

世界に先駆けて様々なHTML5ブラウザ上で高精度な音声認識を利用可能とする技術の開発

ならびに、クラウド型音声認識サービス“SpeechRec for Browser”の販売開始について

～ソフトウェアのインストールなしで、様々なブラウザ上で音声認識が可能なWebアプリケーションの開発を支援～

日本電信電話株式会社(以下NTT、本社:東京都千代田区、代表取締役社長:鶴浦 博夫)は、世界に先駆けて様々なHTML5(※注1)ブラウザ上で事前のソフトウェアのインストールが不要で高精度な音声認識を利用可能とする技術を開発しました。

また、エヌ・ティ・ティ アイティ株式会社(以下:NTT アイティ、本社:神奈川県横浜市中区、代表取締役社長:長谷 雅彦)は、本技術をNTTより取得し、「HTML5ブラウザによる音声認識サービス“SpeechRec for Browser”」として、2014年10月より、国内にて販売を開始する予定です。

1. 様々なHTML5ブラウザ上で高精度な音声認識を利用可能とする技術の開発について

● 開発の背景

これまでWebブラウザで音声認識を利用しようとする、事前にソフトウェアのインストールが必要であり、使い方やブラウザも限られるという問題がありました。また、アプリケーションを開発する側から見た場合、ブラウザ毎のソフトウェアの開発が必要であるなど、複数のブラウザに対応するためには、個別の対応が必要でした。

● 技術のポイント

本技術では、W3C(※注2)で仕様策定が進められているJavaScript APIを組み合わせることで、端末種別、OSの違いによらず、HTML5ブラウザから音声認識を使ったサービスの利用を可能としました。これにより、ウェブ閲覧者は、特別な準備をすることなしに音声認識を使ったWebアプリケーションの利用が可能となります。

さらに、様々な利用環境での使用に対応するため、周囲騒音に頑健な音声区間検出機能(※注3)も兼ね備えており、周囲環境によらず、高精度な音声認識が可能となりました。

● 今後の展開

今後、Webブラウザを搭載した端末が拡大していくことで、モバイル端末に加え、自動車やデジタルサイネージなど、さまざまな利用シーンでの音声の利用も期待されます。従来のWebアプリケーションに、音声認識を始めとする音声入力のインタフェースを容易に追加可能になることで、音声入力をトリガとした様々なサービスの発展が期待されます。

※詳細は【別紙1】参照

2. 「HTML5 ブラウザによる音声認識サービス “SpeechRec for Browser”」の販売開始について

(1) サービスの概要

NTT アイティ はこれまで、多様な環境下で利用可能な高精度日本語音声認識ソリューション”SpeechRec”を展開してきました。“SpeechRec for Browser”は、高機能な音声区間検出機能を備えた音声認識機能をHTML5ブラウザだけで利用することができるため、既存のWebページへの音声認識機能の追加が簡単に行えます。また、他の利用企業との共有型サービスとは異なり、専用の認識辞書を利用することで製品の独特な固有名詞なども精度良く認識することができます。

NTT アイティは、本サービスをいち早く提供することにより、Web ページ作成ベンダと連携した企業や自治体の Web ページの価値向上や、SI ベンダと連携しモバイル環境でのタブレット端末を活用した報告システムの利用率向上など、音声認識の利用の幅を広げシェアの拡大を図っていきます。

さらに、WebRTC(※注4)を活用したアプリケーションとの連携による新たな通信サービスや、対話や翻訳などのサービスと連携したサインージサービスの創出など、音声認識を活用した新たなサービスの創出に貢献します。

(2) 販売開始 : 2014 年 10 月 予定

(3) 販売価格 : NTTアイティにお問い合わせください。

(4) 販売エリア : 日本国内

※詳細は【別紙2】参照

※注1 HTML5: 2014年冬に正式な仕様として勧告が行われる予定のHTMLの次期仕様を指します。より広い意味で、従来では実現できなかったWebアプリケーションを実現可能にするためのWeb関連技術を含めてHTML5と呼ばれる場合もあります。

※注2 W3C(World Wide Web Consortium): Web技術の標準化と推進を目的とした、会員制の国際的な産学官共同コンソーシアムのことです。

※注3 音声区間検出機能: 雑音の混じった音声から、実際に人がしゃべった音声区間のみを抽出する技術

※注4 WebRTC(Web Real-Time Communication): ブラウザ間でリアルタイム通信を実現するための技術です。W3Cなどで標準化に向けた取組みが進められています。

様々なHTML5ブラウザ上で高精度な音声認識を利用可能とする技術開発について

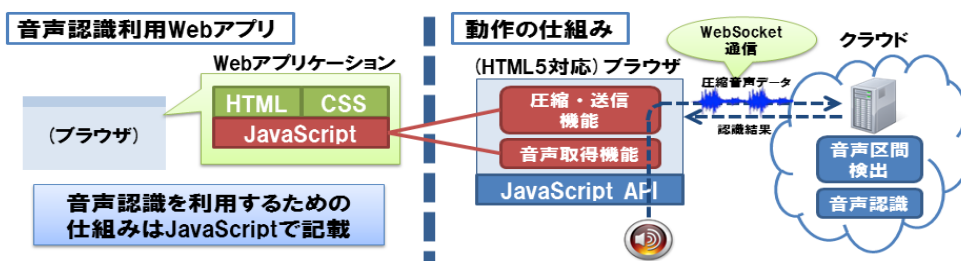
■開発の背景

ブラウザに音声操作のための音声認識機能を導入するためには、専用のアプリケーションを新規に構築する、もしくは予め特定の音声認識エンジンが導入されている特定のブラウザ上のみで動作するWebアプリケーションを構築する必要があり、様々なサービスで音声認識を実現したい事業者にとって大きな障壁となっていました。

■技術のポイント

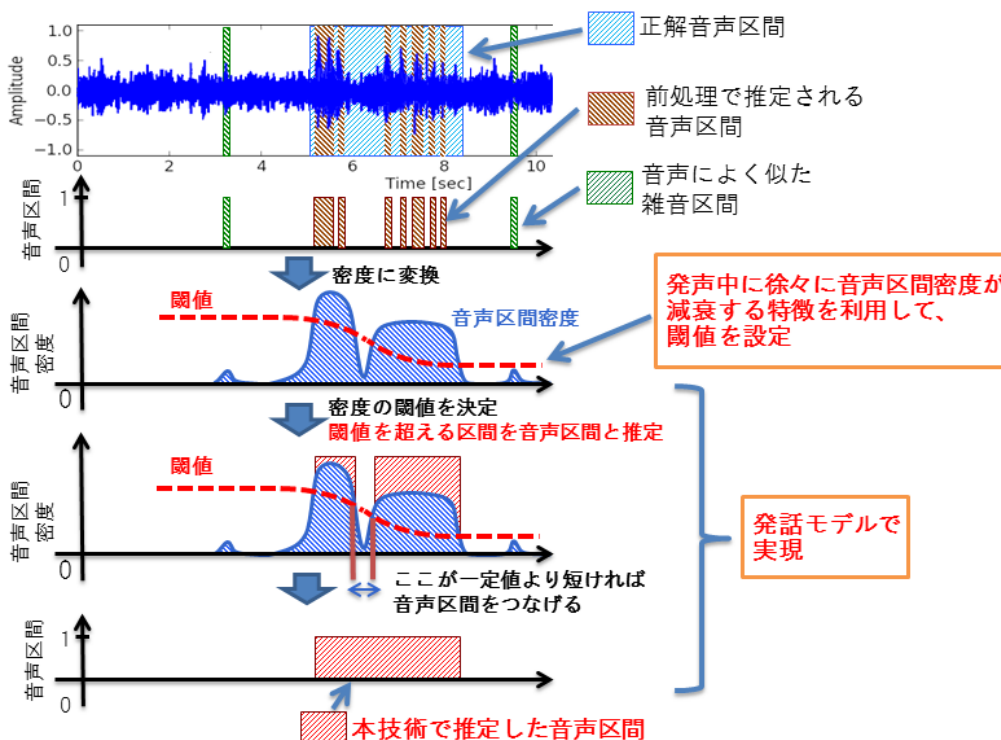
① 幅広いブラウザでの動作を可能とする JavaScript API の選択

各 JavaScript API の標準化に向けた取組み状況を調査した上で、幅広いブラウザで実装が進んでいるものを組み合わせて利用することで、端末のマイクデバイスからの音声取得を実現しています。また、ストリーミング形式による音声認識を実現するために、WebSocket を利用したリアルタイム通信により利用者の体感品質向上に努めています。



② 高精度な音声区間検出

ユーザの発話の経過に応じて特徴的に変化する音声区間密度に注目し、これを独自の発話モデルにより処理することにより、音声区間検出の精度を最大で15%向上、音声と特徴の似た雑音に対しても高い精度で音声区間を検出することに成功し、発話終了の微妙なタイミングを正確に検出することが可能になりました。



「HTML5ブラウザによる音声認識サービス “SpeechRec for Browser”」について

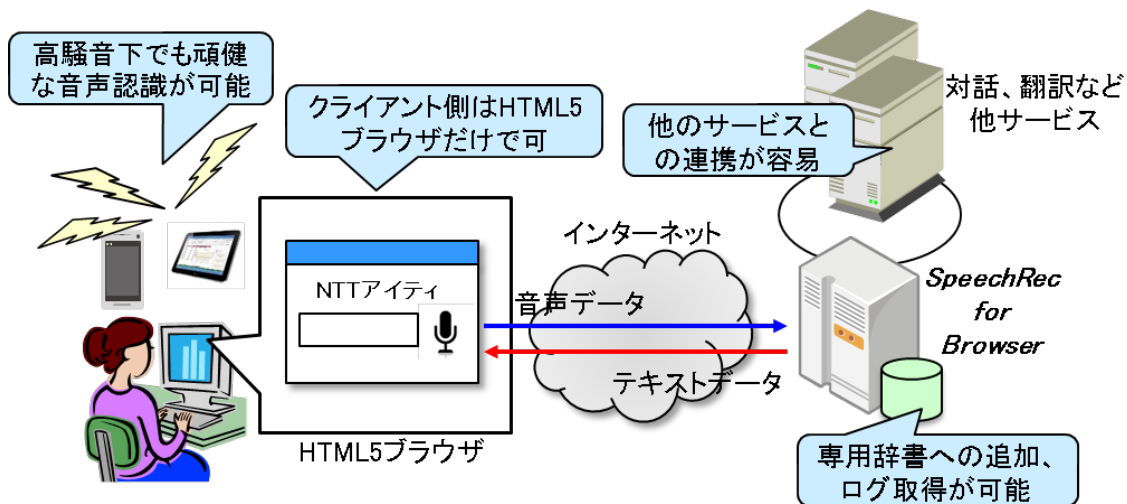
■サービスの概要

本サービスでは、NTTメディアインテリジェンス研究所が開発した高性能な音声認識エンジンであるVoiceRexを採用しています。今回、NTTサービスエボリューション研究所のHTML5コンテンツ上での音声認識実装技術を採用し、新たな利用形態として、HTML5ブラウザで音声認識を利用できるようにするサービスを提供致します。これにより音声認識機能を活用するアプリケーションを簡単に構築できます。

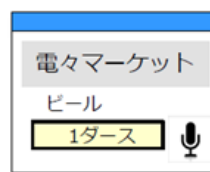
他の利用企業との共有型のサービスと異なり、認識辞書や言語モデルなどのカスタマイズにも対応できます。さらに、ユーザがどのような用語を検索しているかなど、企業にとってマーケティングなどの面で重要なアクセスログの取得・分析も可能です。

NTTが新たに開発した高機能な音声区間検出機能を備えており、高騒音下でも頑健な音声認識が可能となります。これにより、自動車内、駅構内や工場などこれまで音声認識の利用が困難だった環境でも使えるアプリケーションが開発可能となります。

音声データの処理をクラウド側に集約しているため、他の様々なサービスとの連携が容易になります。HTML5ブラウザを用いたサービスとして、WebRTCを利用したWeb会議などが出てきていますが、リアルタイムで議事メモを作成するなどのアプリケーションの構築がより簡単になることが期待されます。また、対話や翻訳などのサービスと連携した新たなサインページサービスの創出も期待されます。



検索語入力に！



通販、予約サイトに！



ブログ、SNSに！

■適用例

- ・検索キーワード入力
- ・通販サイトでの商品名や個数などの項目入力
- ・ブログやSNSなどでのテキスト入力
- ・タブレット端末を用いた業務報告、データエントリシステムなど
- ・WebRTCアプリとの連携
- ・対話型サインページサービス