自律的行動決定モデルに基づくインタラクティブキャラクタ

牛田博英　平山裕司　中嶋宏
オムロン株式会社 新事業開発センタ

1. はじめに

筆者らは人間と機械の新しいコミュニケーション形態として，音声や身振りなど人間にとって自然なインタフェースを利用し，楽しみや親しみなど精神的価値をもたらすコミュニケーションを目指し研究を行っている。このようなコミュニケーションを実現するために，人間の心の働きを模擬する心のモデルを機械に持たせることを提案し，心理学理論に基づく自律的行動決定モデルを開発中である[1]。筆者らは，本モデルを感情を価値判断に用いる心のメカニズムと，反射と熟考のプロセスを処理する意識のメカニズムという２つのアプローチによって開発していることから，MaCモデル（Mind and Consciousness model）と呼んでいる[2]。これまで，触覚センサを中心とした入力機能を備える仮想生物やペットロボットにMaCモデルを適用してきた[1][2][3]。本稿では，音声で対話できるようにMaCモデルを発展させ，さらに身振り認識や顔識別などの機能を付加して，ユーザとインタラクションするキャラクタを試作した例について報告する。

2. インタラクティブキャラクタの概要

インタラクティブキャラクタは，図1に示すように3次元CGで表現された人の姿をしたキャラクタで，名前をステラと呼ぶ。ステラの目的は，ユーザの好きな音楽ジャンルの曲をユーザの指揮に合わせて演奏することであり，ステラはユーザ回答に応じてユーザへの質問を変更することにより，この目的を達成する方向に自律的に対話を進める。また，対話状況に応じて感情を表出することで自分の持つ価値観をユーザに伝える。ユーザの好きな音楽ジャンル名を対話により獲得すると，そのジャンルの曲をMIDIデータを利用して演奏を開始する。演奏曲のテンポと強さをユーザの指揮に合わせて制御でき，3次元CGでステラにキーボードを演奏させることにより，ユーザと協調して演奏しているように表現することができる。このように本キャラクタでは音声など自然なインタフェースを利用すると共に，感情表出や音楽演奏を用いるコミュニケーションにより楽しみや親しみの提供を試みる。

3. キャラクタの入出力機能

3.1. 入力機能

1) 音源方向検出 [3]

キャラクタが表示されるディスプレイの左右に設置した2本のマイクから入力される音声の位相差とマイク間の距離をもとに音源の座標値を計算する。

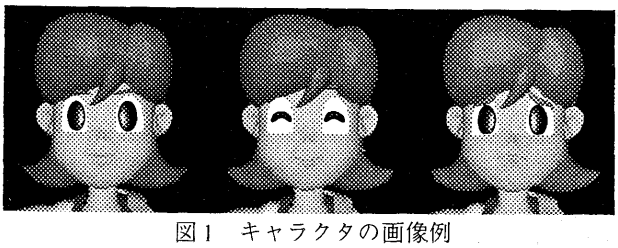

図1　キャラクタの画像例
（左から中立，喜び，困惑の各感情表現を示す）

"An Interactive Character based on an Autonomous Behavior Model",
Hirohide Ushida, Yuji Hirayama, Hiroshi Nakajima,
New Business Development Division HQ, OMRON Corporation.
〒617-8510 長岡京市下海印寺
Phone: 075-953-3880, Fax: 075-952-0411
{ushida, hirayama, nak}@zoo.ncl.omron.co.jp

2) 音声認識

米国Nuance社[4]が開発した音声認識エンジンを用いている。本認識エンジンでは，あらかじめ登録した文章を認識することができる。

3) 顔画像検出 [5]

ユーザの顔方向にステラの顔や視線を向けるためにCCDカメラから取り込んだ画像中からユーザの顔を検出する。顔の特徴点の相互の位置関係を表現したグラフを照合することで顔を検出する。

4) 顔画像識別 [5]

ユーザを顔画像で識別するために前記の顔画像検出から得るグラフ形状の類似度を計算して顔画像を識別する。識別された顔画像には顔識別番号が割り当てられる。

5) 身振り認識 [6]

ユーザの意図を理解するために，時系列の画像に対して連続DPに基づくスポッティング認識を用いて8種類の身振りを認識できるようにした。

6) 振動認識

加速度センサを内蔵した指揮棒を用いて，演奏に対するユーザの指揮の振りの速さと強さを計算する。

3.2. 出力機能

1) 3次元CG

3次元CGを利用して，表情や身振りによる感情表現，音声出力と同期した口の動き，ユーザや物音に対する視線や顔の向きの追従動作，および楽器演奏の表現を行うことができる。

2) 音声出力

録音した音声ファイルを複数用意しておき，後述するMaCモデルの行動選択部で選択した音声を再生する。

3) 音楽演奏

振動認識から得られる指揮の特徴をもとに，演奏する音楽のテンポと音の強さを制御する。複数の曲がMIDIデータとして保存され，対話から得られるユーザの好みとキャラクタの気分に応じた曲が選択される。

4. MaCモデル

MaCモデルの構成を図2に示す。図2において，楕円は記憶モジュールを表し，長方形はタスク処理モジュールを表す。モジュール間の矢印はデータの流れる方向を示す。以下，MaCモデルの主な機能について述べる。なお，前節で説明した入力機能はMaCモデルの感覚器と認識部に，出力機能はMaCモデルの行動生成部に含まれる。

4.1. 意図理解

理解部の機能は２つある。１つは，音声と身振りを入力するユーザを顔画像識別で得られる顔識別番号のユーザとみなす機能であり後述の学習で利用する。もう１つは，音声または身振りの認識結果，および音声や身振りが入力される直前のステラの発話をもとに，ユーザの意図を理解する機能である。例えば，「音楽演奏の指揮をしたいですか？」というステラの質問に対して，ユーザが「はい」と音声で答えるか，首を縦に振ると，理解部は「ユーザは指揮をしたい」と解釈して状況記述に書き込む。

4.2. 感情生成

感情生成部はステラの感情，気分，および評価型態度[7]を生成する。

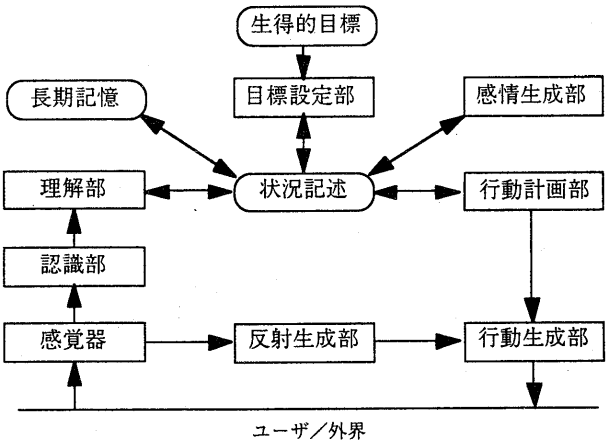

図2　MaCモデルの構成

感情は状況の評価結果から直接生じる短期的な内部状態である。状況記述に書き込まれた理解部の結果をもとに目標が達成されたかを評価し，その評価結果に応じて感情が生成される。例えば，目標が達成されると喜び，紛糾すると悲しみや怒りの感情が起こる。その他に，困惑と驚きの感情も生成する。感情は表情や音声によりユーザへ提示されるほか，価値判断機構として目標設定に影響する。

気分は感情から生じる長期的な内部状態であり上機嫌，不機嫌，落胆の3種類を設けた。気分は感情と相互に影響する。例えば，喜びと上機嫌，悲しみと落胆は，それぞれ相互に促進し，喜びと落胆，悲しみと上機嫌は相互に抑制する。また気分は演奏曲の選曲に影響する。

評価型態度として，ユーザがステラの目標達成に貢献すると，そのユーザを好きになり，ユーザが目標を紛糾させると嫌いになるように設定した。評価型態度は行動に現れ，例えば好きなユーザには親しい言葉で話しかける。評価型態度は顔識別番号と組にして長期記憶に保存される。

4.3. 目標設定

MaCモデルが扱う目標には生得的目標と経験的目標がある。本研究では生得的目標をステラの願望として，ユーザは音楽好きなこと，ユーザはポップス好きなこと，ユーザはネコ好きなこと，音楽演奏することなどを与えた。これらの達成や紛糾に応じてステラが感情表出することにより，ステラの持つ願望をユーザに伝えることができる。

経験的目標は目標設定部で状況に応じて目標項目リストから選ばれる。ステラでは最終目的である音楽演奏を達成するために，必要なユーザ情報を得るための目標がユーザとの対話に応じて選ばれる。例えば，ユーザが音楽好きだと答えると，演奏曲のジャンルを決めるために，「ユーザの好きな音楽ジャンルを知る」という目標を設定するのに対して，ユーザが音楽嫌いと答えると「ユーザが指揮したいかを知る」という目標を設定し，ユーザが指揮したいと答えると「指揮したい音楽ジャンルを知る」という目標を設定する。それぞれの目標には予め異なる重要度を設計者が与えることができ，重要度により対話戦略を変更できる。行動計画部では設定した目標，理解結果，および感情に応じた音声ファイルがステラの発話として選択される。

4.4. 学習

顔画像識別機能から得る顔識別番号と，対話から得るユーザ情報を結びつけて長期記憶に記憶する。例えば，ユーザは音楽やネコを好きか，ユーザの好きな音楽ジャンルは何か，などのユーザ情報を記憶し，同じ顔識別番号を持つユーザと再会した場合に，これらユーザ情報を用いた質問により同じユーザであることを確認する。

4.5. 反射動作

反射生成部の機能として2種類の反射動作を実現した。1つは，音源方向検知機能から得られる音源座標値を用いてCG合成されるステラの顔を音源方向に向けること，もう1つは，顔画像検出機能から得られるユーザの顔の位置座標を用いてステラの顔や視線をユーザの顔方向に向けることである。これら反射動作は，ステラが外界の物音やユーザに注意を払っていることを表現し，ユーザがステラを擬人化することを促進する。

5. 対話例

以下にユーザ（U）とステラ（S）の対話の一例を示す。

S:「こんにちは，初めてお会いする方ですか？」
U:「いいえ」
S:「それじゃあ，顔を思い出すからこっちを見てね」
（顔画像識別により得る番号からユーザ情報を検索する）
S:「ネコとポップスが好きな人ね」
U:「はい」
S:「また来てくれて嬉しいわ」（喜びの表情と声）
S:「ご機嫌いかがですか？」
U:（無言）
S:「大きな声で答えてくださいね」（困惑の表情）
S:「指揮したいですか？」
U:（首を横に振る）
S:「指揮してくれないの？」（困惑の表情）
U:「します」
S:「良かった。ポップスの曲を演奏するから指揮してね」
U:（指揮棒を振る）
S:（ユーザの指揮に合わせて曲を演奏する）

上記の例において，初対面のユーザでない場合，顔画像識別によりユーザの顔識別番号を獲得し，ユーザ情報を検索してネコとポップスを好きなユーザであることを確認する。ユーザが確認を肯定すると再会を喜ぶ。喜ぶ理由は，以前の対話でユーザがネコとポップスを好きと答えたことで，ステラがユーザを好きになったからである。次に社交的対話として機嫌を質問する。ユーザから回答がないため困惑するが，機嫌を知ることの重要度が小さいため同じ質問を繰り返さない。最終目的である音楽演奏の指揮をしたいかという質問にユーザが首を横に振ると，身振り認識を用いてユーザ意図を否定と解釈し，ステラは「指揮してくれないの？」と質問する。演奏は重要度が大きい目標なので，ユーザが一度否定してもステラはもう一度質問して目標の達成を試み，ユーザが同意すると以前の対話でポップス好きと知っているのでポップスの曲の演奏を開始する。

6. おわりに

筆者らが開発している自律的行動決定モデルに基づき音声対話などを行うインタラクティブキャラクタの試作について報告した。今後，過去の対話履歴を利用して行動決定する機能などを追加し本モデルを発展させていく。

参考文献

[1] H. Ushida, Y. Hirayama, H. Nakajima, "Emotion Model for Life-like Agent and Its Evaluation", Proc. of AAAI-98, pp.62-69, 1998.
[2] 牛田，平山，中嶋，田島，齋藤，"心のモデルに基づくインタラクティブエージェント"，第4回知能情報メディアシンポジウム予稿集，1998.
[3] 大角，工藤，齋藤，田島，"感情を持ったインタラクティブ・ペットロボット"，OMRON TECHNICS, Vol.38, No.4, pp.428-432, 1998.
[4] http://www.nuance.com/
[5] E. Elagin, J. Steffens, H. Neven, "Automatic Pose Estimation System for Human Faces based on Bunch Graph Matching Technology", Proc. of the International Conference on Automatic Face and Gesture Recognition '98, pp.136-141, 1998.
[6] 西村，向井，野崎，岡，"低解像度特徴を用いた複数人物によるジェスチャの単一動画像からのスポッティング認識"，信学論，D-II, Vol.J80-D-II, No.6, pp.1563-1570, 1997.
[7] 戸田正直，"感情"，東京大学出版，1992.