

複数人ユーザ会話におけるエージェントの 割り込みタイミングの推定手法の検討

Finding the Timings for a Virtual Conversational Agent to Intervene User-user Conversation

乙木 翔地^{1*} 堀田 怜¹ 黄 宏軒¹ 馬場 直哉² 中野 有紀子² 川越 恭二¹

Shochi Otogi¹ Ryo Hotta¹ Hung-Hsuan Huang¹

Naoya Baba² Yukiko Nakano² Kyoji Kawagoe¹

¹ 立命館大学情報理工学部情報コミュニケーション学科

¹ Department of Information and Communication Science, Ritsumeikan University

² 成蹊大学理工学部情報科学科

² Department of Computer and Information Science, Seikei University

Abstract: In these days, there are more and more opportunities to use conversational agent as the user interface of the systems in public facilities. However, in those situations, the system is often used by groups of visitors rather than the individuals. In the multi-user situation which is much more complex than single user one, specific features are required. One of them is the ability to intervene user-user conversation. If human agent can be able to do that, it can then engage in more natural conversation, i.e. mixed initiative conversations which occur frequently in human-human conversations. Therefore, in this research, we aim to develop the method to estimate the appropriate timings for the agent to do intervention. Firstly, we have conducted a WOZ experiment for data collection. We retrieved data on the direction of face and body, and acoustic information from the experiments. From the analysis result, we propose that there can be five kinds of timings that allow the agent to do intervention. We also propose the directions of the implementation of fully automatic system.

1 はじめに

近年、情報端末等ではユーザインタフェースとして擬人化エージェントを利用する機会が少しずつ増加している。美術館や映画館等の公共施設では、個人で利用するものだけでなく複数人のグループで利用する場面が想定される。このような複数人で利用するエージェントでは、個人で利用するものとは別に、複数人会話に特化した機能が必要となる [1]。我々はこれまでも、エージェントと複数人ユーザの会話に関する研究を行ってきた。ユーザからの問いかけに適切に応答するために、受話者推定手法を提案してきた [2]。しかし、ユーザに対する受動的な対応だけでなく、より柔軟にユーザに対して主体的な対応する必要がある。その一つに、会話への割り込み行動が存在する。従来のエージェントと人間とのインタラクションの対話システムでは、ユーザ主導型、もしくはエージェント主導型の

2種類で成り立っていることが多い。しかし、人対人のコミュニケーションでは、対話場面において互いに割り込みが生じ、主導権が会話の中で移り変わる混合主導型の会話が自然である。複数人ユーザに対しての擬人化エージェントによる情報提供の会話場面において、エージェントがユーザ同士の会話に適切なタイミングで介入し、有用な情報を提供することが可能となれば、より有効なシステムができ、かつ、エージェントの存在感が期待できる。これを実現するためにエージェントが情報提示を行う際にユーザの状態を考慮した対処を行うことが必要となる。本研究では会話への割り込みを行うタイミングを推定することを目的としたコーパス収集 WOZ (Wizard of Oz) 実験を行い、どのような場面で割り込み可能であるかを発話状況や姿勢等の非言語情報の解析を行った。分析結果より、割り込みタイミングを5つ定義し、その結果を報告するとともに、今後、混合主導型エージェントシステムの実現に向けたシステムの検討を行う。

*連絡先：立命館大学情報理工学部情報コミュニケーション学科
E-mail: is0010hr@ed.ritsumeikan.ac.jp

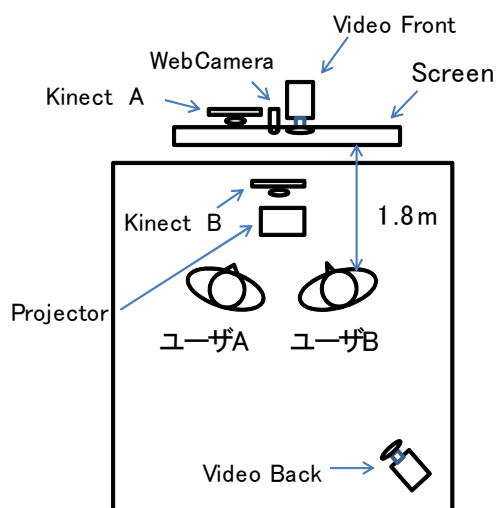


図 1: 実験環境

2 関連研究

対話システムの発話タイミングに関する研究は盛んに行われている。河添ら [3] の研究では、発話タイミングに影響を与える要因として「発話内容」、「相手との関係」、「発話者の性格」を挙げ、これら 3 つからタイミングを決定するアルゴリズムを検討している。また荒井ら [4] の研究では、人間の邪魔をしないという観点からマルコフモデルを利用し、対話の間を見つける手法を提案している。

田中ら [5] は、ユーザの割り込み拒否度を推定し、情報提示を行う手法を提案している。しかし、これらの手法は、邪魔をしないことを目的としタイミングを推定し、システムからユーザに積極的に働きかけを行っていない。そのため、積極的にユーザに対して働きかけを行い、またユーザ数が複数になった際に対処可能な手法が必要となる。

3 コーパス収集実験

複数人ユーザとエージェントの割り込みタイミングを探るためのデータ収集 WOZ 実験を実施した。

3.1 実験概要

実験の内容は、等身大の女性のアニメーションのキャラクターがスクリーンに投影され、前に離れてペアの被験者が並んで立ち、エージェントから情報を求める対話実験を行った。実験環境は、擬人化エージェントが一般的に使用させる場面として、ユーザがエージェン

トの操作を行い、情報を得ることを想定し、各グループに以下の 3 つのタスクをそれぞれ約 10 分間ずつ取り組んでもらった。ユーザ間において議論を活発にすること目的とし、最終的に第 1 希望から第 3 希望までを決めてもらう形とした。

履修登録タスク：実験協力者には、12 の授業のうち一緒に履修したい科目を第 1 希望から第 3 希望まで選ぶように教示をした。協力者は、チュータの役割を持つエージェントに質問することにより、授業に関する情報を得ることができる。エージェントから得られる情報には、授業の概要、担当の先生、単位取得の難易度、授業の時間帯などがある。

アルバイト紹介タスク：全 14 種類のアバイトから、実際にアルバイトをしたらどのアルバイトがしたいかを第 3 希望まで決めてもらった。エージェントより、時給、勤務地、勤務時間等の情報を得ることができ、それぞれの条件から、今の自分と照らし合わせてどのアルバイトがしたいかを決定する。より、現実に近い状況にするために勤務地は、大学の周辺を設定し、アルバイト内容は一般的に大学生が行なっているコンビニエンスストアでの販売等を設定した。

観光案内タスク：実験協力者には、1 泊 2 日の九州旅行への国内旅行を行う設定で、訪問する場所を第 3 希望まで 2 人で話し合って決めるよう教示した。エージェントから得られる情報は、各観光スポットについての簡単な歴史、地元の有名な食べ物、見どころ、観光地などである。より実験協力者に興味を持ってもらうために、あまり知られていない観光地等を設定した。

操作者はエージェントの反応時間を短縮するために、2 時間操作の練習をしてもらった。WOZ システムには、メニューによる発話選択に加え、必要な場合に任意の発話を入力できる機能も用意した。WOZ の操作者はディスプレイ画面を通して、被験者の様子を見ることができるよう環境設定した。

3.2 収集データ

本実験の被験者は、同性の友人ペア 12 組、計 24 人の立命館大学学部生、大学院生に協力をしてもらった。12 組のうち、男性ペアは 8 組、女性ペアは 4 組である。平均年齢は、19.2 歳であった。各グループには、上記の 3 つのコンテンツをそれぞれ 1 回ずつ行なってもらった。1 度目は、エージェントの対話に慣れてもらうため

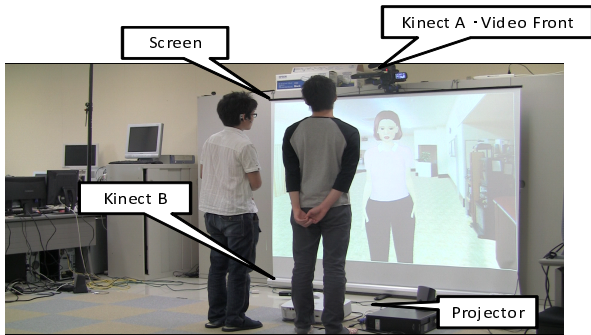


図 2: 実験風景

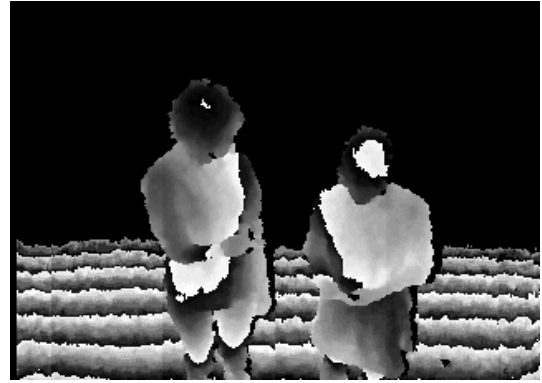


図 3: Kinect 深度情報

表 1: 1 ペアあたりのエージェント・ユーザの発話数

	エージェント	ユーザ	合計
平均	48.0	116.9	164.9
最大	72	182	254
最小	34	62	96

とし、今回は2回目以降のデータを計12セッションを扱った。会話内容は各コンテンツのはそれぞれ4つずつとし、偏りがでないように考慮した。1グループあたりの会話平均時間は、12分43秒であった。

3.3 収集方法

エージェントと被験者とのインタラクションは、前方と後方からの2台のビデオカメラにより録画した。これらのビデオデータに加えて、スクリーン上部に顔認識用のUSBウェブカメラ、姿勢情報取得のためのKinect(A)を設置した。また、音源情報取得のために被験者の前にも1台Kinect(B)を設置した。また、音声データの収録のためBluetoothのワイヤレスヘッドセットを装着してもらった。図1に実験レイアウトを、図2に実験風景を示す。

4 コーパス分析

分析を行うにあたって取得した音声情報、姿勢、顔方向と割り込みタイミングとの対応を見やすくするためにビデオデータのアノテーション手法を用いた。アノテーションツールには、Anvil¹を使用した。

4.1 音声データ分析

発話時間・発話者

音声データでは、Kinectのマイクロフォンアレイを用い発話者、発話時間を取得した。取得したデータを元に発話者（エージェント・ユーザ右・左）の計3種類のラベルを定義し、Kinectから取得したログを元に自動ラベリングを行った。また、データがすべて均一にとれているわけではなく、取得しきれなかった部分に関しては、手作業で補正した。発話数を表1に示す。また、1発話あたりの平均時間は2.55秒であった。

受話者

発話者、発話時間のラベリング結果を元に、受話者について手作業でラベル付を行った。ラベルは、ユーザの発話が誰に向けられたものかを知るために、エージェント・パートナー（もう片方のユーザ）とした、またエージェントの発話はすべてユーザ2人を受話者とした。

4.2 顔方向の分析

顔方向について分析を行った。実験時に、ウェブカメラのビデオデータからリアルタイムで顔認識ソフトウェアFaceAPI²を用いて、顔方向の位置を秒間30フレームで取得した。これらのデータをデータマイニングツールWeka³のJ48決定木により分類し、顔方向を正面、もう片方のユーザ方向、その他の3種類でラベルを定義し、FaceAPIで取得したログより自動でラベリングを行った[2]。しかし、リアルタイムで顔方向が取得しきれなかった部分に関しては、オフラインでスクリーン上に設置したビデオデータを用いデータの再取得を行った。

¹<http://www.anvil-software.org/>

²<http://www.cs.waikato.ac.nz/ml/weka/>

³<http://www.seeingmachines.com/product/faceapi/>

表 2: 1 ペアあたりの平均タイミング数

割り込みタイミングの種類	平均
Stagnation	1.0
Check	3.0
Argument	2.4
Arrangement	0.9
Decision	1.0

4.3 姿勢分析

顔方向だけでなく、ユーザの姿勢情報からも有益な情報が得られると考え、Kinect の深度情報を用いて 320 × 240 の解像度で、秒間 30 フレーム取得した。取得したデータ例を図 3 に示す。今回は、これらの情報から 4 方面（正面、もう片方のユーザに対して小さな方向転換、大幅な方向転換、その他）について手動で、ラベリングを行った。今後、自動化システムの構築のために取得した Kinect の深度情報を顔認識と同様に、手動でラベリングを行ったデータを Weka を使用し、自動ラベリングを可能とする予定である。

5 割り込みタイミング

分析結果より、エージェントが割り込み可能なタイミングを 5 つの種類を定義し、以下の 5 項目についてラベリングを行った。

5.1 割り込みタイミングの分類

Stagnation : 会話の停滞の場面。会話の中でユーザがエージェントへの質問が浮かばない、意思決定に向けてのヒントが見つからない等により、議論が停滞している場面が見受けられた。この場面では、エージェントより「何で悩んでいますか」等のユーザに対して発話を行うことで、会話の活性化が期待できる。

Check : ユーザ同士のエージェントへの質問内容の確認をしている場面。実験では、エージェントに質問を行う前にユーザ同士で質問の内容を確認し、後に情報提供を求める質問を行う場面が見られた。会話の例を以下に示す。

ユーザ A「(エージェントに対して) アルバイトの時給を聞いてみる？」

ユーザ B「そうだね、聞いてみようか」

ユーザ A「時給はいくらですか？」

この場合、エージェントが直接ユーザに対して質問への応答が可能となればユーザ間の確認の手間を省けることが期待できる。

Argument : エージェントの発話内容に対しての議論を行なっている場面。エージェントの直近の発話内容に関して、不明な点があった場合にユーザ間でその発話に関する意味の確認を行なっている場面が見受けられた。この場面では、エージェントが不明な点をユーザに確認し、更に具体的な情報提供を行うことが可能となる。

Arrangement : ユーザ同士がそれまでに得た情報を整理を行なっている場面。ユーザ間で会話の中でエージェントより得られた情報を整理している場面が見受けられた。この場合、情報を提供したエージェントが会話の中に入り、情報整理を協力することで、より正確な情報提供が期待できる。

Decision : エージェントへの意思決定を行うための議論を行なっている場面。ユーザ間で、意思決定を行なっている場面では議論が生じていた。この際にエージェントよりユーザに対して意思決定に有用な情報提供を行うことで、円滑な意思決定を支援することが期待できる。

上記 5 つのラベリングの総数は 12 グループで、計 127 タイミングあった。表 2 に各定義ごとのタイミング数を示す。

5.2 コーパスの分析

割り込みタイミング数は 1 グループあたり 10.6 回あり、平均約 1 分あたり 1 回の割合でエージェント割り込める込めることが示された。これは概ね、多いと考える。図 4 より各グループにおいての定義数については、回数にばらつきが見られた。Ccheck に関しては、数にばらつきがあるものの各グループに見られた。また、1 コンテンツ内にすべての定義が含まれるものは 3 グループであった。これより、グループごとに定義に一定の規則性は見られなかった。しかし、男女別に見ると、女性グループでは定義数が平均 4.5 種類入っていた。男性グループの定義数の平均 3.3 種類と比べると 1

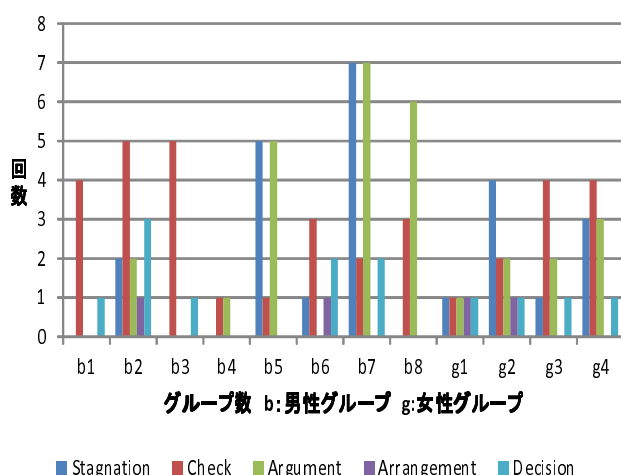


図 4: 各グループにおけるタイミングの数

グループあたり 1.2 種類多く、男性よりも女性の方が 1 コンテンツあたりの定義の種類が多いことを示している。これにより、男性と女性に対して違う検知ルール使う必要がある。また、グループのエージェント・ユーザの総発話数との関係を見ると表 1 で示したグループの平均発話数 164.9 より発話が多いグループは、7 つあり、定義数が平均 10.9 個あった。これは平均発話数を下回る 5 グループの平均 6.0 個を 4.9 個上回ることから、発話数の多さが定義数に影響することが見られた。

6 割込みタイミングの自動検出に向けて

今後、実験から得られた、非言語情報より判別可能な定義について、タイミングを自動で判別するためのシステムを構築を行う。現在、非言語情報のみで判別できる定義として Stagnation, Cheak, その他の 3 つの種類の 3 つについて分類可能であると考える。システムの構築にあたり重要となるのは、適切な判別のアルゴリズムである。アルゴリズムの検討には、実験で収集した音声情報・顔向き情報・姿勢のラベリングの結果を用いる。また、ラベル付を行った部分だけでなく、最後のラベルからの経過時間や直近のラベルだけでなく過去数十秒間のラベル結果を用いることで、定義した場面の発見、分類が可能ではないかと考える。それらの結果より、定義した割り込みタイミングまでのラベルの関係を導き出しアルゴリズムを確立する予定である。アルゴリズムの確立後、順次音声分析処理・顔向き判定処理・姿勢判定処理の構築より、リアルタイムシステムの構築に取り掛かる予定である。

7 おわりに

本稿では、複数人での会話場面において、エージェントがユーザ同士の会話に介入するためのタイミングを非言語情報より推定するために行った WOZ 実験の分析結果を報告した。まず、エージェントの割り込みタイミングを推定するために WOZ 実験を行い、会話データを収集した。次に、収集した音声情報、顔方向、姿勢について分析を行った。その後、分析結果より割り込みタイミングを 5 つ定義した。更に、その定義より、自動検出システムの構築を行うための検討を行った。最後に、今回は非言語情報で分類可能な定義について、システムの検討まで行った。これらから非言語情報のみで、割り込みタイミングの検出が可能であると推定した。今後、リアルタイムでの割り込みタイミング自動検出のためのシステム構築を行なっていく予定である。

8 謝辞

本研究は科研費若手 (B)No.23700183 の助成による。

参考文献

- [1] D.Traum.Issues in multiparty dialogues. In Advances in Agent Communication, International Workshop on Agent Communication Language (ACL'03), pages201-211, 2003.
- [2] 馬場 直哉, 黄 宏軒, 中野 有紀子. グループ会話対応型会話エージェントにおける受話者推定システム. HAI2011
- [3] 河添 麻衣子, 北村 泰彦. 発話タイミングを考慮したマルチエージェント対話システム. 電子情報通信学会技術研究報告”, AI, 人工知能と知識処理 106(617), 53-56, 2007-03-21
- [4] 荒井 考晋, 高橋 桂太, 金子 正秀. 受取人の状態に応じて割込み行動を変化させる配達ロボット. 映像情報メディア学会技術報告 36(8), 139-142, 2012-02-11
- [5] 田中 貴紘, 藤田 欣也. ユーザの割り込み拒否度推定に基づくインタラクション仲介エージェント. HAI2010