

## 科学研究費助成事業 研究成果報告書

平成 26 年 5 月 2 日現在

機関番号：12601

研究種目：基盤研究(C)

研究期間：2011～2013

課題番号：23500170

研究課題名(和文)マイノリティ言語向け合成音声システム開発のための効率的音声適応手法の構築

研究課題名(英文)Study on Adaptation of Speech Synthesis for Minority Languages

研究代表者

巖淵 守 (IWABUCHI, Mamoru)

東京大学・先端科学技術研究センター・准教授

研究者番号：80335710

交付決定額(研究期間全体)：(直接経費) 4,100,000円、(間接経費) 1,230,000円

研究成果の概要(和文)：本研究では、既存の合成音声ソフトの他の言語への適応手法の構築を目指した。最頻出単語に対する音素マッピング手法をベースに、母語話者が代替綴りの変換規則を確認・修正することで音変換のルールを作成した。市販のヒンディー語の合成音声ソフトをベースに評価用の実験システムを作成したところ、確認用の日本語では正しい聞き取りが3割にも満たなかった一方、言語的に近いネパール語では、6～8割の聞き取りが可能であった。以上から、本研究で提案された方法は、言語的・音声学的に近い言語ペア間ではその有効性が期待されるものの、実際のマイノリティ言語に応用するためには、各言語に依存した問題が生じることが明らかとなった。

研究成果の概要(英文)：This study pursued an economical method to create a text-to-speech (TTS) software for a new language using an existing voice synthesizer. The new sound conversion rules were created with manually edited letter-to-sound rules by native speakers for most frequently used words. In order to evaluate the method, Japanese and Nepali TTS systems were created using Hindi TTS that was commercially available. The created Nepali TTS produced well recognized speech with 60-80 percent accuracy. Meanwhile the created Japanese TTS produced speech that was recognized with less than 30 percent accuracy. It was concluded that the method of this study could be usable for a language pair where the original and target minority language are phonetically and linguistically close. However there would still be challenges about language-specific issues, such as prosodic modeling, unknown words pronunciation, and grammatical analysis.

研究分野：総合領域

科研費の分科・細目：情報学・知能情報学

キーワード：自然言語処理 音声合成

### 1. 研究開始当初の背景

近年の飛躍的な性能向上を果たした合成音声技術は、すでに普及の段階を迎え、そのいくつかはパソコンの標準機能として搭載されるに至っているだけでなく、スマートフォン、ナビゲーションシステム、ロボットなど、様々な機器にその利用が広がってきた。情報のグローバル化に対応するため、合成音声技術をさらに多くの言語に適応させる方法についてもこれまで様々な研究が行われてきており、研究開始時点において、40言語を越える合成音声ソフトが存在した。中でもカーネギーメロン大のSchultzら(2007)は、各言語の最頻出単語を中心に発音ルールを決定していく同大Maskeyら(2004)の統計的手法を用いて、効率的に新たな言語向けの合成音声システムを構築するウェブベースのツール「SPICE」の開発にも成功しており、合成音声ソフトが対応する言語の増加が期待されていた。

しかし、世界には3千から8千もの言語が存在すると言われ、その数の多さに加え、マイノリティ言語の多くは言語的・音声学的情報も乏しく、言語処理と合成音声システムの構築に関する知識・スキルを兼ね備えた人材の確保の難しさからも、上記のツールの利用をもってしても新たなマイノリティ言語対応合成音声ソフトの開発は困難である。そこで本研究では、既存の合成音声ソフトを応用し、それを改良することでマイノリティ言語に対応させる、より簡易で時間的・金銭的コストの大幅な削減を可能とする手法の構築を目指した。

### 2. 研究の目的

本研究では、マイノリティ言語向け合成音声システム開発のため、既存合成音声ソフトウェアを効率的・効果的に適応させる手法の理論的枠組みを構築することをその目的とした。その際、本研究の特徴として、マイノリティ言語に対応するための専用の音素データベースを新たに録音・生成せず、代わりに既存合成音声ソフトウェアの発音音を利用する手法を用いた。そのため、技術的課題として、これまでの手法で行われソフトウェアの内部パラメータとして扱われる言語の音素表現に代わり、既存の合成音声ソフトウェアに対して外部から直接入出力できるテキスト情報で表現する必要があり、その適応手法を明らかにすることを目的とした。

### 3. 研究の方法

本研究では、まずその目標となるマイノリティ言語向けの合成音声ソフトウェアについて、調査を交えた要求分析を行った。

次に上記技術的課題としてあげた、対象と

なるマイノリティ言語の基本音素と既存合成音声ソフトウェアの言語の文字表現をつなぐ Sound-to-Letter (音素列を文字列表現に対応させる)の音声変換ルールの候補の探索方法として以下の手順を検討し、その有効性を確かめた。

(1) 既存合成音声ソフトウェアが持つ基本音素と連結音の情報から、この言語における音素と綴りの「対応表」を作成する。

(2) (1)で作成された「対応表」を用いて、新たな対象となるマイノリティ言語のアルファベットに対応する音を、ネイティブスピーカーの研究協力者に手動で割り当ててもらう。この段階で、対象マイノリティ言語の1文字ずつの音が、既存合成音声ソフトウェアのテキスト表現によって表すことのできる音変換の基本「ルール」を得たこととなる。

(3) 新たな対象となるマイノリティ言語のテキスト(標準的な書籍やニュースからの文章等)から最頻出単語リストを作成する。

(4) (3)で作成された最頻出単語リストの上位の語から順に、その発音を(2)で得た「ルール」を基に、システムが予測した候補を提示し、上記ネイティブスピーカーが手動で(1)の「対応表」からの音の選択をもってこれを修正して、正しい発音を割り当てる。

(5) (4)の操作を繰り返す。この中で、それまでに発音を登録された語に含まれる最長の共通スペルの組、ならびに含まれる3文字のスペル(トライグラム)の組の発音の類似性から、統計的に次の単語の予測発音の確からしさを算出することが可能である(この算出方法は、前出のMaskeyらの先行研究によって提案されている)。この算出方法をもとに、次の単語の予測発音候補の数を、確からしさの順に複数提示して、ネイティブスピーカーが選択・修正しやすいようにする。また、順次(2)で議論された音変換の「ルール」の精度を向上させていく。

(6) 発音の予測がある一定のレベルを超えた時点で、発音の登録作業を終了する。

以上の手順について、評価が容易である日本語とヒンディー語、ネパール語の3つの言語への対応を検討し、その中で、本研究の手法の基盤となる理論構築を目指した。

本研究の実験用システムとして、研究協力者のマンチェスター大学Blenkhorn名誉教授がこれまでに開発し、現在は無料で一般公開されるスクリーンリーダーである「Thunder」(<http://www.screenreader.net/>)を改良して利用した。

### 4. 研究成果

要求分析については、本研究に適用される

既存の合成音声技術についての調査を行った。その結果、例えばインドにおける複数の言語に対する合成音声ソフトの開発が進むなど、途上国における合成音声技術の母語対応が徐々にではあるものの進展も見られることが明らかになった。しかし、一方ではマレーシアのように、母語であるマレー語の合成ソフトの研究は存在しても、一般の人がそのソフトウェアを利用する状況には至っていないことも明らかになった。多くの国々では、英語など、既存の合成音声ソフトが対応した言語を理解できる人がその言語の OS 環境でのみ合成音声ソフトを利用することに限られるのが現状である。

一方、すでに合成音声ソフトが存在する国々では、合成音声ソフトに対する極めて高いニーズを持ち、そうしたソフトを日常的に活用している視覚障害のある（特に全盲の）人達がいる。彼らは、スクリーンリーダーと呼ばれる画面上に表示されたテキスト情報を合成音声で読み上げるソフトを利用して文字情報にアクセスしている。日本や英国、インドにおけるスクリーンリーダーの利用者について、専門家や支援団体にインタビュー調査を行ったところ、初心者の中には、スクリーンリーダーの持つ合成音声ソフトの音質の良さ、すなわち人間に近い自然な発音に重要性を感じる一方、その利用に慣れてくるに従い、むしろ機械的な音声の方が好ましいと感じる利用者が増えることがわかった。その理由として、スクリーンリーダーの利用者の多くは、より早く情報にアクセスするために合成音声の読み上げ速度を何倍かに高めており、速度を上げた時は、機械的な音声の方が単調で聞き取りやすいという人が多い。このことは、合成音声ソフトがスクリーンリーダーに応用される場合、他の応用分野で期待される高品質な合成音声を必ずしも必要としないことを意味する。

次に上記「研究の方法」で紹介された既存合成音声ソフトを応用するアルゴリズムの有効性を確認するため、既存のヒンディー語合成音声ソフト（ScanSoft Lekha22）を用いて、日本語テキストを読み上げる実験システムを構築した。具体的には、ヒンディー語のデーヴァナーガリー文字を用いて日本語の50音表の発声対応を作成した後、国立国語研究所が公開する「現代日本語書き言葉均衡コーパス」語彙表（頻度リスト）を用いて、頻出上位語の発音の確認と変換リスト項目の追加・修正を行った。以上で得られた発音ルールを適用した実験システムを用いて、複数の日本語のニュース記事を読み上げさせたものを、日本語を母語とし、聴覚能力に問題のない4名の実験協力者（男性2名、女性2名）に聞かせたところ、各試行とも聞き取れる範囲が全体の15～30%に留まった。この聞

き取れる割合は、読み上げ速度を、上記の元となったヒンディー語合成音声ソフトの標準の読み上げ速度に対して0.5倍から1.5倍に変更しても大きな変化は見られなかった。一方、視覚障害があり、日常的に日本語の合成音声ソフトを用いて情報へのアクセスを行っている実験協力者（男性）に対して、同じ条件で読み上げ音声を聞かせたところ、5～30%と、上記最初の4名の実験と同様の結果が得られた。

次に同じヒンディー語合成音声ソフト（ScanSoft Lekha22）を用いて、ネパール語テキストを読み上げるシステムの構築を試みた。ネパール語は、ヒンディー語と同じデーヴァナーガリー文字で表記され、そのほとんどを共有しており、また、文法的にも2つの言語は近い。しかし、異なる発音も多数含まれる。本研究では、上記の合成音声ソフトが発音するデーヴァナーガリー文字、および数字について、ネパール語を母語とする2名（男性1名、女性1名）の協力者（男性には視覚障害があり、普段より英語の合成音声ソフトを利用していた）により、ネパール語の発音と異なるもの（これは、一般的なヒンディー語とネパール語の違いではなく、合成音声ソフト ScanSoft Lekha22 が出力するネパール語用には不適切な発音）を抽出し、表1のような変換表を作成した。

表1 本研究で採用したネパール語音声への基本変換表

文字	誤った発音	変換先発音
त	Achi chub	Ta
न	Na	Yan
प	Pya	Pa
ब	Bya	Ba
क	Kasa	Chhya
े	E	Ai
२	Do	Dui
६	Chhe	Chha
९	Naw	Nau

構築されたシステムでは、この変換表に含まれる文字が現れた場合、それを「変換先発音」の欄に描かれた文字列に変換して、発音するようにした。その後、Nepali Language Resource Center 等のネパール語最頻出単語リスト等を用いて頻出上位語の発音の確認と変換リスト項目の追加・修正を行った。以上で得られた発音ルールを適用した実験システムを用いて、以上2名のネパール語の母語話者に複数のネパール語のニュース記事を読み上げさせたところ、各試行とも聞き取れる範囲が全体の60～80%になり、内容もよく理解できるとの結果が得られた。このネパール語の例の場合においても、視覚障害に伴う合成音声ソフトの日常的な利用の有無に

よる差は見られなかった。

以上から、本研究で提案された方法は、ヒンディー語からネパール語のように、言語的・音声学的に近い言語ではその有効性が期待されるものの、以下のような各言語に依存した困難が生じることが明らかとなった。

(1)元となる合成音声ソフトが持つアクセントやイントネーションの特徴を制御できない

(2)元となる合成音声ソフトが持たない発声パターンは生成できない

(3)文法の情報に基づく処理を別途追加する必要がある

(2)については、出力に不自然さが発生してしまうものの、部分的に録音音声で補うなどの方法も考えられる。以上の問題点は、言語学的・音声学的の違いに基づく個別性が高く、既存合成音声エンジンとターゲット言語の組み合わせについて、さらなる検討が必要である。

本研究では、既存の合成音声ソフトをベースとして、別の新しい言語に対応させるということもあり、対象となる言語への対応について、ある程度誤った発音が含まれてしまうことは避けがたいことであった。しかし、要求分析の調査にも述べたように、合成音声ソフトの利用ニーズの高い視覚障害のある人のように、ある程度不完全な合成音声システムであってもうまく利用できる人がおり、こうした人に対しては、発音の不十分さが許容されるかとも思われたが、今回、視覚障害に伴う合成音声ソフトの日常的な利用の有無による差は見られなかった。この点についても今後あわせての検討が必要であると考え

## 5. 主な発表論文等

(研究代表者、研究分担者及び連携研究者には下線)

〔雑誌論文〕(計0件)

〔学会発表〕(計2件)

巖淵 守・渡辺崇史(2012) アプリ開発をどのように進めるべきか? , ATAC カンファレンス 2012 東京 , 東京都市センターホテル, 2012年12月22~23日

巖淵 守(2013) 身の回りにあるテクノロジー(アルテク)が生み出す最新の支援技術 ATACカンファレンス 2013 京都, 京都国際会館, 2013年12月14-15日

〔図書〕(計1件)

巖淵 守(2012) 役立つはずなのに使われない... 支援技術の開発と利用の狭間, バリアフリー・コンフリクト(中邑賢龍・福島 智 編), 東京大学出版会, pp.29-45

〔産業財産権〕  
出願状況(計0件)

名称:  
発明者:  
権利者:  
種類:  
番号:  
出願年月日:  
国内外の別:

取得状況(計0件)

名称:  
発明者:  
権利者:  
種類:  
番号:  
取得年月日:  
国内外の別:

〔その他〕  
ホームページ等

## 6. 研究組織

### (1) 研究代表者

巖淵 守 (IWABUCHI, Mamoru)  
東京大学・先端科学技術研究センター・准教授  
研究者番号: 80335710

### (2) 研究分担者 ( )

研究者番号:

### (3) 連携研究者 ( )

研究者番号: