柔軟なモジュール切替が可能なWebベースMMIシステムの開発

工藤 正志† 桂田 浩一† 入部 百合絵† 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

1. はじめに

我々は誰もが手軽にマルチモーダル対話（Multi Modal Interaction：MMI）を利用できるよう，広く一般に使われる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，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 Project)で開発したGalatea Toolkit[1][2]をベースに構築している．実装にはJavaScriptなどの標準技術のみを用いるため，ユーザは特殊なソフトウェアや高性能端末を必要とすることなくMMIを利用できる．しかし，開発したシステムはモジュール間の結びつきが強く，モダリティの変更など，システム改変が容易でないという問題があった．

そこで，今回，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が策定した，MMI 6階層モデル[3]に準拠してシステムを再構築した．6階層モデルに準拠することで，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，システムの拡張性の向上が実現できた．

2. Webベース MMI システム

Webベース MMI システムは，Galatea Toolkit をベースに構築した 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にシステムの構成を，また図2にシステムのスクリーンショットを示す．

本システムでは，図2の左上に示すエージェントに向かいユーザが発話すると，発話内容がブラウザ側のSound Recorderに録音される．録音データはBrowser Controllerによってサーバ上のSession Managerに送信された後，音声認識器Juliusで処理される．この手法はw3voice[4]と同様，ブラウザ上で低負荷に音声認識を実現することができる．音声認識後，認識結果はInput Integratorでキーボードからの入力と統合される．その後，統合結果はXISL Interpreterに送られ，シナリオ記述言語XISLの記述に従い出力内容が生成される．出力内容がエージェントの場合，Agent Managerが顔画像合成器Galatea FSMと音声合成器Galatea Talkの出力を用いてエージェント動画像を生成し，ブラウザ上のAgent Presenterに動画を配信することでエージェントから応答が返される．

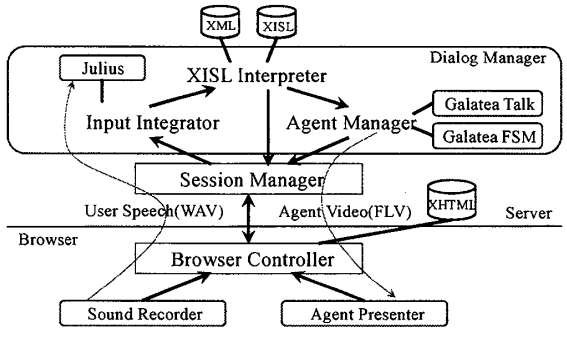

図1　システムの構成

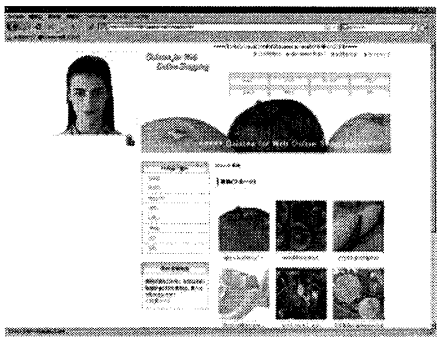

図2　システムのスクリーンショット

3. MMI 6階層モデル

MMI 6階層モデルは，情報処理学会試行標準委員会で提案されたMMIシステムのための階層型アーキテクチャで，MMIの処理レベルに応じてコンポーネントが6つの階層に分割されている．なお，このアーキテクチャでは複数の層を統合して1層として実装するといったことが許されている．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（入出力デバイス層）**：入出力デバイスの制御を行なうラッパープログラムである．
- **第2層（モダリティ依存層）**：音声認識・合成などの各エンジンが，上位層とのインタフェースの整合性を図るためのラッパーがこの層で用意される．
- **第3層（multimodal ⇔ A-modal 変換層）**：入力統合，出力分化，入出力同期制御などを行なう層である．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（A-modal 対話制御層）**：各タスク内の対話制御と応答内容決定を行なう層である．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（タスク制御層）**：対話タスク間の遷移などの制御を行なう層であり，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（アプリケーション層）**：データモデルとアプリケーションロジックを実装する層である．この層では5層に対する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

Flexibly Module Replaceable Web-based Multimodal Interaction System
Masashi Kudo†, Kouichi Katsurada†, Yurie Iribe†, Tsuneo Nitta†
† Toyohashi University of Technology

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．

4. MMI 6 階層モデルに準拠したシステム

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 に示すように MMI 6 階層モデルに準拠する形で再構築した．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，ブラウザのモジュール構成自体は同じであるが，サーバ上の Session Manager 以降（2 層以上）のモジュール構成が大きく異なる．MMI 6 階層モデルに準拠するようモジュール構成を変更したことで，モジュール間の結合度が低くなり，保守性・拡張性が高まった．

なお，今回のシステムでは MMI 6 階層モデルにおける 4 層と 5 層を 1 つの層に統合して実装している．4 層と 5 層を統合した理由は，今回使用した対話シナリオ記述言語（XISL，SCXML[5]）がタスク間遷移とタスク内遷移の双方を記述可能なことから，統合することで実装が容易になったためである．以下，サーバ上のモジュール構成とシステムの機能拡張例について述べる．

4.1 2 層の構成

従来システムでは，Julius，Galatea Talk，Galatea FSM といったエンジンがサーバ内に並立していた．本システムではこれらのエンジンを 2 層に纏めて設置している．特に各エンジンに対するラッパーの置換，追加を容易にするため，ラッパーを Modality Input Manager，Modality Output Manager の配下に置いた．なお，複数のエンジン（音声合成，顔画像合成）が連携して生成するエージェント出力では，Galatea Talk の合成音声を出力するラッパーと，Galatea FSM の合成顔画像を出力するラッパーの 2 つを 2 層の Agent Manager が制御する形で実装した．層間の通信は Modality Input Manager と Modality Output Manager が行なう．これにより，新規のモダリティを追加する際にも，ラッパーとエンジンをマネージャ配下に置くだけで済む．

4.2 3 層の構成

従来システムでは，モダリティ統合と入力統合が同一モジュールで行なわれたため，モジュールが肥大化していた．また，出力管理と対話状態管理についても同様，1 つのモジュールが行なってきたため，モジュールが肥大化していた．

本システムでは 6 層モデルに基づき，肥大化したモジュールの入力統合および出力管理機能を，それぞれ入力統合器 Input Integrator と出力制御器 Output Controller に分離し，3 層に配置した．また，各モジュールを容易に置換・追加できるよう，これらを Input Integrate Manager，Output Control Manager の配下に設置した．これらのマネージャは 2 層と同様，層間通信も管理している．なお入力統合器と出力制御器は XISL を処理するよう設計した．

4.3 4+5 層の構成

この層は MMI 6 階層モデルにおける 4 層と 5 層を 1 つに統合した層であり，対話の状態管理・制御を行なう．ここでは対話記述言語として従来システムと同様に XISL のマルチモーダル入出力に関する記述を利用することを想定している．

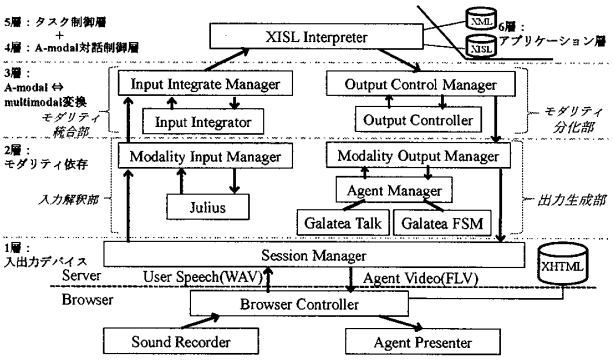

図 3 構築したシステムの構成

4.4 構築したシステムの機能拡張例

4.4.1 出力エンジンの変更

出力エンジンの変更が可能になると，端末環境に応じて高速なエンジン，高品質なエンジンを使い分けることができる．本稿では出力エンジンの切り替え例として，音声合成エンジンを Galatea Talk から Aques Talk[6]に変更した場合，および顔合成エンジン Galatea FSM を自作の連番画像生成エンジンへ切り替えた場合を検証した．エンジン変更に伴うシステム側の変更は，音声合成エンジンと顔画像合成エンジンの配置，およびそれぞれのラッパー構築と追加のみであった．

4.4.2 対話記述言語の変更

対話記述言語を自由に選択できるようになると，開発者は自分の使いやすい対話記述言語を選択できる．本稿では，対話記述言語の変更例として，XISL から SCXML への切替を行なった．SCXML インタプリタのうち，タスク状態管理（4+5 層）には Commons SCXML[7]を用い，またマルチモーダル入出力命令（3 層）には XISL を使用した．この切替によるシステム変更は，インタプリタの構築と置換のみであった．

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．また，構築したシステムの出力エンジンの変更や対話制御および記述言語の変更を行ない，Web ベース MMI システムの拡張性が向上したことを確認した．今後はユーザモデル／デバイスモデルを活用し，利用環境を対話に反映させたり，携帯端末上で対話システムを実現するなどに取り組みたい．

参 考 文 献

[1] 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).

[2] Galatea Toolkit : http://sourceforge.jp/projects/galatea/releases/

[3] 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)

[4] 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).

[5] SCXML : http://www.w3.org/TR/scxml/

[6] Aques Talk : http://www.a-quest.com/aquestalk/

[7] Apache Commons: http://commons.apache.org/scxml