

科学研究費助成事業 研究成果報告書

平成 29 年 6 月 19 日現在

機関番号：62615

研究種目：挑戦的萌芽研究

研究期間：2015～2016

課題番号：15K12071

研究課題名(和文) 統計的音声合成を利用したインタラクティブオーディオブックと集合知への応用

研究課題名(英文) Interactive audiobook using statistical parametric speech synthesis and collective intelligence

研究代表者

山岸 順一 (Yamagishi, Junichi)

国立情報学研究所・コンテンツ科学研究系・准教授

研究者番号：70709352

交付決定額(研究期間全体)：(直接経費) 2,600,000円

研究成果の概要(和文)：電子書籍には音声合成機能が搭載され、ユーザは電子書籍を読んで楽しむだけでなく、聞いて楽しむことも出来る。この技術に、様々な声質や発話様式による音声生成が可能な統計的音声合成を組み合わせれば、合成音声の表現をユーザ自身がインタラクティブに制御可能なプラットフォームに変わる。この目的のため、本研究では音響モデリング技術を音声の変換関数を因子化することで高度化させた。話者の変換行列と感情の変換行列を明示的に因子化させ、ある話者から推定された感情の変換行列を、全く別の話者においても利用できる適応アルゴリズムを提案した。また話者の性別や年齢を因子化したシステムも構築した。提案オーディオブックも試作した。

研究成果の概要(英文)：Nowadays e-book readers have speech synthesis functions and users can enjoy not only reading but also listening to the e-books. If statistical parametric speech synthesis, which can flexibly generate various voice types of synthetic speech in various speaking styles, is combined with the e-book readers, e-books may become a future platform where the users can operate the controls of expression of synthetic speech interactively. For this purpose, we have advanced acoustic modeling techniques by means of factorizations of speech transformation functions. More specifically, we explicitly factorized speaker and emotional transformations and proposed a new adaptation algorithm to transplant emotional transformations estimated from a speaker into another speaker. We also constructed a new system where speaker's gender and age are factorized. A prototype e-book reader based on proposed speech synthesis techniques was also built for demonstrating the new ideas.

研究分野：音声情報処理

キーワード：音声合成 オーディオブック 集合知 機械学習

1. 研究開始当初の背景

統計的音声合成は近年研究が劇的に進んでいる。研究代表者らにより提案、開発された話者適応技術などにより、様々な声質、発話様式、方言による音声出力が実現可能になった。

研究開始当初、電子書籍に音声合成機能が搭載されてはじめており、ユーザは電子書籍を読んで楽しむだけでなく、聞いて楽しむことも出来るようになりつつあった。これはオーディオブックと呼ばれ、オーディオブックを対象とした音声合成の研究は、研究開始当初そして現在も盛んに行われている。世界的な音声合成のコンペである「Blizzard Challenge」でも2012年以降オーディオブックが評価対象である。

2. 研究の目的

オーディオブックに現在利用されている音声合成は、文章をただ読み上げているだけであり魅力に欠けている。しかし研究代表者が開発してきた声を自由に变化させる技術を利用することで、オーディオブックという媒体を、音声を聞くだけでなく、合成音声の表現をユーザ自身がインタラクティブに制御し、魅力的なオーディオブックを自由に創作することが出来るエンターテインメント性の高いプラットフォームに拡張可能と考えた。

また、さらに提案インタラクティブオーディオブックをウェブ上で公開すれば、ユーザがどの様に音声合成を修正したかというメタデータを収集できる。そのメタデータを機械学習に利用することで、これまで高精度な予測が大変困難であり、多くの場合人手で対応せざるをえなかった、テキストを読み上げる話者や方言の自動決定といった未解決難題に対して集合知を利用した新たな方法論を築くことが可能になるであろうと考えた。

そこで、インタラクティブオーディオブックの構築に必要な基盤技術の統合研究と、上記メタデータを集める環境を構築することを行う。

3. 研究の方法

(1) ユーザが電子書籍を読み上げる音声合成の声優を自由に配置し、使用する方言を好みに応じてアレンジし、また合成音声の抑揚など韻律表現もペンなどで制御することを可能にするためには、研究代表者らが長年開発・改善し、研究分野をリードしている幾つかの音響モデル基盤技術を理論的に統合し、さらなる改良を実施する必要がある。

特に、ユーザにより変更することが可能なすべての組み合わせ(例:話者数×方言数×複数の発話様式)を網羅する音声データを収録することはコストが非常にかかり非現実的であるため、音声の変換関数などをより一般化させさせる必要がある。具体的には、音声の変換関数を因子化(Factorization)す

る研究を行う必要がある。

(2) インタラクティブオーディオブックの試作

次に、統合された変換技術を利用することで、音声合成の表現をユーザ自身がインタラクティブに制御し、魅力的なオーディオブックを自由に創作することが可能な斬新なプラットフォームを試作する。ユーザがエンターテインメントとして楽しみながら利用出来る様、ポッドキャストを共同でアノテーションするPodCastle(podcastle.jp)などのデザインを参照にしつつ、インタラクティブオーディオブックのユーザインターフェイスを注意深くデザインし、試作品を開発する。

(3) 話者、方言、発話様式のタグ付け

試作したインタラクティブオーディオブックシステムをアプリ化する。その際、ユーザが文章の一部とテキストを読み上げるキャラクターを色で対応付けられるようプログラムを拡張する。方言や感情表現についても同様である。そして、ユーザによりタグ付けされた、テキストを読み上げるのにふさわしい話者、方言、発話様式等の情報の分析を詳細に行う。

4. 研究成果

(1) 音響モデル基盤技術

音声の変換関数の因子化のため、研究代表者が2009年に提案したCSMAPLRと呼ばれる隠れマルコフモデル音声合成の話者適応技術の枠組みをさらに拡張し、他の異なる話者の音声データから得られた感情表現を、全く別の話者へ「移植」するための新アルゴリズムを提案した。このアルゴリズムでは、話者性を変えずに感情表現のみを移植することを目標とし、話者情報および感情情報の変換方法はどちらも平均声モデルからの話者適応を利用している。ただし、従来の話者適応とは異なり、話者の変換行列と感情の変換行列を明示的に因子化し、これらの二つの変換行列を統合し、隠れマルコフモデルへ掛け合わせることで、話者と感情表現を同時に変換する新アルゴリズムを提案した。これを応用すれば、例えば、「話者性の変換行列を話者Aから推定、感情の変換行列は異なる話者Bから推定した上で、これらの変換行列を掛け合わせることで、(感情音声を収録していない)話者Aの感情合成音声を作り出す」ということが可能になる。

実験結果からは、4種類の感情表現(怒り、喜び、悲しみ、驚き)を表した変換行列を、3名の女性話者および3名の男性話者へ移植するという実験を行った。実験結果から、変換された合成音声の自然性は若干劣化するものの、合成音声には移植した感情表現が適切に含まれることを確認した。

また、隠れマルコフモデル音声合成の話者適応だけでなく、話者、性別、年齢を因子化

し、さらに高度に制御可能な DNN ベースのテキスト音声合成システム及びその話者適応手法も新たに開発した。

大規模な日本語コーパスを用いた実験を行い、年齢が 10 歳から 80 歳までの男性 68 名、女性 70 名の発声したスタジオ品質の音声データを用いて、新 DNN 音声合成システムを構築した。この新 DNN システムにより、(1) 個々の話者の個人性を再現できること、(2) バックプロパゲーションを介して未知話者に適したベクトル表現を推定し、未知話者の音声も合成できること、(3) 話者の性別、年齢を所望した通りに変更できることを実験的に示した。この結果、複数話者の音声を高性能に合成することが可能になり、また未知話者へも少量の適応データにより変換でき、そして話者の性別と年齢も直感的に操作できるようになった。

さらに、このような音声の特徴制御と文章を対応づけることができる様、自然言語処理で利用されている Topic モデルに検出されたトピックベクトルと音声合成システムとを密結合させる研究も行った。

(2) オーディオブックのアプリ試作

文章を読み上げる声優、使用する方言、発話様式等をユーザが自由にかつインタラクティブに配置可能なオーディobook 創作アプリの試作を行った(図1)。試作したアプリは iOS 上のデバイス動作し、ePub フォーマットの電子書籍を統計的音声合成にて読み上げるものである。また、ユーザがどの話者をどのコンテンツに利用したかなどのメタ情報を逐次記録できる様にした。



図 1. 試作したオーディobook 創作アプリ

さらに、隠れマルコフモデル音声合成だけでなく、ディープラーニングに基づく音声合成システムもその重要性が高まっていることから、iOS の GPU API (Metal API) を利用し、DNN 音声合成システムの iOS 実装も行った。そして、より適切な韻律表現を実現できる様、プロの声優が読み上げたオーディobook の音声データから音声合成用音響モデルを構築した。

(3) 話者、方言、発話様式のタグ付け

試作したインタラクティブオーディobook システムにおいて、ユーザが文章の一部とテキストを読み上げるキャラクターを色で対応付けられるようプログラムを拡張した。

しかし当初予定していなかった DNN 音声合成システムの iOS 実装などを行い、予定以上に時間を費やした関係上、メタデータの収集は、簡易的に実行するに留まった。テキストを読み上げるのにふさわしい話者、方言、発話様式等の大規模収集及び、その情報の分析は今後継続して行っていく予定である。

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 1 件)

J. Lorenzo-Trueba, R. Barra-Chicote, R. San-Segundo, J. Ferreiros, J. Yamagishi, and J.M.Montero, "Emotion transplantation through adaptation in HMM-based speech synthesis", Computer Speech & Language volume 34, issue 1, pp.292-307, November 2015, 査読有り
<https://doi.org/10.1016/j.csl.2015.03.008>

〔学会発表〕(計 6 件)

H.T. Luong, S. Takaki, G.E. Henter, and J. Yamagishi, "Adapting and Controlling DNN-Based Speech Synthesis Using Input Codes", The 42nd IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP 2017), New Orleans, USA, 2017 年 3 月 5 日-9 日.

J. Lorenzo-Trueba, R. Barra-Chicote, A. Gallardo-Antolin, J. Yamagishi, and J.M. Montero, "Continuous Expressive Speaking Styles Synthesis based on CVSM and MR-HMM", The 26th International Conference on Computational Linguistics (COLING 2016), Osaka, Japan, 2016 年 12 月 13 日-16 日

Lauri Juvola, Xin Wang, Shinji Takaki, SangJin Kim, Manu Airaksinen, Junichi Yamagishi, "The NII speech synthesis entry for Blizzard Challenge 2016", Blizzard Challenge workshop 2016, California, USA 2016 年 9 月 16 日

〔図書〕(計 0 件)

〔産業財産権〕

出願状況(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：

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〕

該当なし

6．研究組織

(1)研究代表者

山岸順一（YAMAGISHI, Junichi）

国立情報学研究所・コンテンツ科学研究

系・准教授

研究者番号：70709352

(2)研究分担者

該当なし

(3)連携研究者

高木信二（TAKAKI, Shinji）

国立情報学研究所・コンテンツ科学研究

系・特任助教

研究者番号：50735090

(4)研究協力者

該当なし