様　式　Ｃ－１９、Ｆ－１９－１、Ｚ－１９　（共通）

科学研究費助成事業　　研究成果報告書

平成　２９　年　　６　月　２０　日現在

機関番号：　６２６１５
研究種目：　挑戦的萌芽研究
研究期間：　2014～2016
課題番号：　２６５４００９２

研究課題名（和文）耳からの知識獲得otopediaの研究

研究課題名（英文）Human Learning from Ears: Research on Otopecia

研究代表者
佐藤　健（Satoh, Ken）

国立情報学研究所・情報学プリンシプル研究系・教授

研究者番号：００２７１６３５

交付決定額（研究期間全体）：（直接経費）　　2,100,000円

研究成果の概要（和文）：聴覚を利用した学習ツールのため、音声合成において話者を変換する話者適応技術の研究を進め、これまで利用した隠れマルコフモデルでなく、ディープラーニングを用いた場合でも少量の声のサンプルから声のデジタルクローンが可能であることを実験的に示した。また、音声合成における話者適応技術を利用したオーディオブックリーダiOSアプリも試作し、ePubフォーマットの電子書籍を所望の話者により読み上げることを実演した。これらの実験環境の構築により、話者の声の違いによるengagement timeの分析、そしてゆくゆくは、話者の声の違いににによりもたらされる学習や記憶への影響を調べる土台を構築することができた。

研究成果の概要（英文）：For human learning from ears, we developed a method of making speaker-adaptive voice synthesis system. We showed that we can use deep-learning for the purpose with small amount of voice data as the previous Hidden Markov models. We also made a iOS application of audio book reader using the above method and demonstrated that the application could read digital books in a ePub format using speaker's voice. We could use this system for analysis of engagement time and learning performance using various voices.

研究分野：　人工知能基礎

キーワード：　音声合成　話者適応　ディープラーニング　Otopedia

１．研究開始当初の背景

聴覚を利用した学習ツールは、満員電車での通勤時やジョギング時に使えるので、効率的な学習として非常に役に立つ。しかし、最近までは、聴覚を用いた学習手法では、以下の問題があった。

・良質コンテンツを作るためには、通常はアナウンサーなどを使わざるを得ないが、そうなるとそのアナウンサーの話す能力によってコンテンツの生産量が決まってしまい、大量のコンテンツを作ることはできない。

・上記の問題を解決するために音声自動合成プログラムを使うことが考えられる。この場合は、音声モデルを使って高速にコンテンツを作るため、大量のコンテンツを作ることはできるが、今までの音声合成プログラムでは、音声の質が悪く、棒読みのような音声となってしまい、聞くに堪えないコンテンツになってしまう。

しかし、近年、CEVIO(http://cevio.jp/) のような良質の音声合成ソフトが開発されるようになってきた。これからそのようなソフトを使ったコンテンツは増えていくものと思われる。研究代表者はこの CEVIO を使って判例等の読み上げコンテンツを作って、ジョッギング中や通勤中に聞いてみたが、確かに不自然ではないので聞きやすいが、条文の記憶までいたらなかった。これにはいろいろな原因が考えられるが、情報の提示の仕方に大きな問題があるように感じた。

一方アメリカでは、通勤時の運転中で聞く電子ブックのようなアプリケーションソフトがよく売れているという話であるし、日本でも「スピードラーニング」のような聴覚を利用した学習ツールが市販されているので、聴覚学習ツールのニーズは潜在的に大きいと考える。しかし、これらのツールは単に何回も聞き流して覚えるような程度のものでしかなく、その効果を科学的に検証したものはほとんどない。

我々のサーベイにおいて、そのような科学的に検証したツールといえるものは、ATR からのスピンオフ企業が販売している「ATR CALL BRIX」(http://www.atr-lt.jp/products/brix/) という製品しかなかった。ただし、この学習ツールは英語の発音訓練に特化されており、本研究の目標である効果的な情報記憶（判例記憶等）に関しての聴覚ツールの科学的実証研究は、我々が調査した範囲ではない。

２．研究の目的

上の背景を踏まえ、本研究では目的を以下のように設定した

・聴覚を利用した学習教材に関して、記憶に効果的な情報提示法の研究

・そのような情報提示法を利用した学習ツールの効果の科学的実証

なお、研究課題名の中の otopedia とは、oto-(ギリシャ語で「耳」の意味）と pedia(ギリシャ語で「学習」の意味）を組み合わせた研究代表者が提案している造語である。)

３．研究の方法

当初の計画では、聴覚者モデルを作り、そのモデルに応じた、学習モデルを構築し、それに基づいてツールを開発し、有用性を科学的に検証する予定であった。しかし、研究分担者の山岸がオーストリアで行っていた視覚障害者との取り組みを通し、ゲームなどのタスクにおいて、音声合成の話者性を変えた場合、タスクを継続する時間が長くなることが分かったため、それに基づいてツールをつくることとなった。

４．研究成果

平成 26 年度は関係技術についての研究者の研究集会を開催し、関連技術の調査を行った。具体的には、前述の ATR CALL BRIX の開発者の講演聴講や、文字情報の読みやすさの研究、複数人の音声から言葉に対応する要素抽出の研究、物の形状と幼児における表現方法の研究、自然言語処理における自動要約の研究、視覚障害者用ゲームなどのタスクにおける音声合成の話者性の影響調査を行った。その結果、自分の声や馴染みのある声にした場合、ゲームなどのタスクを継続する時間が長くなるという予備実験結果が有望ということがわかった。平成 27 年度では、前述の自分の音声使用によるゲームの継続時間の変化分析を厳密に行い、視覚障害者用ゲームをプレイする際に、音声の話者性に対する馴染みさがゲームにおける集中時間だけでなく、ゲームのクリアに必要なステップ数などパフォーマンスにも影響することを見つけ出した。 使用したゲームは記憶ゲームと迷路ゲームという 2 つのゲームであり、このゲームを、学校の子供たちによる音声、その教師による音声、ラジオの声優による音声で構築し、それぞれの音声を使用した場合の影響について測定した。その結果、生徒が自分の声で構築された合成音声を使用したゲームをプレイした場合、他のケースと比べ、集中持続時間が有意に長くなり、ゲームのクリアに必要なステップ数も有意に減少するということを実験的に確認した。子供たちが合成音声を自分のものとして認識していないにもかかわらず、この結果は観察されており、大変興味深い現象である。平成 28 年度では、自分の声だけではなくなじみのある声であれば同様の効果が出るのではないかと考え、母親の声を合成により子供に提示し、集中力の持続時間の変化を計測する実験を計画していたが、被験者の手配に手間取るとともに、聴覚学習モデルの構築がうまくいっていなかったこともあり、このような実験環境を容易に構築するツールを作成することに計画を変更した。そこで、音声合成において話者を変換する話者適応技術の研究を進め、テキストだけではなく話者、性別、年齢の特徴量

を用いた DNN ベースのテキスト音声合成システムを開発した。このため、大規模な日本語コーパスを用いて、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、以下の 3 つの実験を行った。それらは、(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、である。この結果、提案された DNN システムにより複数話者の音声を高性能に合成することが可能であり、また未知話者へも少量の適応データにより変換できること、そして話者の性別と年齢も直感的に操作できることを示した。その他、音声の自然な韻律を機械学習し、文章から予測する AUTO REGRESSIVE RECURRENT MIXTURE DENSITY NETWORK という新たなモデルの提案及び実験も行った。また、音声合成における話者適応技術により構築された複数話者モデルを利用するオーディオブックリーダ iOS アプリも試作し、ePub フォーマットの電子書籍を様々な話者により読み上げることを実演した。これらの実験環境の構築により、話者の声の違いによる engagement time の分析、そしてゆくゆくは、話者の声の違いによりもたらされる学習や記憶への影響を調べる土台を構築することができた。

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　3　件）

① Pucher, Michael / Toman, Markus / Schabus, Dietmar / Valentini-Botinhao, Cassia / Yamagishi, Junichi / Zillinger, Bettina / Schmid, Erich (2015): "Influence of speaker familiarity on blind and visually impaired children's perception of synthetic voices in audio games", In INTERSPEECH-2015, pp.1625-1629, (2015), 査読有.

② Hieu-Thi Luong, Shinji Takaki, Gustav Eje Henter, Junichi Yamagishi,
"A DNN-based text-to-speech synthesis system using speaker, gender and age codes", The journal of the Acoustical Society of America, vol. 140 (2016), 査読有.
http://dx.doi.org/10.1121/1.4969152

③ Michael Pucher, Bettina Zillinger, Markus Toman, Dietmar Schabus, Cassia Valentini-Botinhao, Junichi Yamagishi, Erich Schmid, Thomas Woltron, "Influence of speaker familiarity on blind and visually impaired children's and young adults' perception of synthetic voices". Computer, Speech, and Language(2017) , 査読有.
https://doi.org/10.1016/j.csl.2017.05.010

〔学会発表〕（計　3　件）

① 山岸順一、「音声合成で良い音を作る！」2015 年春季研究発表会特別企画 音響学シンポジウム 招待講演、 2015 年 3 月 15 日(日)、中央大学理工学部後楽園キャンパス、東京

② Hieu-Thi Luong, Shinji Takaki, Gustav Eje Henter, Junichi 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, LA, USA (2017).

③ Xin Wang, Shinji Takaki, Junichi Yamagishi
"An Autoregressive Recurrent Mixture density Network For Parametric Speech Synthesis", The 42nd IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP 2017), New Orleans, LA, USA (2017).

〔図書〕（計　　件）
山岸順一、徳田恵一、戸田智基、みわよしこ、情報研シリーズ１９「おしゃべりなコンピューター音声合成技術の現在と未来ー」丸善ライブラリ、２０１５年３月

〔産業財産権〕
○出願状況（計　０　件）
○取得状況（計　０　件）
○ホームページ等
なし

６．研究組織
(1)研究代表者
佐藤　健　（SATOH, Ken）
国立情報学研究所・情報学プリンシプル系・教授
研究者番号：00271635

(2)研究分担者
山岸　順一（YAMAGISH, 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

(3)連携研究者
相澤　彰子（AIZAWA, Akiko）
国立情報学研究所・コンテンツ科学研究系・教授
研究者番号：90222447

宮尾　祐介(MIYAO, Yusuke)
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：00343096

坊農　真弓(BONO, Mayumi)
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：50418521