情報処理学会論文誌 Vol.58 No.5 967–980 (May 2017)

多人数会話における調整・介入機能を有する対話ロボット

高瀬 裕[1,a)] 吉野 尭[1,b)] 中野 有紀子[1,c)]

受付日 2016年8月5日, 採録日 2017年2月9日

概要：人と共生するロボットの実現には人同士の会話に対して適切に参入できる機能が必要である．本研究では，複数人のユーザによる会話への円滑な参加と会話による介入を全自動で行うシステムを提案する．その実現のために，会話データ収集を通して会話参加者の振舞いの分析とモデル化を行い，システムへと実装する手法をとった．実現した会話介入システムは，会話中のユーザの優位性や参与役割といった状態を認識し，適切な注視行動を行うことに加え，システム自身に発話権があることを認識することで適切なタイミングでの発話が可能である．評価実験の結果，本研究が提案する優位性や発話権を考慮した会話介入はそうでない介入に比べて優位性が低い参加者の発話を促すことが可能であることが示された．

キーワード：多人数会話，会話優位性，次発話者予測，会話介入，コミュニケーションロボット

Conversational Robot with Conversation Coordination and Intervention Functionality in Multi-party Conversations

Yutaka Takase[1,a)] Takashi Yoshino[1,b)] Yukiko I. Nakano[1,c)]

Received: August 5, 2016, Accepted: February 9, 2017

Abstract: Towards a collaborative environment between humans and robots, robots are required to be capable of properly participating in multiparty conversations. This study proposes an autonomous robot that can participate and intervene into multiparty conversations. First, we collected human conversation data and based on the data analysis, we established participants' behavioral models. Then, we implemented the models into a conversational robot system. The system can estimate each participant's conversational dominance and participation role. It can also produce attentional behaviors and intervene in the users' conversation in a proper timing by predicting a chance that the robot can take the next turn. The results of an evaluation experiment showed that the proposed system, which considers dominance and participation roles in deciding intervention timing, more effectively encouraged less dominant participants to take a turn, compared to that without such functions.

Keywords: multi-party conversation, conversational dominance, next speaker prediction, conversational intervention, communication robot

1. はじめに

近年，人型ロボットが広く認知され始め，店頭や窓口等で見かける機会が増えている．今後，ロボットはより身近なものとなり，我々の生活により深く関わることになると予想されるが，人と共生するロボットを実現するためには，人同士の会話に参入できるコミュニケーション機能が必須となる．たとえば，相談や議論をしている人たちがロボットに情報提供を求めたりする状況において，ロボットは，単に情報検索端末として機能するのではなく，会話の参加者としてユーザに受け入れられることが望ましい．

そこで本研究では，複数人のユーザとの対話に適切に参加しつつ，会話への介入も行うコミュニケーションロボットを提案する．このようなコミュニケーション機能を有するロボットを実現するためには，第1に，ユーザの発話に適切にフィードバックを返しながら，円滑な対話進行に貢献する行動を行うロボット行動決定機構が必要となる．そして，第2に，ロボットが発言してよいタイミングを見

[1] 成蹊大学理工学部
Seikei University, Musashino, Tokyo 180–8633, Japan
[a)] yutaka-takase@st.seikei.ac.jp
[b)] oucgll.5609@gmail.com
[c)] y.nakano@st.seikei.ac.jp

© 2017 Information Processing Society of Japan

つけ，積極的な介入を行い，それがユーザ同士のコミュニケーションに有効に働く必要がある．ユーザとの1対1の対話において，適切なフィードバックを返すことにより，ユーザの発話を促す会話エージェントはこれまでにも提案されているが [1]，本研究では，複数人のユーザとの多人数会話において，ユーザ同士のコミュニケーションに影響を与えることができる会話ロボットの実現を目指す．

具体的には，提案システムでは，対話制御において重要な非言語行動である注視行動に着目し，会話中の各ユーザについて，優位性と各発話における参与役割とを推定し，これらに基づいてロボットの注視行動を自動的に決定する．ここで，優位性とは，その参加者が会話を牽引している度合いのことである [2], [3]．また，本研究では，参与役割として，現ターンの発話者，その受け手となる受話者，そのいずれでもない会話参加者である傍参与者の3つの役割を対象とする [4], [5]．各ユーザの優位性と参与役割に基づき，ロボットが注視行動を変化させることにより，複数人ユーザとの多人数会話において，視線による適切なフィードバックを表出することができるロボットが実現すると考える．また，ロボットによる会話への介入行動を実現するために，次に発話者となる参加者（以降，次発話者と呼ぶ）を予測するモデルを作成し，これを実装することにより，ロボットが発話権を取得してよいタイミングであるか否かを判断する．ロボットが発話権を取得してよいタイミングであれば，そこでユーザの意見を求める等，議論を発展させるための発言を行うことができる．つまり，本研究が目指すロボットは，誰が会話のリーダであり，誰が誰に話しかけているのかを認識しながら，適切に応答しつつ，人同士の会話に介入するロボットである．

以上に述べたロボットのコミュニケーション機能の実現において，実装機能についての設計基盤が明確であることが重要であると考える．そのために，本研究では，人間同士のコミュニケーションを分析・モデル化し，これに基づき，ロボットのコミュニケーション行動を設計，実装するというアプローチを採用する．このようなアプローチをとることにより，どのようなロボット行動を実装する必要があるのか，またそれはなぜなのかといった点について根拠のある説明が可能となり，インタラクションに関する基礎研究からシステムの設計，応用までを統合した研究となる．

本論文は，2章でロボットやエージェントによる会話介入システムについて先行研究をまとめる．3章では，優位性推定モデルについて述べる．4章，5章では適切な注視行動を行うための注視行動モデルと，適切な発話タイミングを図るための次発話者予測モデルについて詳述する．その後，それらをコミュニケーションロボットに統合した会話介入システムについて6章で述べた後，7章でいくつかの動作例を示す．システムの評価と考察を8章で述べ，最後に9章でまとめを行う．

2. 先行研究

参加者が3名以上である多人数会話（multi-party conversation）を行う会話ロボットの研究では，ロボットが円滑に会話に参加し，会話を制御することを目指した研究が行われてきた．Kuno ら [6] は，博物館案内ロボットにおいて，ロボットがターン交代のタイミングを知らせる注視行動を行うことにより，円滑なターン交代が実現することを示している．さらに，Mutlu ら [7] では，ロボットによる注視行動を使い分けることにより，各参加者の会話参与役割 [8] をユーザに知らせることができ，複数ユーザを含む会話における参与役割を適切に制御できることを示している．我々の先行研究 [9] では，優位性と，参与役割を考慮した多人数会話での注視行動を分析，モデル化しコミュニケーションロボットにモデルを実装した．分析の結果得られたモデルからは，優位性の高い参加者とそうでない参加者には注視行動の傾向に差異があることが見て取れた．また，評価実験の結果，優位性の高い参加者は，あまり注視対象を変化させず，発話者を見続けるような優位性が低い場合のロボットに好感を持つことが分かった．本研究においてもここで提案した注視行動モデルを採用する．その詳細は4章で述べる．

また，対話システムが搭載された全自動の会話ロボットの研究においても，このようなターン交代や参与役割を知らせる非言語シグナルを決定する機構を実装することにより，多人数会話を制御することに成功している [10], [11]．Bohus ら [12], [13] では，ディスプレイに顔アニメーションを表示した移動型ロボットを用い，複数人ユーザとの会話におけるターン制御のモデルを提案するとともに，それを対話システム中のターン制御機構として実装している．これらの対話システムでは，ロボットやアニメーションエージェントによる，ターン交代のシグナルの生成に加え，参与役割の認識機能も必要となる．そのために，音声や画像の情報を用いて，現発話の発話者，その受け手となる受話者を推定する方法が研究されている [14], [15]．さらに，次に起こる発話における参与役割を予測することにより，次のターンでのロボットの行動を適切に決定することを目指した研究も行われている [16], [17]．

しかし，以上に述べたコミュニケーションロボットの研究では，円滑なターン交代，対話制御を目的としており，ロボットが積極的に人同士の会話に介入し，影響を与えることを目標としたものではない．

一方，ユーザのコミュニケーション行動に影響を与えることを目的とした人工物の研究も数多く行われている．人の行動に影響を与える人工物を実現するには，人と人工物との間に信頼関係（rapport）を築くことが重要である [18]．Gratch らによるアニメーションのエージェントを用いた研究では，エージェントとユーザの間に信頼関係を築くため

© 2017 Information Processing Society of Japan

には，エージェントによる笑顔や頷き等のフィードバックが有効であることが示されている [1]．また，それによりインタラクションがより長く継続し，より多くの言葉が利用者から引き出されるという報告もある [19], [20]．この考えをさらに発展させ，PTSD によりうつ状態が疑われるユーザの話を聞くカウンセラとして振る舞うアニメーションの会話エージェントの研究も行われている [21]．しかし，これらの研究は，ユーザとエージェントとの 1 対 1 の対話を対象としており，複数人のユーザの会話に参入し，その参加者に影響を与えようとするものではない．

以上，人工物であるエージェントやロボットが人の会話行動に影響を与える力を持つことは，先行研究でも実証されているが，本研究では，この考え方を発展させ，複数人ユーザが参加する多人数会話を適切に制御しつつ，積極的に介入することにより，グループユーザの会話に影響を与える全自動の会話ロボットを実現することを目標とする．

3. 優位性推定モデル

多人数会話にロボットが円滑に参加するために，様々な研究で注視行動の制御が行われてきた [6], [7]．これらによって，参加者にターン交代のタイミングを知らせたり，参与役割を明確にしたりできるといった点が明らかになった．本研究は，こうした円滑な会話への参加に加えて，ロボットによる適切な会話への介入を目的に，会話における優位性推定を基にした会話状態の把握を行う．どの参加者が会話全体を牽引しているリーダ的役割であるのかといった会話における優位性 [2], [3] は注視行動へも影響を与え，より適切な非言語的振舞いを行うロボットの実現や，会話介入における介入対象者の決定にも有用である．本章では，優位性とその推定モデルについて述べ，次章で優位性を考慮した注視行動モデルについて詳述する．

3.1 会話における優位性

社会心理学の分野では，会話における優位性とその影響について広く研究されており，また文献 [3] のように優位性の自動推定を行う試みも行われている．たとえば，会話においてリーダ的役割を持つ参加者は優位性が高く，反対にあまり会話に参加できていないと優位性は低くなる．このような会話における各参加者の会話牽引度合いが推定できれば，優位性が高い会話の牽引者へ新たな情報の提供を行ったり，反対に優位性が低く，会話に参加できていない者に対してロボットが意見を求めたりするような行動をとることが可能となる．

我々の先行研究 [22] では，複数人会話における参加者の優位性（Conversational Dominance）を各参加者の視線特徴と発話特徴から推定するモデルを提案した．本研究では，ここで提案した優位性推定モデルをシステムに実装し各参加者の優位性をリアルタイムに計測したうえで，ロボットに適切な会話介入行動を行わせる．

3.2 優位性推定モデル

文献 [22] で提案した優位性推定モデルは，会話における優位性を表す量として情報提供または情報要求発言数を採用し，この値を各参加者の発話時の発話特徴と視線特徴を利用して推定する線形回帰モデルである．

具体的には，参加者 i の優位性 $D(i)$ は，i の発話時間の累積 $S_{\mathrm{t}}(i)$ sec，i が各発話中に他者を見ていた時間の累積 $G_{\mathrm{o}}(i)$ sec，各発話中に他者と相互注視していた時間の累積 $G_{\mathrm{m}}(i)$ sec，発話権取得回数 $T(i)$ 回から次式 (1) によって推定される．

$$D(i) = 0.94S_{\mathrm{t}}(i) + 0.80G_{\mathrm{o}}(i) + 0.162G_{\mathrm{m}}(i) + 0.256T(i) - 0.25. \quad (1)$$

ここで，発話権取得回数とは，前発話者の発話終了後 2 秒以上の沈黙が続いた後に発話権を取得，すなわち発話を行った回数を表す．本回帰式のあてはまりの良さを示す決定係数は，0.72 である．式 (1) から分かるように，参加者 i の優位性 $D(i)$ は，i が発話を行うたびに変化する．そのため，本回帰式は会話を通した最終的な各参加者の優位性だけでなく，発話ごとの優位性の変化量の推定式として扱える．この場合，各変数は各発話ごとの値となり，発話権取得回数項はその発話が発話権取得条件に合致していた場合にのみ，0.256 が加算される．また，最終項は回帰式の切片を意味するため，本研究での使用目的のように各参加者の優位性を比較する場合には考慮する必要がない．

以上の議論から，各参加者の優位性推定のためには，その視線特徴と発話特徴の継続的な取得が必須となる．

4. 注視行動モデル

前章では，多人数会話に円滑に参加し，会話介入を行うために会話における優位性を推定し，会話状態を把握することを述べた．我々の先行研究 [9] ではまた，人の注視行動が参加者の優位性に依存した遷移を行うことを会話データの収集と分析を通して明らかにし，その注視の遷移モデルを構築した．そこで，本システムではここで提案した優位性を考慮した注視行動モデルをロボットへ実装する．本章ではこの注視行動モデルについての概略を述べる．

4.1 注視行動モデルの構築

4.1.1 データ収集

会話における参加者の注視行動と，各参加者の優位性との関係を調べるため会話収集実験を行った．実験では男性 16 名，女性 4 名の計 20 名の大学生を 4 名ずつ 5 グループとし，各グループに 2 回会話を行わせた．各セッションの会話テーマは 2 種類から重複のないように無作為に与えた．収集したデータは各参加者の会話中の頭部方向データと，

© 2017 Information Processing Society of Japan

図 1 データ収集実験の様子
Fig. 1 Snapshot of data collection.

俯瞰カメラによるビデオ映像である．図 1 に示したように参加者 4 名を正方形のテーブルの各辺に着席させ，各参加者の頭部の回転角度を Microsoft 社の Kinect センサ[*1]によって計測した．このデータは 6 章で述べる顔向き推定モデルの構築に利用する．

4.1.2 特徴量抽出

注視行動の観察を行った結果，会話を牽引するリーダ的参加者は発話時に比較的頻繁に視線先の対象を変える一方で，そうではない参加者は発話中にも視線がある受話者に固定され，他の傍参与者に移り難いといった特徴を持つことが分かった．このことは，優位性の高い参加者は他者に広く気を配るが，優位性の低い参加者はそうした行動をとることが少ないことを示している．このように，会話における注視行動は発話者，受話者といった発話ごとに変化する参与役割のみでなく，会話中の優位性にも依存する．

参加者の優位性を考慮した注視行動モデルの作成のために，俯瞰カメラの映像に人手で注釈付けを行った．その内容は，各参加者ごとに発話区間，発話区間ごとの各参加者の参与役割（発話者，受話者，傍参与者 2 名），各参加者の注視対象者（自分以外の誰か）である．これらのデータから，前章で述べた優位性推定モデルより各参加者のセッションごとの最終的な優位性を求めた．優位性は発話が生じるたびに変化する量ではあるが，こうした多人数会話においてリーダ的役割で積極的に発話する優位性の高い者と，あまり会話に参加できない優位性の低い者が頻繁に入れ替わったり，拮抗したりすることは少ない．以下に述べるように注視行動モデルの構築においては参加者の優位性の高低のみに着目するため，各セッションの最終的な優位性のみを利用した．

4.1.3 モデル作成

注視行動モデルは，参加者を参与役割 3 種類（発話者，受話者，傍参与者）で分類し，各参与役割における注視行動を分析することで作成した．このとき，各参加者の優位性の高低（最終的な優位性が 1 位かそれ以外か）も考慮するため，たとえば，発話者モデルには，(a) 優位性の高い発話者が優位性の低い受話者，傍参与者へ注視行動を行う場合，(b) 優位性の低い発話者が，優位性の高い受話者，低い傍参与者へ注視行動を行う場合，(c) 優位性の低い発話者が，優位性の低い受話者と，高い傍参与者へ注視行動を行う場合の計 3 種類がある．

このように全参加者それぞれの各状態における注視の変化を，誰を，どれだけの時間見ていたかという点から分析し，構築した注視行動モデルを付録図 **A·1** に示した．発話者モデル 3 種類に加えて，受話者モデル 3 種類，傍参与者モデル 4 種類の計 10 種類のモデルとなる．図中の数値はデータ分析によって得られた視線遷移の確率と，その注視対象者を見ていた平均注視時間である．本研究では，この遷移確率と注視時間をロボットの行動モデルとして利用する．たとえば，ロボットが優位性の高い発話者であり，優位性の低い受話者を見ている場合には図 A·1 (c) を適用し，1.34 秒間受話者を見続けた後，79%の確率で続けてそのまま受話者を見続ける．または，21%の確率で傍参与者に視線が移り，それは 0.87 秒継続する．その後，65%で傍参与者を 0.67 秒間見続ける．または 35%の確率で受話者に視線が戻る．といった注視行動を行わせる．

本研究では，会話参加者の優位性と参与役割を推定したうえで以上で述べた注視行動モデルに従い自動的に注視行動を行う会話介入ロボットを実現する．そのためには前章で述べた優位性推定に加えて，各参加者の発言ごとの参与役割をリアルタイムに認識しなければならない．

5. 次発話者推定モデル

前章までで，会話への円滑な参加と適切な介入が可能なロボットの実現のため，優位性に基づいた介入対象者の選択や，注視行動の生成について議論した．ここでは，適切な介入のタイミングを考慮するための次発話者推定モデルについて述べる．

多人数会話において，適切なタイミングで会話介入を行うためにはシステムが自身が発話権を得たことを認識し，そのうえで介入を行うことが適当である．本研究では，非言語情報から簡易的に次発話者を推定可能なモデルを構築し，それをロボットに実現することで，適切なタイミングでの会話介入を行った．

5.1 次発話者推定

言語，非言語行動から会話における発話ターンのやりとりを予測することはすでに様々な手法で行われている [16], [23], [24]．文献 [23] では，ターンの移り変わり時の視線遷移パターンに着目し，4 人による会話において次発話者を 59.2%の精度で推定するモデルを作成した．また，文献 [24] では，発話前の息を吸い込む動作に着目し，呼吸の深さを特徴量とすることで，40.2%で 4 人会話での次発話者予測モデルを構築している．文献 [16] では，ポスター発表の場での発表者 1 人と聴衆 2 人という状況を題材に，

*1 http://www.xbox.com/en-US/xbox-360/accessories/kinect

© 2017 Information Processing Society of Japan

韻律情報や視線情報から 69.7%の精度で次発話者を予測できるモデルを作成した．

本研究では，より簡易的，かつ会話参加者に特殊な機器を装着させずに次発話者を推定するモデルを作成する．特に，ロボットを用いた実際のシステムとしての実装を考えると，言語，非言語行動を取得するためのセンサが増加することはシステムの複雑化を招き好ましくない．また，会話参加者がそのつど様々なセンサを身につけ会話を行うことも現実的ではない．そこで，これまで優位性推定や，注視行動モデルの作成に使用した発話情報や，顔向き情報からのモデル構築を試みた．これらのデータから次発話者予測が可能であれば，センサの増加による煩雑化を防ぎ，かつ利用者に特殊な機器を身につけさせることのないシステムが実現できる．

5.2　多人数会話データからの特徴量抽出

4 章で収集した会話データから，次発話者予測に有用と思われる特徴量をビデオの観察をもとに仮定し，以下の 3 つの特徴量を人手で注釈付けを行い，抽出した．

- 発話終了時の発話者が視線を向けている参加者
- 発話終了後 1 秒前から発話終了までの発話者の各参加者への注視時間
- 発話終了 1 秒前から発話終了までの発話者と各参加者が相互注視した時間

第 1 の特徴量は，発話者にとって，正面，右，左に位置する参加者のうち，いずれの参加者に視線を向けていたかである．第 2，第 3 の特徴量についても同様に，座席位置を規準とした解析を行った．たとえば，第 2 の特徴量は，発話者が正面の参加者を a 秒，右の参加者を b 秒，左の参加者を c 秒注視した．といったようになる．第 3 の特徴量は，a 秒，b 秒，c 秒のうち，相互注視を行っていた時間を表す．このように発話者に対する座席位置を規準に分析を行うことで，全参加者の全発話を画一的に扱うことができる．

5.3　次発話者推定モデル

上記によって得られた特徴量から次発話者を予測するモデルを C4.5 アルゴリズムによる分類器学習を行い作成した．Leave-one-group-out 法によるモデル評価では，精度 0.586，再現率 0.583，F-Measure 0.581 という結果が得られた．これは，現在の発話者を除いた 3 人から発話者を決定する場合のベースラインが 0.33 であることを考慮すれば比較的良い結果である．

本システムでは，これらの特徴量をリアルタイムに取得し，次発話者予測を行うシステムを構築し，適切なタイミングで会話介入可能なシステムを実現する．

6.　会話介入システム

本章では，ここまで述べてきた優位性推定モデル，注視行動推定モデル，次発話者推定モデルを用いたコミュニケーションロボットによる会話介入システムについて詳述する．システムの構成図を**図 2** に示した．以下では，図を参考に処理の流れと，各種推定部からの出力をもとにした会話介入部についての詳細を述べる．

6.1　システム環境

図右下に示したように，本システムは会話介入ロボット 1 台と，参加者 3 人の計 4 人での会話を想定している．システム環境の実際の様子を**図 3** に示した．今回は，4 人で

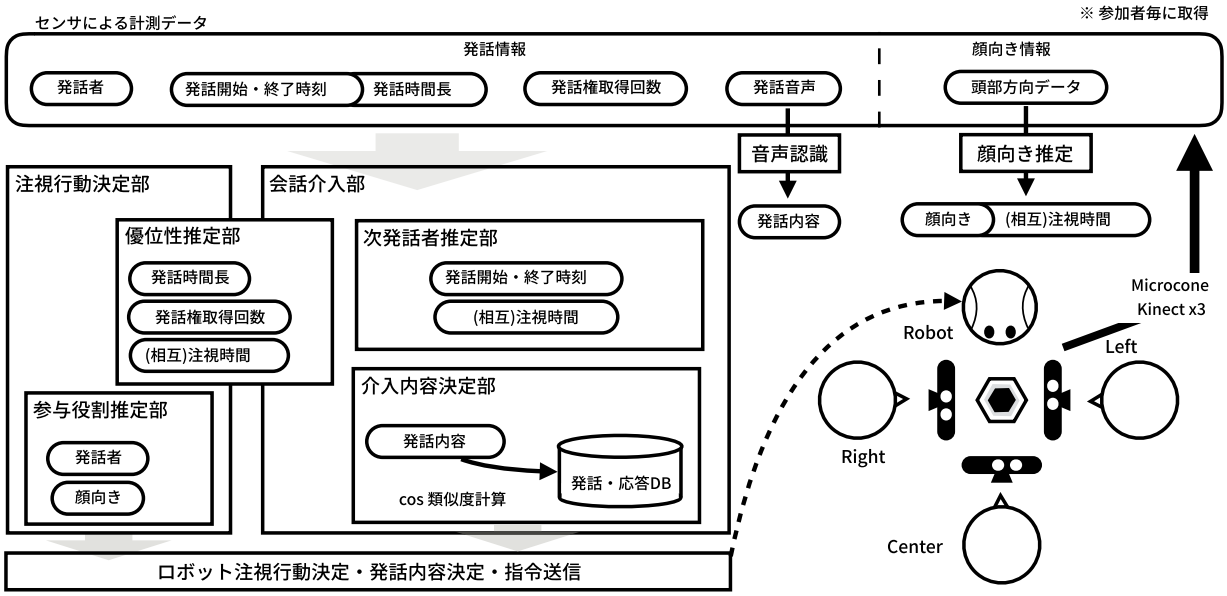


図 2　システム構成図
Fig. 2　System architecture.

© 2017 Information Processing Society of Japan

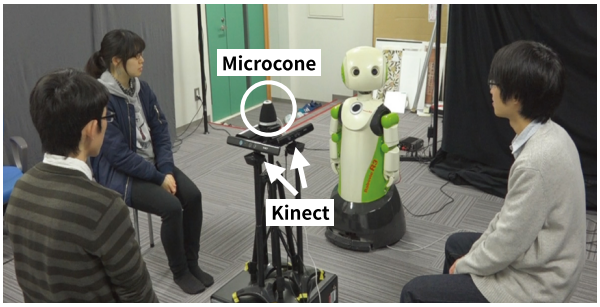


図 3　実装環境
Fig. 3　Overview of the implemeted system.

ある観光地の旅行計画を立てるという会話を想定したシステムの構築を行った．

参加者は中央に置かれたマイクロフォンアレイ（Dev-Audio 社製，Microcone [*2]）を囲んで 2 人ずつ向かい合わせに着席する．また，各参加者の正面には Microsoft Kincet センサを設置する．各 Kinect センサからのデータ収集用計算機と，ロボット動作制御用計算機，Microcone からのデータ収集用計算機に加えて，全体の統合や後述する各種推定部での演算等を行う計算機を用いる．各計算機は同一のローカルネットワーク上に置かれ，ネットワークを通じてデータの送受信を行う．

Microcone は 6 角形の本体の各辺に指向性マイクが内蔵され，それぞれ独立にマイクへの音声入力を検出できる．また，音声入力が行われると，無音区間が生じるまでの録音データを保存できる．本システムでは，ロボットを含めた 4 人の参加者それぞれにマイクを割り当て，個々の発話の検出と，発話区間の音声データの保存に利用した．

Kincet センサからは，FaceTracking API を用いて会話参加者頭部の水平方向の回転角度（Yaw 角）を取得する．

6.2　計測データ

前述した 3 つの推定モデルや，後述するそれにともなう参与役割の認識，注視方向推定，会話介入といった機能の実装に必要な会話参加者の状態を表す特徴量は，すべて上記した Microcone と Kinect センサから取得する．以下，それぞれのデータの取得方法を述べる．

6.2.1　発話者

現在の発話者は，参与役割の推定に用いる．これは，Microcone を利用し，ロボットを含む各参加者の正面に位置するマイクに音が入ったら，その者を発話者と定めた．複数人が同時に話す等で複数のマイクが同時に音を取得した場合には後に話した者が現在の発話者として認識される．これにより，ある時刻に同時に複数の発話者が存在することはない．

6.2.2　発話開始・終了時間，発話時間長

上記によって，マイクに音が入った時刻を発話開始時刻，そのマイクに入る音が途切れたときを発話終了時刻と見なした．この時刻の差をとることで，発話時間長を得る．

6.2.3　発話権取得回数

同様に，いずれのマイクにも音声入力がされていない時間が得られ，それが 2 秒以上になった後，初めて発話した場合に，その対象者の発話権取得回数がカウントされる．

6.2.4　発話音声

Microcone によって個々に保存された音声データは，介入内容の決定のために利用する．本システムでは，取得した音声データを Google Speech API にかけ，発話内容をテキストデータとして保持する．ただし，頷きや，咳払いのようなごく短い音声がマイクに入力された場合は発話とは認識しない．

6.2.5　頭部方向

Kinect センサから得られた Yaw 角をもとに各参加者の頭部方向を求め，これを注視方向であるとした．Yaw 角から注視方向（正面，右，左）の推定するために，4 章で収集した実験データから顔向き推定モデルを作成した．推定に使用する特徴量として，現在時刻 t と，$t-1$，$t-2$ の各時刻の Yaw 値を利用し，C4.5 アルゴリズムによる分類器学習を行った．Leave-one-person-out によるモデル評価の結果，モデルの精度は 0.808，再現率 0.806，F-Measure 0.806 という結果が得られ，十分な精度を有するモデルとなった．以降では，この推定された 3 種類の頭部方向を各参加者の注視方向として扱う．

6.2.6　注視時間・相互注視時間

各参加者が他者を見ていた時間は，上記注視方向から算出する．また，相互注視時間はある対象者 A が参加者 B を注視していたとき，参加者 B も対象者 A を見ていた場合に成り立つと定義し，その時間を各参加者間ごとに求めた．

6.3　コミュニケーションロボット

会話介入のためのコミュニケーションロボットとしてヴィストン社製の Robovie-R3 [*3]（図 3 中央）を使用した．Robovie-R3 には多数のセンサが搭載され，17 自由度の間接を持つが，本システムではロボットは顔向き動作を行うためだけに使用し，先に述べたように会話中の特徴量を測定にはすべて外部センサを利用する．

発話音声は Microsoft Speach API（SAPI）による合成音声を利用する．ロボットの内部にスピーカを設置し，ロボット制御用計算機から発話内容をスピーカに出力する．これによってロボット本体から合成音声が流れる．

6.4　システム構成

ここでは，上記のように取得した各特徴量を使用する各予測モデルを利用した推定機構と，会話介入機構の実装に

*2　http://www.dev-audio.com

*3　http://www.vstone.co.jp/products/robovie_r3/

© 2017 Information Processing Society of Japan

表 1　想定質問と応答対応リストの一部（観光対象地域が神戸の場合）
Table 1　Part of the question and answer database.

Q	ハーバーランドってなにがあるの？
A	神戸ハーバーランドは、神戸港を代表する観光スポットの一つで、多くの商業施設が建ち並んでいます。
A2	神戸にはどのようなイメージをお持ちですか？
Q	神戸祭りは？
A	神戸まつりは、神戸最大の祭典です、サンバなど賑やかなものが多いです。
A2	ところでお祭りは好きですか？
Q	どこがおすすめ？
A	私のおすすめはハーバーランドです
Q	食べ物はどこがおすすめですか？
A	神戸祭りです
Q	もう一回候補地をお願いします
A	候補地は、神戸まつり、イオンモール神戸北、メリケンパーク、神戸海洋博物館、神戸ハーバーランドです。

ついて述べる．

6.4.1　優位性推定部

優位性推定部では会話参加者 4 人のうちいずれかの発話が検出された場合に，発話時間長，発話権取得回数，注視時間，相互注視時間から各参加者の発話ごとの優位性変動量を算出する．参加者の優位性はこの累積となる．

6.4.2　注視行動推定部

注視行動推定部では，現在の参加者の優位性と，参与役割をもとに 10 種類の顔向きモデル（図 A·1，**図 A·2**，**図 A·3**）に従って注視行動を行う．参与役割は，現在の発話者を認識し，その発話開始時の注視対象者を受話者，それ以外の 2 人を傍参与者として扱う．発話終了後も，次の発話が行われ参与役割や優位性の変化が生じるまでは使用するモデルを変化させず，遷移確率に従った行動をとる．すなわち，現在の視線先の平均滞留時間経過後に乱数に従って視線先を遷移させるかどうかを判定し，遷移させる場合にはロボットに対して顔向きの変更を指示し，これを繰り返す．

6.4.3　次発話者推定部

次発話者の推定には，発話者と，発話の開始・終了時刻，注視時間を利用する．発話終了時に発話者が見ていた参加者と，発話者が発話終了 1 秒前から発話終了までに他者を注視していた時間，そのうち，相互注視していた時間を求め，推定モデルから次発話者を予測する．このとき，次発話者がロボットであった場合に，次に述べる会話介入機構から介入内容を決定し，発話を行う．

6.4.4　会話介入部

次発話者予測モデルによって，次発話者がロボットであると判定された場合，本システムは直前の発話内容をもとに会話介入を行う．このとき，ロボットに発話権を受け渡す直前の発話内容は，音声認識結果から得ることができる．本システムでは，このロボットに対して発話権を受け渡した発話をロボットに対する発話であると仮定し，その発話に対する応答として発話を行う．適切な応答を行うためには，こうして認識した介入タイミングに加えて，ロボットに対する発話内容を認識しなければならない．

ロボットによる応答内容は，事前にデータベースとして参加者の想定発話とそれに対する応答を保持しておき，発話権をロボットが取得するたびにデータベースから適切な内容を選択し，発話を行わせる．このためには，想定される参加者の発話内容の収集と，それに対応する応答リストの作成に加えて，ロボットに対して発せられた発話内容と最も近しい想定発話を求め，それに応じた応答を選択しなければならない．

応答データベースの構築　想定発話の収集と，それに対する応答集の作成のために，事前に Wizard of Oz 法による予備実験を行った．実験では，今回構築した，優位性や参与役割に応じた注視行動を行うロボットを利用した．ロボットに発話権が回ってきたと実験者が判断したとき，その直前の発話に対する応答を実験者が考え，それを音声合成を通してロボットから発話させた．この予備実験の参加者は大学生 27 人である．収集，作成した応答データベースの一部を**表 1** に示した．表の Q が質問を意味し，A，A2 はその質問に対してロボットが発話する内容である．A と A2 は続けて発話されるが，発話対象者が異なる．A は直前の発話者，つまり，ロボットに発話権を渡した者に対してその質問への応答として発せられる．一方で A2 はその時点で参加者 3 人のうち優位性の低い参加者に対して（その参加者を注視して）発話される．このとき，優位性が低いとは，その時点での優位性が 1 位でない参加者を意味する．

質問リストとの類似度計算　作成された応答データベースから，参加者の発話に最も近しい想定発話を選択し，それに対する応答を行うために，参加者の発話と，想定発話の類似度計算を行う．具体的には，応答データベースの全文章を形態素解析器 Mecab [*4]にかけ，出現する全名詞を抽出する．各想定発話はこの名詞リストに含まれている名詞の出現頻度によってベクトル化を行う．同様に参加者によ

*4　http://taku910.github.io/mecab/

© 2017 Information Processing Society of Japan

る発話の音声認識結果を形態素解析し，応答データベースから作成した名詞リストを使ってベクトル化する．これによって，全想定発話と，参加者の発話の cos 類似度を計算し，最も類似した想定発話に対する応答を行うことができる．ただし，発話権を譲渡する参加者の発話がとても短く，cos 類似度がすべての質問で 0 になる場合がある．この場合には候補地の名称を乱数によって取得し「(候補地名) についてはどう思いますか」という発話を行う．

以上のように構築した提案システムは，リアルタイムに参加者の優位性，参与役割を認識し，自らに発話権が譲渡された際にその直前の発話内容に応じた会話介入を行う．

6.4.5 システムの制限

提案システムが持つ制限はモデルや会話内容のデータをあらかじめ用意しなければならない点である．ここまでで述べてきたとおり，本システムが利用する注視行動モデルや次発話者推定モデルは 4 人による会話データから作成しており，会話参加人数の拡張や話題の変更を行うためには場合に応じたモデルの作成や介入内容の決定が必要となる．しかしながら，任意の環境において適切な介入を行うために会話参加人数やその発話状態，顔向き状態を把握したり，適切な介入内容をリアルタイムで生成したりすることは難しい．そのため本研究では参加人数や会話内容をあらかじめ決定した環境下で，適切な介入が可能なロボットの実現を目指した．

7. 動作例

前章までで実装を行った提案システムによる動作例を示す．ここでは，次章で述べる本システムの評価実験から特徴的な場面の抽出を行った．

7.1 情報提供の様子

以下の 2 つの例は，ロボットによる質問応答と，情報提供の場面である．前者はロボットに対してお勧めの観光スポットを尋ねた結果，どういった場所なのかを説明することなくスポット名を発話した．このロボットの提案に対して参加者が賛同する様子が見て取れる．また，後者はグランフロント大阪という施設について情報を求めたときの様子である．参加者 A の質問に対してシステムが適切な解答を行った結果，参加者間でグランフロント大阪についての知識共有が行われているのが分かる．

質問応答

参加者 A：	おすすめはどこか教えて下さい
ロボット：	私のおすすめはユニバーサルスタジオジャパンです
参加者 B：	おお，楽しそう
一同：	うん
参加者 B：	遊園地はいいよね
参加者 A：	一回行ってみたい

情報提供

参加者 A：	グランフロント大阪は？
ロボット：	ショッピングモールやレストラン・カフェ、オフィス、ホテル、コンベンション・センター、劇場、超高層マンションから構成されています。
参加者 B：	あー
参加者 A：	それがグランフロント大阪？
参加者 B：	なるほど
参加者 A：	すごい．じゃああべのハルカス的な感じ？

7.2 会話展開の様子

以下は，参加者からの質問に対して，応答することに加えて，優位性の低い参加者に対して話題を発展させる発話を行う例である．ロボットによる応答発話から会話が広がる様子が見て取れる．この例では，参加者 A の質問に応答し，メリケンパークの説明をした後，顔向きを優位性の低い参加者 B へと変更し，発展発話を行った．その結果，参加者 B の発話を引き出すことができ，また加えて，公園という話題について会話が展開している．

会話展開

参加者 A：	メリケンパーク教えて
ロボット：	メリケンパークは神戸ポートタワーが立つ中突堤とメリケン波止場の間を埋め立てて作った海に面する美しい公園です
ロボット：	公園とかには普段行きますか
参加者 B：	公園は..
参加者 A：	公園は久しく行ってないですね
参加者 C：	うん
参加者 B：	近所にあるけど.. みたいな

次章では，こうした介入行動による会話への影響を定量的に評価し，本システム有用性を評価する．

8. 評価実験

以上のように実装を行ったシステムによる会話介入が，どのように多人数会話に影響を与えるかを調査するため評価実験を行った．

8.1 実験目的

本実験の目的は，第 1 に，次発話者予測に基づいて自身の発話権を認識したうえで，直前の発話に対する応答を返す介入が，そうした要因を考慮しない介入に対して会話により影響を与えられるかどうかを評価する点にある．また，第 2 に，会話における優位性を考慮した介入を行うことで，優位性の低い参加者の発話を促すことが可能かどうかもあわせて評価する．

8.2 実験条件

実験被験者として男性 18 人，女性 6 人の計 24 人の大学

© 2017 Information Processing Society of Japan

表 2 各観光対象地域とその候補地.
Table 2 Sight-seeing locations and each tourist spots.

大阪	北海道	神戸
グランフロント大阪	札幌	神戸まつり
大阪城	旭山動物園	イオンモール神戸北
通天閣	函館	メリケンパーク
道頓堀	小樽	神戸海洋博物館
あべのハルカス		神戸ハーバーランド
ユニバーサルスタジオジャパン		

生に協力を依頼し，1 グループ 3 人で 8 グループを構成した．各グループに提案システムによって自動制御された介入ロボットを加えた計 4 人で特定の地域（大阪，北海道，神戸）での観光計画を立てるタスクを行わせた．具体的には，**表 2** に示したように各地域について 4 から 6 つの観光地をあらかじめ用意し，ロボットを含めた 4 人で旅行に行く前提で候補地から訪れたい場所 2 カ所を話し合いで選ばせた．このタスクを各グループ 2 セッション，観光の対象地域を変更して行った．観光対象地域は，同一グループで重複しないことを条件に無作為に選択した．

応答データベースは，各観光対象地域に応じたものを作成し，セッションごとに適切なものに切り替えて使用した．それぞれの想定発話・応答数は大阪 54，北海道 54，神戸 62 個である．

ロボットは，つねに前述した注視行動モデルに従って注視行動をとるとともに，同一のセッション内で以下 2 種類の会話介入を行う．

(1) 次発話者がロボットである場合に，直前の発話内容をもとにした発話を行う発話権考慮介入
(2) 不定期に，決められた発話内容を，無作為に決めた対象に発話する無作為介入

発話権考慮介入は，6 章で実装した会話介入機構に従い，ロボットに発話権を渡した直前の発話者の発話内容に応じて表 1 に示したような返答を返す．先に述べたように表中 A は直前の発話者を注視し発せられるが，A2 は会話参加者のうち優位性の低い参加者に対して発せられる．A2 に設定された発話内容は，「ところでお祭りは好きですか？」や，「神戸にはどのようなイメージをお持ちですか？」といった，受話者に対して発言を求め，会話を発展させる内容となっている．このような A2 文を持つ応答は全体で 37 個存在する．

一方，無作為介入では，不定期なタイミングで，参加者 3 人から介入対象を無作為に決めたうえで，「どこか興味のある所はありましたか？」，「一番興味ある場所はどこですか？」といったあらかじめ決められた 16 種類の発話リストから無作為に 1 つを選び対象に向かって発話を行う．

無作為介入は応答が単純であり，加えて介入のタイミングが不適切な場合が多いと考えられる．そのため，比較対象としてこの介入のみを行うロボットと会話を行うと参加者のロボットへの印象が悪くなり，介入に対する反応が著しく悪化する恐れがある．そこで上記の 2 種類の介入を混在させ，ロボットへの印象に偏りが生じることを防いだ．

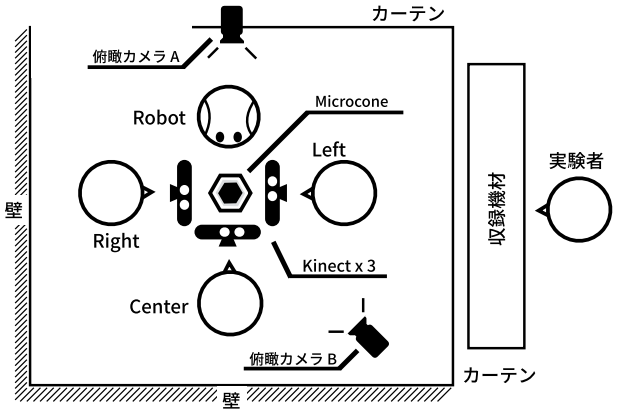


図 4 評価実験環境
Fig. 4 Experimental setting.

各セッションの制限時間は 10 分としたが，早期に行き先が決定した場合はその時点で終了することを認めた．

8.3 実験環境と収集データ

実験環境を**図 4** に示した．参加者周囲は壁とカーテンで囲われ，実験中に実験者の様子をうかがうことはできない．壁面には暗幕や毛布をかけて音の反射を防ぎ，Microcone による発話者の誤検出を防止した．各参加者正面に Kinect センサと，その中央に Microcone を設置した．Microcone から 80 cm 程度離れた位置に椅子を設置し，参加者を座らせた．これは Kinect センサによる顔検出を適切に動作させるためである．したがって，向かいの参加者またはロボットまでは 1.6 m 程度の距離があるが，特にこの距離によって互いの会話に支障がある様子は観察されなかった．

また，参加者の視線を固定することを防ぐため，実験環境にタスクに関する情報の掲示はいっさい行わず，資料の配布も行わなかった．参加者には各セッションの開始前に観光地域とその候補地を口頭で伝え，不明な点や候補地の詳細はロボットに随時自由に尋ねるように指示した．

収集したデータは，システムが利用するための Kinect センサ，Microcone からのデータに加え，図 4 に示すように設置した 2 台の俯瞰カメラによる映像データである．

8.4 実験結果とその分析

収集したデータから，発話権考慮介入と無作為介入それぞれの会話への影響の大きさを調べるため以下のように収集データの分析を行った．なお，8 グループ，各 2 セッションのデータのうち，音声認識の精度が著しく悪くロボットが正常に動作を行っていなかった 2 グループ，計 4 セッションと，優位性推定機構が正常に作動せず，ある参加者

© 2017 Information Processing Society of Japan

表 3 ロボットによる全介入の分類
Table 3 Classification of interventions.

単位：回

介入種類	有効数	連続	合計
発話権考慮	67	11	78
無作為	140	23	163
合計	207	34	241

表 4 無作為介入数と有効介入影響度との相関係数
Table 4 Correlation Coefficient between No. of Random Interventions and Avg of Influence Point of Valid Interventions.

	無作為介入回数	発話権考慮介入
無作為介入回数	1	–
発話権考慮介入	−0.090	1

の優位性が異常値を示した 1 セッションのデータを除き，6 グループ 11 セッションを分析対象とした．

8.4.1 有効な介入発話の選定

表 3 に，実験で観測されたロボットによる介入種類の分類を示した．11 セッションを通して，今回の比較対象である介入は 241 回生じた．このうち，発話権考慮介入は 78 回，無作為介入は 163 回だった．ただし，ある介入発話後，きわめて短い間隔で続けて介入が生じる場合が存在した．後に述べる介入の会話への影響力を示す介入影響度は，各介入によって誘発された発話から求める量であるため，参加者が応答する暇なく連続して介入が行われると算出できない．そこで，発話間隔が 0.5 秒以下でロボットが連続して発話を行った場合には，その最後の発話以外を分析から除外した．これによる除外は表中の「連続」列に示したように，発話権考慮介入で 11 回，無作為介入で 23 回あった．こうして，以降の分析対象となる有効な介入は，発話権考慮介入 67，無作為介入 140 の計 207 介入となった．

8.4.2 介入影響度の定義と付与

上記のように選定した有効介入会話すべてについて，それがその後の会話にどの程度影響を与えたかを定量的に評価するため，介入影響度を次のように定義した．すなわち，介入影響度は，ロボットによる発話がその後の参加者の発話を誘発した場合の影響の強さを表す量とし，介入への応答や，介入をきっかけとした参加者間の会話の中で内容語（名詞，動詞，形容詞）が含まれる発話を行った人数（1 から 3）を各介入の介入影響度とする．具体的には，影響度が加算される参加者の発話内容は，主に以下のようなものが挙げられる

- 介入への応答とその応答への他の参加者の反応
- 介入内容をきっかけとして参加者間で行われた会話
- 介入内容の言い換えやオウム返し

たとえば，「（候補地）には興味がありますか？」という介入に対する「あります」という応答や，それに続く他の参加者からの「楽しそうだよね」というロボットではなく，参加者への発話，また，「興味か」のようなロボットによる介入の一部を言い換えたり，オウム返しを誰にともなく発した場合にも，介入によって促された発話であると判断し，影響度を加算した．これらの条件に該当する発話を介入影響度対象発話と呼ぶ．

また，介入後の参加者の発話でもその介入の影響と判定しない例は以下のような場合である．

- 内容語ではない発話
- 介入前の会話を継続している場合
- 介入内容とは異なる話題で会話が始まった場合

「うん」や「あー」のような相槌は，それがロボットへの反応であるか，ロボットの介入をきっかけとした参加者間での会話でのものかを考慮せず加算を行っていない．また，明らかに介入前から継続した発話や，「（候補地）には興味がありますか？」という介入の後の「そろそろ決めましょうか」のように介入内容を無視した会話が展開した場合も影響度の加算は行わなかった．

参加者間での会話内容が介入をきっかけとしたものから移り変わったり，新たにロボットによる介入が生じたりした場合は，その時点で対象としていた介入に対する影響度の加算を終了した．

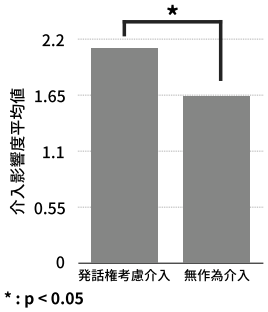
たとえば 7 章で示した動作例での質問応答の様子ではロボットの「わたしのおすすめはユニバーサルスタジオジャパンです」という介入に対して，参加者 B と参加者 A が内容語が含まれる発話しているため 2 点が付与される．参加者 C も同意を意味する「うん」という発話を行うが，これは内容語ではないためカウントを行わない．同様に，情報提供でのロボットの介入発話では，参加者 A のみが内容語を含む発話を行っているため 1 点となる．このように，直接的にロボットの介入を受けて応答を返すことに限らず，ロボットの介入内容に沿った会話が続いた場合，影響度を加算する．

こうして定義した介入影響度を分析対象となる全有効介入発話について人手で付与した．

8.4.3 無作為介入による介入影響度への影響の分析

本実験は，ロボットは発話権考慮介入と無作為介入の両方を 1 つのセッション中に行う．無作為介入が介入のタイミングや発話内容を考慮しないことを考えると，8.2 節でも触れたように参加者のロボットへの印象が悪くなり，発話権考慮介入への応答も悪化していく場合が想定される．そこで，各セッションごとの無作為介入の回数と，発話権考慮介入の介入影響度平均値の相関分析を行った．その結果を**表 4** に示した．表から分かるように無作為介入の回数と発話権考慮介入の影響度平均値には相関が見られなかったことから，今回の実験では無作為介入によって発話権考慮介入への応答が悪化することはないといえる．

© 2017 Information Processing Society of Japan



発展発話(A2)の有無	優位性最低の参加者の発話割合
有	92.8%
無	71.0%

図 5 介入手法の違いによる介入影響度の平均値の差（左）と，発話権考慮介入で A2 発話の有無による優位性の低い参加者の発話割合（右）

Fig. 5 Left: Avg Score of the Intervention Influence Point. Right: Utterance Ratio of Less Dominace Prticipant in Presence/Absence of A2 Utterance.

表 5 発話権考慮介入の分類

Table 5 Distribution of Interventions considering Speaking Turn.

単位：回

A2 発話	顔向き（低）	顔向き（低以外）	合計
有	14	22	36
無	–	–	31

8.4.4 介入影響度の検定

介入影響度の平均値をセッションごと，介入行動ごとに求め，等分散を仮定しない t 検定を行った．結果を**図 5** 左に示した．2 つの介入手法による発話の介入影響度の平均値は発話権考慮介入が 2.13，無作為介入が 1.65 となり，その差には $p < .05$（$t = 2.60$）で有意差が認められた．このことから，介入対象と介入タイミングを考慮しない無作為な介入に比べて，発話権を認識し，適切な相手に対して会話介入を行う方が，会話に対して影響を与え，発話を促したり，会話の話題提供が可能であることが示された．

8.4.5 優位性を考慮した介入による影響

最後に，前述した会話の発展を促す発話（A2）を優位性の低い参加者に行うことで，発話対象である優位性の低い参加者が発話の機会を得られたかどうかを次のように分析した．**表 5** に，有効な発話権考慮介入の内訳を示した．すなわち，全 67 回の介入のうち，36 回が A2 発話のある介入であった．このうち，優位性が最も低い参加者を注視して発話を行ったのは 14 回だった．この 14 回の介入のうち，注視された優位性が最低の参加者が介入影響度対象発話を行った割合と，A2 発話がなく，発話権を譲渡した参加者に対して応答を行った 31 回の介入のうち，同じく優位性最低の参加者が介入影響度対象発話を行った割合を求めた．その結果が図 5 右の表である．A2 ありの場合には，92.8%の確率で優位性の最も低い参加者が発話を行うのに対して，そうでない場合には 71.0%であった．この比率の差に有意差があるかどうか検定を行った結果 $p < .10$ で有意傾向が認められた．このことから，本システムは優位性の低い会話参加者へ話題を振ることで，その参加者の発言の機会を増やすことが可能であることが示唆された．

以上より，本研究が提案する，優位性，参与役割に基づいた注視行動をとりながら，発話権を取得した場合に会話介入を行うロボットは，そうでない介入システムに比べて話題を提供し，それに応じた会話や応答を誘発することが可能であり，また優位性に基づいた介入を行えば会話にあまり参加できていないような参加者に発話を促す効果がある可能性が認められた．

8.5 議論

本研究では，参加者の優位性と参与役割に基づいた注視行動をとりながら，自身の発話権を認識し，会話参加者の発話に応じた介入を行うロボットの実装と評価を行った．本システムが持つ注視行動モデルや発話権認識モデル，応答データベースからの応答内容の決定等の機能はここまで見てきたように，多人数会話に限らずロボットが人間と対話によるコミュニケーションをとるうえで欠かせない要素である．本章では，これらの機能を実装した介入ロボットがそうでないシステムに比べて有意に大きく会話参加者の発話を誘発することが可能であることを示した．述べてきたとおり，提案システムの振舞いはこれらの機能が統合され組み合わされた結果として表出するものであるため，各機能それぞれの貢献度合いを測定することは難しい．しかしながら，1 つ 1 つの機能，あるいはその組合せによるロボットの行動を生成可能なシステムを構築し，評価実験を繰り返すことで多人数会話に影響をあたえるのに重要な要因を選定することができるだろう．

加えて，今回の実験では介入に対する参加者の応答内容を考慮していない．たとえば発話権考慮介入に対するポジティブな反応も無作為介入に対するネガティブな反応もどちらも介入によって誘発された発話として同様に扱い，ここで定義した介入影響度はその介入が会話に対して良い影響を与えたのかどうかまでを判断できない．しかし，多人数会話への適切な介入によって会話のコントロールを目指すことを考えると，発話誘発の要因だけでなく，介入によってどういった効果が得られるのか，参加者のロボットへの印象がどのように変化するのかという点についても認識しそれに応じて行動を変化させるシステムとしていく必要がある．

9. おわりに

本研究では，多人数会話に円滑に参加し，会話介入を行う全自動会話ロボットの実現を目的に，人同士の多人数会話におけるコミュニケーションを，優位性，参与役割，注視行動，発話権交代といった視点から分析，モデル化し，それを実装した会話介入システムを提案した．

© 2017 Information Processing Society of Japan

提案システムは，上記にあげたような時々刻々と変化する会話参加者の状態を全自動で認識し，適切に介入タイミング，介入対象，介入内容を選択し行動することが可能なシステムである．

本システムによる発話権を考慮した会話介入手法は，無作為に介入するよりも有意に大きく発話のきっかけや話題の提供が可能であることが分かった．また，優位性を考慮し，会話の発展を促すような発話を行うことで，優位性の低い対象者の発話機会が増えることが示された．

今後は，さらに詳細な評価実験を行うことでシステムがどの程度会話を制御可能かどうかを調査する．たとえば，今回の実験では会話を発展させるような発話を優位性の低い参加者に対して行っていたが，優位性の高い参加者に対して会話の収束を促したり，意見集約を求めることで実際に多人数での意思決定の速さや，会話の質に影響を与えられるかどうかを調査していく．加えて，会話参加者による主観的評価を行い，介入ロボットの存在が会話やその成果である意思決定にどのように影響したと感じたかを評価する．

将来的に，提案システムが会話に影響を与え，その制御を十分に行うことが可能となれば，本システムは多人数会話に参加できるだけでなく，議論のファシリテーションシステムとして応用できる．人間らしい振舞いで会話に介入するロボットは議論に溶け込みながら，情報提供だけでなく議論の流れをコントロールし，質の高い議論へと参加者を導くような役割を担うことになるだろう．

参考文献

[1] Gratch, J., Okhmatovskaia, A., Lamothe, F., Marsella, S., Morales, M., van der Werf, R.J. and Morency, L.-P.: Virtual rapport, *Intelligent Virtual Agents*, Vol.4133, pp.14–27, Springer Berlin Heidelberg (2006).

[2] Hung, H., Jayagopi, D.B., Ba, S., Odobez, J.-M. and Gatica-Perez, D.: Investigating automatic dominance estimation in groups from visual attention and speaking activity, *Proc. 10th International Conference on Multimodal Interfaces – IMCI '08*, New York, New York, USA, p.233, ACM Press (online), DOI: 10.1145/1452392.1452441 (2008).

[3] Jayagopi, D.B., Hung, H., Yeo, C. and Gatica-Perez, D.: Modeling Dominance in Group Conversations Using Nonverbal Activity Cues, *IEEE Trans. Audio, Speech, and Language Processing*, Vol.17, No.3, pp.501–513 (online), DOI: 10.1109/TASL.2008.2008238 (2009).

[4] Goffman, E.: *Forms of Talk*, Conduct and Communication, University of Pennsylvania Press, Incorporated (1981).

[5] Clark, H.H.: *Using Language*, Cambridge University Press (1996).

[6] Kuno, Y., Sadazuka, K., Kawashima, M., Yamazaki, K., Yamazaki, A. and Kuzuoka, H.: Museum guide robot based on sociological interaction analysis, *Proc. SIGCHI Conference on Human Factors in Computing Systems – CHI '07*, New York, New York, USA, p.1191, ACM Press (online), DOI: 10.1145/1240624.1240804 (2007).

[7] Mutlu, B., Shiwa, T., Kanda, T., Ishiguro, H. and Hagita, N.: Footing in human-robot conversations, *Proc. 4th ACM/IEEE International Conference on Human Robot Interaction – HRI '09*, New York, New York, USA, p.61, ACM Press (online), DOI: 10.1145/1514095.1514109 (2009).

[8] Birdwhistell, R.L.: *Kinesics and Context: Essays on Body Motion Communication*, Conduct and Communication, University of Pennsylvania Press, Incorporated (2010).

[9] Nakano, Y.I., Yoshino, T., Yatsushiro, M. and Takase, Y.: Generating Robot Gaze on the Basis of Participation Roles and Dominance Estimation in Multiparty Interaction, *ACM Trans. Interactive Intelligent Systems*, Vol.5, No.4, pp.1–23 (online), DOI: 10.1145/2743028 (2015).

[10] 松坂要佐, 東條剛史, 小林哲則：グループ会話に参与する対話ロボットの構築, 電子情報通信学会論文誌．D-II, 情報・システム, II-パターン処理, Vol.84, No.6, pp.898–908〈オンライン〉, 入手先〈http://ci.nii.ac.jp/naid/110003184138/〉(2001).

[11] Moubayed, S.A., Edlund, J. and Beskow, J.: Taming Mona Lisa, *ACM Trans. Interactive Intelligent Systems*, Vol.1, No.2, pp.1–25 (online), DOI: 10.1145/2070719.2070724 (2012).

[12] Bohus, D. and Horvitz, E.: Facilitating multiparty dialog with gaze, gesture, and speech, *International Conference on Multimodal Interfaces and the Workshop on Machine Learning for Multimodal Interaction on – ICMI-MLMI '10*, New York, New York, USA, p.1, ACM Press (online), DOI: 10.1145/1891903.1891910 (2010).

[13] Bohus, D. and Horvitz, E.: Facilitating multiparty dialog with gaze, gesture, and speech, *International Conference on Multimodal Interfaces and the Workshop on Machine Learning for Multimodal Interaction on – ICMI-MLMI '10*, New York, New York, USA, p.1, ACM Press (online), DOI: 10.1145/1891903.1891910 (2010).

[14] Katzenmaier, M., Stiefelhagen, R. and Schultz, T.: Identifying the addressee in human-human-robot interactions based on head pose and speech, *Proc. 6th International Conference on Multimodal Interfaces – ICMI '04*, New York, New York, USA, p.144, ACM Press (online), DOI: 10.1145/1027933.1027959 (2004).

[15] Baba, N., Huang, H.-H. and Nakano, Y.: Identifying the Addressee using Head Orientation and Speech Information in Multiparty Human-Agent Conversations, *Trans. Japanese Society for Artificial Intelligence*, Vol.28, No.2, pp.149–159 (online), DOI: 10.1527/tjsai.28.149 (2013).

[16] Kawahara, T., Iwatate, T.K.T.: Prediction of Turn-Taking by Combining Prosodic and Eye-Gaze Information in Poster Conversations, *INTERSPEECH* (2012).

[17] Ishii, R., Otsuka, K., Kumano, S. and Yamato, J.: Prediction of Who Will Be the Next Speaker and When Using Gaze Behavior in Multiparty Meetings, *ACM Trans. Interactive Intelligent Systems*, Vol.6, No.1, pp.1–31 (online), DOI: 10.1145/2757284 (2016).

[18] Bickmore, T.W. and Picard, R.W.: Towards caring machines, *Extended Abstracts of the 2004 Conference on Human Factors and Computing Systems – CHI '04*, New York, New York, USA, p.1489, ACM Press (online), DOI: 10.1145/985921.986097 (2004).

[19] Gratch, J., Wang, N., Gerten, J., Fast, E. and Duffy, R.: *Creating Rapport with Virtual Agents*, pp.125–138, Springer Berlin Heidelberg (2007).

[20] Wang, N. and Gratch, J.: Don't Just Stare at Me!, *Proc. SIGCHI Conference on Human Factors in Computing*

© 2017 Information Processing Society of Japan

Systems, CHI '10, New York, NY, USA, pp.1241–1250, ACM (online), DOI: 10.1145/1753326.1753513 (2010).

[21] DeVault, D., Artstein, R., Benn, G., Dey, T., Fast, E., Gainer, A., Georgila, K., Gratch, J., Hartholt, A., Lhommet, M., Lucas, G., Marsella, S., Morbini, F., Nazarian, A., Scherer, S., Stratou, G., Suri, A., Traum, D., Wood, R., Xu, Y., Rizzo, A. and Morency, L.-P.: SimSensei Kiosk : A Virtual Human Interviewer for Healthcare Decision Support, *International Conference on Autonomous Agents and Multi-Agent Systems*, No.1, pp.1061–1068 (2014).

[22] Nakano, Y. and Fukuhara, Y.: Estimating conversational dominance in multiparty interaction, *Proc. 14th ACM International Conference on Multimodal Interaction – ICMI '12*, New York, New York, USA, p.77, ACM Press (online), DOI: 10.1145/2388676.2388699 (2012).

[23] Ishii, R., Otsuka, K., Kumano, S., Matsuda, M. and Yamato, J.: Predicting next speaker and timing from gaze transition patterns in multi-party meetings, *Proc. 15th ACM on International Conference on Multimodal Interaction – ICMI '13*, New York, New York, USA, pp.79–86, ACM Press (online), DOI: 10.1145/2522848.2522856 (2013).

[24] Ishii, R., Otsuka, K., Kumano, S. and Yamato, J.: Analysis of Respiration for Prediction of "Who Will Be Next Speaker and When?" in Multi-Party Meetings, *Proc. 16th International Conference on Multimodal Interaction – ICMI '14*, New York, USA, pp.18–25, ACM Press (online), DOI: 10.1145/2663204.2663271 (2014).

付　　録

A.1 注視行動モデル

4 章で述べた注視行動モデル [9] は，対象の参与役割（発話者，受話者，傍参与者）によって，それぞれ複数のモデルが構築される．以下に参与役割に応じた全モデルを示す．本注視行動モデルを適用する対象（ロボット）が発話者である場合は，発話者モデルを利用する．以下受話者，傍参与者においても同様である．

A.1.1 発話者モデル

図 A·1 は発話者を対象とした注視行動モデルである．対

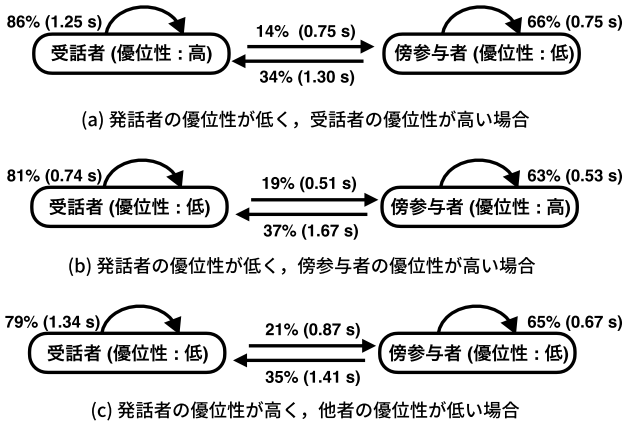


図 **A·1** 発話者の注視行動モデル

Fig. A·1 Speaker gaze behavior models.

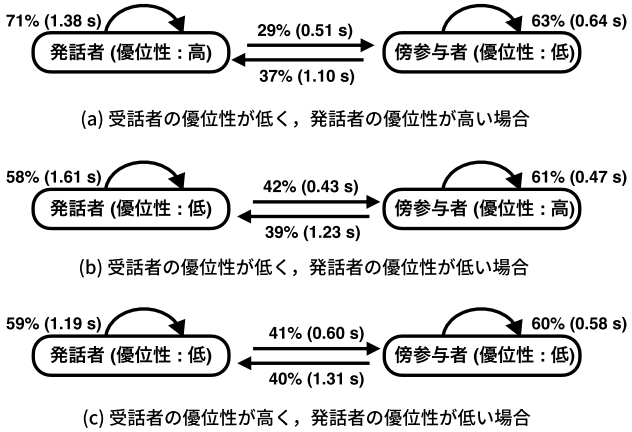


図 **A·2** 受話者の注視行動モデル

Fig. A·2 Addressee gaze behavior models.

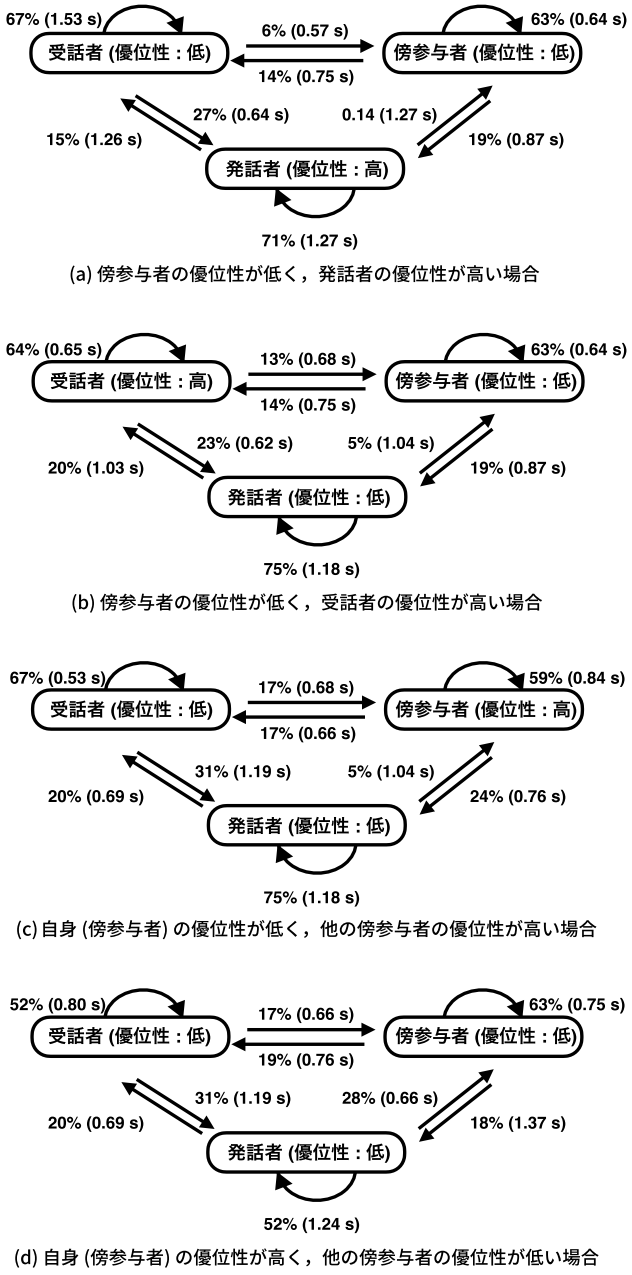


図 **A·3** 傍参与者の注視行動モデル

Fig. A·3 Third participant gaze behavior models.

© 2017 Information Processing Society of Japan

象である発話者か，そうでない受話者，傍参与者それぞれの優位性が高い場合の 3 種類のモデルとなる．

A.1.2　受話者モデル

図 A·2 は受話者を対象とした注視行動モデルである．発話者対象のモデルと同様に 3 種類のモデルとなる．

A.1.3　傍参与者モデル

図 A·3 に示した傍参与者モデルは，受話者，発話者それぞれの優位性が高い場合に加えて，対象である傍参与者（自身）の優位性が高い場合と，他の傍参与者が高い場合の計 4 種類のモデルとなる．図中に示した “傍参与者” とは，対象である自身ではない，もう一方の参加者を意味する．

高瀬 裕

2006 年東京都立大学理学部物理学科卒業，2008 年総合研究大学院大学物理科学研究科天文科学専攻修了，2014 年東京工業大学知能システム科学専攻博士修了．同年成蹊大学理工学部情報科学科助教，現在に至る．ヒューマンイータフェースの研究に従事．博士（工学）．

吉野 尭

2014 年成蹊大学理工学部情報科学科卒業．2016 年同大学大学院修士課程修了．同年ニッセイ情報テクノロジー株式会社入社，現在に至る．

中野 有紀子　（正会員）

1990 年東京大学大学院教育学研究科修士課程修了．同年日本電信電話（株）入社．2002 年 MIT Media Arts & Sciences 修士課程修了．同年より JST 社会技術研究開発センター専門研究員，東京農工大学大学院工学府特任准教授，成蹊大学理工学部情報科学科准教授を経て，現在，成蹊大学理工学部情報科学科教授．知的で自然なユーザインタフェースの実現に向けて，人との言語・非言語コミュニケーションが可能な会話エージェントの研究に従事．博士（情報理工学）．ACM，人工知能学会，電子情報通信学会各会員．

© 2017 Information Processing Society of Japan