社団法人　情報処理学会　研究報告
IPSJ SIG Technical Report

2003－ＳＬＰ－49　(56)
2003／12／19

# 擬人化音声対話エージェント基本ソフトウェアの開発プロジェクト報告

嵯峨山茂樹 *1　伊藤克亘 *2　宇津呂武仁 *3　甲斐充彦 *4　小林隆夫 *5
下平　博 *6　伝　康晴 *7　徳田恵一 *8　中村　哲 *9　西本卓也 *1
新田恒雄 *10　広瀬啓吉 *1　峯松信明 *1　森島繁生 *11　山下洋一 *12
山田篤 *13　李晃伸 *14

*1 東大　*2 名大　*3 京大　*4 静岡大　*5 東工大　*6 北陸先端大　*7 千葉大　*8 名工大　*9 ATR
*10 豊橋技科大　*11 成蹊大　*12 立命館大　*13 ASTEM　*14 奈良先端大
E-mail: sagayama@hil.t.u-tokyo.ac.jp

**あらまし**　擬人化音声対話エージェントのツールキット "Galatea" の開発プロジェクトについて報告する．Galatea の主要な機能は音声認識，音声合成，顔画像合成であり，これらの機能を統合して，対話制御の下で動作させるものである．研究のプラットフォームとして利用されることを想定してカスタマイズ可能性を重視した結果，顔画像が容易に交換可能で，音声合成が話者適応可能で，対話制御の記述変更が容易で，更にこれらの機能モジュール自体を別のモジュールに差し替えることが容易であり，かつ処理ハードウェアの個数に柔軟に対処できるなどの特徴を持つシステムとなった．この成果はダウンロード可能となっており，一般に無償使用許諾している．
**キーワード**　擬人化エージェント，音声対話システム，顔画像合成，ソフトウェアツールキット

# Development of Anthropomorphic Spoken Dialogue Agent Toolkit

Shigeki Sagayama *1 , Katsunobu Itou *2 , Takehito Utsuro *3 , Atsuhiko Kai *4 ,
Takao Kobayashi *5, Hiroshi Shimodaira *6, Yasuharu Den *7, Keiichi Tokuda *8,
Satoshi Nakamura *9, Takuya Nishimoto *1, Tsuneo Nitta *10, Keikichi Hirose *1,
Nobuaki Minematsu *1, Shigeo Morishima *11, Yoichi Yamashita *12, Atsushi Yamada *13, and
Akinobu Lee *14

*1 Univ. of Tokyo　*2 Nagoya Univ.　*3 Kyoto Univ.　*4 Shizuoka Univ.　*5 Tokyo Inst. of Tech.
*6 JAIST　*7 Toyohashi Univ. of Tech.　*8 Chiba Univ.　*9 Nagoya Inst. of Tech.　*10 ATR
*11 Seikei Univ.　*12 Ritsumeikan Univ.　*13 ASTEM　*14 NAIST
E-mail: sagayama@hil.t.u-tokyo.ac.jp

**Abstract**　This paper describes the outline of "Galatea," a software toolkit of anthropomorphic spoken dialog agent developed by the authors. Major functions such as speech recognition, speech synthesis and face animation generation are integrated and controlled under a dialog control. To emphasize customizability as the dialog research platform, this system features easily replaceable face, speaker-adaptive speech synthesis, easily modification of dialog control script, exchangeable function modules, and multi-processor capability. This toolkit is ready for download with an open-source and license-free policy.
**Key words**　anthropomorphic agent, spoken dialog system, animated face image, software toolkit

## 1. はじめに

著者らは，擬人化音声対話エージェントのソフトウェアツールキット "Galatea" を開発し，2003年8月より公開している．本成果は商業利用も含めた無償使用を認めている [1] [2] [3] (注1)．

近年，音声対話エージェントの研究が盛んになって来ており，ツールキットの開発もなされている [4]～[8] が，多くはコンピュータグラフィクスで顔画像を合成し，手作業で組み立てた仮想人物であり，いわばゲームソフトの1キャラクタのように，その容貌や合成音声を変えることは容易でない場合が多い．また，音声認識，音声合成，顔画像生成などのモジュールまで含み，それらをオープンソースで，無償使用できるツールは，少なくとも日本語については存在しなかった．このような途上の技術を用いてさまざまな音声言語対話の用途を開発するためには，カスタマイズ可能性やソースレベルでの変更可能性が重要である．これが本ツールキットを開発した主要な動機である．

本報告では，開発の経緯とシステムの特徴，各サブモジュー

(注1)：http://hil.t.u-tokyo.ac.jp/~galatea/ にダウンロード方法やサポートなどの情報を掲載している．

図 1　Galatea システムとの対話風景

ルの詳細と今後の予定について述べる．

## 2. プロジェクトの概要

### 2.1 開発の経緯

この計画の源流は，1994 年の情報処理学会 音声言語情報処理研究会の発足時に行われた「なぜ音声認識は使われないか」と題する議論である．この中で，音声認識性能の向上だけでは音声認識技術利用が進まないのではないかとするさまざまな意見が出された．その後，同研究会の「マルチモーダルツールワーキンググループ」(1998-2000) で，今後の音声研究者の研究目標を議論し，次世代の研究ターゲットとして擬人化エージェントを構想し，その研究プラットフォームを研究者の共同作業により構築して公開する計画を持った．この構想は，2000〜2002 年度に情報処理技術振興協会 (IPA) の支援を受け，主に大学の十数名の研究者が協力体制を作って実行した．

開発は音声認識，音声合成，顔画像合成，統合の各要素ごとに分担して行われた．第 1 年度 (2000 年度) には各要素の機能と通信に関する仕様を作成し，グループごとに開発を行なった．第 2 年度 (2001 年度) には，全てのグループが集合して合宿での作業を行ない，仕様に従って開発されたサブモジュールを結合させて Linux 環境での動作確認を行なった．その後，第 3 年度 (2002 年度) にかけて，各要素の改良を行なうと同時に，特に統合制御機能の実装や Windows 環境への移植などを行なった．さらにバグ修正や動作検証などを行ない，2003 年 8 月に Linux 版および Windows 版のツールキットを公開した．

2003 年 11 月末の時点で登録ユーザは 300 人を越えている．学会発表だけでなく，マスメディアや検索エンジンなどで本ツールを知ったユーザからのダウンロードも多い．利用目的を問うアンケートを実施したところ，個人の趣味での利用から大学・企業における研究開発まで多岐に渡っている．システムの改造や具体的な応用に関する問い合わせ等も増えている．本プロジェクトが目指した音声対話エージェント研究の出発点としての役割は達成されつつあると考えられる．

### 2.2 Galatea の全体構成と特徴

本ツールキットは，音声認識，音声合成，顔画像合成の 3 基本機能を統合し，対話制御のもとでユーザと対話するエージェント，およびその開発環境を提供するものである．特徴としては

- 高いカスタマイズ可能性 (顔，合成音声，認識文法，対話制御など)．
- 標準化動向に対応 (VoiceXML, JEIDA-62-2000 など)．
- 簡明なモジュール間通信．部品交換が容易．モジュール別に別々の PC に分散して実行可能．
- 最新の高度な技術内容を実現．特に，初の無償の日本語テキスト音声合成システムが含まれている．
- ソース公開，無償での使用許諾．

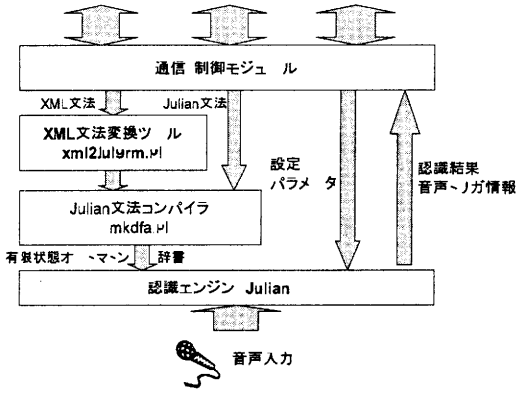


図 2　音声認識モジュールの構成

などを挙げることができる．

本ツールキットは Mobile Pentium III 1.2GHz の CPU と 512MB のメモリを搭載したノート PC 1 台での動作が確認されている．システム画面，およびシステムとの対話風景を図 1 に示す．

## 3. 音声認識モジュール

### 3.1 開発構想

音声対話システムのための音声認識システムには，認識性能が高精度かつ高速であるだけでなく，探索用パラメータの制御，認識処理の中断・再開，結果の逐次出力，複数文法の切り替え，および出力形式の選択などの様々な機能が求められる．

### 3.2 技術開発内容

対話システム用音声認識モジュールとして基本的に以下の機能を実現した．

- 文法に基づく音声認識
- 発話中の逐次的な認識結果出力
- 認識処理の動的な制御（中断，文法切り替えなど）

音声認識モジュールの構成を図 2 に示す．音声認識モジュールは音声認識エンジン，通信・制御モジュール，文法変換モジュールの 3 つのサブモジュールからなる．

文法切替えは，通信・制御モジュールを経由して随時行うことができる．認識処理を行っている間は，送られてきた文法はすべて順にキューに入れられて認識が続行され，文法の切り替えタイミングは音声入力の切れ目ごととなる．上述のような逐次出力，出力形式や文法の切替えなどは，全て外部インタフェースを通じて随時設定を行うことができる．

文法仕様としては Julian [9] 形式と XML 形式の 2 種類の文法・辞書の記述形式に対応する．本モジュールで新たに導入した XML 形式による文法・辞書の記述仕様は，W3C による Speech Recognition Grammar Specification の仕様案を参考にしており，基本は文献 [10] における XML 形式の文法の仕様に準じている．すなわち，文法・辞書の記述は主に “トークン (token)” 及び “書き換え規則 (rule)” の並びの定義及び参照からなる．単語の読み情報を付与するため，token タグに phoneme 及び syllable の属性を独自に拡張している．図 3 に記述の一部の例を示す．

### 3.3 今後の課題

今回開発した機能以外の対話システム用の機能である，認識結果の棄却，不要語やポーズへの対応などの機能の実装は今後

```
<rule id="siritai">
<one-of>
<item><token phoneme="sh;i;r;i;t;a;i;">
  知りたい</token></item>
<item><token phoneme="k;i;k;i;t;a;i;">
  聞きたい</token></item>
</one-of>
</rule>
```

図 3　XML 形式による音声認識文法の記述例

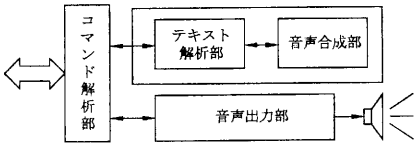


図 4　GalateaTalk の構成

の課題である．

## 4. 音声合成モジュール

### 4.1 開発構想

音声対話システムを構成するための音声合成システムには，様々な韻律や声質での音声合成，韻律や声質の外部からの制御，発話を中断する出力制御などの機能が求められる．顔画像出力も伴うマルチモーダルな出力生成では，さらに，他出力モジュールとの同期も必要となる [11]．これまでにも様々な音声合成システムが開発されているが，音声合成を用いた音声対話システムを構築しようとしたときに，使い勝手の良い音声合成システムは見当たらないのが現状である．そこで，本プロジェクトでは，日本語 TTS (Text-to-Speech) システムとしても動作し，対話音声合成に必要な機能を持った音声合成システムをフリーウェアとして開発することを目指した．

### 4.2 技術開発内容

Galatea における対話音声合成モジュール (GalateaTalk) は，四つのモジュール，コマンド解析部，テキスト解析部，音声合成部，音声出力部から構成されており，図 4 の構成をとる．

音声合成部では，HMM に基づいた音声合成 [12] [13] [14] により，合成波形を生成する．音声合成部で必要となる話者の音響モデルとして，男女各 1 名の基本話者のモデルと，平均的な声質を男声話者 1 名の声で適応化を行なった男声話者モデルの合計 3 話者のモデルが提供されている．テキスト解析部は，すでに開発されていた “茶筌 [15]” を解析エンジンとして用い，アクセント情報を付加した辞書を新たに作成することによって実現した．辞書のエントリ数は約 23,000 である．アクセント情報については，先行研究 [16] に基づいたアクセント処理を行なうために，アクセント型，結合アクセント価，アクセント結合様式が必要に応じて与えられている．“茶筌 [15]” で形態素解析したのち，序数詞などの読みを修正した後，アクセント処理を行ないアクセント句を決定する．

### 4.3 GalateaTalk の機能

GalateaTalk は，日本語テキスト音声合成を行う基本的な機能

(1) 形態素解析
(2) 読み，アクセント型情報の付与
(3) 韻律生成
(4) 合成波形生成
(5) 合成音声出力

```
<SPEECH> <VOICE OPTIONAL="male1">
これは<PRON SYM="アイピーエー">IPA</PRON>のプロジェクトで
開発された<EMPH>対話</EMPH>音声合成システムです.
</VOICE> </SPEECH>
```

図 5　JEIDA-62-2000 による発話文の記述例

に加えて，顔画像生成をともなう音声対話システムを構成するための音声合成モジュールとして，以下の機能

(6) 出力発話 (合成音声) における各音素の継続時間長の出力
(7) 埋め込みタグによる韻律の制御
(8) 音声出力の途中停止，および中断における既出力音素列の出力

を持つ．(6) は顔画像出力における口唇の動きと合成音声を同期させるために用いられる．

GalateaTalk では，発話文の表現形式として，プレインテキストによる漢字仮名混じり文に加えて，(7) の機能を実現するために，(社) 日本電子工業振興協会「日本語テキスト音声合成用記号の規格 (JEIDA-62-2000)」[17], [18] におけるテキスト埋め込み制御タグおよび仮名レベルの韻律記号に準拠したタグ付きテキストを受け付ける [20]．この記述例を図 5 に示す．GalateaTalk による合成が行なえる WWW のサイト [19] を設けている．興味のある方は是非試されたい．

### 4.4 今後の課題と計画

GalateaTalk では，読みやアクセント型の決定精度が不十分，音声出力が開始できるまでの処理時間が長い，などの問題があり，今後改善を図る必要があると思われる．

## 5. 音声合成のためのテキスト解析

### 5.1 開発構想

日本語音声合成システムにおいて，任意の日本語テキストから合成音声を生成するためには，入力テキストを解析して，当該文脈における現代日本語の標準的な読み方やアクセント情報を付与する必要がある．さらに，電子化テキストに見られる算用数字についても，その用法に応じた読み方が自動的に与えられることが望ましい．このような目的のもとで，入力テキストに対して読みやアクセント情報を付与するプログラムを開発し，辞書を整備した．

### 5.2 技術開発内容

入力テキストを形態素解析することにより，各語に読みを与えることとした．形態素解析においてはタギングと品詞の認定が主務であるため，我々の目的に照らし合わせて，形態素解析だけでは解決しない部分を前処理及び後処理で対応した．

#### 5.2.1 前処理

算用数字列をその用法に応じて，位取りの有無を区別した上で，漢数字列に変換することにより，算用数字が正しく読まれるようにした．

#### 5.2.2 形態素解析辞書の構築

音声合成に必要となるのはいわゆる読み仮名ではなく，実際に発声される音に関する情報 (発音形) である．このために発音情報およびアクセント情報を格納した辞書を新規に開発した．形態素解析プログラムとしては「茶筌」を用い，これを「茶筌」用の辞書として実装した．辞書には各語 (第 1 層語) 毎に，発音形，アクセント型，アクセント結合型，音韻交替型，音韻交替結合型に関する情報を格納している．また，後段の処理におけるハンドリングを容易にするため，出力を XML 形式にした．

#### 5.2.3 後処理

形態素解析結果を品詞情報に基づいて部分的にまとめあげる

ために，ルールベースのチャンカを作成した．チャンカで纏め上げられた単位 (第 2 層語) で，音韻交替処理を行う．これは，前接，後接の語のもつ音韻交替型，音韻交代結合型の情報をもとにして，発音形の音韻交替を行うものである．具体的には数詞，助数詞に関してこれらの型を整備して，数詞と助数詞の接続に伴う発音形の変化に対応できるようにした．このために開発したプログラムが ChaOne である．

### 5.3 成果物の説明

算用数字の読み方に関しては，設定ファイルでデフォルトの挙動を設定可能になっている．算用数字連続を位取りして読むか否か，英字連続を英単語として読むか否か，ISO 8601 に基づく日付表記を読むか否か，ISO 8601 に基づく時刻表記を読むか否かを設定できる．辞書は茶筌でコンパイル可能なので，ユーザが自ら新たな語を登録することができる．音韻交替プログラム ChaOne で用いる各種データも XML ファイルの形で宣言的に記述されているので，ユーザが自由に手を加えることができる．

### 5.4 今後の計画と発展，応用

現状では，同表記で異なる発音をするものに対する手当てが不十分である．この読み分けの問題については語形選択機能を実現することで対処する予定である．また，ルールベースのチャンカから，SVM ベースのものに変更することにより，第 2 層ベースの処理が確実に行えるようにする．辞書の拡充も今後の課題である．

## 6. 顔画像合成モジュール

擬人化音声対話エージェントシステムにおけるエージェントの自然かつリアルな表情表出や振る舞いは，対話者の直接的な印象として与えられる為，システムの一機能である顔画像合成の重要性は非常に大きい．本稿で述べるエージェント顔生成技術の最終目標は，人間の顔のみならず表情や印象も含めた人間らしさを追及する点であり，この目標に向け研究が進められている．顔合成技術は大きく次の 2 つに分類できる．1 枚の顔画像からエージェントの顔 3 次元モデル化を行う「エージェント生成ツール」，そしてこのツールで生成エージェントを用いた対話ツール「顔画像合成モジュール」である．

### 6.1 エージェント生成ツール

予め用意した正面顔画像からエージェントを生成する．このツールには 3 角形ポリゴンで構成された標準顔モデルが含まれており，このモデルを変形させ顔画像に整合させて個々の幾何モデルを生成する．整合の際，正面画像に関して，顔の各部位に存在する複数の特徴点を動かすことで，1 点ごとに顔モデルの頂点を移動することなく短時間で整合が完了する．口内や目に関しては正面顔画像のみで表現することが難しいため標準顔モデルに付属させ，口形状・目の変化に対応できるよう考慮されている．

### 6.2 顔画像合成モジュール

ツールによって生成されたエージェントはシステム統合部からさまざまな動作要求に答えられるよう，以下の機能が備わっている．

#### 6.2.1 表情変化

人間の顔表情は顔の各部位の動きを組み合わせることにより表現することができる．人間の顔表情を画面内のモデルに表現させるためには顔の各部分の動きを定量的に与える表情記述規則が必要である．顔の表情変化を表現する方法として FACS(Facial Action Coding System) [21] を導入している．これは，顔表面に現れる顔面筋の位置及び動きの方向を解剖学的に考慮した表情記述方法である．FACS は解剖学的に分類された 44 種類の運動単位 AU(Action Unit) から成り立ちこの AU の組み合わせにより様々な表情を表現することが可能とされている．この AU の移動量および移動方向をパラメータとして 3 次元モデルを変形させ表情合成を行う．表情変化は 3 次元モデルの各格子点を AU の強さによって移動させる．

この基本動作を組み合わせ 6 つの基本感情（怒り，喜び，悲しみ，驚き，嫌悪，恐れ）を用意した．またユーザが独自の表情を追加定義できるよう配慮されている．図 6 に表情合成結果の一例を示す．

#### 6.2.2 口形状変化・リップシンク

発話時の口の形状を規定する口領域の変形パラメータを表現するためには AU とは異なる口領域の変形に限定したパラメータを用いる．パラメータ化の際，擬人化音声対エージェントの話音声認識モジュールで認識できる音素すべてをパラメータとした．これらの口形状を決定するため，口形編集ツールを用いてカスタマイズを行う．口形状は基本的に 13 個の唇の厚みと形状を表現するパラメータ (Viseme) によって記述される．対応した個々のスライダーバーを制御することによって任意の 3 次元口形状を編集可能である．実際に合成される口形状が画面上から Preview 可能で回転表示もできるため，奥行き方向の形状も含めて精密な編集をインタラクティブに実現できる．図 7 に典型的な母音の口形を示す．唇は厚みを持ち，さらに先述した口内のモデル持っているため，リアルかつ微妙な口の形状表現が可能となっている．

#### 6.2.3 振る舞い・自律動作

本モジュールでは統合・制御モジュールから送信されたパラメータによって頭部の制御が可能となっている．また瞬き，首をかしげる，横に振る等振る舞いも制御でき，より自然なエージェント構築が可能となっている．またこれらの動作を繋ぎ合わせ自律動作させることが可能である．

エージェントのキャラクタ変更・追加に関しては登場させたいエージェントの数だけデータを製作し，モジュール起動前にそれらデータを読み込み，要求に応じて切り替えを行う．データ生成は先述したエージェント生成ツールを用いることで簡単に構築可能である．

### 6.3 ま と め

音声対話擬人化エージェントの実現に向けた顔画像合成に関する技術について紹介した．これら紹介した技術は本システムのみならず多方面での運用が考えられる．例えば使用した口形・表情パラメータのみを使用した顔画像通信システムである．通常，動画像通信は画像自身を圧縮しそれを相手先に伝送するが，本手法を用いることで情報圧縮の限界を目指すことも可能である．また，実写との融合を行う研究も進めており，例えば主人公の顔部分を置換して表情合成を行う手法 [22] やオリジナルの音声を認識し，機械翻訳を行い表情合成した声と再度リップシンクさせるビデオ翻訳の実現のため，口周辺部のみを実画像と置換させ合成する手法 [23] についても検討している．

## 7. 統　　合

### 7.1 プロトタイピングツール (Galatea - IB)

Windows 上で動作する Galatea MMI システムを対象に，ラピッドプロトタイピングを実現するツール Galatea-Interaction Builder (IB) を開発した [24]．入力モダリティとして，音声のほかマウスとキーボードを，また出力モダリティとして，音声 (TTS)，顔画像（表情，リップシンク）のほかウィンドウへのコンテンツ表示を使用することができる．各モジュールはソケットで接続される．

Galatea-IB は，Galatea プロジェクトで開発した各モジュー

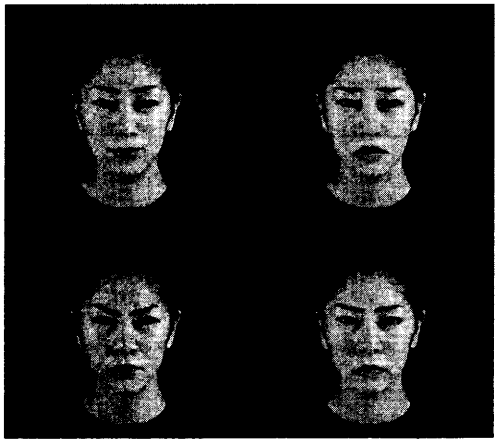

図 6　表情合成結果の一例

図 7　典型的な母音の口形

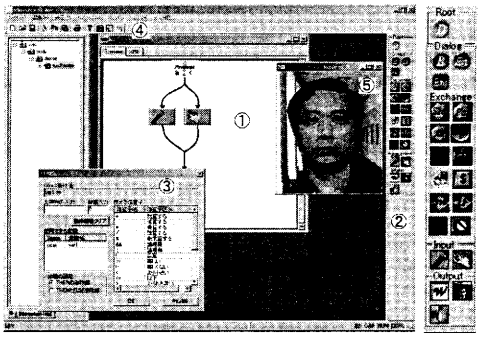


図 8　Galatea-IB の実行画面と対話部品バー

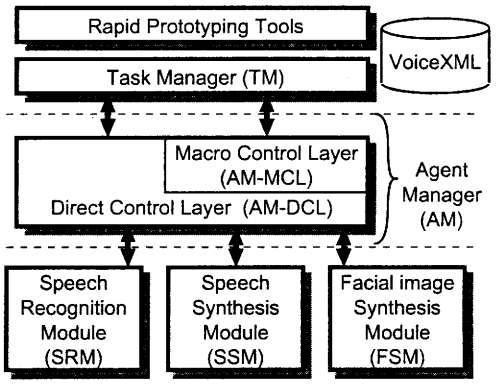


図 9　AM と各モジュールとの接続関係

ルを対話部品とし，GUI 環境でマルチモーダル対話シナリオ (MMI 記述言語 XISL [25] [26] で記述) を自動生成することを可能にする．また，製作したプロトタイプシステムをテストし，実際に動作させる MMI システム環境も提供している．

図 8 にこの動作画面を示す．Galatea-MMI システム開発者は，作成する対話シナリオに沿って，まず対話部品バー（同図 (2)，その右は拡大図）からモダリティ部品をシナリオビュー（同図 (1)）の上に drag & drop する．同時に，各部品間の連携（モダリティの動作（順次的，並行的，選択的）や割り込み可否など）を必要なら指定する．続いて，各部品（音声認識，顔画像合成，音声合成，タッチボタンなど）の属性を dialog box（同図 (3)）を開き，認識文法の指定，合成用テキストと読み上げ時の話者・速度・ピッチ，エージェント顔画像の種類（人，表情）を指定する．エージェント動作と音声合成は，dialog box 内で部品単位のテストを行うことができる（同図 (5)）．また，対話シナリオ全体の動作を確認するため，テストモードを用意している（同図 (4) はその移行ボタン）．

IPA Galatea プロジェクトで開発し提供したパッケージには，航空チケット予約案内の開発事例が入っている．

### 7.2　モジュール統合

Linux 環境では，各機能モジュールはモジュール統合処理部 (Agent Manager: AM) が統合し，タスク制御モジュール (TM) の下で動作する．

AM と各モジュールとの接続関係は図 9 に示すように，大きく分けて 2 つの機能レイヤーで構成される．Direct Control Layer (AM-DCL) は，各モジュールの規定するコマンドセットを直接制御する事を可能とするレイヤーであり，多くのモジュールはこのレイヤーを介して他のモジュールとの通信を行う．

Macro Control Layer (AM-MCL) は，主に対話タスク記述などを管理するタスク管理部 (TM) 向けのレイヤーで，良く使われる一連のコマンドセットをまとめたマクロコマンドとして再定義したり，モジュール間の同期管理などの低レベルなモジュール制御を請け負うことで，TM から見た利便性を向上させる．例えば，エージェントの音声発話時における合成音声と合成画像中の口の運動の同期 (以後 Lip Sync と呼ぶ) は，マクロコマンドとして実現した．

簡明で実現が容易なモジュール間の通信方式として，次のような方式を採用した．

- 各モジュールは，UNIX の標準入出力を通して通信する．モジュール間の通信は，必ず AM を介して行なう．
- 通信はコマンド形式で行う．
- 各モジュールは，機能をスロットとして定義した仮想マシンモデルとして扱う．

これにより，モジュールの分散並行処理ができ，モジュール追加などが容易になった．また，モジュールの単体利用・開発・試験が容易になった．今後は，各種センサなどの視聴覚情報を統合するプラットフォームとしての機能を充実させていく予定である．

### 7.3　対話マネージャ

Linux 上で動作し，音声対話システムの試作や評価実験を想定した VoiceXML 処理系 (Galatea DM) を実装した [27] [28]．これはタスク制御モジュール (TM) の一種に位置付けられる．前

```
<form id="main">
<field name="place">
<prompt> <emotion type="HAPPY">場所をどうぞ．</emotion>
</prompt>
<prompt count="3">
<emotion type="SAD">東京と京都のどちらですか？</emotion>
</prompt>
<grammar><rule><one-of>
<item><token sym="とうきょう">東京</token></item>
<item><token sym="きょうと">京都</token></item>
</one-of></rule></grammar>
</field>
</form>
```

図 10　VoiceXML による対話の記述例

述した基本モジュールに GUI 対応などいくつかのモジュールを追加し，その上で Galatea DM が動作する．VoiceXML 2.0 [29] の主要な機能を実装し，音声合成の制御のための JEIDA-62-2000 の各要素，顔表情などの制御のための emotion 要素および native 要素が使用できる．VoiceXML による対話の記述例を図 10 に示す．エージェントとの音声対話に必要な各種情報を一つのファイルにまとめて，状態遷移に基づいて記述することができる．

本システムを用いて，マルチモーダル入出力，実験者の対話への関与，ログの表示などが容易に実現できることを確認した．利用例として，擬人化エージェントを用いた音声対話における視線や表情などの視覚的な情報の役割の検討などが可能であり，身体動作が可能なエージェントへの拡張も試みている [30] [31]．

統合システムの開発や評価を通じて，「人間そっくりの機械と自然な対話をできる夢の技術」を実現するために，音声や表情の自然性，実時間性，自律性など，さまざまな研究課題が明らかになっている．今後は，ヒューマンインタフェースの改善につなげていくために，適切なエージェント制御を行なうと同時に，音声合成や顔画像合成などの機能で実現される個性の表現と利用についても検討していく．

## 8. おわりに

擬人化音声対話エージェント開発ツールキット “Galatea” の開発プロジェクトについて報告した．

今後は，情報処理学会・音声言語情報処理研究会 (SLP) を母体とする音声対話技術コンソーシアム (ISTC) において，本プロジェクトおよび連続音声認識コンソーシアム (CSRC) の成果を発展・拡充すると同時に，技術講習会やセミナーなどの活動を行なっていく予定である．

**謝辞**

本プロジェクトにおける川本真一氏と四倉達夫氏の多大な貢献に感謝する．

### 文　献

[1] 嵯峨山 他: “擬人化音声対話エージェント開発とその意義,” 情報処理学会研究報告 2000-SLP-33-1, Oct. 2000.

[2] 嵯峨山 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 2003-SLP-45-10, pp.57-64, Feb. 2003.

[3] H. Prendinger, M. Ishizuka (Eds.): Life-Like Characters – Tools, Affective Functions, and Applications, Springer, 2003.

[4] D. W. Massaro, M. M. Cohen, J. Beskow and R. A. Cole, “Developing and evaluating conversational agents,” in J. Cassell, J. Sullivan, S. Prevost, & E. Churchill (Eds.) *Embodied conversational agents*, Cambridge, MA: MIT Press, 2000.

[5] J. Gustafson, N. Lindberg and M. Lundeberg: “The August Spoken Dialogue System,” Proc. of Eurospeech99, pp.1151-1154, 1999.

[6] S. Seneff, E. Hurley, R. Lau, C. Pao, P. Schmid and V. Zue: “GALAXY-II: A Referece Architecture for Conversational System Development,” Proc. ICSLP98, pp.931–934, 1998.

[7] 土肥, 石塚: “Face-to-face 型擬人化エージェント・インタフェースの構築,” 情報処理学会論文誌, Vol.40, No.2, pp.547-555, Feb. 1999.

[8] 向井, 関, 中沢, 綿貫, 三吉: “非言語情報を用いたマルチモーダル対話インタフェースの試作,” Interaction2001, pp.139-140, 2001.

[9] 李 晃伸, 河原 達也, 堂下 修司, “文法カテゴリ対制約を用いた A*探索に基づく大語彙連続音声認識パーザ,” 情報処理学会論文誌, Vol.40, No.4, pp.1374–1382 (1999).

[10] Speech Recognition Grammar Specification for the W3C Speech Interface Framework - W3C Working Draft 20 August 2001, http://www.w3.org/TR/2001/WD-speech-grammar-20010820/.

[11] 山下洋一：“対話システムにおける音声合成”, 情報処理学会研究報告, SLP-33-4, pp.19-24 (2000).

[12] 益子貴史, 他：“動的特徴を用いた HMM に基づく音声合成”, 信学論, **J79-D-II**, 12, pp.2184–2190 (1996).

[13] 益子貴史, 他：“多空間確率分布 HMM によるピッチパタン生成”, 信学論, **J83-D-II**, 7, pp.1600–1609 (2000).

[14] http://hts.ics.nitech.ac.jp/

[15] http://chasen.aist-nara.ac.jp/

[16] 匂坂芳典, 佐藤大和：“日本語単語連鎖のアクセント規則”, 信学論, **J66-D**, 7, pp.849–856 (1983).

[17] 赤羽誠, 養輪利光, 板橋秀一：“音声合成用記号の標準化について”, 音響誌, 57, 12, pp.776-782 (2001).

[18] (社) 日本電子工業振興協会：日本語テキスト音声合成用記号の規格, JEIDA-62-2000 (2000).

[19] http://kt-lab.ics.nitech.ac.jp/~demo/gtalk/demo.php

[20] 山下洋一, 他：“マルチモーダルコミュニケーションのための音声合成プラットホーム”, 情報処理学会研究報告, SLP-40-12, pp.67-72 (2002).

[21] Ekman, P. and Friesen, W.V. , Facial Action Coding System. Consulting Psychologists Press Inc., 1978.

[22] 森島,” The Fifteen Seconds of Fame　－視聴者参加型インタラクティブ映画の提案－”, フォーラム顔学, 第 3 回日本顔学会大会予稿集, 1998.

[23] 緒方, 中村, 森島, “ビデオ翻訳システム-自動翻訳合成音声とのモデルベースリップシンクの実現-”, インタラクション ’2001, pp203-210, 2001.

[24] 佐藤, 桂田, 山田, 新田：マルチモーダル対話作成支援ツール Galatea-IB の機能強化, 情報処理学会研究報告 SLP-48-2, pp.7-12 (2003-10 月)

[25] 桂田, 中村, 山田, 山田, 小林, 新田：MMI 記述言語 XISL の提案, 情処論 Vol.44, No.11, (2003).

[26] http://www.vox.tutkie.tut.ac.jp/XISL/XISL.html

[27] 西本, 荒木, 新美: 擬人化音声対話エージェントのためのタスク管理機能, 音講論, pp.29-30, 2002-03.

[28] 西本, 嵯峨山: 擬人化エージェント Galatea のための VoiceXML 処理系, 第 17 回人工知能学会全国大会, 2C2-04, 2003-06.

[29] http://www.w3.org/TR/voicexml20/

[30] 西本, 中沢, 嵯峨山: 音声対話における擬人化エージェントの利用効果の検討, 情処研報 2003-SLP-47, pp.25-30, 2003-07.

[31] 中沢, 西本, 嵯峨山: アイコンタクト機能を持つ擬人化音声対話エージェント, 音講論, pp.43-44, 2003-09.