

## K-51 マルチモーダル対話システムのためのプロトタイプングツールの開発 Development of a Prototyping tool for Multi-Modal Interaction systems

足立 裕秋† 桂田 浩一† 山田 博文‡ 新田 恒雄†  
Hiroaki Adachi Kouichi Katsurada Hirobumi Yamada Tsuneo Nitta

### 1. はじめに

我々は、MMI 記述言語 XISL を提案すると共に[1][2], XISL を用いてマルチモーダル対話(MMI)システムを構築する際の支援ツール、Interaction Builder(IB)の開発を進めている。IB は GUI 操作によって XISL を記述することを目的としたプロトタイプングツールで、完成後、ツールの一部は情報処理振興協会(IPA)の擬人化音声対話エージェント基本ソフトウェア[3]に含まれる予定である。

今回開発したツールは、IPA プロジェクトで開発中の顔画像合成モジュール、音声認識モジュール、及び市販の音声合成ソフト(最終的には IPA プロジェクト開発のモジュールに置き換え予定)を組み込んだフロントエンド(以降 IPA 版フロントエンドと呼ぶ)を利用すると共に、オンラインショッピングシステムを例題に、マルチモーダル対話システムの開発を支援することを目指している。

本報告では、MMI システム(以降 IPA-MMI システムと呼ぶ)の概要を述べた後、IB の機能を説明する。

### 2. IPA-MMI システムの概要

図 1 に IPA-MMI システムのアーキテクチャを示す。フロントエンドは、ユーザからの入力を受け付ける入力インタフェースと、ユーザへの出力を行なう出力インタフェースからなる。IPA 版フロントエンドは PC 端末上で動作し、表 1 に示すモダリティを使用することができる。なお、IPA プロジェクトでは、各モジュールが分散した環境下でも動作するよう、エージェントマネージャが用意されているが、今回のプロトタイプシステムでは、同一システム上にすべてのモジュールを置き、Windows 上で動作することを前提に開発した。以下では、フロントエンドの各モジュールについて説明する。

#### 2.1 音声認識モジュール

音声認識エンジンには、IPA の日本語ディクテーション基本ソフトウェアに含まれる Julius for SAPI を用いる。これは、京都大学音声メディア研究室で開発された大語彙連続音声認識エンジン Julius[4] に、Microsoft 社の SAPI を付け加えたものである。今回は Julius for SAPI が出力した書き下し文をスポッティングすることで、ユーザの発話音声を特定している[5]。なお、IPA プロジェクトでは、対話制御部から送られる文法を直接利用できるモジュールを開発中で[3]、将来はこの組み込みを予定している。

#### 2.2 顔画像合成モジュール

IPA プロジェクトで開発し顔画像合成モジュール[6]の Windows 版を使用する。このモジュールは、任意の顔画像と、予め用意されたワイヤーフレームモデルとの整合を行なうことにより、顔モデルを生成する。表情とその継続時間の設定、および音声合成との同期をとるリップシンク等を行なうことができる。

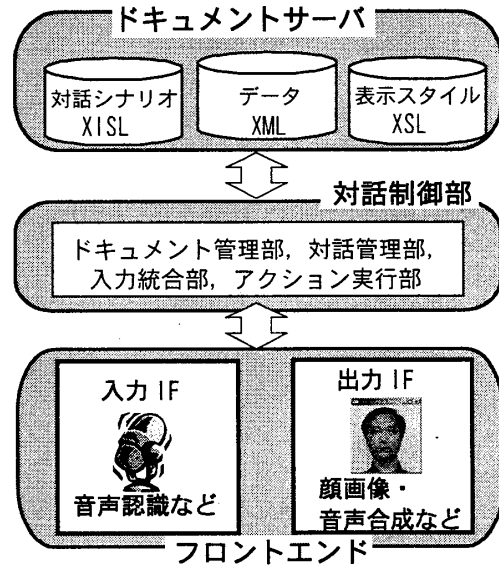


図 1 IPA-MMI システムのアーキテクチャ

表 1 IPA 版フロントエンドで利用可能なモダリティ

モダリティ	受け付け可能なイベント または出力動作	
入力	マウス	クリック, ダブルクリック等
	音声	音声認識結果
	キーボード	キー入力
出力	ウィンドウ	コンテンツの表示
	音声合成	テキストの読み上げ
	顔画像合成	顔モデルの変更, 表情, リップシンク等

#### 2.3 音声合成モジュール

合成エンジンには、今回、東芝 LaLaVoice™2001 を使用した。出力の音量、声質、読み上げキャラクタ等を指定できる。IPA プロジェクトでは、Linux 版音声合成モジュールを開発しているが[7]、この Windows 版も検討中であり、完成後このモジュールに置き換える予定である。

## 3. Interaction Builder の概要

### 3.1 IB の開発経緯

IB はこれまで、MMI オンラインショッピングシステム(以降 OLS システムと呼ぶ)のための XISL を生成するツールとして開発されてきた[8]。このツールを使用すると、実際の OLS システムの画面を見ながら対話の枠組みを構成すると共に、OLS システムで利用可能なモダリティの属性(語彙・文法、声質、表情など)を設定しながら、マルチモーダル対話のシナリオを生成することができる。

### 3.2 IPA-MMI システム用 IB

今回は、IPA-MMI システムで実行可能な XISL を生成するため、顔画像、音声認識、音声合成の各モジュールを制

†: 豊橋技術科学大学 大学院工学研究科  
‡: 豊橋技術科学大学 工学部

御する機能を追加した。図 2 に IPA-MMI システムを開発中の IB 画面の例を示した。以下に IB の機能を述べる。

### 3.2.1 シナリオ ビュー

図 2-①は、対話の流れを図式的に表示するためのシナリオビューを示している。受け付ける入力イベントやそれに対する出力動作が、それぞれ一つのブロックで表わされ、対話の流れに沿って配置される。これにより、構築中の対話の流れを容易に理解することができる。

### 3.2.2 タグツールバーおよびモダリティツールバー

図 2-②に示すバーは、対話を作成する際に利用する部品(対話の枠組み、組み合わせ、対話の遷移等)を並べたタグツールバーである。また、モダリティの属性を設定するボタンを並べた部分を特にモダリティツールバーと呼んでいる(拡大図 3 の右半分に相当)。ツールバーの中で、使用できない部品・ボタンは反転表示され(図 3 の左から 1, 3, 4 個目と 8 個目以降)、選択することができないことが示されている。

マルチモーダル対話を作成するには、ツールバーから目的の対話部品を選択し、シナリオビューに drag & drop すればよい。シナリオビューに対話部品が置かれると、図 2-③に示すダイアログボックスが表示され、属性値など必要な情報を入力するよう促す。

### 3.2.3 実行例

図 4 はシナリオビューを拡大したものである。この例では、音声かタッチ(マウス)のどちらかが入力されるのを待ち受ける対話を示している(alternative モード)。例えば、上記の条件に合う入力イベントが検出されたとき、擬人化エージェントが笑顔で、“ご用件は何でしょうか?”と応答する局面を記述するには以下の手順に沿って操作を行えばよい。

- イ) モダリティツールバーから顔画像を選択する。
- ロ) 選択したアイテムを、シナリオ中の追加したい個所に drag&drop する。
- ハ) モダリティに応じて、必要な属性値を入力するダイアログボックスが表示されるので、応答内容と表情を指定する。
- 二) ダイアログボックスを閉じると、シナリオビューには追加された擬人化エージェントの動作を示す対話部品が表示される。(図 4-(B)参照)

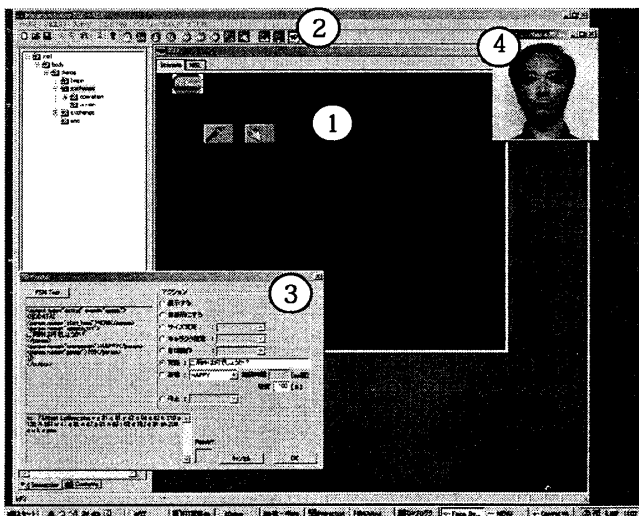
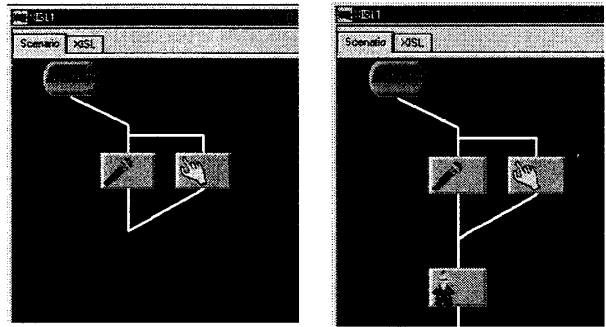


図 2 IB の実行画面例



図 3 タグの追加とモダリティを付与するためのツールバー



(A) (B)  
図 4 シナリオビュー(拡大図)

ハ)の時点で設定した属性値が、実際にどのような動作を示すかは、図 2-④に示す擬人化エージェント動作確認ウィンドウでチェックすることができる。

## 4. まとめ

本報告では、IPA-MMI システムのためのプロトタイプングツールについて述べた。今回の開発により、IPA-MMI システムで利用可能な入出力モダリティに属性を指定すると共に、対話シナリオを記述して XISL ドキュメントを生成することが可能になった。IB 及び、IPA-MMI システムは現在も改良が進んでいる。今後は、変数管理、条件式の生成等の機能を追加すると共に、シナリオ作成途中のテスト機能についても検討したい。

本研究の一部は、情報処理進行事業協会(IPA)「独創的情報技術育成事業」の支援を受けた「擬人化音声対話エージェント基本ソフトウェアの開発」プロジェクトの一部として行なわれている。

## 参考文献

- [1] 桂田浩一他：“音声対話記述言語 VoiceXML と MMI 記述言語 XISL の比較”，情報処理学会研究報告 2001-SLP-37, pp.43-48 (2001).
- [2] <http://www.vox.tutkie.tut.ac.jp/XISL/XISL.html>
- [3] 川本真一他：“カスタマイズ製を考慮した擬人化音声対話エージェントの設計” 情報処理学会論文誌 Vol.43 No.7 pp.2249-2263 July (2002).
- [4] 河原, 李, 小林他：“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.43-48 (2001).
- [5] 新田恒雄他：“汎用 LVCSR を用いた対話音声の認識”，情報処理学会研究報告 2002-SLP-41, pp.69-74 (2002).
- [6] 四倉達夫, 森島繁生：“擬人化音声対話エージェントツールキットのための表情合成技術”，情報処理学会研究報告 2001-SLP-40-11, Feb. (2002).
- [7] 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フォーム”，情報処理学会研究報告 2002-SLP-40-12, Feb.(2002).
- [8] 足立裕秋他：“Interaction Builder -マルチモーダル対話システムのためのプロトタイプツール”，情報処理学会第 64 回全国大会講演論文集, 5C-01, (2001).