

優位性と参与役割に基づく視線制御モデルの コミュニケーションロボットへの応用

吉野 堯[†] 高瀬 裕[‡] 中野 有紀子[‡]
成蹊大学 理工学研究科 理工学専攻[†]
成蹊大学 理工学部 情報科学科[‡]

1. はじめに

対面コミュニケーションにおいて会話を円滑に行うためには非言語情報の使用が重要であるが、非言語情報の中でも特に注視行動は多人数会話を円滑に運ぶための重要なコミュニケーションシグナルであり、その重要性は人対ロボットの多人数会話においても実証されている[1]。本研究では、会話の参与役割と優位性(優位性とは会話の牽引する度合い)によって、注視行動に違いがあることを明らかにする。参与役割として、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」、会話には参加しているが発話者でも受話者でもない「傍参与者」の3種類を取り上げる[2]。

次に、参与役割と優位性により注視行動を変化させるコミュニケーションロボットを実装するために、人同士での会話収集実験を行い、参与役割と優位性を考慮した注視行動モデルを構築する。そしてリアルタイムでの会話参加者の参与役割の認識方式を提案するとともに、注視行動モデルをコミュニケーションロボットに実装する。

2. データ収集

コミュニケーションロボットに実装する注視行動モデルを構築するために、人同士の会話を収集する実験を行った。

4人を1グループとし、ある2つのテーマについて5分間ずつの会話を行ってもらった。実験参加者は男性16名、女性4名の計20名、計5グループである。会話データの総時間数は約50分である。

各参加者の正面にKinectを設置し、フェーストラッキングデータを計測した。また、2台のビデオカメラを用いて、会話の様子を俯瞰する映像を録画した。録画した実験映像に対し、アノテーションソフトELANを用いて、各実験参加者の注視対象(右、左、正面の参加者)の注釈づけを行った。さらに、各発話に対して、参与役割(発話者、受話者、傍参与者)の認定を人手により行い、参与役割に関する注釈データとした。

3. 優位性の推定

次に注釈付けにより得た注視対象データと音声データから各参加者の優位性を推定した。これには、以下に示す重回帰式による優位性推定モデル[3]を使用した。

$$\begin{aligned} \text{優位性} = & (0.80) * \text{注視時間} + (0.162) * \text{相互注視時間} \\ & + (0.94) * \text{発話時間} + (0.256) * \text{発話権取得} \\ & + (-0.25) \end{aligned}$$

注視時間: 発話中に他の参加者を見ていた時間の累計

相互注視時間: 発話中に他の参加者と相互注視を行っていた時間の累計

発話時間: 発話時間の累計

発話権取得回数: 前話者の発話終了後2秒以上の沈黙が続いた後、ターンを取得した回数

4. 注視行動モデル

2, 3章で得られたデータや優位性の推定結果を利用し、優位性と参与役割を考慮した注視行動モデルを構築する。

4.1. 分析手法

まず3章で推定した優位性に基づき、各グループにおいて優位性の推定値が最も高い人を「優位性の高い参加者」、優位性推定値の順位が3, 4位の人を「優位性の低い参加者」とした。これに2章で注釈づけを行った参与役割を組み合わせて、優位性の高い発話者/受話者/傍参与者と優位性が低い発話者/受話者/傍参与者、全ての組み合わせについて注視行動を分析した。

例えば、優位性低い受話者の注視モデルの場合、注視対象が発話者でも、優位性が高い発話者(優位性推定値順位が1位)を見る場合と、優位性の低い発話者(優位性値ランキング3 or 4位)を見る場合に分けることが出来る。

注視モデルは、視線遷移割合とそれに対する平均注視時間により構成される。視線遷移割合とは、現在の注視対象から次の注視対象への視線(顔向き)遷移確率である。例えば、現在、優位性の低い傍参与者が優位性の高い発話者を見ている場合、次に向ける視線先が優位性の低い受話者、優位性の低い傍参与者または視線が遷移しない確率を求める。平均注視時間とは、注視対象が遷移した後、次に遷移が起こるまでの平均時間(つまり注視継続時間)である。

4.2. 分析結果

傍参与者の注視行動モデルの例として、傍参与者の優位性が低く、発話者の優位性が高い場合を表1に、傍参与者の優位性が高く、他の参加者の優位性が低い場合表2に示す。

表1ではモデル化対象の傍参与者の優位性が低く、発話者の優位性が高いため、現在の注視先が発話者である場合、次の注視先は71%の確率で再度発話者であり、そこで1.27秒間発話者を注視する。一方表2では、モデル化対象の傍参与者の優位性が高いため、現在の注視対象が発話者の場合、次の注視対象が再び発話者である確率は52%であり、表1に比べて視線を他の参加者に遷移させない割合が低くなっていることが分かる。優位性が高いということはその会話の中心的役割を担っているため、全ての人に視線を配っていることを示している。同様に、

Applying a Gaze Model based on Dominance and Participation Roles to Communication Robots

[†] Takashi Yoshino, Graduate School of Science and Technology, Seikei University

[‡] Yutaka Takase and Yukiko Nakano, Dept. of Computer and Information Science, Seikei University

現在の注視対象が受話者の場合でも、優位性が高い場合には視線を遷移する確率が高くなっていることが分かる。

表 1 モデル対象(傍参与者)の優位性が低く、発話者の優位性が高い場合

次の注視対象 現在の注視対象	発話者 (優位性高)	受話者 (優位性低)	傍参与者 (優位性低)
発話者 (優位性高)	71% (1.27sec)	15% (1.26sec)	14% (1.13sec)
受話者 (優位性低)	27% (0.64sec)	67% (0.53sec)	6% (0.57sec)
傍参与者 (優位性低)	19% (0.87sec)	14% (0.75sec)	66% (0.97sec)

表 2 モデル対象(傍参与者)の優位性が高く、他の参加者の優位性が低い場合

次の注視対象 現在の注視対象	発話者 (優位性低)	受話者 (優位性低)	傍参与者 (優位性低)
発話者 (優位性低)	52% (1.24sec)	20% (0.69sec)	28% (0.66sec)
受話者 (優位性低)	31% (1.19sec)	52% (0.80sec)	17% (0.66sec)
傍参与者 (優位性低)	18% (1.37sec)	19% (0.76sec)	63% (0.75sec)

5. 顔向き推定方式

Kinect により計測されたフェーストラッキングデータから、リアルタイムに各会話参加者の注視方向を認識するための顔向き推定モデルを構築する。

2 章で行った注視対象の注釈を教師データとし、データ収集実験において、Kinect のフェーストラッキング機能を用いて計測した顔の回転角度: pitch, yaw, roll の値のうち、yaw(t-2), yaw(t-1), yaw を学習データとして C4.5 アルゴリズムによる分類器学習を行った。Leave-one-person-out 法によるモデル評価を行ったところ、適合率:0.808、再現率:0.806、F-Measure:0.806 の性能を得た。十分な精度が得られたと判断し、本研究では、Kinect からの計測データをリアルタイムに本モデルに適用することにより、各参加者の注視対象を認識する。

6. 実装

3 章で構築した視線モデルを適用したシステムアーキテクチャを図に示す。各モジュールについて説明する。

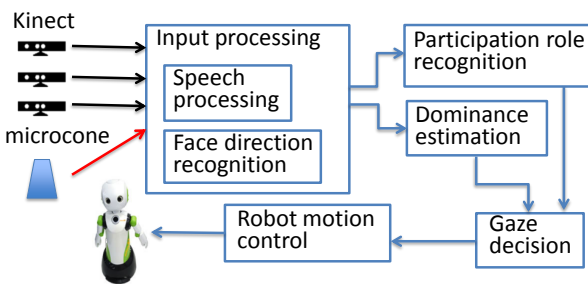


図 1 システムアーキテクチャ

Input Processing: Kinect から得られるフェーストラッキング結果とマイクロフォンアレイを搭載した

microcone¹ から得られる音声データを基に、発話者と各参加者の顔向き推定を行う。まず、Kinect からは各実験参加者の顔の回転角度 yaw が入力される。また microcone は全方向の音源情報を得ることが可能であり、コミュニケーションロボットの前に microcone を設置することで、音源方向を取得することができる。これを用いて Speech processing モジュールではその音源方向にいる参加者を発話者と推定する。Face direction recognition モジュールでは kinect から受信した参加者の顔の回転角度から 5 章で述べた顔向き推定方式を用いて顔向きを推定する。

Participation role recognition: 発話者と各参加者の顔向き情報から参与役割を推定する。発話者は先ほどの Speech processing モジュールで推定されている。受話者推定では発話者推定直後にその発話者が視線を向けている参加者を受話者とする。そして、発話者、受話者でない会話参加者を傍参与者と推定し、各会話参加者の参与役割を明確にする。

Dominance estimation: 発話者と各顔向き情報から各会話参加者の優位性を計算する。優位性の推定方法として、3 章で説明した優位性推定モデルを適用する。まず、注視時間、相互注視時間、発話時間、発話権取得回数を計算し、その結果から優位性推定モデルを用いて各会話参加者の優位性を計算する。

Gaze decision: 4 章で構築した注視行動モデルを用いて、コミュニケーションロボットが次に向ける視線先を決定する。そのために各会話参加者の優位性と参与役割の情報から注視行動モデルを用いて、次に向ける視線先を計算し、その視線先の情報を Robot motion control モジュールに送る。そして平均時間経過後、再び確率を計算し次の視線先を決定する。

Robot motion control: Gaze decision から入力された次の視線先の情報に基づきコミュニケーションロボットに顔向きの命令を送信し、ロボットの視線を動かす

7. まとめ

参与役割と優位性により人は注視行動を変化させるため、本研究では優位性と参与役割を考慮した視線モデルを構築した。まずは人同士の会話収集実験を行い、得られたデータから優位性と参与役割を考慮した視線分析を行い、注視行動モデルを構築、実装した。今後は評価実験を行い、この視線モデルの妥当性を検証する。

文 献

[1] Mutlu, B., Shiwa, T., Kanda, T., Ishiguro, H., and Hagita, N. 2009. Footing In Human-Robot Conversations: How Robots Might Shape Participant Roles Using Gaze Cues, in the 4th ACM/IEEE International Conference on Human-Robot Interaction (HRI '09), p. 61-68.

[2] Goffman, E. 1981. Forms of Talk, Philadelphia, PA: University of Pennsylvania Press.

[3] Nakano, Y. I. and Fukuhara, Y. 2012. Estimating Conversational Dominance in Multiparty Interaction, in 14th ACM International Conference on Multimodal Interaction (ICMI2012), p. 77-84.

¹ <http://www.dev-audio.com/products/microcone/>