまな大規模言語モデルをチャット形式で利用できるサービスであるが、GPT-4 などを含む現在の大規模言語モデルの多くは同社の Radford ら[6]による研究に基づいて開発されている。Radford らは、それまでの言語モデル開発に必要とされていた学習データの人力によるラベル付けの負担を大幅に軽減し、それまで困難とされていた大規模な学習による汎用的な言語モデル作成へのハードルを大幅に引き下げた。

現在では、汎用的な性能を獲得した大規模言語モデルを活用する動きが広がっている。例として、橋

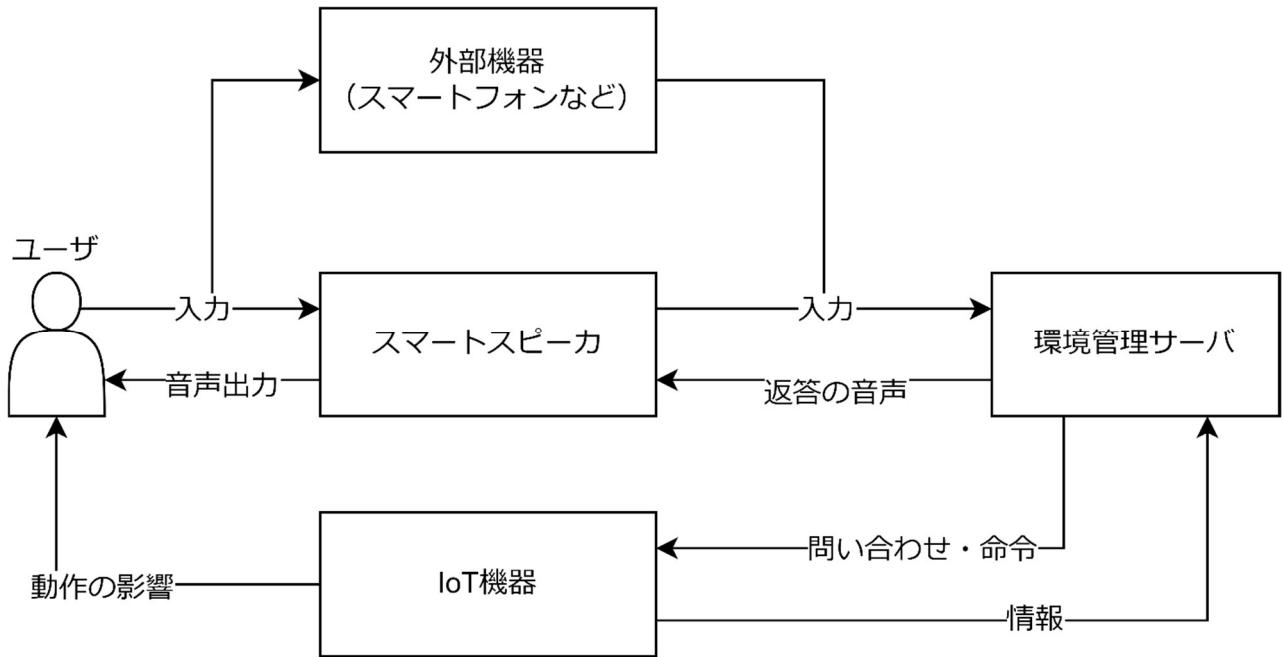


図 5 動作想定環境概要図

本らによる就職支援の研究[7]が挙げられる。現代日本における就職活動では、就職後の会社と就職者のミスマッチを防ぐことを目的として自己分析を行うことがある。しかし、自己分析にはワークシートなどを用いても高いメタ認知が必要とされ、個々のユーザや状況に応じたカスタマイズもなされないため非常に困難とされていた。橋本らはこの問題に対して、大規模言語モデルによって価値観を揺るがすような質問を投げかけることにより、ユーザに自身の価値観と向き合って思考する機会を与えるシステムを開発した。検証実験の結果として、システムの動作には、ユーザのキャリア意識や自己理解を向上させる効果が確認され、長期的なキャリア視点やその方向性を明確にする助けとなることが確認された。

3. 動作想定環境の構築

3.1 環境の概要

本研究を進めるに先立ち、対話エージェントの実装先として想定される環境の構築を行った。図 5 は動作環境の概要図である。環境にはスマートスピーカ、環境管理サーバ、IoT 機器が設置されている。スマートスピーカ、センサ機能を備えた IoT 機器、その他通信が可能な外部機器から入力が発生すると、必要に応じて環境管理サーバがユーザに対するレスポンスの決定と命令を行い、スマートスピーカによる発話と IoT 機器の動作を行う。

3.2 スマートスピーカ

環境内には Amazon の Alexa Echo Show 10 (以降 Echo) 2 台および自作したスマートスピーカ 1 台を設置している。Echo はマイクによる入力の受け付け、ディスプレイによる視覚情報とスピーカによる音声による出力の機能を備えている。自作したスマートスピーカは 8 つのマイクによる多人数からの同時音声入力の受け付け、スピーカによる音声出力の 2 点を備えている。Echo は Python による限定的な機能拡張、自作のスマートスピーカは内蔵された Raspberry Pi のカスタマイズによる広範な機能拡張が可能である。

3.3 環境管理サーバ

環境管理サーバはネットワークによってスマートスピーカおよび IoT 機器と接続されているほか、スマートフォンなどの外部機器からの入力を受け付ける機能を備えている。Windows OS によって動作しており、接続する機器からの入力などに応じて判断と命令の送信を行う機能を搭載できる。

3.4 IoT 機器

動作想定環境中には各種の IoT 機器が設置されている (表 4)。温湿度計が設置されているほか、ライトなどの各種 IoT 機器からは現在の動作状況などを取得することができる。また、環境管理サーバなどから命令を送信することによって遠隔操作が可能で

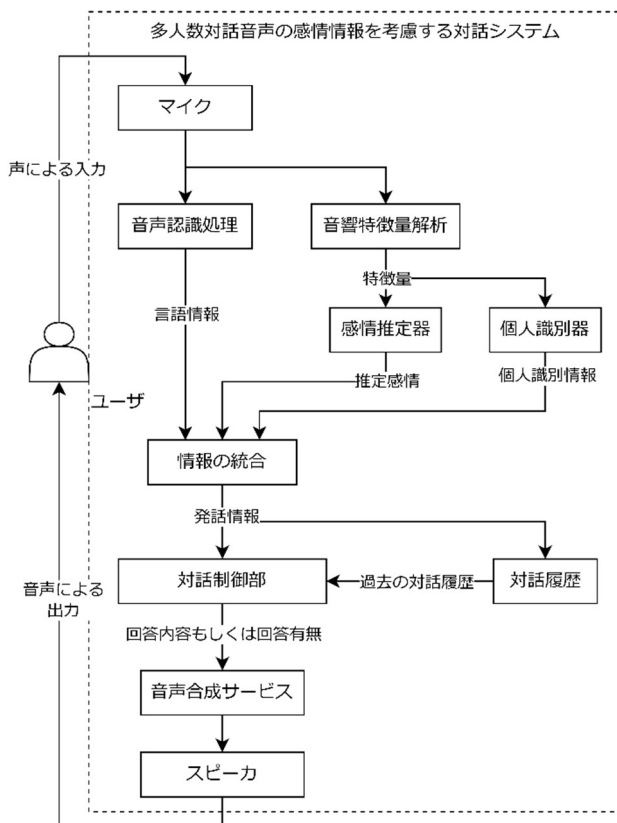


図 6 提案システム概要図

ある。

4. 多人数対話音声の感情情報を考慮する対話システム

4.1 提案システム概要

マイクが音声の入力を受け取ると、音声入力に対して音声認識、音声の特徴量 (表 5) と素性値 (表 6) の取得とそれに基づく音響的特徴 (表 7) の計算を行う。音響的特徴の各種数値は感情推定器及び個人識別器に受け渡され、推定感情と個人識別情報を音声認識結果と合わせて対話制御部に受け渡される。対話制御部は受け取った情報をもとに発話可否と発話内容を生成し、合成音声としてユーザに向け出力する。

4.2 感情推定

感情推定器では、Tang らの研究[4]で利用されていた、表 7 の各種数値をサポートベクターマシン (以下 SVM) で分類する手法を想定する。SVM の学習には感情評定値付きオンラインゲーム音声チャットコーパス[8]を利用し、Plutchik[9]の 8 感情に平静を

表 5 音声から取得する特徴量

特徴	説明
RMSenergy	音量の二乗平均平方根値
MFCC	1 次～12 次メル周波数ケプストラム係数
voiceProb	その時点での音が声である確率
pcm_zcr	波形のゼロ交差率 (および上記各特徴量の一次微分)

表 6 音声から取得する素性値

素性値	説明
max	データ中の最大値
min	データ中の最小値
maxPos	最大値を出力した位置
amean	算術平均
linregc1	線形近似の勾配度
linregc2	線形近似のオフセット

表 7 感情推定に有効な音響的特徴

感情推定の有効性順位	音響的特徴
1	$\Delta RMSenergy - max$
2	$RMSenergy - linregc2$
3	$MFCC[1] - linregc2$
4	$\Delta MFCC[10] - max$
5	$\Delta MFCC[10] - min$
6	$\Delta MFCC[8] - linregc1$
7	$\Delta MFCC[9] - linregc2$
8	$\Delta voiceProb - maxPos$
9	$\Delta MFCC[12] - linregc2$
10	$\Delta MFCC[4] - linregc1$
11	$ZCR - amean$

表 8 感情分類と記号

	分類名	記号
1	受容	ACC
2	怒り	ANG
3	期待	ANT
4	嫌悪	DIS
5	恐れ	FEA
6	喜び	JOY
7	悲しみ	SAD
8	驚き	SUR
9	平静	NEU

加えた、表 8 の 9 感情に分類する。

4.3 個人識別

表 9 対話履歴の例

識別情報	発話内容	推定感情
User A	渋谷 昨日の宿題 今日持ってくる約束だったよね	ANG
User B	あ 忘れてました	NEU
User A	忘れてましたじゃないよ	ANG
User A	昨日あれほど言ったじゃないか	ANG
System	渋谷、謝るべきだよ。	なし
User B	すみません 昨日は忙しくて	NEU

表 10 発話可否と発話条件

	状況	発話条件
1	絶対に発話すべき	無条件で発話
2	発話することが望ましい	直前4回の履歴にシステム発話が含まれない
3	どちらともいえない	直前4回の履歴にシステム発話が含まれない
4	発話することが望ましくない	直前4回の履歴にシステム発話が含まれない
5	絶対に発話してはならない	発話を行わない

提案システムの個人識別には、表 5、表 6、表 7 の各種数値をもとに、i-vector という手法によってあらかじめ学習しているユーザに分類する手法[10]を用いる。この手法は、メル周波数ケプストラム係数（以降 MFCC）を入力とすることで動作する点で感情推定に用いる技術との相性がよく、同時に様々な研究によって高い実用性が認められている技術でもある[11]。

4.4 対話履歴の管理

対話履歴には、ユーザによる発話とエージェントによる発話が表 9 に示すような形で、発話が終了した順に格納される。ユーザ発話が発生した際は、音声認識、感情推定、個人識別のすべてが完了した時点で、発話者の識別情報、発話内容、推定感情の3点が格納される。エージェントが発話する際は、発話が成功したと判断され次第、エージェントであることを示す識別情報、発話内容の2点が格納される。

4.5 発話可否の判断と発話生成

対話履歴にユーザによる発話の情報が入力される

と、対話制御部は発話可否の判断と発話生成を行う。発話可否と発話内容の生成には、Open AI 社[12]の GPT-4-turbo（以降 GPT API）を利用している。

多人数対話場面においては、エージェントが発話すべきと考えられる状況から、エージェントによる発話を避けてユーザ自身の発言を待つべき状況まで様々な状況がある。そこで本研究においては、「絶対に発話すべき状況」から「絶対に発話すべきでない状況」までの5段階の状況（表 10）があるものとし、GPT API は判断基準と対話履歴をもとに判断を行う。また同時に、絶対に発話すべきでない状況と判断された場合を除き、判断条件などに基づいて適切だと考えられる発話内容を生成する。

発話内容の生成までが完了したあと、状況ごとに指定された発話を行う条件を満たしていた場合、生成された発話内容を合成音声としてスピーカから出力することで、ユーザから入力を受けた際の動作が完了となる。

5. 提案システムの対話制御部動作に対する第三者評価実験

5.1 実験概要

本研究では、多人数対話におけるエージェントの発話が、エージェント自身を含む対話参与者への印象向上に寄与することを目的としている。そこで、多人数対話へのエージェントの介入における感情推定の効果、およびエージェントの振る舞いによるエージェント自身と他の会話参与者に対する印象変化の2点を調査することを目的とし、対話制御部の動作に対する第三者評価実験を行った。なお、本実験は東京工芸大学の研究倫理審査委員会の承認を得て実施された(承認番号：倫 2023-20)。

5.2 実験手順

本実験では、システム動作に基づく動画作成、第三者による印象評価アンケート、アンケート結果の集計・分析の3つを行った。

まず、2名のユーザの全15発話からなる識別情報、発話内容、推定感情の履歴を用意した。次に、用意した対話履歴を発話ごとに順次対話制御部に入力し、その出力を記録する作業を、感情推定機能とエージェントの立場に関する条件ごとに行った。その後、対話履歴とエージェントの動作結果を組み合わせ、音声対話の動画にした。なお、システム動作に含まれない叱責者と被叱責者の声については、声優に依頼し感情を演技する形で音声を作成している。



図 7 映像における叱責者の発話



図 8 映像における仲裁者の発話



図 9 映像における非叱責者の発話（画像中の数字はアンケート確認用）

表 11 対話参加者の一覧

役名	立場	備考
緑沢	叱責者	終始怒っており、声に怒りの感情が入っている。
洪崎	被叱責者	反省しておらず、声は常に平静である。
鎌場	仲裁者	合成音声での発話となっている。

動画に対する評価アンケートは、クラウドソーシングを行うことができる 18 歳以上を対象とする。実験参加者にはまず事前アンケートを受けてもらい、その後動画を見ながら事後アンケートに回答しても

表 12 エージェントの動作条件

記号	立場	感情推定
A (被叱・感情)	被叱責者に寄る	ある
B (被叱・なし)	被叱責者に寄る	ない
C (中・感情)	中立	ある
D (中・なし)	中立	ない
E (叱・感情)	叱責者に寄る	ある
F (叱・なし)	叱責者に寄る	ない

らった。事前アンケートでは年齢と性別、叱責に関する過去の経験を回答してもらった。事後アンケートでは、動画中の発言について選択式の印象評価を行ってもらった。

最後に、条件ごとにユーザによる印象評価を集計・分析し、条件間での比較を行った。なお、エージェントを含む各対話参加者の発話内容に限って印象評価を行ってもらうために、動画中の映像は発話内容と話者のわかる文字と図形のみで構成されている（図 7, 8, 9）。

5.3 場面設定

本実験で用いられる対話場面は叱責の場面であり、叱責を行う担任教師（以降叱責者）、叱責を受ける生徒（以降被叱責者）、および仲裁を行う副担任教師（仲裁者）の 3 名で構成される（表 11）。被叱責者は宿題忘れの常習犯であり、昨日も宿題を忘れてきた。叱責者はこのことを重く受け止め、必ず宿題を持ってくるように約束したが、約束の時間になっても宿題は提出されなかった。動画は叱責者が宿題の未提出を指摘する場面から始まり、叱責者と被叱責者の会話に適宜仲裁者が入り込む形で 1 分程度続く。

なお、対話参加者の役名については、同じ苗字を持つ実験参加者への心理的影響を考慮し、集中を阻害するような不自然さがなく、なおかつ実在しない苗字を使用している。

5.4 動作条件

提案システムは本実験中、感情推定機能に関する 2 条件、エージェントの立場に関する 3 条件を組み合わせ、全 6 条件（表 12）で動作を行う。条件間で実験参加者の重複はない。

感情推定の有無は、対話履歴から対話制御部へのプロンプト入力となされる際に、感情に関する情報を組み込むかどうかで変更しており、その他のプロンプトには変更を加えていない。エージェントの立場に関する条件は、対話制御部から GPT API への入力となされる際のプロンプトを一部書き換えるものであり、エージェントが中立の立場、もしくははず

表 13 評価に用いた形容詞

分類	項目
好感因子	外交的 真似したい 格好いい 感じがいい 仲良くなれそう また会いたい 優しそう 共感が持てる 気が利く
性格的要因	真面目そう おとなしい やんちゃ 不真面目
機械・生物	機械らしい 生物らしい
その他	楽しめそう イライラする 規則的

れか一方の対話参与者に与するような立場からの発話を行うように設定を行う。

なお、本実験においては対話の開始から終了に至るまでに動作条件を変更する実験は行わない。また、エージェントの発話に変化したことによる印象の変化を観察することが目的であるため、エージェントを除く対話参与者の発話はすべての条件で同一である。

5.5 評価基準

本実験では高吉らの研究[13]を参考にし、エージェントを含む対話参与者3名に対する各18項目(表13)の7件法による評価を行う。全9項目からなる好感因子は対話参与者に対する親しみやすさと強い関連性があり、すべての対話参与者に対する好感因子の向上が見られるエージェントのふるまいを探ることが本実験の目的である。4項目の性格的要因は実験参加者が対話参与者に対して抱く性格的な印象であり、場面に応じたふるまいの適切性を判断する一材料となりうる。

5.6 実験結果

本実験では、クラウドソーシングによる募集により、1条件あたり20名の全120名が評価を行った。集計したアンケート中の印象評価について、動作条件に基づく6群間の差について Kruskal-Wallis 検定および Steel-Dwass 検定を行い、実験参加者の過去の

経験に基づく2群間の差について Mann-Whitney の U 検定を行った。

図 10, 11, 12 は、動作条件ごとの対話参与者に対する印象評価の平均値をグラフ化したものであり、Steel-Dwass 検定により有意差の認められた項目に印をつけている。全体の傾向として、条件により対話参与者に対する印象が一変するような影響は見られなかった。対話参与者ごとに見ると、叱責者、被叱責者への評価には条件ごとの有意な差はみられなかった ($p>0.05$)。一方で、仲裁者への評価では、「仲良くなれそう」の項目における条件 B (被叱・なし) と条件 D (中・なし) の間、「また会いたい」の項目における条件 C (中・感情) と条件 D の間に有意な差があり ($p<0.05$)、いずれも条件 D (中・なし) のほうが低い評価を受けていた。

過去の経験に基づく比較では、表 14 に示す項目で有意な差が確認された ($p<0.05$)。「当時は納得できなかったが現在では納得できる理由の叱責」を受けた経験については、あると答えた実験参加者の仲裁者に対する「格好いい」の評価が高くなっている。

「理由の不明な叱責」を受けた経験については、被叱責者に対する「やんちゃそう」という印象評価が経験あり群が低くなっており、「真面目そう」という印象は経験あり群が高くなっている。理不尽な理由による叱責をした経験の有無では、仲裁者に対する「イライラした」という項目で経験あり群が高かった。

5.7 考察

6群間の有意な差がほとんど出なかったことから、本実験における仲裁者の発話の違いはいずれの対話参与者に対しても印象の変化が生じるほどのものではなかったことが示された。原因としては、動画中における仲裁者の発言が2回から4回程度しかなく印象を持てるだけの情報がなかった、プロンプトの変更によるシステム動作の変化が小さかった可能性がある。また、仲裁者に対する印象の変化と比べ、叱責者および被叱責者に対する印象の変化が小さかったことから、仲裁者の発話そのものが叱責者や被叱責者の印象に与える影響は限りなく小さいことが示唆された。実験参加者の過去の経験に基づく比較では、叱責を受けた時点で納得のいかなかった経験や理由の不明な叱責を受けた経験がある人物は、それぞれの経験がない人と比べ仲裁者と被叱責者に対して好意的な印象や真面目そうという印象を受けていた。このことから、過去に受けた叱責の経験のあるユーザが、被叱責者や被叱責者をかばうような仲裁者に対してより好意的になる可能性がある。

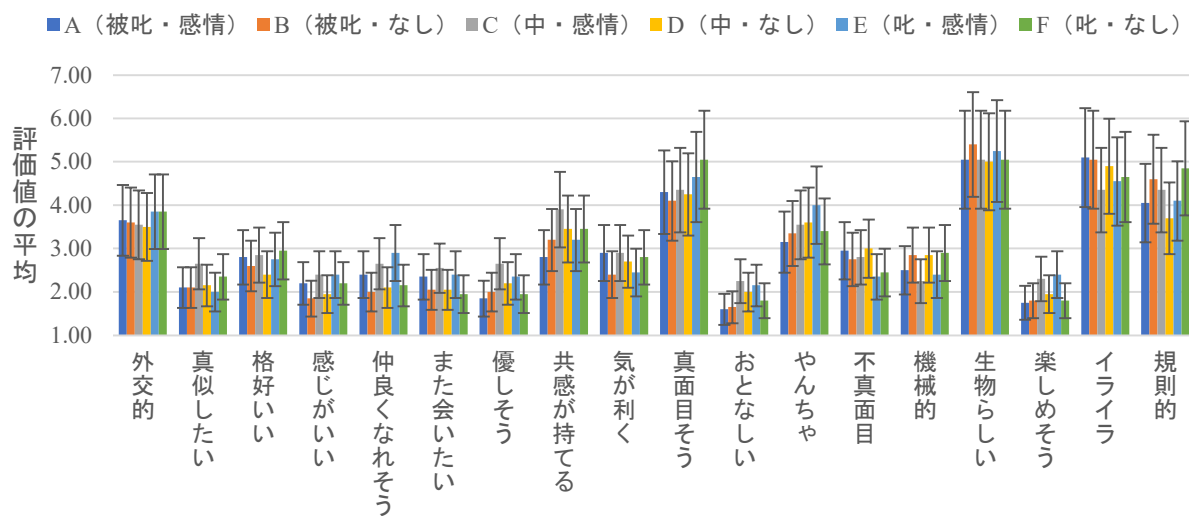


図 10 叱責者に対する印象

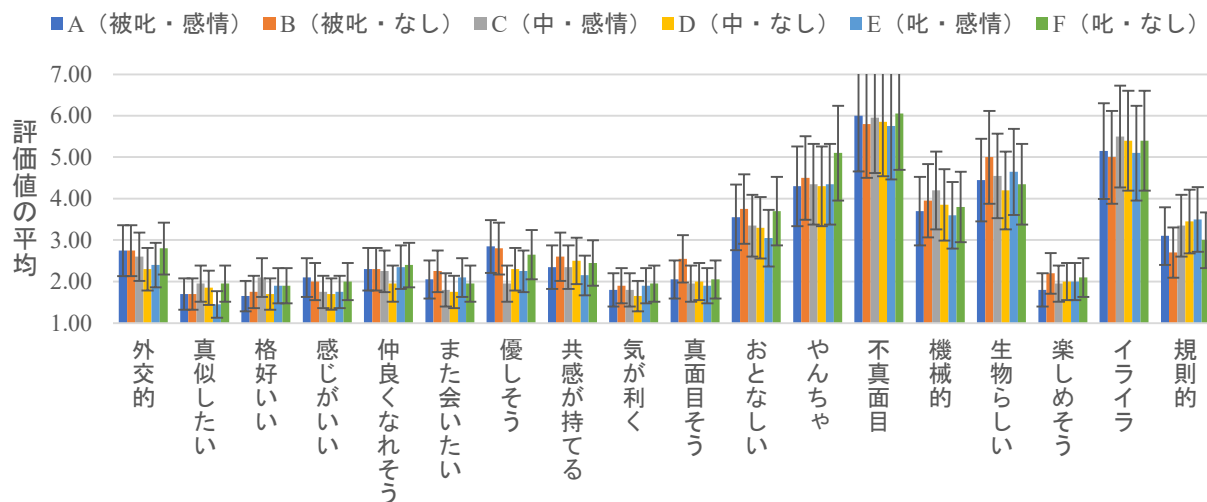


図 11 被叱責者に対する印象

* : $p < 0.05$

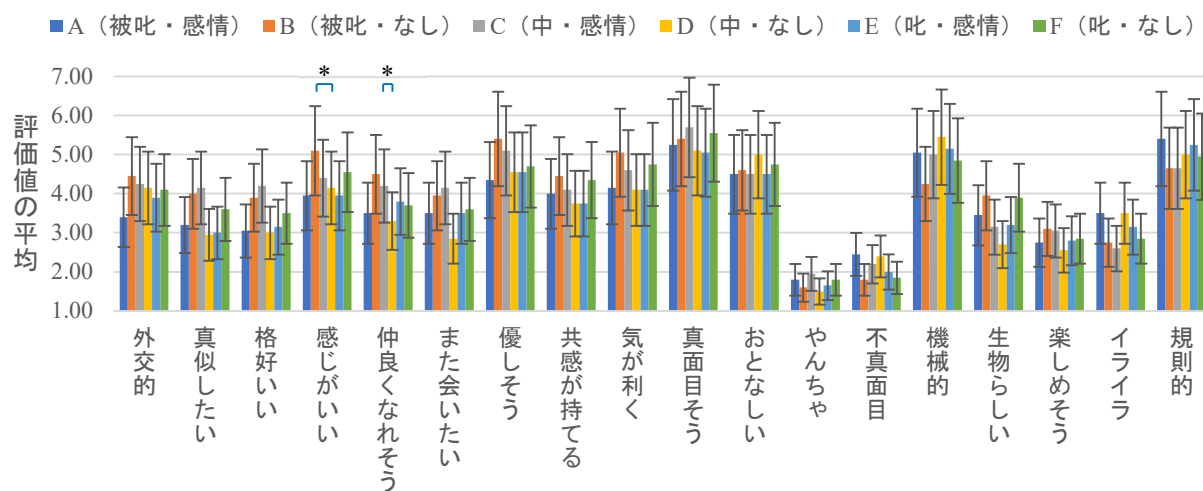


図 12 仲裁者に対する印象

表 14 過去の経験により有意差の生じた印象項目

経験	人数	対象	形容詞	平均値 (有効数字 2 桁)		標準誤差
				あり	なし	
当時は納得できなかった理由による叱責を受けた	あり：62名 なし：58名	仲裁者	格好いい**	あり	3.85	0.18
				なし	3.05	0.19
理由の不明な叱責を受けた	あり：11名 なし：109名	被叱責者	やんちゃ*	あり	3.55	0.47
				なし	4.58	0.14
		仲裁者	真面目そう**	あり	6.27	0.30
				なし	5.25	0.17
理不尽な理由による叱責をした	あり：22名 なし：98名	仲裁者	イライラ*	あり	3.73	0.36
				なし	2.91	0.15

* : $p < 0.05$, ** : $p < 0.01$

また、過去に理不尽な理由による叱責を行った経験がある実験参加者は、そうでない実験参加者と比べ仲裁者に対する苛立ちを強く示していた。原因として、過去に行った自身の叱責について後悔などの感情を抱いている場合、納得のいく理由による叱責に介入する仲裁者に対して不快感を覚えた可能性がある。

6. おわりに

本研究では、多人数対話に対する適切な介入を行う対話システムの開発を目標として、多人数対話音声の感情を考慮して動作する対話システムの開発を行った。個人の識別情報と推定される感情情報を付与した形で対話履歴を貯蔵する構造を構築し、文脈と感情を考慮して動作する対話エージェントを作成した。また、対話エージェントによる感情考慮の効果、エージェントがとる立場による対話参与者への印象変化の調査を行った。

印象評価実験の結果、実施した 6 条件の間では大半の印象項目で有意な差が確認されず、エージェントの発話が発生するだけではエージェントを除く対話参与者への印象が変化しないこと、実験中の 6 条件によるエージェントの発話の違いが不十分である可能性などが示唆された。また、実験参加者の叱責に関する過去に基づく比較では、過去に納得のいかない理由による叱責や理由の不明な叱責を受けた実験参加者には被叱責者や仲裁者に同情的な傾向が確認され、理不尽な理由による叱責をした経験のある実験参加者は仲裁者に対して不快感を示した。このことから、同様の対話場面に対しても、ユーザの過去の経験と対話場面の受け取り方に関係性があることが示唆された。

今後は発話回数や発話内容の変更、対話時間の違いなど、より多くの条件による印象影響の調査、ユ

ーザの過去の経験と発言に対する印象の調査を行うことを考えている。

謝辞

本研究は、東京工芸大学後援会研究奨励費の助成を受けました。記して感謝いたします。

参考文献

- [1] tender|Zoom の運営を自動化するサービス : <https://tender.party/> (参照日 2024-1-14)
- [2] 会話 AI ロボット Romi (ロミイ) : <https://shop.romi.ai/> (参照日 2024-1-12)
- [3] 黄宏軒, 乙木翔地, 堀田怜, 川越恭二: 多人数会話において積極的に情報提示ができるガイドエージェントの実現に向けての介入場面の検討, 実行知能学会論文誌, 2016
- [4] Tang B. N., 目良和也, 黒澤義明, 竹澤寿幸: 音声に含まれる感情を考慮した自然言語対話システム, HAI シンポジウム 2014, pp. 87-91, 2014
- [5] Yoshua B., Réjean D., Pascal V., and Christian J. : A Neural Probabilistic Language Model, Journal of Machine Learning Research, Vol. 3, pp. 1137-1155, 2003
- [6] Alec R., Karthik N., Tim S., Ilya S. : Improving Language Understanding by Generative Pre-Training, 2018
- [7] 橋本慧海, 柳楽浩平, 水本武志, 白松俊: GPT-4 を用いた就活生の自己分析支援コーチングエージェントの開発, 第 99 回言語・音声理解と対話処理研究会, pp. 56-58, 2023
- [8] 有本泰子, 河津宏美: 音声チャットを利用したオンラインゲーム感情音声コーパス, 日本音響学会 2013 年秋季研究発表会講演論文集, pp. 385-388, 2013
- [9] Robert P. : Emotion : A Psychoevolutionary Synthesis,

Harper & Row, 1980

- [1 0] 塩田さやか：音声を用いた生体認証技術，話者照合ソフトウェア入門，アイサイ研究者のための音声情報処理ソフトウェア入門 特集号, Vol. 62, No. 2, pp. 63-68, 2018
- [1 1] Najim D., Patrick K., Reda D., Pierre D., and Pierre O. : Front-End Factor Analysis For Speaker Verification, IEEE Transactions on Audio Speech and Language Processing, Vol. 19, pp. 63-68, 2018
- [1 2] OpenAI : <https://openai.com/> (参照日 2024-1-14)
- [1 3] 高吉幸治, 田中俊也：ロボットの振る舞いと知性・性格の印象の関係，電子情報通信学会技術研究報告, Vol. 160, pp. 43-48, 2007