

コンテキスト推定に基づいて予測的認知を誘発する 非自然言語エージェント

奥岡 耕平[†] 大澤 正彦^{‡§} 今井 倫太[†]

慶應義塾大学 理工学部[†] 慶應義塾大学 大学院理工学研究科[‡]

日本学術振興会 特別研究員 (DC1) [§]

1. はじめに

スマートスピーカーに代表されるように、音声対話エージェントの需要が拡大している。しかし、自然言語による破綻しない円滑な対話の継続は膨大な語彙や対話の流れを理解することが必要であり、未だ困難である。

従来研究では、清丸ら[1]が明確な意味を持たない非自然言語を用いるエージェントを提案している。非自然言語に対して人が自発的に意味を解釈することによって、対話を破綻させずに成立し得る可能性を実験により示唆している。

清丸らの研究ではエージェントの発話を人の解釈に委ねることで対話の破綻を防ぐことはできるが、エージェントが人に情報を伝達したい時に正確に伝達できない可能性がある。しかし、正確に情報を伝達するための非自然言語音声エージェントの振る舞いについての検討は不十分である。

本稿では、非自然言語エージェントが正確に情報を伝達するための振る舞いをデザインする指針として記号接地に着目した。人の中で、ある特定の非自然言語という記号が特定の物体や事象に接地すれば、非自然言語音声による正確な情報が伝達できると考えられる。

そこで、人の記号接地を誘発する非自然言語エージェントの振る舞いをデザインし、提案する振る舞いによって特定の物体に対して人の記号接地を誘発することを検証した。

2. 記号接地を誘発する振る舞い

記号接地を誘発する振る舞いの1つとして接地する対象の物体に目線や指差しなどを用いて人の注意を向け、対応する音声を発話すること



図1 実験の様子

が考えられる。この時、発話した非自然言語音声は物体に対応する記号であると人が解釈をする必要がある。

しかし、Schererら[2]が示すように非自然言語音声だけでも特定の感情を表現することができることを考慮すると、特定の感情表現に繋がる音声では物体に対応する記号ではなく、感情表現として受け取られる可能性がある。本稿では、特定の感情表現に繋がる非自然言語音声は記号接地の誘発に及ぼす影響を調査する。

3. 実験

特定の感情表現に繋がる非自然言語音声の発話の印象と、記号接地への影響を評価する実験を行った。記号接地を誘発する振る舞いを行うロボットを実装し、撮影した映像を実験参加者に視聴してもらいアンケートに答えてもらうことで評価を行った。ロボットには川崎ら[3]が開発したコミュニケーションロボットを用い、発話する音声は清丸ら[1]が収集した音声データを利用した。

3.1 予備実験

予備実験として、清丸ら[1]の音声データから特定の感情表現に繋がる音声の選別を行った。平均年齢 25.1 歳の 16 名の男女に音声データを

Non natural language agent with inducing predictive cognition based on context estimation

[†]Kohei Okuoka, Michita Imai

Faculty of Science and Technology, Keio University

[‡], [§]Masahiko Osawa

Graduate School of Science Technology, Keio University

Japan Society for the Promotion of Science, Research Fellow

(DC1)

聴いてもらい、音声に対してポジティブまたはネガティブな印象のどちらを抱くかを7段階評価してもらった。各音声の評価平均値を元にポジティブ、ネガティブ、どちらでもないニュートラルの3つの条件に分類し、評価平均値と分散を基に各条件につき3つの音声を選別した。

3.2 評価用映像

図1に示すように、ロボットの目の前に3つのオブジェクトが置かれている環境を用意した。ロボットは図左から順に1つずつ、指さしによってオブジェクトに注意を誘導し、対応する音声を1度発話するという動作を行った。この映像をロボットが発話する音声条件を変えて3本撮影した。

3.3 実験手順

3.2章で製作した動画を視聴してもらい、ロボットの発話した内容について、抱いた印象を表1に示す選択肢から選んでもらった。

表1 発話内容の印象についての質問の選択肢

選択肢	発話内容
性質	指をさした物体の性質
名前	指をさした物体の名前
動作	指をさした物体に対する動作
印象	指をさした物体に対する印象

続いて、実験参加者の中で記号接地が正確に行われたかどうかを評価した。動画で利用した3つの音声を音声のみで聴いてもらい、それぞれの音声に対応すると思ったオブジェクトを動画に提示した3種類から1つ選んでもらった。

上述の試行を音声の条件を変えて1人につき3条件全てで行ってもらった。また、順序効果を考慮してカウンターバランスをとり、実験を行った。実験参加者は平均年齢26.9歳で男性11名、女性4名の合計15名であった。

3.4 実験結果

ロボットの発話内容に対する印象について各音声条件で比較した結果を図2に示す。カイ二乗検定を行ったところ、5%有意水準で有意に音声条件と発話内容に関連があることが示された。結果、ニュートラルな音声を用いることで物体の名称を発話している印象を人に与える傾向が高いことが示唆された。

次に、記号接地の正答率を各音声条件で比較した結果を図3に示す。分散分析を行った結果、

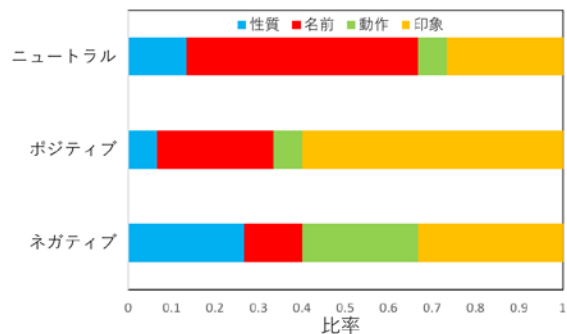


図2 発話に対する印象の比較結果

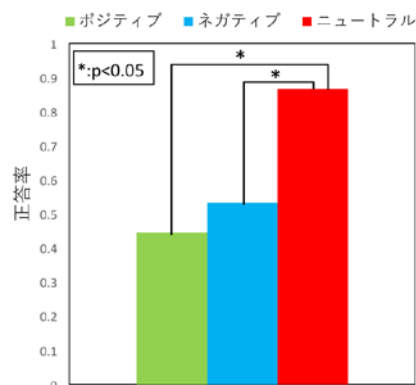


図3 記号接地の正答率の比較結果

ニュートラルな音声における正答率が他の条件と比較して5%有意水準で有意差が認められた。

特定の感情表現に寄与しない音声が発話した音が物体の名称であるという印象を人に与えることに適していることが示された。その結果、音声と物体の関係を暗記しようとする人の動作を誘発することになり、正確な記号接地に繋がったと考えられる。

4. まとめ

本稿では、非自然言語音声エージェントにおける記号接地を誘発する振る舞いとして適切な非自然言語音声について検討した。実験の結果、特定の感情表現に繋がらないニュートラルな非自然言語音声が発話した音が人の記号接地を誘発する振る舞いには適していることが示唆された。

参考文献

[1] 清丸寛一 et al.: 予測的認知を用いた非自然言語による言語的コミュニケーション, SIG-AGI 6th, (2017)
 [2] Scherer, Klaus R: "Acoustic Concomitants of Emotional Dimensions: Judging Affect from Synthesized Tone Sequences." (1972).
 [3] 川崎邦将 et al.: 認知的制約付き擬人化キャラクターに着目したコミュニケーションロボットの設計と開発, SIG-AGI 6th, 2017