

Wizard-of-Oz による情報提供エージェントとの

多人数対話における言語・非言語行動の分析

An analysis of verbal and nonverbal behaviors in
multiparty conversations with Wizard-of-Oz kiosk agents

武田 信也^{*1}
Shinya Takeda

中野 有紀子^{*2}
Yukiko Nakano

^{*1} 東京農工大学 工学府 情報工学専攻
Tokyo University of Agriculture and Technology

^{*2} 成蹊大学 理工学部
Faculty of Science and Technology, Seikei University

The goal of this study is to build a kiosk agent that can manage multiparty conversations with two users. Since little has been studied about how to design agent's behaviors in multiparty conversational interfaces, first, this study collects multiparty conversation corpus using a Wizard-of-Oz setting. In addition to audio and video data, head pose data are collected using a face-recognition technology. We will analyze the data using statistical methods to reveal the relationship between verbal and nonverbal behaviors specifically in terms of turn-taking.

1. はじめに

情報提供端末として利用されるエージェントは、複数ユーザと応対できることが望ましいが、その設計の基礎となる理論はまだ確立していない。多人数対話における人間の会話行動を調べた研究として、榎本らは、3人会話における発話交代の仕組みを視線の動きや顔きの中に見出し、非言語行動と話者交代との関係を分析した [1]。また、坊農はポスター発表会場での3人会話における発話交代の仕組みを、視線配布パターンや発話の統語形式に見出し、言語・非言語行動と話者交代の仕組みを分析している [2]。しかし、これらは自由対話を対象とした分析であったり、ポスター発表に特化した分析であり、本研究が目指す情報提供端末用のエージェントを実装するためには、その設計指針を得るためのデータ分析が必要であると考えた。

そこで本研究では、エージェントと2人のユーザとの3者対話の実現を目指し対話収集実験を行った。本稿では3人の人間同士での対話と、Wizard-of-Ozのエージェントと被験者2人の3者の対話における言語・非言語の関係の分析、および顔認識技術を利用した、顔動作の統計的分析により判明した結果を報告する。

2. 対話収集実験

情報提供エージェントとユーザとで多人数会話を行うシステムを実現するための設計指針となる、話者交代や会話中に行われる言語・非言語行動の仕組みを解明するために、その基礎となるデータ収集を目的とした実験を行った。

2.1 状況設定

本実験は、大型ショッピングセンターに来店した2人組みの顧客が、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にした。情報提供エージェントが設置される場所は、駅構内、美術館、アミューズメント施設など敷地内が複雑に入り組んでいて分かりづらい場所であると想定されるが、多くの地域に存在していて、将来多くの使用が見込まれる大型ショッ

ピングセンターでの場面設定にした。

2.2 実験目的

本実験の目的は、会話相手がエージェントであるか人間であるかによって言語・非言語行動にどのような差異が生じるかを明らかにすることである。

2.3 実験条件

エージェント条件:案内役のアニメーションエージェントと客役の2人の被験者の、3者間で会話を行う

3人会話条件:案内役1名と客役となる被験者2人の、3人で会話を行う

エージェント条件の場合、音声合成により作成された案内エージェントの発話やそれに伴う動作は、人間と同程度には自然なものにはならないので、3人会話条件(案内役が人間の場合)と比べ被験者が違和感を感じてしまうのではないかと予想される。一方、3人会話条件の場合、案内役の会話・身体表現がエージェントよりも大幅に広いバリエーションを取ることができるが、そのことがインタラクションの質に大きく影響することを避けたいと考えた。そこで、3人会話条件の案内役は、エージェントのように機械的な応対を行い、挨拶や場所の指さし程度の身体動作だけを行うようにした。

2.4 被験者

被験者は、エージェント条件と、3人会話条件それぞれつき、同性の友達同士のペア、5組ずつ、計10組20名の大学生、大学院生である。

2.5 収集データ

本実験で収集するデータは、会話全体の様子を収録したビデオデータ、被験者の上半身(特に顔や視線の動き)の様子を収録したビデオデータ、会話の言語情報を収録した音声データ、および被験者の顔データである。顔データの数値化に際しては、Okao Visionを使用した。Okao Visionを用いることにより、2人の被験者の視線の向き、頭部の向き、目や口の開き具合などを測定できる。なお、この機能は複数人にも対応している。Okao Visionでの測定の様子を図1に示す。



図1 Okao Vision での測定

3. 実験内容

3.1 実験課題

各被験者ペアに与える課題は共通であり、以下に示す 3 種類の課題であった。実験開始前、被験者には案内役に質問可能な項目として、各店舗ごとに質問可能な基本項目(店舗概要、営業時間、お勧め商品など)と、特定の商品にのみ質問可能な項目(サイズ、色、素材など)があるという情報を与えた。実験中はこれらの質問可能な項目に加え、店内のフロアマップも同時に閲覧できるようにした。

(A)大型ショッピングセンターが新規オープンしたので、友人と 2 人で買物に来たが、店内にどのような店舗が存在するか分からないので、案内役と会話を行い、店舗についての情報を聞き出す。必須質問として、ショッピングセンターで行われているイベント情報を聞く。

(B)この大型ショッピングセンターには来店経験がある 2 人で、今回は服と靴、そして家にお土産を買って帰ろうと思っている。そこで、案内役と会話を行い、どの店舗へ行くか意思決定の参考にしてもらう。必須質問として、店舗ごとに行われているキャンペーン情報を聞く。

(C)大型ショッピングセンターに買物に来た 2 人だが、1 人が急にお手洗いに行きたくなったので、案内役に場所を尋ね、分かり次第お手洗いへ向かう。必須質問として、お手洗いの場所を尋ね、同時に急いでいる気持ちになって取り組んでもらう。

なお、課題は(A)→(B)→(C)の順に実施し、対話は課題 A と課題 B では約 5 分、課題 C では 1 分程度で収録を行った。

3.2 実験環境

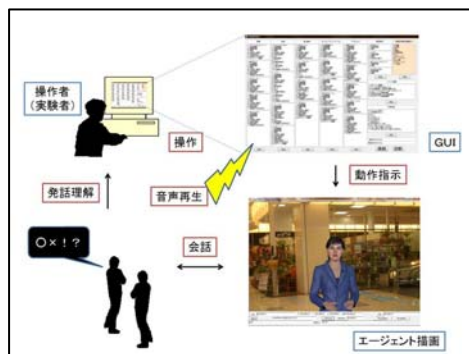


図2 実験システム

実験システムの概要を図 2 に示す。エージェント条件では、操作者(実験者)が GUI を操作することにより、エージェントの音声発話とそれに付随する動作アニメーションが出力される。被験者 2 人はエージェントの発話を聞き、相談を行い、エージェントに再度質問をする。さらに、その発話を操作者が理解して、GUI を操作する、という一連の流れが会話終了まで続く。なお、3 人会話条件では、GUI とエージェントアニメーションは使用されない。

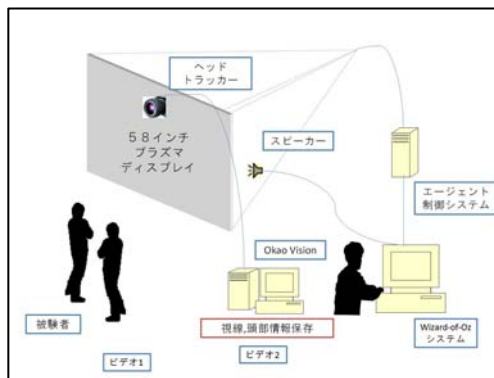


図3 システム構成

次に、実験システムの構成を図 3 に示す。エージェント条件において、操作者(実験者)が GUI を操作することにより選択されたエージェントの発話はスピーカから出力され、それに付随する動作はエージェント制御システムに送られる。エージェントアニメーションは 58 インチプラズマディスプレイ上で提示される。被験者 2 人の視線・頭部情報は Okao Vision を通じて数値化され、XML 形式で記録される。視線・頭部の動き、および会話中の言語情報は 2 台のビデオカメラにより収録される。

4. 機械学習

被験者がどの方向に顔を向けているかを自動的に判定するために、Okao Vision により取得した顔データを用いて、決定木学習を行った。決定木学習は、データ集合を属性値に応じて部分集合に分割することを再帰的に繰り返すことで決定木を生成する手法であり、決定木が分類モデルとなる。学習にはデータマイニングツール Weka の J48 を用いた[3]。用いた特徴値はヘッドトラッカーの設置方向に対する顔の左右方向の角度を示す LeftRight, 上下方向の角度を示す UpDown, 顔の回転角度を示す Roll, 視線の左右方向の角度を示す GazeLR, 上下方向の角度を示す GazeUD の 5 つである。なお、数値が正の場合は右、または上方向に対応し、数値が負の場合は左、または下方向に対応している。教師データには、案内役がエージェントである場合・人間である場合それぞれ 2 ペア合計 8 人のビデオについて、被験者の顔の向き(正面(front), 下(down), 左(left), 右(right)の 4 種類)のラベル付けを行った結果を用いた。ラベリング作業には、ビデオアノテーションツール Anvil4.5 を用いた[4]。

決定木学習の結果、サイズが 37、葉の数が 19 のモデルが生成され、10 回の交差検定における分類精度は 93.4623%であり、十分な精度が得られた。

5. 実験結果と考察

次に、4 節で得られた決定木により自動判定された 2 人の被験者の顔の向きを、ペアとしてどのような状態にあるかによって

5 種類に分類し、それらをペアの顔動作状態にした。なお、正面を向いていると思われる状態には front 以外に down も含まれる。これは被験者同士の身長差が front と down の値に大きく関係してきたため、機械学習では完全な判別が困難であったためである。

■ 正面→正面

被験者 2 人が正面を向いている状態から変わらず、そのままであることを示す。

■ 正面→左右

被験者 2 人が正面を向いている状態から、被験者の 1 人がパートナーの方を向いている状態に移行することを示す。

■ 左右→正面

被験者の 1 人がパートナーの方を向いている状態から、被験者 2 人が正面を向いている状態へ移行することを示す。

■ 顔合わせ状態への移行

被験者の 1 人がパートナーの方を向いている状態から、被験者 2 人がお互いに顔を向きあっている状態に移行することを示す。

■ 顔合わせ状態からの移行

被験者 2 人がお互いに顔を向きあう状態から、被験者の 1 人が正面を向きあっている状態に移行することを示す。

5.1 発話状態による差異

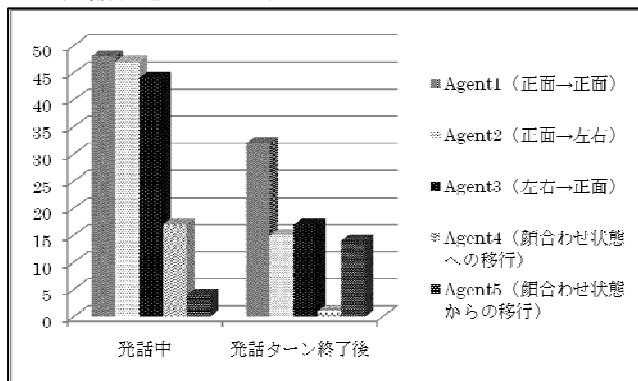


図 4 案内役がエージェントの場合における顔動作状態の変化が生じた時点の分布

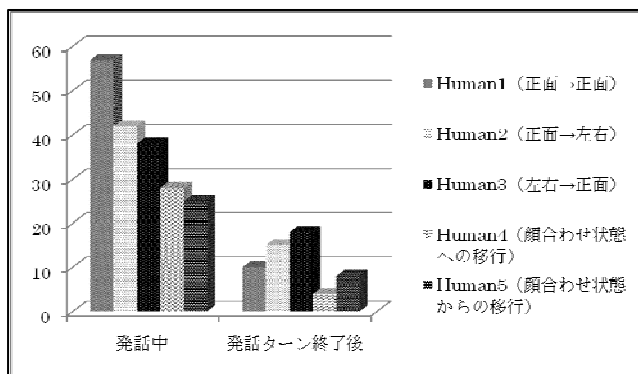


図 5 案内役が人間の場合の場合における顔動作状態の変化が生じた時点の分布

発話状態による被験者の顔動作状態の分布を、図 4 および図 5 に示す。この際、顔動作状態が変わった時の発話状態は、以下の 2 種類に分類される。

■ 発話中

案内役、もしくは被験者が何らかの類の発話を行っている状態を示す。

■ 発話ターン終了後

誰も発話を行っていない状態のことを示す。

➤ 考察

案内役がエージェントの場合と人間の両方において、発話中と発話ターン終了後で顔動作状態の分布が異なっていることが見て取れる。特に、発話中において、正面→正面、正面→左右、左右→正面の頻度が高くなっていることが分かるが、これについては、次節で詳しく分析する。

5.2 発話者による差異

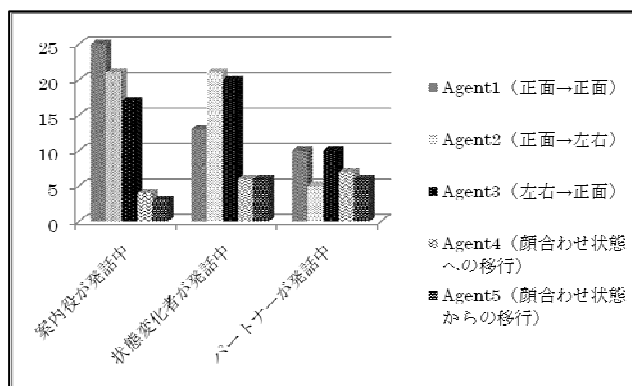


図 6 案内役がエージェントの場合における発話者の違いによる分布

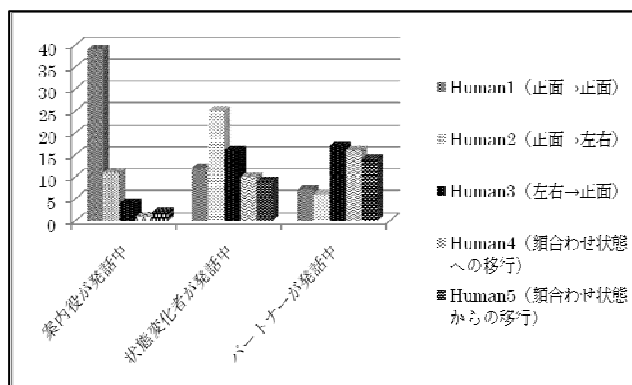


図 7 案内役が人間の場合における発話者の違いによる分布

前節の分析結果より、発話中と発話ターン終了後で分布が異なることが分かったが、発話中という状態は、誰が発話者であるかによって、さらに以下の 3 種類に分けることができる。発話者による分析結果を図 6 および図 7 に示す。

■ 案内役が発話中

案内役が被験者に向けて発話を行っている状態で、被験者のどちらかに顔動作に変化が生じた場合を示す。

■ 状態変化者が発話中

発話を行っている被験者の顔動作に変化が生じた場合を示す。

■ パートナーが発話中

発話を行っていない被験者の顔動作に変化が生じた場合を示す。

➤ 考察

案内役が発話中の場合、被験者は 2 人とも正面を向くことが多いが、案内役がエージェントの場合では、被験者が左右を向く場合も頻りに観測される。それに対し、案内役が人間の場合は正面を向くことのみが圧倒的に頻度が高い。これは、案内役がエージェントの場合被験者がその存在を軽視しているためだと推測される。また、状態変化者が発話中の場合、正面から左右、および左右から正面へと被験者の顔動作状態が移行する傾向が強いことが分かった。これは、パートナーの様子をうかがいながら発話を行っているためであると考えられる。さらに、パートナーが発話中の場合、発話していない被験者の顔動作状態が左右から正面へと移行する傾向が強いことが分かった。これは、パートナーが発話を始めたことを受けて、相手の話を聞きつつ案内役の回答に注目するという行動をとるのだと推測される。

5.3 発話内容による差異

発話内容によって、顔動作状態の分布がどのように変化するかを調べるために、まず、発話内容を以下の 6 種類に分類した。

- 説明
案内役が被験者に行う、店舗についての情報提示。
- 質問
被験者が案内役に行う、店舗情報についての質問。
- 相談
被験者同士が行う、次の取るべき行動の選択に関する議論。
- 理解
被験者による、案内役の店舗についての説明に対する納得の表明。
- コメント
被験者同士が行う、店舗とは無関係の雑談。
- その他
上記のいずれにも該当しない発話(例:笑い)。

発話内容による被験者の顔動作状態の分布を、表 1 および表 2 に示す。

表 1 案内役がエージェントの場合における発話内容による分布

	説明	質問	相談	理解	コメント	その他
Agent1(正面→正面)	46%	24%	39%	20%	6%	1%
Agent2(正面→左右)	37%	19%	22%	11%	44%	8%
Agent3(左右→正面)	2%	1%	25%	13%	11%	2%
Agent4(顔合わせ状態への移行)	10%	5%	4%	2%	28%	5%
Agent5(顔合わせ状態からの移行)	6%	3%	10%	5%	11%	2%

表 2 案内役が人間の場合における発話内容による分布

	説明	質問	相談	理解	コメント	その他
Human1(正面→正面)	69%	38%	30%	12%	0%	0%
Human2(正面→左右)	18%	10%	10%	4%	34%	12%
Human3(左右→正面)	7%	4%	50%	20%	11%	4%
Human4(顔合わせ状態への移行)	2%	1%	0%	0%	37%	13%
Human5(顔合わせ状態からの移行)	4%	2%	10%	4%	17%	6%

➤ 考察

発話内容が説明の場合は、基本的に被験者は正面を向いていることが多い(Agent:46%, Human:69%)が、同時に正面から左右を向くことも多くなっている(Agent:37%, Human:18%)ことが分かった。これは、案内役の説明を聞くのと同時に、パートナーの様子をうかがっているためであると考えられる。また、発話内容が質問の場合も同様に、基本的に被験者は正面を向いていることが多い(Agent:39%, Human:30%)が、質問時の特徴として左右から正面へと顔動作状態を移す傾向が強い(Agent:25%, Human:50%)ことが分かった。これは、案内役の質問に対する返答に注目するためであると考えられる。さらに、発話内容が相談の場合、被験者の顔動作状態が正面から左右、もしくは顔合わせ状態に移行する傾向が強い(正面から左右→Agent:44%, Human:34% 顔合わせ状態に移行→Agent:28%, Human:37%)ことが分かった。この結果は、相談をする際には相手の方を向くということを考えれば、ごく自然な結果であるといえる。また、相談が行われる際、被験者の両方が正面を向いていることがまず無いのも特徴的である。その他の傾向として、発話内容が理解の場合は、すべての顔動作状態がほぼ均等に出現しているが、これは、相手の発話内容に同意する場合には、理解表明者の顔が様々な方向に動くことを示している。また、発話内容がコメントの場合は、正面から左右と顔合わせ状態に移行の割合が高く、相談の場合に近い顔動作状態を示すという特徴が見られた。

5.4 発話ターン終了後における前発話内容による差異

発話ターン終了後における前発話内容は、発話中と同様に 6 種類に分類した。なお、発話ターン終了後の前発話内容による被験者の顔動作状態の分布をまとめた表は、サンプル数が十分ではなかったため割愛する。

➤ 考察

サンプル数は少なかったが、発話ターン終了後における前発話内容による分布は、発話中における発話内容による分布と近い傾向を示した。その他の割合が多くなっていることが見受けられるが、この大部分は、発話ターン終了後に笑いが起こる場合であった。

6. おわりに

本論文では、エージェント、もしくは人間と 2 人のユーザとの 3 者対話における言語・非言語の関係の分析、および顔認識技術を利用した、顔動作の統計的分析により判明した結果を報告した。今後は、顔動作状態の時間的な関係、継続長等についても詳しく調べていく予定である。

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 技術を利用しています。本研究の一部は科研費基盤(S)(課題番号:19100001)の助成による。

参考文献

- [1] 榎本美香・伝康晴・松坂要佐:3 人会話における談話行為と受け手のラベリングとその基礎的分析, 人工知能学会研究会資料, SIG-SLUD-A502, 2005.
- [2] 坊農真弓:日本語会話における言語・非言語表現の動的構造に関する研究, ひつじ書房, 2008.
- [3] Weka: <http://www.cs.waikato.ac.nz/ml/weka/>
- [4] Anvil: <http://www.dfki.de/~kipp/anvil/description.html>