カスタマイズ性を考慮した擬人化音声対話ソフトウェアツールキットの設計

川本真一[†1] 下平博[†1] 新田恒雄[†2]
西本卓也[†3] 中村哲[†4] 伊藤克亘[†5]
森島繁生[†6] 四倉達夫[†6] 甲斐充彦[†7]
李晃伸[†8] 山下洋一[†9] 小林隆夫[†10]
徳田恵一[†11] 広瀬啓吉[†14] 峯松信明[†14]
山田篤[†12] 伝康晴[†13]
宇津呂武仁[†2] 嵯峨山茂樹[†14]

本論文では，擬人化音声対話エージェントを将来のヒューマンインタフェースの重要な技術要素として位置づけ，研究開発の共通プラットフォームとなりうる高いカスタマイズ可能性を備えたソフトウェアツールキットの実現を目指し，それに必要な要素とその実現技術について論じる．今後のヒューマンインタフェース技術において，コンピュータがあたかも一個の人間として振る舞い，人間の顔や姿を持ち，ユーザと音声言語で対話するようにすることは，大きな目標の１つである．このような研究開発を進めるにあたっては，多分野の協力が必要であり，研究成果を集積していくための共通プラットフォームが必要である．それには，音声認識，音声合成，画像合成，対話制御などの基本モジュールと，それらを統合制御する仕組みが必要である．さらに，個性の表現や広い応用などのためには，各モジュールは高い基本機能のみならずカスタマイズ可能性が重要である．このため，筆者らは，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，さら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るなどの特徴を持つ擬人化音声対話エージェントシステムを構想し，実装した．いくつかの簡単な対話タスクについてエージェントを試作し，必要な機能に関する達成度を確認した．

Design of Software Toolkit for Anthropomorphic Spoken Dialog Agent Software with Customization-oriented Features

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

This paper discusses the design and architecture of a software toolkit for building an easily customizable anthropomorphic spoken dialog agent (ASDA). Human-like spoken dialogue agent is one of the promising man-machine interface for the next generation. Simply combining, however, the existing software modules for speech recognition, speech synthesis, face-animation synthesis and dialogue control do not lead to a satisfying agent system as might be expected. ASDA requires more sophisticated functions of the modules than those when the modules are used independently, as well as the integration mechanism. Another problem with ASDA was that it required great customization effort for any user-system interaction task. Therefore, developing an easy-to-customize software platform for ASDA is quite meaningful, though it is still a great challenge in both research and development aspects. This paper discusses basic and essential requirements for ASDA systems, and software modules for the system are designed to fulfill the requirements. Using this software toolkit, A prototype agent system has been developed on a UNIX-based system using this software toolkit. Finally, we discuss current achievements of the toolkit.

1. はじめに

今後のヒューマンインタフェース技術において，コンピュータがあたかも一個の人間として振る舞い，人間の顔や姿を表現し，ユーザの音声言語で話し聞くような擬人化音声対話エージェントは，大きな目標の1つであり，さまざまな研究ならびに開発が進められている[1)~7)]が，人間同士の対話に比べるとまだきわめて初歩的である．人間のコミュニケーションに関連する分野はきわめて広く，心理学，自然言語処理，知識処理など多分野の多面的な協力が必要になるだろう．それらの研究促進のためにも，また成果の集積のためにも，多くの研究開発者が容易に使用・開発参加することができるような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が，共通の研究プラットフォームとして公開・提供されることが望ましい．

このようなシステムを構築するには，単に既存の音声認識，音声合成，画像合成，対話制御などのソフトウェアモジュールを組み合わせただけでは，機能性の面においてもレスポンス性の面においても不十分である．そこで，擬人化音声対話エージェントに適するように各モジュールの機能を高度化することに加えて，これらのモジュールが互いに連携して円滑に動作するような仕組みの構築が必要となる．しかし，各モジュールの機能の高度化については研究的にも未解決の問題が多いうえ，実際のシステムの開発には多大な労力を要する．

また，それには柔軟なカスタマイズ可能性が必要である．多種多様な仮想人物を作り出すには，顔，声，動き，反応などが，容易に多種多様にカスタマイズできることが必要である．従来しばしば見られた，固定した少数のキャラクタの動画像，固定した少数の音声合成，あるいは非公開のソフトウェアなどは，研究開発のための共通のプラットフォームとして大きな貢献はできない．

1つの目的（タスク）用に開発したシステムを別の目的に利用するには，設計の大幅な変更を余儀なくされる場合が多々ある．従来は，たとえば，タスクや対象ユーザが変われば，音声認識や対話管理機能の変更が多くの場合必要であった．エージェントの顔や音声を変更する際にも，しばしば当該モジュール自体の変更が必要であった．さらにエージェントの個性を重要視する場合には，個性を変更するために，顔の表情や仕草，しゃべり方などの変更が必要であった．このような問題を解決するためには，システムを複数の独立した基本要素（モジュール）の組合せによって構成し，各モジュールに高い基本機能とカスタマイズ性を持たせることが重要である．さらに，モジュールが容易に交換可能であったり，新たな機能のモジュールの追加が簡単に行えたりするとともに，モジュール単独あるいは複数のモジュールでの動作が可能であることは，今後の技術進歩を持続的に取り入れる観点からも望ましい．従来，擬人化音声対話エージェントを開発するためのソフトウェアツールキット開発の研究[8)~10)]はいくつか行われている．また，人間と機械のより自然な対話を目指した研究として，ロボットの自然な対話を目指した興味深い研究[11)]や，市販のソフトウェアを各要素技術として利用した擬人化音声対話エージェントの例もあるが[1)]，後述するような高度な機能と高いカスタマイズ性を持たせたソース無償公開のソフトウェアはまだ提供されていない．

筆者らは，擬人化音声対話エージェントを用いた研究開発のための共通のプラットフォームとなる，カスタマイズ可能なソフトウェアツールキットの開発を行っており，現在，その基本動作を確認した段階にある．本論文では，多様な擬人化音声対話エージェントを構築するためのソフトウェアツールキットとして，必要とされる機能，それを実現するための問題点と解決法

†1 北陸先端科学技術大学院大学
Japan Advanced Institute of Science and Technology
†2 豊橋技術科学大学
Toyohashi University of Technology
†3 京都工芸繊維大学
Kyoto Institute of Technology
†4 国際電気通信基礎技術研究所
Advanced Telecommunications Research Institute International
†5 産業技術総合研究所
National Institute of Advanced Industrial Science and Technology
†6 成蹊大学
Seikei University
†7 静岡大学
Shizuoka University
†8 奈良先端科学技術大学院大学
Nara Institute of Science and Technology
†9 立命館大学
Ritsumeikan University
†10 東京工業大学
Tokyo Institute of Technology
†11 名古屋工業大学
Nagoya Institute of Technology
†12 京都高度技術研究所
The Advanced Software Technology and Mechatronics Research Institute of Kyoto
†13 千葉大学
Chiba University
†14 東京大学
The University of Tokyo

について議論する．さらに，著者らが実現したシステムについて述べ，目的の達成について評価し，今後の技術的課題を議論する．

2. 擬人化音声対話エージェントへの要求条件

コンピュータがあたかも人間のように振る舞い，人間の顔や姿を表現し，ユーザの音声言語で話し聞くようなインタフェース実現のための擬人化音声対話エージェント・ソフトウェアツールキットに求められる要素について議論する．

2.1 人間らしい対話実現のための要素技術

人間同士の音声対話では，キーボード対話などではあまり見られない特有の現象，たとえば，相手の話の途中で「ええ」と相槌を打つ,「えっ？」などのように短い応答で聞き直す，最後まで聞かずに割り込むなどが見られる．また，韻律によって，感情や意図や状況（同意していないなど）を表現することが多い．これらが実現できる基本機能を人間・機械間においても実現すれば，両者のコミュニケーションを人間らしくできるのみならず，音声対話の特性を生かした効率的なコミュニケーション実現のプラットフォームとしても意味がある．

従来の音声認識では発話が終了してから認識結果を確定する方式が多く，また音声合成では対話的な制御機能を備えていないことが多く，このような用途に活用できる音声認識および音声合成のシステムは多くない．音声認識における漸次的な認識結果出力機能や，音声合成の対話的な制御機能などを実現すれば，人間・機械間のコミュニケーションの研究にも寄与することができよう．

また，音声対話に関しては，効率的で対話しやすいリズムを確保すべく，ユーザに対する素早いレスポンスの実現が必要である．さらに，エージェントとの対話の自然性を確保するために，各要素技術の高い次元でのバランスも重要である．特にシステムを構成する技術のうち，どれか 1 つでも能力が不足している場合，その不足分ばかりが目立ち，システム全体に対する印象を悪化させる可能性がある．たとえば，システムとの対話の評価・分析において性能の不足をカバーするために Wizard-of-Oz 法を利用した研究もあるが，これは，音声認識性能が不十分の場合，自然な対話現象として得たい情報が得られないため，音声認識の代用として用いるものである．これを裏返せば，システムを構成する技術のどれか 1 つの能力が不足すれば自然な対話が損なわれることを意味しており，システムとの自然な対話を実現するソフトウェアツールキットを提供するためには，それぞれの要素技術はつねに高い性能を目指す必要がある．また，一方では多くの対話システムが存在し，現状の技術をうまく組み合わせた構成が実現されている．これらの対話システムの多くは，人間同士の対話と比較するとまだまだ初歩的ではあるものの，対話システムとしての完成度が高いものも存在する．つまり，自然な対話に近づけるために要素技術としてはつねに高い性能を目指しつつ，ソフトウェアツールキットで作られる対話システムの完成度をあげるためには，それぞれの性能のバランスを考慮する必要があると考えられる．

対話における時間要素に関していくつか研究例はあるが[12),13)]，システムとして実現されているものはまだ少なく，多くの研究課題が残されている．この対話における時間要素に関して検討できるソフトウェアツールキットが実現すれば，研究開発の共通プラットフォームとして有用である．

2.2 カスタマイズが容易な構成

人間的なヒューマンインタフェースの研究の共通基盤の 1 つとしてツールキットを提供する以上，その利用者の広い要求に応えられることが望ましい．そのうちの重要な 1 つの要素は，エージェントの顔や音声のカスタマイズ可能性である．両者とも，適用分野，利用目的などにより望みどおりに容易に変えられることが望ましい．たとえば，複数の開発者がそれぞれ独自の機能を実現した擬人化音声対話エージェントが，同じ顔と同じ声を持たざるをえないとすれば，この分野の将来の可能性を制約するだろう．あるいは，多くの擬人化音声対話エージェントが混在するような環境を想定した場合，各エージェントの機能がそれぞれ異なるにもかかわらず，同じ顔や声ならば，利用者は混乱する．さらに，エージェントの機能ごとに異なる顔や声などのインタフェースを設定する際，利用者の好みに応じて容易にカスタマイズできることが望ましい．また，電子秘書などの場合，ユーザが好みに応じてエージェントの顔と声を変えられるようにすることも望ましい．

2.3 機能部品のモジュラリティ

擬人化音声対話エージェントの応用において，標準で準備されている音声認識や音声合成，顔画像合成のモジュールに代えて，一部に別のモジュールを使用したい場合がありうる．また，これらの機能モジュールに関して研究開発する場合においても，標準部品と性能や機能を比較するために，たとえば音声合成のみの入れ換えなど，部分的に差し替えられることが望ましい．また，開発した新しいモダリティなど，新機能の

モジュールや新たな入出力の追加などを扱う場合は，それらを容易に取り込むための枠組みが必要である．

また，音声認識や音声合成，顔画像合成などの要素技術研究における参照システムとしての利用を考えるとき，それぞれのシステムは単独で利用可能であることで効率的な要素技術の参照が行える．また，各要素技術の開発においても，単独利用できることは，開発効率を向上させることができるだろう．

さらに，擬人化音声対話エージェントの運用に関しても，各モジュールが分散して処理することでさらに高速，かつ効率的に動作する可能性があり，そのためには，機能単位のモジュール化とそれらを統合する枠組みが必要であろう．

以上により，機能別にモジュールを提供し，それらは容易，かつ統一的な方式で制御できることが望ましいといえる．

2.4 ソース無償公開のソフトウェア

新しいヒューマンインタフェース技術分野の発展の基盤として，擬人化音声対話エージェントのソフトウェアがソースコードも含めて，すべてが無償で公開されることは，きわめて有意義であろう．この技術は，まだまだ発展途上であるため，多くの研究開発者による改良を歓迎するものでなければならない．同時に，研究開発において，種々の応用に無償で使用できることも意義が大きい．このソフトウェアをソースコードも含めて無償公開することにより，産業応用や研究開発においてこのソフトウェアの利用・応用の促進や擬人化音声対話エージェント，および関連技術の研究開発の促進に寄与できる．

このような技術を実現するためには，既存の要素技術を高度に統合することに加えて，新たな研究開発が必要である．また，人間らしいコミュニケーション能力を備えた擬人化音声対話エージェントの実現に関連する分野には，多くの難しい研究課題が含まれており，音声情報処理技術や画像情報処理技術の研究者のみならず，知識処理や心理学などさまざまな研究分野から多くの研究者が参加して総合的に進められる学際的研究領域となるだろう．つまり，人間らしい対話実現のための研究を進めるためには，多くの分野の研究者が擬人化音声対話エージェントを使った研究に参入しやすい環境の整備が必要であり，擬人化音声対話エージェント研究のための理解と利用が容易な共通の研究開発プラットフォームの提供が重要である．

2.5 従来のソフトウェアにおける要求達成度

それぞれの要素技術が研究段階のものであり，多くの難しい問題を含んでいるため，現状では上記の要求をすべて満たすようなカスタマイズ性に優れたソフトウェアツールキットはまだ見られない．たとえば，漢字仮名混じり文を明瞭に読み上げることを実現し，ソースコードまで無償公開された音声合成ソフトウェアは，現時点では存在しない．

2.6 その他の要求条件

擬人化音声対話エージェントを具体的な対話タスクで利用するうえで不可欠である対話記述は，従来は専門外の開発者にとって必ずしも容易ではなかった．これを極力容易にするとともに，対話記述の標準化に沿って，音声対話のために書かれた対話記述の再利用や部分利用が可能な方法を模索することは，さまざまな研究者にこのシステムの利用を促すうえで望ましい．同時に，擬人化音声対話エージェントを用いて音声認識・合成の研究成果の効果の実証・検討や，研究・開発した技術のデモンストレーションを行う際，適した対話記述やデモンストレーションを作成する労力は極力省きたい．これらのために，対話タスクを容易に記述・カスタマイズできることが望ましい．

ただし本論文では，対話タスクの記述・カスタマイズに関しては議論の対象としない．

3. 構成要素モジュールの設計

先に述べた擬人化音声対話エージェント研究のための共通の研究開発プラットフォームを実現するには，ソフトウェアツールキットを構成する基本ソフトウェア（以下モジュールと呼ぶ）にはどのような機能の実現が必要かについて議論する．

擬人化音声対話エージェントのプラットフォームとして，先に議論した要件を満たすためには，少なくとも，擬人化音声対話エージェントの顔や声，対話タスクをカスタマイズ可能にするための 3 つの対話部品モジュール（音声認識モジュール，音声合成モジュール，顔画像合成モジュール）と，各モジュールの統合処理を実現するエージェント管理部が必要であり，本論文では**図 1** のような構成を想定する．

3.1 音声認識モジュール

筆者らは，対話音声にも利用可能な音声認識モジュールとして，IPA「日本語ディクテーション基本ソフトウェアの開発」（1997.4〜2000.3）において大語彙連続音声認識エンジン Julius[14] の開発や，SPOJUS[15] の開発を進めてきた．

Julius では，言語モデルとして統計的言語モデルを用いていたが，対話記述や簡単な対話アプリケーションの作成・変更を容易にするためには，文脈自由文法などの形式言語による言語モデル記述，および状況に

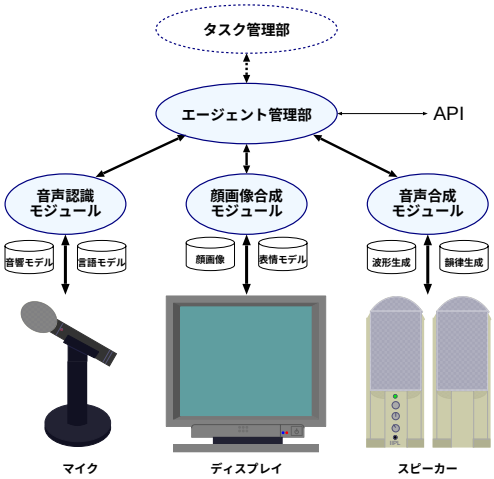

図 1 擬人化音声対話エージェントのソフトウェアツールキット
Fig. 1 Configuration of anthropomorphic spoken dialog agent.

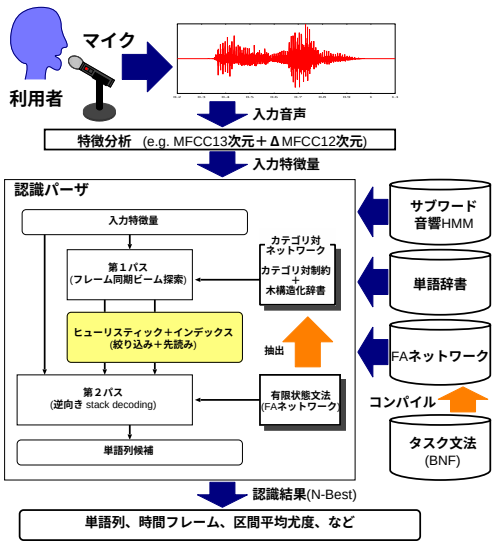

図 2 音声認識モジュール
Fig. 2 Speech recognition module.

応じた動的な言語モデルの切替えは重要である．さらに，インタラクティブな対話を実現するためには，まずはユーザ発話に対する素早いレスポンスの実現が必要である．

この問題に対処するため，Julius を基盤として有限状態文法に基づく連続音声認識パーザ Julian[16]（**図 2**）を開発し，文法の動的な切替えや漸次的音声認識結果の出力を実現できるように設計することとした．

3.2 音声合成モジュール

音声合成モジュールに求められる機能は，特定の人の声で漢字仮名混じり文を明瞭に読み上げる機能と，

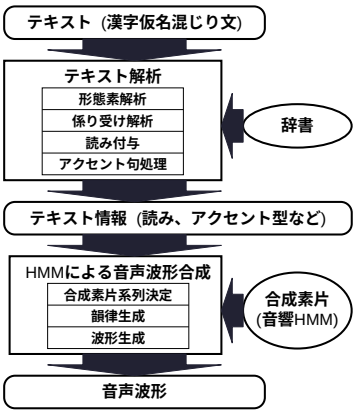

図 3 音声合成モジュール
Fig. 3 Speech synthesis module.

声および発話内容がカスタマイズ可能であることである．

特に，声（話者性）のカスタマイズ可能性を重要と考え，多様な声質や発話スタイルによる音声合成を容易に実現できる可能性がある HMM 音声合成方式[17),18)]（**図 3**）を採用する．ただし声のカスタマイズに関して，主に韻律のカスタマイズについて議論し，話者適応などに関しては本論文では議論の対象としない．

さらに対話の自然性を確保するためにエージェントの音声発話時において，合成音声と合成画像中の口および唇の運動の同期（以後 LipSync と呼ぶ）を精密に行うために，合成音声生成時の発話音素や継続長などの情報を共有し，他のモジュールとの連係をとりながら，処理を進めることとした．

ここで想定する LipSync の精度は，音素単位の同期，および音素継続長の同期である．音声合成モジュール側では，顔画像合成モジュールに提供する発話音素と音素時間長を用いて合成しているため，原理的には音素境界でのずれはない．しかし，音声の DA（ディジタル–アナログ）変換出力の開始時刻と画像の表示開始時刻のずれによる微小な同期のずれは存在しうる．

また，発話する漢字仮名混じり文の対する部分的な強調や韻律制御などの指定は，日本電子工業振興協会の作業グループによって制定された「日本語テキスト音声合成用記号の規格：JEIDA-62-2000」[19)]をベースに拡張を検討する．

3.3 顔画像合成モジュール

エージェントに任意の顔を持たせるためには，顔画像のカスタマイズが容易であることが重要である．コンピュータグラフィクス（CG）合成によるアニメーションの顔を用いる場合，複数の顔を準備するのは必ずしも容易でなく，まして相槌などの頭の動きや精密

な LipSync を実現するための 3 次元 CG となると，各ユーザがそれぞれ複数の顔を準備するのは困難であることが多い．

一方，IPA「感性擬人化エージェントのための顔情報処理システムの開発」（1995.6〜1998.3）で筆者らが開発したソフトウェア[20] では，正面方向から撮影した 1 枚の顔画像と標準ワイヤフレームモデルを整合させ，各個人の 3 次元頭部モデルを生成しているため，顔画像さえ準備すればエージェントの顔をカスタマイズできる．また，このソフトウェアはソースコードも含めて，すべてが無償公開されており，今回開発する顔画像モジュールの基盤として適している．

さらに，音声対話擬人化エージェントの顔画像合成モジュールとして必要な，以下のような機能を実現する．

- 後述するエージェント管理部（AM）で規定する対話的なコマンドに対応して，顔画像に任意の表情変形をリアルタイムに行う機能．
- 音声合成モジュールと連係し，精密な LipSync を行う機能．

また，LipSync 実現に際し，音声合成モジュールから提供される発話音素，および音素継続長の情報を利用して，以下の手順で口形状アニメーションを生成する．

(1) 音素と視覚素（Viseme）[21] の対応関係を参照し，音素に対応した口形状アニメーション用のキーフレームを準備する．本論文において基本とする Viseme の口形状は，**表 1** に示す 14 種に分類し，使用する．

(2) 受け取った音素を口形状アニメーション用のキーフレームとし，音素間の口形状は音素と対応した口形パラメータを線形補間することで各時間の口形状を定め，1ms 単位の音素継続長分のデータベースを作成する．

(3) 音声発話時には，口形状パラメータのデータベースから発話経過時間に応じた適切なパラメータを取得し，口形状の表示に反映させる．

この方法により，使用するコンピュータの性能で描画フレームレートが変化したとしても，コンピュータの性能に応じた口形状を再現することができる．

図 4 に顔画像合成モジュールの概要を示す．

3.4 エージェント管理部

音声認識，音声合成，顔画像合成などの要素技術のモジュラリティとカスタマイズ容易性を確保するため，各モジュールの通信を管理し，各モジュールの低レベルの制御を実装する必要がある．具体的には，各モジュールのインタフェースを共通化し，各モジュールの扱いを統一化することで，新たなモジュールの追加や機能拡張を容易にする．また，モジュールは独立したプロセスとして，単一のコンピュータ，もしくは複数のコンピュータ上で並行に動作することを想定し，エージェント管理部を介してモジュール間の入出力を扱うことで，モジュール間の入出力の扱いを簡単化し，モジュールの独立性を高める．さらに，エージェントの音声発話における LipSync の機能など対話の自然性を確保するために頻繁に使う機能は，低レベルな制御のマクロ処理コマンドとして提供する．

これらの各モジュールが連動して 1 つの対話システムとして円滑に動作するには，分散環境におけるシス

表 1 発話音素に対する口形状の分類
Table 1 Conversion table from phonemes to mouth shapes.

口形状 No.	音声合成モジュールの提供する音素表記
0	/h/, /y/, /cl/, /pau/, /sil/
1	/r/, /ry/
2	/m/, /b/, /p/, /my/, /by/, /py/
3	/t/
4	/n/, /d/, /ny/
5	/k/, /g/, /ky/, /hy/, /gy/, /N/
6	/f/
7	/s/, /sh/, /ch/, /ts/, /z/, /j/, /dy/
8	/w/
9	/a/, /A/
10	/i/, /I/
11	/u/, /U/
12	/e/, /E/
13	/o/, /O/

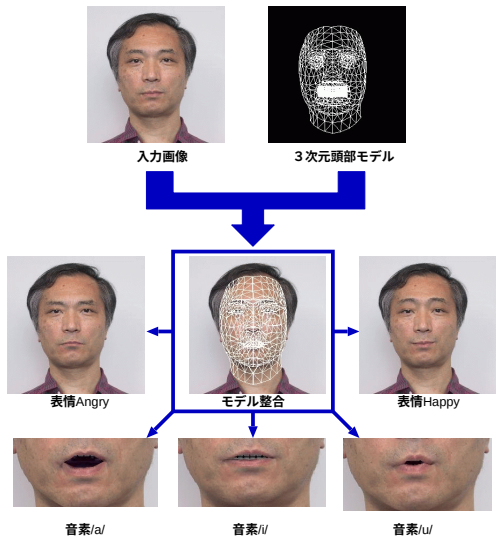

図 4 顔画像合成モジュール
Fig. 4 Facial image synthesis module.

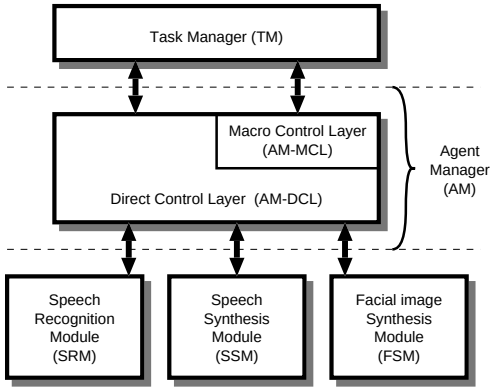

図 5 エージェント管理部と各モジュールとの基本構成図
Fig. 5 Basic configuration of agent manager and modules.

テム制御，情報管理などが必要となり，これらの開発例として，MIT で開発された Galaxy-II[8] を基礎とした DARPA のコミュニケータ・プログラム[9] におけるシステムや，SRI の開発した Open Agent Architecture（OAA）[10] などがある．このようなシステムの標準化の動向は参考にすべき点が多くある．一方，汎用的なシステム構成を実現するために，機能が複雑化・多様化することにより，システムの利用が困難になる．これらの標準化の動向に注目しつつ，最小限の変更でモジュール交換や機能追加などのカスタマイズを実現可能にするモジュール統合の設計でなければならない．

以下では，エージェント管理部の実装上の構成，各モジュールのインタフェースの設計，およびエージェント管理部と各モジュールとの入出力について述べる．

3.4.1 エージェント管理部と各モジュールとの基本構成

エージェント管理部（AM）は大きく分けて 2 つの機能レイヤで構成される．エージェント管理部と各モジュールとの基本構成を**図 5** に示す．

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ことを可能とするレイヤであり，多くのモジュールはこのレイヤを介して他のモジュールとの通信を行う．Macro Control Layer（AM-MCL）は，主にタスク管理部（TM）向けのレイヤで，頻繁に使われる一連のコマンド列をまとめたものを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ったりすることで，タスク管理部から見た利便性向上を目的としている．

音声認識モジュール（SRM），音声合成モジュール（SSM），および顔画像合成モジュール（FSM）は，原則として AM-DCL との間で通信を行う．これにより各対話部品モジュールは AM との通信のみを考慮して開発を進めることができ，開発拠点が分散化している本プロジェクトにおいて，開発の効率化に役立っている．

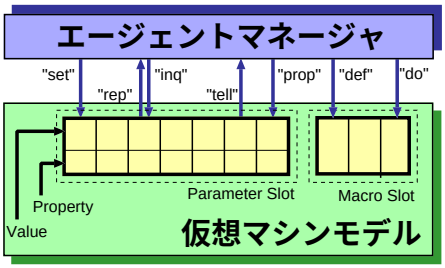

図 6 エージェント管理部と仮想マシンモデルとの関係図
Fig. 6 Communication protocol between agent manager and virtual machine model.

TM は基本的に AM-MCL との間で通信を行うが，必要に応じて他のモジュールと同様に AM-DCL との通信も行うことができる．各モジュールからのすべての出力は TM に提供される．これにより TM では，各対話部品モジュールから得られるすべての情報を必要に応じて取捨選択して利用することができる．

AM-DCL を介して通信を行う場合，以下のようにコマンドの送信先を指定することで，モジュールに直接コマンドを送信する．

```
to @送り先 送りたいコマンド
```

また，AM-MCL で提供されるコマンドに関しては，送り先は AM となるため送信先指定はない．たとえば，LipSync をともなうシステム発話を実現するマクロコマンドは以下のように送信する．

```
set Speak = 発話内容
```

3.4.2 仮想マシンモデル

各モジュールを**図 6** に示すような仮想マシンモデルとして扱うことで，インタフェースの統一化をはかる．

各対話部品モジュールは，各入出力パラメータをパラメータスロットとして管理する．また，各対話部品モジュールにおけるマクロコマンド定義も，同様にマクロスロットとして管理する．これらのスロットは，AM-DCL との間で共有され，これらのスロットを介して，AM-DCL と各対話部品モジュールとの間の通信を行う．

各スロットはそれぞれ値と属性を持ち，仮想マシンの制御パネルに配置されたスイッチに見立てて扱われる．また，各スロットは共通のコマンドにより操作される．スロットの値により，動作状態の監視や，動作の開始・終了の指示，動作環境の設定などを操作でき

表 2 仮想マシンモデルの基本操作に用いるコマンド名とその機能
Table 2 Names and functions of basic operation commands in virtual machine model.

名前	機能
set	パラメータスロットの値の設定
inq	スロットの値の問い合わせ
prop	スロットの属性の設定
save	現在のスロットの値を別名をつけて蓄積
rest	save で蓄積されたスロットの値の復帰
del	save で蓄積されたスロットの値の解放
def	マクロスロットの値の設定
do	マクロスロットの値
	もしくはファイルの内容の評価

表 3 スロット値などモジュールからの出力の際に付加される識別子名とその機能
Table 3 Identifiers preceding slot values reported from function modules.

名前	機能
rep	スロットの値の出力
tell	その他の値の出力

る．スロットの値の変更は，基本的には即座に動作に反映される．つまり，たとえば動作指示のスロット値を変更することは，即座に動作につながることを意味する．

このようなモデル化により，各対話部品モジュールの扱いを統一化することができる．コマンドはモジュールに依存しない通信仕様であり，パラメータスロットは通信には依存しないモジュール依存の仕様である．つまり，各対話部品モジュール間の違いは，準備されるパラメータスロットの機能の違いのみとなる．また，各対話部品モジュールを仮想マシンモデルとして抽象化して扱うことで，機能拡張や対話部品モジュールの追加が容易になる．たとえば，新しい機能を追加することは，仮想マシンモデルのパラメータスロットを追加することで実現できる．対話部品モジュールの追加も，他の対話部品モジュールと同様に仮想マシンモデルの考えに基づき，パラメータスロットを定義することで実現できる．仮想マシンモデルの基本操作に用いるコマンドを**表 2** に示す．スロット値などをモジュールから送信する際は，**表 3** に示すような 2 種類の識別子をデータに付加する．

3.4.3 音声合成・唇動画像合成の同期（LipSync）の実現

システムの音声発話時に連係が必要となる音声合成モジュールと顔画像合成モジュールとの同期管理を行うモジュールは，AM 以外の上位モジュールでも実現可能である．また，同期専用のモジュールを新たに定義・実装することも考えられる．しかし，システム発話は音声対話において頻繁に利用される機能であり，システムの重要な基本機能であるため，エージェント管理部が提供するマクロコマンドとして実現した．

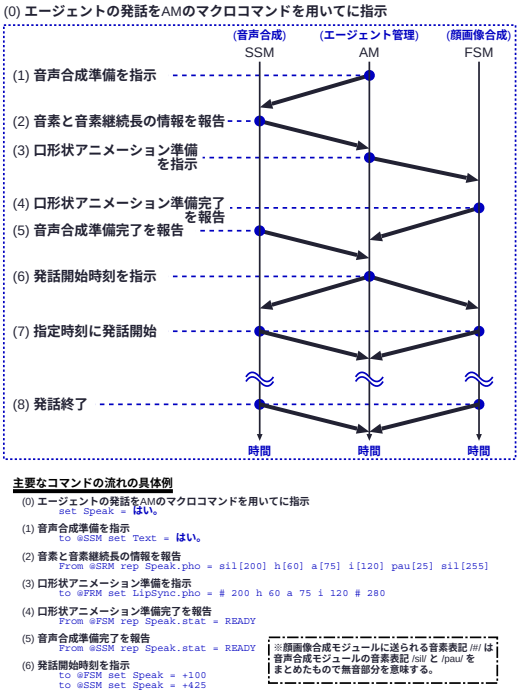

図 7 システム発話時における AM，SSM，FSM の処理の流れ
Fig. 7 Processing flow between AM, SSM and FSM when agent speaks.

システム発話時の LipSync において最低限必要な情報は，発話音声を構成する発話音素と音素継続時間である．この発話音素と音素継続時間は，音声合成モジュールに問い合わせることで取得できる．また，実際に音声合成と顔画像合成の両モジュールの発話準備が整ったかどうかを確認してから，発話を開始する必要がある．この処理は次のように実現する．エージェント管理部は，まず両モジュールに発話準備を指示する．次いで，発話準備が整ったことを示す情報が両モジュールから返信されたことを確認後，実際の発話開始を指示する．具体的なコマンドの流れを**図 7** に示す．

3.4.4 エージェント管理部と各モジュールとの入出力接続

モジュール間の入出力は，UNIX システムで使われている標準入出力を用いる簡便な方法で実装する．これによって，各モジュールは UNIX 上の 1 コマンドとして開発することができるため，個々のモジュールの単独開発・デバッグが可能になると同時に，各モジュール単独での利用が容易となった．エージェント管理部と各モジュールとのプロセス間通信は，UNIX の代表

図 8 擬人化音声対話エージェントの画面表示例
Fig. 8 Screenshot of anthropomorphic spoken dialog agent.

的なメッセージパッシング手段の 1 つであるパイプを用いて実現した．

さらに，各モジュールは UNIX 上の 1 コマンドとして扱うことができるため，UNIX 上の rsh（Remote Shell）コマンドや ssh（Secure Shell）コマンドの併用により，複数の計算機上で各モジュールを独立して動作させることができる．

3.4.5 類似するシステムとの比較

本提案手法と類似したシステム構成をとるものとして，Galaxy Communicator[8),9)] がある．同システムでは本提案システムのエージェント管理部に対応するハブ（Hub）を中心に各種機能モジュール（サーバ）をスター型結合した構成となっている．そのため，本提案システム同様，高いモジュール性を実現している．Galaxy Communicator と本提案システムとの単純な比較は難しいが，通信方式について見ると，前者がアプリケーションプログラムにおいては 1 回の通信においても専用の関数呼び出しによる複数の手続き（たとえばフレーム領域の確保や開放）が必要であるのに対して，後者は UNIX の標準入出力関数を用いる簡便な方式となっている．このため，後者はモジュール単独でのデバッグ作業が直感的で比較的容易に行えるのに対して，前者は，Hub や他のモジュールとの通信を模擬するプログラムが必要となり単独での作業が煩雑になる傾向がある．また，前者には 600 種類以上にのぼる関数が存在してマニュアルも膨大であるのに対して，後者は基本コマンドは表 2 にあるようにわずか 8 種類であり，システム開発者が通信方式を理解するのが容易である．このため，提案システムは小規模の実験システムを手軽に構築することができるという利点がある．

図 9 擬人化音声対話エージェントとの対話風景
Fig. 9 An example of user-system interaction.

4. 擬人化音声対話エージェントシステムの実現

前章で述べた構成要素モジュールの設計に従って，モジュールに要求される機能を実現し，音声対話擬人化エージェントのソフトウェアツールキットを作成し，その機能を検証するためにこれを用いた対話システムを試作した．システム画面，およびシステムとの対話風景を**図 8**，および**図 9** に示す．

このシステムは，エージェントの音声発話時における音声と顔画像の同期を中心に，音声認識モジュール（SRM），音声合成モジュール（SSM），顔画像合成モジュール（FSM），およびエージェント管理部（AM）の基本機能の動作確認のために以下に示す 4 つの対話タスクについて実現した．いずれのタスクも単語数 100 以下，単語パープレキシティ 10 以下の非常に簡単なタスクである．それぞれのタスクと，その目的を次に述べる．

(a) オウム返しタスク： 音声認識結果を合成音声によって復唱するだけのタスク．音声認識結果の確認，および LipSync を確認した．

(b) 面会確認確認タスク： 数名の関係者への来訪者スケジュールを確認するタスク．コマンド通信により，対話中にリアルタイムに表情を変化させられることを確認した．

(c) 連絡先問合せタスク： 数名の関係者の連絡先を問い合わせるタスク．後述する複数の文法の動的な切替え機能が実現されていることを確認した．

(d) 八百屋タスク： 八百屋で購入する野菜・果物を選択するタスク．野菜・果物の選択に対する認識結果が漸次的に出力されていることを確認した．

上記のタスクはいずれも AM の提供するコマンドを直接利用するプログラムで実現した．同程度に簡単なタスクであれば，同様に AM のコマンドを直接利用

表 4 Julian の認識性能と処理速度[22]（CPU: Pentium III 866 MHz）
Table 4 Recognition accuracy and average processing time[22] (CPU: Pentium III 866 MHz).

文法	語彙数	単語 perp.	最適ビーム幅	単語正解精度	平均処理速度
PG-M	806	91.6	800	93.7	0.4×RT
PG-L	4,439	257.1	2000	88.8	0.8×RT
SG-M	833	28.7	1000	97.6	0.5×RT
SG-L	5,023	76.0	3000	95.3	1.8×RT

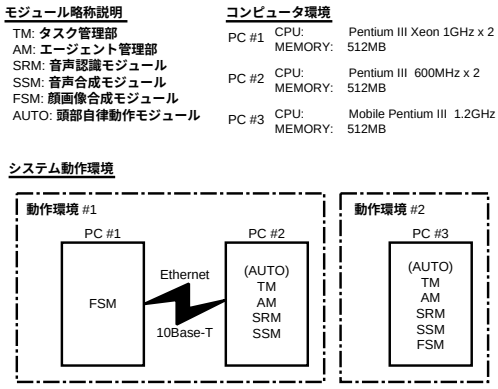

図 10 擬人化音声対話エージェントの動作環境
Fig. 10 Hardware configuration of anthropomorphic spoken dialog agent.

するプログラムでタスク記述を実現することができると考えられる．なお，これらの対話タスクの一部については，WWW の `http://iipl.jaist.ac.jp/IPA/` で，対話の様子のビデオデータを公開している．

図 10 に試作システムのハードウェア構成を示す．

4.1 音声認識モジュール

有限状態文法に基づく認識に加え，語彙・文法の動的な切替え機能を実現した．従来，文法などの切替えのためにモジュールの再起動が必要だったが，文法の動的な切替え機能の実現により，モジュール再起動の時間を削減でき，文法切替えの処理時間を短縮できた．また，音声認識結果の漸次的な出力を実現した．この機能は，認識結果に対する漸次的な応答の生成に利用できる．さらに，認識結果の出力形式や，認識時のパラメータなどをコマンドによって対話的にカスタマイズする機能を実現した．

認識性能と処理速度に関しては，すでに**表 4** に示す性能を実現した[22]．音響モデルとしては，IPA'99 より 16 混合モノフォンモデルを用いた[22]．また，タスクドメインは個人スケジュール管理[16] である．

4.2 音声合成モジュール

特定の人物の声による任意の漢字仮名混じり文章に対する明瞭な読み上げを実現した．韻律制御に関しては，JEIDA-62-2000[19] の規格の部分的な実装を行い，疑問文発話の際の文末表現などを実現できることを確認した．たとえば，通常の発声に比べピッチ周波数を x 倍にするタグ `<PITCH LEVEL="x">...</PITCH>` を用いて，

何時からのお約束です`<PITCH LEVEL="1.5">`か`</PITCH>`

と表記することで，語尾を強制的に高いピッチで発声することができるようになった．音声合成の準備時間は，対話タスク (a) におけるシステムの 200 発話について図 10 の動作環境#2 で測定したところ，平均 2.6 秒の発話時間に対して，平均 1.3 秒で発話の準備ができることを確認した．同時に，HMM による波形生成処理に要する時間が，発話準備時間の大半を占めており，発話準備時間は発話時間にほぼ比例していることを確認した．

さらに，顔画像合成モジュールと連係することで，システム発話時の LipSync が音素単位での同期が実現できていることを視察により確認した．

4.3 顔画像合成モジュール

対話的なコマンドに対応して，顔画像に任意の表情変形をリアルタイムに行う機能を実現した．さらに，口腔環境内の歯のモデルを加え，より精密な口周辺の合成画像が得られるようになった．また，顔画像合成モジュールの動作速度は，ハードウェアアクセラレーション機能を使い，図 10 の動作環境#1 で平均描画更新レート 20 [frame/sec]，動作環境#2 にて平均描画更新レート 15 [frame/sec] を達成した．

システム発話時の LipSync における口形状のアニメーションでは，フレームレートにあわせて口形状パラメータを選択することで，合成音声との同期を実現できていることを視察により確認した．発話音素と音素継続長から 1 ms 単位の口形状データベース生成に要した時間は，先の音声合成準備時間の測定と同時に図 10 の動作環境#2 で測定したところ，平均 178 ms であった．実際には音声合成の波形生成処理と並行して処理されているため，発話開始時間の遅れへの影響はほとんどない．

4.4 エージェント管理部

図 10 の両動作環境において，先に示した 4 つの対話タスクの動作，各モジュールの動作，およびシステ

ム発話時の LipSync の実現を確認した．また，図 10 の動作環境#1 においては，各モジュールを並列して動作させることで，各モジュールの効率的な処理を実現した．

4.4.1 モジュールの追加実験

顔画像合成モジュールを別のモジュールから制御し，頭部の自律動作を実現し，その動作を確認した．この頭部自律動作モジュールは自律動作の開始・停止を制御するスロットを持ち，顔画像合成モジュールの頭部の回転コマンドを生成することで，自律動作を制御する．エージェント管理部への追加は，登録モジュール定義の記述に頭部自律動作モジュールの名前と実行コマンドを 1 行追加することで実現できる．頭部自律動作モジュールは他のモジュールと同様に並行して動作しており，音声認識中やシステム発話中にも自律動作が可能であることを確認した．

5. 考　　察

試作システムとの対話を通じて見られた開発目標の達成度と機能拡張の必要性について，以下に考察する．

5.1 カスタマイズ性

音声認識モジュールとして利用している Julian に関しては，文法の動的な切替えを実現した．この機能は対話中の音声認識モジュールの文法制御を容易にするとともに，文法切替えの処理時間を短縮できた．また，認識結果の出力形式や，認識時のパラメータについても，コマンドによって対話的にカスタマイズすることを可能にした．

音声合成に関しては，特定の話者について，任意の漢字仮名混じり文章に対する明瞭な読み上げを実現した．さらに，JEIDA-62-2000[19] の規格の部分的な実装により，韻律制御を実現した．これにより，声と発話内容のカスタマイズが可能になった．

顔画像合成に関しては，正面方向から撮影した 1 枚の顔画像と標準ワイヤフレームモデルを整合させることで，エージェントの 3 次元頭部モデルが生成できた．この正面画像の整合作業は，慣れれば 1 枚の顔画像につき 10 分ほどで完了する．また，画像を整合した 3 次元頭部モデルを利用して，瞬きや笑顔などの表情の生成を実現した．これにより，顔画像データを準備できれば，任意の人の顔画像でエージェントの表情を生成する顔・表情のカスタマイズ機能を実現した．

5.2 機能部品のモジュラリティ

エージェント管理部は，各モジュールのインタフェースを統一化して扱い，モジュールの追加などのカスタマイズが容易な設計を実現した．実際新たなモジュールとして，顔画像の頭部自律動作を頭部回転コマンドを直接用いて制御するモジュールの追加を容易に実現できた．さらに，モジュール間の入出力はすべて AM を介し，通信に UNIX システムで使われている標準入出力を用いた結果，個々のモジュールの単独開発・デバッグが可能となり，高いモジュールの独立性を実現した．

今後は，Julian と切替え選択可能な音声認識エンジンとして SPOJUS[15] も使えるように作業をする予定であり，さらにモジュラリティの検証が進むと考えられる．

5.3 人間らしい対話の実現

対話処理の速度に関して，各モジュールの処理速度については先の章に示した速度を実現した．これにより，今回試作した比較的簡単なタスクであれば，ほぼリアルタイムに動作することを確認した．また，音声合成モジュールと顔画像合成モジュールが連係し，システム発話時の LipSync はほぼ問題なく動作した．さらに，口腔環境内の歯のモデルを加え，より精密な口周辺の合成を実現した．これにより，エージェントの音声発話の自然さを向上させることができた．

よりインタラクティブな対話を実現するためには，音声認識も漸次的に行える必要があり，音声認識から漸次的に出力される認識結果を利用した相槌によるシステム状態の開示が検討されている[23),24)]．筆者らは，このようなインタラクティブな対話に必要な情報として，音声認識モジュールにおける音声認識時の第 1 パスの結果の漸次的な出力を実現した．また，各モジュールを並行に動作させており，エージェントが音声発話中でも音声認識可能である．これらの情報をうまく利用することで，たとえばエージェントの対話理解状態の開示のための相槌やユーザの発話に対するエージェントの割込みなど，より効率的な対話を実現できると考えられる．

5.4 今後の展望

今後，VoiceXML[25] インタプリタによる対話管理モジュールやプロトタイピングツールとの連係したエージェントシステムの動作を考えるとき，VoiceXML で標準的な文法として提供されているもの（ビルトイン文法）を音声認識モジュールに持たせることで，VoiceXML との親和性が向上すると考えられる．たとえば，対話で頻繁に用いられる「はい」もしくは「いいえ」の二者択一の認識文法などは，音声認識モジュール側で準備しておくようにすると，タスクの記述が容易になるため，標準的に提供することが望ましい．

また，VoiceXML は主に電話を利用した音声応答

サービスの提供を支援する言語である．より擬人化音声対話エージェントの機能を活用するために，擬人化エージェント環境に固有なイベントの定義，音声出力時のエージェントの顔表情制御や合成音声の感情属性などの記述方法の定義などの拡張が考えられる．これによって，擬人化エージェントの表情や発話の強調など多様な表現の指定が可能となり，エージェントにより表現力を持たせることができるようになる．

音声合成モジュールにおける声のカスタマイズに関しては，少量の音声データを用いて，任意の発声者の音声を生成できるようにするために，声質変換，話者適応の技術を応用することができる．さらに，エージェントの個性を豊富にするために，発話の強調，多様な声質・韻律・感情表現・発話スタイルなどの細かな制御ができるようになることが望ましい．これらを実現するにはまだ多くの研究課題があるので，さらに音声合成の研究・改良を行う予定である．

対話においてよく見られる肯定や否定，疑問の動作など頭部のより自然で多様な動きを実現するには，オクルージョンにより顎の裏側や耳の裏側などの情報が欠損することが問題となる．この問題に対しては，任意の方向から撮影された複数の顔画像に対して標準ワイヤフレームモデルとの整合を行うことで，より忠実で自然な 3 次元頭部モデルを実現できる．さらに複数方向から撮影した頭部画像に対して，標準ワイヤフレームモデルを整合させ，3 次元の個人モデルを容易に作成するための GUI の構築することで，そのカスタマイズも容易になる．

エージェントの発話に関して，発話内容の部分的な強調が必要な場合，それに対応する表情や頭部動作のタイミングを制御するすることで，より人間らしい発話に近づけることができる．また，よりインタラクティブな対話を実現するためには，ユーザとシステムとが双方向から割込み可能なシステムであることが望ましい．音声認識部の漸次的な出力を利用することで，ユーザの発話中にエージェントが割り込んで対話を進行することも可能である．さらに今後は，エージェントの発話中にユーザが割り込んで話し始める barge-in への対応を行うとともに，話し言葉で見られる不要語やポーズなどへの対応を行うことで，より人間らしい対話を実現することができる．

エージェントが話を聞いている状態や発話が中断された状態，発話中に関しても強調などの動作が口形状アニメーションや表情だけではなく，頭部動作も連係した振舞いとして合成するようにすると，より明示的なシステムの状態開示を行うことができ，人間らしい擬人化音声対話エージェントの実現に役立つと考えられる．

今後は，各モジュールにさらに機能拡張と改良を行い，標準的な分散オブジェクト環境アーキテクチャ CORBA[26] などの導入も検討しながら，開発を進める予定である．

6. おわりに

本論文では，擬人化音声対話エージェントを将来のヒューマンインタフェースの重要な技術要素として位置づけ，その分野の共通プラットフォームとなりうる高いカスタマイズ可能性を備えたソフトウェアツールキットに必要な要素とその実現技術について論じた．さらに，実際に構築したソフトウェアツールキットを使って，簡単な対話タスクによって目標の達成度を評価した．

本論文で示した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は，情報処理技術振興協会（IPA）のサポートを受け，2000 年 3 月より十数カ所の研究機関の共同で開発が進められており[27)~31)]，ソースとともに無償公開する予定である．

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」の支援を受けた．

参考文献

1) 土肥 浩，石塚 満：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 (1999).
2) 牛田博英，平山祐司，中島 宏：自律的行動決定モデルに基づくインタフェースエージェント，電子情報通信学会論文誌，Vol.J82-D-II, No.10, pp.1655–1665 (1999).
3) 長谷川修，坂上勝彦，速水 悟：実世界視覚情報を対話的に学習・管理する人間型ソフトウェアロボット，電子情報通信学会論文誌，Vol.J82-D-II, No.10, pp.1666–1674 (1999).
4) 向井理朗，関 進，中沢正幸，綿貫啓子，三吉秀夫：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).
5) 河野恭之，鈴木 薫，知野哲朗，田中克己，屋野武秀，金澤博史：入力モダリティの多様化とその統合・利用について，情報処理学会研究報告，98–SLP–23–3 (1998).
6) 河野 泉，久寿居大，吉坂主旬，上窪真一：擬人化キャラクタを利用した知的対話システム，情報処理学会研究報告，98-SLP-23-4, pp.13–18 (1998).
7) Gustafson, J., Lindberg, N. and Lundeberg, M.: The August Spoken Dialogue System, *Eu-*

roSpeech, pp.1151–1154 (1999).
8) Seneff, S., Hurley, E., Lau, R., Pao, C., Schmid, P. and Zue, V.: GALAXY-II: A Referece Architecture for Conversational System Development, *ICSLP-1998*, pp.931–934 (1998).
9) DARPA: Communicator Program (1998). http://fofoca.mitre.org/
10) OAA (The Open Agent Architecture). http://www.ai.sri.com/~oaa/
11) 松坂要佐, 東條剛史, 久保田千太郎, 田宮大介, 古川賢司, 早田啓介, 中野裕一郎, 小林哲則：複数話者による対話システム，*Interaction'99*, pp.33–34 (1999).
12) Ehsani, F., Hatazaki, K., Noguchi, J. and Watanabe, T.: Interactive Speech Dialogue System Using Simultaneous Understanding, *ICSLP-1994*, pp.879–882 (1994).
13) 河口信夫, 松原茂樹, 外山勝彦, 稲垣康善：発話同時理解に基づくマルチモーダル図形エディタ, 情報処理学会研究報告, 98-SLP-22, pp.1–6 (1998).
14) 河原達也, 李 晃伸, 小林哲則, 武田一哉, 峯松信明, 伊藤克亘, 山本幹雄, 山田 篤, 宇津呂武仁, 鹿野清宏：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, 情報処理学会研究報告, 99-SLP-26-6, pp.39–46 (1999).
15) 甲斐充彦, 中川聖一：冗長語・言い直し等を含む発話のための未知語処理を用いた音声認識システムの比較評価, 電子情報通信学会論文誌, Vol.J80-D-II, No.10, pp.2615–2625 (1997).
16) 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).
17) 吉村貴克, 徳田恵一, 益子貴史, 小林隆夫, 北村 正：HMM に基づく音声合成におけるスペクトル・ピッチ・継続長の同時モデル化, 電子情報通信学会論文誌, Vol.J83-D-II, No.11, pp.2099–2107 (2000).
18) Yoshimura, T., Tokuda, K., Masuko, T., Kobayashi, T. and Kitamura, T.: Speaker Interpolation for HMM-based Speech Synthesis System, *J Acoust. Soc. Jpn.* (*E*), Vol.21, No.4, pp.199–206 (2000).
19)（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 (2000). JEIDA-62-2000.
20) 森島繁生, 八木康史, 金子正秀, 原島 博, 谷内田正彦, 原 文雄：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, pp.129–136 (1998).
21) 緒方 信, 中村 哲, 森島繁生：ビデオ翻訳システム — 自動翻訳合成音声とモデルベースリップシンクの実現, *Interaction2001*, pp.203–210 (2001).
22) 李 晃伸, 河原達也, 鹿野清宏：記述文法に基づく高性能連続音声認識エンジン Julian, 日本音響学会研究発表会講演論文集, 3-1-10, pp.111–112 (2001).
23) 平沢純一, 川端 豪：音声対話システム Noddy — ユーザ発話途中でのうなずき・相槌生成, 情報処理学会研究報告, 98-SLP-20-9, pp.203–210 (1998).
24) 横山真男, 青山一美, 菊地英明, 帆足啓一郎, 白井克彦：人間型ロボットの対話インタフェースにおける発話交替時の非言語情報の制御, 情報処理学会論文誌, Vol.40, No.2, pp.487–496 (1999).
25) VoiceXML (Voice eXtensible Markup Language Ver1.0) (2000). http://www.voicexml.org.
26) CORBA (The Common Object Request Broker Architecture). http://www.corba.org/
27) 嵯峨山茂樹, 中村 哲：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, pp.1–6 (2000).
28) 甲斐充彦, 伊藤克亘：対話システムにおける音声認識, 情報処理学会研究報告, 2000-SLP-33-2, pp.7–12 (2000).
29) 森島繁生, 四倉達夫：対話システムにおける顔画像生成, 情報処理学会研究報告, 2000-SLP-33-3, pp.13–18 (2000).
30) 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, 2000-SLP-33-4, pp.19–24 (2000).
31) 新田恒雄, 下平 博, 西本卓也：対話システムにおけるモジュール統合とプロトタイピング, 情報処理学会研究報告, 2000-SLP-33-5, pp.25–30 (2000).

(平成 13 年 11 月 20 日受付)
(平成 14 年 4 月 16 日採録)

川本 真一

1998 年九州工業大学情報工学部卒業. 2000 年北陸先端科学技術大学院大学博士前期課程修了. 現在, 同博士後期課程在学中. 擬人化エージェントによる音声対話に関する研究に従事.

下平　博（正会員）

1982 年東北大学工学部電気工学科卒業．1984 年同大学大学院博士前期課程（情報工学）修了．1988 年同博士後期課程修了．同年同大学工学部情報工学科助手．1992 年北陸先端科学技術大学院大学情報科学研究科助教授，現在に至る．工学博士．音声，文字，画像の認識処理およびヒューマンインタフェースに関する研究に従事．

新田　恒雄（正会員）

1969 年東北大学工学部電気工学科卒業．1998 年より豊橋技術科学大学大学院工学研究科教授．音声認識・合成・文字認識，およびマルチモーダル対話システムの研究に従事．工学博士．

西本　卓也（正会員）

1993 年早稲田大学理工学部卒業．1995 年同大学大学院理工学研究科修士課程修了．1996 年京都工芸繊維大学工芸学部助手．音声対話システム，ヒューマンインタフェースの研究に従事．

中村　哲（正会員）

1981 年京都工芸繊維大学工芸学部電子工学科卒業．1981 年〜1994 年シャープ（株）中央研究所および情報技術研究所に勤務．1986 年〜1989 年 ATR 自動翻訳電話研究所に出向．1992 年 4 月京都大学博士（工学）．1994 年〜2000 年奈良先端科学技術大学院大学情報科学研究科助教授．2000 年 4 月より ATR 音声言語通信研究所および音声言語コミュニケーション研究所第一研究室長，音声認識，多次元信号処理等の音声・音響情報処理，マルチモーダル情報処理，ヒューマンンインタフェースの研究に従事．

伊藤　克亘（正会員）

1988 年東京工業大学工学部情報工学科卒業．1993 年同大学大学院博士課程修了．同年電総研入所．現在産総研主任研究員．音声認識，音声対話の研究を行う．日本音響学会会員．博士（工学）．

森島　繁生（正会員）

1982 年東京大学工学部卒業．1984 年同大学大学院修士課程修了．1988 年同博士課程修了．工学博士．2001 年成蹊大学工学部教授，現在に至る．1994 年から 1 年間トロント大学客員研究員．1996 年から通信放送機構 3 次元空間共有プロジェクトサブリーダー．ATR 音声言語コミュニケーション研究所，ATR メディア情報科学研究所客員研究員を併任．コンピュータグラフィックス，ビジョン，マルチモーダルインタフェース等の研究に従事．

四倉　達夫

1998 年成蹊大学工学部卒業．2000 年同大学大学院修士課程修了．現在同博士課程在学中．コンピュータグラフィックスによる顔モデルの構築・仮想空間上でのコミュニケーションシステムに関する研究に従事．

甲斐　充彦（正会員）

1996 年豊橋技術科学大学大学院博士後期課程修了．同年豊橋技術科学大学工学部助手．1999 年静岡大学工学部講師．2000 年同助教授．音声認識を中心とした音声言語処理と対話処理に興味を持つ．博士（工学）．

李　晃伸（正会員）

2000 年京都大学大学院情報学研究科博士課程修了．同年より奈良先端科学技術大学院大学情報科学研究科助手．主として音声認識・理解の研究に従事．博士（情報学）．

山下 洋一（正会員）

1982 年大阪大学工学部電子工学科卒業．1984 年同大学大学院修士課程修了．同年大阪大学産業科学研究所文部技官，1993 年同助手，1994 年同講師，1997 年立命館大学理工学部助教授，2001 年同教授，現在に至る．博士（工学）．音声情報処理に関する研究に従事．

小林 隆夫（正会員）

1982 年東京工業大学大学院博士課程修了．同年東京工業大学精密工学研究所助手．同助教授を経て現在東京工業大学大学院総合理工学研究科物理情報システム創造専攻教授．工学博士．音声分析・合成・符号化・認識，マルチモーダルインタフェース，ディジタル信号処理の研究に従事．

徳田 恵一（正会員）

1984 年名古屋工業大学工学部電子工学科卒業．1989 年東京工業大学大学院博士課程修了．同年同大学電気電子工学科助手．1996 年名古屋工業大学知能情報システム学科助教授．工学博士．音声分析・合成・符号化・認識，マルチモーダルインタフェースの研究に従事．

広瀬 啓吉（正会員）

1972 年東京大学工学部電気工学科卒業．1977 年同大学大学院博士課程修了．工学博士．同年東京大学工学部電気工学科講師．1994 年同電子工学科教授．1996 年東京大学大学院工学系研究科電子情報工学専攻教授．1999 年 4 月より新領域創成科学研究科基盤情報学専攻教授．

峯松 信明（正会員）

1995 年東京大学大学院工学系研究科電子工学専攻博士課程修了．博士（工学）．現在同大学院情報理工学系研究科助教授．音声認識，音声分析，音声応用，音声知覚，および音声合成の研究に従事．

山田 篤（正会員）

1986 年京都大学工学部情報工学科卒業．1988 年同大学大学院修士課程修了．1991 年同大学院博士後期課程研究指導認定退学．京都大学工学部助手，奈良先端科学技術大学院大学助教授を経て，現在（財）京都高度技術研究所情報メディア研究室長．1998 年より京都大学大学院情報学研究科客員助教授および通信総合研究所専攻研究員．博士（工学）．言語処理系の研究に従事．

伝 康晴

1988 年京都大学工学部電気工学第二学科卒業．1993 年同大学大学院工学研究科博士後期課程研究指導認定退学．ATR 音声翻訳通信研究所研究員，奈良先端科学技術大学院大学情報科学研究科助教授を経て，2000 年 10 月より千葉大学文学部助教授．京都大学博士（工学）．専門は，音声談話分析，心理言語学，計算言語学．

宇津呂武仁（正会員）

1989 年京都大学工学部電気工学第二学科卒業．1994 年同大学大学院工学研究科博士課程電気工学第二専攻修了．京都大学博士（工学）．同年奈良先端科学技術大学院大学情報科学研究科助手．2000 年豊橋技術科学大学工学部情報工学系講師，現在に至る．自然言語処理の研究に従事．

嵯峨山茂樹（正会員）

1974 年東京大学大学院工学系研究科計数工学専攻修士課程修了．同年日本電信電話公社に入社，武蔵野電気通信研究所にて音声情報処理の研究に従事．1990 年 ATR 自動翻訳電話研究所音声情報処理研究室長として自動翻訳電話プロジェクトを遂行．1993 年 NTT ヒューマンインタフェース研究所にて音声認識・合成・対話の研究開発に従事．1998 年北陸先端科学技術大学院大学情報科学研究科教授．2001 年東京大学大学院工学系研究科のち情報理工学系研究科教授．博士（工学）．