

ユーザとの信頼関係構築を目的としたロボット対話における  
同調的音声合成の評価

Evaluation of empathetic speech synthesis in robot dialogue  
to build a trust relationship with users

中村 卓矢<sup>†</sup>      西村 祥吾<sup>†</sup>      川波 弘道<sup>‡</sup>      神原 誠之<sup>†</sup>      萩田 紀博<sup>†, §</sup>  
Takuya Nakamura   Shogo Nishimura   Hiromichi Kawanami   Masayuki Kanbara   Norihiro Hagita

## 1. はじめに

ライフスタイルの変化に伴い、年齢問わず独居世帯の数が増加している[1]。それによって、日常生活における会話が減少している。日常会話が不足すると、脳機能の低下や、生活環境の悪化によりうつ病や認知症の原因になると考えられている。そのため、このような日常会話が不足している人の発話機会を増やすための対策が重要な課題となっている。

このような課題の対策として、日常的に対話ロボットを利用してコミュニケーションを活性化させる取り組みは従来から数多く研究されてきた。神田ら[2]は、ロボットが社会的な存在として人々に受け入れられるためには人とロボットが友好的な関係を築く必要があると述べている。このように、日常的に対話ロボットを利用してもらうためには人とロボットの友好的な関係が重要であると考えられる。

そこで我々は、ロボットとの対話を人にとって好ましいものにするために、心理学や介護福祉分野で用いられるペーシング[3]と呼ばれる技術に着目した。ペーシングとは、相手の話し方や状態、呼吸などのペースに合わせることである。これにより、相手の警戒心をとり安心感を与えることができるだけでなく、相手の自分への肯定感や重要感を満たし、ラポールを形成することができると考えられている。そこで本研究では、このペーシング技術を人とロボットの会話に応用することによって人とロボットの信頼関係の構築を目指す。

## 2. 関連研究と本研究の位置づけ

本章では関連研究について、2.1.にて対人関係におけるラポール形成に関する心理学的な関連研究を紹介し、2.2.にてヒューマンロボットインタラクション分野における先行研究を紹介する。また、2.3.にてそれらを踏まえた本研究の位置づけを示す。

### 2.1 心理学的な関連研究

対人社会心理学の分野において、同調傾向は多くの研究者によって、円滑なコミュニケーションの指標としてみなされてきた。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、同調傾向はラポールをもたらしたりポジティブな対人印象をもたらしたりすることが示されてきた。このことは、同調傾向がコミュニケーションにおいて何らかの重要な役割を果たしていることを示唆している[4]。

また、ラポールを確率するためのテクニックとしては、

- ・ミラーリング
- ・ペーシング
- ・バックトラッキング

などがある[5]。ミラーリングとは、相手の無意識的な行動パターンを、鏡に映したように真似して、相手にフィードバックするプロセスである。バックトラッキングとは、相手のコミュニケーション内容を、相手にそのままフィードバックするプロセスである。

### 2.2 HRI/HAI における関連研究

ヒューマンロボットインタラクション分野において、ロボットと人間のインタラクションによるラポール形成に関する先行研究を紹介する。

山野ら[6]は人間とロボットのインタラクションにおいて、人間の情動に同調させた際に同調しない場合よりもインタラクション時間が長くなり、ポジティブな印象となることを示した。さらに人間の心理状態が快適な方向に変化したことを示した。

Huangら[7]は RapportAgent と呼ばれるエージェントに対し、信頼関係を構築するための3要素(明るさ、自然な注視、強調)に基づいた会話エージェントを開発した。エージェントと人間の2者間の対話において、人間の発話の有無、視線、頷き、笑顔等をリアルタイムに分析して自らの反応をパターン化できる。この結果、相槌、話者交代等の動作にいくつかの動作パターンが設定され、ラポールの形成に貢献していることが示された。このロボットはミラーリングによる同調的なインタラクションが実装されている。

以上から、ヒューマンロボットインタラクション分野においても、心理学的分野で対人に使用される技術を用いることでラポール形成に有効であることが分かる。

### 2.3 本研究の位置づけ

本研究では、ロボットとの対話において、同調的音声合成を用いた時の信頼関係の構築を目的とする。また、評価方法として主観的評価のみでなく、客観的評価も用いる。同調的音声合成を用いた時の信頼関係構築を客観的指標で評価した事例は少なく、本研究は今後の研究のための予備実験的な位置づけとする。

## 3. 同調的音声合成を用いた時のロボットに対する印象評価実験

本章では3.1.で今回の実験の概要を述べ、3.2.で今回の実験で用いた同調的音声合成システムについて説明し、3.3.で同調的音声合成システムを用いた時のユーザの印象評価手法について述べる。

<sup>†</sup> 奈良先端科学技術大学院大学 情報科学研究科

<sup>‡</sup> 津山工業高等専門学校 総合理工学科

<sup>§</sup> 国際電気通信基礎技術研究所 知能ロボティクス研究所

### 3.1 実験概要

本実験では、同調的音声合成を用いた時にユーザがロボットに対してどのような印象を持つか評価する。今回、20代の男性 6 名を被験者とした。具体的には、ユーザにロボットと対話を行ってもらい、ユーザの発話から平均的なペーシングパラメータを推定する。推定したペーシングパラメータを用いて同調的音声合成を実現し、ユーザと 4 分程度の対話を行ってもらう。なお、今回の対話では、Wizard of Oz(Woz)法[8][9]を用いて、遠隔操作によってロボットとユーザの自然な対話を実現した。また、パラメータを固定値に設定した場合にも 4 分程度の対話を行ってもらい、それぞれの評価値を比較する。この際、それぞれの対話を行う順序は被験者によってランダムとする。実際の実験の概観を図 1 に示す。

なお、今回用いた対話内容については事前に作成したシナリオを 6 つ組み合わせるものを使用することにする。図 2 に実際に対話に用いたシナリオの一例を示す。また、それぞれの対話における内容は抑揚の少ない単調なものとし、内容はそれぞれ異なるものとする。

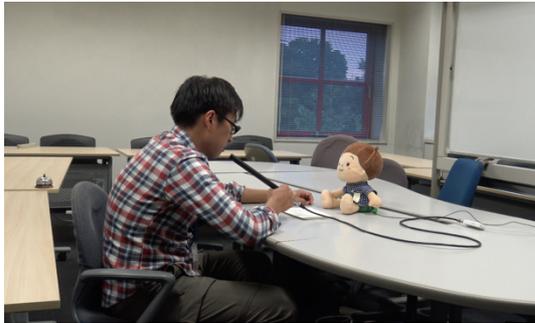


図 1 実験の概観

口： こんにちは！  
 被： こんにちは！  
 口： 好きな食べ物は何ですか？  
 被： 私はラーメンが好きです。  
 口： 何ラーメンが好きなんですか？  
 被： 私は醤油ラーメンが好きです。  
 口： 醤油ラーメンだったらおすすめのお店を知ってますよ。  
 被： どこにあるんですか？  
 口： 大阪に美味しいお店を知っています。  
 被： 何が美味しいんですか？  
 口： チャーシューとスープがとても美味しいです。今度、一緒に行きましょう。  
 被： 行きましょう！

口=ロボット、被=被験者

図 2 対話シナリオの一例

### 3.2 同調的音声合成システム

本研究で用いた同調的音声合成システムについての模式図を図 3 に示す。本システムの構成にあたって Julius[10]、SPTK[11]、OpenJTalk[12]の 3 つのフリーソフトを用いた。対話ロボットはユーザの発話に基づく音声データから

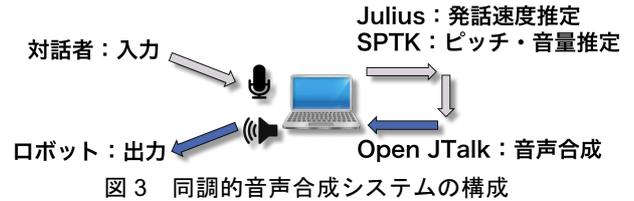


図 3 同調的音声合成システムの構成

Julius, SPTK を用いて発話速度、ピッチ、音量を推定する。ユーザとの対話のやり取りを複数回行い、それぞれの対話から推定された発話速度、ピッチ、音量の平均値を取る。この値を音声合成ソフトである Open JTalk のパラメータに変換して、ペーシングパラメータとして用いることで同調的音声合成を実現する。

### 3.3 ユーザのロボットに対する印象の評価

本実験では、アンケートによる主観的評価と生理指標による客観的評価を行う。特に、生理指標に関しては被験者への負担が少ない LF/HF・HF、唾液アミラーゼを用いた。

#### 3.3.1 アンケートによる主観的評価

ユーザのロボットに対する印象の主観的評価においては、表 1 の 6 種類の評価項目に対して 7 段階のリッカート尺度 (1: 全くそう思わない-7: 非常にそう思う) で質問を行った。これをペーシングがある場合とない場合それぞれの対話において回答してもらう。

表 1 ロボットの印象についてのアンケート項目

- ①ロボットは親しみやすかったですか？
- ②ロボットと一緒にいて楽しかったですか？
- ③ロボットはあなたの話を聞いている感じがしましたか？
- ④ロボットと話していて不快でなかったか？
- ⑤ロボットと話していて違和感を感じなかったか？
- ⑥ロボットを継続的に利用したいと感じましたか？

#### 3.3.2 唾液アミラーゼによる客観的評価

車酔いの客観的評価の手法の一つとして唾液アミラーゼの活性値を用いる。唾液アミラーゼの活性値は、ストレスを定量的に評価する生理指標として一般的に用いられている[13]。唾液アミラーゼの活性は、ストレスによって交感神経系の興奮信号を励起し、自己防衛反応として高まると考えられている。本研究においては、それぞれの対話の前後に唾液アミラーゼ値を測定する。なお、今回の実験においてはペーシングありの対話の方がなしの場合に比べて唾液アミラーゼの活性が小さいという仮定のもと実験を行った。

#### 3.3.3 LF/HF による客観的評価

ユーザのロボットに対する印象の客観的評価の 1 つとして LF/HF・HF を用いる。LF/HF・HF は一般的に自律神経系の活性を定量的に評価する指標である。LF(Low Frequency Components)は 3, 4 秒程度の周期を持つ呼吸を信号とする変動波、HF(High Frequency Components)は約 10 秒周期の血圧変化を信号とする変動波のことである。両者とも心電図から心拍感覚を計算した心拍変動時系列データからパワースペクトル密度を計算することで得られる。ストレス状態時には交感神経の活性により LF/HF の値が上昇し、

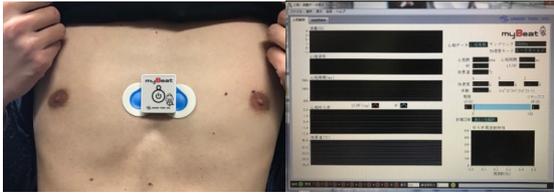


図 4 心拍センサ (左), 解析用ソフトウェア (右)

リラックス状態時には副交感神経の活性により HF の値が上昇する。なお、今回の実験においては対話時のストレス状態の変化を見るために LF/HF の値のみを用いた。これらを踏まえ、ペーシングありのときの対話の方がなしの時に比べ、交感神経の活性が低くなるという仮定のもとに実験を行った。また、LF/HF・HF の値の取得に使用した心拍センサ、解析用ソフトウェアを図 4 に示す。

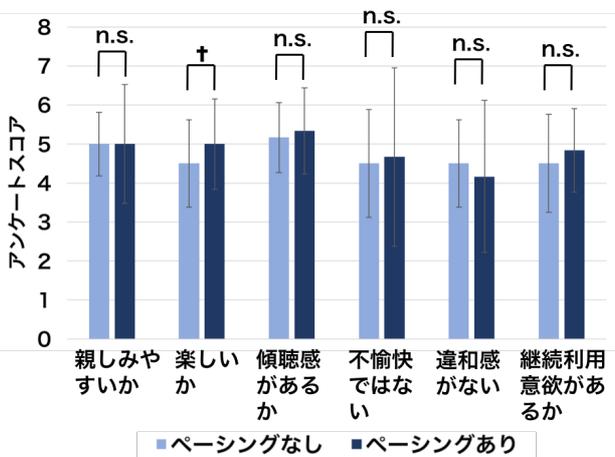
### 3.4 実験結果

#### 3.4.1 アンケートによる評価結果

アンケートによる評価結果として、それぞれの項目の点数の平均、分散および p 値を表 2 に示す。図 5 は、それをグラフ化したものである。表 1 における項目②に関しては有意傾向が見られた。それ以外の項目に関しては、ペーシングありの場合となしの場合で等しい、もしくはほとんど差がなかった。

表 2 同調的音声合成システムの構成

	ペーシングあり		ペーシングなし		p値
	平均	分散	平均	分散	
親しみやすいか	5.00	1.53	5.00	0.82	1.00
楽しいか	5.00	1.15	4.50	1.12	0.08
傾聴感があるか	5.33	1.11	5.17	0.90	0.79
不愉快ではない	4.67	2.29	4.50	1.38	0.79
違和感がない	4.17	1.95	4.50	1.12	0.58
継続利用意欲があるか	4.83	1.07	4.50	1.26	0.36



n.s.: 有意差なし(0.1<P) †: 有意傾向(0.05<P<0.1)

図 5 アンケート結果

#### 3.4.2 LF/HF による客観的評価

図 6 は、実験における LF/HF の時系列データにおいて、初めと終わりの 1 分間における平均値を求め、初めの値に対する終わりの値の増加率をグラフ化したものである。被験者①, ②, ④, ⑤, ⑥に関してはペーシングありの場合に LF/HF の値が減少していることが読み取れる。さらに、被験者⑤以外ではペーシングありの場合の方がペーシングなしの場合に比べて LF/HF の値が減少している。

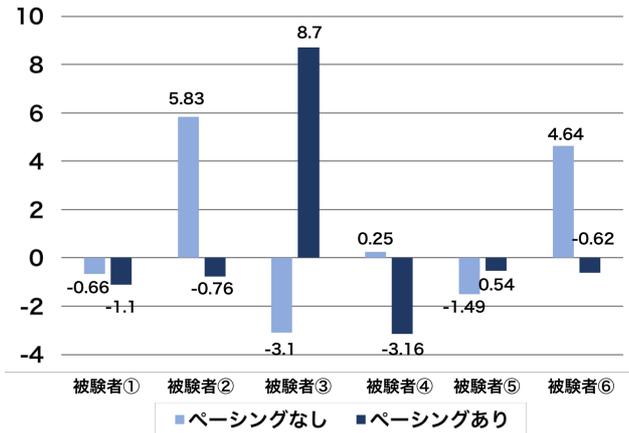


図 6 LF/HF の増加率

#### 3.4.3 唾液アミラーゼによる客観的評価

図 7 は実験開始前の唾液アミラーゼ値の変化をグラフ化したものである。被験者②, ③, ⑥はペーシングなしの場合にアミラーゼ値が増加しており、被験者⑤は双方の場合にアミラーゼ値が減少しているが、ペーシングなしの場合の方が減少量が多い。また、被験者①, ④に関してはどちらの場合もほとんど値が変化していない。

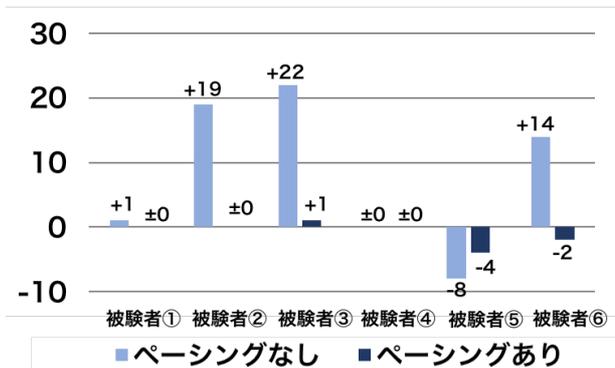


図 7 唾液アミラーゼ値の増加量

### 4. 考察

図 5 のアンケートの結果より、ユーザはロボットに対しペーシングあり、なしの場合においてほとんど印象は変わらないことが読み取れる。これは、ペーシングはユーザの潜在意識に働きかけるテクニックであるため、主観的評価においては差が出にくいと考えられる。

また、図 6 の LF/HF における評価の結果より被験者 6 名の内 4 名においてペーシングありの場合の方がストレス状態が軽減されていることが読み取れる。さらに図 7 の唾液アミラーゼ値の評価の結果より、半数の被験者においてペー

ーシングありの時のの方がストレスを感じにくく、2名の被験者においては会話の影響がほぼ出ていないことが読み取れる。これは人によってアミラーゼ活性の値が出やすい人と出にくい人がいるためであると考えられる。

## 5. おわりに

本研究ではロボットとの対話時にペーシングを利用した時のユーザの印象を評価した。実験の結果、主観的評価においてはペーシングを用いた対話の影響はほとんどないことが確認できた。客観的評価では特に LF/HF においてペーシングありの場合の方がストレスが軽減され、信頼関係の構築に有効であると考えられる。今後は LF/HF、唾液アミラーゼ以外の客観的指標を用いて更に実験を行う。また、対話毎にペーシングを行った時のユーザの印象を評価する。

## 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] 内閣府, “高齢者の生活環境”, in 平成 28 年版高齢社会白書, (2016).
- [2] 神田崇行, 佐藤留美, 才脇直樹, and 石黒浩, “対話型ロボットによる小学校での長期相互作用の試み”, ヒューマンインターフェース学会論文誌, Vol. 7, No. 1, (2005).
- [3] 永野浩二, “人間の存在(presense)を体験する授業の試み” 追手門学院大学教育研究所紀要, no. 30, (2012).
- [4] 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, (2006).
- [5] 加藤雄士, “プロセス・モデルのコーチングに関する一考察-日本における「ニューコード NLP コーチング」の可能性について-”, 産研論集, No. 44, (2017).
- [6] 山野美咲, 薄井達也, 橋本稔, “情動同調に基づく人間とロボットのインタラクション手法の提案”, Human-Agent Interaction Symposium (2008).
- [7] L. Huang, L. Morency, and J. Gratch, “Virtual Rapport 2.0” In 10th International Conference on Intelligent Virtual Agents, (2011).
- [8] Green, A., Huttenrauch, H. and Eklundh, K.S.: Applying the Wizard-of-Oz Framework to Cooperative Service Discovery and Configuration, Proc. 13<sup>th</sup> IEEE International Workshop on Robot and Human Interactive Communication (Ro-Man 2004), (2004).
- [9] Walters, M.L., Dautenhahn, K., Boekhorst, R., Koay, K.L., Kaouri, C., Woods, S., Nehaniv, C., Lee, D. and Werry, I.: The Influence of Participants' Personality Traits on Personal Spatial Zones in a Human-Robot Interaction Experiment, IEEE International Workshop on Robot and Human Communication (Ro-Man 2005), (2005).
- [10] A. Lee and T. Kawahara, “Recent development of open-source speech recognition engine Julius”, APSIPA, (2009).
- [11] SPTK working group, “Examples for Using Speech Signal Processing Toolkit Ver. 3.9”, (2015).
- [12] “Japanese TTS system: Open JTalk”, <http://opentalk.sourceforge.net>
- [13] 山口昌樹, 花輪尚子, 吉田博, “唾液アミラーゼ式交感神経モニタ基礎的性能”, 生体医工学, Vol.45, No.2, (2007).