

## 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

西村 良太<sup>†1</sup> 中川 聖一<sup>†1</sup>

現在の音声対話システムは、一人のユーザに対して一つのエージェントが対応する1対1の対話を扱っているが、本論文ではシステム側のエージェントを複数にすることで多人数対話を実現するシステムを構築する。今回は、2つのエージェントを扱った、三者対話システムの開発を行う。本システムは、我々がこれまでに構築した1対1対話の音声対話システムを拡張しており、ユーザに対する応答のタイミングや、応答の種類（あいづちなど）の制御を決定木を用いて行っている。また、ユーザからの非流暢な発話に対しても頑健に応答する事が可能になっている。エージェントは、2Dのキャラクタを、それぞれ2台のディスプレイに1つずつ表示し、出力音声も別々のスピーカから出力されるようになっていく。

### Spoken Dialog System using Multiple Dialog Agents

RYOTA NISHIMURA<sup>†1</sup> and SEIICHI NAKAGAWA<sup>†1</sup>

Almost all present spoken dialog systems have treated dialog that one user talks with one agent. In this paper, to achieve the multiparty conversation (polylogue, many participates conversation), the number of system agents is increased. Three person's conversation system that treats two agents is developed. This system is enhancing the spoken dialog system of two person's conversation that we have developed so far. The response timing to the user and response type are controlled by using the decision tree. The system also reacts robustly to the user's disfluencies. The agent is displayed by the 2D character on two displays respectively one by one. The speech outputs are also output from two different loud-speakers.

<sup>†1</sup> 豊橋技術科学大学 情報工学系

Department of Information and Computer Sciences, Toyohashi University of Technology

### 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対話現象を模倣する音声対話システムを構築した<sup>1)</sup>。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。

その他にも、東中ら<sup>2)</sup>は、テキストベースの対話システムにて、self-disclosure(自己開示)、empathy(共感・感情移入)についての分析を行っており、収集したコーパスにて、[self-disclosure/empathy]と[closeness(親近感)/satisfaction(満足感)]との間の相関を調査している。タスクは「動物の好きな所・嫌いな所」である。分析の結果、ユーザの共感発話は、親近感や満足感を表しており、システムの共感発話は、ユーザの共感発話につながっていた。つまり、システムの共感発話を増加させることで、ユーザの満足度を上げることができる。また、ユーザからの自己開示については、ユーザがトピックに好感を持っているときには自己開示が多くなっている。また、松阪らは、複数ユーザと一人のエージェント（ロボット）との多人数対話を扱い、エージェントの視線の制御、うなずきなどの機能を導入し、自然な対話システムの構築をしている<sup>3)</sup>。Zhengらは、コンピュータ上の仮想空間（美術館）にて、複数のシステムエージェントとユーザとの対話環境を構築している。ユーザも仮想空間上に表示されており、グループの行動などをモデル化して、多人数対話を実現させている<sup>4)</sup>。

このように、対話システムにおいては、ユーザ満足度を上げ、システムに対して親近感を持たせることが重要である。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する。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。また、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も期待できる。

浅井ら<sup>5)</sup>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た。

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<sup>6)</sup>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。そして、エージェント同士の対話によって、ユーザに情報提供をするシステムの開発を目指している。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。動作に制約が無い漫才においても、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。視線が「誰と話すか」を表し、姿勢が「誰に伝えるか」を表しているのではないかとこの仮説を立てている。

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。

これらのことから、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う。ユーザを対話に引き込み、ユーザの満足を得られる対話システムの構築を目指す。

**2. 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」に拡張する。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ている。また、韻律素性を決定木

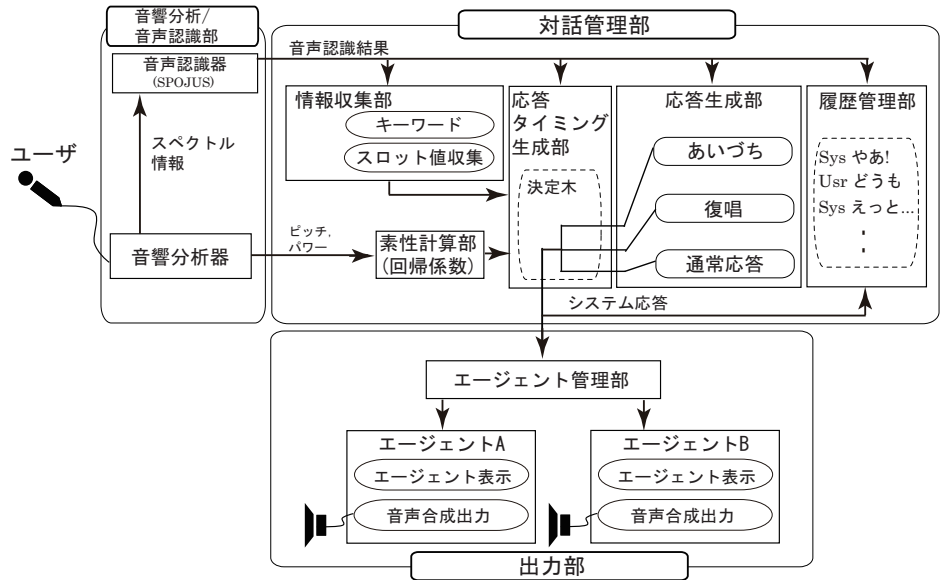


図1 三者対話システムの概略図

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。詳細については、以下の節で述べる。

**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ている。

二人のエージェントが、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。この際、エージェント間で意見を対立させ、ユーザをどちらかの意見に引き込むことや、エージェント間の意見を揃えて、ユーザを特定の意見に引き込むといった戦略も考えられる。

好き嫌いデータベースは Web から収集する予定であるが、現段階では人手で作成している。

## 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<sup>7),8)</sup> を用いる。SPOJUS には、2 つのバージョンがあり、一つは 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。

SPOJUS は、音響特徴量として 12 次元の MFCC (Mel-Frequency Cepstrum Coefficients) とその 1 次・2 次微分である  $\Delta$ MFCC と  $\Delta\Delta$ MFCC、 $\Delta$  パワー、 $\Delta\Delta$  パワーを用いており、音声のサンプリング周波数は 16kHz、分析窓はハミング窓であり、フレーム長は 25ms、フレームシフトは 10ms である。音響モデルとしては HMM を用いており、4 状態、5 ループ、各状態には 4 混合のガウス混合モデルを用いており、全共分散行列を用いている。また、文脈依存型の音節 HMM モデルを用いており、モデル数は 928 モデルである。認識の途中結果もリアルタイムに出力させることが可能であり、システムではその結果を用いて復唱応答の準備などを行うことが可能となっている。

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。

## 2.3 対話管理部

対話管理部は、5 つのサブコンポーネントから構成されており、音声認識結果と韻律情報を用いて応答を生成する。1 つのコンポーネントである応答タイミング生成部では、決定木を用いており、韻律情報から得られる素性に基づいて応答タイミングの決定を行っている。ユーザ発話中のピッチ・パワーの概形を素性として用いており、これらは、F0 と対数パワーの回帰係数として求められる。

音声認識器 SPOJUS による音声認識結果は、情報収集部へと送られ、情報スロットに必要な情報を格納する。情報スロットに格納された情報は、応答生成部へと送られ、応答生成に用いられる。応答生成部では、音声認識結果と情報スロットを用い、ELIZA<sup>9)</sup> 型のテンプレートによる手法によって応答が生成される。そして、システムで扱う応答現象 (あいづち、復唱、一般的な応答) の為の応答を、同時に平行して用意し、用意された応答の中から適切な応答を、決定木によってリアルタイムに適切なタイミングで選択する。選択された応答は、出力部へと送られ、エージェントにより応答としてユーザに提示される。

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とになる。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ようにしている。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。

### 2.3.1 情報収集部

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になっている。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。システムエージェントのそれぞれのキャラクターの名前や、ユーザの名前を格納するスロットが用意されており、これを活用することで、名前前でユーザやエージェントに呼びかけることが可能になる。対話の中でユーザの名前を聞き出して、対話に盛り込むことも可能になる。ただし、音声合成器にてリアルタイムに応答を出力する場合は良いが、合成による遅延時間を解消する為に、予め音声ファイルを出力しておく場合には、ユーザの名前に当たる部分は、「君」などに固定しておく必要がある。

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」「もう一方の食べ物が嫌いな理由」などの情報を認識結果から抽出し、対話を行う。

### 2.3.2 素性計算部<sup>1)</sup>

ここでは、音響分析器から得られた音響分析結果を元に、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韻律素性を計算している。

システムでの応答タイミング生成に用いる素性としては、発話の最後 100ms 部分と 500ms 部分のピッチ・パワーの一次線形回帰係数などである。対象となる 100ms の区間を、3 つの区間に分け (窓幅は 55ms、フレームシフトは 25ms)、その区間での傾きを計算し、素性として用いる。また、発話の最後の 500ms の区間に関しても、同様にピッチ・パワーの傾きを計算し、素性として用いており、この場合には、窓幅 100ms、フレームシフト 100ms の 5 つの区間に分割している。これらの素性については、計算コストが低く、リアルタイムで計算することが可能である為、今回のシステムのような、リアルタイム性が重要となる場合においては適した素性である。

### 2.3.3 応答生成部

本システムでの応答生成には、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、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、テンプレートの例を示す。

```
[first prompt]
@ (.*)
= ; うどんとラーメンどっちが好き?;initiate:L,subtask:1,sentence:1;
[topic]
@ (food)
= subtask:1,sentence:1;A+なんで?;sentence:2;
= subtask:1,sentence:3;A+そっかー。;sentence:4;nod
@ (prefer)
= subtask:1,nameU:.,ownL:.,ownR:.,sentence:2;A+そっかー。$2 もうどんが好きだよ。B+$3 はラーメン派だ。$1 はどう?;sentence:3,initiate:CHANGE;
@ (うん)
= subtask:1,sentence:4;A+うどんとラーメンどっちが好き?;sentence:1;
```

記述方法は、以下の通りである。

```
@ 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。一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。また、スロット条件の部分は「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、出力を行うというものである。出力文については、最初の記号は、エージェント間での主導権を表し、「A（主導権がある側）」「B（主導権が無い側）」を示す。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。アニメーションコマンドは、「nod」でうなずきを行うなど、エージェントの動作を記述することができる。

上記の例から生成される対話例を以下に示す。

システムL：うどんとラーメンどっちが好き？  
ユーザ：えっと、やっぱりうどんかな。

システムL：なんで？  
ユーザ：美味しいからだよ。  
システムL：そっかー。僕もうどんが好きだよ。  
システムR：私はラーメン派だ。君はどう？  
ユーザ：だから、僕はうどんが好きだよ。  
システムR：そっかー。  
ユーザ：うん。

マッチングルールにマッチすると、「=」行のスロット条件が判定され、条件を満たしている場合には応答文を出力し、スロット値を書き換え、エージェントアニメーションをさせる。上記のマッチ文の例は、1つのサブタスクを示している。最初にシステムを起動すると、[first prompt]の中から、文を選択し出力する。今回の場合は、システムから「うどんとラーメンどっちが好き？」と発話し、その際に、右に記した3つのスロットの値を書き換えている。次に、ユーザの発話がなされたあと、その発話とのマッチングをとる。[topic]では、ユーザ発話内容と、サブタスク番号、サブタスク内の文番号によって、シナリオの展開を記述している。

このようにすることで、システム同士の掛け合いが可能となる。

現在のシステム同士の対話の発話タイミングは、相手のエージェントの発話が終了した直後になっている。ユーザ入力に対するエージェントからの応答については、決定木でタイミングを決定しているため、将来的にはエージェント間の対話のタイミングの制御も行いたいと考えている。

### 2.3.4 応答タイミング生成部<sup>1)</sup>

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<sup>1)</sup>。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できる。

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。応答タイミングの決定に関しては、以下の素性を用いている。

- ユーザ発話開始時点からの経過時間 (F1)
- ユーザ発話終了時点からの経過時間 (F2)
- システム発話終了時点からの経過時間 (F3)

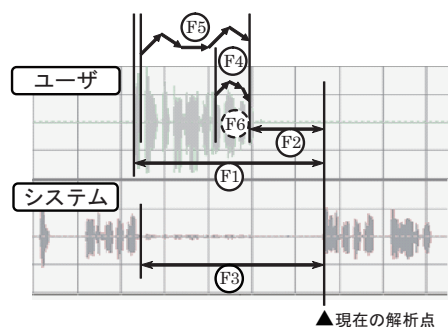


図 2 決定木で用いる素性

- ユーザ発話の最後の 100ms 区間のピッチ・パワーの傾き (F4)
- ユーザ発話の最後の 500ms 区間のピッチ・パワーの傾き (F5)
- 認識結果 (途中結果も含む) の最終単語の属性 (F6)

図 2 に、各素性の関係を図示する。

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、共同補完と一般的な応答は 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。

決定木の学習には、RWC コーパス<sup>10)</sup>を用いており、このコーパスには、対話ドメインとして「自動車販売」「海外旅行計画」の 2 種類の対話が合計 48 対話含収録されている。対話は二者対話である。データ量としては、コーパス全体で 6.5 時間分あり、各対話の時間は 10 分程度である。また、発話数は、16,399 発話である。一方の話者は、実際の自動車販売員、または旅行代理店員であり、もう一方の話者は、12 人の一般人である。このコーパスを用いて決定木の学習を行った。決定木の学習器には、C4.5<sup>11)</sup>を用いた。

### 2.3.5 履歴管理部

対話履歴を保存しておき、後に参照して文脈を考慮した対話戦略を実現する為の部分である。この部分については現段階で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する対話を行いたいと考えている。

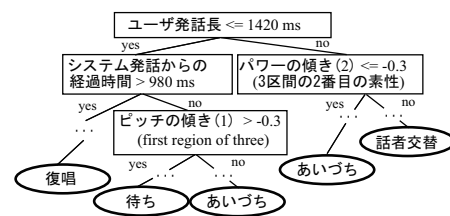


図 3 決定木の一部

## 2.4 出力部

出力部では、各エージェントが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 エージェント管理部

エージェント管理部では、対話管理部から送られてくる出力結果を、各エージェントに振り分け、各エージェントを制御する作業を行っている。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、どちらのエージェントに向けての出力か (エージェントとしての主導権がある側、無い側) の情報が記述されており、それを元に、情報を振り分ける。

### 2.4.2 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、表示するエージェントについては、リアルな人間に近い 3D のものではなく、2D の絵で表示を行う。頭と体の部分を別々に管理しており、それぞれがアニメーション可能となっている。アニメーションは、定義ファイルにて設定を行い、特定のコマンド (うなずきなら nod) に応じて、登録されているアニメーションが表示されるようになっている。待ち状態の場合には、瞬き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、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く、単純に口を開閉する動作のみを行う。

岡元ら<sup>6)</sup>は、エージェントを制御する際のポイントとして、以下のことを挙げている。

- ユーザに明示的に語りかける場面においては非言語チャネルの指向性を同調させる (内部指向性)
- 対話相手のエージェントに向けた発話では基本的に視線を同調させ (対話相手へ向け) つつ姿勢をユーザに向ける (外部指向性)

この指摘から、複数エージェントによる対話では、エージェントの視線と姿勢の制御が重要であり、これらを表示・制御できるものが好ましい。頭部のみを表示するもの (Galatea Toolkit に含まれる顔合成ソフトなど) では、体が表示されていないため、姿勢の制御が出来ない。これらのことから、今回は、本システムの目的に合ったエージェント表示部の実装を行った。

アニメーションコマンドは、「lip:1000 nod:1000」のように記述し、この場合には、1000 ミリ秒間口パクアニメーションを行い、その後でうなずきを 1000 ミリ秒繰り返すというも

のである。アニメーション途中であっても、stop コマンドを送信することで、アニメーションを停止させることができる。その他には、「首を横に振る、もう一方のエージェント側を向く」というコマンドが用意されている。

現在は、実装されていないが、表情機能として、喜び・笑い/怒り・悲しみなどのような表情を出力できるようにしたいと考えている。

#### 2.4.3 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、擬人化音声対話エージェントのツールキット Galatea Toolkit<sup>12)</sup> に含まれる音声合成器 GalateaTalk 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。今回は、エージェントとして男女の2種類を用意している為、性別により音声を変更している。

ここで、音声合成には、500ms 程度のタイムラグが発声する為、システムで用いる際には、想定されるシステム応答を予め全て音声ファイルに出力しておき、それを再生することで音声応答出力を行う。

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、出力音声もそれに合わせて、男女両方のものを用意した。

### 3. ま と め

本論文では、これまでに我々が開発してきた二者音声対話システムを拡張した、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。ユーザ入力に対するシステム応答については、応答タイミングを決定木にて制御しており、適切なタイミングであいづちを返すことも可能になっている。エージェント表示に関しては、2D のアニメーションにより、うなずきや口パクなどを表現するシステムを実装し、これを2台のディスプレイに表示させることで、それぞれのエージェントを別なものとしてユーザが識別できるようにした。また、出力音声についても、別々のスピーカ（PC）から出力している。

本システムでは、ユーザとシステムとの間の対話においては、応答タイミングが考慮され

ているが、システム同士の掛け合いのタイミングについては考慮されていない。今後は、全体的なリズムの制御も行いたいと考えている。また、このシステムを用いた被験者実験を行い、ユーザ満足度や、ユーザの引き込まれ度など、三者対話の有効性について、調査・分析を行いたい。

### 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) Higashinaka, R., Dohsaka, K. and Isozaki, H.: Effects of Self-Disclosure and Empathy in Human-Computer Dialogue, 2008 IEEE Workshop on Spoken Language Technology (SLT 2008), pp.109-112 (2008).
- 3) 松坂要佐, 東條剛史, 小林哲則: グループ会話に参加する対話ロボットの構築, 電子情報通信学会論文誌, Vol.J84-D-II, No.6, pp.898-908 (2001).
- 4) Jun, Z., Xiang, Y. and San, C.Y.: Designing multiparty interaction support in Elva, an embodied tour guide, AAMAS '05: Proceedings of the Fourth International Joint Conference on Autonomous Agents and Multiagent Systems, pp.929-936 (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) Weizenbaum, J.: ELIZA - A computer program for the study of natural language communication between man and machine, Communications of the ACM, Vol.9, No.1, pp.36-45 (1965).
- 10) 田中和世, 速水 悟, 山下洋一, 鹿野清宏, 板橋秀一, 岡 隆一: RWC 計画における音声対話データベースの構築, 情報処理学会音声言語情報処理 1 1-7 (1996).
- 11) J.Quinlan, R.: C4.5: Programs for machine learning, Morgan Kaufmann (1992).
- 12) 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).