

自然な対話を行う擬人化エージェントシステム

小川 均

情報理工学部

概要 対話に密接に関係するインタフェースと対話部に関して、自然な対話に必要な機能を提案し、試作した。インタフェースに関しては、入力として音声による入力とキーボードによる入力の両方が可能であるようにした。出力としてテキストに加えて、音声出力を行い、擬人化エージェントにより表情や身振りによる状況表現の補助を行った。対話部ではつぎの5つの機能について検討した;挨拶, 文脈に合わせた応答, 多様な応答表現, 時間帯発話, ユーザ未応答対応。試作システムを実現するために, NetPeople (iNAGO社), Arcadia SpeeCAN-TTS SDK (アルカディア社), および, Microsoft Speech SDK 5.1 (マイクロソフト社) を利用した。実験では, 5つの機能の実現が確認できた。

Multi-modal Interaction Project:

A Natural Interactive System Using Personified Agent

Hitoshi Ogawa

Faculty of Information Science and Engineering

Abstract: In this paper, several functions were proposed, that are necessary for the interface and the dialog control part to have a natural conversation. Both speech recognition and input from keyboard are necessary as functions for the input interface. Expressions and motions of a personified agent were proposed to assist the representation of situations adding to text output and voice synthesis. The following five functions for dialog were proposed; greetings, suitable replies for a context, various expressions for reply, adaptable utterances for time, and utterances in the case of no answer.

1. はじめに

近年、情報技術の発展により、パーソナルコンピュータや情報家電など様々な機能を持つ製品が普及している。しかし、機能が複雑だったり、操作するためにキーボードやマウスなどを駆使しなければならなかったりする事が多く、決して「使いやすい」とは言えないシステムが多い。そこで、機械との対話という手段によって利用しやすいインタフェースを実現しようとする研究が盛んに行われている^[1]。その中でも、日常生活において人間が機械を便利に利用するために、擬人化エージェントやロボットなどの自然な対話研究が

進められている^[2]。しかし、自然な対話においても、擬人化エージェントにおける外見上の研究^[3]であったり、あいづち^[4]であったりと、対話上一部分の研究にしか過ぎず、対話全体の流れを考慮していないものである。これでは、人間対人間の会話と比較すると自然な対話であるとは言いがたい。

そこで、本研究では、人間が情報家電等を扱う際など、機械が人間に対して情報を要求するために対話を行なう事を前提とし、その上でより自然な対話を可能にするシステムを考案する。

2. 自然な対話システム

機械と人間の対話においては、Q&Aシステムのように目的に応じた会話をするのが通例であった。しかし、人間同士の対話において、自然の対話とは同じ質問に対しても柔軟な対応を返したり、相手が何もしていない状態であれば何かしら反応を起こしたりするものである。人間対機械のコミュニケーションにおいてそれらの人間的な言動や行動を実現する事によって、より自然な対話が可能になる。

2. 1 システムの概要

一般的にユーザと自然に対話するシステムの構成は、図1に示すようにインタフェース、対話部、および、推論部から構成される。それぞれについて簡単に説明する。

(1) インタフェース

対話の時に使用する機器やメディアを提供する。計算機のようにキーボード・マウスの入力、テキストやグラフの出力だけでなく、日常生活に使用しているメディアを利用できるようにする必要がある。

(2) 対話部

対話内容とユーザの特徴（年齢、性別、嗜好など）を考慮し、目的指向の発話ではなく、ごく自然な発話内容を生成する部分である。

(3) 推論部

ユーザの要求、嗜好、環境の制約等を考慮して、対象システムへの操作を決定する部分である。

2. 2 自然な対話に必要な機能

対話に密接に関係するインタフェースと対話部に関して必要な機能を述べる。

2.2.1 インタフェース

(1) 入力部

正確にテキストを入力するにはキーボードが適切であるが、一般の使用者には不慣れである。サービスの窓口で人間が対応する時、ほとんどの場合、音声による入力となる。したがって、本研究で開発する対話システムでは、音声による入力とキーボードによる入力の両方が可能であるようにする。

(2) 出力部

計算機からの正確な出力は、テキストとグラフであるが、モニターを注意深く見る必要がある。長時間の使用には適していない。本対話システムでは、テキスト出力に加えて、音声出力を行なう。さらに、擬人化エージェントにより表情や身振りによる状況表現の補助を行なう。たとえば、答えられない質問に関しては、しばらく考える身振りをしたあとで返答を行なうことにより、答えられなくて困っていることを直感的に表現する。

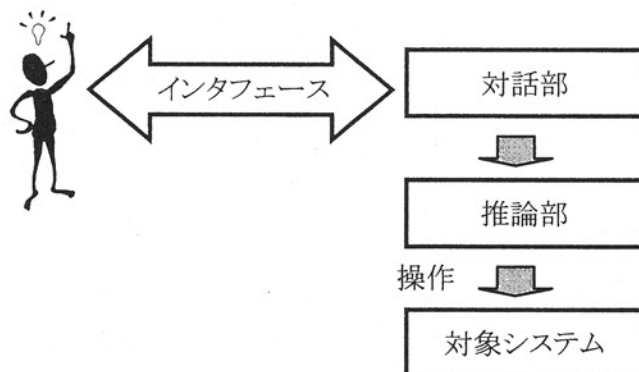


図1 対話システムの構成

2.2.2 対話部

自然な対話を行なうためには、目的に依存した発話以外につきのような機能が必要である。

(1) 挨拶

会話の始めに挨拶を行なう。人間同士の対話では一般的には、挨拶やそのときの時勢について2、3の会話を交わした後に、本題に入る。

基本的には、会話を始める時間帯に応じて、朝、昼、夜に適切な挨拶を行なうことが必要である。

(2) 質疑の流れに適応した応答

・文脈に合わせた応答

同じ単語や文章が入力された時、質疑の流れにより、すなわち、文脈により応答が異なる。たとえば、「場所はどこですか？」という質問に対して、話題が病院であればその位置、ゲートボールであればそのゲートボール場を返答する必要がある。

・多様な応答表現

ユーザからの同じ質問が繰り返されたとき、常に同じ応答をするのではなく、同様の意味で言葉を変更することも自然な対話に必要なことである。

(3) イベント駆動型応答

単にユーザの質問（入力）に応答するのではなく、ある出来事に対応して、自動的に発話する機能も必要である。

・時間帯発話

指定された時間帯に発話を行なう。たとえば、朝、昼、夜にそれぞれの食事や薬の話題を尋ねることにより、対話の推進を図る。

・ユーザ未応答対応

対話中にユーザが発言や操作を一定時間行わない場合特別な反応を返す。すなわち、対

話を継続するよう促したり、対話の終了を尋ねたりする。

3. マルチモーダルシステム

2章で述べた自然な対話システムのうち、インタフェースと対話部を実現するために、NetPeople (iNAGO) ^[5]、Arcadia SpeeCAN-TTS SDK (アルカディア) ^[6]、および、Microsoft Speech SDK 5.1 (マイクロソフト) ^[7] を使用した。

3.1 システム構成

マルチモーダルシステムの全体構成は図2に示すようである。NetPeopleは、擬似エージェント表示とテキスト入出力を行なうクライアントと対話部を実現するサーバの2つに分かれる。音声認識にはMicrosoft Speech SDK 5.1を使用し、認識結果はNetPeopleクライアントに直接テキストとして入力される。また、音声による返答は、NetPeopleサーバにより指定され、Arcadia SpeeCAN-TTS SDKによりテキスト-音声変換され、スピーカより出力される。

3.2 NetPeople

NetPeopleシステムは、NetPeopleエージェントという返答システムを作成するためのプログラムとコンポーネントの複合体である。NetPeopleテクノロジーは、“ブレイン”、“ボディー”及び“環境”の組み合わせであり、これらの3つの要素を同時に使用することにより、リアルなユーザーとのやり取りを実現する。

Netpeopleの裏側にはゴールベースの自然言語処理が内蔵されており、人とのやりとりにおいて、外見上だけでなくキャラクターの内面から人間らしさが提供できる。更に、音声合成を用いた自然言語インターフェースにより、音声認識を導入することで、NetPeopleとの会話はリアルタイムで行われるだけでなく、人と人とのやり取りのように感じられる。特に、Netpeopleでは、対話内容に詰まったときや一定時間反応が無かったときに、今出来る事を選択肢として表示する事が可能であるの

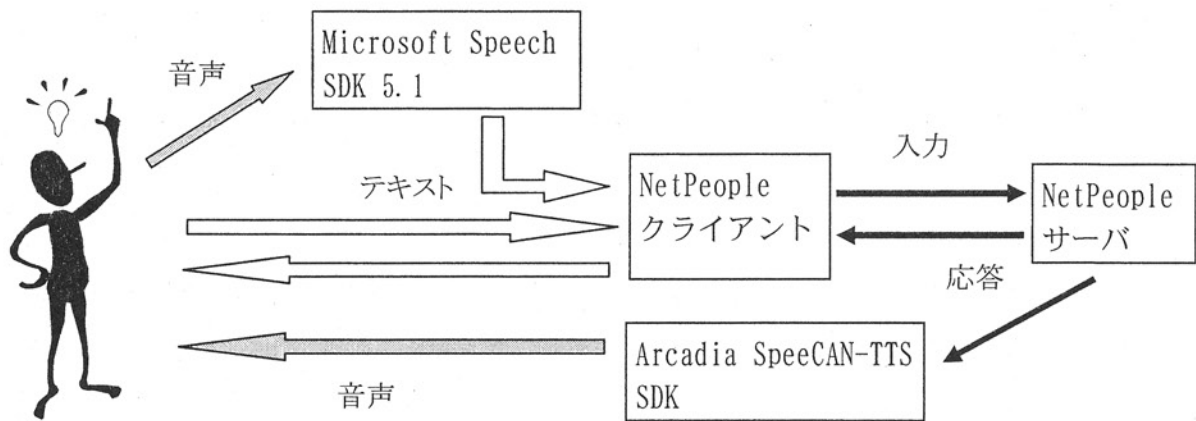


図2 マルチモーダルシステムの全体構成

で、スムーズな対話の実現可能である。Netpeopleの対話機能を利用し、キーボードやマウスといった面倒なデバイスを使用せずともさまざまな家電や機械を扱えるようにする事により、人間にとって快適な機能を利用しやすい状態で提供する事が出来るようになる。

NetPeopleは人工知能、学習処理、コミュニケーションのための自然言語処理、イベントに対する返答、そしてエージェントボディとの関係、ボディやユーザー動作との相互作用などを扱う。例えば、イベントデータベースにおいては、グリーティングという対話の最初に必ず挨拶したり、名前を聞くとしたユーザーの情報を取得したりするイベントを起こす事ができ、自然な対話には必要な機能が実現できる。また、外部アプリケーションインタフェースやブラウザインタフェースも備えているので、ユーザーと対話しながらブラウザのホームページを切り替えたり、別のエージェントやアプリケーションを動かしたりという事も実現できる。そして、ユーザーの情報をデータベースで保存しておく事が可能であるので、最初にユーザーの名前を取得した後、それぞれのユーザーに合った環境を構築する事も可能である。

3.3 Arcadia SpeeCAN-TTS SDK

アルカディアに開発されたテキスト文を入

力対象とし、音声データを出力するプログラム（音声出力エンジン）をアプリケーションに実装させるための開発キットです。自動応答電話の可変ガイダンス部分、文字放送の読上げなど、リアルタイムにテキスト～音声変換が必要となるシーン全てである。

3.4 Microsoft Speech SDK

マイクロソフトが開発した、音声認識および合成システムである。本システムでは音声認識部分を利用した。

4. 実行例

自然な対話に必要な機能について、実現結果を例を用いて示す。

(1) 挨拶

システムを起動したときに、その時間帯に合わせて挨拶をする。図3、4にそれぞれ朝、昼にシステムに起動したときの画面を示す。システムは対話する前に時刻を確認し、その時刻に合わせた発話を行なう。

また、対話を終了するときには図5のような発話を行なう。

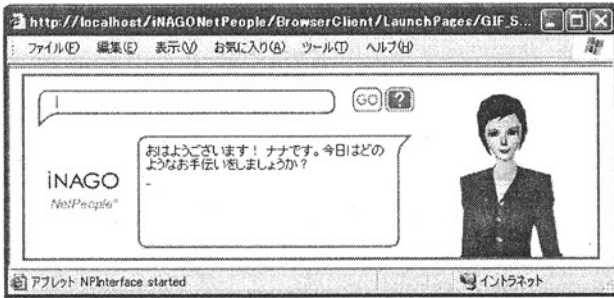
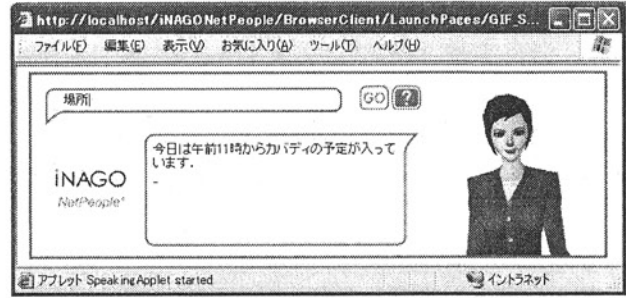


図3 朝に起動した時の画面



(a) 予定の回答

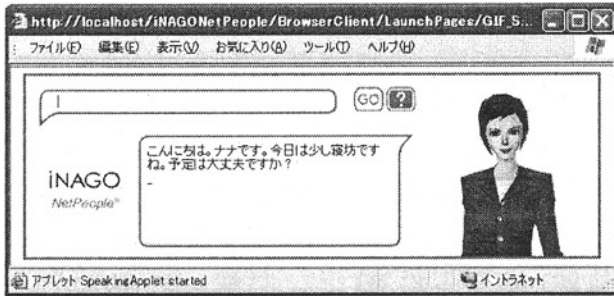
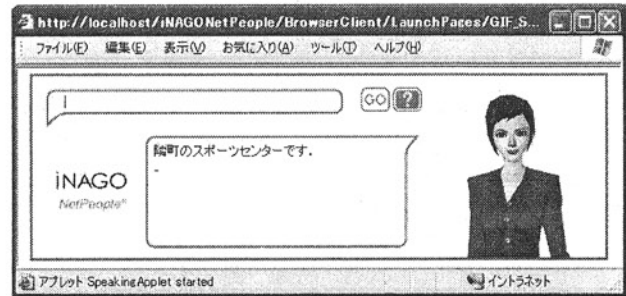


図4 昼に起動した時の画面



(b) 場所の回答

図6 スポーツセンター

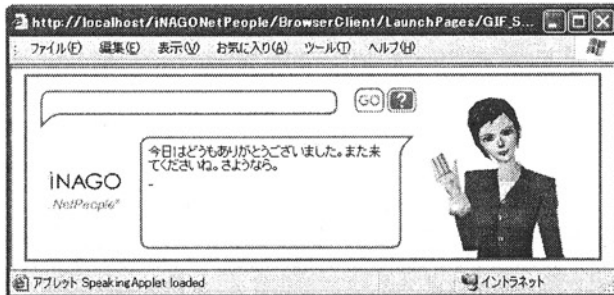
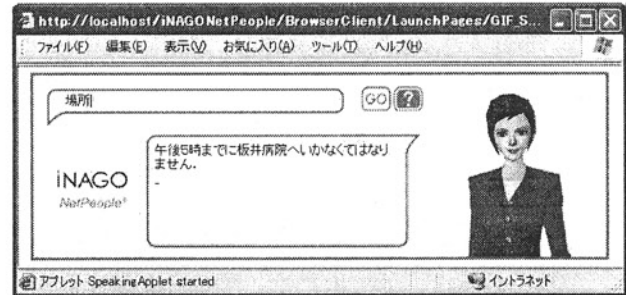


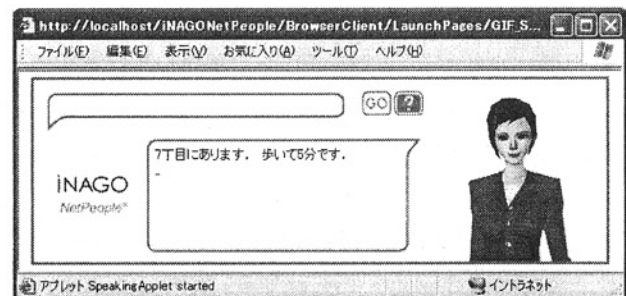
図5 システムが終了する時の発話



(a) 予定の回答

(2) 文脈に合わせた応答

同じ単語を用いて質問した場合でもその時の文脈に応じて適切な回答を行なう。図6 (a)、図7 (a) は予定を質問した場合の応答を示している。午前と午後で行く場所が異なるため、両方とも同じ単語「場所」を用いて質問しているが、それぞれ図6 (b)と図7 (b)のように応答を行なっている。結果的に適切な場所を回答している。



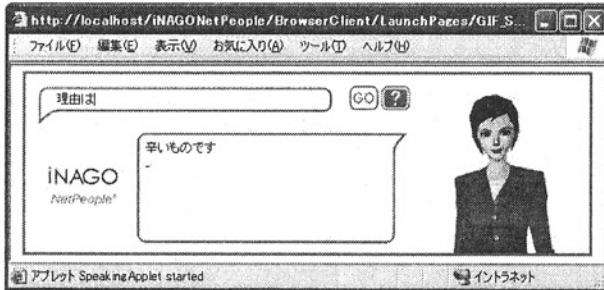
(b) 場所の回答

図7 病院

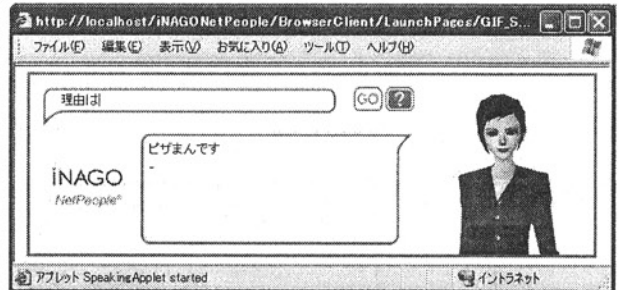
同様の例を、図8(a)(b)と図9(a)(b)に示す。両方とも「嫌いなものは」に対する回答であるが、その「理由は」の質問について適切な回答を行なっている。

(3) 多様な応答表現

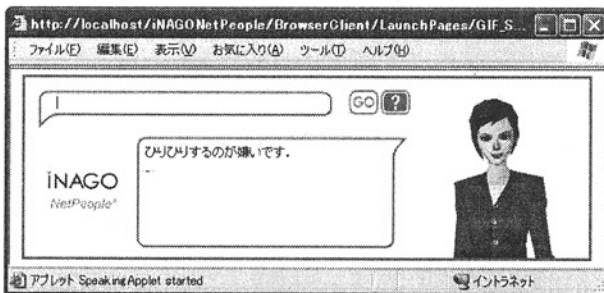
同じ発話に対して時刻が変われば、その時刻に合わせた対応を行なう。図10(a),(b)にはそれぞれ朝と昼に「おはよう」に対する応答を示す。



(a) 「嫌いなもの」に対する回答

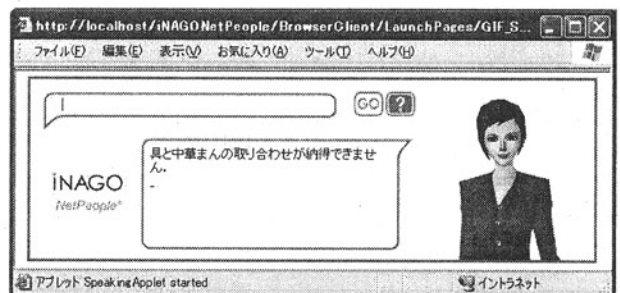


(a) 「嫌いなもの」に対する回答



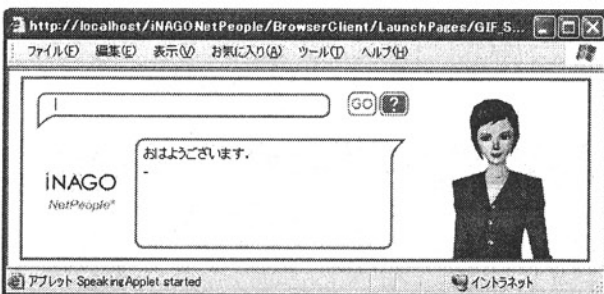
(b) 「理由は」に対する回答

図8 「辛いもの」が嫌いな場合

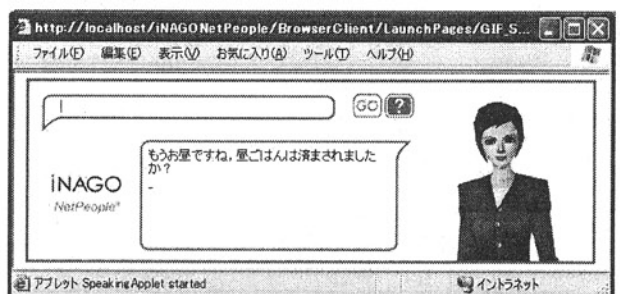


(b) 「理由は」に対する回答

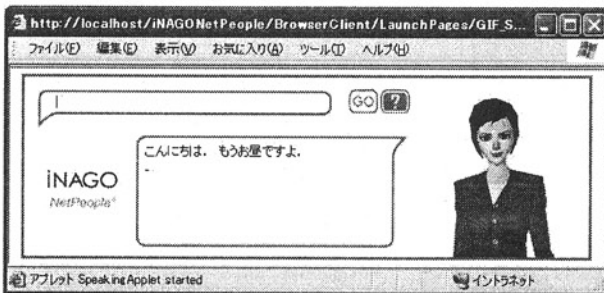
図9 「ピザまん」が嫌いな場合



(a) 朝の対応

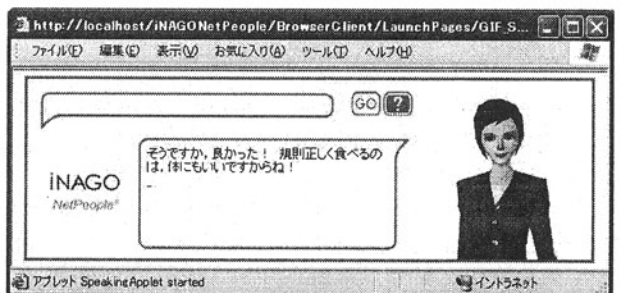


(a) 昼に行なわれる発話



(b) 昼の対応

図10 時間帯による対応の違い



(b) 「はい」に対する応答

図11 時間帯発話の例

(4) 時間帯発話

ユーザから入力がない場合でも、指定した時刻がきた場合に適切な質問を行なう。また、その時点でのユーザの応答にも適切に対応することができる。図11(a)(b)に例を示す。

(5) ユーザ未応答対応

ユーザが対話中、または、長く入力しない場合は、図12のような発話を行い、対話を促す。

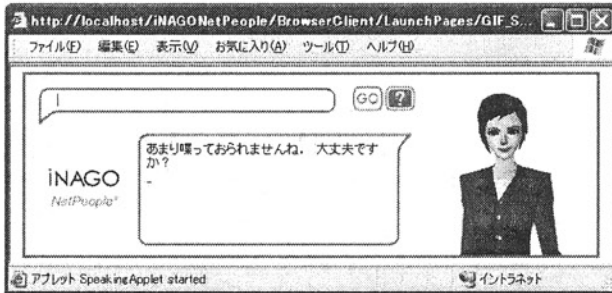


図12 ユーザ入力が途切れた場合の発話

5. むずび

今後、対話例を増加し、単身生活者を対象とした実証実験を行なう。

参考文献

- [1] 三好秀夫: 人とコンピュータとの自然な対話, 映像情報メディア学会誌, vol. 55, No11, pp1403-1406 (2001).
- [2] 米村俊一, 細谷未生, 西山茂: 人と擬人化エージェントとの対話設計法に関する検討, 電子情報通信学会信学技報, HCS99-70, 29-35 (2000).
- [3] 芥子育雄他: デジタル情報家電のインタフェースエージェント技術の開発, シャープ技報, vol. 77, No9, pp15-20 (2001).
- [4] 土肥浩, 石塚満: WWWと連携する擬人化エージェントとのHAI, 人工知能学会誌, vol. 17, 693-700 (2002).
- [5] NetPeople PLATFORM,
http://www.inago.co.jp/iNAGONetPeople/iNAGO_website/html/np/index.html.
- [6] SpeeCAN,
<http://www.arcadia.co.jp/SpeeCAN/>.
- [7] Microsoft Speech SDK 5.1,
<http://www.microsoft.com/speech/download/sdk51/>