擬人化エージェントとの円滑なマルチモーダル対話のための強化学習を用いた割り込み制御の検討

Barge-in Control with Reinforcement Learning for Efficient Multi-modal Spoken Dialogue Agent

盧 迪 [i]　中沢 正幸 [1]　西本 卓也 [1]　嵯峨山 茂樹 [1]

Di Lu[1], Masayuki Nakazawa[1], Takuya Nishimoto[1], Shigeki Sagayama[1]

[1] 東京大学大学院 情報理工学系研究科

[1]Graduate School of Information Science and Technology, the University of Tokyo

Abstract: To make the dialogue between the agent and the user smoother, we propose a multi-modal user simulator that could be widely used in real-time agent control for multi-modal dialog agent with reinforcement learning. We also implemented the prototype system that utilized the result of reinforcement learning.

1. はじめに

さまざまな機能を持つWebベースのアプリケーションが広く使われるようになった現在、「電話をかけて相手と話せば済んでいた」仕事の多くがコンピュータの操作に置き換わりつつある。たとえその一部でも音声インタフェースを介して利用できることの意義は大きい。ブラウザのフォームに情報を埋める作業を繰り返していると、もっと効率よく、あるいは、キーボードやマウスに頼らずに操作したい、と感じるのではないか。日常的操作の多くをキーボードに依存している現状はユニバーサルデザインという観点からも好ましくない。機械との音声コミュニケーションが重要である。

この目的のためには、インタフェースシステムを統合技術と捉えつつ、音声入出力の特質を踏まえ、マルチモーダルインタフェースとしての合理的な設計を行わなくてはならない。嵯峨山 [1] によってこのような問題提起がなされ、西本 [2] はマルチモーダルシステムのための「インタフェースの原則」を提案し、使いやすい音声インタフェースを試行錯誤に頼らず合理的に設計できることを示した。

擬人化音声対話エージェントを用いることの意義もまた「コミュニケーションの効率と質を高める」ことである。つまり、人間は相手の表情から反応を読むことができる。一方が話している間にも頷いたり首をかしげたり、聞き取りにくければ直ちに「え？」と聞き返すことができる [3] 。

人間同士のコミュニケーションの「分かっているのか分かっていないのか反応がある人とは会話がしやすい」という特長を生かすことは、音声インタフェースの有効な利用につながる。この問題は「インタフェースの透過性」として音声インタフェースの研究者に広く知られるようになった。

もうひとつの問題は「音声認識の処理速度」である。一般に音声認識アプリケーションは、応答の遅れによって、ユーザに不満を与えたり不安を感じさせたりしている。これに対して、人間の対面コミュニケーションでは、相手が口を開いた瞬間に、あるいは何かを言い終わる前に、言いたいことが相手に伝わってしまうことが多々ある。話者同士の状況、相手の表情や仕草など、人間はさまざまなモダリティからリアルタイムに情報を得ている。

このような検討の末、以下の仮説に至った：

仮説「マルチモーダル情報を常に受け取り、意味のある反応をリアルタイムに行う擬人化音声対話エージェントシステムは、効率的なインタラクション実現のために有効である。」

例えば、発話中の割り込みや聞き返しに対する制御、相槌や頷きの生成や応答などは、こうした仮説を支持する提案となり得る。しかしこのような制御モデルの構築は、個別の対話タスクに依存する複雑な問題である。

我々は「効率的なインタラクションのためのマルチモーダルなやりとり」そのものは単純なユーザや環境のモデルに基づいていると仮定する。そして、対話タスク、対話場面、ユーザなどモデルが与えられれば、リアルタイムのエージェント制御モデルを

機械学習によって構築可能である、という立場に立つ。特に本報告では強化学習を用いた実装について、予備的な検討の結果を報告する。

2. 本研究の位置づけ

2.1. Galatea Project と関連研究

我々は Galatea Project としてカスタマイズ性を考慮した擬人化音声対話ソフトウェアツールキット [4] の研究開発に取り組んできた。すでに音声合成、顔画像合成、音声認識などの要素技術を統合する Galatea Toolkit（図 1）が完成しオープンソースソフトウェアとして公開されている[5] 。

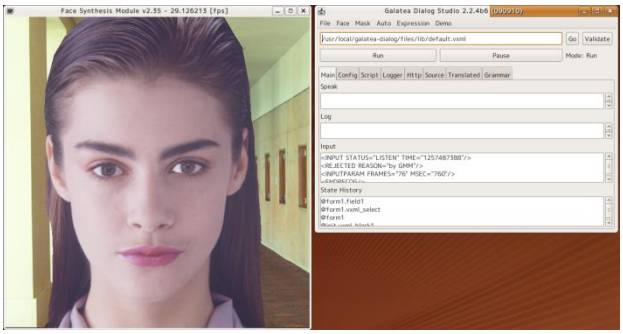

図 1　Galatea Toolkit の画面

擬人化エージェントシステムによって機械に向かって話しかけることへの心理的抵抗をやわらげることに成功している事例もある。しかし、いわゆる「メディアの等式」の議論が示唆するように、人間は機械を擬人化するために、必ずしも人間の姿を必要としない。単に人間の姿をしたインタフェースを実現しただけでは不十分である。

音声対話システムにおける「使いにくさの解消」とは「人間同士のやりとりの様式を模倣する」ことではなく「ユーザを楽にする」ことである。具体的には、少ない所要時間や操作回数で効率よくタスクが達成できること、あるいは操作時の記憶や注意などの心的負荷を軽減できること、などが「楽である」ことの指標となる。

近年はキーボードを 1 回叩くごとに何らかの反応するインクリメンタル検索のインタフェースが普及し、ユーザの支持を得ている。そこで西本 [6] は「インクリメンタル音声検索インタフェース」を試作し、音声入力に対する実時間の反応がインタフェースの透過性を改善することを示した。

擬人化エージェントの身体動作によってインタフェースの透過性を高めることも有望である。西本 [7] は「ユーザ発話開始」「音声認識完了」などのイベントから「ユーザの不安」を推定しこれを解消することを試みた。

しかし、音声認識の途中で得られる情報は断片的かつ不完全である。ユーザ発話の終了を待たずに音声認識エンジンから情報を得る手法も提案されている。ユーザの発話中に頷いたり相槌を打ったりする音声対話システムも提案されてきた。しかし対話内容が限定されていたり、有効性が必ずしも明確ではなかったりするなど、汎用的な制御手法につながっていない。

人間はお互いに相手の顔を見ながら話している。話しながら「合理性、必然性のあるタイミングで相手がうなずいている」ことを確認している。もし相手の反応が不適切であったら自分から言い直しを行うなど、確実で効率的なコミュニケーションを取る工夫をしている。

このような合理性、必然性のあるリアルタイム制御を行うためには、マルチモーダル入力の活用や、タスク知識およびコンテクストの高度な利用が欠かせない。こうした振る舞いは規則によって記述することが困難である。またタスクによって規則を変えなくてはならない。

さらに、このような制御には、人間の性格や好みも影響する。人間にも個性があり、相手の話を聞きながら豊かなリアクションを行う人もいれば、そうでない人もいる。また話し手も、聞き手のリアクションに対する好みを持っている。

対話場面の影響も考慮しなくてはならない。例えば、暗くて相手の顔が見えにくいときは、相手の声による相槌が重要な情報となる。逆に、騒がしくて声が聞こえにくいときは、相手の顔や視線の動きから得られる情報が重要になる。このように対話場面によって「相手の顔」「相手の声」の情報をどの程度信用すべきか、といったことが異なってくる。

2.2. 本研究の着眼点

従来、音声対話システムの制御には状態遷移機械やスロットフィリングなど決定論的なモデルが多く用いられてきた。現在 Galatea Toolkit が採用している VoiceXML もこうした枠組みの技術である。

しかし近年、音声対話システムの制御に、確率モデルに基づく機械学習を用いる提案がある[8] [9] 。隠れマルコフモデル(HMM)、部分観測隠れマルコフ過程(POMDP)、ベイジアンネットワークなどの手法に基づき、ユーザシミュレータと強化学習 [10] を用いるこれらの手法は、「頑健な対話制御の規則は記述することが難しいが、ユーザモデルは比較的簡単に記述できる」という立場を取り、頑健な対話制御を実現している。特に POMDP は、音声認識結果を不完全または部分的な観測として扱えるため、音声対話システムに有効とされる。

我々は実時間のエージェントの動作制御、特にシステムやユーザの発話中に行われる相槌や割り込みなどを生成に同じ枠組みが適用できるのではないか、と考えた。すなわち「この話題のこのタイミングで相手が頷くことは合理的である」「相手の頷き方に矛盾を感じたら、自分から言い直したり、相手の理解の確認をしたりする」といったルールは比較的明確に記述ができる、という立場である。

このユーザモデルに基づいて強化学習を行い、実時間でエージェントが取るべき行動を判断させることができる。強化学習の際には「タスクが遂行できたこと」「タスク遂行に必要な時間や発話数が少なくユーザが楽であったこと」「ユーザが不安を感じたり混乱したりせず、心的負荷が少なかったこと」といった尺度で報酬を与える。

3. 提案手法

3.1 強化学習の概要

強化学習とは教師なしの機械学習の一手法である。ある行為を選択したとき、環境から得られる期待報酬を、すべての状態において実際に探索しながら求めることによって、状態と期待報酬を最大化するように行為をマッピングし、最適な戦略を学習していく。本研究においては、エージェントが環境との相互作用から学習して目標を達成することに相当する。

強化学習には四つの主な構成要素がある。つまり、方策、報酬関数、価値関数、環境のモデルである。

- 方策とは、ある時点での環境の状態 S に応じて、エージェントが行動 A を取ることに相当する。
- 報酬関数とは、ある時点でのエージェントによって観察された環境の状態と、それに基づいて選んだ行動（状態行動対）から得られる報酬である。
  強化学習におけるエージェントの目的は、最終的に受け取る総報酬の最大化である。
  報酬関数はこの状態と行動の二要因に基づいて設計される。報酬関数は各エピソード（本研究では 1 回のタスク遂行における対話）の最後に評価される。タスク達成の所用時間が少なく、ユーザに多く情報を与えることができ、ユーザが感じる不安や混乱が少ない、といった効率的で快適な対話に対して、より多くの報酬を与える。
- 価値関数とは、各エピソードの中で、最終的にどのような対話が行われるべきか、という基準を示す。つまり、ある状態または行為に対して、ある時点から対話の終了までの間に蓄積することを期待する報酬の総量である。
  我々はある状態 s において、行動 a を取ることの価値を$Q(s,a)$と表わし、その期待報酬を

$$Q(s,a) = E\{R_t | s_t = s, a_t = a\}$$

$$= E\left\{\sum_{k=0}^{\infty} \gamma^k r_{t+k+1} | s_t = s, a_t = a\right\}$$

  として定義する。これは一般に行動価値関数（action-value function）と呼ばれる。
- 環境モデルは環境の状態を模倣する。本研究では、マルチモーダル拡張された汎用的なユーザシミュレータの利用を提案する。
  環境モデルは対話タスクの記述や対話場面、ユーザなどから構成される。エージェントとユーザのやりとりに応じて報酬を与える基準も環境モデルの一部である。
  機械と人間の実際の対話を通じて学習がなされる場合は、人間も環境モデルの一部となる。しかし実際にはユーザの一般的な行動選択を模擬するユーザシミュレータが、強化学習においては有用である。
  ユーザシミュレータは、エージェントのある状態と行動を観測した場合に、ユーザの次の行動を決定する。可能な次の状態が複数存在する場合は確率的に選択される。
  視覚と聴覚のいずれの入力を重視して報酬の評価を行うか、といった対話場面の考慮も、この環境モデルに含まれる。

例えばエージェントがユーザに情報を提供する案内型の対話タスクにおいて、この学習の目的は、対話を効率的に終了させることと、ユーザにすべての内容をきちんと伝えることである。

上記の定式化により、エージェントに効率よく対話や動作生成をさせるための学習問題は、エージェントの利得（報酬の総計）を最大化する対話全体の方策を探索する、最適化問題に帰着される。

一般的には、強化学習はマルコフ決定過程として定式化される。しかし実環境中の対話システムにおいては、外界のノイズなどの影響があり、観測状態に不完全性や不確実性があるため、部分観測マルコフ決定過程による定式化がなされる。

3.2 対話エージェントの実時間制御

ここでは基本的なインタラクションとして、説明型の対話タスクにおける、ユーザからシステムに対する割り込みの要求に着目する。

エージェントがユーザに道案内を行うタスクを取り上げる。例えば屋外に置かれた電子案内板に擬人化エージェントの姿が現れて、立ち寄ったユーザと

以下のような会話を行う。

- ユーザ「安田講堂はどこですか？」
- エージェント「まっすぐ行って、2 番目の交差点で左に曲がって右側にあります」

この例の中でユーザは「2 番目の」「左に」「右側に」といったキーワードを聞き漏らす可能性がある。その場合ユーザはエージェントに対して「え？」「何番目？」などと聞き返す発話を行う。このように、ユーザとエージェントは、互いに相手が聞き返しを行う可能性を常に考慮する。

ユーザの顔は例えば画像認識によって処理され、頷きなどが検出可能であるとする。またエージェントは顔画像の表示能力を持っており、首を上下左右に動かして「頷く」「首をかしげる」といった動作を表出でき、ユーザはエージェントの顔表示の変化を知覚できるものとする。

ユーザまたはエージェントが特定の状況やタイミングで頷いたり首をかしげたりした場合に、相手は互いに一般的な規則に基づいて「合理的である」ことを判断できるとする。例えば、ユーザが何か発話を行っている途中や発話を行った直後に、エージェントが頷くことを知覚できた場合は、ユーザは「自分の言いたいことが伝わっている」という合理的な解釈が可能である。

このような前提において、具体的に「エージェントがユーザに長い説明を行い、ユーザは聞き取れなかった指示をエージェントに随時聞き返す状況」に着目する。図 2 に説明タスクの状態遷移モデル例を示す。左から順に「まっすぐ行って」「2 番目の交差点で」といった（聞き返しの単位として予想される）情報の断片が、左から順に状態と対応している。一つの情報の断片を伝えることが状態間の遷移、すなわち動作 A に対応する。

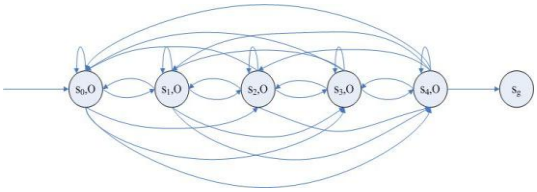

図 2　説明タスクの状態遷移モデル例

状態 S には以下が含まれる。

- 対話状態 S_d（state-dialog）
- ユーザの発話状態 S_s (state-speak)
- ユーザの動作の状態 Sh (state-head)

対話状態 S_d は、エージェントが現在どの情報に関する発話をしているのか、という状態を示す。エージェントの発話がインデックス i の場合には状態は S_{di} となる。

ユーザの発話状態 S_s は、音声認識エンジンが検出可能なユーザの発話状態に対応する。実際の音声対話システムにおいては、ユーザの発話終了を検出した時点で最終的な仮説の探索が開始され、やや遅れて音声認識結果が確定するのが一般的である。これらのステップと S_s が対応している。

エージェントはユーザの割り込みを予測して、できるだけ早く対応することが望まれる。一般的に、音声認識の結果が確定するまでの間に、エージェントにとって使用できる情報は、ユーザが発話しているか否かという状態だけである。

ここでは割り込みや相槌に関するユーザの発話を四種類に分けて、「割り込み」と総称する。この道案内タスクで受理できるユーザ発話を以下のように分類する。

1. 対話を進める割り込み発話例「はい」「うん」
2. 対話を前の状態に戻す割り込み発話例「もう一回」「え？」「さっきの」
3. 対話を止める割り込み発話例「ちょっと待って」「すみません」
4. 対話を先にスキップさせる割り込み発話例「パス」「次は？」
5. 対話を終了する発話「ありがとう」「わかった」

従って S_s は｛黙る、しゃべっている、音声認識の結果１～５｝と定義できる。

ユーザの動きの状態 S_h は、ここでは簡単のため{頷いている、頷いていない｝の２状態とする。

エージェントの動作 A には以下が含まれる。

- 発話動作 A_s（action-speak）
- 頭の動作 A_h（action-head）

エージェントの発話インデックスを i（i は 1～N）とすると、A_{si} は i 番目の情報の断片を発話し始める動作を意味する。As_{N+1} は対話を続ける動作を意味する。As_{N+2} は発話を止める、あるいは黙り続ける動作を意味する。

頭の動き A_h は簡単のために２状態 A_h={頷く、頷かない}と定義する。

3.3 環境モデルの詳細設計

環境モデルはユーザシミュレータ (ユーザモデル) などを含み、ここでは以下の二つの役割を果たしている。

1. ある時刻、エージェントの状態と今回の対話の全体の状態によって、ユーザの動作を決定する。この動作はエージェントが観測できる状態に対応する。ただし POMDP として定式化する場合にはエージェントからの観測は確率的になる。

2. エージェントが観測の状態によって決定した動作に対して、報酬を与える。

環境モデルには、起こりうる対話（エピソード）に関するすべての可能性を生成するための確率モデルを持たせる。

ユーザの発話が発生する確率は前述したユーザ発話分類 1～5 それぞれに対して P(continue), P(repeat), P(wait), P(pass), P(end)と定義される。

実環境では、ユーザがエージェントの発話をすべて聞き取って理解できるとは限らない。ここではユーザはある確率でエージェントの発話を聞き取れるものとする。この確率も P(understand)として定義できる。あるいは、連続時間系のモデルとして、エージェントの発話継続時間に対するユーザの聴取時間の比として、了解度を模擬する方法も考えられる。すなわち 1.0 秒の発話は最初の 0.5 秒間で 50%の了解度をもたらす、といった考え方である。

対話において、ユーザは発話だけではなく、様々な振る舞いをする可能性がある。その中には合理的に解釈ができる振る舞いが含まれる。このようなユーザの振る舞いも環境モデルで生成できる。例えば、ユーザがエージェントの発話を聞き取れた場合には、ある確率 P(nod)でユーザが頷くものとする。

3.4 強化学習のアルゴリズム

本研究の目標である「常に情報を受け取り意味のある動作を行う」というインタラクションを扱うために、発話交替を単位とした処理ではなく、例えば 0.5 秒周期といった時間間隔（インターバル）の概念を導入する。つまり、ある時間間隔 t のインターバルごとに行動の決定、状態の決定、報酬の評価を行う。

各インターバルにおいて、エージェントはユーザの状態を観測し、探索戦略によってある動作（発話など）を決定し、エージェントからの出力とする。環境モデルはエージェントからの出力（観測される動作）に基づいて確率的に行動と報酬を決定する。そしてエージェントは報酬を受け取り、以下のような Q 学習が行われる。報酬が収束すれば学習を終了する。

Q(s, a)を初期化
各エピソードに対して繰り返し：
　s を初期化
　エピソードの各ステップに対して繰り返す：
　　s 状態で Q に基づいて行動 a を選択する
　　行動 a を取る、r, s’を観察する
　　$Q(s,a) \leftarrow Q(s,a) + \alpha[r + \gamma \max_{a'} Q(s',a') - Q(s,a)]$
　　$s \leftarrow s'$
　s が終端状態ならば繰返しを終了

4. 予備実験

4.1 実験条件

擬人化エージェントが音声認識、音声合成、顔画像生成、顔画像認識によるユーザの頷き検出の機能を持つという想定で、マルチモーダル対話のシミュレーションと強化学習を行う実験系を Python 言語で実装した。

対話の内容としては、ユーザからの道案内の要求があったという前提で、エージェントが正しい道を案内する。エージェントの応答は 3 つの発話

- 「交差点で左に曲がって」
- 「右側にあります」
- 「さようなら」

に限定する。ユーザが可能な発話は

- 対話を進む割り込み「はい」「うん」
- 対話を前に戻す割り込み「もう一回」「え？」「さっきの」
- 対話を終了する発話「ありがとう」「わかった」「さようなら」

である。エージェントが対話を終了する発話を行ったら、終端状態に遷移し、対話（一つのエピソード）は終了する。エージェントおよびユーザ発話には 5 モーラ／秒と仮定した。

各インターバルにおいては、エージェントが正しい順序で説明の発話を行っている場合にのみ、正の値の報酬を与える。

またエピソード終了時点での報酬は $\alpha \times$（聞き取れた発話数／総発話数）$+ \beta \times$（タスク達成に最低限必要な発話の所要時間／実際の対話の所要時間）$+ \gamma \times$（エージェントの適切な頷き回数／エージェントの総頷き回数）（ただし α, β, γ は重み係数）である。

4.2 強化学習の結果と考察

前述の条件で 300 回のエピソードに関する強化学習を行った。学習回数と報酬の関係を図 3 に示す。報酬の値は約 200 エピソードで収束している。その後も報酬が一時的に低下することがあるが、これは今回用いた ε グリーディ行動選択において ε の値を一定（$\varepsilon = 0.1$ ）にしているためである。

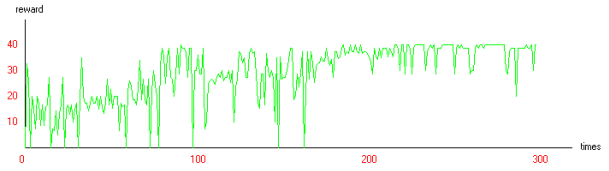

図 3　学習回数と報酬の関係

図 4 に学習結果に基づく対話例を示す。ただし横軸は時間（インターバルは 0.5 秒）で、各黒線は上から順に、エージェントの発話状態、エージェントの頷きイベント、音声認識のイベント（発話未検出、発話検出済み、認識結果計算中）を意味する。またこれに対応するイベントを表 1 に示す。これらの結果より、エージェントが適切な順序で説明を行えるような方策を獲得したことがわかる。またエージェント頷き（B）（D）はユーザの発話中に生成された「合理的な頷き」である。ただし（A）（C）は本来望ましくない頷きである。

これらの結果はまだ十分とは言えないが、基本的な手法の妥当性を示唆する結果と考えられる。

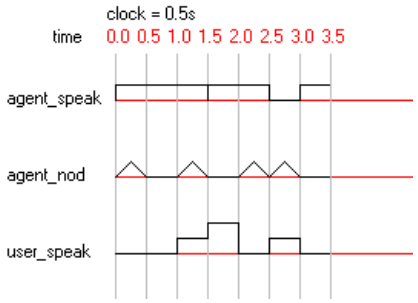

図 4　学習結果に基づく対話例

表 1　各インターバルにおけるイベント

時間t	エージェント発話	エージェント頷き	ユーザ発話
0	「交差点で左に曲がって」	True（A）	Idle
0.5	continue	False	Idle
1	continue	True（B）	発話中「はい」
1.5	「右側にあります」	False	音声認識中
2	continue	True（C）	認識結果「はい」
2.5	idle	True（D）	Idle
3	対話終了	False	発話中「はい」

5. まとめ

本報告では、近年普及してきた強化学習に基づく音声対話の枠組みを発展させ、マルチモーダル拡張された汎用的なユーザシミュレータを用いて、エージェントの実時間制御を行うモデルを自動学習する手法を提案した。今後は擬人化音声対話エージェントへの実装・評価を行う予定である。

謝辞

日頃御議論いただいている嵯峨山・小野研究室の皆様、ISTC（音声対話技術コンソーシアム）の皆様、Galatea Project の皆様、情報処理学会試行標準 IPSJ-TS 0012「マルチモーダル対話のための記述言語 Part1 要求仕様」WG メンバーの皆様から多くの示唆を得たことに謝意を表する。

参考文献

[1] 嵯峨山 茂樹: "なぜ音声認識は使われないか・どうすれば使われるか？" 情報処理学会研究報告, 94-SLP-1, Vol. 94, No. 40, pp. 23-30, (1994)

[2] 西本 卓也, 志田 修利, 小林 哲則, 白井 克彦: "マルチモーダル入力環境下における音声の協調的利用—音声作図システム S-tgif の設計と評価—," 電子情報通信学会論文誌 D-II, Vol.J79-D-II, No.12, pp.2176-2183 (1996)

[3] 嵯峨山 茂樹, 西本 卓也, 中沢 正幸: "擬人化音声対話エージェント," 情報処理学会誌, Vol.45, No.10, pp.1044-1049, (2004)

[4] 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, 嵯峨山 茂樹: "カスタマイズ性を考慮した擬人化音声対話ソフトウェアツールキットの設計," 情報処理学会論文誌, vol.43, no.7, pp.2249-2263, (2002)

[5] http://sourceforge.jp/projects/galatea/

[6] 西本 卓也, 岩田 英三郎, 櫻井 実, 廣瀬 治人: "探索的検索のための音声入力インタフェースの検討," 情報処理学会研究報告 2008-HCI-127(2), pp.9-14, (2008) http://www.youtube.com/watch?v=g6xYvRj3E3I

[7] 西本 卓也, 中沢 正幸, 嵯峨山 茂樹: "音声対話における擬人化エージェントの身体動作表現の利用," 2004 年度人工知能学会全国大会論文集, 2C2-01, (2004)

[8] Williams, J.D., Poupart, P., Young, S.J., “Factored partially observable Markov decision processes for dialogue management,” Proc. Workshop on Knowledge and Reasoning in Practical Dialog Systems, Int. Joint Conf. on Artificial Intelligence (IJCAI), Edinburgh. (2005)

[9] Jason D. Williams, Steve Young: “Partially observable Markov decision processes for spoken dialog systems,” Computer Speech and Language, Volume 21 , Issue 2, pp. 393-422 (2007)

[10] Richard S. Sutton, Andrew G. Barto（三上 貞芳, 皆川 雅章 訳）: 強化学習, 森北出版 (2000)

[i] 連絡先:東京大学大学院情報理工学系研究科
〒113-8656 東京都文京区本郷 7−3−1
E-mail: {lu, nishi} @hil.t.u-tokyo.ac.jp