実世界情報システムプロジェクト〜ネオサイバネティクス研究グループ

実世界環境のための擬人化対話エージェントに関する研究

酒向 慎司
情報理工学系研究科特任助手（NC グループ）

概 要

我々は，擬人化対話エージェントをベースとして，実世界環境で人間とインタラクションできるロボットの開発にむけて研究を行っている．本年度は，そのロボットの実現を目指し，我々の研究グループの保持する様々なセンシング技術との統合のため，大学院の講義を中心としてその検討を行った．また，音声対話に不可欠である音声合成システムの改良に取り組み，合成音の品質改善を行った．

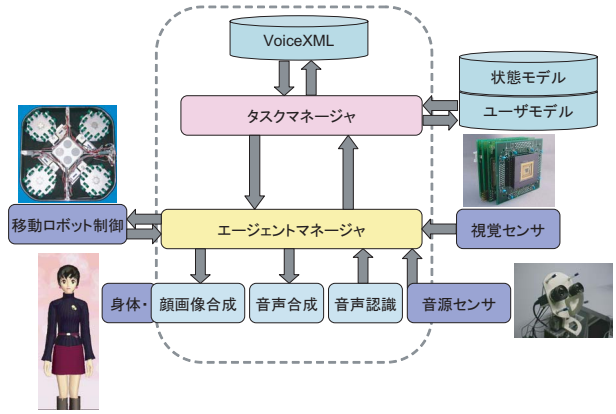

図 1: 案内ロボットの構成

1 はじめに

我々の研究グループでは，様々な研究機関とともに擬人化エージェントによる対話音声システムを開発してきた．これは，音声認識・音声合成・顔画像認識などの要素技術と，それらを統合し，対話を制御する各種コンポーネントから構成されている．一方，当研究グループには，高速ビジョンシステムや音源定位センサーに代表される様々な視聴覚センシングのための優れた技術を有しており，これらは実世界に適応したシステムを構築する上で非常に有用な技術である．

本プロジェクトでは，擬人化エージェントの枠組みを元に，実環境から得られる様々な視聴覚情報を統合するためのプラットホームとして拡張することで，実世界で人間と対話できるロボットの実現を目指している．プロジェクトの 3 年目である本年度では，展示会等を案内できる対話型の案内ロボットを実現するため，そのシステムの統合に向けた検討と設計を進めた（図. 1）．

2 擬人化対話エージェントとセンシング技術の統合

人間は実に様々な感覚器官を通して情報を取得している．聴覚と視覚から得られる情報だけを見ても実に多彩である．それらすべてを機械によって実現するのは困難だが，人間の知覚行動をよりよくサポートするための技術として，様々な視聴覚情報のためのセンシング技術を統合することはできないだろうか．

ここでは，全方向型の移動型ロボットに等身大のエージェントを表示させ，人や実環境から得られる情報を元に対話制御や移動可能なシステムを構築する上で，我々のもつ視聴覚センシング技術を統合し，より実世界環境でユーザをサポートできるロボットの開発を目指している．

2.1 センサー情報の効率的な統合処理

前述の通り，人間は環境から得た様々な情報を同時に処理しながら知的活動を行っている．しかし，それらの多種多様な情報を同時に処理しているわけではないことは容易に想像がつく．それは，時間的な解像度が必要な情報であったり，空間的な解像度が必要である情報など，必要に応じて情報の取捨選択が行えるように感覚機器が発達し，同時にそれらを処理する脳が適応しているからだと考えられる．

機械では，必要であれば時間的・空間的の両面で解像度の高い情報が得られるが，それらをすべて網羅的に処理を行うのは現実的ではない．視聴覚音声の研究分野で扱われるバイモーダル性を考えても，例えばマガーク効果に代表されるように視覚と聴覚の情報は互いに密接した関係にあり，その依存性を考えても個々の視聴覚情報の効率的な処理が必要であることが伺える．

そこで，様々なセンシング機器から得られる情報を，いかにして統合して円滑な対話システムを構築するかは本研究の重要なテーマとなり，多数の情報を効率的に扱うことが可能な方式としてブラックボードシステムに着目して，システム実装のための検討を進めている．

3 音声合成システムの改良

音声対話の出力部分を担う音声合成システムは，人間がシステムの表現力を直接的に感じる最も重要なコンポーネントの一つである．従って，音声としての品質のみならず人間らしい表現力が要求されている．ここでは，合成音声の品質改善と，話者の多様性を実現するための検討を行った．

3.1 合成音声品質の改善

Galatea Talk で用いられている音声合成システムでは，隠れマルコフモデル (Hidden Markov Model) に基づいた音声合成手法が採用されており，音声は分析合成系によって変換されたスペクトルパラメータによって統計的にモデル化されている．これにより，音韻の特徴を表すフォルマント形状が平滑化され，結果として合成音声の明瞭性が損なわれるという問題が生じやすい．そのため，なだらかな変化になったフォルマントの形状を鋭角化させるポストフィルタが有効となる．ここでは，話者の持つ声の特徴は一定ではなく，各話者モデルに適切なポストフィルタを設計することで，合成音声の品質改善を確認した．

一方，話者による違いは，ポストフィルタ係数のみならず，モデル化のための音響パラメータに変換する際の条件にも合成音声の品質に影響を与えている（例えば性別の違いなど）．そこで，さらに話者モデルごとに独自の分析条件を適用させるための検討を行っている．

3.2 話者モデル

擬人化エージェントに要求される技術要素として，柔軟なカスタマイズ性能があげられる．例えば，Galatea Toolkit では一枚の顔写真からその人独自のエージェントを構成することが容易に可能である．音声合成に関しても，様々な人の声を合成させたいという要求は高い．

Galatea Toolkit の音声合成システムでは，人間によって発声された音声データと，その韻律・言語的な情報を記述したラベルデータによって自動学習されているが，独自のモデル構築のための枠組みについては，十分な知識が必要とされており，明確なガイドラインが存在していなかった．そこで，音声の収録から合成音声用のモデル構築までの一連の手続きを検討し，少量の音声データからでも，了解性のある合成音声が得られるモデル学習法を示した．これにより，ユーザーが独自に音声合成用モデルを構築することができ，エージェントに自分の声を喋らせるといったことが実現可能となる．

4 まとめ

本年度では，案内ロボットの実現のための，各種センシング技術の統合方法の検討と，音声合成システムの品質改善とその柔軟性について検討を行った．来年度以降は，対話エージェントと各種センサーとの技術を統合し，案内ロボットの実装を進めていきたい．