

# 講演3 W3Cにおける音声・マルチモーダルインタフェースへの取り組み – Voice Browser & Multimodal Interaction の活動 –

芦村 和幸<sup>†</sup>

<sup>†</sup>World Wide Web Consortium (W3C)

## 1 はじめに

World Wide Web Consortium (W3C) 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を確保するために必要な各種プロトコルの開発を行なっている。

本稿では、まず、W3C の組織構成について概説した上で、W3C の Voice Browser Working Group および Multimodal Interaction Working Group で取り組んでいる、音声・マルチモーダル技術を利用した Web アクセスに関する各種仕様策定の動向について説明する。

音声・マルチモーダル技術は、様々な環境や条件における人々を対象に Web アクセスの利便性を向上させることが可能な、次世代のヒューマンインタフェース技術として期待されている。しかし、携帯端末、カーナビゲーションシステム、情報家電など各種機器を利用した情報アクセス手法は、各開発ベンダごとに異なる部分が多いのが現状であり、開発やメンテナンスのためのコスト削減という観点からも、各種プロトコルの国際標準化が急務である。

## 2 W3C について

W3C は、会員組織、専任スタッフ、そして一般の皆様が一丸となって Web 標準の策定に取り組む、国際的な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の 3 ホスト機関により共同運営されている。また、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オフィスが、世界各地に設置されている [1]。なお、W3C の組織構成に関する詳細に関しては、W3C の Web サイト [2, 3, 4] を参照されたい。

**W3C 慶應ホスト** W3C を共同運営する 3 つのホスト機関のうち、東アジア地区を担当するホストであり、会員および一般向けに様々なサービスを提供している。特に、日本国内向けには、報道発表の日本語版や、日本会員会議など、日本語による情報提供も行なっている。2006 年 9 月現在、10 名のスタッフが標準化活動

に従事している。

## 3 国際標準化の重要性

近年、携帯電話による電子メールや Web へのアクセスはパソコンをしのいでいるが、小型かつ多機能化が進むほどに、そのメカニズムおよび操作方法はますます複雑化し、開発サイドにとって、安定動作を保証するために必要な、様々な操作パターンの組合せを考慮した網羅的試験を行なうことが困難となっている。それに加え、事業者間 (キャリア、メーカ) の競争激化により、音楽機能、外部メモリ、TV 放送の受信など、さらに複雑な新機能を持つ機種を短期間のうちに開発・市場投入する必要もある。

各メーカは新機種の開発・販売に膨大な投資を要求されるため、開発コスト高騰および製品不良などのリスクをメーカが独力で全て負担するのは非常に困難となりつつある。したがって、何らかの形でリスクを分散する方法が必要とされ、他社との共同開発によりコスト削減と開発期間の短縮に取り組むメーカも見られる。

今後は、こうした開発統合の動きが活発化する可能性が高いと考えられるが、今日のグローバル経済社会においては、同一のハードウェアおよびソフトウェアから構成される基本プラットフォームを複数の製品および国際市場で流用することも多く、長期的に安定した相互運用が可能な製品の開発に必要な、品質保証や仕様統一に関する国際標準化の枠組が、ますます重要性を増すと考えられる。

## 4 W3C による国際標準化の取り組み

W3C による国際標準化の取り組みについて、以下に概説する。

### 4.1 組織構成

W3C は、以下に示す W3C Member と W3C Team により構成される。

#### W3C Member (W3C 会員):

W3C における世界規模での Web 技術標準化に賛同し、会員登録をする企業、研究所、大学、政府機関、NGO、NPO、ユーザ団体などの組織。2006 年 8 月 28 日現在 420 組織 (うち日本会員は 35 組織)。

#### W3C Team (W3C Team スタッフ):

W3C の仕様策定に関わる、(1) WG の活動を支援、監督する技術的専門家、(2) 組織運営およびその補佐、および (3) 広報、システムサポート担当者。米国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)、他に本部を置く欧州情報処理数学研究コンソーシアム (ERCIM)、

W3C's Standardization Activity on Multimodal Interaction & Voice Browser

<sup>†</sup> Kazuyuki Ashimura (ashimura@w3.org)

Multimodal Interaction WG & Voice Browser WG, World Wide Web Consortium (W3C) (<sup>†</sup>)  
W3C/Keio University, Keio Research Institute at SFC,  
5322 Endo, Fujisawa, Kanagawa 252-8520, Japan

慶應義塾大学 (SFC 研究所) のいずれかのホストに所属する。2006 年 8 月 28 日現在 66 名 (うち慶應ホストには 10 名)。

#### 4.2 ワーキンググループ (WG) による仕様策定活動

W3C の標準仕様策定活動は、ワーキンググループ (WG) 単位で取り組まれる。WG は、具体的な技術仕様やガイドラインの策定を行う活動の最小単位であり、会員企業から参加する技術者と、Team から参加する技術スタッフ (Team スタッフ) で構成される。

各 WG はそれぞれ、一人もしくは二人の議長により、仕様策定のスケジュールや、各種会議の進行を管理される。議長は主に会員側参加者から選出され、Team スタッフは Team Contact として議長を補佐し、WG の円滑な活動を支援する。なお、議長および Team Contact の判断にもとづき、仕様策定活動に必要な外部の専門家を Invited Expert として招聘することも可能である。仕様策定活動への参加者は、各 WG へ個別に参加登録を行なうため、その構成人数は、グループごとに数名から数十名規模までさまざまに異なる。

WG には世界中の W3C 会員企業等から技術者が参加しており、直接の打ち合わせを頻繁に開催することが困難であるため、標準仕様策定にあたっては、主に、週に 1, 2 回程度開催される電話会議を利用して意識合わせや進捗確認を行なう。また、2, 3 か月程度に 1 回は、いずれかの WG 参加者が主催者となって一堂に会し、綿密な打ち合わせを行なう (Face-to-Face 会議)。なお、日常の議論や情報交換においてはメーリングリストおよび IRC が活用され、各種情報の蓄積、共有および閲覧には Web が利用される。

**W3C Team Contact** Team Contact は、W3C Team から WG に参加する技術スタッフであり、議長を補佐しながら、WG メンバによる円滑な仕様策定活動を支援する。Team Contact の主な業務は以下の通り。

- WG 運営管理のための各種ドキュメント作成
  - Activity Statement [16, 20]
  - WG Charter [17, 21]
  - 一般公開用 Web page [19, 23]
  - WG 内部用 Web Page (W3C 会員様限定; 一般非公開)
- W3C 勧告など各種ドキュメントの公開
- 電話会議, Face-to-Face 会議, ワークショップなど WG の活動への参加
- その他, WG 運営に関する業務全般

#### 4.3 W3C Recommendation Track

W3C で策定する標準仕様は、以下の 6 段階を経て W3C 勧告として公開される。

##### 公開草案初版 (First Public Working Draft; FPWD):

当該活動領域における仕様策定を W3C 内外に示すために最初に公開される草案で、W3C 特許方針に基づく特許請求除外期間が設定される。

##### 草案 (Working Draft; WD):

FPWD から LCWD に至る過程で更新される草案で、LCWD や CR, PR の段階で問題が生じた場合に差し戻されることもある。

##### 最終草案 (Last Call Working Draft; LCWD):

一通り技術的な問題を解決して仕様が決まり WG 内で合意が得られた最終草案で、これ以降は技術的な変更は行われない。

##### 勧告候補 (Candidate Recommendation; CR):

実装試験期間。

##### 勧告案 (Proposed Recommendation; PR):

会員による投票期間。

##### 勧告 (Recommendation; Rec):

技術統括責任者 (Director) である Tim Berners-Lee の承認を経て、勧告として公開される。

#### 4.4 W3C の活動への参加方法

W3C は会員制の産業コンソーシアムであり、WG で行なう仕様策定活動に参加するためには、参加企業等の本部が置かれている場所にもとづき、W3C の 3 箇所のホスト (MIT/ERCIM/慶應) のいずれかに会員登録を行なう必要がある。各 WG による具体的な標準化作業に参加するためには、各標準仕様ごとに W3C Royalty Free Licence に同意する必要があるため、会員登録後、別途、各 WG への参加登録が必要である。

W3C 会員には、企業/研究機関等の規模に応じて、以下に示す通り、Full / Intermediate / Affiliate という 3 種類の種別があるが、これらは会員企業等の規模に応じて年会費が異なるのみで、会員としての権利の違いは一切ない。なお、以下の年会費額は、2006 年 1 月以降に日本会員として登録した場合の金額である。

##### Full Member:

総収入が年間 230 億円もしくは、それ以上の営利組織。会費年額は 730 万円。

##### Intermediate Member:

総収入が年間 57 億 5000 万円もしくはそれ以上、かつ 230 億円未満の営利組織。会費年額は、292 万円。2005 年 4 月から 2008 年までの 3 年間、暫定的に設置されている会員種別であり、当該期間中に、新規に 20 組織以上が本会員種別に登録すると、正式に運用される予定である。

##### Affiliate Member:

非営利組織および、上記に当てはまらない組織 (営利組織を含む)。会費年額は、73 万円。

#### 4.5 会員企業等の特許に関する留意点

仕様規定に本質的な技術 (normative technology) は全て、W3C 勧告として公開される時点において、W3C Royalty Free Licence にもとづいて提供されなければならない。そのため、各 WG に参加登録する時点で、

各会員企業等は、W3C Royalty Free Licence への同意が必要である。

なお、Royalty Free Licence の詳細については、W3C サイト上の関連資料 [24, 25, 26, 27, 28] を参照されたい。

## 5 音声・マルチモーダル技術に関する仕様策定

音声・マルチモーダル技術の標準化に関しては、VoiceXML など、基本的なマークアップ言語仕様が W3C の Voice Browser WG で標準化され、対話型音声応答 (IVR) サービスを中心に利用されている。また、マルチモーダル対話 (Multimodal Interaction; MMI) の記述についても、Multimodal Interaction WG において標準化へ向けた作業が現在進行中である。

各 WG の活動について、以下に概説する。

### 5.1 Voice Browser WG の活動

Voice Browser WG (表 1) は、Web 技術と音声技術を統合することにより、世界中のあらゆる電話 (音声端末) から音声もしくは簡単なキー操作を用いて、さまざまなサービスにアクセスできるようにするために必要とされる、各種標準仕様 (W3C Speech Interface Framework; 図 1) の策定を目的としている。

Voice Browser WG の活動の詳細については、W3C Web サイト上の関連資料 [17, 19] を参照されたい。

表 1: Voice Browser WG

参加者	90 人 (38 組織; W3C 最大の WG)
策定対象の仕様	<ul style="list-style-type: none"> <li>• VoiceXML 2.1 [5]</li> <li>• VoiceXML 3.0 (DFP Overview [6])</li> <li>• SSML [7]</li> <li>• SRGS [8]</li> <li>• PLS [9]</li> <li>• SISR [10]</li> <li>• CCXML [11]</li> <li>• SCXML [12]</li> </ul>
会議	<ul style="list-style-type: none"> <li>• 電話会議 (週 6 回)</li> <li>• F2F 会議 (年 3-4 回)</li> </ul>
情報共有	<ul style="list-style-type: none"> <li>• Mailing List (一般公開用 [18] / 内部用)</li> <li>• Web (一般公開用 [19] / 内部用)</li> </ul>

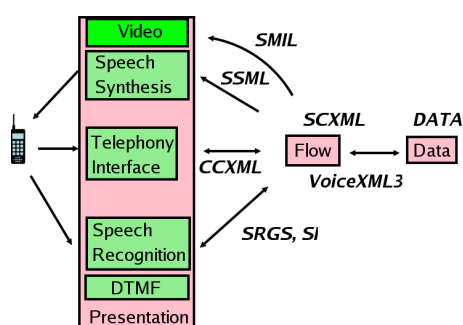


図 1: Speech Interface Framework

SSML の国際化対応 近年、中国やインドを含むアジア、中東および東欧など、Web 利用においてこれまで十分に考慮されてこなかった様々な言語を用いた Web コンテンツの著しい増加が見込まれている。W3C Voice Browser WG は、音声合成記述言語 SSML (Speech Synthesis Markup Language) の仕様策定に取り組んでいるが、北京語、日本語、韓国語、ヒンディ語、アラビア語、ロシア語、ヘブライ語など、世界中のより広範な言語に、よりきめ細かく対応できるように、SSML 1.0 の更新版である SSML 1.1 の策定に着手した [29, 30]。

SSML 1.1 策定活動の一貫として、2005 年 11 月 2 日 および 3 日、北京において第 1 回目 [31] の、そして 2006 年 5 月 30 日 および 31 日、ギリシャのクレタ島において第 2 回目の「SSML の国際化に関するワークショップ」 [32] が開催され、音声合成のさらなる品質改善に必要な課題の洗い出しが行われた。

上記ワークショップでは、(1) 音調、(2) 音節単位での強勢やアクセント、あるいは (3) 音の長さなど、音声言語の特徴を表現する機能を機械処理可能な方法で標準へ追加する必要性が指摘されており、今後は、世界中のより多くの言語をこれまで以上に適切に取り扱えるよう、SSML の機能を拡張するための機構に関する検討が進められる。

### 5.2 Multimodal Interaction WG の活動

Multimodal Interaction WG (表 2) は、GUI、音声、画像、ペン入力、ジェスチャー、触覚など、多様なアクセス方法に対応するよう Web を拡張するとともに、誰でも、どこでも、どのような装置からでも、そしていつでも、利用者にとって最も好ましい方法で Web にアクセスし、装置、利用者および環境要因に応じた最適なサービスを手に入れる (図 2) ことを可能とするために必要な、各種標準仕様の策定を目的としている。

Multimodal Interaction WG の活動の詳細については、W3C Web サイト上の関連資料 [21, 23] を参照されたい。

表 2: Multimodal Interaction WG

参加者	53 人 (29 組織)
策定対象の仕様	<ul style="list-style-type: none"> <li>• Authoring</li> <li>• Modality Component Interface Framework [13]</li> <li>• EMMA [14]</li> <li>• InkML [15]</li> </ul>
会議	<ul style="list-style-type: none"> <li>• 電話会議 (週 4 回)</li> <li>• F2F 会議 (年 3-4 回)</li> </ul>
情報共有	<ul style="list-style-type: none"> <li>• Mailing List (一般公開用 [22] / 内部用)</li> <li>• Web (一般公開用 [23] / 内部用)</li> </ul>

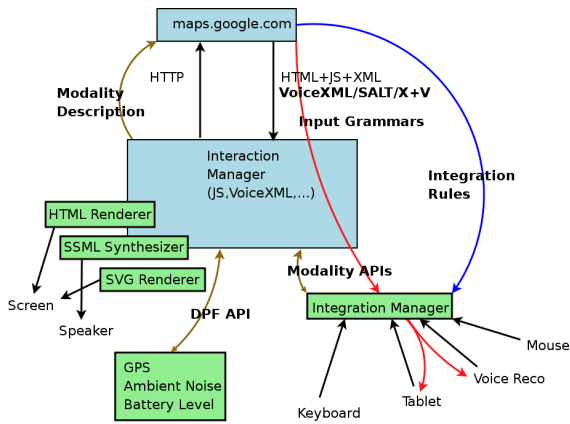


図 2: Multimodal Architecture

多様なアプリケーションへの対応 マルチモーダル・アプリケーション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、複雑な処理が必要である。また、既存の Web ブラウザと同様に、利用者の個人情報を特定したり保護する必要もあるため、話者識別などの応用についても検討を要する。さらには、利用者の環境情報(位置、気象情報など)をいかに処理に反映するかという、今まで取り扱われてこなかった、新たなデータ処理も必要となる。

## 6 日本の技術と国際標準

我が国においても、様々な組織で音声・マルチモーダル対話、およびその標準化に関する検討が進められており、その中には、電子情報技術産業協会 (JEITA) の音声入出力方式標準化委員会 (以下、単に JEITA) や、音声対話技術コンソーシアム (ISTC) の MMI 記述言語検討ワーキンググループ (SIG-MMI-WG) など、各団体が独自に行なった標準化のノウハウを W3C にフィードバックしている例もある。JEITA は、2005 年 11 月に北京で開催された第 1 回の「SSML 国際化に関するワークショップ」に参加し、(1) 日本語音声合成用記号、(2) モーラ単位の時間指定、および (3) ルビ情報の活用に関する提案を行なった [31]。また、SIG-MMI-WG では、独自にユースケース抽出、要求仕様洗い出し、および MMI アーキテクチャの提案等、MMI 記述言語の策定作業を進めているが、その成果は、W3C 会員である豊橋技術科学大学によって、今後、W3C Multimodal Interaction WG へ提案される予定である。

我が国の優れた研究開発のさらなる発展のためにも、仕様標準化および品質保証のためのワークベンチとして、日本の企業の皆様に W3C の標準化活動へ積極的にご参加いただき、優れた技術およびアイデアをタイムリーに国際標準に反映していただくことが非常に重要であると考えられる。

## 7 おわりに

W3C の組織構成について概説した上で、Voice Browser WG および Multimodal Interaction WG で取り組んでいる、音声・マルチモーダル技術を利用した Web アクセスに関する各種仕様策定の動向につい

て説明した。

音声・マルチモーダル技術は、様々な環境や条件における人々を対象に Web アクセスの利便性を向上させる、次世代のヒューマンインタフェース技術として期待されており、我が国でも、様々な研究開発の取り組みが行なわれているが、現状では、各開発ベンダごとに異なる部分が大きく、開発やメンテナンスのためのコスト削減という観点からも、各種プロトコルの国際標準化が急務である。

お問い合わせ先 (ご質問, ご参加希望等)

慶應義塾大学 SFC 研究所 W3C

〒252-8520 神奈川県藤沢市遠藤 5322

電話: 0466-49-1170 Fax: 0466-49-1171

email: keio-contact@w3.org

担当: 平川, 小野塚, 芦村

## 参考文献

- [1] <http://www.w3.org/Consortium/Offices/staff.html>
- [2] <http://www.w3.org/>
- [3] <http://www.w3.org/Consortium/>
- [4] <http://www.w3.org/2002/03/new-to-w3c>
- [5] <http://www.w3.org/TR/2005/CR-voicexml21-20050613/>
- [6] <http://www.w3.org/Voice/2006/DFP>
- [7] <http://www.w3.org/TR/speech-synthesis/>
- [8] <http://www.w3.org/TR/2004/REC-speech-grammar-20040316/>
- [9] <http://www.w3.org/TR/2006/WD-pronunciation-lexicon-20060131/>
- [10] <http://www.w3.org/TR/2006/CR-semantic-interpretation-20060111/>
- [11] <http://www.w3.org/TR/2005/WD-cxhtml-20050629/>
- [12] <http://www.w3.org/TR/2006/WD-scxhtml-20060124/>
- [13] <http://www.w3.org/TR/mmi-arch/>
- [14] <http://www.w3.org/TR/emma/>
- [15] <http://www.w3.org/TR/InkML/>
- [16] <http://www.w3.org/Voice/Activity.html>
- [17] <http://www.w3.org/2004/11/voice-charter.html>
- [18] <http://lists.w3.org/Archives/Public/www-voice/>
- [19] <http://www.w3.org/Voice/>
- [20] <http://www.w3.org/2002/mmi/>
- [21] <http://www.w3.org/2004/03/mmi-charter.html>
- [22] <http://lists.w3.org/Archives/Public/www-multimodal/>
- [23] <http://www.w3.org/2002/mmi/>
- [24] <http://www.w3.org/2005/10/Process-20051014/>
- [25] <http://www.w3.org/2005/07/pubrules/>
- [26] <http://www.w3.org/Consortium/Patent-Policy-20040205/>
- [27] <http://www.w3.org/2003/12/22-pp-faq.html>
- [28] <http://www.w3.org/2004/01/pp-impl/>
- [29] <http://www.w3.org/2006/08/ssml-pressrelease.html.ja>
- [30] <http://lists.w3.org/Archives/Public/www-voice/2006JulSep/0000>
- [31] <http://www.w3.org/2005/08/SSML/ssml-workshop-agenda.html>
- [32] <http://www.w3.org/2006/02/SSML/minutes.html>