

日本人風英語発音へのリアルタイム変換システム

Real-time conversion System to English pronunciation like the Japanese

基盤ソフトウェア学講座 0312011158 山本忠弘

指導教員：澤本潤 杉野栄二 瀬川典久

1. はじめに

2020 年の東京オリンピックに向けて、日本人のグローバル化を目指していくために、2013 年に「英語教育改革案」が文部科学省から発表され、2014 年に新たなカリキュラムが施行された。

その改革案の中に「コミュニケーション能力の素地を養う」とあるが、長らく日本人は文法理解や読解・英作文といった読み書きに比べ、会話を苦手に行っていると言われている。英語でコミュニケーションをとるにあたって問題となるのが、聞く、喋る際の発音の問題である。一般に多くの人は、義務教育に加え、高校や大学入試を通じた英語教育で語彙は十分にある。しかし語彙だけでは相手の話す内容が理解できないし、話すこともできないと考える。

そこで西田氏は、日本人音声認識と TTS を組み合わせてネイティブの英語の発音を日本人風の発音へ変換するシステムを提案したが、リアルタイムな変換を用いていなかったため、コミュニケーションの場で使用する際に相手の発話が終わるまで変換が開始されず、発言のたびにボタンを押して認識を開始しないといけないため、実際の使用は難しいと考えた。

そこで本研究は、先行研究にリアルタイムな変換の機能を加え、コミュニケーションの場での使用を目的とした実験用プログラムを使用し、ネイティブとの英語でのコミュニケーションの際、日本人に聞き取りやすくなることによる英語学習の意欲の促進を目指す。

2. システムの手法

本研究では音声認識プログラムとテキスト読み上げプログラム(text-to-speech;TTS)を用いることで日本人風の発音への変換を行う 2 つの方法(図 1, 2)を実装した。どちらの方法も自動的に同時翻訳者の

ように話している長い発話を短い文節に分割して変換したあと、音声再生される。音声認識には Google 社の WebspeechAPI, TTS には, Googlechrome に備わっているものをそれぞれ利用した。

2.1. 変換方法

一つ目の変換方法は、ネイティブの話者の音声を認識して得られた英語テキストを日本語 TTS に読み上げさせることで日本人風の発音に変換する方法である(図 1)。

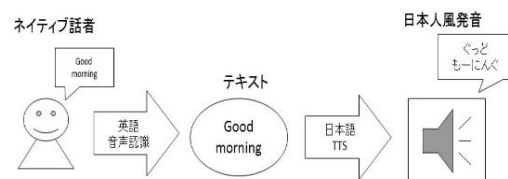


図 1. 変換方法 1: 英語音声認識の出力するテキストを、日本語 TTS に読み上げさせる方法

この方法は実装が非常に容易であるが、変換方法 1 の日本語 TTS を利用することで日本人が各単語をローマ字読みしているように変換することができる。しかしこの方法は英語を日本語 TTS で読み上げる際に、“I am” を “あいえーえむ” と読むなど、間違っただけの変換をしてしまう場合がある。

そこで、二つ目の方法として、音声認識プログラムの出力する文字列を文字列置換(表 1)によって変換し、変換した文字列を TTS に指定して読み上げさせる方法を用意した(図 2)。

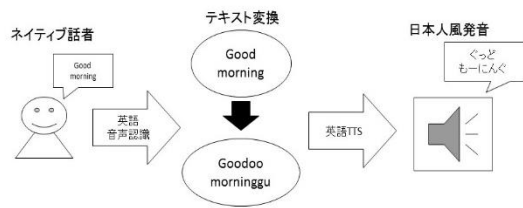


図 2. 変換方法 2: 語音声認識の出力するテキストを変換し、英語 TTS に読み上げさせる方法

文字列の置換例		単語末尾の置換例	
変換前	変換後	変換前	変換後
zy (lazy)	zii (lazii)	d (good)	ddo (goodoo)
ph (phone)	f (fon)	f (roof)	fuu (roofuu)
the	za	g (dog)	ggü (doggu)
		k (book)	kku (bookku)
		le (needle)	ruu (needruu)

※()内は単語例を示す

表 1. 文字列置換のルール表

この方法では、変換方法 1 ほど忠実に日本人風の発音を再現できていないが、日本人が英語をしゃべる際の特徴として挙げられる語尾の母音を強調する点を忠実に再現出来る。

この方法を用いることで、変換方法 1 の時に生じた自分が望んでいない発音への変換をすることが少なくすることができる。

2.2. システムの構成

本システムを利用する際の OS は Android を推奨とし、図 1 に示すように Googlechrome でのインターネットブラウザで利用することを目的とする。

本システムの利用方法はページ上部のボタンを押すと音声認識がはじまり、左のボックスに認識した文字を表示し、右のボックスに変換後の文字列が表示され、TTS によって読み上げられるといったものである。機能としては、TTS の声の高さ、早さ大きさをスライドバーによりそれぞれ調節できるようにした。

3. 評価と考察

日本人の成人男性と成人女性それぞれ 2 名に使用してもらい、このシステムを使うことで実際にネイ

ティブの発話の意味を理解できるかどうかの評価を行った。

10 種類のリスニング音源を用いて、システムを用いた際の意味の正答率、用いなかった場合の意味の正答率は以下表 2 の通りであった。

システムを用いていない場合の正答率

被験者	A	B	C	D
正答率 (%)	60	40	50	40

システムを用いた場合の正答率

被験者	A	B	C	D
正答率 (%)	100	90	100	80

表 2. 実験結果

表 2 の結果から、システムを用いた場合の正答率が、用いなかった場合の正答率に比べて高くなっていることが明らかになった。

4. おわりに

本研究では音声認識と TTS を用いる事で、ネイティブとの英語でのコミュニケーションの際、発話を聞き取れやすくなる日本人風英語発音へのリアルタイム変換システムを提案し、実験用システムの実装と評価を行った。評価の際、大体的意味はわかるが日本人の発音に聞こえない部分もあるとの意見があり、今後はこの問題について解決していく必要がある。

5. 参考文献

- 1) 西田 健志 (2013) 「ネイティブ英語発話の日本人風の発音への変換による国際的な意識の促進」『No. 73 第 21 回インタラクティブシステムとソフトウェアに関するワークショップ(WISS2013) 論文集』pp. 1-6
- 2) 清水宏晃, Graham Neubig, Sakriani Sakti, 戸田智基, 中村哲 (2013) 「同時通訳データを利用した自動同時通訳システムの構築」