

## 確率的な振舞を伴う擬人化対話エージェント

川本 真一, 井波 暢人, 加藤 裕, 槻尾 洋志, 藤永 勝久  
山崎 義人, 横山 国宏, 中井 満, 下平 博, 嵯峨山 茂樹  
北陸先端科学技術大学院大学 情報科学研究科

〒 923-1292 石川県能美郡辰口町旭台 1-1

Tel: 0761-51-1225 / Fax: 0761-51-1149

E-mail: skawa@jaist.ac.jp

URL: http://www-ks.jaist.ac.jp/

### 1 はじめに

筆者らは、心のこもったコミュニケーションの実現や、より親しみやすいシステムの構築を目指して研究を行っており、これらを演出する要素として「人間らしさ」というキーワードを取り上げ、その実現の為に擬人化対話エージェントの構築を行っている。

本稿では、このシステムの人間らしさを生み出すメカニズムとして確率的に振舞う感情モデルを埋め込んだシステムを試作した例について報告する。

### 2 対話のモデル化

2人の話者A(ユーザー、外部), B(システム)の対話において、Aの発話に対して、Bが応答・振舞を決めるとき、「Bの現在の状態や経験」を基に、「Aから得られた情報」を利用して、「適切な応答・振舞を決定」する問題を考える。

この問題に対して、中身は見えないけど人間らしく振舞うモデル、つまり人間の内部状態がある程度ブラックボックス的に扱うモデルを考える。

また、人間の心が状態の変化で表現されると仮定するならば、Aの発話に対して、Bが応答・振舞を決めることは、「Bの状態・モデル」を基に、「Aからの入力系列」を利用して、「適切な応答・振舞を尤度最大の基準で推定する」問題ととらえることで、隠れマルコフモデル (Hidden Markov Model; HMM) を適用できると考えた。

HMMは連続量の非定常信号を扱うことのできる数理的なモデルであり、音声認識をはじめとして広くその実用性が認められている。これは、人間が表現する様々な連続で非定常な時系列情報について、実用的なモデル化が期待できる。また、確率的な学習ができるという利点を持ち、実データから個性を真似るモデルとしての発展が期待できる。

従来研究で、Ergodic HMMを用いて対話の機能を分析、モデル化した研究 [1] があるが、今回は感情などの生成・制御のモデルとして HMM を適用した点が異なる。モデルの概念図を図1に示す。

ここで想定する対話(図2)では、システムが入力を受理し、それを反映して状態を変化させ、出力を与えたと考え、HMMは確率的に入出力の変換関数を切替えるモデルととらえる。これはユーザの入力を受理することで状態が変化し、尤度最大化基準における適切な出力を生成するモデルであり、出力を生成する状態が入力によって変化する。ま

た、モデルをブラックボックスとして見るならば、ユーザの入力に対するシステムの出力を生成するモデルである。

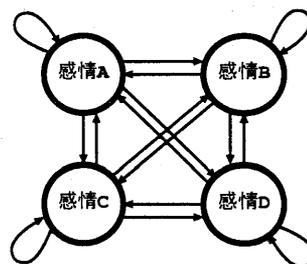


図1: 感情を生成・制御する HMM の概念図

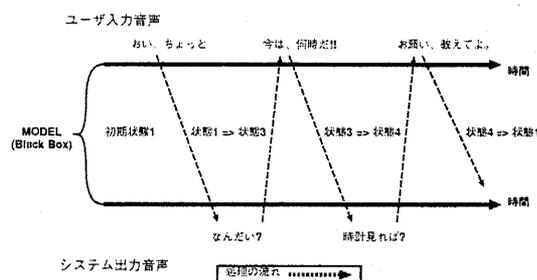


図2: 想定する対話の流れ

### 3 システム概要

システム構成図を図3に示す。モジュール間の通信は出来るだけ簡素にわかりやすくする為、テキストによる最小限の通信で実現している。また各モジュールは取り換え可能なように、標準入出力でデータを送受信し、各モジュールに wrapper をかぶせることで、モジュール間の通信仕様に合わせている。

#### 3.1 音声認識部

音声認識部は情報処理振興事業協会 (IPA) が実施した「独創的情報技術育成事業」の研究成果物である大語彙連続音声認識アコーダ Julius [3]、ならびにそれに付随する音響モデルを使用し、言語モデルはタスクを限定した統計的言語モデルを生成し、使用している。

#### 3.2 音声合成部

音声合成部は、録音再生方式の音声合成で簡易的に実現した。限られたタスクの発話について、複数のニュアンスの音声を録音し、ユーザの発話内容やシステムの感情状態によって音声を選択、再生する。

"Anthropomorphic Conversation Agent with Probabilistic Behavior," by Shin-ichi Kawamoto, Nobuhito Inami, Yutaka Kato, Youji Tsukio, Katsuhisa Fujinaga, Yoshito Yamasaki, Kunihiko Yokoyama, Mitsuru Nakai, Hiroshi Shimodaira and Shigeki Sagayama, School of Information Science, Japan Advanced Institute of Science and Technology.

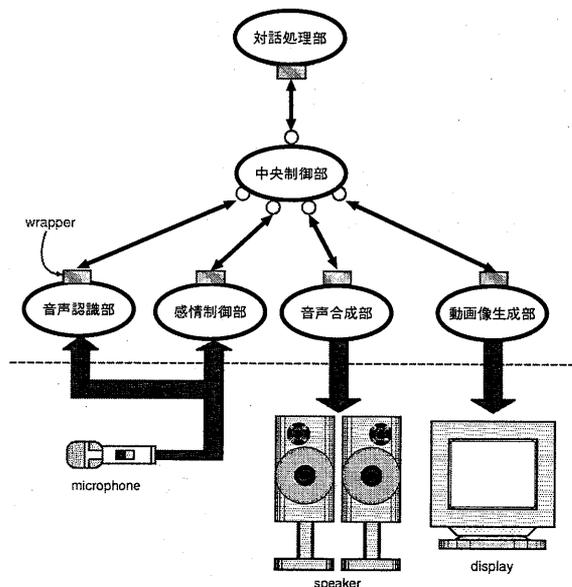


図 3: システム構成図

### 3.3 対話処理部

対話処理部は、限られたタスクで受理する文章について、応答文を生成し、受理した文章によって感情状態の補正を行うものである。感情状態の補正については、ヒューリスティックな値を使用している。

### 3.4 動画像生成部

動画像生成部は、情報処理振興事業協会 (IPA) が実施した「独創的情報技術育成事業」の研究成果物である表情合成ツール facetool[4] を使用している。対話が進むにしたがって、まばたきやうなずきなどを行うとともに、機嫌が良い・悪いなどを表現する表情を表出する。

### 3.5 システム制御部

システム制御部は各モジュール間の標準入出力を統括し、データフローを制御している。扱うデータ形式はテキストで、XML に似た以下のようなタグによってデータを識別している。

データの開始 ... < 識別子 >

データの終了 ... < /識別子 >

また、音声などの比較的データ量の大きなものについては、システム制御部を経由せず、直接モジュール間で送受信する。

### 3.6 感情制御部

感情制御部は、扱う情報を音声に限定し、感情などの非言語情報が含まれると報告される [2] 韻律的情報についてモデル化する。つまり、ユーザの発話する音声のマクロ的な韻律的情報 ( $F_0$  やパワー) によって状態が変化するモデルを考える。基本周波数  $F_0$  やパワーは、調波構造がきれいに現れている (有声音) 区間のみを扱うこととする。

この状態によって、各モジュールの振舞いに変化するよう設計してある。

## 4 対話例

ここでは、対話タスクとして、新聞購読勧誘の 1 シーンを考えた。システム (S) は自宅にて対応する客、ユーザ (U) を家庭を訪問する新聞販売員と設定した。以下に対話例を示す。

U: ごめんください。

S: はい。

U: 新聞の者ですが、今、どちらか新聞お読みになってますか?

S: いいえ。でも、うちは新聞読まないんですよ。

U: そこを何とか、3 か月くらいいかがですか。商品券もお付けしますよ。

S: 今までも一度もとらなかったから、本当に結構です。(怒った顔に変化し、声も怒った声)

U: 残念だなあ。他のところではなかなかこんなサービスないんだけどな。

S: 本当に結構です。お帰り下さい。

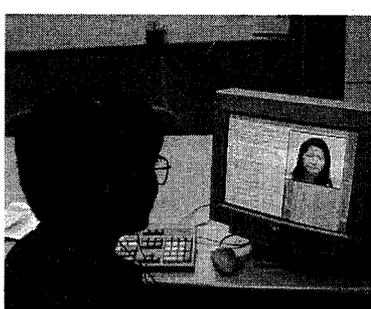


図 4: 対話の様子



図 5: 対話画面

ユーザの発話の変化によって、システムも状態を変化させ、様々な表情を表出し、システムの発話のニュアンスも変化させている。

## 5 おわりに

「人間らしさ」を演出する要素として確率的に振舞う感情モデルを埋め込んだ擬人化対話エージェントを試作した例について報告した。今後は他の特徴量も含んだ総合的な個性を演出するモデルへと発展させていく。

## 参考文献

- [1] 森川, 中里, 白井, “エルゴードック HMM による対話のモデル化と評価,” 日本音響学会講演論文集, 1-3-2 (1996.9).
- [2] 藤崎, “音声の韻律的特徴における言語的・パラ言語的・非言語的情報の表出,” 電子情報通信学会技術報告, HC94-37 (1994.9).
- [3] 河原 他, “日本語ディクテーション基本ソフトウェア (98 年度版) の性能評価,” 情報学研報, 99-SLP-26-6 (1999-05).
- [4] 森島 他, “顔の認識・合成のための標準ソフトウェアの開発,” 信学技報, PRMU97-282 (1998-03).