

多人数会話における参与役割を考慮したコミュニケーションロボットの視線制御

Determining Robot's Gaze according to Participation Roles in Multiparty Conversations

吉野 堯[†] 林 佑樹[‡] 中野 有紀子[†]
Takashi Yoshino Yuki Hayashi Yukiko Nakano

1. まえがき

対面コミュニケーションでは、視線や顔向きは話し手へのフィードバックとして重要な非言語行動である。特に、多人数会話では、2人での対話に比べて会話の参与役割が複雑になり、参与役割に応じて顔向き行動も異なる。多人数会話における参与役割は、発話を行っている発話者、その発話が向けられている受話者、そして、現発話が向けられている対象ではないが、会話参加者である傍参与者からなり、発話ターンが切り替わるたびに会話参加者は自分の参与役割に応じて、非言語シグナルの使用を適応的に切り替えてる

複数人のユーザとのグループ会話に参加できるロボットを実現するには、人同士の多人数会話と同様、ロボットは参与役割に応じた顔向き行動を適切に決定することが必要となる。しかし、人間同士の参与役割を考慮した顔向き行動モデルはまだ十分に解明されておらず、ロボットに実装すべき顔向き決定モデルの提案も数少ない[1,2]。そこで、本研究では、人同士の会話を分析することにより多人数会話におけるエージェントの顔向きモデルを構築し、これを実装することにより、参与役割に応じて顔向きを効果的に制御することができるコミュニケーションロボットを実現することを目的とする。

2. データ収集実験

2.1 実験の実施

多人数会話における顔向きモデルを構築するために、対話収集実験を行い、顔向きデータを収集した。

グループで話し合っ、外出先を2か所決めることを課題とし、各グループ3セッション行ってもらった。実験協力者は、20代の男性11名、女性7名、計18人の大学生である。この18人を3人ずつのグループにし、全てのグループに4人目の参加者として、候補となる場所の情報をよく知っている情報提供者となる人が加わった。全てのセッションで同じ人が情報提供者となり、その他の3人の参加者は情報提供者に質問することにより、場所や施設の情報を得ることができた。以上、4人のグループ、6グループで、3セッションずつ会話を行ってもらい、計18会話を収集した。

2.2 収集データ

実験の様子を俯瞰した映像に加え、ヘッドセットマイクによる音声データを収集した。また、webカメラを帽子に装着し、各被験者視点に近い映像を撮影した(図1)。

各被験者の顔向きデータが必要であるが、これを手作業でのアノテーションにより収集しようとする大変な作業

[†] 成蹊大学, Seikei University

[‡] 大阪府立大学, Osaka Prefecture University

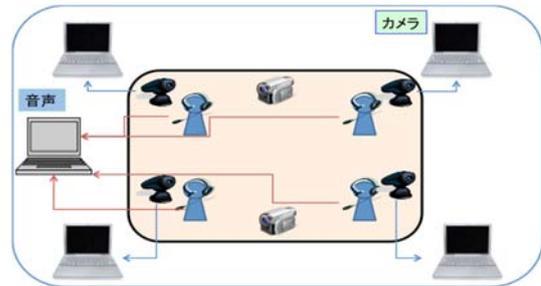


図1 実験環境

量となる。そこで、各被験者の体の前にボードを立ててARマーカを表示し(図2)、Webカメラの映像がこのマーカをとらえ、ARToolkitでマーカが認識されれば、そのマーカの人物の方に顔を向けているものとみなし、マーカを認識した時間、マーカの種類の情報を取得した。この方法により、自動的に顔向きデータを収集することができた。また、4パターンのマーカをあらかじめ登録しておくことで4人を識別した。



図2 ARマーカの設置

2.3 顔向きアノテーション

ビデオアノテーションソフトELANを用いて、注釈付けを行った。得られた音声データから音声認識エンジンJuliusのadinツールを用いて音声区間を取得した。また視線データはARToolkitから得られたデータを判別することにより自動的に注釈付けを行い、顔向きの判定が正確であるかどうかを手で確認した。受話者の認定は手動で行った。また、同時に2つのマーカが認識されてしまうことがあったが、その場合は、後から認識された方に注意を移動した判断し、後から認識されたマーカの対象を顔向き方向とした。

3. データ分析

収集したデータより平均顔向き停留時間、顔向きの遷移先の割合を分析し、顔向き状態遷移モデルを作成した。本研究では、情報提供者となるロボットの顔向きを決定するためのモデルを作成することが目的であるため、情報提供

表1: 発話者の顔向き分析結果

| 次注視対象 / 現注視対象 | 受話者 (正面) | 受話者 (左, 右) | 傍参与者 (正面) | 傍参与者 (左, 右) |
|---------------|-------------|-------------|-------------|-------------|
| 受話者 (正面) | 0.57 (0.61) | 0 | 0 | 0.43 (6.45) |
| 受話者 (左, 右) | 0 | 0.58 (4.74) | 0.23 (1.68) | 0.19 (3.56) |
| 傍参与者 (正面) | 0 | 0.11 (0.66) | 0.89 (6.04) | 0 |
| 傍参与者 (左, 右) | 0.33 (0.33) | 0 | 0 | 0.67 (6.60) |

者である人物の顔向き行動に焦点を当てて分析を行った。また、顔向き行動は参与役割によって異なると考えられるため、情報提供者が、発話者、受話者、傍参与者の場合、それぞれについて顔向き分析を行った。情報提供者が発話者である場合の顔向きの遷移先の割合と平均注視時間を表1に示す。正面にいる受話者を見ている場合、次の注視対象が再度正面にいる受話者である割合が0.57、平均停留時間は0.61秒であり、同様に左右にいる傍参与者に顔向きを変える割合が0.43、平均停留時間は6.45秒であった。

この分析結果をロボットの顔向き遷移確率として利用し、顔向き決定のマルコフモデルを作成した(図3)。

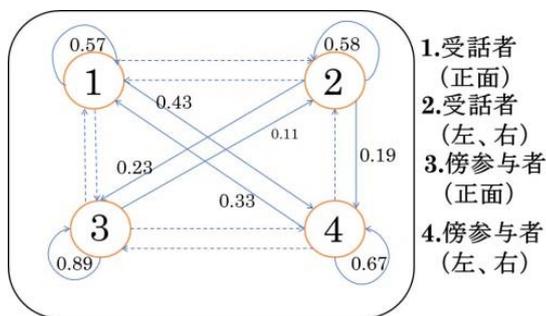


図3 話者の顔向き決定マルコフモデル

4. 顔向きモデルのロボットへの実装

3章で提案した顔向きモデルをヴィストン社製のRobovie-R3を用いて実現する。Robovie-R3はロボットの目がカメラになっているため、このカメラを用いて各被験者のARマーカを認識することが可能である。実装したシステムの構成図を図4に示す。

4.1 発話者推定

ロボットが受話者である場合、ロボットが発話者を特定できれば、その他の人は傍参与者となる。本システムでは、Kinectのマイクアレイを利用して音源方向を特定することにより、誰が話しているかを推定した。

4.2 顔向きの決定

ここでは、3章で提案した顔向き決定モデル(図3)を参照

し、次に誰に視線を向けるかを確率的に決定する。

4.3 ユーザの頭部位置座標の取得

遷移先の顔向きが決定すると、その対象となる人物の顔を見るようにロボットの頭部をどの角度に動かせばよいかを計算する必要がある。図4中の②ユーザ頭部位置取得部では、Kinectのボーン情報から各関節の座標を取得し、その中から頭の位置のx、y座標を対象者の頭の位置として利用した。

4.4 ロボットの顔向き制御

4.3節から顔向き対象者の頭部位置情報が得られると、これをロボットへの動作命令のパラメータとして、ロボットの顔向きを制御する。Robovie-R3は27個の関節を持つが、各モータを同時に動かすことによりスムーズな動きを実現した。

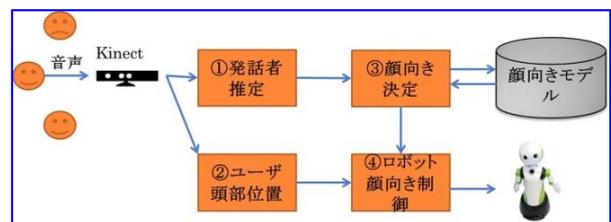


図4 システム構成図

5. まとめ

本研究では、多人数対面会話実験で得られた顔向きデータを分析し、参与役割に応じた顔向き決定の状態遷移モデルを提案した。さらに、これをロボットに実装し、動作を確認した。

本研究ではより確実に顔向きデータを取得できる方法として、ARマーカを用いて顔向き方向を推定したが、このようなマーカを使わない方法、例えば精度の高いヘッドトラッカ等を用いる方法に改良してゆく予定である。また、今後は、情報提供者以外の視線データを分析することにより、ロボットが情報提供者でない場合の顔向き決定モデルも構築してゆく予定である。

<参考文献>

- [1] Samer Al Moubayed, Jonas Beskow, Gabriel Skantze, Björn Granström, "Furhat: A Back-projected Human-like Robot Head for Multiparty Human-Machine Interaction", Cognitive Behavioural Systems, Lecture Notes in Computer Science Volume 7403, pp 114-130, 2012.
- [2] Bilge Mutlu, Toshiyuki Shiwa, Takayuki Kanda, Hiroshi Ishiguro, Norihiro Hagita, "Footing In Human-Robot Conversations: How Robots Might Shape Participant Roles Using Gaze Cues", HRI '09 Proceedings of the 4th ACM/IEEE international conference on Human robot interaction, pp 61-68, 2009.