

MMI システム構築のためのプロトタイピングツールの開発

足立裕秋* 桂田浩一* 山田博文** 新田恒雄*

*豊橋技術科学大学 大学院工学研究科 知識情報工学専攻
**豊橋技術科学大学 マルチメディアセンター
〒441-8580 愛知県豊橋市天伯町雲雀ヶ丘 1 - 1
Email: adachi@vox.tutkie.tut.ac.jp

あらまし：我々は，MMI 記述言語 XISL(Extensible Interaction Scenario Language)を提案し，XISL で記述された対話シナリオを実行するための MMI システムを開発している．本報告では，マルチモーダル対話(MMI)システム構築支援のためのプロトタイピングツールについて述べる．MMI システムを構築するための支援ツール，Interaction Builder(IB)により，利用できるモダリティの属性値(音声認識文法や応答文など)の設定が容易になり，対話シナリオの効率的な記述が可能になる．今回の開発では，IPA 擬人化音声対話エージェント基本ソフトウェアプロジェクト(Galatea プロジェクト)で開発中の各種モジュールに対応するために，新規モダリティのサポート等，機能の追加を行なった．これにより同プロジェクトで開発中の音声認識，顔画像合成，音声合成モジュールを組み込んだ MMI システムを，対話記述言語 XISL に関する知識なしでも開発することが可能になった．

キーワード： マルチモーダル対話，対話記述言語，XISL，プロトタイピングツール，Galatea プロジェクト

Development of a Prototyping Tool for MMI Systems

Hiroaki ADACHI*, Kouichi KATSURADA*, Hirobumi YAMADA ** and Tsuneo NITTA*

*Graduate School of Technology, Toyohashi University of Technology
** Multimedia Center, Toyohashi University of Technology

1-1 Hibarigaoka, Tempaku-cho, Toyohashi 441-8580, JAPAN
Email: adachi@vox.tutkie.tut.ac.jp

Abstract : This paper presents a prototyping tool for multi-modal interaction(MMI) systems. We have proposed an MMI description language XISL, which enables system developers to describe MMI scenarios. Moreover, we have developed a prototyping tool "Interaction Builder (IB)" for constructing MMI systems. IB provides such functions as set-up of attribute values of modalities (speech grammar, etc.) and efficient description of MMI scenarios. In this paper, we introduce a new version of IB that supports new modalities developed by "Open-source Software for Developing Anthropomorphic Spoken Dialog Agents Project"(Galatea Project). A MMI system of the Galatea Project is consist of a speech recognition engine module, a facial image synthesis module, a speech synthesis module and a multi-modal interaction module. System developers can implement these MMI modules without the knowledge of MMI description language XISL by the support of IB.

Key words : multi-modal interaction, interaction language, XISL, prototyping tool, Galatea Project

1. はじめに

我々は、マルチモーダル対話(MMI)記述言語 XISL(Extensible Interaction Scenario Language)を提案すると共に[1][2], XISL で記述した対話シナリオに沿って、ユーザと対話を行なうことができる MMI システムを開発している。同時に、XISL を用いて MMI システムを構築するための支援ツール、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ピングツールで、これまでのバージョンでは、我々のグループが開発してきた 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)[3]を対象としている。IB を使用することにより OLS システムで利用できるモダリティの属性値(音声認識文法など)の設定や、対話シナリオ記述を効率的に行なうことができる。

今回のバージョンでは、新たに情報処理振興協会(IPA)の擬人化音声対話エージェント基本ソフトウェアプロジェクト(以降 IPA プロジェクトと呼ぶ)[4]で開発中の顔画像合成モジュール、音声認識モジュール、音声合成モジュールを組み込んだ Windows 版 MMI システム(以降 IPA-MMI システムと呼ぶ)を対象とし、IPA-MMI システムで利用できるモダリティの属性値設定等、機能の追加を行なった。

以下では、始めに XISL を概観し、IPA-MMI システムの概要を説明する。続いて、IB の機能について述べる。尚、本 IB は完成後、一部が IPA プロジェクト(Galatea Project)が提供するオープンソフトウェアに同梱される予定である。

2. MMI 記述言語 XISL

XISL は XML ベースの MMI 記述言語であり、対話シナリオの可読性が高く、入出力の制御や割り込み対話の記述が容易であるといった特徴を持つ。したがって XISL の利用により、システム開発者は MMI 機能を持つ Web サービスを容易に提供できる。

XISL は、一組の対話シナリオを表す dialog の集合から構成される。OLS システムの例では、ユーザ ID の認証を行なう対話、商品閲覧のための対話などが一つの dialog として記述される。

各 dialog は、対話の最小単位である exchange の集合を内部に含み、各 exchange は、operation と action をそれぞれ一つずつ持つ。operation は

```
hello.xisl
<?xml version="1.0" encoding="Shift-JIS" ?>
<!DOCTYPE xisl SYSTEM "xisl.dtd">
<xisl version="1.0">
  <head>...</head>
  <body>
    <dialog id="hello World">
      <exchange>
        <operation target="hello.xml"> -
          <input type="touch" event="click"
                match="/page"> -
        </operation>
        <action>
          <output type="speech" event="tts-speech">
            <![CDATA[
              <param name="speech-text">
                Hello World!
              </param>
            ]]>
          </output>
        </action>
      </exchange>
    </dialog>
  </body>
</xisl>
```

図1 XISL の記述例

ユーザの入力記述部で、例えば音声やポインティングの入力待ち受けが内部の input に記述される。一方、action はシステムのアクション記述部で、合成音声や擬人化エージェントによる応答を内部の output に記述することができる。

図1に XISL の例を示す。この例では、hello.xml の<page>の内容が、XSL によりブラウザに表示されていることを想定し、表示指定箇所がクリックされたときに、システムが“Hello World”と合成音声で応答する対話の記述を示している。

3. IPA-MMI システム

図2に、IPA-MMI システムのアーキテクチャを示す。システムは、フロントエンド、対話制御部、及び WWW 上のドキュメントサーバから構成される。ドキュメントサーバには、システムのドキュメント(対話シナリオ(XISL)、データ(XML)、表示スタイル(XSL))が保持される。三つのドキュメント群をそれぞれ独立に取り扱うことで、データの保守性が向上するといった利点を持つ。また、対話シナリオや表示スタイルを独立に変更できることは、異なるユーザインタフェースを持

つ端末からの、シームレスなサービスを可能にする[5]。これらのドキュメントは、対話制御部に渡される。対話制御部は、対話シナリオの解釈・実行、フロントエンドからの入力情報の処理、フロントエンドへの出力命令の送信を行なう。フロントエンドは、IPA プロジェクトで開発中の各モジュールを利用した入出力インタフェースを持つ。入力インタフェースがユーザからの入力を受け付けると、その内容を対話制御部に送信し、一方、対話制御部からの出力命令を受けると、出力インタフェースがユーザへの出力を行なう。

IPA-MMI システムは、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ている。

以下に、フロントエンドの各モジュールについて説明する。

3.1 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[6] 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[7]。なお、IPA プロジェクトでは、CFG 文法の利用が可能な音声認識モジュールを開発中で[4]、将来はこの組み込みを予定している。

3.2 顔画像合成モジュール

IPA プロジェクトで開発した顔画像合成モジュール[8]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。

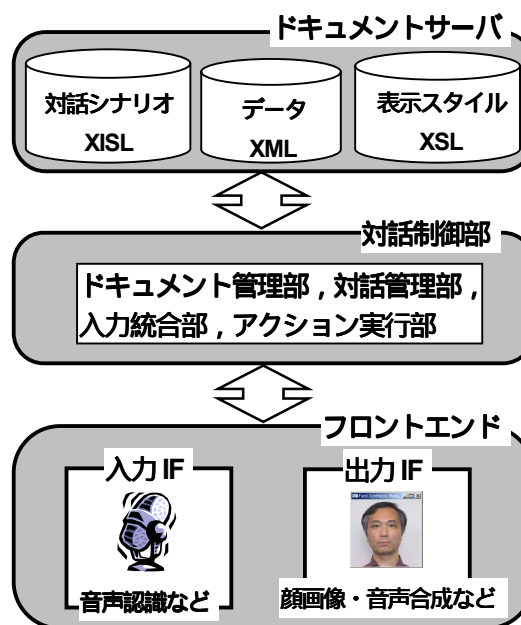


図 2 IPA-MMI システムのアーキテクチャ

表 1 IPA-MMI システム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

3.3 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[9]、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。

4. Interaction Builder の概要

以前に報告した IB では OLS システム用の XISL の生成が可能であり、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し、同システムで利用できるモダリティの属性(語彙・文法、声質、表情など)を実際の動作を確認しながら設定できた[10][11]。

今回は、IPA-MMI システムを対象に、同システムで実行可能な XISL を生成するため、音声認識、顔画像合成、音声合成モダリティの属性を設定で

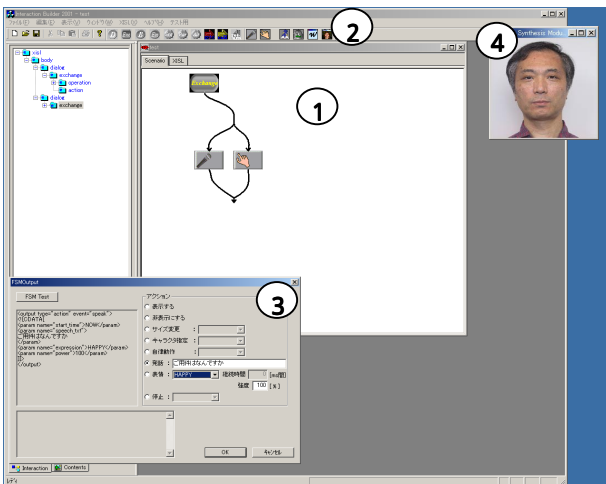


図3 IBの実行画面例



図4 タグの追加とモダリティを付与するためのツールバー

きる機能を追加した。図3に対話シナリオを記述中のIB画面の例を示す。以下では、IBの機能について説明し、動作画面を基に実行例を示す。

4.1 シナリオビュー

図3- 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。対話の枠組みや対話の遷移、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。

4.2 ツールバー

図3- 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)を並べたツールバーである。図4にその拡大図を示す。ツールバーの左半分はXISLのタグ追加用のボタンを並べた部分であり、タグツールバーと呼ぶ。一方、右半分はモダリティの属性を設定するボタンを並べた部分であり、モダリティツールバーと呼ぶ。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4の左から1,3,4個目と8個目以降)、選択できる部品を用意に識別できるようになっている。

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。

4.3 実行例

ここでは、以下に示す対話シナリオを、実際にIBを用いて生成する過程を、画面例を示しながら説明する。

- シナリオ -
音声による入力を待ち受け、指定した文法にマッチする入力を受け付けると、エージェントに“ご用件はなんですか？”と発話させる

Step 1 :

IBを起動し、新規作成メニューを選択すると、図5に示すように、空白のシナリオビューが表示される。このビュー上に、タグツールバーから目的の対話部品を drag&drop しながら、対話シナリオを構成していく。

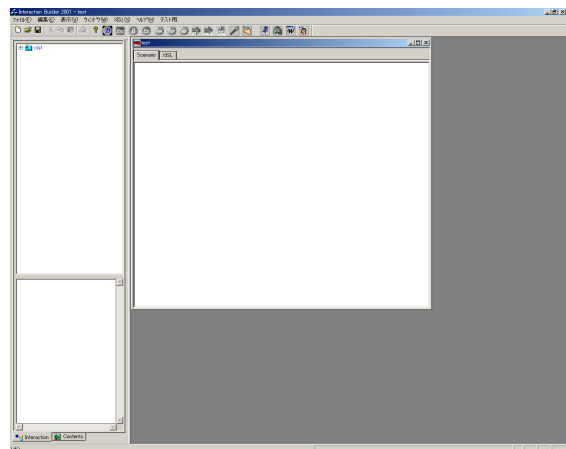


図5 Step1の実行画面

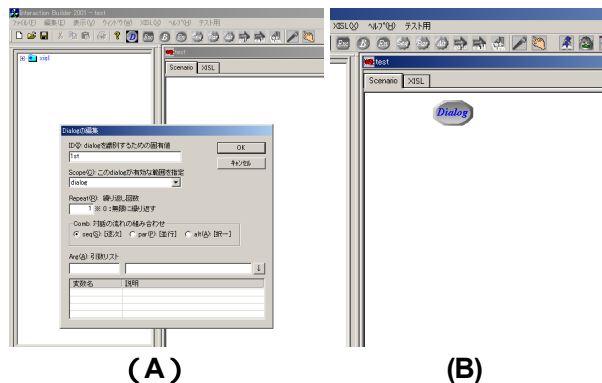


図6 Step2の実行画面

Step 2 :

対話の枠組みを構築するため、タグツールバーから、dialog を選択し、シナリオビューに drag&drop すると、図 6-(A)に示すように、dialog に必要な情報(ID, 引数, 繰り返し回数など)を入力するダイアログボックスが表示される。OK ボタンを押すと、図 6-(B)に示すように、dialog を示すブロックがシナリオビューに追加される。

Step 3 :

dialog を右クリックすることにより表示されるメニューから、図 7-(A)に示すように dialog 内部の構造を表示することができる。ここで、受け付ける入力やそれに対する応答を記述するための枠である exchange をタグツールバーから選択し、dialog 上に drag&drop すると、図 7-(B)に示すように、exchange を示すブロックが追加される。この exchange に対し、音声入力とエージェントの出力を追加する。

Step 4 :

追加した exchange をダブルクリックすることにより、exchange 内部の表示に切り替わる。ここで、音声入力を待ち受ける記述をするには、モダリティツールバーから音声認識を選択し、

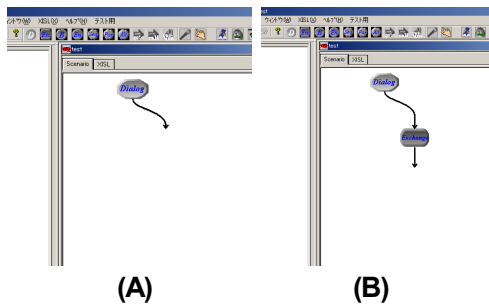


図 7 Step3 の実行画面

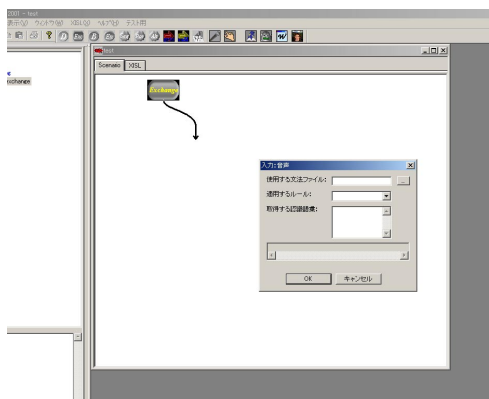


図 8 Step4 の実行画面 (1)

exchange 上に drag&drop すればよい。すると、図 8 に示すように、音声認識モダリティに必要な属性(使用する文法ファイル, 適用するルールなど)を入力するためのダイアログボックスが表示される。属性を設定し OK ボタンを押すと、図 9 に示すように、音声入力を待ち受けることを示すブロックが表示される。この後に、エージェントの動作を追加すれば、目的のシナリオが完成する。

Step 5 :

音声入力に対応する、エージェント出力を記述する。モダリティツールバーから、エージェントを選択し、drag&drop すると、図 10 に示すように、必要な属性(表情とその継続時間, 発話内容など)を入力するダイアログボックスが表示される。また、画面右上に表示される実際のエージェントを利用して、設定した属性値がどのような動作をするかを確認することができる。ここで、発話内容の項目に、「ご用件は何ですか?」と入力し OK ボタンを押すと、図 11 に示すようにシナリオビューにエージェントの動作を示すブロックが表示さ

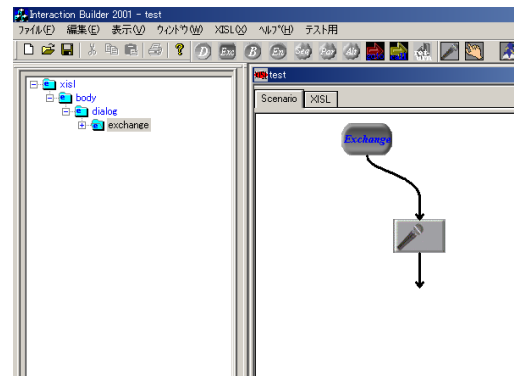


図 9 Step4 の実行画面 (2)

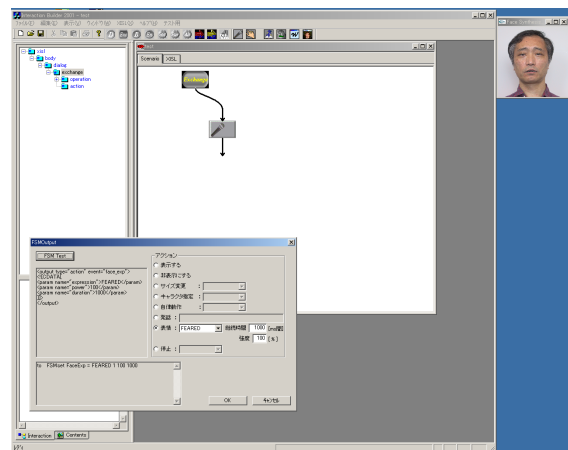


図 10 Step5 の実行画面

れる。

一方、図 12 に示すように、ビューを切り替えることで、実際の XISL のソースを確認することもできる。

5. まとめ

本報告では、IPA-MMI システムのためのプロトタイプツールについて述べた。今回の開発を通して、IPA-MMI システムで利用可能な顔画像エージェント等の入出力モダリティ、およびその属性を簡単に指定できるようになり、XISL に関する知識なしでも、MMI の対話シナリオを記述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(Galatea プロジェクト)の一部として行なわれている。

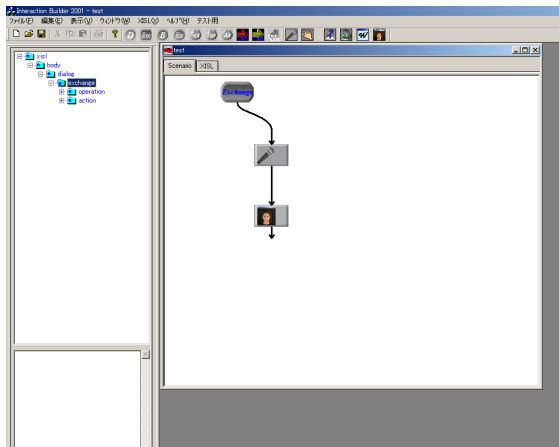


図 11 Step5 の実行画面

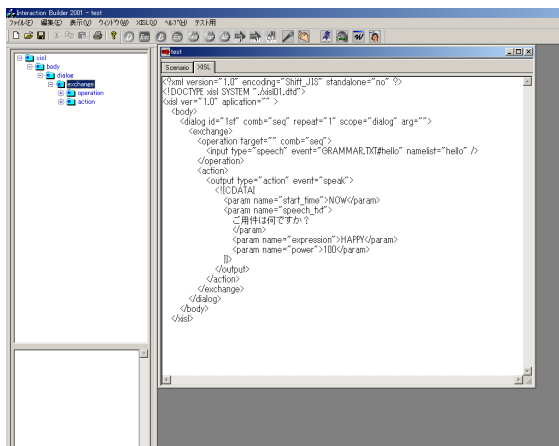


図 12 XISL のソースを表示

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] 小林剛典，足立裕秋，桂田浩一，小林聡，山田博文，新田恒雄：“MMI 記述言語 XISL を用いたオンラインショッピングシステムの試作”，情報処理学会第 64 回全国大会講演論文集(分冊 4)，pp.107-108 (2002)
- [4] 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)。
- [5] 桂田浩一，大谷佳彦，中村有作，小林聡，山田博文，新田恒雄：“多様な端末からのアクセスを可能にする MMI アーキテクチャ”，情報処理学会研究報告書 2002-SLP-40，pp.51-56 (2002)
- [6] 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)。
- [7] 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)。
- [8] 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)。
- [9] 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)。
- [10] 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)。
- [11] 足立裕秋他：“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の開発”，情報処理学会 第 1 回情報科学技術フォーラム，k-51