

# 会話エージェントによる多人数会話の実現を目指して —複数人ユーザの動作データの収集と分析—

Towards the Multiparty Communication Capability in Conversational Agents:  
Collecting and Analyzing Human Behavioral Data in Wizard-of-Oz Multiparty Conversations

中野 有紀子\*<sup>1</sup>  
Yukiko Nakano

\*<sup>1</sup> 成蹊大学 理工学部情報科学科  
Faculty of Science and Technology, Seikei University

The goal of our research is to build conversational agents that can communicate with multiple users, and serve as receptionists or information clerks. To build such agents, the system not only needs to be able to keep track of the conversational situation which dynamically changes, but also it should be able to decide the proper timing to make the agent join the human conversations. To discuss the design basis of the agents with multiuser capability, this paper reports corpus collection experiments in two settings: two users communicating with a WOZ agent, and four users communicating with a WOZ agent. Then, we will analyze the data using statistical methods to reveal the relationship between verbal and nonverbal behaviors in human-agent multiparty conversations.

## 1. はじめに

会話エージェントは、具現化された形が実世界におけるロボットであっても、仮想世界のヴァーチャルキャラクタであっても、ユーザである人間とコミュニケーションを行うことを目的とするという点において共通している。一方、想定される使用場面が異なるために、実現しようとするコミュニケーションの形態には若干違いがあるように思われる。コミュニケーションロボットの研究では、ロボットが人間と共生することが重要な研究課題であったために、ユーザが複数人いることが想定された研究が比較的多いが、ヴァーチャルキャラクタを用いた複数人のユーザに対応できるエージェントについてはまだ研究が始まったばかりである [Bohus 2009].

しかし、どちらのタイプのエージェントにおいても、人間同士の会話に唐突に割り込んだり、話を混乱させるような介入を行うことなく、人間同士の会話のルールに従って、会話に加わることができるエージェントを実現するには、人間同士の会話の状態を詳細に観測し、適切なタイミングでコミュニケーション行動を実行するための判断や行動選択の機構が必要である。そこで本稿では、複数のユーザに対応し、ユーザ同士の会話に参入できるヴァーチャルエージェントの実現を目指し、Wizard-of-Oz システムを用いた複数ユーザとヴァーチャルエージェントとの会話収集実験について報告するとともに、いくつかの分析例を紹介する。

## 2. ユーザ同士の会話とユーザ対エージェントの会話

### 2.1 データ収集

大型ショッピングセンターに来店した 2 人組の客が、店内の案内役と会話を行うことで、必要な情報を得るという場面設定で会話の収集を行った。図 1 に示すように、Wizard-of-Oz での案



図 1 : 2 人ユーザによる会話収集実験

内役エージェントと会話をするペア 5 組、人間の案内役と会話をするペア 5 組、計 10 組がそれぞれ 2 会話行い、これらの計 20 会話を分析対象とした。

### (1) 収録データ

会話全体の様子と被験者の上半身を映したビデオデータ、会話音声データ、及び OKAO Vision による 2 人の被験者の視線の向き、頭部の向きのデータを収録した。

### (2) ヘッドトラッカデータを用いた顔向きの自動アノテーション

OKAO vision により計測したデータから各被験者の顔向き (正面, 右, 左, 下) を自動的に判別する決定木を作成し、これを利用して顔向きのアノテーションを自動化した。顔向き判別の決定木の精度は 93.5% であり、十分な分類精度が得られた。この方法により、人手によるビデオ分析を行うよりもアノテーションのコストが大きく削減できた。

## 2.2 分析

会話内容とユーザペアの顔向きとの関係を分析した結果を図 2 に示す。会話内容が案内役による説明である場合は、被験者は 2 人とも正面 (つまり案内役の方) を向く、あるいはペア相手の方に向くことが多い、それに対し、被験者のうちどちらか一方が案内役に質問している場合は、質問を行う被験者が正面を向くことが多い、また、被験者同士が相談や雑談をしている場合は、被験者同士がお互いに向きあうことが多く観察された。これらの

連絡先: 中野有紀子. 成蹊大学理工学部情報科学科. 〒180-8633 東京都武蔵野市吉祥寺北町 3-3-1. e-mail: y.nakano@st.seikei.ac.jp

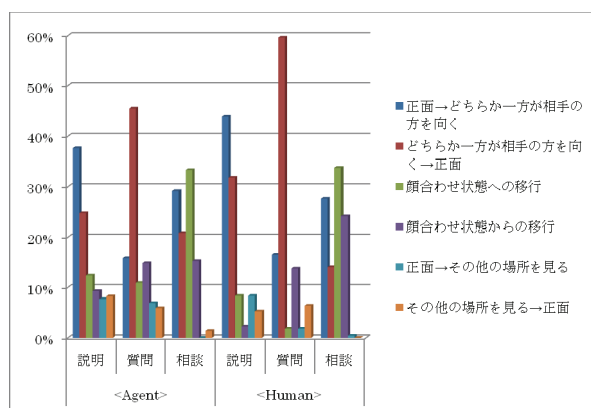


図2: 会話の状態と被験者ペアの顔向きとの関係

傾向は案内役が人間である場合とエージェントである場合でほぼ共通していた。

### 2.3 会話状態の推定

被験者ペアの現在の会話の状態が質問・相談・雑談のうちいずれであるかをリアルタイムで判定するために、コーパスの分析結果を用いて決定木学習を行った。会話の状態が説明の場合は、必ず案内役の発話であり、予測する必要が無いため、予測対象からは除外した。用いた特徴量としては、被験者ペアの顔向き・被験者ペアの1つ前の顔向き・発話者・発話前の無言状態の長さ・発話開始からの経過時間・7つ前までの会話状態のタイプである。データマイニングツール Weka の J48 に AdaBoostM1 を適用することで決定木を構築した結果、10回の交差検定における分類精度は、相談が74.5%、質問が68.8%、雑談が66.7%、全体での分類精度は70.5%であった。

以上の分析結果より、ヘッドトラッカから得られた頭部の動きと会話音声データから、ユーザがエージェントとコミュニケーションしようとしているのか(説明, 質問), ユーザ同士で会話をしているのか(相談, 雑談)を自動的に判別できる可能性が高いことが示唆された。この知見に基づき、会話状態推定方式を会話エージェントによるユーザへの介入タイミングの決定に利用した自律的会話エージェントを実装した[武田 2010]。

### 3. ユーザ間の多人数会話

多人数対話の参与枠組みは、話し手(Speaker:S), 発話の受け手(Hearer:H), 会話に参加しているが発話の受け手ではない傍参与者(Side-Participant:SP), 付近にはいるが会話に参加していない傍観者(Bystander)からなる[Clark 1996]。本実験では、動的に変化する参与役割と非言語行動の関連性を明らかにするために、4人の被験者と Wizard-of-Oz エージェントとの会話を収録した。

#### 3.1 データ収集

4人1組のグループ、計3組12人の大学生が実験に参加した。週末の外出先を話し合っ決めてもらうという課題について、2種類の条件設定で会話を行ってもらった。尚、被験者はエージェントに質問することにより行き先の場所や施設の情報を得ることができると教示した。

##### (1) 収録データ

会話全体の様子を撮影したビデオデータ, 会話音声, モーションキャプチャシステム OptiTrack による被験者の上半身と頭部

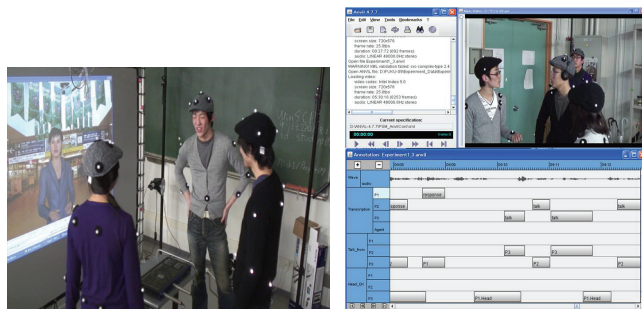


図3: 会話収録の様子

図4: コーパス例

動作のデータを収録した。実験の様子を図3に示す。あらかじめマーカが装着されたセータや帽子を用意し、これらを被験者に着用してもらうことにより、モーションキャプチャの設定を簡便化した。また、各課題終了後には会話の中心となった人物の順位付けのアンケートをおこなった。

##### (2) 作成したコーパス

ビデオデータと音声データから発話区間, 発話者, 発話の受け手のラベル付けを行った。また、モーションキャプチャから得られた各被験者の頭部と上半身の位置と回転角度の情報により各被験者が誰を注視しているかの推定をおこない、注視対象の自動ラベリングを行った。アノテーションの結果を図4に示す。

現在は、コーパスの整備が完了した段階であり、今後このコーパスについて詳細な分析を行っていく予定である。

### 4. まとめ

本稿では、ユーザ同士の会話の状況を感じることにより、適切なタイミングで会話に参加する複数人ユーザ対応型会話エージェントの実現を目指し、基礎的なデータの収集と分析について報告した。

本稿では、ヴァーチャルキャラクタを用いた Wizard-of-Oz 実験を行ったが、同様の課題設定でもロボットを用いた Wizard-of-Oz では被験者の反応が異なる可能性がある。それぞれのエージェントにおける身体性をユーザがどのように区別し、認識するのか、さらには、どのような会話において、エージェントにどんな役割をさせるには、どちらのタイプのエージェントがより適切であるのかに関しても、実験的に検討する必要があると考える。

謝辞: 本研究における頭部姿勢情報の収集には、オムロン株式会社の OKAO Vision 技術を利用させていただいた。

本研究の一部は科研費基盤(S)(課題番号:19100001), および科研費特定領域研究「情報爆発時代に向けた新しい IT 基盤技術の研究」(課題番号:21013042)の助成による。

### 参考文献

[Bohus 2009] Bohus, D. and Horvitz, E. Learning to Predict Engagement with a Spoken Dialog System in Open-World Settings. in SIGdial'09. (2009). London, UK.  
 [武田 2010] 武田信也, 中野有紀子 情報提供エージェントとの多人数対話における対話制御方式, 第72回情報処理学会全国大会, 5V-2, 2010.  
 [Clark 1996] Clark, H.H., Using Language. (1996), Cambridge: Cambridge University Press.