グループディスカッション参加ロボットの頭部動作の印象評価

Impression Evaluation of Head Movement of Group Giscussion Participating Robots

木村 清也 [1,2] 黄 宏軒 [1,2,3] 桑原 和宏 [2] 西田 豊明 [2,3]
Seiya Kimura[2] Hung-Hsuan Huang[1,2,3] Kazuhiro Kuwabara[1] Toyoaki Nishida[2.3]

[1] 国立研究開発法人理化学研究所革新知能統合研究センター
[1] RIKEN Center for Advanced Intelligence Project
[2] 立命館大学大学院情報理工学研究科
[2] College of Information Science and Engineering, Ritsumeikan University
[3] 京都大学大学院情報学研究科
[3] Graduate School of Informatics, Kyoto University

Abstract: 近年の日本企業が社員に求める能力としてコミュニケーション能力があげられる．それに伴って，多くの企業が就職採用選考おいて，就職活動者のコミュニケーション能力を評価するためにグループディスカッションを取り入れている．こうした背景から我々は現在グループディスカッションに参加可能なロボットの開発を進めている．本論文では，これまで提案していたロボットの頭部動作のアテンション対象と頷きモデルの印象評価について結果を報告する．印象評価実験では，クラウドソーシングサービスで募った参加者に，本論文で提案されているモデル条件，その他の確率条件，正解条件で動作するロボットの映像を閲覧し，自然さを評価してもらった．その結果，頷きに関する評価項目にてモデル条件が最も優れている評価を得た

1 はじめに

近年，就職採用の場において，グループディスカッションを採用している企業が増加しており，コミュニケーション能力の方が専門スキルよりも重視されるというケースもある．企業の採用担当者はグループディスカッションの中で，求職者の個性やコミュニケーション能力を評価する．したがってコミュニケーション能力を向上させることによって就職活動を成功させる可能性を高めることができるのである．コミュニケーション能力は反復練習によって向上させることができると考えられているが，反復練習を行うためには相応のパートナーが必要となっており，主な就職活動者である学生にとって訓練パートナーを用意することは簡単ではない．そこで我々は仮想エージェントやロボットとグループディスカッションを行うための訓練システムの開発を目指している．グループディスカッション訓練システムでは複数のユーザーとロボットがコミュニケーションをとる必要がある．複数のユーザーとロボットとのコミュニケーションに着目した研究はこれまでにもなされてきたが，その多くは人間のユーザーとロボットが対話中の役割が異なっている [1]．一方我々の研究ではエージェントやロボットを他の人間の参加者と同様の立場としてディス カッションに参加させることに焦点を当てているため，それ専用の動作モデルが必要となる．グループディスカッションではロボットの他に 2 人以上の会話パートナーが存在するため，ロボットは注視行動や頷きはスムーズなコミュニケーションを行うためには必要不可欠である．そこで本稿ではこれまで提案してきたロボットの頭部動作のアテンション対象と頷きモデルの印象評価の結果を報告する．

2 関連研究

人間同士のコミュニケーションに関する研究では，視線は対面での会話において重要なコミュニケーション行動であることが示されている．発話者は受話者を見て態度や話の理解度などをモニタリングしている．一方で受話者の発話者に対する視線についても，発話者に対して肯定的な反応を返す意味もつことが示されている [2]．また対話相手の目を凝視することもまた重要な意味を持っており，発話者が発話を終了し，発話ター

ンを譲る場合にも次の発話者の目を凝視する [3]．また，Vertegaal[7] は，視線情報は受話者の推定に有用であることが主張されている．また，人間が対話中に行う頭部による頷き行動は発言に対する肯定や強調，意図の提示等といたフィードバックだけではない様々な機能を有していることが示されている [4]．

複数人会話であるグループディスカッションに参加できるロボットを実現するためには，1 対 1 の対話にはない課題を乗り越える必要があり，Traum[8] では，複数のユーザーとエージェントとのインタラクションを実現する上での主要な課題について検討されている．複数のユーザーとエージェントとのインタラクションについての研究はいくつか存在するが，その多くは人間のユーザーとはエージェントは異なる立場として扱われている [10, 11]．このほかに，Schiavo[9] らはグループメンバーの非言語行動を観察することによって，参加型のディスプレイを通じて自動的に参加者に指示を与え，グループ活動のコミュニケーションの流れをサポートするシステムを提案した．

これまでになされてきた複数のユーザーとエージェントとのインタラクション関する研究の多くはエージェントが対話の調整役，あるいは聴き手としての対話に参加するなどエージェントと人間の立場が対等でない．一方我々の研究では，人間とエージェントが対等な立場でグループディスカッシを行うことを目標としているため，それに向けた実験，あるいはシステムの構築に取り組んでおり，これまで注視行動に関するモデルと頷き行動に関するモデルを提案してきた．

3　提案モデル

この章ではこれまでの我々の研究 [13] で提案してきた「アテンション対象モデル」と「頷きモデル」について述べる．これらのモデルはグループディスカッションに参加しているロボットの周りの参加者からマルチモーダルな情報を収集し，その情報を基に頭部動作を決定するモデルである．

3.1　データコーパス

MATRICS コーパスと同じ実験手順に従いデータコーパスの収集を行った [12]．実験に参加した 40 人の大学生のうち 30 人が男性，10 人が女性で全員が日本人のネイティブスピーカーとなっている．参加者は互いに顔見知りでない学生 4 人を 1 グループとした 10 グループの合計 40 人分の データによって構成されている．また，今回の実験では文化祭実行委員のメンバーとなって配布した資料の中に記載された 15 名の有名人の中から収益や集客を考慮し，最適だと思われる人物の順位をつけていく「学園祭に招待する 有名人ランキング」を議題として設定した．実験参加者は正方形の机の周りに座り，議論を行う．

3.2　アテンション対象モデル

本研究ではロボットの眼球方向と頭部方向の総合的な組み合わせを「注視対象」として扱う．これは人工的なエージェントはやロボットはアクチュエータが人間と同じように機能しないことに起因する．このモデルでは周りの参加者の行動を基にロボットが左側の参加者，正面の参加者，右側の参加者，および配布された資料が置かれているテーブルの 4 つのうちから注視する対象を決定する．またロボットがグループディスカッションに参加した場合， 話し手やあるいは聴き手などの役割を担いつつ議論に参加することになる．そこでロボットがグループディスカッションに参加した場合に起こりうる 3 種類のシチュエーションを想定し，Speaking : ロボット自身が発言している場合．Listening : ロボット以外の参加者が発言している場合．Idling : ロボットを含む参加者全員が発言していない時のこれら 3 種類のモデルを作成した．モデルを作成するにあたり，グループディスカッションに参加している 4 人の参加者のうち 1 人をロボット見立て，その参加者を「センター参加者」として定義し，その他の 3 人の参加者の行動から，言語，注視対象，発話ターン，韻律，頭部動作の 5 つのモダリティで合計 122 の特徴量 0.1 秒ごとに抽出した．モデルの生成には 10 グループ 40 人分のデータを使用し RBF カーネル (Gaussian カーネル) を用いた support vector machine (SVM) によって 3 種類のシチュエーションの予測モデルを生成した．モデルの評価は leave-one-person-out 法にて交差検証を行った．それぞれのモデルの F-Measure 値は，Speaking モデル : 0.460，Listening モデル : 0.580，Idling モデル : 0.528 となっている．モデルはチャンスレベル (25 %) よりも大幅に優れており，グループディスカッション参加者の注視対象に傾向があることを示している．また，モデル精度が Listening > Idling > Speaking となっていることから，参加者が発話を行っていない場合は他の参加者から受ける影響が大きくなることを示している．

3.3　頷きモデル

このモデルはロボットがグループディスカッション中に他者の発言に対して，或いは自分の発言に対して頷くべきかどうかを判定するモデルである．このモデルにおいてもシチュエーションに応じて別々のモデルを生成するが，一つの発話に対して判定を行うモデルであるため，Speaking と Listening モデルの 2 種類の

モデルを作成する．特徴量は注視対象モデルと同様に 5 種類のモダリティが存在し，データコーパスについても注視対象モデルと同様であるが，特徴量は判定を行う発話区間でのみ抽出する．モデルの生成には注視対象モデル同様，RBF カーネル SVM によって用い，leave-one-person-out 法にて交差検証を行った．それぞれのモデルの F-Measure 値は，Speaking モデル:0.648，Listening モデル：0.683 となっている．このモデルは 2 クラス分類であり，チャンスレベルは 50 %であることを考慮すると，パフォーマンスは中程度であるといえる．モデル精度が Listening > Speaking となっており，参加者が発話を行っているときに頷く理由は多様で，他の参加者の行動から予測するのがより困難であることを意味している．

4　印象評価実験

4.1　概要

我々の研究ではグループディスカッション訓練システムの開発を目標としており，訓練システムのロボットは人間の代替としてグループディスカッションを行う．そのためロボットの振る舞いは自然であることが望ましい．そこでこの実験では，我々がこれまで提案してきた注視行動に関するモデルである「アテンション対象モデル」と「頷きモデル」を基に動作するロボットの自然さとその他の条件に基づいて動作するロボットの自然さを第三者の主観によって比較評価を行う．なお，この実験における「自然さ」とは「ロボットがグループディスカッション中に他の参加者の話を聞く場合や或いは発言している場合の振る舞いとして適切でかどうか」と定義する．また，ロボットは人間と比べて自由度が低く，人間のような微妙な注視行動を再現できない可能性がある．そこで本研究ではロボットの頭部方向或いは眼球方向の総合的な組み合わせを「注視行動」として扱う．

実際のグループディスカッションによって評価実験を行った場合，評価者は比較する条件ごとに異なる状況でそれらを評価することになる．そのため他の要因に影響され，注視行動と頷きの自然さのみを比較評価できない可能性がある．そこで評価者は同じ参加者の同じ会話内容でロボットの振る舞いのみが違う動画を 3 本見て，それぞれの動画で比較しつつロボットの振る舞いの自然さについて比較しつつ評価するという方法をとった．実験に使用する動画はグループディスカッションを行う 4 人の参加者の顔映像を用いて，そのうち 1 人の映像を条件に基づいて動作するロボットの顔映像に置き換えたものである．図 1 に実験映像のスクリーンショットとグループディスカッション参加者の実際の配置を示す．

4.2　頭部動作モデル

グループディスカッションの参加者は正方形のテーブルの周りに座って議論をしており，ロボットは左側の参加者，正面の参加者，右側の参加者，および配布された資料が置かれているテーブルの 4 つのうちのいずれかの方向への注視行動と各方向を向きながらの頷き行動を行う．これらの決定するための頭部動作モデルを 3 種類用意した．モデルは議論中 1 フレーム (0.1 秒) ごとに注視する対象の決定と頷きを行うか否かの判定を出力する．なお，今回は純粋にロボットの頭部動作のみを評価するために元々の参加者の音声をそのまま使用し，評価者にはロボットが発話しているものとして評価するように指示をした．以下に各行動モデルの詳細を述べる．

1．提案モデル条件
　前章にて述べた提案モデルを基に頭部動作を行う．

2．正解モデル条件
　置き換えられた元々の参加者と同様の頭部動作を行うモデルである．元々の参加者に対して注視対象と頷きのアノテーションを行い，その結果を基に頭部動作を行う．

3．確率モデル条件
　確率によって頭部動作を決定するモデルである．提案モデルを生成するために使用されたデータコーパスに対して行った注視対象と頷きのアノテーション結果から，各注視方向に向いていた時間割合と発話終了時に頷く確率の平均値を算出し．その確率によってモデルの出力を決定する．

4.3　動作フィルタ

モデルでは 1 フレームごとに注視する対象の決定と頷きを行うか否かの判定を出力する．しかし，ロボッ

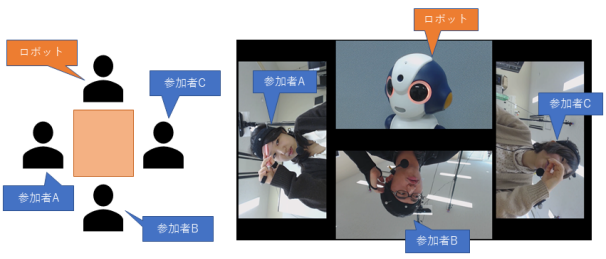

図 1: グループディスカッション参加者の実際の配置 (左) と実験映像 (右)

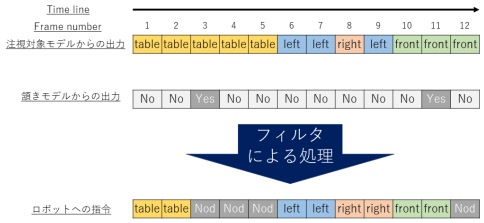

図 2: 動作フィルタの処理例．注視切替と頷き中は注視行動モデルからの出力を無視し，頷きモデルからの出力は注視切替が終了次第，頷きを開始する．

トには物理的な制約があり，注視方向を変える場合にも一定の時間が必要となる．そのため，1 フレーム毎のモデルの決定に対応することができない可能性がある．また，頷きを行っている間は注視対象を変更できないなど．モデルよる頭部動作の決定をロボットが対応できるように加工する必要があるそこで今回は，注視方向の切り替えにかかる時間と頷きの継続時間がどのような時間でも対応出来るような動作フィルタを作成し，処理を行った．図 2 にフィルタ処理の例を示す．図では例として注視の切り替えに 2 フレーム頷き 3 フレーム分を要する場合の例を示している．

4.4 評価

ロボットはグループディスカッション中に注視行動と頷きを行う．今回の評価実験では，それらの頭部動作をモダリティごとに自然さを評価するための項目と，両方の頭部動作を含めた全体的な自然さを評価するために以下の評価項目を設けた．

1．注視する対象とそれを切り替えるタイミングの自然さ

議論中のロボットの注視方向に対する印象について．議論の流れに則した適切な方向への注視やその切り替えのタイミングが適切であるか評価する．

2．頷きタイミングの自然さ

議論の流れに則した適切なタイミングで頷きを行えているかどうかについて評価する．

3．以上の両方を含めた全体的な自然さ

ロボットの注視行動や頷き行動の両行動を踏まえた全体的なロボットの振る舞いに対する印象について評価する．

評価者は 3 条件の動画を比較しつつこれらの評価項目を 10 段階で評価を行う．

対話実験 20 人分データを使用して，その 1 人分である条件ごとの 3 本の動画を 1 セクションとし，ランダムに 3 セクションを組み合わせたものを 1 セットとした．これを 20 セットを用意した．評価者にはランダムに 1 セットを配布する．評価者は 1 セクションごとに 3 本の動画比較しそれぞれに上記の項目に従い評価を行う．それを 3 セクション繰り返し評価を行う．

4.5 実験結果

実験には合計 140 人（男性 84 人女性 56 人）平均年齢 39.3 歳（標準偏差=8.8）に参加してもらった．1 セット当たり 7 人に回答が得られ，その結果 1 セクションあたり 21 件，合計で 420 件のデータが得られた．表 1 には獲得した評価の平均値と標準偏差，セクションごとの平均評価値と動作回数の相関係数，また，三群以上対応ありのノンパラメトリック検定であるフリードマン検定を行った結果の有意確率を示す．注視行動の評価項目ではモデル条件は確率条件に次いで平均評価値が高かったが，条件ごとに有意な差はみられなかった．頷きの項目ではモデル条件が最も平均評価値が高く有意な差がみられた．全体の評価の項目においても，モデル条件の平均評価値が最も高かったが，有意な差は見られなかった．

5 考察

正解条件は人間行動を基にロボットを動作させたものの，高い評価を受けることができなかった．これは参加者が本来行っている動作が「注視」「頷き」の二種類に単純化されたために，条件間での差が出にくかったものと思われる．注視行動では，3 条件間での評価値に有意な差が見られなかったが，提案条件および正解条件では，注視対象の切替回数と評価に正の相関がでている．今回の実験では評価にかける時間が 1 分間だったが，実際のグループディスカッションではさらに長い時間議論を行うことが予想されるため，確率条件よりも優れた評価を受けることが期待できる．また頷きの項目ではモデル条件が最も評価が高く，有意な差も見られたため、頷きモデルの有用性も示されたといえる．全体の評価という面では条件間での有意な差は見られなかったが，確率条件では動作回数が多ければ多いほど評価が悪くなるという傾向が見られた．この結果から確率で動作するロボットを用いてグループディスカッションを行った場合，動作を重ねるに従って他の参加者に与える違和感が大きくなる可能性がある．一方でモデル条件では頭部動作回数が多ければ多いほど総合評価が高くなっているため，動作の繰り返しによって与える違和感は小さいと考える．このことから

表 1: 各条件ごとの評価値の統計結果

評価項目	提案	正解	確率	検定
	4.82	4.80	4.95	n.s.
注視	2.15	2.22	2.00	
	0.58	0.52	-0.1	
	5.06	4.66	4.88	*
頷き	2.15	2.23	1.97	
	0.34	0.72	0.36	
	4.91	4.78	4.90	n.s.
総合	2.07	2.18	1.89	
	0.60	0.60	-0.21	

上段：平均，中段：標準偏差，下段：相関係数
(*：p<0.05, .n.s.：有意差なし)

ロボットがグループディスカッションに参加する場合には，根拠なく動作するロボットよりも大量のデータを用いてモデルを構築し，一定の根拠を基に動作するモデルを用いたマルチモーダルフレームワークを実装したロボットが適していると考える．

6 終わりに

本研究では，コミュニケーション能力の向上を支援するシステムの構築を目指し，グループディスカッションに参加するロボットのためのマルチモーダルフレームワークを提案した．グループディスカッションの会話実験を行い収集したコーパスを基に「注視対象モデル」と「頷きモデル」をロボットがグループディスカッションに参加した場合に起こりうるシチュエーションをごとに構築した．それらのモデルを用いてマルチモーダルフレームワークを実装したロボットの頭部動作の「自然さ」とその他の条件に基づいて動作するロボットの「自然さ」を第三者の主観による比較評価を行った．結果は頷きの評価項目において提案しているフレームワークが最も評価が高く，有意な差がみられた．今後の展望としては注視対象と頷きの両モデルの更なる精度向上を目指すとともに本稿で提案したマルチモーダルフレームワークをリアルタイムに動作可能な形でロボットへ実装し，実際のグループディスカッションに参加させることが挙げられる．

参考文献

[1] MarynelVazquez,ElizabethJ.Carter, Braden McDorman,Jodi Forlizzi Aaron Steinfeld,and Scott E. Hudson: Towards robot autonomy in group conversations Understanding the effects of body orientation andgaze. In12thACM/IEEE Internationa lConferenceon Human-Robot Interaction pp. 42-52 (2017)

[2] Adam Kendon.:some functions of gaze direction in social interaction. Acta Psychologica 26, pp. 22-63.(1967)

[3] Starkey Duncan ： Some Signals and Rules for Taking Speaking Turns in Conversations. Journal of Personality and Psychology 23, pp. 283?292. (1972),

[4] Tomio Watanabe, Ryusei Danbara, and Masashi Okubo. Effects of a speech-driven embodied interactive actor interactor on talker's speech characteristics. In Proceedings of IEEE International Workshop on Robot-Human Interactive Communication, pp. 211-216, (2003).

[5] Michael Katzenmaier, Rainer Stiefelhagen, and Tanja Schultz:Identifying the Addressee in Human-Human-Robot Interactions based on Head Pose and Speech. In Proceedings of the 6th international conference on Multimodal interfaces (2004).

[6] Hung-Hsuan Huang, Naoya Baba, and Yukiko Nakano:Making Virtual Conversational Agent Aware of the Addressee of Users' Utterances in Multi-user Conversation from Nonverbal Information. In 13th International Conference on Multimodal Interaction, pp. 401-408 (2011)

[7] Roel Vertegaal, Robert Slagter, Gerrit van der Veer, and Anton Nijholt:Eye gaze patterns in conversations: there is more to conversational agents than meets the eyes. In Proceedings of the SIGCHI conference on Human factors in computing systems. pp. 301?308 (2001)

[8] David Traum : Issues in Multiparty Dialogues:In Advances in Agent Communication, International Workshop on Agent Communication Languages pp. 201-211(2003)

[9] Gianluca Schiavo, Alessandro Cappelletti, Eleonora Mencarini, Oliviero Stock, and Massimo Zancanaro ： Overt or Subtle? Supporting Group Conversations with Automatically Targeted Directives, In Proceedings of the 19th international conference on Intelligent User Interfaces, pp. 225-234 (2014)

[10] Iolanda Leite, Marissa McCoy, Monika Lohani, Daniel Ullman, Nicole Salomons, Charlene Stokes, Susan Rivers, and Brian Scassellati:Emotional Storytelling in the Classroom: Individual versus Group Interaction between Children and Robots. In 10th ACM/IEEE International Conference on Human-Robot Interaction, pp. 75-82 (2015)

[11] Marynel Vazquez, Elizabeth J. Carter, Braden McDorman, Jodi Forlizzi Aaron Steinfeld, and Scott E. Hudson:Towards Robot Autonomy in Group Conversations: Understanding the Effects of Body Orientation and Gaze, Proceedings of the 2017 ACM/IEEE International Conference on Human-Robot Interaction, pp. 42-52 (2017)

[12] Fumio Nihei, Yukiko I. Nakano, Yuki Hayashi, HungHsuan Huang, and Shogo Okada : Predicting In?uential Statements in Group Discussions using Speech and Head Motion Information. In Proceedings of the 16th International Conference on Multimodal Interaction, pp. 136-143 (2014)

[13] Huang, H.H., Kimura, S., Kuwabara, K., and Nishida, T.: Proposal of a Multimodal Framework for Generating Robot ' s Spontaneous Attention Directions and Nods in Group Discussion, Proc. FAIM/ISCA Workshop on Artificial Intelligence for Multimodal Human Robot Interaction, pp. 15-18 (2018)

[14] N. V. Chawla, K. W. Bowyer, L. O. Hall, and W. P. Kegelmeyer. Smote: Synthetic minority oversampling technique.Journa lOf Arti?cial Intelligence Research, 16 pp. 321-357 (2002)