

Galatea: 音声対話擬人化エージェント開発キット

Galatea: A DTK for Developing Anthropomorphic Spoken Dialogue Agents

新田恒雄¹ 西本卓也² 川本真一³ 下平博³ 森島繁生⁴ 四倉達夫⁵ 山下洋一⁶ 小林隆夫⁷ 徳田恵一⁸
広瀬啓吉² 峯松信明² 山田篤⁹ 伝康晴¹⁰ 宇津呂武仁¹¹ 伊藤克亘¹² 甲斐充彦¹³ 李晃伸¹⁴ 中村哲⁵ 嵯峨山茂樹²
(1豊橋技科大・2東大・3北陸先端大・4成蹊大・5 ATR・6立命館大・7東工大・8工大
9 ASTEM・10千葉大・11京大・12名大・13静岡大・14奈良先端大)

E-mail: nitta@tutkie.tut.ac.jp

1. 緒言

”Galatea”は音声対話システムをより自然な形で利用することを目的に制作された。この開発キットは modularity を考慮して設計され、マルチモーダル対話の研究開発支援プラットフォームを提供している[1]。

2. Galatea の構成と特長

図1に開発キットの全体構成を示した。各モジュールは独立のプロセスとして設計されており、対話システムは単一 PC (Note PC (Mobile Pentium III 1.2GHz, 512MB) 上で動作確認済)、あるいは分散環境 (複数 PC による並行動作) で使用することができる。以下に各モジュールを概説する。

(1) 音声認識: Julian[2]をベースに音声対話システムで要求される、(a) 文法に基づく音声認識、(b) 発話中の逐次的な認識結果出力、(c) 認識処理の動的制御(中断、文法の切替等)の諸機能を提供している。

(2) 音声合成: 日本語テキスト音声合成に必要な基本機能(形態素解析(茶筌[3])、読み・アクセント付与、韻律生成、合成波形生成)のほか、(a) 音素継続時間長を出力し顔画像の口唇との同期が可能、(b) テキスト埋め込みタグ(JEIDA規格準拠)による韻律制御が可能、(c) 合成音を出力途中で中断可能(barge in等)といった特長を持つ。合成器はHMMに基づく方式[4]を採用し、男女各1名の話者モデルを提供。

(3) 顔画像合成: 標準ワイヤフレームモデル中の代表点と正面写真中の対応点を、短時間(5-10分)のマウス操作で整合させるだけで、表情変化が可能[5]。表情は怒り、喜び、悲しみ、驚き、嫌悪、恐れの6種を用意。音声対話のため、LipSyncのほか、自律的な動作(うなづき、瞬き等)を提供。

(4) エージェントマネージャ: 対話部品が個々に規定するコマンドセットを使用して直接制御するレイヤと、対話管理に便利なマクロコマンドを利用して制御できるレイヤの二つを提供している(Unix版の場合[5]。Windows版では対話マネージャが各対話モジュールとソケット通信)。

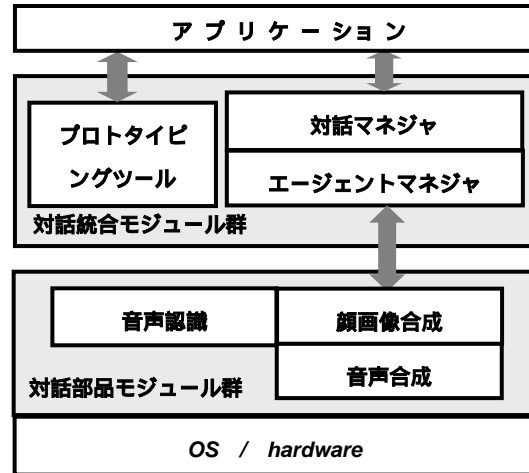


図1 Galatea 開発キットの全体構成

3. 開発支援ツール

対話記述言語として、(a) VoiceXMLにGUIのためのタグを付加したもの(主にUnix版で使用[6])と、(b)モダリティの追加が可能なマルチモーダル対話向け言語(XISL[7]; Windows版で使用)の二つを提供している。現在、(a)では簡単なGUIツールが、(b)ではラビッドプロトタイプングツール(Interaction Builder (IB)[7])が開発キットに含まれている。

4. 結言

Galateaはオープンソース化を前提に開発された。現在、CDROM配布、セミナー・講習会開催を通して関連研究と応用システム開発を支援するコンソーシアムの結成を計画している。今後、多くの方々に参加頂き、開発キットの改良・普及を継続的に行っていききたい。

謝辞

本研究の一部は情報処理技術振興協会(IPA)の支援を受けた。

参考文献

- [1] 嵯峨山ほか: 情処研報, SLP-45-10, pp.57-64 (2003).
- [2] 住吉ほか: 情処研報, SLP-37-16, pp.91-96 (2001).
- [3] <http://chasen.aist-nara.ac.jp/>
- [4] <http://hts.ics.nitech.ac.jp/>
- [5] 川本ほか: 情処論誌, vol.43, no.7, pp.2249-2263 (2002).
- [6] 西本ほか: 人工知能学会全大, 2C2-04 (2003).
- [7] 足立ほか: 情処研報, SLP-43-2, pp.7-12 (2002).