＊＊＊研究紹介＊＊＊

マルチメディア情報を活用した音声対話システムの研究：利用者の well-being と満足度の向上

橋本 真幸 [1*]

A Research on Dialogue Systems Utilizing Multimedia Information: Enhancing User Well-being and Satisfaction

Masayuki HASHIMOTO[1*]

1. はじめに

筆者は、2023 年 4 月に東洋大学に赴任した。それまでの 26 年間は企業の研究所に所属し、画像符号化や画像認識などのマルチメディア情報処理とその健康医療分野への応用、並びに、コンテンツセキュリティや対話 AI の研究に従事してきた。これらを背景に、本学で新たに研究プロジェクトを立ち上げている。本稿では、その新規研究プロジェクトに関し、研究方針を中心に紹介を行う。

2. マルチメディア情報を活用した対話システム

近年、コミュニケーションロボットやスマートスピーカーといった AI（人工知能）搭載の自律的エージェントが日常生活に浸透してきた。これらのエージェントは、今後人間の日常において重要な役割を果たす存在となっていき、その関係性は単なる道具としての利用を超え、人間と共存する存在となっていくことが考えられる。ここでは、システムと人間とのインタラクションとして音声対話システムに注目する。前述の人間との共存を考えた場合、明日の天気を尋ねる、オンラインサイトで物品を購入するなど、所定のタスクの遂行を目的とする対話（タスク指向型対話）だけでなく、所定のタスクの遂行を目的としない非タスク指向型対話（いわゆる雑談）を行う音声対話システムの重要性が増すことが考えられる。そのため、非タスク指向型対話システムに関する研究が盛んにおこなわれている [1)2)]。

通常、人間同士で行われる雑談は、人間関係構築手段の一つとして考えられている。また、人間関係や社会的繋がりは人々の well-being に大きな影響を与えると多くの心理学的研究で指摘されている [3)]。

本研究では、先述の自律的エージェントが単に対話システムとして機能するだけでなく、マルチメディア情報をうまく活用して、人間同士の雑談のように会話を行うことで、利用者の well-being※や満足度を向上させるための技術に関する研究を行う。（※well-being という語は、一般的に、心身と社会的な健康を意味する概念として用いられるが、ここでは、持続的に幸福感が得られている状態を指す語として用いる。）図 1 に、本研究の概念図を示す。

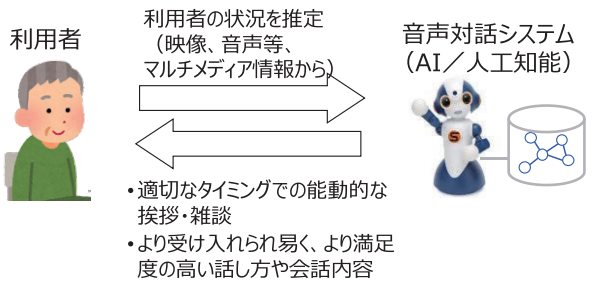

図 1. 本研究の概念図

3. 研究のアプローチ

近年、大規模生成モデルを用いた対話生成技術が注目され、その技術的発展が著しく、対話生成の品質向上に関する研究が盛んに行われている。本研究では、対話内容の生成に関してはこれら多くの既存研究に委ねることとし、利用者の well-being 向上と満足度向上に寄与する他の要因に焦点を当てる。

利用者の well-being 向上について考えた場合、人間同士の会話では、日常的に相手から挨拶されたり、雑

[1] 東洋大学 理工学部 電気電子情報工学科
Toyo University

*Corresponding Author: hashimoto065@toyo.jp

談を持ちかけられたりすることで、自分が存在として認識されていることが確認でき、安心感や幸福感が増すと考えられる。そこで、システム側から日常的に、プロアクティブ(能動的)に挨拶や雑談を行うことで、これと同様の効果を得るための研究を行う（下記4節に記述)。

次に、利用者の満足度向上について考えた場合、対話システムに対する満足度は、発話内容だけでなく、話し方や対話のテンポにも大きく影響されることが容易に想像でき、その一部は先行研究により実証されている[4]。そこで、本研究では、より詳細に、システムの発話の韻律、抑揚、音声の質感など、話し方が利用者の満足感にどれだけ寄与するのかを明らかにすることを目的とする（下記5節に記述)。

4. Well-being向上につながる状況認識技術

コミュニケーションロボットやスマートスピーカーなどの音声対話システムは、人々の生活空間に存在することで、利用者のwell-being向上に寄与する可能性を秘めている。例えば、朝の挨拶や帰宅時の挨拶など、気にかけの一言が利用者に心理的な安心感をもたらすことが期待される。このような細やかなコミュニケーションが、利用者の日常の中での小さな幸せやwell-beingの向上に寄与する可能性がある。

しかし、一方で、これを実現するには課題が多いことも事実である。利用者にとって不適切なタイミングでの話しかけや、誤った状況認識に基づくアクションは、利用者に不快感を与えることにつながる。そこで、利用者の状況をできるだけ正確に認識することが重要となる。コミュニケーションロボットのカメラが利用できる場合、利用者が別の事象に集中しているため話しかけるべきではないことを認識することができるかもしれない。ここで重要となる技術は、映像からの利用者の視線検出技術である。視線検出に関する既存研究としては、顔の向いている方向を推定するHead Pose Estimationに関する数多くの既存研究がある[5]。これらの技術のいくつかはソフトウェア実装として一般に利用可能であり[6]、図2に示すように、顔画像から目鼻の位置と顔の向きを検出できる。本研究においてもこれらの技術を活用する。一方で、既存技術では両目が映っていない場合には、顔方向を正確に検出できないという課題がある。これに対し本研究では、新たにこのような課題を含む画像データセットを作成して学習モデルを作成するなどの手法を通じて、課題解決を図っていく。

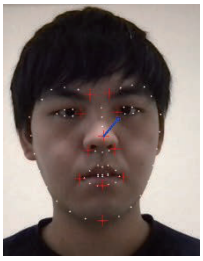

図2. 目鼻位置と顔方向の推定結果

図3. 卓上カメラ画像の例（PC画面を見ている人物）

また別の課題として、特に卓上のコミュニケーションロボットのカメラのように、制約された撮像環境では、利用者の全体像や背景情報の取得が難しく、状況認識の精度が低下することがあげられる。例えば、図3のように、PCの画面を見ているがPCの画面は画像には映っていない場合や、スマホを見ているが手に持っているスマホは映っていない場合などがそれに当たる。本研究では、このような課題に対して、新たにこのような課題を含む画像データセットを作成して学習モデルを作成する手法や、事前に部屋の広範囲を撮影した画像を利用するなどの手法を通じて、課題解決を図っていく。

5. 満足度向上につながる音声発話技術

対話システムの対話に対する満足度は、対話内容だけでなく、話し方や、やり取りのテンポにも大きく影

響される。そこで、これらの要素に関する研究を行う。先行研究によれば、人間同士の対話では、お互いの話速や声の高さなどの韻律が同調する傾向がある。この「引き込み現象」を音声対話システムに応用することで、利用者の満足度の向上が期待される。既存研究では、リアルタイムにユーザ話速に応じてシステム話速を制御するシステムを構築した研究を行い、その効果が確認されている[4)]。本研究では、これをさらに発展させ、話速だけでなく、声の高さ、声の大きさについても制御し、より満足度の高いシステムの要件について明らかにする。

また、現在の音声対話システムでは応答遅延が大きな課題の一つとなっており、人間同士のような自然な会話を妨げる要因となっている。この応答遅延を低減する手法についても研究し、これにより自然でスムーズな対話の実現への貢献を目指す。

6. おわりに

我々の生活の中でAIとの共存が進む中、その関係性は単なる道具としての利用を超えたものとなる可能性がある。現在行われている様々な制度面や倫理面での議論と並行して本研究を実施し、利用者の幸福感やwell-beingを向上させる技術の実現に貢献できれば幸いである。

参考文献

1) 東中竜一郎:『AIの雑談力』, KADOKAWA(2021)

2) 東中竜一郎・稲葉通将・酒井和紀・石黒浩他:小特集「人間機械共生を目指した対話知能システム学の取り組みと今後の展開」, 人工知能, vol.38, no.5, pp.699-737(2023)

3) 遠藤由美・柴内康文・内田由紀子:「人間関係はいかにwell-beingと関連するか」, 関西大学 経済・政治研究所 調査と資料, 第105号 現代社会における人間関係の諸相 第1章, pp.1-28(2008)

4) 三原寛哉・李晃伸:「ユーザフレンドリィな音声対話システム実現のためのユーザ話速および発話内容に基づくシステム話速制御手法の検討」, 情報処理学会研究報告, Vol. 2016-SLP-112, No.15(2016)

5) A. Asperti and D. Filippini: “Deep Learning for Head Pose Estimation: A Survey”, SN Computer Science, 4, 349(2023)

6) V. Kazemi and J. Sullivan: “One millisecond face alignment with an ensemble of regression trees”, 2014 IEEE Conference on Computer Vision and Pattern Recognition, pp.1867-1874(2014)