

ユーザとの信頼関係構築を目的とした対話ロボットのための
ユーザの情動に同調する対話手法

The Method of synchronizing with the emotion of a User
for a Chat Robot to Obtain the User's Confidence

西村 祥吾[†] 中村 卓矢[†] 川波 弘道[‡] 神原 誠之[†] 萩田 紀博^{†, §}
Shogo Nishimura Takuya Nakamura Hiromichi Kawanami Masayuki Kanbara Norihiro Hagita

1. はじめに

インターネットの普及や独居高齢者の増加に伴い、年齢問わず日常的な会話機会が減少している[1], [2]。日常的な対話機会が不足すると、QoL の低下やそれに伴う生活環境の悪化、身体機能の低下、うつ病等の問題を引き起こし、特に高齢者の場合は脳機能低下による認知症に発展する可能性があり、社会問題となっている。この問題を解決する対策の一つとして、対話ロボットによる日常的なコミュニケーションの活性化を目指す取り組みが注目されている。しかし、そのような目的を持つロボットを、ユーザが自発的に日常生活の中で利用する意欲（以下、継続利用意欲）を高めるための取り組みは少ない。

Miyazawa ら[3]は音声対話システムにおいてユーザの継続利用意欲を向上する要因を明らかにしており、ロボットに対し対話の間楽しさやソーシャルスキルの高さを意味する社会性と、意外性や新しい情報を提供する新奇性の大きく 2 つを導入する重要性を示した。Minami ら[4]は、実況掲示板等のソーシャルメディア上のコメントを使ったテレビ視聴型雑談ロボットを提案した。ロボットがテレビ番組をユーザと共に視聴するという限定された状況において、同じテレビ番組を視聴している不特定多数の SNS ユーザがリアルタイムに書き込んだコメントを用いて発話を行い、機械的生成が困難であった臨場感やユーモアのある新奇性の高い内容の発話を継続的に提供可能とした。また自然な応答速度を実現するため、ロボットがユーザの発話に対して相槌、復唱を行うことで社会性を高めており、ユーザの継続利用意欲の向上に貢献した。

本研究では Minami らと同様にユーザと対話ロボットがテレビの共同視聴を行う環境を想定する。特にスポーツ中継などの場面においては、加点やファインプレーといったシーンで盛り上がりが発生する。この際にユーザとロボットが情動体験を共有することを雰囲気共有と定義し、ロボットの社会性を向上するための機能として提案する。雰囲気共有を実現するためのアプローチとして、人との信頼関係構築を目的とした対話手法である同調対話に着目し、人とロボットの対話において実装する。ユーザの時系列的な情動変化を模倣することによってロボットとの信頼関係を構築し、ユーザの対話ロボットに対する継続利用意欲の向上を目指す。

2. 関連研究と本研究の位置付け

本章ではユーザと対話ロボット間の雰囲気共有を実現す

るための同調対話に関する関連研究について説明し、本研究の位置付けを述べる。

2.1 同調に関する心理学的研究

対人社会心理学の分野において同調傾向は多く研究されており、円滑なコミュニケーションの指標のひとつされてきた。同調傾向とは、相互作用相手との間でコミュニケーション行動が連動し、パターンが類似化していくことを指す。さらには、同調傾向は他者とのラポール形成やポジティブな対人印象をもたらすことが示されてきた[5]。対話中の同調を用いたラポール形成のためのテクニックとして、相手の身振りや動作を合わせるミラーリングや相手の話し方や状態、呼吸などのペースを合わせるペーシング、相手の言ったことを返すバックトラッキングなどが挙げられる。

2.2 対話ロボットの感情表現に関する研究

カウンセリングを行う CG エージェントの研究によると、ユーザ（カウンセリングを受ける人）の表情と同じ感情表現をエージェントが行うことで、ユーザとエージェント間のラポールが形成されることが示された[6]。Matsumoto ら[7]は姿勢、ジェスチャ、音声による感情表現が可能なロボットを用いて、情動を想起させる映像に対して適切な感情表現を行ない、ロボットが映像に合った感情を表現することで、映像から喚起されるポジティブ（興奮・リラックス）な情動は強化され、ネガティブな情動（恐怖）は弱化されることを示した。一方でロボットがシーンに不一致な感情表現や一貫性のない感情表現を行うことでユーザに無視されるなど、ユーザのロボットに対する信頼性が失われる可能性も示唆されている。

Takahashi ら[8]はテレビ視聴時に 1 ユーザ対 3 雑談ロボットという状況でテレビ視聴時を行った際、1 ユーザ対 1 雑談ロボットの場合に比べ、場の雰囲気が盛り上がっているように感じるなど、ユーザにポジティブな印象を与えることが示された。

2.3 本研究の位置付け

2.1, 2.2 から人対人のインタラクションと同様に人対ロボットインタラクションも同様に同調傾向がラポール形成に貢献する可能性が考えられる。対話ロボットを用いた関連研究では、同一のコンテンツをユーザと共視する環境で感情表現や対話を行なっているが、ユーザの心的状態の変化を反映してロボット動作を同調させている例は少ない。

そこで本研究では、ユーザの時系列的な心的変化をセンシングし、その変化に同調して動作を行う対話ロボットを用いることで、ユーザに与える印象の評価を行う。以下に本本研究の目的とアプローチをまとめる。本誌ではユーザの音声に着目し、韻律情報の取得と同調について述べる。

[†] 奈良先端科学技術大学院大学 情報科学研究科

[‡] 津山工業高等専門学校 総合理工学科

[§] 国際電気通信基礎技術研究所 知能ロボティクス研究所

目的：対話ロボットの社会性を高め、ユーザとの信頼関係構築を通しユーザの継続利用意欲の向上を目指すロボットインタラクションの評価を行う。

アプローチ：雰囲気共有を対話ロボットで実現するために、ユーザと同じコンテンツを共に視聴する際に、ユーザの韻律情報、ジェスチャ、姿勢をユーザと同調させ、ロボットの動作を動的に変化させる。

3. ユーザの韻律情報の分析と同調

本誌では雰囲気共有の実現に必要な要素の 1 つである韻律の動的な同調について述べる。ユーザの音声を入力とし、音声から抽出された韻律情報を対話ロボットの音声合成に反映させ、合成された音声の妥当性について検証を行う。

3.1 韻律情報の抽出

ユーザの発話から韻律情報を抽出するために、フリーの音声認識ソフトである Julius[9]と音声解析ソフトである SPTK[10]を用いる。韻律情報には話速、ピッチ、ボリュームの 3 つのパラメータによって音声合成による韻律模倣を行う。マイクから入力された音声を Julius によって音声区間検出および音素数を求めることで話速を導出する。具体的には Julius の音声認識処理の前処理で得られる発話区間長で音素数を除算して得られる、一秒あたりの平均音素数を発話速度を表す特徴量として用いた。ここで簡単のため二重音素も一音素とみなした。

ピッチとボリュームの導出には SPTK を用いる。対話者の発話から取得した wave ファイルの F0 (基本周波数) を抽出し、その対数領域で平均値を求めることにより話者の平均ピッチを推定する。そこから 10 [msec] 毎に分析窓をシフトして F0 を出力する。ここから得られた F0 値から無声音区間を取り除いた F0' の平均値を話者の平均ピッチとして推定する。ボリュームを導出する際に関しても簡単のために音声区間のパワー項の平均値を用い、これを対話者の平均音量として推定する。

3.2 音声合成による韻律模倣

音声合成には Open JTalk [11]を用いて行う。Open JTalk は話速、ピッチ、音量のそれぞれが可変となっており、各パラメータをユーザの音声入力から抽出された韻律情報にそれぞれ対応するよう変換を行う必要がある。3.1 にて述べた韻律の各値から音声合成用の各パラメータへの変換式を作成し、同調の妥当性について検証する。主観的な評価によって、韻律の同調を確認した。

4. おわりに

本誌ではユーザとの信頼関係を構築し継続利用意欲を高める対話ロボットのためのインタラクションとして、雰囲気共有によるユーザとの同調を提案した。雰囲気共有実現に向け、ユーザの音声から韻律を分析し模倣するための韻律情報の推定を動的に行い、推定されたパラメータを元にユーザ発話の韻律情報に同調する合成音声を作成した。合成された音声の主観評価により、韻律情報の模倣の妥当性を検証した。

今後はテレビ視聴型雑談ロボットに雰囲気共有を適用し、テレビ番組におけるシーンの特徴に応じて対話ロボットの韻律やジェスチャを動的に変化させた際のユーザの心理的な変化を主観的・客観的な指標を用いて評価を行う。

謝辞

本研究の一部は JSPS 科研費 18H03274 および、国立研究開発法人科学技術振興機構研究成果展開事業世界に誇る地域発・実証拠点 (リサーチコンプレックス) 推進プログラムの助成を受けたものである。

参考文献

- [1] Cabinet Office Government of Japan Annual Report on the Aging Society, (2016).
- [2] Cabinet Office Government of Japan Annual Report on the Child and Youth (2016).
- [3] K. Miyazawa, et. al., "Factors of Interaction in the Spoken Dialogue System with High Desire of Sustainability," Trans. IEICE (A), Vol. 95, No. 1, pp.27- 36, (2012).
- [4] H. Minami, et. al., "Chat robot coupling machine responses and social media comments for continuous conversation," MUSTEH, (2016).
- [5] 長岡千賀, "対人コミュニケーションにおける非言語行動の 2 者相互影響に関する研究." 対人社会心理学研究 6, 101- 112, (2006).
- [6] T. Bickmore and A. Gruber, "Relational agents in clinical psychiatry," Harvard Review of Psychiatry, vol.18, no.2, pp.119- 130, 2010.
- [7] T. Matsumoto, et. al., "An Emotion-Expressing Robot for Enhancing User's Emotional Experience", IEICE (A), Vol. J99- A No. 1, pp. 45- 55, (2016).
- [8] T. Takahashi, et. al., "A Social Media Mediation Robot to Increase an Opportunity of Conversation for Elderly: Mediation Experiments Using Single or Multiple Robots", IEICE, Vol. 113, No. 84, pp.31- 36, (2013).
- [9] A. Lee and T. Kawahara, "Recent development of open- source speech recognition engine Julius", APSIPA, pp.131- 137, 2009
- [10] SPTK working group, "Examples for Using Speech Signal Processing Toolkit Ver. 3.9", 2015.
- [11] "Japanese TTS system: Open JTalk", <http://open-jtalk.sourceforge.net>