

対話型エージェントが意見伝達を仲介する際に適した音声の高さに関する調査

渡辺 好汰¹ 丸山 葉² 大西 俊輝² 大串 旭² 呉 健朗² 大澤 正彦¹ 宮田 章裕^{1,a)}

概要: 多人数のコミュニケーションにおいて、他者の意見を伝達するシーンが発生する。このようなシーンにおいて、他者の意見を伝達する「仲介」はコミュニケーションで重要な役割を果たしており、日常的に用いられる手段である。これまで我々は、エージェントが他者の意見を伝達する仲介を行う研究を行ってきた。しかし、先行研究ではエージェントの出力はテキストに限定されており、エージェントの音声について扱ってこなかった。そこで本稿では、エージェントの音声の特徴に焦点を当て、特に意見伝達の仲介に適した音声の高さに注目する。対話型エージェントの音声の高さの違いによる対照実験を行うことで、どのような高さの音声か意見伝達の仲介に適しているか明らかにする。対照実験の結果、通常の高さの音声を持つ対話型エージェントが意見伝達を仲介する際に、意見の受信者は意見に対し好印象を抱く可能性があることが分かった。

1. はじめに

多人数のコミュニケーションにおいて、他者の意見を伝達するシーンが発生する。たとえば、A・B・Cの3者がコミュニケーションを行う際、AがCに直接意見を伝えるのではなく、AがBに意見を伝え、BがAからの意見をCに伝えることがある。このように、仲介は我々のコミュニケーションで重要な役割を果たしており、日常的に用いられる手段である。

ここで、昨今の対話型エージェントの社会への広がりについて注目したい。エージェントの対話能力は日々向上しており、フォーマル/インフォーマルを問わず、人とエージェントが混在した形態のコミュニケーションが関心を集めている。近い将来、この形態のコミュニケーションが日常的なものになることは想像に難くない。仲介の文脈においても、エージェントは重要な役割を果たすと考えられている。その考えに基づいて、我々は対話型エージェントが意見伝達の仲介をする研究を行ってきた [1][2]。しかし、先行研究ではエージェントとの対話の出力がテキストに限定されており、音声での出力については研究の対象外としてきた。社会に普及している対話型エージェントの多くは、音声で出力することが当たり前に行われている。そこで本稿では、エージェントの音声に焦点を当て、意見伝達の仲介に適している音声はどのような特徴を持っているのか明

らかにすることを旨とする。

本稿の貢献は、対話型エージェントが音声で意見伝達を仲介する際に、音声の高さの異なる条件間で対照実験を行い、意見を受信するユーザの意見に対する印象がどのように変化するか調査したことである。

2. 関連研究

2.1 仲介を行う対話型エージェント

野口ら [5] は、遠隔コミュニケーションを仲介するロボットが高齢者の自己開示に与える影響について調査している。その結果、他者との会話に仲介するロボットの社会的振舞いは、内面性の高い話題の伝達を助けることが示唆されている。また、仲介ロボットは高齢ユーザに warmth と competence を強く知覚させ、これらの知覚量と開示量には正の相関があることが明らかにしている。上記より、仲介ロボットの要求仕様として、ロボットには対話者という擬人化された存在として認識されるための振舞いが必要であり、自己開示を促すためにはより warmth と competence が強く知覚されることが重要であることを明らかにしている。加えて、野口ら [6] は、社会性ロボットが高齢者とその家族・友人高齢者との遠隔コミュニケーションを支援する役割を果たすための設計指針の提案も行っている。高齢者が他者に話すことに抵抗を感じる話題において、社会的仲介ロボットが高齢者の自己開示を促進する可能性があることがわかっている。また、社会的仲介ロボットの効果的な性格特性を調査し、類似性-惹起性-反発性に関して、

¹ 日本大学文理学部

² 日本大学大学院総合基礎科学研究科

^{a)} miyata.akihiro@acm.org

詳細な知見を得ることを確認している。Kobayashi ら [7] は、既存のソーシャルメディアを介した高齢者と若い世代の双方向コミュニケーションに利用できるソーシャルメディア仲介ロボットを提案している。人間型ロボットに搭載したシングルボードコンピュータを利用して、スマートフォンを使わずに音声によるソーシャルメディア上の情報取得・発信ができるシステムを開発している。Tahir ら [9] は、ヒューマノイドロボット Nao を社会的仲介者として用いるためのユーザスタディを行っている。まず、実験参加者が Nao を介して伝えられるフィードバックメッセージの理解度を評価している。次に、実験参加者は、Nao ロボットが仲介役となるシナリオに基づいた 2 者対話を行う。Nao を介して会話を分析した結果をフィードバックするシステムを利用し、参加者は受け取ったフィードバックの内容、適切さ、タイミングなどの評価を行っている。また、Godspeed アンケート [8] により、社会的仲介者としての Nao に対する総合的な認識を評価している。その結果 Nao は対話におけるフィードバックの提供に効果的に利用できることが示されている。

2.2 音声表現に関する研究

内田ら [10] は音声表現に関する研究として、話者の発話速度と高さによる影響を調査している。人は話者の発話速度と声の高さによって、話者への印象が変化することが分かっている。これは話者が人である時に限らず、人間とエージェントの対話においてエージェントの音声は重要な特徴になりつつあり、人々がどのようにエージェントと対話し、反応し、エージェントの音声に影響されるかを理解することが重要であると言われている [11]。橋本ら [12] は、人工音声を使用し実験を行い、発話速度と高さが特性推論に及ぼす影響を検討している。その結果、音声のパラ言語的特徴の違いによって、異なる性格特性が推論されることが示された。Tolmeijer ら [13] は、音声アシスタントの性別に関する実験を行い、音声アシスタントに対して暗黙のステレオタイプ化が起こることを明らかにしている。加えて、性別が曖昧な音声と性別がある音声では、信頼関係に有意な差がないことが明らかにしている。

3. 研究課題

1 章で述べたように、対話型エージェントは仲介の文脈において重要な役割を果たすと考えられる。その考えに基づいて、我々は意見伝達を仲介するエージェントに関する調査を行ってきた [1][2]。しかし、これらの先行研究では、エージェントの出力をテキストに限定していた。本研究を推進するうえで、対話の出力を音声で行わせることが考えられるが、エージェントの意見伝達を仲介する際に適した音声の高さがどのようなものであるか明らかにされていない。

この問題を踏まえ、本稿では、対話型エージェントが意見伝達を音声出力で仲介する際に適した音声の高さを、意見を受信するユーザの意見に対する印象が良くなる音声の高さとし、それを明らかにすることを研究課題とする。

4. 音声の作成方法

本稿では、先行研究であるエージェントの音声の高さを変える対照実験を行う方法を利用する [3]。音声を作成するために、音読さん [14] と Praat [15] を使用した。音読さんは、文章を入力すると好みの音声で読み上げてくれる音声読み上げソフトである。Praat は、音声进行分析、変換、合成することができるフリーソフトである。

4.1 基準とする音声の特徴

本稿では、国内で高いシェアを誇っている Apple 製品に搭載されている音声アシスタントの Kyoko を参考に基準とする音声特徴を設定した。Kyoko は macOS・iOS に搭載されている日本語のテキスト読み上げ機能用の音声アシスタントである。Kyoko が“おはようございます”という文章を読み上げた際の音声の速さと高さを Praat を用いて計測したところ、音声の声帯振動の開始から終了までの時間が 1.04 秒、平均基本周波数が 224Hz で読み上げていた。この声帯振動の開始から終了までの時間と平均基本周波数を基準とする。音読さんを使用して“おはようございます”という文章を読み上げた際、高低の設定を - 4.6、速度を $\times 1.0$ とすると、音声の声帯振動の発生から終了までの時間が 1.01 秒、平均基本周波数が 224Hz であった。その後、Praat を使用して発話速度を遅くし、音声の声帯振動の開始から終了までの時間が 1.04 秒、平均基本周波数が 224Hz の音声を作成した。上記の方法で作成した音声を本稿では通常の高さ/速さの音声とした。

4.2 実験で使用する音声の作成方法

本研究で行う実験では、全条件、速さは一緒にし、音の高さが異なる 3 種類の条件を作成する (表 1)。音声の高さを変化させる方法について述べる。通常の高さより高い音声と低い音声の設定方法については、内田らの研究 [10] を参考に設定した。通常とする高さ (平均基本周波数) を 100% とした際に、高さが 156.25% の場合を高い音、高さが 64% の場合を低い音声となるように Praat で編集を行う。この値は通常の高さに対して対数軸上で等間隔になっている。

5. 検証実験

本稿では、対話型エージェントが初対面同士での一般的な評価基準を持つ課題に対する評価の伝達を仲介する際に、対話型エージェントの音声の高さの違いによる対照実験を行う。

表 1 手法ごとの音声の高さ

手法	音声の高さ
手法 1	通常の高さ
手法 2	通常の高さより高い
手法 3	通常の高さより低い

表 2 登場人物・エージェント

実験参加者への教示	実際の人物
実験者	実験者
実験参加者 A	実験参加者
実験参加者 B	実験者
伝達エージェント	実験者

5.1 実験の目的

本実験では、対話型エージェントが初対面同士での普遍的な評価基準を持つ課題に対する評価の伝達を仲介する際に、音声の違いによって、評価を受信するユーザのポジティブな評価・発信者・エージェントに対する印象がどのように違うのか明らかにする。

5.2 実験条件

実験参加者は、20代の学生7名（男性：4名，女性：3名）である。本実験では、各手法で評価の受信者のポジティブな評価・発信者・エージェントに対する印象にどのような違いがあるか検証するため、Wizard of Oz法 [4]を用いて、表1に示した3つの手法（表1）による比較を行う。実験参加者は実験参加者間でペアを組み、実験参加者は意見の受信者として、各手法を相手を変えて、それぞれ1回ずつ行う。各手法において、実験者は、意見の発信者・仲介エージェントとして実験に参加するが、実験参加者には、意見の発信者は別の実験参加者であると教示し、エージェントは自律したエージェントであると教示する（表2）。

5.3 実験の手順

本実験では、実験前アンケート、対話後アンケート、実験後アンケートの3種類を行う。アンケートの詳細を表3に示す。Q1~Q3が実験前アンケート、Q4~Q14が対話後アンケート、Q14~Q20が実験後アンケートとなっている。

5.3.1 事前準備

事前準備は下記の手順で行う。

Step 1: 実験者は実験者参加者に実験について説明を行う。

Step 2: 実験参加者は、実験前アンケートに回答する。

5.3.2 実験時

実験時は下記の手順で行う。

Step 1-1: 実験者と実験参加者 A は実験実施場所で着席する

Step 1-2: 実験者は実験参加者 A に本実験で使用するエージェント・実験参加者 B とユーザの関係性、シー

ンについて説明する。

Step 2: 実験参加者 A と仲介エージェント（実験者）は、1つのテキストチャットルームに入室する。実験参加者 B（実験者）と仲介エージェントは、上記と異なる1つのテキストチャットルームに入室する。

Step 3: 実験参加者 A・B は、2桁2項の四則演算を行うタスク（電卓等を使わない）を1分間行う。タスク終了後、実験参加者 A・B はタスク結果の画像（解答用紙）をエージェントに共有する。

Step 4: 仲介エージェントは、実験参加者 A に実験参加者 B のタスク結果の画像（解答用紙）を、実験参加者 B に実験参加者 A のタスク結果の画像（解答用紙）を共有する。

Step 5: 実験参加者 A は、実験参加者 B のタスク結果に対しての評価を、仲介エージェントにテキストチャットで共有する。

Step 6: 仲介エージェントは、実験参加者 B からの実験参加者 A のタスク結果に対する評価を、実験参加者 A に音声で伝える。

Step 7: 実験参加者 A は、現在の手法に対する対話後アンケートに回答する。

Step 8: Step 2 から Step 6 を3手法すべてで行う。

Step 9: 実験参加者は、実験後アンケートに回答する。Step1-2では、本実験で使用するエージェントは実体がなく、機械であることを説明した。Step3からStep7は順序効果を相殺するため手法の順番をランダムで行った。Step6では、実験参加者がエージェントの発言をしっかりと聞けるように、2回再生した。実験者 B からの実験参加者 A のタスク結果に対する評価は、事前に褒める評価に関するアンケートで収集した評価を選定し、ランダムに出力した（表4）。

5.4 実験結果・考察

この実験での評価指標は、意見の受信者である実験参加者がポジティブな評価に好印象を抱いたか、発信者に好印象を抱いたか、エージェントに好印象を抱いたかの3種類となる。

まず意見の受信者である実験参加者がポジティブな評価に好印象を抱いたかという指標に関して述べる。アンケート Q4 の結果の分布を図1に示す。5以上の回答した人は手法1から3の順に、42.9%、42.9%、14.3%となった。これを裏付けるようにアンケート Q8 の結果（表5）で、評価に対して最も好印象を抱いたと回答された数が多かったのは手法1であった。アンケート Q9 より、“高すぎない音声だと、正しく評価されていると感じた”という回答や、“満足にいく結果の時に高すぎる声と文章が単調すぎると機械的に対応されている感が強かった”という回答があった。上記より、評価に対して最も好印象を抱いたかという指標

表 3 質問内容

質問内容	回答方法
Q1 対話型エージェント (Siri や alexa など) を普段からどのくらいの頻度で利用していますか?	利用していない: 1日1回以上: 週に2~6回: 月1~2回
Q2 (対話型エージェントを利用している人) 利用している対話型エージェントへの入力手段はどれですか?	テキスト: 音声: その他
Q3 (対話型エージェントを利用している人) 利用している対話型エージェントへの出力手段はどれですか?	テキスト: 音声: その他
Q4 あなたのタスク結果に対する評価 (相手ではない) に好印象を抱いた	7: とてもそう思う~1: 全くそう思わない
Q5 あなたのタスク結果に対して評価した相手 (評価や仲介者ではない) に好印象を抱いた	7: とてもそう思う~1: 全くそう思わない
Q6 あなたのタスク結果に対する評価の仲介者 (評価した人ではない) に好印象を抱いた	7: とてもそう思う~1: 全くそう思わない
Q7 その他に感じたことがあれば自由に記述してください	自由記述
Q8 あなたのタスク結果に対する評価 (相手ではない) に最も好印象を抱いた音声の高さはどれですか?	手法1~3: その他 (自由記述) から1つ選択
Q9 対話型エージェントの音声の高さの違いについて, あなたのタスク結果に対する評価 (相手ではない) の観点で感じたことを教えてください	自由記述
Q10 あなたのタスク結果に対して評価した相手 (評価や仲介者ではない) に最も好印象を抱いた音声の高さはどれですか?	手法1~3: その他 (自由記述) から1つ選択
Q11 対話型エージェントの音声の高さの違いについて, あなたのタスク結果に対して評価した相手 (評価や仲介者ではない) の観点で感じたことを教えてください	自由記述
Q12 あなたのタスク結果に対する評価の仲介者 (評価した人ではない) に最も好印象を抱いた音声の高さはどれですか?	手法1~3: その他 (自由記述) から1つ選択
Q13 対話型エージェントの音声の高さの違いについて, あなたのタスク結果に対する評価の仲介者 (評価した人ではない) の観点で感じたことを教えてください	自由記述

表 4 褒める評価の出力例

条件	褒める評価内容
そのまま使う	短時間でこんなに解けるのすごい 正確に解けていてすごい
n を回答結果によって変更する	n 問正解はすごい n 点はとても素晴らしいと思います n 点だね! すごい n 問以上あってるね! すごい n つしかミスして無いね。すごい

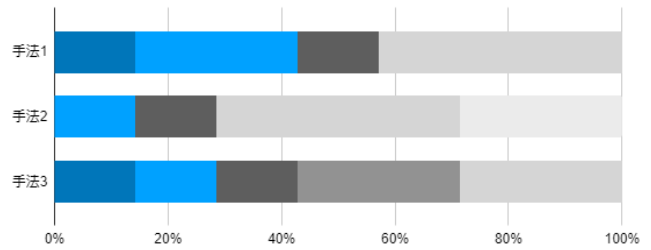


図 2 アンケート Q5 の回答結果 (N=7)

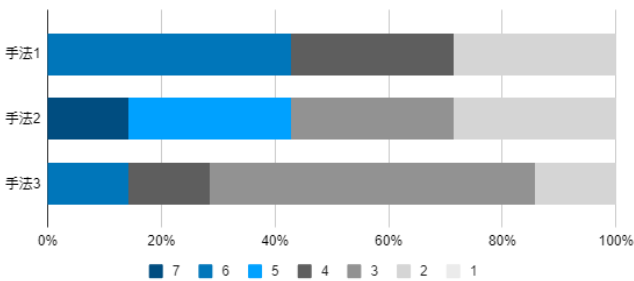


図 1 アンケート Q4 の回答結果 (N=7)

表 5 アンケート Q8,10,12 の回答結果 (N=7)

	手法 1	手法 2	手法 3	その他
Q8	71.4%	0%	28.6%	0%
Q10	42.9%	28.6%	0%	28.6%
Q12	28.6%	28.6%	14.3%	28.6%

に関しては、通常の高さである手法 1 が適している可能性がある。

次に意見の受信者である実験参加者が評価に好印象を抱いたかという指標に関して述べる。アンケート Q5 の結果の分布を図 2 に示す。5 以上の回答した人は手法 1 から 3 の順に、42.9%、14.3%、28.6% となった。これを裏付けるようにアンケート Q10 の結果 (表 5) で、評価した相手に対して最も好印象を抱いたと回答された数が多かったのは

手法 1 であった。アンケート Q11 より、“高い声のほうが結果による感動をそのまま伝えているように感じました”という回答や、“高いほうが人間らしく感じた”という回答があった。上記より、評価した相手に対して最も好印象を抱いたかという指標に関しては、通常の高さである手法 1 もしくは手法 2 が適している可能性がある。

次に意見の受信者である実験参加者がエージェントに好印象を抱いたかという指標に関して述べる。アンケート Q6 の結果の分布を図 3 に示す。5 以上の回答した人は手法 1 から 3 の順に、28.6%、14.3%、14.3% となった。これを裏付けるようにアンケート Q12 の結果 (表 5) で、エージェントに最も好印象を抱いたかと回答された回数が多かったのは手法 1, 2 であった。アンケート Q13 より、“高すぎず低すぎない声でやってもらおうと相手とやり取りしてる感がある気がする”という回答があった。しかし、“仲介者に対する印象は声の高さで変わらないように感じた”という回答や、“手法間に差はないと感じた”という回答があった。上記より、エージェントに最も好印象を抱いたかという指標に関しては、通常の高さである手法 1 または、通常の高さよりも高い手法 2 が適している可能性があるが、どの高さにおいてもエージェントへの印象は変わらない可能性がある。

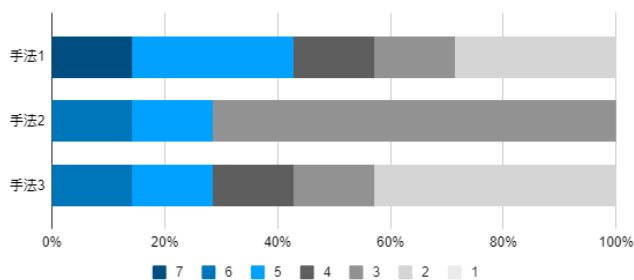


図3 アンケート Q6 の回答結果 (N=7)

最後に、意見の受信者である実験参加者がポジティブな評価に好印象を抱いたか、発信者に好印象を抱いたか、エージェントに好印象を抱いたかという指標を合わせて考える。手法1はポジティブな評価に好印象を抱いたか・評価した相手に好印象を抱いたか・エージェントに好印象を抱いたかという指標に関して最も適している可能性がある手法であった。上記より、全ての指標に関して考えた際に、通常の高さである手法1が適している可能性がある。

6. おわりに

本稿では、対話型エージェントが意見伝達の仲介発言を音声で出力する際の、意見伝達を仲介する際に適した音声の高さの調査を行った。その結果、全ての指標に関して考えた際に、通常の高さである手法1が適している可能性があった。また実験参加者に人数が7名と少なかったため、引き続き実験参加者を増やして検証を続ける必要がある。

今後は、異なるシーンでの適した音声の高さの検証を行っていく。

参考文献

- [1] 丸山葉, 大西俊輝, 大串旭, 呉健朗, 大澤正彦, 宮田章裕: 意見伝達を仲介するエージェントの基礎検討. HAI シンポジウム 2022 予稿集, G-17 (2022).
- [2] 丸山葉, 大西俊輝, 大串旭, 呉健朗, 大澤正彦, 宮田章裕: 意見伝達を仲介する対話型エージェントに対する利用意欲の調査. マルチメディア, 分散, 協調とモバイルシンポジウム 2022 論文集, pp.369-373 (2022).
- [3] 得田舜介, 呉健朗, 大西俊輝, 大串旭, 宮田章裕: 対話型エージェントのユーモア表現に適した韻律的特徴の調査. マルチメディア, 分散, 協調とモバイルシンポジウム 2022 論文集, pp.570-575 (2022).
- [4] Kelley, J.F.: An iterative design methodology for user-friendly natural language office information applications. ACM Transactions on Information Systems (TOIS), Vol. 1, No. 2, pp.26-41 (1984).
- [5] 野口洋平, 上出寛子, 田中文英: 遠隔コミュニケーションを仲介するロボットが高齢話者の自己開示に与える影響. ヒューマンインタフェース学会論文誌, Vol. 20, No. 1, pp.67-78 (2018).
- [6] Noguchi, Y., Kamide, H. and Tanaka, F.: Personality Traits for a Social Mediator Robot Encouraging Elderly Self-Disclosure on Loss Experiences. ACM Transactions on Human-Robot Interaction (THRI), Vol. 3, No. 9, pp.1-24 (2020).
- [7] Kobayashi, T., Katsuragi, K., Miyazaki, T. and Arai,

K.: Social media intermediation robot for elderly people using external cloud-based services. In 5th IEEE International Conference on Mobile Cloud Computing, Services, and Engineering, pp.31-38 (2017).

- [8] Bartneck, C., Kulić, D., Croft, E. and Zoghbi, S.: Measurement instruments for the anthropomorphism, animacy, likeability, perceived intelligence, and perceived safety of robots. International journal of social robotics, Vol. 1, No. 1, pp.71-81 (2009).
- [9] Tahir, Y., Dauwels, J., Thalmann, D. and Magnenat Thalmann, N.: A user study of a humanoid robot as a social mediator for two-person conversations. International Journal of Social Robotics, Vol. 12, No. 5, pp.1031-1044 (2020).
- [10] 内田照久, 中畝菜穂子: 声の高さと発話速度が話者の性格印象に与える影響. 応用心理学研究, Vol. 75, No. 5, pp.397-406 (2004).
- [11] Seaborn, K., Miyake, P.N., Pennefather, P., OtakeMatsura, M.: Voice in Human-Agent Interaction: A Survey. ACM Computing Surveys, Vol. 54, No. 81, pp.1-54 (2021).
- [12] 橋本和奈実, 古谷 健: 発話速度と声の高さが特性推論に及ぼす影響—二段階推論仮説に基づいて—. 応用心理学研究, Vol. 45, No. 1, pp.15-25 (2019).
- [13] Tolmeijer, S., Zierau, N., Janson, A., Wahdatehagh, J., Leimeister, M. J., Bernstein, A.: Female by Default?—Exploring the Effect of Voice Assistant Gender and Pitchon Trait and Trust Attribution. Extended Abstracts of the 2021 CHI Conference on Human Factors in Computing Systems (CHI EA '21), No. 455, pp.1-7 (2021).
- [14] 音読さん, 入手先 (<https://ondoku3.com/ja/>) (last visited on 2023/12/21).
- [15] Boersma, P., Weenink, D.: Praat: doing phonetics by computer, 入手先 (<https://www.fon.hum.uva.nl/praat/>) (last visited on 2023/12/21).