

E-039

MMI 6階層モデルに準拠した Web ベース MMI システムの開発

Web-based Interaction System based on an MMI Six-Layered-Model

工藤 正志†
Masashi Kudo

桂田 浩一†
Kouichi Katsurada

入部 百合絵†
Yurie Iribe

新田 恒雄†
Tsuneo Nitta

1 はじめに

マルチモーダル対話 (Multi-Modal Interaction: MMI) は、次世代ヒューマンインターフェース技術として、様々な分野での活用が期待されている。これまで我々は、誰もが手軽に MMI を利用できるようにするため、広く一般に使われている Web ブラウザ上で動作可能な Web ベース MMI システムを開発してきた。

このシステムは、JavaScript などの標準技術のみを用いているため、特別なソフトウェアのインストールや高性能端末の装備を必要とすることなく MMI を利用するのが特長である。しかし、このシステムはモジュール間の結びつきが強いため、使用するモダリティを変更する際など、システム改変が容易でないという問題があった。

そこで我々は、MMI システムの標準アーキテクチャとして情報処理学会試行標準委員会で提案された MMI 6 階層モデル¹⁾に準拠するようシステムを再構築することにした。MMI 6 階層モデルでは、入出力および対話制御の処理が粒度 (モダリティ制御、タスク遷移...)毎に階層化されており、階層間で厳密にモジュール分割されている。このモデルに従うことで、モダリティの変更や対話記述言語の切り替えが容易になり、システムの拡張性を向上させることができる。

以下、2. で本システムのベースとなる Web ベース MMI システムについて、3. で MMI 6 階層モデルについて説明した後、4. では MMI 6 階層モデルに準拠した Web ベース MMI システムの構成について述べる。

2 Web ベース MMI システムの構成

Web ベース MMI システムは、Galatea tool kit²⁾を基にして構築された MMI システムである。このシステムは Ajax および Comet の技術を利用してサーバとブラウザを連携させ、高負荷な処理をサーバ上で、低負荷な処理をブラウザで行なうよう設計している。図 1 にシステム構成を、表 1 にシステムが扱うことができるモダリティを示す。以下では Web ブラウザ上での処理とサーバ上での処理に分けて説明する。

2.1 Web ブラウザでの処理

2.1.1 ユーザからの入力取得

音声認識は複雑な処理を必要とするため、ブラウザ単体で低負荷に実装するのは困難である。そのため録音をブラウザで、認識をサーバ上で行なうことにより、ブラウザへの負荷が少ない音声認識を実現した。この手法は西村らの w3voice³⁾で実用性が確認されている。ブラウザ上での録音は Java Applet を用いた音声録音器 Sound Recorder によって行なわれる。録音した WAV 音声データ

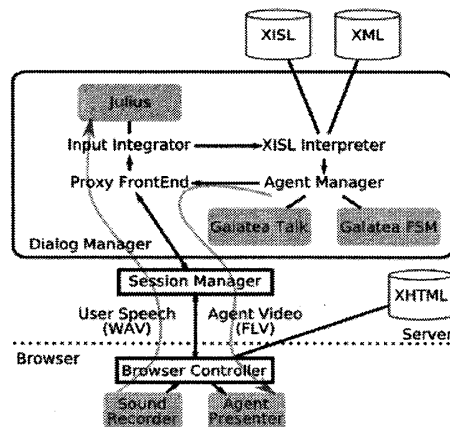


図 1 従来システムのモジュール構成

表 1 利用可能なモダリティー一覧

	利用可能なモダリティー
入力	音声, ポインティング, キーボード
出力	ブラウザ, エージェント(動画), 合成音声

は Base64 エンコードされた後、JavaScript で構築されたブラウザ制御器である Browser Controller によってサーバに送られ、音声認識処理にかけられる。また、ポインティングなど音声以外の入力についても Browser Controller によって取得され、サーバに送信される。

2.1.2 ユーザへの出力

顔画像合成と音声合成によるエージェント出力は高負荷な処理となる。そこで我々は、顔画像と音声をサーバ上の Agent Manager において 1 つの FLV 形式の動画に結合し、ブラウザ上でその動画を再生することでエージェント出力を低負荷に実現する方法を採った。動画再生には Adobe Flash を用いた動画再生器である Agent Presenter を使用した。また、ページ遷移などのブラウザ出力は Browser Controller によって実現している。

2.2 サーバ上の処理

ブラウザからのマルチモーダル入力データは、図 1 に示すサーバ上の Session Manager が受け取り、統合前処理 (音声入力は音声認識結果に変換) の後、Dialog Manager 内の Input Integrator に送られ、統合処理が行なわれる。統合結果は、MMI 対話シナリオ記述言語 XISL で記述された文書を解釈する XISL Interpreter に渡され、対話シナリオに沿った出力命令が生成される。出力がエージェントの場合は、Agent Manager で動画が生成され、Session Manager を通じてブラウザに送信される。

その他の出力 (Web ページの表示など) の場合、そのまま Session Manager を通じてブラウザに Web ページ表示などの命令が送られる。

† 豊橋技術科学大学
Toyohashi University of Technology

3 MMI 6階層モデル

MMI 6階層モデル¹⁾とは、情報処理学会試行標準委員会にて提案された、MMIシステムのための階層型アーキテクチャである。このアーキテクチャは図2に示すように、MMIシステムの処理の単位ごとに、コンポーネントが階層化されている。層間の通信に用いるイベントも規定されており、各層の独立性が高くなっているのがこのモデルの特徴である。

MMI 6階層モデルの各層の仕様は以下の通りである。

- 第1層：基本的に単独モダリティの入力・出力を行なうコンポーネントである。入出力の両方の機能を持つモジュール(例：Webブラウザ)もこの層に該当する。
- 第2層：画面表示、音声入出力、擬人化エージェント制御といったモダリティの制御を行なうコンポーネントである。
- 第3層：入力統合、出力分化、入出力同期制御などを行なうコンポーネントである。逐次入力や同時入力の解釈、逐次出力や同時出力の同期制御などがこの層で実行される。
- 第4層：タスク内の対話制御と応答内容の決定を行なうコンポーネントである。フォーム処理の充足判定や、タスク内の対話遷移処理などがこの層で行なわれる。
- 第5層：対話タスクの全般的な制御を行なうコンポーネントである。また、アプリケーション層との入出力通信はこの層が行なう。
- 第6層：データモデルとアプリケーションロジックを実装するコンポーネントである。この層にはこれらに対するAPIを定義することが要求される。
- ユーザモデル/デバイスモデル：外部オントロジーで定義されるユーザモデル/デバイスモデル変数を管理し、2~5層のコンポーネントに対してAPIを提供する機能を持っている。

6階層モデルに準拠することで、モダリティの入れ替えや対話記述言語の切り替えが容易になり、システムの拡張性を高めることが可能になる。

4 MMI 6階層モデルに準拠したシステムの構成

これまで我々が開発してきたWebベースMMIシステム(以後、従来システムと呼ぶ)を図3に示すような、MMI 6階層モデルに準拠する形で再構築した。図3中の4層は、MMI 6階層モデルにおける4層と5層を一つの層に統合したものになっている。

従来システムと今回開発したシステムの処理を比較すると、Webブラウザのモジュール構成は同様であるが、サーバ上のモジュール構成が大きく異なる。

従来システムの入力統合および統合前処理の音声認識は、全て入力統合器 Input Integrator で行なわれていた。これに対して、今回開発したシステムの入力部は、統合前処理の音声認識を2層の入力管理部 Modality Input Manager で行ない、入力統合を3層の統合管理部 Input Integrate Manager で行なうように設計している。

また、従来システムの対話管理および出力管理は、対話管理部である XISL Interpreter が逐次出力などのタイミングを管理し、エージェント生成器 Agent Manager に出力の指示を出していた。これに対して、今回開発したシステムでは XISL Interpreter には対話管理だけを行なわせ、

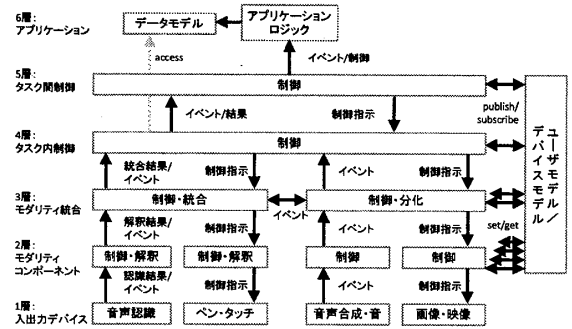


図2 MMI 6階層モデルのアーキテクチャ

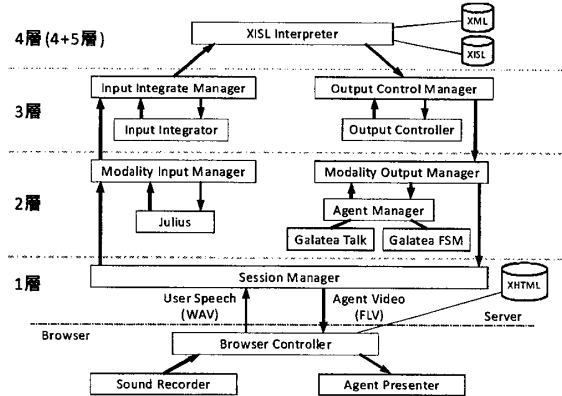


図3 6階層モデルに準拠したシステムのモジュール

出力タイミングの管理は3層の出力管理部 Output Control Manager が行なうように設計した。また Agent Manager については、2層の出力部 Modality Output Manager のモジュールの一つとして構築した。

以上の改良によって、モダリティ追加・変更などの拡張性が向上したといえる。例えば、音声認識エンジンを変更する場合は認識モジュールを差し替えるだけでよい。このような柔軟な変更が可能なのは、各層の独立性が高く、一つの層に対する変更が他の層に影響を与えないという特徴がMMI 6階層モデルにあるためである。

5 おわりに

本稿では、MMI 6階層モデルに準拠したWebベースMMIシステムについて述べた。MMI 6階層モデルに準拠することで、これまで我々が開発してきたWebベースMMIシステムの拡張性向上を実現している。今後は、これまでのシステムで扱っていなかったモダリティ(Webカメラなど)の追加等の、システム拡張を行ないたい。

参考文献

- 1) 新田恒雄, 他, "マルチモーダル対話システムのための階層的アーキテクチャの提案", 情報処理学会研究報告, 2007-SLP-68-2, pp.7-12(2007).
- 2) S. Kawamoto, et al., "Galatea: Open source software for developing anthropomorphic spoken dialog agents", in Life-Like Characters, ed. H. Prendinger and M. Ishizuka, pp.187-212, Springer-Verlag (2004).
- 3) 西村竜一, 他, "音声入力・認識機能を有するWebシステム w3voice の開発と運用", 情報処理学会研究報告, 2007-SLP-68-3, pp.13-18 (2007).