

複数の対話エージェントを用いた 音声対話システムの分析と評価

藤堂 祐樹¹ 西村 良太² 山本 一公¹ 中川 聖一¹

概要：現在のほとんどの音声対話システムは、ユーザとシステムの1対1の対話を扱っているが、我々はシステム側のエージェントを2人にした三者対話システムの開発を行ってきた。対話ドメインは従来の「うどんとラーメンのどちらが好きか」の他に「北海道旅行と沖縄旅行のどちらに行きたいか」を追加し、それぞれ同じドメインの二者対話システムと三者対話システムをユーザに使用し、評価実験を行った。被験者実験の結果、三者対話システムは、対話の弾み具合の向上や、対話の雑談らしさの印象を被験者に与えるという結果が指示された。

キーワード：音声対話システム，多人数対話，複数エージェント，雑談対話

Analysis and Evaluation of Spoken Dialog System with Multiple Agents

YUKI TODO¹ RYOTA NISHIMURA² KAZUMASA YAMAMOTO¹ SEIICHI NAKAGAWA¹

Abstract: Almost all current spoken dialog systems have treated dialog that one user talks with one agent. On the other hand, we investigated the multiparty dialog system which treated two agents. We developed the three person's dialog system and two person's dialog system in one domain, which treated the same dialog tasks "Which do you prefer *udon* and *ramen* ?". We also developed a new dialog system in another domain, "Which do you prefer trip to *Hokkaido*(snowy region) and *Okinawa*(tropical region) ?", and compared user's behavior/satisfaction. According to the results of the experiments, the three person dialog system performed better in terms of lively conversation, and user can talk with the agents more like chatting.

1. はじめに

近年、音声認識技術を用いたインターフェースの需要が高まっており、それに伴って音声対話システム、特に雑談対話システムの開発が行われてきている [1], [2]。我々も、これまでに音声対話システムの開発を行ってきており、より自然な対話を実現することが重要であると考え、人間同士の雑談対話中にて生じる種々の対話現象を模倣する音声対話システムを構築した [3]。本研究では、ユーザを対話に引き込み、より楽しく対話ができる環境の構築を目指す。その為に、これまでのユーザ対システムという1対1の対話を、1ユーザ対多エージェントとの対話に拡張した [4]。これにより、新しい形態の対話システムを構成することができ、これまで実現不可能であった対話を実現させることが期待される。例えば、エージェント間の上下関係や、

ユーザ専属のエージェント、エキスパートエージェントなど知識の差別化を図ることや、考えの異なるエージェントとの対話に発展させることによってユーザに新たな考えをうながす効果が期待できる。

多人数対話の先行研究として、人同士の多人数対話の場合、Dielmannら [5] は、多人数対話での Dialog Act を自動で付与するためのモデルの学習を行っている。また複数のユーザと単一のシステムとの対話の場合、藤江ら [6] は、視線の制御やうなずきの機能を備えた対話ロボットを使用し、デイケアセンターにて多人数対話を試みている。対話ドメインは漢字の読みを当てるクイズであり、対話ロボットが複数のユーザから活発な発話を引き出している。浅井ら [7] は、複数の人間と複数の対話エージェントによる多人数対話において、対話エージェントが状況に応じた働きかけを行うことで、全体のコミュニケーションを活性化させている。対話はテキストベースの対話システムで行われており、2名のユーザと、2つのエージェントが対話に参加している。対話ドメインは、人物当てクイズなどをユーザ

¹ 豊橋技術科学大学
Toyohashi University of Technology
² 名古屋工業大学
Nagoya Institute of Technology

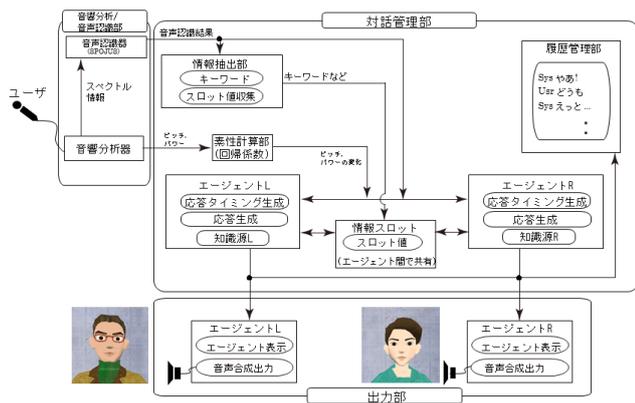


図 1 三者対話システムの概略図

に対して出題し、ヒント文を順に提示するというものである。2つのエージェントは、出題エージェントと回答エージェントに分かれており、両方が共感的発言や自己中心的発言を行う。対話実験の結果、ユーザの満足度やユーザの発言数を増加させる効果があることが示され、エージェントからの共感的発言がユーザ満足度を更に向上させ、対話を活性化させている。

このように、複数のエージェントとの対話はユーザ満足度の向上や対話の活性化に繋がることが示唆されている。

本稿では、従来の「うどんとラーメンのどちらが好きか」という対話ドメイン以外に、「北海道旅行と沖縄旅行のどちらに行きたいか」というドメインに対して、単一の対話エージェントと、複数の対話エージェントとでそれぞれ対話実験を行い、複数の対話エージェントが被験者に与える印象、満足度について分析した結果を報告する。

2. 三者対話システム

これまで我々が開発してきた音声対話システムは、ユーザ対システムでの1対1の対話を扱ったものであったが、これを、“性格の異なる2つのエージェント(システム)とユーザとの3人対話”に拡張した[4]。エージェント間では、実際に発話した内容以外にも、すべての情報が共有できる為、様々な対話制御が可能となり、広い応用が考えられる。今回構築した三者対話用の音声対話システムの概略図を図1に示す。このシステムでは、音声認識の結果から、テンプレートマッチングによって応答文を生成し、韻律素性を決定木に入力することで、応答の種類とタイミングを決定している[3]。

2.1 対話ドメイン

2.1.1 うどんとラーメン [4]

システムとの対話内容としては、誰でも気軽に対話ができ、また、三者対話において、ユーザの引き込みを実現させることができるのが好ましい。このことから、2つの物・事柄の好き嫌い・賛成反対の話題を扱う、「うどんと

ラーメンのどちらが好きか」といった話題で対話を行うようにした。

2.1.2 北海道旅行と沖縄旅行

今回は、(1) 三者対話システムの拡張性・汎用性 (2) 対話ドメインによるユーザ印象の変化を示すため、別のドメインを扱う対話システムを構築した。もう1つのドメインとして「北海道旅行と沖縄旅行、どちらに行きたいか」を扱う。

二者対話システムでは、1人のエージェントが旅行(麺類)好きであり、北海道旅行(うどん)と沖縄旅行(ラーメン)両方をユーザに薦める。三者対話システムでは、それぞれのエージェントが北海道旅行(うどん)好き、沖縄旅行(ラーメン)好きであり、それぞれ良い点・悪い点をユーザに示して対話を進めていく。

2.2 音響分析・音声認識部

本システムで用いる音声認識器には、本研究室で開発されたSPOJUSを用いる[8]。登録単語数は、「うどんとラーメン」システムで約270単語、「北海道と沖縄」システムで約430単語であり、エージェントの応答文の数はそれぞれ約400と800である。音声認識と同時に、本システムでは、音響分析として韻律情報の抽出も行っており、ピッチ・パワー情報を抽出して応答タイミング生成部へ送信している。これは、決定木の素性として用いている[3]。

2.3 対話管理部

2.3.1 素性計算部 [3]

ここでは、音響分析器から得られた音響分析結果を元に、韻律素性を計算している。素性としては、フレーム毎にピッチ(F0)とパワーの回帰係数を求め、これを応答タイミング・応答種類制御をする決定木の入力として用いる。

2.3.2 情報抽出部

ここでは、音声認識器からの認識結果から、必要な情報を抽出し、スロットに格納している。スロットに格納された値は、応答生成に用いられる。これにより、ある程度文脈を考慮した対話が可能となっている。対話ドメインが「北海道旅行と沖縄旅行、どちらに行きたいか」の場合、スロットの例としては、「ユーザが好きな旅行」「その旅行が好きな理由」「いま話している観光地」などがあり、これらの情報を認識結果から抽出し、対話を行う。

2.3.3 情報スロット

対話中の重要な情報がスロットに格納されており、これらについては、エージェント間で情報を共有している。この情報を参照して、ユーザの嗜好に合わせた共感発話を行い、対話を盛り上げる方向に進める。また、共有している情報を元に、対話の流れ(シナリオ)を変化させ、情報を応答に盛り込み、結論の誘導を行うことが出来る。

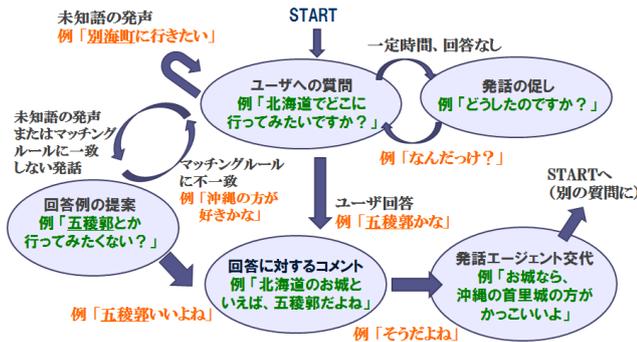


図 2 対話の状態遷移(三者対話)

2.3.4 応答生成部

本システムでの各エージェント内の応答生成には、各知識源に基づくテンプレートマッチングを用いている。入力された音声を音声認識し、その結果と応答用テンプレートとのマッチングを行って、マッチするものに対して、それに対応した応答文を出力として用意する。出力文を生成する際には、スロット情報も用いて、文脈を考慮した応答文生成を行うことができる。また三者対話システムにおいて、ユーザがどちらのエージェントと対話を行うかについても、応答用テンプレートで決定している。

三者対話での対話例を以下に示す。システム L が沖縄を、システム R が北海道をそれぞれユーザに薦めている。

システム L : 沖縄旅行と北海道旅行、どっちに行きたいですか?
 ユーザ : 沖縄かな。
 システム L : 僕も沖縄派なんです。
 どんなところに行ってみたいですか?
 ユーザ : やっぱり首里城とか見たいよね。
 システム L : やっぱり沖縄と言えば、首里城だよな。
 ユーザ : そうだねえ。
 システム R : お城なら、北海道の五稜郭だって、負けてないですよ。
 ユーザ : まあそうかもね。
 システム R : 北海道旅行もいいですよ。どんなところに行ってみたいですか?
 ユーザ : 旭川かな。
 システム R : 旭川と言えば、さいきは旭山動物園が有名ですね。

三者対話での、対話の状態遷移を図 2 に示す。対話の状態遷移は応答生成部に該当する。状態遷移の円の中にある発話がエージェントの発話であり、円の外にある発話がユーザ発話である。START から、エージェントが“ユーザーへの質問”を行う。一定時間、ユーザの回答がなければ“発話の促し”を行い、ユーザの発話がいずれのマッチングルールにもマッチしなかった場合、“回答例の提案”を行う。ユーザが未知語を発声し、他の単語に誤認識されてルールにマッチしなかった場合も、同様に提案を行う。こ

のようにエージェントが“回答例の提案”を行う事により、ユーザが発話に困らずスムーズな対話を行うことが出来る。

ユーザ発話がマッチングルールにマッチすると、エージェントが“回答に対するコメント”を行い、さらに“発話エージェント交代”を行ってコメントする。最後に START に戻り、エージェントがユーザへ別の質問を行う。この繰り返して対話が進んでいく。誤認識によって、エージェントが別の単語に対してコメントを行っても、ユーザが適当なあいづちを打つことで対話が進むようになっている。これにより途中で対話が途切れることなく、誤認識に頑健な対話を行うことが出来る。また“回答例の提案”や“回答に対するコメント”において、同じようなエージェント発話が続かないように、前述の情報スロットを使用している。

2.3.5 応答タイミング生成部

今回構築したシステムで用いる応答タイミング生成の手法は、我々が先行研究で用いていた手法と同じものである [3]。このシステムでは、ユーザの発話中・ポーズ中に関わらず、全てのセグメント (100ms 毎) に対して、応答するかどうかの判定を行っており、ユーザ発話にオーバーラップする応答を返すことが出来る。今回システムが行う応答の種類は「あいづち・一般的な応答・待ち」の 3 種類とした。

2.4 出力部

出力部では、対話管理部から送られてくる出力結果を、各エージェントから出力する。対話管理部から送られてくる出力結果には、エージェントの発話内容、アニメーション内容の情報が記述されており、それに基づいて映像、音声にて出力する。各エージェントはそれぞれ別々の画面 (PC) に表示される。また、音声も別々のスピーカ (PC) から出力される。以下に詳細を述べる。

2.4.1 エージェントの表示方法

エージェントの表示方法としては、2つの画面に個別に表示する手法を用いる。エージェントの表示には TVML (TV program Making Language) [9] を用いた。表示するエージェントについては、アニメキャラクターのような 3D モデルを用いた (TVML オプションパック内の「abeno(男性)」と「suyama(女性)」)。

2.4.2 音声出力部

音声出力は、音声合成器を用いて行う。「うどんとラーメン」システムの音声合成には、TVML インストールプログラムに含まれている GalateaTalk (擬人化音声対話エージェントのツールキット Galatea Toolkit [10] に含まれる音声合成器) を用いている。また「北海道と沖縄」システムには、同じ HMM 音声合成である OpenJtalk [11] を用いた。今回は、音声合成をリアルタイムで行うことが難しい

ため、あらかじめ応答文の音声波形をファイルとして用意しておいた。

2.5 三者対話システムからの二者対話システムの構築

図1の三者対話システムから、エージェントをひとつ取り除き、二者対話システムを構築した。三者対話システムの2つのエージェントを、1つのエージェントで共有する形となり、対話内容については、矛盾が生じない程度に三者対話システムの内容とほぼ同じとした。エージェントについては、「うどんとラーメン」システムでは、三者対話システムの片方のエージェント(abeno(男性))を用い、「北海道と沖縄」システムではもう一方の女性エージェント(suyama)を用いた。認識文法や語彙は三者対話システムと同じものを使用した。三者対話システムにおいて、発話エージェントが交代してコメントする部分は、1人のエージェントが2回続けてコメントを行うこととした。

3. 被験者実験

3.1 実験内容

開発した二者対話システムと三者対話システムを用いて、被験者対話実験を行った。「うどんとラーメン」システムの被験者を従来の8名から20名に増やし、「北海道と沖縄」システムは12名の被験者が使用した。被験者は全て男性で、両方のドメインのシステムを使用した被験者はいない。まず被験者は始めに対話システムのデモを視聴し、数分程度、システムに慣れるために対話システムを使用した。その後、1名毎に二者対話システムと三者対話システムをそれぞれ5分程度使用してもらい、対話を途中で打ち切ってアンケートに記入をした。アンケート項目については、各被験者是对話前に確認を行い、半分の被験者は二者対話、三者対話システムの使用の順番を入れ替えた。現在の対話システムには対話の終了状態がなく、合図をするまで被験者には対話を続けてもらった。アンケートは次の項目で行われた。

- (1) どちらのシステムが話しやすかったか。(二者(1 2 3 4 5)三者 以下同)
- (2) どちらのシステムの方が、エージェントから色々な意見が聞けたと感じたか。
- (3) どちらのシステムの方が、エージェントの意見に親しみが持てたか。
- (4) どちらのシステムの方が話題に興味は持てたか。
- (5) どちらのシステムの方が、対話は弾んだと感じたか。
- (6) どちらのシステムとの対話が雑談のように感じたか。
- (7) システムの応答内容と応答速度が、人間と同程度に自然だった場合、どちらのシステムを再度使いたいと思うか。

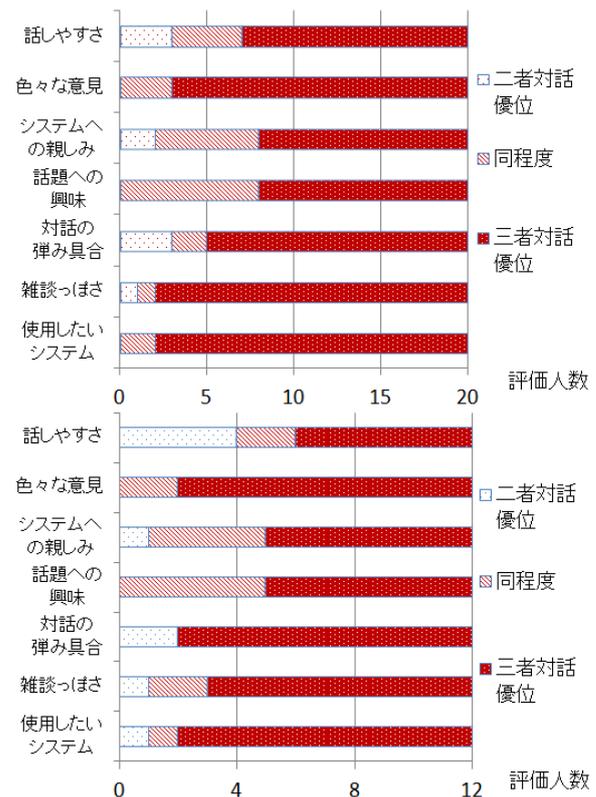


図3 相対評価: 評価に1または2を付けた被験者数を“二者対話優位”として表し、4または5を付けた被験者数を“三者対話優位”として表す。3を付けた被験者数を“同程度”とする。上段が「うどんとラーメン」システム、下段が「北海道と沖縄」システム

3.2 実験結果

3.2.1 主観評価

(a) 相対評価

実験結果として、被験者からのアンケートの結果を図3に示す。質問(2),(5)~(7)については、三者対話システムに高評価が付けられた。「三者対話優位」の評価人数と「二者対話優位+同程度」の評価人数を比較すると、これらの質問項目において有意な差が見られた(両側z検定, $p < 0.05$)。質問(2)について、「うどんとラーメン」システムでは20人中17人が、「北海道と沖縄」システムでは12人中10人が三者対話システムの方が色々な意見が聞けたと回答し、自由筆記形式の回答を参照すると、「3人で会話しているので、より多くの意見が聞けたように感じたから」などが挙げられていた。質問(5)については、「2者の場合より、3者の方が話題が豊富だったように感じた」などが挙がっていた(実際の話題の豊富さは両システムで同じ)。また質問(6)については、「うどんとラーメン」システムで20人中18人が、「北海道と沖縄」システムで12人中9人が、三者対話システムとの対話がより雑談のように感じたと答えた。「2人のエージェント同士が話しているときに(雑談のように)強く感じた」、「(二者対話システムだと)事務的な雰囲気になってしまうから」などが回答として挙

がっており、対話エージェントを2つにすることで、ユーザがより自然な対話を行うことが出来たと考えられる。質問(7)については、8割以上の被験者が三者対話に高評価を示し、「色んな相談やひまつぶしになりそう」「違う意見を持っている方が対話が盛り上がる」などが挙げられた。

逆に、質問(3),(4)では有意な差は現れず、質問(1),(5)では、二者対話システムと三者対話システムで評価が分かれた。三者対話システムに高評価をつけた被験者は、「(三者対話システムは)ポンポン言葉が飛んでくる感じがしたから」、「3人で話した方がテンポよく感じたから」と回答している。二者対話システムに高評価をつけた被験者は、「三者対話でも、結局1対1でしゃべるので、2者で十分」、「(三者対話は)2画面を視線が行き来するので大変だった」、などと回答した。これらは人間同士の対話でも、多人数対話となると発話のタイミング(主導権の移動など)が難しくなることから、ある程度予想できる回答である。質問(1)で二者対話システムに高評価をつけた被験者2人は、質問(5)においても二者対話システムに高評価を付けている。前者については、今回の三者対話システムでは、エージェント同士の対話のごく限られているからであり、対話シナリオの拡充などによって、エージェント間の対話を活発に行う必要がある。後者については、現在のシステムではエージェント間の発話タイミングに固定値を用いているためである。これについては、対話全体のリズムを制御する必要がある[3]。

他の自由筆記の回答として、「三者対話の場合、音声認識誤りがあっても、あまり違和感がなかった(ストレスがなかった)」などがあった。これについては、対話エージェントが交代することで、ユーザの音声認識誤りのストレスを軽減させているのではないかと考えられる。被験者からのシステムの改善案としては、「エージェントに動き(ジェスチャーなど)があればいいと思う」、「(好きな観光地や食べ物について)1つの話題でもっと話せるようにしてほしい」などが挙げられた。

(b) 絶対評価

上記の相対評価に加え、被験者は、(1)~(6)の質問で二者対話システム、三者対話システムをそれぞれ絶対評価した。評価は例として、「(1)対話システムは話しやすかったか」に対して「そう思わない(1~5)そう思う」のような形で5段階評価で行った。絶対評価の結果を図4に示す。「うどんとラーメン」システムでは、「話しやすさ」や「色々な意見」、「対話の弾み具合」、「雑談っぽさ」の項目で、三者対話システムは二者対話システムより高評価となっている(両側t検定, $p < 0.05$, 「話しやすさ」のみ $p < 0.1$)。「北海道と沖縄」システムでは、同じく「色々な意見」、「雑談っぽさ」の項目で、三者対話システムが高評価となっている(両側t検定, $p < 0.05$)。またこれらの質問とは別に、「エージェントの応答内容はどうだったか。

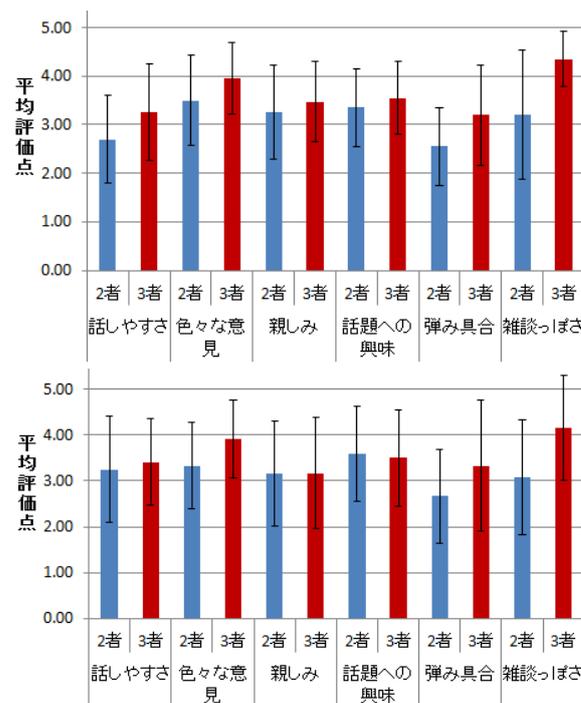


図4 絶対評価: 質問に対する評価値の平均値と標準偏差。上段が「うどんとラーメン」システム, 下段が「北海道と沖縄」システム

不自然(1~5)自然”という質問をしたところ、「うどんとラーメン」システムでは平均で2.5点、「北海道と沖縄」システムでは3.5点だった。「北海道と沖縄」システムでは、情報スロットを用いて対話エージェントが同じような発話をするのを避けたため、より自然な対話が出来たものと思われる。

注目すべき結果として、「雑談っぽさ」の項目では、相対評価と同じく絶対評価でも、三者対話システムが非常に有意な評価を得られており、三者対話システムは我々の目標通りに、ユーザに雑談対話の印象を与えていることがわかる。

3.2.2 客観評価

客観的な実験結果として、「うどんとラーメン」システムを使用した被験者の音声認識率(単語正解率: Cor)と未知語率(OOV), 対話現象頻度を表1と2に示す。上から平均の音声認識率が高い話者順に並んでおり、図1は20名のうちの上位4名と下位4名、また特徴的な結果の被験者を示す。当然ながらOOVが高い被験者ほど、音声認識率が低く、大きな負の相関がある。各話者で対話時間に差があるのは、5分程度の切りの良いところで対話を打ち切る際に、対話があまりスムーズに進んでいない被験者は、その分だけ長く対話を行ったためである。同じ話者で二者対話と三者対話の対話時間はほぼ同じ長さとなっている。音声認識率とユーザターン数に高い負の相関があり、これは被験者が誤認識の時に言い直しを行ったり、対話があまり進んでいない場合に対話時間が長くなったりしたためと考

表 1 音声認識率 (Cor) と対話現象頻度 (「うどんとラーメン」システム)

| 話者 | 音声認識率 [%] | | OOV [%] | | 対話時間 | | ユーザターン数 | | システムターン数 | |
|-----------|-------------|-------------|------------|-------------|-------|-------|---------|-------|-----------|-----------|
| | 二者 | 三者 | 二者 | 三者 | 二者 | 三者 | 二者 | 三者 | 二者 | 三者 |
| 1 | 72.5 | 82.8 | 4.5 | 0.0 | 4'21" | 4'14" | 38 | 35 | 62 | 52 |
| 2 | 73.5 | 81.3 | 2.9 | 4.5 | 4'18" | 4'50" | 44 | 46 | 59 | 63 |
| 3 | 80.7 | 73.6 | 2.1 | 4.9 | 4'23" | 4'32" | 44 | 45 | 62 | 60 |
| 4 | 70.4 | 76.2 | 2.4 | 5.4 | 4'48" | 5'03" | 66 | 52 | 79 | 72 |
| 7 | 67.6 | 63.8 | 3.7 | 1.7 | 4'57" | 4'44" | 34 | 32 | 55 | 77 |
| 17 | 49.0 | 54.1 | 2.1 | 1.3 | 5'42" | 6'00" | 49 | 52 | 70 | 76 |
| 18 | 49.4 | 44.0 | 10.0 | 9.5 | 5'11" | 5'30" | 66 | 66 | 82 | 78 |
| 19 | 45.3 | 44.1 | 10.3 | 7.1 | 4'43" | 4'48" | 59 | 56 | 81 | 77 |
| 20 | 55.4 | 27.9 | 7.7 | 17.7 | 5'58" | 5'50" | 48 | 48 | 67 | 63 |
| 平均 | 62.7 | 61.3 | 4.6 | 6.3 | 4'56" | 5'04" | 50.2 | 48.0 | 70.0 | 69.4 |
| 音声認識率との相関 | | | -0.46 | -0.65 | | | -0.41 | -0.40 | | |

表 2 音声認識率 (Cor) と対話現象頻度 (「北海道と沖縄」システム)

| 話者 | 音声認識率 [%] | | OOV [%] | | 対話時間 | | ユーザターン数 | | システムターン数 | |
|-----------|-------------|-------------|---------|-------|-------|-------|---------|-------|----------|------|
| | 二者 | 三者 | 二者 | 三者 | 二者 | 三者 | 二者 | 三者 | 二者 | 三者 |
| 21 | 57.3 | 71.3 | 1.6 | 0.0 | 5'00" | 4'49" | 45 | 41 | 61 | 58 |
| 22 | 65.3 | 62.6 | 2.0 | 2.2 | 5'04" | 5'48" | 47 | 51 | 65 | 70 |
| 23 | 72.3 | 47.1 | 1.1 | 3.7 | 4'55" | 4'46" | 44 | 45 | 55 | 58 |
| 24 | 61.2 | 56.8 | 3.8 | 2.3 | 5'11" | 4'48" | 43 | 41 | 58 | 45 |
| 25 | 61.9 | 55.9 | 1.8 | 1.0 | 4'19" | 4'46" | 38 | 43 | 48 | 52 |
| 26 | 51.2 | 63.1 | 5.6 | 2.7 | 5'18" | 5'33" | 39 | 40 | 53 | 53 |
| 27 | 56.8 | 52.9 | 0.0 | 4.7 | 5'31" | 5'53" | 49 | 51 | 62 | 70 |
| 28 | 56.7 | 47.3 | 5.0 | 5.8 | 4'48" | 5'14" | 45 | 52 | 76 | 67 |
| 29 | 42.6 | 50.9 | 14.6 | 8.4 | 5'14" | 5'09" | 49 | 50 | 72 | 78 |
| 30 | 43.3 | 49.5 | 1.0 | 2.0 | 5'19" | 5'05" | 56 | 52 | 80 | 70 |
| 31 | 42.8 | 50.0 | 4.9 | 7.5 | 6'18" | 4'59" | 53 | 38 | 79 | 58 |
| 32 | 26.5 | 27.2 | 5.7 | 3.0 | 5'46" | 5'54" | 60 | 60 | 74 | 65 |
| 平均 | 53.1 | 52.9 | 3.9 | 3.6 | 5'14" | 5'14" | 47.3 | 47.0 | 65.3 | 62.0 |
| 音声認識率との相関 | | | -0.49 | -0.38 | | | -0.75 | -0.66 | | |

えられる。システムターン数は、一般的な応答以外にあいづちも含めた数である。

話者 7, 20, 23 は、前述の相対評価において、二者対話の方が話しやすく、対話が弾んだと答えた被験者である。話者 7 は、三者対話でのシステムターン数が二者対話と比べて極めて多く、これによって三者対話システムに話しにくさを感じたようである。また話者 20 と 23 は三者対話での音声認識率が極端に低く、このため話しやすさにおいて三者対話に低評価を付けたと思われる。

また、音声認識率と、図 4 のアンケート評価「雑談っぽさ」の項目において、二者対話システムでは 0.40 (「うどんとラーメン」システム) と 0.51 (「北海道と沖縄」システム) の大きな相関が見られた。一方、三者対話システムでは、相関係数はそれぞれ 0.13 と -0.21 と小さかった。このことから、二者対話システムは、音声認識率が良い場合には雑談対話の印象をユーザに与えるが、三者対話システムは音声認識率に関係なく、雑談対話の印象をユーザに与え

ると言える。

3.2.3 対話ドメインの比較

対話ドメインによる違いに注目すると、アンケートの絶対評価「システムとの対話は雑談のように感じたか。そう思わない(1~5) そう思う」において違いが見られた。「うどんとラーメン」システムでは、三者対話システムに 2 点以下をつけた被験者は 20 人中 1 人もいなかったが、「北海道と沖縄」システムでは、12 人中 2 人が 2 点(雑談のようにやや思わない)をつけた。これは「うどんとラーメン」という話題が被験者にとって身近な話題であり、より雑談的な対話のように感じたためと思われる。「北海道旅行と沖縄旅行」では観光地や特産物の話題が中心であり、知識の獲得を目的とした、やや目的指向対話の印象を持ったと考えられる。実際に、上記の被験者は「北海道と沖縄」システムのアンケートにおいて、「場所についての知識を得るためのシステムのように感じた」と回答した。より雑談らしい対話を行うには、エージェントは情報を提供するだ

けでなく、ユーザに感想を求めたり、ユーザの意見に積極的に同調したり、またユーザの視点に立って観光地についてわざと知らないふりをするなどの工夫が必要である。

4. まとめ

本稿では、1 ユーザ対 2 システムエージェントによる三者対話が可能な音声対話システムの開発を行い、二者対話と三者対話についてユーザに与える印象・満足度について調査を行った。本対話システムでは、ユーザの嗜好(「うどんとラーメン」または「沖縄旅行と北海道旅行」)についての話題を通して、ユーザを対話システムに引き込む戦略をとっている。対話シナリオとしては、二者対話システムのエージェントには、ユーザにうどん(北海道旅行)とラーメン(沖縄旅行)両方を薦めさせ、三者対話システムのエージェントには、それぞれのエージェントが好きなものをユーザに薦めさせた。被験者実験の結果、三者対話システムは、“色々な意見”、“対話の弾み具合”、“対話の雑談らしさ”の印象を被験者に与えることが示された。

今後の発展として、エージェントの音声合成を人間の録音音声に変更した場合の、ユーザの印象を分析することがある。また今回は、三者対話システムのエージェントが対立関係となるように対話シナリオを作成したが、協調関係や上下関係としたときの調査も必要である。協調関係の場合、2人のエージェントが協力して、1つの物・事柄を薦めるといったシナリオもありうる。他に、三者対話システムにおいて、さらにエージェント同士の対話を活発にした場合の、三者対話システム同士の比較も考えられる。

参考文献

- [1] 大村祐司, 川端豪: 雑談可能な目的達成型音声対話システム, 情報処理学会 音声言語情報処理 (SLP) 研究報告, Vol. 2012-SLP-94, No. 9 (2012).
- [2] 尾崎 健太郎ら: 複数の車内機器操作と雑談を扱えるマルチタスク音声対話システムのユーザビリティの向上, 情報処理学会 音声言語情報処理 (SLP) 研究報告, Vol. 2010-SLP-80, No. 6 (2010).
- [3] 西村良太, 中川聖一: 応答タイミングを考慮した音声対話システムとその評価, 情報処理学会 音声言語情報処理 (SLP) 研究報告, Vol. 2009-SLP-77, No. 22 (2009).
- [4] 藤堂祐樹, 中川聖一ら: 単一对話エージェントと複数対話エージェントを用いた音声対話システムの分析と評価, 情報処理学会 音声言語情報処理 (SLP) 研究報告, Vol. 2012-SLP-90, No. 19 (2012).
- [5] Dielmann: DBN Based Joint Dialogue Act Recognition of Multiparty Meetings, Proceedings of ICASSP '07, pp. 133-136 (2007).
- [6] S. Fujie and T. Kobayashi et al, “Conversation Robot Participating in and Activating a Group Communication”, Proceedings of the Interspeech 2009, 264-267.
- [7] 浅井亮太, 堂坂浩二: 多人数対話における対話エージェントのコミュニケーション活性効果, 言語処理学会第 15 回年次大会発表論文集 (2009).
- [8] 甲斐充彦, 中川聖一: 日本語連続音声認識システム SPOJUS-SYNO の改良と評価, 電子情報通信学会技術

- 報告, SP93-20(1993).
- [9] <http://www.nhk.or.jp/str1/tvml/>
 - [10] 嵯峨山茂樹: 擬人化音声対話エージェントツールキット Galatea, 情報処理学会 音声言語情報処理研究報告, 2002-SLP-45, No.10 (2003).
 - [11] <http://open-jtalk.sourceforge.net/>