

Web と機器を透過的につなぐ Multimodal Interaction フレームワークの構築

芦村 和幸^{†1} 一色 正男^{†1} 中田 潤也^{†1}
中島 博敬^{†1} 小松 健作^{†2}

様々な入出力機器を Web と透過的に連携させるためのフレームワークである Multimodal Interaction (MMI) 技術の国際標準化, 特に World Wide Web Concorcium (W3C) のにおける取り組みについて概説するとともに, 各種機器等から, より簡単に MMI 技術を利用するために必要となるオープンソフト・ワークベンチの構築について議論する.

Multimodal Interaction Framework Towards Transparent and Smarter Integration of the Web and CE Devices

KAZUYUKI ASHIMURA,^{†1} MASAO ISSHIKI,^{†1}
JUNYA NAKATA,^{†1} HIROTAKA NAKAJIMA^{†1}
and KENSAKU KOMATSU^{†2}

This paper summarizes global standardization activity within the World Wide Web Consortium (W3C) on Multimodal Integration (MMI) technology as a framework for transparent and smarter integration of the Web and various Consumer Electronics devices, and then proposes a universal opensource workbench for MMI-ready systems so that people can use those systems more easily and interaoperably.

1. はじめに

World Wide Web Consortium (W3C)¹⁾ は, Web の可能性を最大限に導き出すべく Web の発展と相互運用性確保に必要な各種プロトコルの開発を行なう産業コンソーシアムである. 本稿では, まず, Web 技術の特徴について簡単に触れた上で, Web の国際的な相互運用性を保証するために必要な技術国際標準化に関する取り組み, 特に Web 技術発明者である Tim Berners-Lee が創設した国際的標準化団体である W3C が取り組む各種標準仕様策定活動の最新動向について概観する. さらに, 近年ますます高度化する Web アプリケーションの課題について議論した上で, 「マルチモーダル・マッシュアップ」という新しいフレームワークを提案するとともに, その実現に必要な課題について議論する.

2. Web 技術とその特徴

World Wide Web (以下, 単に「Web」)²⁾ は, インターネット上で提供される, 相互に接続されたハイパーテキスト文書により構成される情報空間, およびそれにもとづくサービスである. Web へのアクセス手法として, 通常「Web ブラウザ」というソフトウェアが用いられ, Hypertext Markup Language (HTML) 等の文書に記述されたテキスト情報, 画像情報, 動画情報等のマルチメディア情報を取り扱うことができる. 「Web ページにアクセスする (もしくは, Web ページを見る)」とは, 何らかの URI (Universal Resource Identifier) を指定することにより, Web 上に存在する Web サーバに蓄積された文書の位置を特定することに始まり, HTTP (Hypertext Transfer Protocol) 等のプロトコルに則って, 当該文書を PC 等のローカルクライアント (Web クライアント) 上に取得した上で, 文書構造を解析し, ディスプレイ画面等に表示することを意味する.

Web は世界中で共通的に利用されるがゆえに, その相互運用性 (Interoperability), 国際化 (Internationalization), 入出力方法の多様性 (Multi-Modality), および利用者への親和性 (Accessibility) を保証する必要がある, そのために, World Wide Web Consortium (W3C)

^{†1} W3C/慶應 (慶應義塾大学 SFC 研究所)

W3C/Keio (Keio Research Institute at SFC)

^{†2} NTT コミュニケーションズ株式会社

NTT Communications Coporation

等による、国際的な技術標準化が行われている。

3. W3C による Web 技術国際標準化

W3C は、会員組織、専任スタッフ、そして Web 技術の国際標準化に興味を持つ一般の技術者が一丸となって標準仕様の策定に取り組む国際的な産業コンソーシアムであり、Web の相互運用性を確保するとともに、Web を堅牢、スケーラブルかつ適応性のある情報基盤とすることを目的として各種標準仕様や指針の策定に取り組んでいる。

W3C は、アメリカのマサチューセッツ工科大学計算機科学人工知能研究所 (MIT/CSAIL)、フランスに本部を置く欧州情報処理数学研究コンソーシアム (ERCIM)、および日本の慶應義塾大学という 3 つのホスト機関により共同運営されている。また、ホスト機関の存在しない世界各国および地域における標準化活動の普及を推進するための拠点となる W3C オフィスが世界各地に設置されている³⁾。

3.1 W3C の使命

W3C は、ウェブの可能性を最大限に引き出すべく、世界中のあらゆる環境、機器における Web の相互運用性 (interoperability) を確保するとともに、あらゆる機器から (デスクトップ PC から携帯端末からも) 共通で単一の情報にアクセスすることを可能とする (One Web) ことを使命としている。そのために、特定のベンダに依存しない中立な立場 (Vendor-neutral) で、以下のような目標のもとに、各種 Web 技術の標準化を行なっている。

- 誰もが使える (Web for Everyone)
- あらゆる機器からアクセスできる (Web on Everything)
- 全地球規模の知識ベース (Web as Knowledge Base)
- 信頼できる (Trust and Confidence)

4. Web インタラクションの高度化

近年、携帯電話の契約件数は一億件を越え、電子メールの送受信、および経路案内サービス等を含めた Web へのアクセス数はパソコンをしのいでいる。また、デジタルテレビに代表される各種情報家電、および車載用経路案内システム等においても、全地球測位システム (GPS) 等との連携を含めて、やはり、Web 上における位置情報やサービス情報 (番組情報、

店舗情報、各種ニュースなど) のやりとりがさかんである。各サービスベンダごとに、あるいは国内の事業者間で、ある程度のデータ流通形式標準化は進みつつあるものの、あらゆる機器から同じ単一の情報にアクセスし情報を共有し、全ての人々が全ての情報にアクセスできるという「One Web」の観点からは、さらなる標準化が必要であると考えられる。

例えば、携帯電話やタブレットを利用して Web にアクセス場合、通常のデスクトップ PC に備わっている、キーボード、マウス、大型ディスプレイといった Graphical User Interface (GUI) を前提することはできず、音声等限られたモダリティのみを利用する必要があり、音声モダリティを制御するための国際的標準が必要となる。

4.1 HTML5 による、Web のアプリケーション・プラットフォーム化

Web 上の文書記述のために主に利用されている形式として、HyperText Markup Language (HTML)^{4),5)} が挙げられる。HTML は、<html> というように、タグをブラケットで囲んだ「要素」によって記述され、テキスト、表、画像、音声等の情報を含めることができる。Web ブラウザは、HTML で記述された文書を解析した上で、その内容をディスプレイ上に表示したり音声を再生するが、動画情報等、より高度なマルチメディア情報を取り扱うためには、プラグインと呼ばれる拡張を必要としており、各種プラグイン・ソフトウェアの具体的実装内容がブラウザの開発元ごとに異なることから、標準的な取扱いに問題があった。そのため、様々な機器やモダリティに対応し、より表現力豊かな Web アプリケーションを実現するための標準的な Web アプリケーションのプラットフォームとして HTML5^{6),7)} が提案されている。

HTML5 は、HTML が元来持つ基本的な GUI 制御機能に加え、よりきめ細かな描画形式の制御、および動的 (Dynamic) かつ Web と利用者との相互的なやりとりに対応 (Interactive) したものとなっている。HTML5 で新たに追加された機能の代表例として以下が挙げられる。

Video & Audio:

プラグイン・ソフトウェアを使わずに動画や音声を再生

Canvas:

2次元画像の描画

Drag & Drop:

画像やテキストの移動・動的編集

Web Storage:

Web クライアント上でのデータ保存

Web Sockets:

HTTP にもとづく通常の Web サーバとの通信と異なる、双方向かつ全二重を可能とするリアルタイム通信

Web Workers:

マルチプロセス制御

上記機能に関する具体的なデモンストレーション^(8),9) が Web 上で公開されているので参照いただきたい。

4.2 マルチモーダル・インタラクション

より表現力豊かな Web アプリケーションを実現するためのプラットフォームとなるべく、HTML5 には、既存の HTML に対して非常に多種多様な機能追加が行なわれており、標準化された手法にもとづいて、動的かつインタラクティブな Web アプリケーションを構築することが可能となりつつある。その一方で、いくつかの技術的課題も指摘されており、特に、以前の HTML (XHTML 1.1¹⁰⁾ 等) では可能とされていた、XML の「名前空間」の仕組みを使った機能拡張や、様々な機器・モダリティへの対応に課題があるとされている。現状では、例えば、ベクトル画像 (SVG¹¹⁾) や数式 (MathML¹²⁾) については、HTML5 の語彙として追加する方向で検討が進んでいる、また、携帯電話、タブレット、デジタルテレビ等の機器操作については、HTML5 の文法体系とは切り分け、JavaScript の API として取り扱う方針で検討が進んでいるが、様々な機器 (例えばテレビやデジタルカメラ) をどのように分類整理し制御するか議論が必要とされている。

このような問題点に対する対策の一つとして、W3C では、Web におけるマルチモーダルインタラクションに関する標準化が進められている。マルチモーダルインタラクションでは、Web アクセスのための入出力手段として、GUI に加え、音声や手書き入力、ジェスチャ等、様々なモダリティを組み合わせる利用が想定されている。

なお、当初、マルチモーダルインタラクションの典型として、HTML による GUI と VoiceXML による音声インタフェースの統合化が検討されていた (XHTML+VoiceXML 等¹³⁾) が、近年では、携帯電話やタブレット、さらにはデジタルテレビ、ゲーム機器、エアコン

等の各種家電をも視野に入れ、Web 上で利用される様々な機器を透過的かつ統一的に取り扱うためのフレームワークとして、Multimodal Interaction Architecture (MMI Architecture; MMI アーキテクチャ)¹⁴⁾ が検討されている。

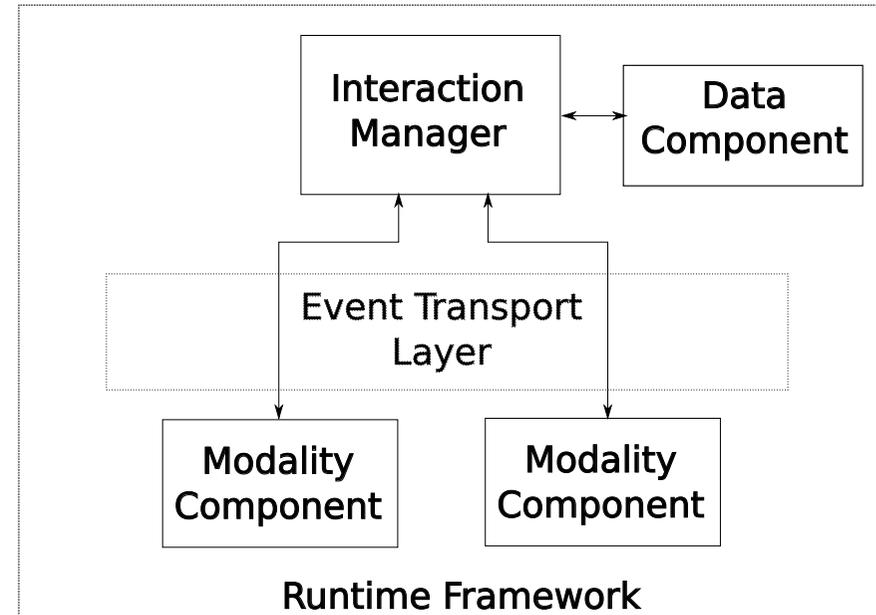


図 1 W3C MMI アーキテクチャ
Fig. 1 W3C MMI Architecture

- MMI アーキテクチャの特徴として、以下が挙げられる。
- 構成コンポーネントのカプセル化:
 - 各構成コンポーネントの実装内容について特定の仮定をしないこと
 - 分散環境への対応:
 - 特定のハードウェアおよびソフトウェア上に実装されて機能のみならず、ネットワーク上に分散した機能を透過的に統合して利用できること
 - 容易な拡張性:

HTML, VoiceXML, SVG, MathML 等, 既存の Web アプリケーションのための標準的な記述言語を用意に再利用できること

再帰的利用:

複数のコンポーネントの集合を新たに一つのコンポーネントとして規定できる「入れ子構造」を可能とすること

モジュール構造:

データ, 制御, 表示の機能を明確に分離できること

図 1 に MMI アーキテクチャの概念図を示す. また図中の各構成コンポーネントについて以下に概説する.

Runtime Framework (ランタイム・フレームワーク):

MMI アーキテクチャの動作環境

Modality Component (モダリティ・コンポーネント):

ネットワーク上に分散する各機器の入出力機能を制御. MMI アーキテクチャにおいて, 当該機器をクライアントとしてとして仮想化. 礼として, HTML を処理する Web ブラウザや VoiceXML を処理する音声ブラウザが挙げられる.

Interaction Manager (インタラクション・マネージャ):

MMI ライフサイクルイベントと呼ばれるイベントの集合を用いて, Web アプリケーションの処理の流れ (アプリケーション・ライフサイクル) を制御.

Data Component (データ・コンポーネント):

Web アプリケーションの動作に必要な各種データを蓄積

5. マルチモーダル・マッシュアップの提案

近年, 地図サービスや音声案内サービス等, 複数の Web サービスの API を有機的に連携させ, あたかも一つの複雑な Web サービスであるかのように機能させる手法 (マッシュアップ) が広く利用されている. JavaScript 等により, Web 上で公開されている各種 API を組み合わせることで, 比較的容易かつ短期間で複雑な Web アプリケーションを構築することができることから注目されているが, 各 API は開発元やサービス提供者ごとに異なるのが現状であり, 様々な機器やサービスを制御するための標準的な API が必要とされている. そ

こで, 我々は, MMI アーキテクチャのライフサイクル・イベントに加え, コンポーネント間でやりとりされるデータ形式として, マルチモーダル Web アプリケーション向けの汎用的データ形式である EMMA¹⁶⁾ を利用した「マルチモーダル・マッシュアップ」を提案する.

マルチモーダル・マッシュアップ手法により, 図 2 に示す通り, 携帯電話等に搭載されたインタラクションマネージャを用いて, ネットワーク上に分散する各種機器およびサービスを透過的かつ統合的に制御することが可能となる. マルチモーダル・マッシュアップの特徴として以下が挙げられる.

- ネットワーク上に分散する複数の機器上の様々なモダリティを, ライフサイクル・イベントという統一的なイベント操作の仕組み, および EMMA という標準的なデータ形式により連携させることが可能
- デジタルテレビやビデオ等, 自宅のホームネットワーク (基本的にプライベートなネットワーク) 上の機器を, Web アプリケーション (いわゆるクラウド・サービス) と連携させることが可能
- 既存のハードウェアおよびソフトウェアに対して, ライフサイクル・イベントの操作機能を付加することにより, モダリティ・コンポーネントとして再利用することが可能
- 通常の HTTP ベースの Web アプリケーションでは取扱いが困難な, アプリケーションのライフサイクル (処理の開始と終了の規定, 双方向かつリアルタイムなメッセージング等) がライフサイクル・イベントにより明確かつ簡潔に規定可能

6. マルチモーダル・マッシュアップ実現のための課題と対策

マルチモーダル・マッシュアップを実現するためには, 多様な入力形態, 入出力の同期, ある処理から派生する別処理との通信など, ネットワーク制御に関する複雑な処理に関する知識が必要であり, 既にシステム実装のためのガイドラインやオーサリング・ツールの充実している通常の Web アプリケーション構築に比べると難易度が高い. そこで, MMI アーキテクチャの仕様策定に取り組む W3C Multimodal Interaction Working Group¹⁷⁾ では, 専門のタスクフォース (Interoperability Testing Task Force) を設立し, 相互運用性の検証, および実装ガイドラインの作成に取り組んでいる. 具体的には, 複数の開発元が分担して各種コンポーネント (インタラクションマネージャおよびモダリティコンポーネント) を開発し,

MMI Architecture and EMMA for Multimodal Mashup

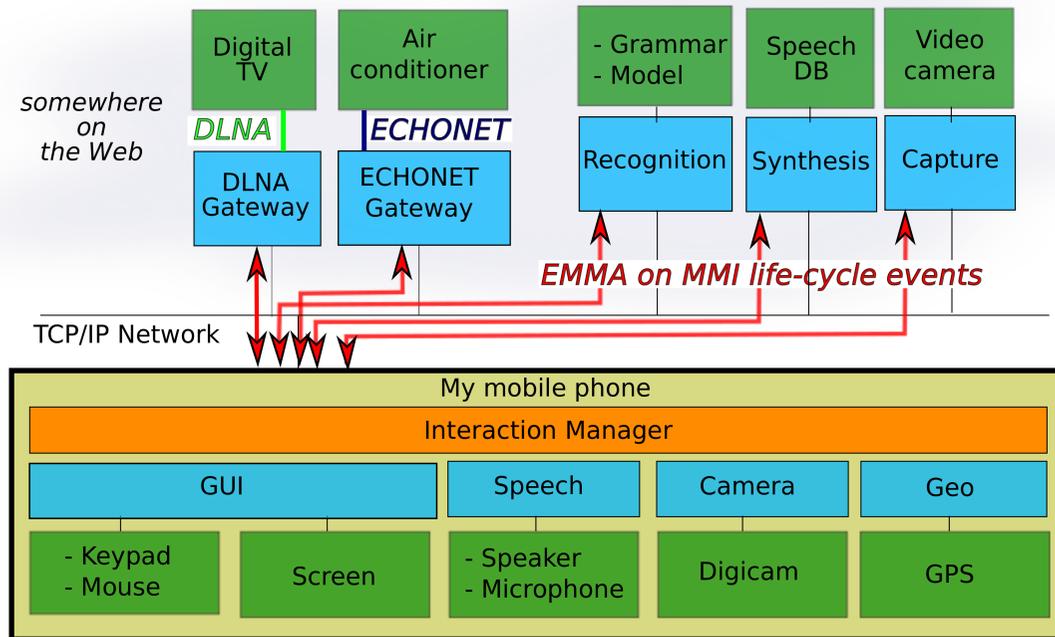


図 2 マルチモーダル・マッシュアップの例
Fig. 2 Example of Multimodal Mashup

実際にマルチモーダル・マッシュアップ・システムとして結合した上で、動作確認および問題点の洗い出しを行なっている。

一方、著者らは、新世代の Web アプリケーション・プラットフォームである HTML5 に対応した各種 Web ブラウザをモダリティコンポーネントとして MMI アーキテクチャに組み込むべく、上記タスクフォースの中でイベント操作機能のライブラリ化に取り組んでいる。このライブラリでは、HTML5 の新機能の一つである Web Sockets を応用することにより、

Web ブラウザからライフサイクル・イベントを利用することが可能となり、ブラウザを透過的にマルチモーダル・マッシュアップの枠組に組み込み、ネットワーク上に分散する複数機器上のモダリティを統合的かつ標準的な方法で取り扱うことができるようになる。

なお、上記、ガイドライン、相互運用性確認用コンポーネント、およびイベント操作機能ライブラリについては、随時、オープンソースのマルチモーダル・ワークベンチとして一般に公開し、マルチモーダル Web アプリケーション構築の一助とするとともに、その有効性につ

いても検証していく予定である。

7. おわりに

本稿では、まず、Web 技術の特徴について簡単に触れた上で、Web の国際的な相互運用性を保証するために必要な技術国際標準化に関する取り組み、特に Web 技術の国際的標準化団体である W3C が取り組む各種標準仕様策定の最新動向について概観した。またさらに、近年高度化する Web アプリケーションの課題について議論した上で、「マルチモーダル・マッシュアップ」という新しいフレームワークを提案し、その実現に必要な課題としてシステム実装のためのガイドラインやオーサリング・ツールの不足を挙げた。さらに具体的対策として W3C Multimodal Interaction Working Group の取り組みに触れるとともに、著者ら自身の取り組みとして、HTML5 対応 Web ブラウザを MMI アーキテクチャのモダリティコンポーネントとするためのライブラリ開発について触れた。著者らは、引き続き、マルチモーダル・マッシュアップ実現のために、上記ライブラリの完成を目指すとともに、ライブラリの実装が進み次第、その有効性についての検証も行なう予定である。

参 考 文 献

- 1) W3C: W3C Top Page, available from (<http://www.w3.org/>)
- 2) Wikipedia: World Wide Web, available from (http://en.wikipedia.org/wiki/World_Wide_Web)
- 3) W3C: W3C Offices, available from (<http://www.w3.org/Consortium/Offices/staff.html>)
- 4) Wikipedia: HTML, available from (<http://en.wikipedia.org/wiki/HTML>)
- 5) W3C: HTML 4.01 Specification, available from (<http://www.w3.org/TR/html401/>)
- 6) Wikipedia: HTML5, available from (<http://en.wikipedia.org/wiki/HTML5>)
- 7) HTML5 Specification, available from (<http://www.w3.org/TR/html5/>)
- 8) Google: HTML5 ROCS, available from (<http://www.html5rocks.com>)
- 9) Mozilla: Audio Data API, available from (https://wiki.mozilla.org/Audio_Data_API)
- 10) W3C: XHTML Specification, available from (<http://www.w3.org/TR/xhtml11/>)
- 11) W3C: SVG Tiny 1.2 Specification, available from (<http://www.w3.org/TR/SVGTiny12/>)
- 12) W3C: MathML Version 3.0 Specification, available from (<http://www.w3.org/TR/MathML/>)
- 13) W3C: XHTML+Voice Profile 1.0 Note, available from (<http://www.w3.org/TR/xhtml+voice/>)
- 14) W3C: MMI Architecture Specification, available from (<http://www.w3.org/TR/mmi-arch/>)
- 15) Wikipedia: Mashup, available from ([http://en.wikipedia.org/wiki/Mashup_\(web_application_hybrid\)](http://en.wikipedia.org/wiki/Mashup_(web_application_hybrid)))
- 16) W3C: EMMA Specification, available from (<http://www.w3.org/TR/emma/>)
- 17) W3C: Multimodal Interaction Working Group, available from (<http://www.w3.org/2002/mmi/>)