

画像処理を用いた対話ロボットとの 対話意欲の推定

武田 怜[†] 大和 淳司[†]

[†] 工学院大学情報学部

1. はじめに

非タスク指向型対話システムにおいて、回答に対する不自然な返答、ユーザからの質問を無視、同じ話題の繰り返し等の問題があり、人の対話意欲を低下させてしまう問題がある。一方で、人間同士の会話では自然と非言語情報の共有を行っている。そこで、非言語情報を読み取り、会話に反映させれば対話意欲の維持につながると考える。そのために、非言語情報から対話ロボットと対話している対話者の対話意欲の推定を行う必要があると考える。

2. 目的

非言語情報には無意識的に表れるものの中で、主に表情、視線、体の姿勢、身体動作といった種類がある。このうち、表情の情報は、湯浅らの研究[1]によれば、発話の意欲につなげるのは難しく、身体動作情報のほうが重要であると考えられている。視線、身体動作の計測方法にはアイトラッキングカメラや光学式モーションキャプチャーやなどの技術が使用される[2]。この光学式モーションキャプチャーでは、高精度で高密度なデータが得られる反面、物理的に大きな空間が必要であり、マーカからの測定したデータでは関節間の関係性が成り立たないためデータの修正が必要であった。そのためにマーカレスで少ないカメラでも非言語情報の測定を行えるようにする。

3. 方法

計測を行う情報は、視線と身体動作であり、身体動作を計測する方法として、単眼カメラを使用した。視線情報を計測する方法として、Pupil Coreというアイトラッキングカメラを使用した。そして、ユーザにはコミュニケーションロボット 2 台と対話してもらい、そのユーザの動作を正面から単眼カメラとマイクを使用して発話の様子の録画を行い、OpenPose(姿勢推定ライブラリ)を使用して映像から骨格推定を行った。また、Pupil Core(視線計測デバイス)、Pupil Capture(視線計測ソフトウェア)、Pupil Player(視線解析ソフトウェア)を用いて録画、視線情報の計測を行った。さらに、視線と言っても、その時に何を見ているのかという情報が大事であるため、コミュニケーションロボット 2 台のどちらを見ているかについて色による検出および識別を行った。

4. 結果

非言語情報を計測するにあたり、単眼カメラとマイク、Pupil Core および Pupil Capture を使用して対話の様子の録画を行う必要があるため、簡単に計測が行えるように Python スクリプトを使用した計測システムを作成した。単眼カメラの録画にはライブ配信用ソフト OBS Studio を使用した。OBS Studio と Pupil Capture ではそれぞれの録画タイミングにずれがあるため、タイムスタンプ情報を残すことで出来るだけずれを修正できるようにした。計測システムの概要を図 1 に示す。



図1. 計測システムの構成

実験では、単眼カメラの録画映像を OpenPose に入力することで骨格推定が出来た。あわせて、マイクの音声から発話のタイミングを確認することが出来た。また、Pupil Player によって Pupil Core の録画データから、視線情報を得ることが出来た。しかし、Pupil Core の録画データについて、30 fps(1 フレームあたり約 0.033 秒)を想定して録画を行っているが、フレーム毎の時間が大きくずれる場合があったため OBS Studio による録画データと時間が一致しない箇所が発生した。そのため、時間を一致させるために、各関節位置または、視線情報を補間する必要があると考える。

5. まとめ

本研究では、非言語情報と対話の関係性を見出すために、非言語情報の計測を実現した。今後は、対話システムにある2台のロボットと対話する人の視線や身体動作の計測を行い、発話の様子と視線や身体動作の関係性を見出して行く予定である。

参考文献

- [1] 湯浅将英, 徳永弘子, 武川直樹, “発話マインドに基づく発話交替モデル-気持ちが読める会話インタフェースを目指して”, 情報処理学会インタラクション 2008 論文集, pp.1-8, March 2008.
- [2] 増田 琢弥, 大和 淳司, 杉山 弘晃, “複数台ロボット環境における HRI 分析のためのユーザ動作計測,” 信学技報, vol. 121, no. 93, CNR2021-1, pp. 1-3, 2021, July 2021.