声の好みによって確率的に振舞が変化する擬人化対話エージェント

川本 真一 †, 山崎 義人 †, 高坂 大輔 †, 坂井 伸圭 †, 田村 紗雪 †,
松下 善則 †, 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: 0761-51-1699 (内線 1381), **Fax:** 0761-51-1149
URL: http://www-ks.jaist.ac.jp/

‡ 東京大学大学院 工学系研究科
〒 113-8656 東京都文京区本郷 7-3-1
Tel: 03-5841-6900, **Fax:** 03-5841-6953
URL: http://www.hil.t.u-tokyo.ac.jp/

あらまし

システムがユーザの声に対する “好み” を持つ事により, 振舞を変化させる音声対話システムを提案する. すなわち, ユーザの発声した声をシステムがどの程度好むかによって, システムの表情や相槌の頻度, 応答の傾向などが変化する感情モデルを組み込んだシステムである. 本稿ではこの感情モデルを実装した擬人化対話エージェントの試作例について報告する.

1. はじめに〜機械への個性の付与の必要性〜

将来の人間と機械とのコミュニケーションにおいては, 機械があたかも一個の人間のように話し, 聞き, 振舞うことの実現が必要であると考える. 例えば, 人間が伝えたい情報を十分に引き出すためには, 機械も人間のように存在感があり, 人格を持ち, 個性や感情を反映して振舞うような人間らしさが必要であると筆者らは考えている.

また, 近年デジタルペットなどさまざまな個性豊かなキャタクタが登場し, エンターテイメント市場において, その地位を確立しつつある. これはキャラクタの工学的な利用価値だけではなく, 個性の付与によって生み出される面白さに価値が見出されている一例といえる.

機械に個性を付与する問題に対して, 従来ではヒューリスティックに与えたり, ルールとして定義することが多かったが [1], 筆者らは機械の振舞を数理的モデルで表現することで, 振舞から見出される感情や個性をパラメトリックに扱うことを試みている [2]. このモデルでは, 図 1 の様に, 学習によって内部のモデル（個性）を構築する. 対話中には隠れた内部の状態の変化に応じて, 表面的な振舞を変化させる事で, 人間を模擬することを目指している.

2. 声の嗜好モデルによる応対モデルの切替え

既報 [2] のモデルでは, どのユーザに対しても同じ 1 つのモデルで応対していた. 試作したシステムは音声と映像を出力するが, システムが受け取る情報は音声のみであるので, ユーザ側の映像を伝送しないテレビ電話のような状況である. 人間同士の対話の場合, このような電話による対話でも, 相手の声によって応対が変わりうる. そこで, ユーザの声によってシステムの応対モデルの切替機構を新たに実装した. 応対の変化要因もさまざま考えられるが, 今回は声の好みに着目する.

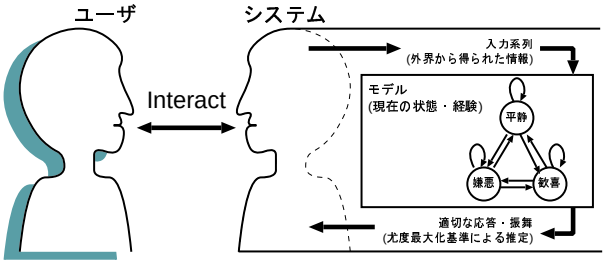

図 1: 感情を生成・制御するモデルの概念図

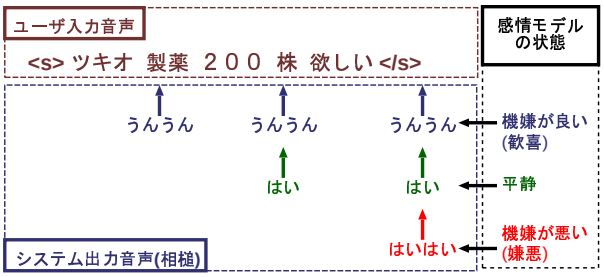

図 2: 想定する対話の流れ

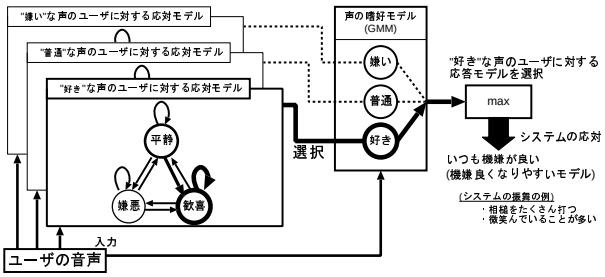

図 3: 声の嗜好モデルによって応対モデルを切替える感情モデルの概念図

その際, システムにとっての声の好みを認識する必要がある. この認識には, 話者認識で実績のある GMM (Gaussian Mixture Model)[3] を利用する. ある人の声の好みをモデル化するために, **1)** さまざまな話者の音声を 1 人の被験者に聞いてもらい, その声が好みかどうかを主観評価し, **2)** 主観評価値の同じ音声データを集めて GMM を学習する. これにより, 被験者の音声の好みに対応するモデルを生成する. また, 各主観評価値に対応した応対モデルも生成する.

応対モデル, および声の嗜好モデルは尤度を基準とするモデルであるため, 実際にユーザの音声が入力されたときに, 声の嗜好モデルの内, 尤度が最大のモデルに対応する応対モデルを用いる. さらに, 応対モデルの尤度が

“Anthropomorphic Conversation Agent that Changes its Behavior Probabilistically by the Taste for User’s Voice,” by Shin-ichi Kawamoto†, Yoshito Yamasaki†, Daisuke Kousaka†, Nobuyoshi Sakai†, Sayuki Tamura†, Yoshinori Matsushita†, Mitsuru Nakai†, Hiroshi Shimodaira† and Shigeki Sagayama†‡, †School of Information Science, Japan Advanced Institute of Science and Technology, 1-1 Asahidai, Tatsunokuchi, Ishikawa, 923-1292 JAPAN, ‡ Graduate School of Engineering, University of Tokyo, 7-3-1, Hongo, Bunkyo-ku, Tokyo 113-8656 Japan.

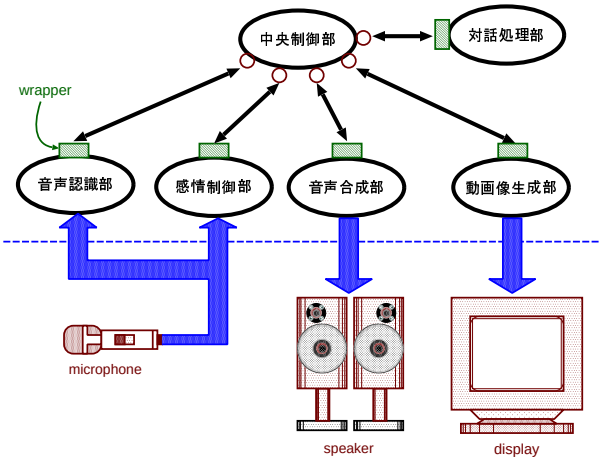

図 4: システム構成図

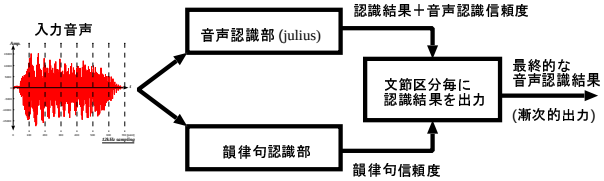

図 5: 韻律句信頼度を用いた漸次的音声認識システムの構成図

最大の状態よりシステムの振舞を生成する.

このモデルの導入により, 例えば図 2 に示すようにユーザの声がシステムにとっての好みの声ならば, 相槌の頻度が多くなり, 機嫌良くなりやすく, ユーザの声が好みでなければ, 相槌の頻度は少なく, 不機嫌になりやすいような, システムの振舞が実現できる可能性がある.

3. システム概要

システム構成図を図 4 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た. 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た.

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する. 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. **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. **音声合成部**は, 録音再生方式により簡易的に実現している. **動画像生成部**は, 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.

3.1. 韻律信頼度を利用した漸次的音声認識部

音声認識部は大語彙連続音声認識デコーダ Julius [5], 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.

さらに音声認識時の音響尤度を用いた音響信頼度と, 韻律句境界推定を用いた韻律信頼度を併用し, 漸次的音声認識の実現と, その認識結果の出力タイミングを利用したシステムの状態開示のための相槌の生成を行っている [6].

3.2. 声の嗜好モデルによる応対モデル切替えを利用した感情制御部

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 [7] 韻律的情報についてモデル化した. 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. この状態によって, 各モジュールの振舞が変化するように設計した.

また, 声質による初期モデルの切替えは, LPC ケプストラム 16 次, パワー, Δ LPC ケプストラム 16 次, Δ パワーの計 34 次元ベクトルを特徴量として, 複数の GMM の内, 尤度の高いモデルを使用した.

各特徴量は, 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した.

4. おわりに

ユーザの声質によって, 表情や相槌などシステムの応答の傾向が変化する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.

謝辞

本研究の一部は情報処理振興事業協会 (IPA) 独創的情報技術育成事業「擬人化音声対話エージェント基本ソフトウェアの開発」, および文部省科学研究費補助金 (奨励研究 A) 課題番号 12780270 の支援を受けた.

参考文献

[1] 牛田 他, “デジタルペット —心を持った機械達—,” 情報処理学会 会誌, Vol.41, No.2, pp.127-136 (2000-02).

[2] 川本 他, “音声対話システムにおける擬人化エージェントの挙動の数理的モデル,” 情報処理学会研究報告, 2000-SLP-32-13, pp.63-68 (2000-07).

[3] Reynolds D. A., “Speaker Identification and Verification Using Gaussian Mixture Speaker Models,” Speech Communication, vol.17, pp.91-108 (1995-08).

[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,” 電子情報通信学会技術報告, PRMU97-282 (1998-03).

[5] 情報処理振興事業協会 独創的情報技術育成事業「日本語ディクテーション基本ソフトウェアの開発」.

[6] 山崎 他, “韻律信頼度を用いた漸次的音声認識出力の検討,” 平成 12 年電気関係学会北陸支部大会講演論文集, F-37, p.338 (2000-09).

[7] 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).