

## 音声対話機能を持つパーソナルガイドエージェント

土井 俊介<sup>†</sup> 角 康之<sup>†</sup> 間瀬 健二<sup>†</sup> 中村 哲<sup>‡</sup> 鹿野 清宏<sup>‡</sup>

<sup>†</sup>(株)ATR知能映像通信研究所

<sup>‡</sup>奈良先端科学技術大学院大学 情報科学研究科

### 1. はじめに

近年、携帯端末の性能が向上し、ネットワーク接続機能や位置検出機能、音声インターフェースを備えたシステムの開発が可能となってきている。筆者らは、これまでに携帯端末を用いたエージェントを介した音声メッセージシステム<sup>[1]</sup>や、C-MAP<sup>[2]</sup> (Context-aware Mobile Assistant Project) システムを試作し、展示会場を対象とした個人化ガイドや新たなコミュニケーション支援手法の提案を行ってきた。そこで筆者らは、これまで開発してきたシステムのインターフェースとして音声対話の可能性を探ることと、音声インターフェースを持ったシステムの形態の違いがユーザにどのような影響をもたらすかを調べることを目的とし、「音声対話機能を有するパーソナルガイドエージェント」を実験システム上に実装した。本稿では試作した本実験システムの構成や特徴について述べ、また、音声対話インターフェースの形態によって、ユーザへ与える影響についても述べる。

### 2. システムの概要・特徴

実験システムは、従来のガイドシステムと比較し、次にあげる特徴を持っている。

#### ● 音声対話インターフェースの使用

音声認識(単語)と音声合成を使用し、音声対話を通じてガイドを行う。

#### ● 音声メッセージ機能

伝言板と同じような働きを音声メッセージ形式で残し、ユーザ情報に応じて提示することで、非同期・オープン型の新たなコミュニケーション形態を提案している。

#### ● 文脈情報の使用

位置・ユーザの興味・対話の応答内容に応じてガイドやルート提案を行う。

### 3. システム構成

#### 3.1 システム構成・ネットワーク構成

ガイドシステム端末は、無線LANを通じて、現在のユーザ位置や個人情報、音声メッセージ、ガイド情報などが格納されたデータベースにアクセスをする。そして、現在のユーザの状況に従い展示の位置に応じたガイドや、音声メッセージの取得、ユーザ情報(興味)の更新を行う。ユーザの位置は天井に取り付けた位置検出センサで赤外線バッジのIDを取得し、位置情報データベースを更新する。端末は音声入出力とペン入力GUIを持ったパッド形式と音

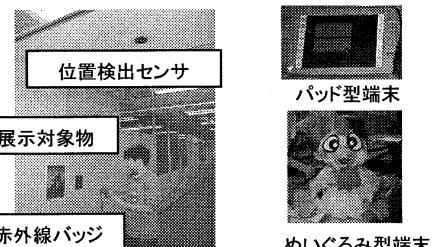


図1. 実際の使用風景と端末の形態

声入出力のみのぬいぐるみ型を用意した。図1に使用風景と端末を示す。

### 3.2 音声対話部アルゴリズム

認識部には東芝音声システムを用いた。東芝音声システムは、音声合成と音声認識機能を持ち、SAPI を介して使用できる。筆者らによる性能を評価した結果を表1に示す。

表1. 使用した音声システムの性能評価

バージョン	: 東芝音声システム・音声認識コントロール 1.0
システム設定値	: 認識精度 50%, タイムアウト 300msec, 感度 100
データベース	: ATR音声バランス216単語 (話者MHT, 12kHz, モノラル)
実験方法	: テキストで対象となる216単語を登録し認識
認識結果	: 94.9%

本システムの音声対話部の流れを図2に示す。対話は基本的にユーザの発話に対してシステムが応答を行う形式だが、認識精度の向上をはかり、ユーザが話しかけない時に対応でき、さらに、機械との対話という印象を軽減するため、次に示す方法を考案し実現している。

#### ● 語をグループ化する

認識・合成で使用する単語は、類似した意味毎に階層化したグループにまとめる。グループ化することで、同じ意味のグループの単語であれば認識が可能である。また、グループ毎にいくつかの応答文を用意しておき、その中から乱数で選ぶことである程度揺らぎを持った応答が可能になる。

#### ● ユーザに関する情報の使用

位置や興味などのユーザの文脈にあわせてガイドの開始や、ルート提案を自発的に行う。

#### ● 次の発話は前発話に依存

認識時、認識対象となる単語(グループ)および応答文は一つ前にシステムが発話した単語に依存するように辞書を構成する。

#### ● 話題に応じた処理

ユーザの要求(例えばガイドの要求など)、ユーザの発話語から関連付けた話題を決定する。話題に沿ったス

Personalized Guide Agent System with Voice Interaction

Shunsuke Doi<sup>†</sup>, Yasuyuki Sumi<sup>†</sup>, Kenji Mase<sup>†</sup>,

Satoshi Nakamura<sup>‡</sup>, Kiyohiro Shikano<sup>‡</sup>

<sup>†</sup>doi,sumi,mase}@mic.atr.co.jp, <sup>‡</sup>{nakamura,shikano}@js.aist-nara.ac.jp

ATR Media Integration & Communications Research Laboratories

<sup>‡</sup>Graduate School of Information Science, NAIST

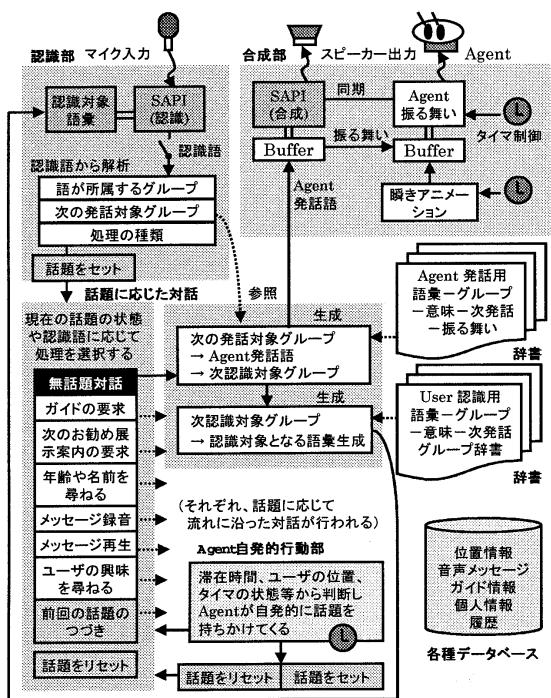


図2. 本システムの音声対話部の模式図

クリプト処理を行う。認識対象単語は話題に合ったグループ集合で構成することで、認識対象語のしづら込みを行う。また、無発話時間が長い場合は、話題をリセットする。

#### ● 認識対象以外の語を認識対象にする

発話内容の短い語や、話題に合っていない認識対象語彙を「話題・誤認識」というグループとして、認識対象語彙に加える。「話題・誤認識」単語が多く発話された場合、システム側の話題とユーザ側の話題とか一一致していないと考え、話題の切り替えや話題のリセットなどを行う。

#### 3.3 音声メッセージ

展示場所に音声でメッセージを保存し、他の人がそのメッセージを聞くことで、共感を促したり、非同期の連鎖コミュニケーションを活性化することが可能になると考え、音声メッセージ機能を付加した。今回は、システムとの対話を通じてメッセージを残し、また、興味が一番近い人のメッセージを再生している。

#### 3.4 個人の興味、位置情報の取得、反映

本システムは表2に示す文脈情報を取得し、ガイドのルート提案等に反映している。

表2. 本システムで使用・取得可能な文脈

[ユーザ情報]	●氏名、年齢、興味キーワード
[位置情報]	▲位置情報 ▲展示場所の滞在時間
[ガイド情報]	●展示者名・案内メッセージ
[対話から取得]	●展示に関するキーワード ▲ガイドを行った内容の深さ ▲展示のキーワードをユーザ情報に付加する ■Agentの質問に答えた興味の有無 (●:明示的に記述 ▲:状況から取得 ■:対話から取得)

#### 4. インタフェースの違いと主観評価の結果

開発したシステムを異なった5種類のインターフェース形態にして評価実験を行った。実験では認識率の導出、ユーザの満足度、印象等の主観評価アンケート実験を行った。図4にインターフェースの違いによるユーザの主観評価の結果を示す。被験者は19名である。音声の平均認識率は64%、ユーザの平均発話数は30語、平均使用時間は280秒である。グラフは質問に対して「一番そう思う」を2点、「一番そう思わない」を-2点として平均をとったものである。その結果、便利さと確認機能を持った音声入出力との相関が強い傾向が得られた。また、システムAはBと比較して、音声入出力における確認が便利さだけでなく、愛着・親密性を向上させている。また、ぬいぐるみの形態が愛着を増すことも確認できた。

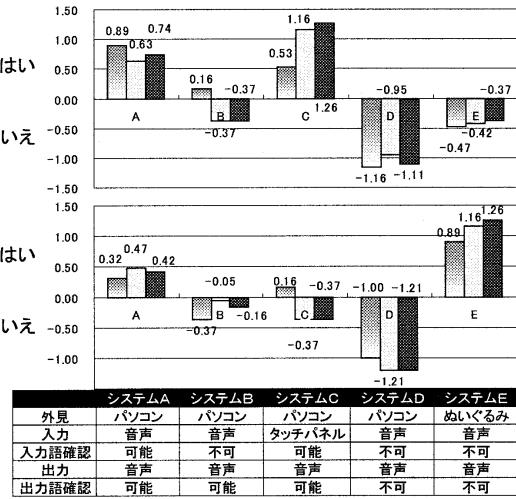


図3. インタフェースの違いによる主観評価実験の結果

また、結果からペット飼育経験の無いユーザは、音声認識を使用したシステムよりもタッチパネル方式のシステムの方が便利であると答える傾向が得られた。(図4)

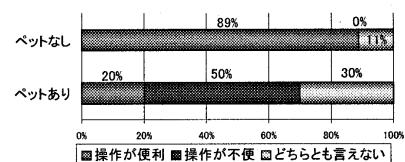


図4. ペット経験とタッチパネル方式の操作性の評価値

#### 5.まとめ

音声対話システムを開発し、インターフェースの形態を変え評価実験を行った。その結果、同一のシステムでもインターフェースの違いによって異なる印象をもつ結果が得られた。音声対話型システムでは、ユーザが持つ印象はインターフェースの形態に影響されると考えられる。今後、実験結果より、認識率、システムへの没入具合、ユーザの評価、ユーザ自身の経験などとインターフェースの形態との関係を導こうと考えている。

#### 参考文献

- [1] 土井,角,間瀬,中村,鹿野. 展示見学時の知識流通を目的としたcontext-awareな音声メッセージシステム,1999年人工知能学会全国大会,33-03,pp.488-491, 1999.6
- [2] 角,江谷,シドニー,ニコラ,小林,間瀬.C-MAP : context-awareな展示ガイドシステムの試作. 情処論, Vol.39, No.10, pp.2866-2878, 1998.