実世界情報システムプロジェクト
視聴覚研究グループ嵯峨山研究室
個性的音声対話擬人化エージェント研究ユニット

西本 卓也， 篠田 浩一，嵯峨山 茂樹
情報理工学系研究科システム情報学専攻

概 要

本研究ユニットでは，個性を持ち，高度な音声対話ができて，自律的に振舞う擬人化エージェントの実現を目標としている．5 年間のプロジェクトの 1 年目である本年度は，擬人化エージェントの基本システムを構築した．また，対話における個性を表現するための基本要素を，バーチャルキャラクタの制御コマンドと，音声対話マネージャの制御コマンドとして定義した．これらは，次年度以降の，自律的に学習し個性を持って動作するエージェントの構築における技術的なフレームワークとなる．

1 はじめに

マルチメディア通信技術の発達により，メディアを介した人間対人間のコミュニケーションは一般的になっている．しかし，時間や空間を越えるコミュニケーション手段の多くにおいては，参加者同士が場を共有しているという臨場感を感じることが難しく，ぎこちなさやコミュニケーション効率の悪さが問題になっている．

また，コンピュータ・グラフィックス技術などの発達は，メディアを介した人間対機械の知的作業を実現させている．しかし，多くのシステムには使いにくさなどの課題を依然として抱えており，より人間中心のインタフェース設計が望まれている．

これらの情報メディアにおける共通の問題は，人間対人間および人間対機械のコミュニケーションというものを画一的に捉えすぎていることである．ユーザが行おうとしているタスクを深く理解し，あるいは，個性を持つ人間同士のコミュニケーションの本質的な要素やバリエーションを深く理解することから，より快適な情報メディアを構築するための指針が得られる．

人間が一人一人持つ「個性」は，従来のコミュニケーションのモデル化においては切り捨てられていた要素であった．しかし，我々は，利用者が単に目的を達成するだけでなく，利用者に満足を与えるような機械を実現するために,「個性」に積極的な役割を持たせることを目指している．人間対人間のコミュニケーションにおいては，通信の場に臨場感をもたらし雰囲気を醸し出すための要素として，人間対機械のコミュニケーションにおいては，人間が生得したルールを自然に適用できる社会的存在として機械を扱えるための手段として,「個性」を応用することが本研究の目標である．

人間のような概観を持つシステムはヒューマノイドと呼ばれる．人間に近い顔や声を備えることにより，外見において個性を持つ機械は，すでにさまざまな手法で実現されている．しかし，擬人化システムの真の目標は，人間が対人行動を引き起こす対象となり得るような機械を実現することである．従来の「人間を模して作られたシステム」はこの点の考慮が十分ではない．

同様な立場からの従来研究としては，人間はあらゆるメディアに対して対人行動的な振る舞いをするという「メディアの等式」論があり，外見がリアルでないものの方が対人行動を引き起こしやすい，といった主張もある．しかし，人間そっくりの外見を持ち，人間そっくりの振る舞いを実時間で行うことができるシステムは未だに実現されておらず，そのようなシステムとのインタラクションにおける対人行動の現れ方や個性の役割については十分な検討がなされていない．また，それらを工学的に扱って定量的なモデルを構築する試みもほとんど行われてい

ない．

我々は，フォトリアルな顔画像と表情を持ち，リアルな動作を伴い，自律動作，反応動作などを行う擬人化エージェントの実現を目標としている．特に，高度なエージェントは個体ごとに個性が感じさせるものでなくてはならない，という観点から，対人的な振る舞いにおいて個性を感じさせる要因を明らかにし，個性の獲得につながる動作や反応の自動学習アルゴリズムの研究も目標としている．

2 Usherette バーチャルキャラクタの開発

2.1 概要

身体動作が可能な 3 次元アニメーションキャラクタの表示システム Usherette (図 1) を構築した．本システムは Linux OS 上で C++言語によって開発されており，3 次元グラフィックス・アクセラレータを有する高性能 PC 上で動作している．実行できる機能は，表情の制御（笑う，驚く，怒る，など），口の制御（「あ」「い」「う」「え」「お」の母音に対応），部分動作（手を握る，指差す，頭を振る，頭を傾ける，など），全体動作（うなずく，お辞儀をする，肩をすくめる，腕を組む，など），自律動作（眼球運動，体の細動）などである．

キャラクタの動作はすべて外部からのコマンドによって制御できる．個々のアニメーションはスムーズに連続して行うことができる．また，表情や口の制御とさまざまな動作は組み合わせて実行することができる．

2.2 動作記述とコマンド定義

Usherette におけるキャラクタの動作は，全体動作（お辞儀など），部分動作（まばたきなど），変形動作（表情，発話，など）に分類される．これらの動作記述には，シーンデータ，動作データ，コマンド定義データの 3 種類のデータを使用する．

各動作は動作系統という属性を持ち，同じ系統の動作であれば滑らかに補完することとし，異なる系統の動作であれば加算して実行される．また，繰り返し動作や，100%ではなく途中まで動作を行う，といった処理を実現するために，動作の記述データに

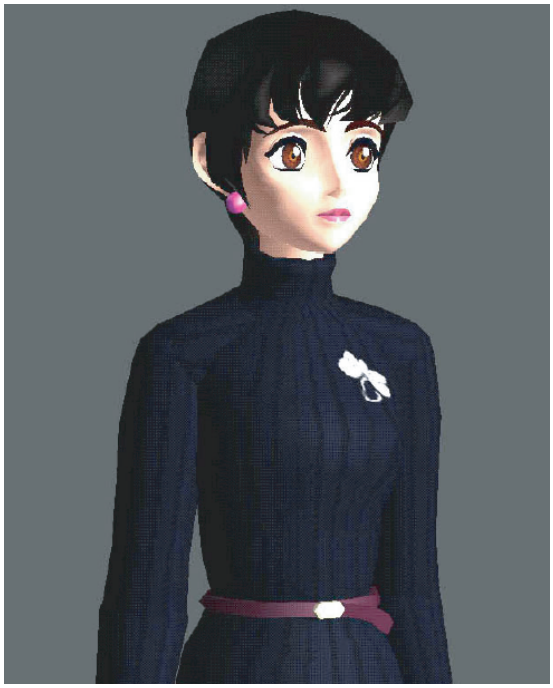

図 1: Usherette の画面

は，ニュートラル位置，ピーク（順方向，逆方向），繰り返し起点および終点，などの属性が含まれている．Usherette の動作コマンド体系を表 1 に示す．

3 Galatea DM の開発

3.1 概要

擬人化音声対話エージェントの制御にはさまざまな抽象度が考えられる．前章で述べたバーチャルキャラクタのコマンド体系は，詳細な動作を比較的詳細に表現できるが，基本的には各コマンドは時刻と対応付けて管理されるため，大局的な立場での記述や分析が困難である．これは，音楽に例えるならば電子楽器の制御情報（MIDI 信号）のようなものだと言える．

これに対して，音声対話を含むエージェントの振る舞いを，人間と機械の互いの会話のやり取りの形で記述することも必要になる．これは，エージェントの個性や会話の場における雰囲気などの詳細を割愛し，人間やエージェントが伝達したい意図にのみ注目した記述である．音楽に例えるならば，具体的な演奏情報ではなく，抽象化された楽譜情報に相当する．

このような抽象的な対話記述の手段として，我々

表 1: Usherette の動作コマンド体系

種別	機能
TOEND	動作を終了するまで（最後のニュートラルに戻るまで）1 回行う
TOPEAK	途中のピーク状態まで動作して停止
LOOP	繰り返し動作を行う
TOARG	ある値の位置まで動作
NEUTRAL	繰り返しや途中まで行われた動作を終了させる
AUTOON/AUTOOFF	自発動作の ON/OFF
MOUSEON/MOUSEOFF	マウスポインタに視線を追従させる動作の ON/OFF
SPECIAL	カメラ位置の切替，再生速度

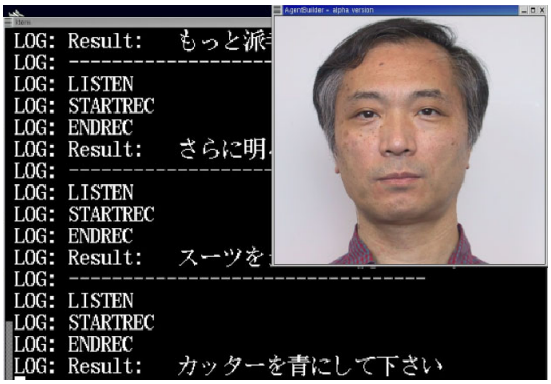

図 2: Galatea の画面

はVoiceXML をベースにした対話記述言語を検討している．本年度は，実写画像を用いたリアルな顔画像によりリップシンクや感情表現などが可能な擬人化音声対話エージェント Galatea (図 2) のための対話マネージャGalatea DM を実装し，特に VoiceXML をベースとした対話制御の基本部分を実装した．

3.2 VoiceXML の概要

VoiceXML は，音声を使ってインターネットに電話でアクセスする技術として開発された．現在は W3C (World Wide Web Consotium) によって VoiceXML 2.0 の標準化が行なわれている．VoiceXML によって，音声対話によるユーザインタフェースと，Web 上で提供されるサービスを分離することによって，一つのセッションの中で複数の音声サービスを切替えながら使用する「ボイスポータル」が実現できる．

対話制御機能としての VoiceXML は，選択肢を列挙して音声で選択させる <menu> と，スロットフィリング型の情報入力を効率良く記述する <form> の 2 つの対話要素によって成り立っている．また，対話は状態遷移によって制御され，各状態はそれぞれの対話要素に対応する．

また，各対話要素に対応する ECMAScript オブジェクトモデルが定義され，音声入力によって ECMAScript オブジェクトへの値の代入が行なわれる．

VoiceXML は，音声を用いたスロットフィリング型対話の詳細な処理を隠蔽し，「聞き返し」などの処理を効率的に記述できる．しかし，対話によって扱おうとしているコンテンツと，出力の詳細な属性や制御の詳細などのインタラクションが混在している部分もある．また，VoiceXML が目指しているのは，熟練者が慣れれば効率的に利用できる音声インタフェースの実現であり，音声操作に不慣れなユーザが自発的に会話できるインタフェースを実現するには不十分である [1]．

3.3 システム構成

擬人化音声対話エージェントのツールキット Galatea は，顔画像合成 (FSM)，音声合成 (SSM)，音声認識 (SRM) の各モジュールの機能を Agent Manager (AM) を介して呼び出す設計になっている．例えばアプリケーションが「発話を行なう」というコマンドを AM に送ると，AM はリップシンクを考慮しながら FSM と SSM を制御する．その詳細はアプリケーションからは隠蔽される [2]．

我々は，この AM の機能のさらに上位のレイヤーとして対話マネージャ(DM) を実装した [3]．我々の DM は VoiceXML を解釈し，PDOC (Primitive Dialog Operation Commands) と呼ばれる独自の形式に変換する．実際の対話は PDOC 形式に基づいて実行

表 2: PDOC で用いられる要素

要素	機能
<state>	状態を定義．<cmd>，<catch> を含む
<cmd>	コマンド列．<cmd>，<add>，<next> ，<goto> を含む．
<add>	出力キューに追加．<voice> <break> <face> を含む．
<script>	スクリプトの実行
<next>	次の状態遷移先を設定
<goto>	出力キューの内容を破棄して次の状態に遷移する
<catch>	イベント駆動型の入力処理

される．

DM を構成する機能を下記に示す．

- 前処理部: VoiceXML を解釈して PDOC に変換する．
- 対話実行部: PDOC で記述された対話を実行し，状態遷移を管理する
- 出力キュー管理部:対話実行部からコマンドを受け取り，適切なタイミングで入出力部に送る．
- 入出力部: AM との通信を行なう．出力キュー管理部から受けとった出力コマンドを実行し，入力を処理する．

対話実行部が処理する PDOC は，VoiceXML と同様に状態遷移と ECMAScript に基づいているが，状態をより詳細に分割し表 2 のような要素を定義した．

3.4 タスク記述

Galatea DM のための試作タスクとして，目的指向の対話タスクとパフォーマンス指向のシナリオを VoiceXML によって記述した．前者はレストラン情報の検索タスクであり，後者は喜怒哀楽を含むモノローグのシナリオである．これらのタスクが Galatea DM によってスムーズに実行できることを確認した．

4 まとめと課題

我々は擬人化エージェントの基本システムとして，Usherette および Galatea DM の 2 種類のシステムを構築した．また，対話における個性を表現するための基本要素を，バーチャルキャラクタの制御コマンドと，音声対話マネージャの制御コマンドとして定義した．

今後は，複合的な動作の加算や移行を考慮した Usherette の制御モデルを，音声対話を前提とした Galatea DM のコマンドに統合する予定である．また，対話における個性を表現するためのパラメータを，対話タスク (VoiceXML) と分離して記述する手法について検討する．

また，自然で効率的な対話を実現するためには，対話制御に確率モデルの考え方を導入し，実際に行なわれた対話履歴などによるモデル学習手法について検討する必要がある．

次年度以降は，拡張 VoiceXML 記述定義による柔軟な対話制御記述を実現し，対話者の動作に反応するエージェントの自律動作とその自動学習について検討する．また，エージェントとロボットが統一的個性を持つようなメディア環境を構築する予定である．

参考文献

[1] 西本 卓也. 音声インタフェースは本当に人に優しいか? 人工知能学会研究会資料, Vol. SIG-SLUD-A202-05, , 2002.

[2] 嵯峨山 茂樹, 西本 卓也, 他. 擬人化音声対話エージェントツールキット Galatea. 情報処理学会研究報告, Vol. 2002-SLP-45-10, , 2003.

[3] 嵯峨山 茂樹, 西本 卓也. 擬人化音声対話エージェント Galatea のための VoiceXML 処理系. 第 17 回人工知能学会全国大会, 2003（予定）.