

# Voice Extensible Markup Language (VoiceXML) 3.0 の話者認識機能

## W3Cの取り組む音声アプリケーション記述言語の国際標準化動向

芦村 和幸†

† W3C/慶應 (慶應義塾大学 SFC 研究所) 〒252-8520 神奈川県藤沢市遠藤 5322

E-mail: †ashimura@w3.org

**あらまし** 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う国際的標準化団体である。本稿では、W3C の組織構成について概説した上で、W3C の音声ブラウザ・ワーキンググループ (Voice Browser Working Group) が取り組む音声アプリケーション記述のための各種マークアップ言語のうち、対話記述のための次世代フレームワークである Voice Extensible Markup Language (VoiceXML) 3.0 策定の動向、特にその話者認識機能について説明する。

**キーワード** World Wide Web Consortium (W3C), 国際標準化, 音声ブラウザ・ワーキンググループ, VoiceXML 3.0, 話者認識技術

## Speaker Identification and Verification Functionality of Voice Extensible Markup Language (VoiceXML) 3.0

### W3C's International Standardization Activity on a Markup Language for Speech Applications

Kazuyuki ASHIMURA†

† W3C/Keio (Keio Research Institute at SFC) 5322 Endo, Fujisawa, Kanagawa, 252-8520 Japan

E-mail: †ashimura@w3.org

**Abstract** The World Wide Web Consortium (W3C) is an international consortium where Member organizations, a full-time staff, and the public work together to develop Web standards. W3C's mission is "To lead the World Wide Web to its full potential by developing protocols and guidelines that ensure long-term growth for the Web." This document first explains the basic structure of the consortium, and then describes the Voice Extensible Markup Language (VoiceXML) 3.0, the next generation dialog framework developed by the Voice Browser Working Group, and its Speaker Identification Verification (SIV) functionality.

**Key words** The World Wide Web Consortium (W3C), International standard, The Voice Browser Working Group (VBWG), The Voice Extensible Markup Language (VoiceXML) 3.0, Speaker Identification Verification (SIV)

#### 1. ま え が き

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 の音声ブラウザ・ワーキンググループ (Voice Browser Working Group) が取り組む音声アプリケーション記述のための各種マークアップ言語のうち、対話記述のための次世代フレームワークである Voice Extensible Markup Language

(VoiceXML) 3.0 策定の動向について説明する。

音声モダリティは、様々な環境や条件下において 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多様な応用が期待されている。

また、近年、携帯電話、携帯情報端末、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全地球規模での情報取得、配信が可能となりつつある一方、具体的なデータへのアクセス手法は、今なお、各機器やサービスごとに異なる部分が多いのが現状であり、あらゆる機器から (デスクトップ PC から携帯端末からも) 同じ単一の情報にアクセスすることに

より、全ての人々が全ての情報にアクセスできるという「One Web (ひとつの Web)」の観点から見て、情報共有のための国際的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 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 [2]。なお、組織構成の詳細に関しては W3C の Web サイト [1], [3], [4] を参照されたい。

### 2.1 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)、共通で単一な情報にアクセスすることにより、全ての人々が全ての情報にアクセスできるようにする (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)

### 2.2 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。

## 3. 国際標準化の重要性

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することにより、全ての人々が全ての情報にアクセスできるという「One Web (ひとつの 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、携帯電話のように音声を主なモダリティとして利用する機器のみならず、既存のデスクトップ PC や、車載端末等を含めて、音声モダリティを制御するための、国際的に利用可能な標準的データ形式、およびその制御コマンド等を制

定する必要がある。

## 4. W3C による国際標準化

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、以下に示す通り、W3C 会員 (W3C Member) と W3C チーム (W3C Team) により構成される。

#### 4.1.1 W3C 会員 (W3C Member)

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。2009 年 3 月現在、409 組織 (うち日本会員は 30 組織)。

#### 4.1.2 W3C チーム (W3C Team)

W3C の仕様策定を支援するために雇用され、以下の業務に携わる担当者。

- W3C 会員が行なう標準化活動を支援する技術的専門家
- 広報担当、システムサポート担当
- 組織運営およびその補佐

W3C チーム員は、コンソーシアムのホスト機関である、MIT CSAIL, ERCIM, もしくは慶應義塾大学のいずれかに所属する。2009 年 3 月現在 63 名 (うち W3C/慶應所属は 7 名)。

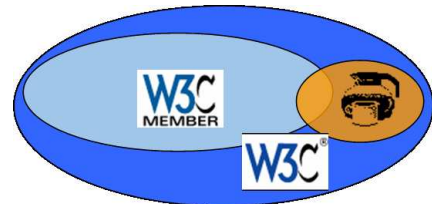


図 1 W3C Member and W3C Team

### 4.2 ワーキンググループ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) 単位で行なわれる。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、チームから参加する技術スタッフで構成される。

#### 4.2.1 ワーキンググループの構成

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、チーム員はチーム・コンタクト (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、議長およびチーム・コンタクトの判断にもとづき、仕様策定活動に必要な外部の専門家を招聘専門家 (Invited Expert) として WG に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 で取り組む標準化活動への興味や必要性に応じて個別に参加登録を行なう。各 WG の構成人数は、グループごとに数名から数十名規模まで様々に異なる。

#### 4.2.2 ワーキンググループの活動

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、各 WG で取り組む標準仕様書ごとに、週に 1, 2 回程度の (国際電話による) 電話会議を開催して意識合わせや進捗確認を行なう。また、2, 3 か月に 1 回程度の頻度で、いずれかの WG 参加者が幹事となって WG 参加者が一堂に会し、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やチャット等の Web アプリケーションが活用され、各種情報

の蓄積、共有および閲覧のために必要なメーリングリストアーカイブ等の Web ページが、W3C の Web サイト上で公開されている。

#### 4.2.3 チーム・コンタクト (Team Contact)

チーム・コンタクトは、W3C チームから 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 による円滑な仕様策定活動を支援する。チーム・コンタクト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・管理
  - 活動報告書 (Activity Statement) [5]
  - WG 憲章 (Charter) [6]
  - 一般公開用メーリングリスト [7]
  - WG 内部用メーリングリスト (W3C 会員限定)
  - 一般公開用 Web ページ [8]
  - WG 内部用 Web ページ (W3C 会員限定)
- W3C 勧告等、各種ドキュメントの公開
- 各種会議への参加 (電話会議, Face-to-Face 会議等)
- その他, WG 運営に関する業務全般

### 5. W3C 標準仕様作成のステップ

W3C で策定する標準仕様は、「W3C Process Document」[9] に記述される通り、以下の段階を経て W3C 勧告として公開される。

- 公開草案初版 (First Public Working Draft; FPWD)
- 最終草案 (Last Call Working Draft; LCWD)
- 勧告候補 (Candidate Recommendation; CR)
- 勧告案 (Proposed Recommendation; PR)
- 勧告 (Recommendation; Rec)

なお、公開草案初版以降、最終草案に至る過程で、各 WG は、一般に一つもしくは複数の草案の更新版を公開していく。

#### 5.1 公開草案初版の公開

当該活動領域における仕様策定を W3C 内外に示し、広く査読を要請するために公開される文書のうち最初のもので、W3C 特許方針に基づく特許請求除外期間が設定される。

W3C は、ディレクタである Tim-Berners Lee の名のもとに、W3C の内外に対して公開草案初版の公開を公示する。

#### 5.2 最終草案の公示

各標準仕様の要件が満たされたとともに、他の仕様との依存性を含む様々な技術的問題が解決され、WG 内で合意が得られた草案で、これ以降は技術的な変更は行われない。

各 WG は、W3C の内外に対して最終草案の公開を公示し、査読を要請する。査読期間として、最低でも公開日より 3 週間以上が設定される。

#### 5.3 勧告候補にもとづく実装要請

最終草案の査読期間満了後、仕様の内容が広く W3C 内外よりの査読を受け、WG の設定した要件を満たすとみなされた場合、当該仕様文書を勧告候補として公開し、W3C の内外に実装を要請する。

W3C は、ディレクタである Tim-Berners Lee の名のもとに、W3C 会員組織に対して実装要請を告示するとともに、公開メーリングリストおよび Web ページを通じて、広く W3C 外に対しても仕様の実装を要請する。

なお、勧告候補の公開にあたって、実装が困難と考えられる機能については、あらかじめ「features at risk」として指定しておくことができる。「features at risk」として指定された機能については、実装事例が得られなかった場合に仕様から削除することが認められる。ただし、その他の機能について実装事例が得られなかった場合、本来の次ステップである勧告案へ進むことは許されず、実装事例が得られるまで待つか、当該機能を

削除した上で再度草案として公開しなおす必要がある。

#### 5.4 勧告案にもとづく査読要請

最終草案の査読、および勧告候補にもとづく実装を経て、十分に内容が安定しているとともに、実装可能性が証明された仕様であり、W3C は、ディレクタである Tim-Berners Lee の名のもとに、W3C 会員組織に対して最終的な承認を要請する。査読期間として、最低でも公開日より 4 週間以上が設定される。

#### 5.5 勧告の公開

W3C 会員組織、およびディレクタである Tim-Berners Lee の最終的な承認を得た仕様 (もしくは指針) であり、他の標準化団体の公開する「標準」に相当する。

W3C は、ディレクタである Tim-Berners Lee の名のもとに、W3C 会員組織に対して勧告の公開を告示するとともに、公開メーリングリストおよび Web ページ (報道発表を含む) を通じて、広く W3C 外に対しても情報を公示する。

### 6. W3C 標準化活動への参加方法

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、各参加組織の本部が置かれている場所にもとづき、W3C の 3 ホスト機関 (MIT/ERCIM/慶應) のいずれかと会員契約を結ぶ必要がある。また、各 WG における具体的な標準化作業に参加するにあたっては、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 会員および Affiliate 会員という 2 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、2009 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。

#### 6.1 Full 会員

総収入が年間 57 億 5 千万円もしくは、それ以上の営利組織。会費年額は 740 万円。

#### 6.2 Affiliate 会員

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 85 万円。

### 7. 特許に関する留意点

W3C 標準仕様の規定にあたって本質的な技術 (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 にもとづき無償で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 サイト上の各種関連資料 [10]~[14] を参照されたい。

### 8. 音声関連技術に関する仕様策定の動向

音声関連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 の音声ブラウザ・ワーキンググループ (Voice Browser Working Group; VBWG) で標準化され対話型音声応答 (IVR) サービスを中心に国際的に利用されている。

以下では、音声ブラウザ・ワーキンググループの活動について簡単に触れた上で、VoiceXML 3.0 における話者認識機能の標準化動向について概説する。なお、活動内容の詳細については W3C の Web サイト上で公開されている関連資料 [6], [8] を

参照されたい。

### 8.1 音声ブラウザ・ワーキンググループの活動

W3Cの音声ブラウザ・ワーキンググループは、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、Web上の様々なサービスにアクセスできるようにするために必要な各種標準仕様（W3C Speech Interface Framework [15]; 図2）の策定を目的としている。

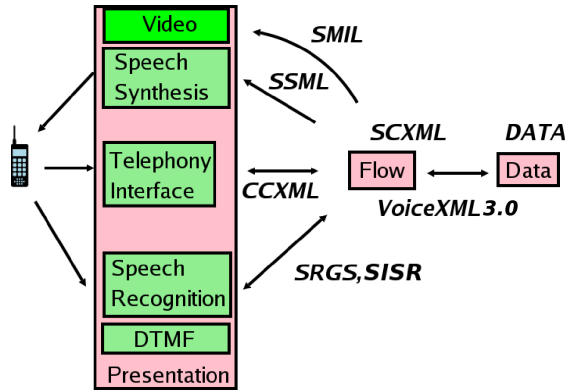


図2 W3C Speech Interface Framework

以下の表1に、音声ブラウザ・ワーキンググループの活動をまとめる。

表1 音声ブラウザ・ワーキンググループ

参加者	77人 (23組織からの参加)
策定対象の仕様	<ul style="list-style-type: none"> <li>• VoiceXML 3.0 [16]</li> <li>• SSML 1.1 [17]</li> <li>• CCXML [18]</li> <li>• SCXML [19]</li> </ul>
会議	<ul style="list-style-type: none"> <li>• 電話会議 (週6回)</li> <li>• F2F 会議 (年数回)</li> </ul>
情報共有	<ul style="list-style-type: none"> <li>• メールングリスト (公開用 [7] / 内部用)</li> <li>• Web ページ (公開用 [8] / 内部用)</li> </ul>

## 8.2 VoiceXML 3.0 仕様の策定状況

### 8.2.1 VoiceXML 3.0 公開草案初版

音声アプリケーション記述のための次世代フレームワークである VoiceXML 3.0 の公開草案初版 [16] が、2008 年 12 月 19 日に公開された。VoiceXML 3.0 は、(VoiceXML2.0/2.1 と異なり) モジュール化された XML ベースのマークアップ言語であり、音声合成、音声認識およびトーン信号 (DTMF) 認識、テレフォニー・サービス、混合主導型対話システム、およびオーディオ、ビデオを含む様々なメディア形式の記録・再生に対応している。

VoiceXML 3.0 公開草案初版では、VoiceXML 3.0 の中核的な機能について、意味論的側面および文法的側面の両面からの記述がなされている。なお、ここで、「意味論的側面」とは、VoiceXML 3.0 の実装者が利用するであろう中核的機能の定義であり、また、「文法的側面」とは、上記の「意味論的側面」にもとづいて規定された各種機能を、アプリケーション作成者が実際に利用するために必要となる、XML 言語としての要素 (element) および属性 (attribute) のことである。

VoiceXML 3.0 においては、全ての機能が、関連する特性をもつモジュールとして分類されており、各モジュールを組み合わせることにより、具体的な言語プロファイルを形成することができる (VoiceXML 2.1 互換プロファイル等)。

また、VoiceXML 3.0 では、「W3C Speech Interface Framework」[15] を用いて、音声アプリケーションをより容易に構築するためのフレームワークとして、モデル-ビュー-コントローラ (MVC) デザイン・パターンの考え方にもとづいて、データ層 (MVC のモデルに相当)、フロー層 (コントローラに相当)、およびプレゼンテーション層 (ビューに相当) という 3 階層を明確に分離した「データ/フロー/プレゼンテーション (DFP)・フレームワーク」[20], [21] を採用している。なお、このフレームワークは、音声関連モダリティを中心に考えた、マルチモーダル対話処理のためのアーキテクチャ (Multimodal Interaction Architecture; MMI Architecture) [22] の一例であるといえる。

### 8.2.2 VoiceXML 3.0 の話者認識機能

残念ながら、上記した VoiceXML 3.0 公開草案初版には含まれていないが、VoiceXML 3.0 の要件 [23] には、VoiceXML 3.0 が満たすべき話者認識機能 (Speaker Identification and Verification; SIV) について記述されている。VoiceXML 3.0 要件は、(これもまたやはり公開草案初版として) 2008 年 8 月 8 日に公開されたが、その第 2.3 節「Speaker Verification」[24] に、「必ず実装しなければならない機能 (must have)」として以下の項目 (図 3) が記述されている。

2.3.1	The markup language MUST provide the ability to verify a speaker's identity through a dialog containing both acoustic verification and knowledge verification.
2.3.1.1	VoiceXML 3.0 MUST support SIV for end-user dialogs
2.3.1.2	SIV features MUST be integrated with VoiceXML 3.0.
2.3.1.3	VoiceXML 3.0 MUST be able to be used without SIV.
2.3.1.4	SIV MUST be able to be used without other input modalities.
2.3.1.5	SIV features MUST be able to operate in multi-factor environments.
2.3.1.6	SIV-specific events MUST be defined.
2.3.1.7	SIV-specific properties MUST be defined.
2.3.1.8	The SIV result MUST be available in the result structure used by the host environment (e.g. VoiceXML 3.0, MMI).
2.3.1.8.1	VoiceXML 3.0 SIV result MUST be representable in EMMA.
2.3.1.9	SIV syntax SHOULD adhere to the W3C guidelines for security handling.
2.3.1.10	(欠番)
2.3.1.11	SIV features MUST support enrollment.
2.3.1.12	SIV features MUST support verification.
2.3.1.13	SIV features MUST support identification.
2.3.1.14	SIV features SHOULD support supervised adaptation.
2.3.1.15	SIV features MUST support concurrent SIV processing.
2.3.1.15.1	SIV features SHOULD support other concurrent audio processing.
2.3.1.16	SIV features MUST be able to accept text from the application for presentation to the user.
2.3.1.16.1	SIV SHOULD be architecturally agnostic

図3 VoiceXML 3.0 話者識別認証機能の要件

### 8.2.3 話者識別認証機能に関するワークショップ

VoiceXML 3.0 要件に記述された上記の項目を具体的に仕様化していくにあたって、音声ブラウザ・ワーキンググループでは、様々な環境、プラットフォーム下における各種音声アプリケーションの開発に必要とされる話者識別認証機能を明確化するために、2009 年 3 月 5-6 日、アメリカのカリフォルニア

州メンロ・パークにて、「VoiceXML 3.0 の話者識別認証機能に関するワークショップ (Workshop on Speaker biometrics and VoiceXML 3.0)」(以下、本ワークショップ) [25] を開催する。本ワークショップの目的は、VoiceXML に必要とされる話者識別認証 (SIV) 機能標準化の方向性を明確化し優先度を整理すること、そしてそれを通して、現在およびこれからの音声アプリケーション市場にとって VoiceXML 3.0 の持つ SIV 機能をより有益なものとするところにある。そのために、本ワークショップでは、主として以下の三点が考慮される。

- VoiceXML 3.0 の SIV 機能が備えるべき要件の明確化
- VoiceXML 3.0 に含めるべき SIV 関連標準の明確化
- 既存の、もしくは現在策定進行中の関連標準との統合

また、本ワークショップの議題ページ [26] にある通りの項目 (図 4) が議題として議論される予定である。なお、第 17 回バイオメトリックシステムセキュリティ研究会が開催される 3 月 23 日に本ワークショップの議事録が公開される予定なので、会議の結論 [27] および議事詳細 [28] については、研究会当日にご報告したい。

#### 8.2.4 今後の展望

音声ブラウザ・ワーキンググループの活動憲章 [6] は 2009 年 1 月末で一旦満了しており、現在、次期の活動憲章を採択する手続きのため、2009 年 4 月末まで、暫定的に活動が延長されている。次期憲章期間においては、引き続き SSML 1.1 および Voice Browser Call Control (CCXML) の W3C 勧告化にも取り組むが、主として、VoiceXML 3.0 および State Chart XML (SCXML) の仕様策定に注力する予定である。

第二版の草案は、SIV 機能を含めて 4 月初旬に公開される予定であり、その後も継続して、3 か月程度ごとに草案の更新版が公開される予定である。

#### 8.3 日本の技術と国際標準

我が国においても、様々な組織において音声関連技術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) [30] や、情報処理学会の試行標準専門委員会 WG4 (音声言語処理インタフェース; 以下、単に WG4) [31] など、W3C の国際標準化と協調を取りながら日本国内における標準化作業を進めている例もある。

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 W3C ワークショップ」[32] 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 [33] 以降、招聘専門家として、継続的に音声ブラウザ・ワーキンググループの標準化活動に参加している。

また、WG4 では、独自にマルチモーダル・システム開発のためのユースケース抽出、要求仕様洗い出し、およびアーキテクチャ提案等、マルチモーダル対話処理記述言語の策定作業を進めているが、その成果は、W3C 会員である京都工芸繊維大学によって W3C の標準化活動ヘフィードバックされている。

## 9. むすび

W3C の組織構成について概説した上で、音声ブラウザ・ワーキンググループで取り組んでいる、VoiceXML 3.0 策定の動向、特にその話者認識機能について説明した。

音声関連技術は、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、各開発ベンダおよびサービスベンダごとに仕様異なる部分が大きく、「全て

#### SIV アプリケーションの概説

- SIV 機能に関連する人々は誰か
- 既存のシステムでは、どのように扱われているか
- どのようなベンダ・ユーザが利用しているか
- IVR サービス以外の用途も考慮すべきか

#### VoiceXML 3.0 の SIV 機能

- VoiceXML 2.0/2.1 との相違点
- 今までのメカニズムにおける制約事項
- VoiceXML 3.0 概要
- VoiceXML 3.0 の SIV 機能に関する VBWG 内での議論

#### SIV の具体的なユースケース

- 当該アプリケーションは、どのような SIV 機能を必要とするか
- SIV 機能を当該アプリケーションに適用する際の問題は何か
- テキスト依存およびテキスト非依存、いずれの技術を利用するか
- 複数の識別エンジンや入出力モダリティを利用する場合の問題

#### SIV 機能の利用者 (複数のサービスベンダを想定)

- 複数のベンダが共存する際に必要となるシステム設計方針
- 複数の不確実な認識結果を統合するためのメカニズム
- VoiceXML ベンダおよびアプリケーション開発者の責任範囲

#### SIV 機能で用いるべきオーディオフォーマット

- SIV 機能で利用されるべきオーディオフォーマット決定の指針
- Web 上で利用するための互換性

#### SIV 機能と MRCP V2 との関係

- MRCP V2 概説
- MRCP V2 と SOAP の相違点
- IVR サービスと SIV 技術を統合する際の利便性
- ストア型利用とストリーミング型利用

#### SIV 機能が必要とするアーキテクチャ

- 各講演者の提案するアーキテクチャとその機能
- ストア型利用とストリーミング型利用のサポート
- 利用者の受け入れ/拒否の決定を、いつどこで行なうか
- 音声データベースを誰が管理するか
- 音声登録は誰が管理するか
- VoiceXML 処理系とアプリケーションサーバとの役割分担
- HMM, ニューラルネット等、特別な処理が必要か
- 連続的な認証に対応するか
- 認証対象外者リストに対応するか
- 具体的な API およびやりとりされるパラメータ

#### マルチモーダル処理のためのデータ形式

- マルチモーダル処理のために必要なデータ形式
- EMMA [?] は有用か
- EMMA 以外に有用なデータ形式があるか
- VoiceXML 3.0 の SIV 機能は、EMMA をサポートすべきか
- 複数のモダリティを統合/分割する処理の扱い

#### 他のバイオメトリクス関連標準との関係 1: Common Biometric Exchange Format Framework (CBEFF) および International Committee for Information Technology Standards (INCITS) 456

- CBEFF とはなにか
- INCITS 456 とはなにか
- CBEFF および INCITS 456 をサポートすべきか

#### 他のバイオメトリクス関連標準との関係 2: Biometric Identity Assurance Services (BIAS) プロジェクトおよび BioAPI

- BIAS プロジェクトとはなにか
- BioAPI とはなにか
- BIAS および BioAPI に適合するか
- BioAPI を取り込むべきか、あるいは仕様策定の参考にすべきか

#### セキュリティおよびアイデンティティの管理

- 個人情報、セキュリティ、およびリスク管理をどう行なうか

図 4 話者識別認証機能に関するワークショップ議題

の人が全ての情報にアクセスできるようにする (One Web)」  
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。

我が国の優れた研究開発成果を国際的に活用し、さらなる技術発展を促進するためにも、是非、日本の技術者および研究者の皆様に、W3C 標準化活動へ積極にご参加をいただき、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。

お問い合わせ先(ご質問、ご参加希望等の受付)

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: 一色, 石倉, 芦村

## 文 献

- [1] <http://www.w3.org/>
- [2] <http://www.w3.org/Consortium/Offices/staff.html>
- [3] <http://www.w3.org/Consortium/>
- [4] <http://www.w3.org/2002/03/new-to-w3c>
- [5] <http://www.w3.org/Voice/Activity.html>
- [6] <http://www.w3.org/2006/12/voice-charter.html>
- [7] <http://lists.w3.org/Archives/Public/www-voice/>
- [8] <http://www.w3.org/Voice/>
- [9] <http://www.w3.org/2005/10/Process-20051014/tr.html#rec-advance>
- [10] <http://www.w3.org/2005/10/Process-20051014/>
- [11] <http://www.w3.org/2005/07/pubrules/>
- [12] <http://www.w3.org/Consortium/Patent-Policy-20040205/>
- [13] <http://www.w3.org/2003/12/22-pp-faq.html>
- [14] <http://www.w3.org/2004/01/pp-impl/>
- [15] <http://www.w3.org/Voice/#q8>
- [16] <http://www.w3.org/TR/2008/WD-voicexml30-20081219/>
- [17] <http://www.w3.org/TR/speech-synthesis11/>
- [18] <http://www.w3.org/TR/ccxml/>
- [19] <http://www.w3.org/TR/scxml/>
- [20] <http://www.w3.org/Voice/2006/DFP>
- [21] <http://www.w3.org/TR/2008/WD-voicexml30-20081219/#DFP>
- [22] <http://www.w3.org/TR/mmi-arch/>
- [23] <http://www.w3.org/TR/2008/WD-vxml30reqs-20080808/>
- [24] <http://www.w3.org/TR/2008/WD-vxml30reqs-20080808/#funct-siv>
- [25] <http://www.w3.org/2008/08/siv/cfp.html>
- [26] <http://www.w3.org/2008/08/siv/agenda.html>
- [27] <http://www.w3.org/2008/08/siv/summary.html>
- [28] <http://www.w3.org/2008/08/siv/minutes.html>
- [29] <http://www.w3.org/TR/emma/>
- [30] <http://home.jeita.or.jp/is/committee/tech-std/std/com03.html>
- [31] <http://www.itscj.ipsj.or.jp/senmon/05sen/ipsjts.html>
- [32] <http://www.w3.org/2005/08/SSML/ssml-workshop-cfp.html>
- [33] <http://www.w3.org/2005/08/SSML/ssml-workshop-agenda.html>