

Enhanced Web Speech API: ブラウザ非依存なエンジン選択・拡張が可能な HTML5 用音声 ウェブアプリケーション開発用プラットフォームの提案と実装*

小本勇貴[†]中野鐵兵[†]小林哲則[†]早稲田大学[†]

1 はじめに

HTML5 に対応したブラウザの普及が進み、音声認識・音声合成機能を備えたウェブアプリケーションの開発が困難でなくなってきた。HTML5 で記述するウェブアプリケーションで音声認識・音声合成機能を使用するためには、W3C によって策定されている Web Speech API [1] を使用する方法が一般的である。しかしながら、Web Speech API を用いた場合、音声認識・音声合成エンジンの使用がブラウザ依存となり自由に制御することが困難である。音声認識・音声合成エンジンベンダーが提供するプロプライエタリなライブラリを用いる方法もあるが、開発コストの増大という問題が発生する。

そこで本研究ではこれらの問題を Web DB を利用したエンジン情報の共有とその情報によるクライアントライブラリの適応により解決し、ブラウザに非依存でかつエンジンの選択・拡張が可能な HTML5 用ウェブアプリケーション開発のためのプラットフォームを提案する。

2 Web Speech API を用いたウェブアプリケーション開発とその問題

Web Speech API は W3C の Speech API Community Group によって策定されている JavaScript の API であり、一般的な音声アプリケーションのユースケースで想定されるインタフェースが定義されている。Web Speech API を使用して記述されたウェブアプリケーションは、WebSpeech API をサポートするブラウザ上であれば特別なソフトウェアをインストールすることなく、音声認識・音声合成機能が利用可能となる。しかしながら、Web Speech API を使用して音声アプリケーションを開発する場合、音声認識・音声合成エンジンの選択をウェブアプリケーション開発者が行うことができないため、下記の問題が発生する。

- 提供される音声認識・音声合成機能の品質が一定でない（品質保証問題）
- 特殊用途に構成された音声認識・音声合成システムなど、独自エンジンの導入が困難（独自エンジン問題）。

品質保証問題は、実際に提供される音声認識・音声合成サービスがブラウザ及び OS に依存していることによって発生する。Web Speech API では、例えば合成音の種類やパラメータをウェブアプリケーション側で自由に変更することが可能である。しかしながら、利用者が使用しているブラウザの違いにより、実際に利用者に提供される音声の品質を均一にすることは出来ない。あるブラウザの利用者にとって不満の残る品質の音声サービスであったとしても、それを改善することは難しい。品質保証問題を解決するためには、ブラウザ・OS 非依存で導入可能な独自の音声認識・音声合成サービスを用いたウェブアプリケーションを開発する必要がある。しかしながら、独自エンジン問題があるためにその導入は必ずしも簡単ではない。

独自エンジン問題は、Web Speech API を使用して標準的に使用される音声認識・音声合成エンジンを、開発者が自由に選択することや拡張することが出来ないことによって発生する。ブラウザの開発元によって決められているエンジンを使用しない場合、ウェブアプリケーション開発者はプロプライエタリなライブラリを利用する必要がある [2]。これらのライブラリのすべてが必ずしも互換性を持つとは限らないため、開発コストの増大は避けられない。

また独自エンジンの問題には、ブラウザに組み込まれていない高性能な独自エンジンの使用が妨げられるという問題を含んでいる。例えば、特殊用途目的で非常に高い性能を発揮するエンジンが開発されたとしても、多くの音声アプリケーション開発者に認知してもらい、使用してもらうことは困難である。エンジン開発者にとっては、プロプライエタリなライブラリの開発や、サーバサイドプログラムの実装も容易ではない。利用可能な環境を広げるためには、Media Resource Control Protocol

* Enhanced Web Speech API: Proposal and implementation of Browser-independent, engine selection and expansion possible voice web application development platform for HTML5

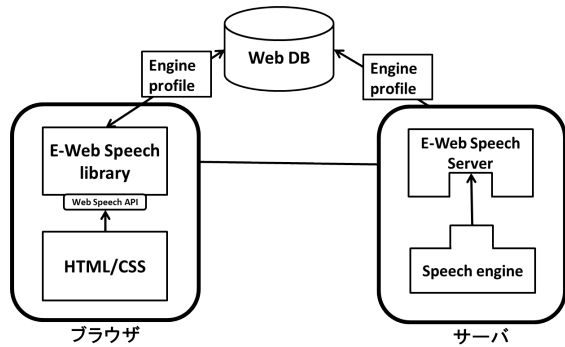


図1 本システムの構成図

(MRCP) などの標準プロトコルに対応してサーバを構築する方法も考えられるが、エンジン開発者がすべてを用意するのは現実的ではない。

3 提案手法

提案手法は、ブラウザで動作する JavaScript ライブラリとサーバサイドで動作するフレームワーク及びそれらの接続を管理する Web DB によって構成される。(図1)

3.1 Web Speech API をもとにしたエンジンの選択・拡張が可能なライブラリの提供

ブラウザにロードすることで環境に依存せず利用可能なライブラリを提供する。Web Speech API によって実装することで標準的なインターフェースでの利用を可能とする。

後述する Web DB から Engine profile を取得する。Engine profile は接続先サーバの情報やエンジン側で取り扱えるデータ形式などの情報を持ち、それを基に使用するエンジンに適した形へライブラリの再構成を行うことでエンジンの選択を可能とし、さらにエンジンごとの独自の拡張パラメータなどへの対応も可能とする。

3.2 独自エンジンをプラグインとして利用できるサーバサイドフレームワークの提供

ライブラリからの接続を可能とするサーバ機能や音声認識・合成機能を司るプラグインを登録する環境を提供する。登録されたプラグインから Engine profile を取得し、Web DB へ自動的にアップロードする。また、実際にライブラリから送信されたデータをエンジンが取り扱える形式への変換を行い、またエンジンが出力したデータをライブラリが取り扱える形式への変換を行う。

3.3 Web DB を利用したエンジン情報の共有

エンジンの情報についてまとめたものを engine profile と定義し、Web DB を介してリアルタイムでの共有を行う。Web DB では現在稼働中の利用可能なエンジンの一覧を提供し、Engine profile の配信を行う。

4 音声合成機能の実装

表1に示すブラウザ・OS に対応する実装を行った。

表1 Enhanced Web Speech API 音声合成の対応状況

	IE	Chrome	Firefox	Safari
Windows 8.1	○	○	○	—
OS X 10.9	—	○	○	○
iOS 8.1.2	—	×	—	×
Android 4.4.2	—	×	○	—

4.1 クライアントライブラリ

JavaScript ライブラリとして実装を行った。使用するエンジンの EngineURI を指定することにより Web DB から Engine profile を取得する。その情報から接続するサーバやデータの入出力形式などを決定する。そのため、使用するエンジンの変更に伴いウェブアプリケーションの実装を変更する必要はない。

4.2 サーバフレームワーク

サーバフレームワークは、Node.js により実装を行った。Node.js の C++ addon を利用して、エンジン側の変更を最小限のまま、プラグインとできるように実装した。エンジン開発者がプラグインを登録すると、サーバフレームワークは自動的に Engine profile を Web DB へアップロードする。これにより、エンジン開発者への負担を最小限に本枠組での利用を可能とする。

4.3 Engine profile

Engine profile は、JSON 形式でサーバの接続先やエンジンの種類、入出力の形式、パラメータなどエンジンの情報をもつ。Web DB として Firebase^{*1}を利用し、ライブラリとフレームワーク間でのリアルタイムでの共有を可能としている。また、ウェブアプリケーション開発者が利用可能な Engine profile の情報が閲覧できるウェブアプリケーションを提供する。このアプリケーションでは Web DB に登録されている利用可能なエンジンを表示し、エンジン選択に必要な情報を確認でき、またエンジン開発者は多くのウェブアプリケーション開発者へ自らのエンジンを認知してもらうことができる。

5 まとめと今後の予定

Web Speech API の問題点として品質保証問題と独自エンジン問題を挙げ、それらを解決する Enhanced Web Speech API を提案し実装を行った。今後は動作しない環境への対応や音声認識についての実装を行っていく。

参考文献

- [1] Glen Shires, Hans Wennborg, Web Speech API Specification, <https://dvcs.w3.org/hg/speech-api/raw/tip/speechapi.html>, 19 October 2012.
- [2] 槇優一, 鎌土記良, 藤村滋, 青野裕司, 中山文二, 阪内澄字, 山田智広. 音声認識機能を有する Web アプリケーションの実装と評価. 情報処理学会研究報告. マルチメディア通信と分散処理研究会報告, 2014(6), 1-6.

*1 Firebase:<https://www.firebase.com>