Vol.2012-MUS-94 No.19
Vol.2012-SLP-90 No.19
2012/2/4

単一対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

藤堂祐樹[†1]　西村良太[†2]
山本一公[†1]　中川聖一[†1]

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，本報告ではシステム側のエージェントを 2 つにした三者対話システムの開発を行った．また二者対話システムと三者対話システムをそれぞれユーザに使用してもらい，システムがユーザに与える影響・満足度についての分析を行った．対話ドメインは「うどんとラーメンのどちらが好きか」とし，二者対話システムのエージェントにはユーザにうどんとラーメン両方を薦めさせた．三者対話システムのエージェントには，それぞれうどん好き，ラーメン好きという個性を与え，それぞれ自分の好きな物をユーザに薦める対話形式とした．被験者実験の結果，三者対話システムは，エージェントへの親しみや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．

Analysis and Evaluation of Spoken Dialog System with One Agent and Multiple Agents

Yuki Todo,[†1] Ryota Nishimura,[†2]
Kazumasa Yamamoto[†1] and Seiichi Nakagawa[†1]

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, which treated the same dialog task "Which do you prefer *udon* and *ramen* ? ", and compared user's behavior/satisfaction.As a result of the experiments, the three parson's dialog system achieved better results in familiarity and frankness.

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の開発も行われてきている．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[1)]．このシステムでは，応答として，あいづち，復唱，共同補完などを扱っており，決定木を用いて応答種類と応答タイミングを決定している．また，このシステムは，ユーザからのオーバーラップ発話（バージイン）やユーザからの非流暢な発話に対しても頑健に応答することが可能になっている．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[2)]．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，Dielmann ら[3)] 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．Ginzburg ら[4)] は，二話者対話プロトコルを，多人数対話にスケールアップする方法についての研究を行っている．多人数対話では，質問に対する応答発話や確認発話などが，二者対話に比べて遠い距離で（3 発話以上あとに）現れる場合が多くある．これに対応する為に，スタックを用いた対話処理を行っている．

浅井ら[5)] 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

†1 豊橋技術科学大学 情報知能工学系
Department of Computer Sciences and Engineering, Toyohashi University of Technology
†2 名古屋大学大学院 工学研究科 電子情報システム専攻
Department of Electrical Engineering and Computer Science, Nagoya University

Vol.2012-MUS-94 No.19
Vol.2012-SLP-90 No.19
2012/2/4

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．しかし，浅井らの実験はテキストベースのシステムで行われており，音声対話システムでの効果は分からない．

岡本ら[6] は，複数エージェント対話システムを構築する際の，エージェント同士の自然な対話を実現するために，どのような非言語動作をどの時点で取るべきかを明らかにしようとしている．分析には漫才を用いている．この理由としては身体動作への制約が最小限であり，対話のみで情報伝達が行われているからである．分析の結果，対話全体として，エージェントの視線が相方，姿勢が観客である場合が多かった．動作に制約がない漫才においても，観客への姿勢配分が大きくなることから，姿勢（ポスチャ）に注目する必要性がある．

岡本らの指摘からは，エージェントの表示と，姿勢・視線の制御が必要であることが示されている為，複数エージェントの対話システムを構築する際には，この条件を満たすエージェント表示部も必要になる．

これらのことをふまえ，我々は，複数の対話エージェントを扱う音声対話システムの開発を行ってきた[2]．本報告では，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[2]．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，音声認識した結果から，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．

2.1 対話ドメイン

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ものが好ましい．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．今回は,「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．二者対話システムでは，1 人のエージェントが麺類好きとし，うどんと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うどん好き，ラーメン好きとし，うどんとラーメンについてそれぞれ良い点・悪い点を示して対話を進めていく．

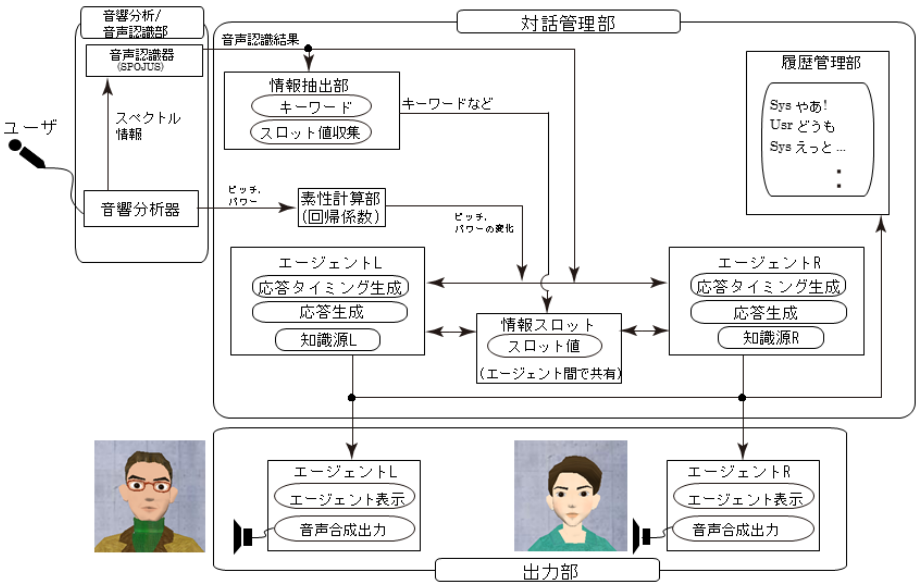

図 1 三者対話システムの概略図

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[7),8)] を用いる．SPOJUS には，2 つのバージョンがあり，1 つは n-gram を用いた大語彙連続音声認識用のもの，もう一つは CFG (Context Free Grammar) を用いたものがあり，今回は，CFG 版の SPOJUS を用いている．

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．

Vol.2012-MUS-94 No.19
Vol.2012-SLP-90 No.19
2012/2/4

2.3 対話管理部

対話管理部は，以下に示すサブコンポーネントから構成されている．

2.3.1 素性計算部[1)]

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．また，名前やエージェントの一人称などを保持しておくことで，応答テンプレートの汎用性を高めている．

今回は，対話ドメインが「うどんとラーメンについての話」であることから，スロットの例としては,「ユーザが好きなもの」「その食べ物が好きな理由」「いま話している食べ物」など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．

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．また，応答戦略として，サブタスク（サブシナリオ）を定義することで，文脈を考慮した対話が可能になっている．以下に，三者対話システムのテンプレートの例を示す．

```
[first prompt]
@（.*)
うどんとラーメンだったらどっちが好き？ initiate:L,subtask:1,sentence:1
[topic]
@（ラーメン）
＝ subtask:1,sentence:1,initiate:L;
   僕もラーメンが好きです．どんな種類のラーメンが好きなの？;
   sentence:2,nowTopic:ラーメン,likeU:ラーメン;
@（しょうゆラーメン）
＝ subtask:1,sentence:2,initiate:L; しょうゆラーメンは，基本中の基本だよね．;
   sentence:3,nowFood:しょうゆ,initiate:CHANGE;
@（.*)
＝ subtask:1,sentence:3,nowFood:しょうゆ,initiate:R;
   ラーメンは全体的に，味が濃すぎると思いますね．;
   sentence:4,nowTopic:うどん;
@（.*)
＝ subtask:1,sentence:4,nowTopic:.+;
   $2 もおいしいよ．$2 では何が好きなの？;sentence:2;

@ マッチングルール
＝ スロット条件; 出力文; スロット書き換え; アニメーションコマンド
```

マッチングルールは，正規表現で記述する．1 つのマッチングルールに対して，出力文（「＝」行）はいくつでも記述することができる．その場合，いずれかの出力文がランダムで選択される．また，スロット条件の部分が「subtask:1,sentence:2」となっている場合には，subtask スロットの中身が「1」，sentence スロットの中身が「2」になっていれば条件と一致するとして，出力を行う．スロット書き換えについては，スロット条件と同様の記述法でスロットの値を書き換えることができる．

二者対話の対話例を以下に示す．

システム　：うどんとラーメンだったらどっちが好き？
ユーザ　　：ラーメンが好きです．
システム　：どんな種類のラーメンが好きなの？
ユーザ　　：しょうゆラーメンかな．

Vol.2012-MUS-94 No.19
Vol.2012-SLP-90 No.19
2012/2/4

システム　：しょうゆラーメンは，基本中の基本だよね．
ユーザ　　：そうですね．
システム　：ラーメンは全体的に，味が濃い感じだよね．
ユーザ　　：そうかもしれない．
システム　：うどんもおいしいよ．うどんでは何が好きなの？
ユーザ　　：きつねうどんとか好きかな．
システム　：きつねうどんは，揚げがおいしいよね．

三者対話での対話例を以下に示す．この対話は上記のテンプレートから生成されている．
システム L がラーメンを，システム R がうどんをそれぞれユーザに薦めている．
システムＬ：うどんとラーメンだったらどっちが好き？
ユーザ　　：ラーメンが好きです．
システムＬ：僕もラーメンが好きです．どんな種類のラーメンが好きなの？
ユーザ　　：しょうゆラーメンかな．
システムＬ：しょうゆラーメンは，基本中の基本だよね．
ユーザ　　：そうですね．
システムＲ：ラーメンはちょっと，味が濃すぎると思いますね．
ユーザ　　：そうかもしれない．
システムＲ：うどんもおいしいよ．うどんでは何が好きなの？
ユーザ　　：きつねうどんとか好きかな．
システムＲ：きつねうどんは，揚げがおいしいですよね．

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し，subtask スロット，sentence スロットに格納された値がひとつの状態に対応する．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，ユーザの発話が未知語であるか，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．二者対話の場合は，１人のエージェントが 2 回コメントを行う．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．この繰り返しで対

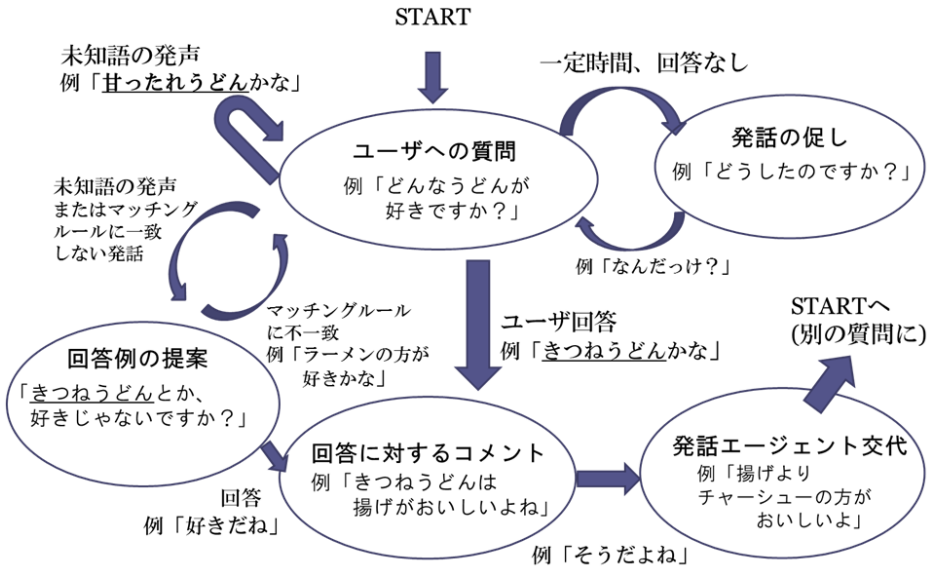

図 2　対話の状態遷移（三者対話）

話が進んでいく．ここでも，“ 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．現時点では，文脈を考慮した応答に情報スロットを使い，履歴管理部については対話に利用していないが，今後は対話履歴の情報を活用し，より文脈を考慮した対話戦略を実現したいと考えている．

2.3.5　応答タイミング生成部[1)]

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[1)]．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．

応答タイミング生成器は，決定木にて韻律素性を用いて応答タイミングを生成する．また同時に，応答生成器にて生成された応答の中から適切な応答を選択する．

決定木では，応答生成器にて応答が準備できているかどうかも素性として用いる．各応答種類毎に一つの応答が準備される．各素性は，100ms 毎に決定木に入力され，応答すべき

かどうかの判定と，応答する場合には適切な応答種類の判定を行う．選択される応答の種類には,「あいづち・復唱・一般的な応答・待ち」がある.「待ち」の場合には，応答を出力しない．応答の回数は，1 回のユーザ発話に対して 1 回のシステム応答に制限されているが，あいづちと復唱に関してはこの制限はない．つまり，1 回のユーザ発話に対して，一般的な応答は 1 回応答することができ，あいづち・復唱は何度も応答することができる．今回は「復唱」については使用せず，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」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

今回は，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．また，エージェントの表示には，NHK 放送技術研究所にて開発された 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)」）．待ち状態の場合には，体が少し揺れたりするなどのアニメーションを行うことも可能になっている．また，音声出力を行っている間は，音声合成器から発話時間を取得し，その時間に合わせて口をパクパクと動かして，喋っていることを表現することもできる．この場合のアニメーションは，厳密なリップシンクではないが，出力音声の大きさに応じて，口を開く大きさが変化するようになっている．なお，現在のエージェントは，いつも発話している相手の方を向くようになっている．エージェント L は，エージェント R が喋ればエージェント R の方を向き，ユーザが喋ればユーザの方を向く．発話しているエージェントは，発話内容に応じて，呼びかける相手の方を向くようになっている．

2.4.2 音声出力部

音声出力は，音声合成器を用いて行う．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．この音声合成器は，発話者タイプ（男女など）の変更や，抑揚・話速を自由に変更できる．本システムでは，対話エージェントを 2 つ扱っており，差別化を図るために，エージェントは，それぞれ男と女のエージェントとしており，出力音声もそれにあわせて変更している．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前述の対話例のように，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，三者対話システムの片方のエージェント (abeno(男性)) を用い，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. 応答文は，不自然にならない程度に三者対話システムと同じとした．

3. 被験者実験

3.1 実 験 内 容

開発した二者，三者対話システムを用いて，被験者対話実験を行った．被験者は 8 名の男性であり，音声関連の研究室の学生である．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，三者対話システムと 5 分程度の対話を行い，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対話前に確認を行い，半分の被験者は使用するシステムの順番を入れ替えた．また，うどんの種類などの登録単語はアンケート用紙に掲載されており，被験者は登録単語を確認しながら対話を行った．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．対話はアンケートは以下の項目で行われた．

(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話題（うどんとラーメンについて）に興味は持てたか．
(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，対話は弾んだと感じたか．
(5) どちらのシステムとの対話が雑談のように感じたか．
(6) 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．
(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．

Vol.2012-MUS-94 No.19
Vol.2012-SLP-90 No.19
2012/2/4

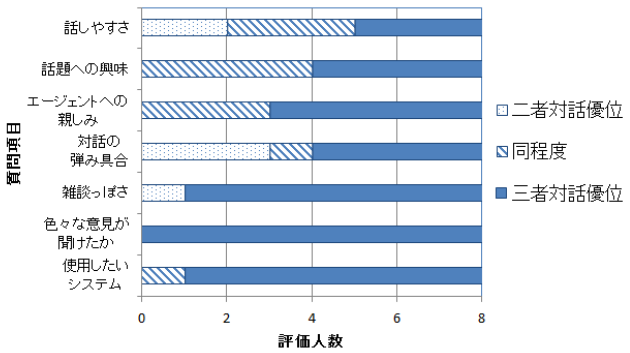

図 3 相対評価：評価に 1 または 2 を付けた被験者数を “ 二者対話優位 ” として表し，4 または 5 を付けた被験者数を “ 三者対話優位 ” として表す．3 を付けた被験者数を “ 同程度 ” とする．

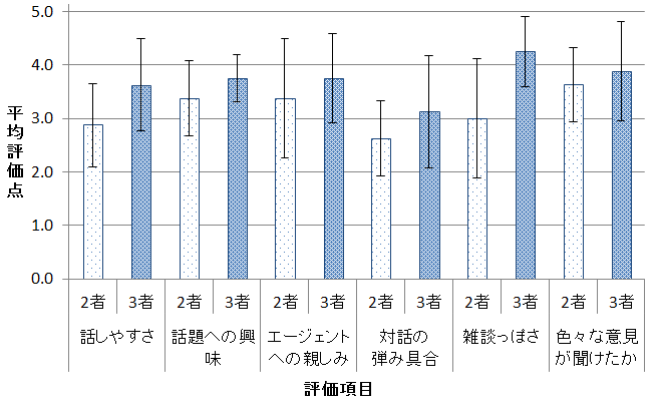

図 4 絶対評価：質問に対する評価値の平均値と標準偏差

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），(3)，(5）～(7)，については，三者対話システムに高評価が付けられている．質問（2）については，8 人中 4 人が三者対話システムの方が話題に興味を持てたと回答し，自由筆記形式の回答を参照すると，「否定的な意見も知ることが出来た」などが挙げられていた．質問（3）については，「エージェント 1 つ 1 つの役割がはっきりしていたから」などが挙がっていた．また質問 (5) については，8 人中 7 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．「二者対話は質疑応答のように感じた」，「(三者対話は）普段している雑談に近い形式だった」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（6）については，被験者全員が三者対話に高評価を示し，「(三者対話は) 否定的な意見も知ることができたため」などが挙げられた．質問（7）については 8 人中 7 人が三者対話システムを使いたいと答えた．

逆に，質問（1），(4）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，「（二者対話システムは）変に身構えてしまった」，「うどん派とラーメン派で意見がぶつかっていたので，目的を持って話せた」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，「(三者対話は）2 人から質問ぜめを受けているように感じた」，「三者対話の場合，エージェント間のやりとりを待ってしまう」，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（4）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，対話の流れで，エージェント同士の対話に繋がらなかったためであり，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．

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，「雑談らしく，(質問→回答のような形式でない）無駄なやりとりがもっと出来ると面白い」，「(三者対話システムは）エージェント同士でもっと会話をさせても面白くなると思う」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

Vol.2012-MUS-94 No.19
Vol.2012-SLP-90 No.19
2012/2/4

表 1 二者対話での音声認識率 (Cor) と対話現象頻度

話者	音声認識率 [%]	OOV[%]	対話時間	ユーザターン数	ユーザオーバーラップ数	システムターン数	ユーザ同一発話回数	システム同一発話回数
話者 1	70.8	3.1	5’17”	56	1	72	10	8
話者 2	68.0	2.1	4’55”	47	2	62	0	4
話者 3	67.6	3.7	4’57”	34	2	**55**	5	3
話者 4	51.6	2.3	5’03”	62	12	73	7	6
話者 5	62.4	0.7	5’27”	55	0	69	14	12
話者 6	49.4	10.0	5’11”	66	6	82	15	16
話者 7	45.3	10.3	4’43”	59	5	81	17	13
話者 8	**55.4**	**7.7**	5’58”	48	0	67	6	6
平均	58.8	5.0	5’11”	53.4	3.5	70.1	9.3	8.5

表 2 三者対話での音声認識率 (Cor) と対話現象頻度

話者	音声認識率 [%]	OOV[%]	対話時間	ユーザターン数	ユーザオーバーラップ数	システムターン数	ユーザ同一発話回数	システム同一発話回数
話者 1	73.9	1.8	5’01”	51	0	67	8	5
話者 2	63.4	6.5	4’59”	40	3	60	3	10
話者 3	63.8	1.7	4’44”	32	5	**77**	9	2
話者 4	62.7	4.5	5’07”	50	12	71	7	6
話者 5	51.6	6.5	5’52”	53	1	69	11	5
話者 6	44.0	9.5	5’30”	66	1	78	10	8
話者 7	44.1	7.1	4’48”	56	3	77	16	11
話者 8	**27.9**	**17.7**	5’50”	48	0	63	9	8
平均	53.9	6.9	5’14”	49.5	3.1	70.3	9.1	6.9

価の結果を図 4 に示す．全ての項目で三者対話システムが高評価を得られているが，特に “ 話しやすさ ”[⋆1]や “ 話題への興味 ”，“ エージェントへの親しみ ” 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．中でも “ 対話が雑談のように感じたか ” 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有意に高い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，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 1,2 に示す．上から，平均の音声認識率が高い話者順に並んでいる．当然ながら OOV が高い被験者ほど，音声認識率が低いことが表からわかる（話者 6，7，8）．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．ユーザオーバーラップ数においては，対話履歴を見て第一著者が確認を行った．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が，ほぼリアルタイムで 5 分間程度，総ターン数 120 回程度対話が続いていることがわかる．

同一発話回数とは，ユーザもしくはシステムが同じ発話を連続して行った回数である．表から，ユーザは全ターン数の約 2 割程度，システムは約 1 割程度，音声誤認識・音声理解誤りのために同じ発話を連続して行っていることがわかる．このことについて，話者 3，5 以外の被験者に “ 苦痛であったかどうか ” を 5 段階評価で回答したもらったところ，話者

⋆1 相対評価では，二者対話と三者対話で評価が分かれた

4，8 は 5(苦痛だった) と答え， 他の 4 人は 4（どちらかと言えば苦痛）と回答した．このことから、（被験者は少なく断定はできないが）同一発話に対する苦痛の程度は，同一発話の出現頻度とは関係が見られなかった．今後，より自然な対話を実現するには，何らかの対策が必要である．

話者 3 と話者 8 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 3 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 8 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，話者 4 は二者対話，三者対話ともにオーバラップ応答が多いが，アンケート結果で特徴的な回答は見られなかった．

また，ユーザターン数の平均，ユーザオーバーラップ数の平均については，若干二者対話システムの値が大きいが，大きな差は見られなかった．また前述のアンケート結果とも関連性は現れなかった．理由として，本システムでは，短いユーザの発話に対してもシステムが割り込んで応答する場合があり，それによって被験者が発話を控えてしまう傾向にあるためと考えられる．これを解決するには，対話の主導権がユーザとシステムのどちらにあるのかを推定する主導権の推定機構が必要である．またシステムの応答テンプレートの不足により，ユーザの発話を促せなかったことも考えられる．

4. ま と め

本報告では，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）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．システムは，ユーザ入力から重要な情報を抽出（スロットフィリング）して，それを応答に組み込み，対話を行うことができる．また，このスロットフィリングを行うことによって，ユーザ入力に対して頑健に応答を返すことが可能になっている．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，ユーザにうどんとラーメン両方を薦めさせ，三者対話システムのエージェントには，それぞれエージェントが好きなうどんとラーメン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，エージェントへの親しみや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が，エージェント間の発話タイミングを制御する必要があるなどの課題も残った．

今後の発展として，システムの対話ドメインを変更することが考えられる．今回は「うどんとラーメンのどちらが好きか」としたが，他に「ブログとツイッター」,「日本料理と中華料理」などの対話ドメインが考えられる．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．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) 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，音声言語情報処理（SLP）研究報告，Vol.2009-SLP-77, No.22 (2009).
2) 西村良太，中川聖一：複数の対話エージェントを扱う音声対話システムの開発，音声言語情報処理（SLP）研究報告，Vol.2010-SLP-080, No.7 (2010).
3) Dielmann, A. and Renals, S.: DBN Based Joint Dialogue Act Recognition of Multiparty Meetings, Proceedings of ICASSP ’07，pp.133–136 (2007).
4) Ginzburg, J. and Fernández, R.: Scaling up from Dialogue to Multilogue: Some Principles and Benchmarks, *Proceedings of the 43rd Annual Meeting of the Association for Computational Linguistics (ACL’05)*, pp.231–238 (2005).
5) 浅井亮太，堂坂浩二，東中竜一郎， 南泰浩，前田英作：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).
6) 岡本雅史，大庭真人，榎本美香， 飯田仁：対話型教示エージェントモデル構築に向けた漫才対話のマルチモーダル分析 (< 特集 > ソーシャルインテリジェンス)，日本知能情報ファジィ学会，Vol.20, No.4, pp.526–539 (2008).
7) 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).
8) Zhang, J., Wang, L. and Nakagawa, S.: LVCSR based on context dependent syllable acoustic models, *Asian Workshop on Speech Science and Technology, SP2007-200*, pp.81–86 (2007).
9) http://www.nhk.or.jp/strl/TVML/.
10) 嵯峨山茂樹，川本真一，下平 博，新田恒雄，西本卓也，中村 哲，伊藤克亘，森島繁生，四倉達夫，甲斐充彦，李 晃伸，山下洋一，小林隆夫，徳田恵一，広瀬啓吉，峯松信明，山田 篤，伝 康晴，宇津呂武仁：擬人化音声対話エージェントツールキット Galatea，情報処理学会研究報告 (2002-SLP-45-10) (2003).