

科学研究費助成事業 研究成果報告書

令和 4 年 5 月 21 日現在

機関番号：56101

研究種目：若手研究

研究期間：2018～2021

課題番号：18K18170

研究課題名（和文）プレゼンテーションスキルの向上を支援する模範音声の自動生成システム

研究課題名（英文）Automatic generation system for exemplary presentation speech to improve presentation skills

研究代表者

太田 健吾（Ohta, Kengo）

阿南工業高等専門学校・創造技術工学科・准教授

研究者番号：80712801

交付決定額（研究期間全体）：（直接経費） 2,400,000円

研究成果の概要（和文）：本研究では、人工知能の一分野である音声言語処理の技術を用いて、発表者の模範となる理想的な発表音声を自動生成し、効果的な自主練習を支援するシステムの構築に取り組んだ。このシステムを構築するための要素技術として、冗長表現を発表音声から自動検出・除去する技術、不適切な単語や言い回し等をより簡易的で発表に適切なものに変換するための技術、発表音声の発話内容を高精度に音声認識するための技術、発表音声の韻律を改善し、話者の個人性を維持しながら模範音声を生成するための技術についてそれぞれ検討した。

研究成果の学術的意義や社会的意義

大学や高等専門学校などの教育機関において、基礎的な学力だけでなく、「分野横断的能力」の一つとしてプレゼンテーションスキルの養成が重要視されるようになってきた。こうした要求に対し、学生が主体的に発表や議論を行うアクティブラーニング型の授業が各校で展開されているが、プレゼンテーションの指導を行う教員の負担は大きく、学生の主体性を活かしながら、かつ、効果的に学生のスキルを伸ばす枠組みの構築が喫緊の課題となっている。本研究で提案するシステムにより、プレゼンテーション教育における教員の負担軽減と学生の効果的な自主練習の実現が期待できることから、社会的意義は大きい。

研究成果の概要（英文）：In this research, we have been working on the development of an automatic generation system for exemplary presentation speech that supports effective independent practice by automatically generating an ideal presentation speech to serve as a model for presenters, using the technology of spoken language processing and natural language processing. The elemental technologies for building this system include: automatic detection and removing redundant expressions such as filled pause from speech, semi-automatic conversion of inappropriate words and phrases into simpler and more appropriate ones for presentation, highly accurate speech recognition of the content of speech, improvement of the prosody of speech and generation of exemplary speech while maintaining the individuality of the speaker.

研究分野：音声言語処理

キーワード：音声言語処理 自然言語処理 発表音声 プレゼンテーションスキル 音声認識 音声合成

1. 研究開始当初の背景

大学や高等専門学校などの教育機関において、基礎的な学力だけでなく、コミュニケーションスキルをはじめとする「分野横断的能力」の養成が重要視されるようになってきた。国立高等専門学校機構で策定されたモデルコアカリキュラムでも、コミュニケーションスキルに関する教育目標が設定され、「説明の方法を工夫しながら、自分の意見や考えをわかりやすく伝える」といったプレゼンテーションの能力がこれに含まれている。こうした要求に対し、学生が主体的に発表や議論を行うアクティブラーニング型の授業が各校で展開されているが、プレゼンテーションの指導を行う教員の負担は大きく、学生の主体性を活かしながら、かつ、効果的に学生のスキルを伸ばす枠組みの構築が喫緊の課題となっている。

一方で、音声言語処理の分野では、コミュニケーションスキルの向上を支援する研究として、話し方に関するアドバイスを行う対話型のシステムや、話し方の自動採点を行うシステムが提案されている。しかし、対話型のシステムでは、「もっとゆっくり話して」や「もっと声を大きく」といった定性的なアドバイスを行うことはできるが、具体的にどの程度の速さや大きさを発話すればよいのかをユーザは定量的に知ることができないため、最適な話し方に調整するためには対話を繰り返し行わなければならない。また、自動採点を行うシステムでは、ユーザは前回の採点結果と比べることで自分の発表が良くなったのかどうかを知ることができるが、さらに良くするために具体的にどのような発表を目指せばよいのかを知ることができない。このように、既存のシステムは、ユーザが目標とすべき模範を具体的に示していないという点で、直接的な支援ではなく、間接的な支援に留まっている。より直接的な支援を行うためには、与えられた音声がある特定の観点において良いか悪いかを判定する識別的なアプローチではなく、複数の観点から見て最も良い音声を具体的に定義するという生成的なアプローチが必要とされている。

2. 研究の目的

本研究の目的は、音響的な観点と言語的な観点の両方において最適化された「最もわかりやすい話し方」を具体的に定義するという工学的問題を解くことにより、プレゼンテーション教育における教員の負担軽減と学生の効果的な自主練習を実現することである。

本研究の工学としての独自性は2点ある。1つ目は、話す速度や声の大きさといった音響的な特徴だけでなく、語句の使い方や単語の係り受けといった言語的な特徴も考慮に入れた「わかりやすさ」の尺度を提案する点である。2つ目は、与えられた音声がある特定の観点において良いか悪いかを判定する識別的なアプローチではなく、複数の観点から見て最も良い音声を具体的に定義するという生成的なアプローチを取る点である。

また、本研究の学習支援システムとしての独自性は、従来のようなアドバイスや自動採点といった間接的な支援ではなく、目標とすべき模範音声を話者に具体的に提示するという直接的な支援を行う点である。話者はシステムの出力した模範音声に倣って発話することで、従来よりも効果的な自主練習を行うことができる。

3. 研究の方法

本研究で提案するシステムの概要図を図1に示す。



図1 提案システムの概要

図1のようなシステムを実現するための要素技術として、(1)冗長表現を発表音声から自

動検出・除去する技術、(2) 不適切な単語や言い回し等をより簡易的で発表に適切なものに変換するための技術、(3) 発表音声の発話内容を高精度に音声認識するための技術、(4) 発表音声の韻律を改善し、話者の個人性を維持しながら模範音声を生成するための技術についてそれぞれ研究開発および検討を行った。

4. 研究成果

(1) 冗長表現を発表音声中から自動検出・除去する技術

「えっとー」「あー」などの間投詞や言い淀みといった冗長表現を発表音声から自動検出するアルゴリズムの研究開発を行った。具体的には、日本語話し言葉コーパスに含まれる講演音声を学習データとして、音声から抽出された音響的な特徴量(基本周波数やゼロ交差率、およびそれらの統計量など)および言語的な特徴量(word2vecに基づく単語分散表現)に基づいて冗長表現を検出する検出器を、深層学習の手法の中でも特に時系列データの扱いに有効な Long short-term memory (LSTM)ネットワークによって学習した。評価実験を行った結果、標準的な機械学習を用いた検出器よりも高い検出精度を達成することができた。

(2) 不適切な単語や言い回し等をより簡易的で発表に適切なものに変換するための技術

発表音声に含まれる用法の不適切な語彙や、難解で理解しにくい語句を、より平易かつ適切な語句に変換するアルゴリズムの研究開発も行った。具体的には、やさしい日本語コーパスに含まれる日本語の言い換えペアを学習データとして、入力された文中の語句を平易に言い換える変換器を、深層学習の手法の一つである注意機構付き LSTM によるエンコーダ・デコーダモデルを用いて学習した。これにより、一部の語句について、よりわかりやすい語句に変換を行えることが確認できた。

また、発表者がより適切な語彙を用いて発表を行えるように、発表原稿の推敲を支援するモジュールの研究開発を行った。具体的には、入力された発表原稿の文章に対し、日本語の概念シソーラスを用いて原稿中の内容語の類義語を検索・提案する。ユーザとなる発表者が効率的に推敲作業を行えるように、GUIに基づくインタフェースも提案した。類義語との類似度を定量的に扱うために、単語の分散表現を用いる方法についても検討した。

(3) 発表音声の発話内容を高精度に音声認識するための技術

システムの要素技術である冗長表現の自動検出・除去や語彙の最適化モデルをさらに高精度化するために、これらの入力となる音声認識結果の精度改善に取り組んだ。具体的には、発表音声に含まれる話し言葉特有の問題(フィラーや言い淀み)を考慮した End-to-End 音声認識モデルを学習するために、こうした非流暢の音響特徴に対応する書き起こし箇所をラベルに置き換える手法を提案した。また、発表音声に含まれる特定の話題に関連した単語や言い回しを高精度に認識するために、Density Ratio Approach の考え方を応用して、特定のドメインに特化した学習データを音声認識モデルの学習において効果的に利用する手法を提案した。

(4) 発表音声の韻律を改善し、話者の個人性を維持しながら模範音声を生成するための技術

発表音声の韻律(ピッチや話速)をより聞き取りやすく自動調整するアルゴリズムについて検討を行った。具体的には、音声分析変換合成の技術を用いて、入力された発表音声のピッチと話速を変換するモジュールを構築した。変換された音声の聞き取りやすさを被験者実験によって評価し、最適なピッチと話速に関する知見を得た。

また、発表音声を自動生成するために、音声合成について検討を行った。具体的には、深層学習に基づく音声合成器をベースとし、入力文の音素列だけを入力するのではなく、外部の解析器によって抽出されたアクセントに関する情報も同時に入力することで、より自然な合成音声を得られることを示した。加えて、発表音声と同じ声質の自然な合成音声を生成する手法について検討した。具体的には、前述の日本語のアクセントの特徴を考慮した音声合成手法を拡張し、合成音声の話者性を制御可能な音声合成モデルを構築する手法について検討した。複数の話者を含む音声言語データベースから学習された、話者の特徴を表す埋め込み表現を利用し、この埋め込み表現に基づく話者ベクトルの線形演算により、多様な声質の合成音声を実現できる可能性を示した。

さらに、発表者にとって模範となる理想的なポーズ(無音による間)を提示するために、入力された文に対し、適切な長さのポーズを、文中の適切な箇所に挿入するアルゴリズムについて検討を行った。提案するアルゴリズムでは、単語の分散表現を入力とする深層学習により、入力文中の各単語の直後にポーズを挿入すべきかどうかと、挿入するポーズの長さを同時に推定する。提案手法によって挿入されたポーズの自然性を被験者実験によって評価した結果、従来手法より有意に優れた結果を得ることはできなかったが、さらなる改善に向けて重要な示唆を得ることができた。

5. 主な発表論文等

〔雑誌論文〕 計0件

〔学会発表〕 計16件（うち招待講演 0件 / うち国際学会 1件）

1. 発表者名 松田 美海, 太田 健吾
2. 発表標題 話者ベクトルの線形演算に基づく仮想エージェント用音声合成
3. 学会等名 計測自動制御学会四国支部学術講演会論文集, PS1-09
4. 発表年 2021年

1. 発表者名 森 大輝, 太田 健吾, 西村 良太, 小川厚徳, 北岡 教英
2. 発表標題 タスク外音響情報を付加したEnd-to-End音声認識モデルの設計
3. 学会等名 日本音響学会春季講演論文集, 2-3Q-2
4. 発表年 2022年

1. 発表者名 堀井 こはる, 福田 芽衣子, 太田 健吾, 西村 良太, 小川厚徳, 北岡 教英
2. 発表標題 非流暢ラベルを用いた言い淀み整形End-to-End音声認識
3. 学会等名 日本音響学会春季講演論文集, 1-3-5
4. 発表年 2022年

1. 発表者名 堀井 こはる, 福田 芽衣子, 太田 健吾, 西村 良太, 北岡 教英
2. 発表標題 言い淀みを考慮したEnd-to-End音声認識
3. 学会等名 第19回情報学ワークショップ(WiNF2021), S-5-2
4. 発表年 2021年

1. 発表者名 森 大輝, 太田 健吾, 西村 良太, 北岡 教英
2. 発表標題 Encoder-Decoder音声認識モデルにおける暗黙的言語情報の置換法
3. 学会等名 第19回情報学ワークショップ(WiNF2021), S-5-5
4. 発表年 2021年

1. 発表者名 森 大輝, 太田 健吾, 西村 良太, 小川 厚徳, 北岡 教英
2. 発表標題 Encoder-Decoder音声認識モデルにおける暗黙的言語情報の置換法
3. 学会等名 日本音響学会秋季講演論文集, 1-3-1
4. 発表年 2021年

1. 発表者名 堀井 こはる, 福田 芽衣子, 太田 健吾, 西村 良太, 北岡 教英
2. 発表標題 言い淀みを考慮した自由発話のEnd-to-End音声認識
3. 学会等名 日本音響学会講演論文集, 1-3-3
4. 発表年 2021年

1. 発表者名 森大輝, 太田健吾, 西村良太, 小川厚徳, 北岡教英
2. 発表標題 End-to-end音声認識モデルにおける暗黙的言語情報の置換法
3. 学会等名 音学シンポジウム
4. 発表年 2021年

1. 発表者名 堀井こはる, 福田芽衣子, 太田健吾, 西村良太, 北岡教英
2. 発表標題 言い淀みラベル付けによる非流暢発話のEnd-to-End音声認識
3. 学会等名 音学シンポジウム
4. 発表年 2021年

1. 発表者名 Daiki Mori, Kengo Ohta, Ryota Nishimura, Atsunori Ogawa, Norihide Kitaoka
2. 発表標題 Advanced language model fusion method for encoder-decoder model in Japanese speech
3. 学会等名 Proc. APSIPA ASC, pp. 503-510
4. 発表年 2021年

1. 発表者名 Koharu Horii, Meiko Fukuda, Kengo Ohta, Ryota Nishimura, Atsunori Ogawa, Norihide Kitaoka
2. 発表標題 End-to-end spontaneous speech recognition using hesitation labeling
3. 学会等名 Proc. APSIPA ASC, pp. 1077-1081 (国際学会)
4. 発表年 2021年

1. 発表者名 坂野陽一, 太田健吾, 杉野隆三郎, 岡本浩行
2. 発表標題 応答発話を聞きやすくするためのポーズの自動挿入に基づく話し言葉変換
3. 学会等名 令和2年度計測自動制御学会四国支部学術講演会
4. 発表年 2020年

1. 発表者名 市原大和, 太田健吾
2. 発表標題 プレゼンテーション訓練システムのための韻律の自動調整手法に関する研究
3. 学会等名 令和元年度計測自動制御学会四国支部学術講演会
4. 発表年 2019年

1. 発表者名 増田晃平, 太田健吾
2. 発表標題 プレゼンテーション発表者の発表原稿推敲を支援する類義語検索システムの開発
3. 学会等名 令和元年度計測自動制御学会四国支部学術講演会
4. 発表年 2019年

1. 発表者名 元木洋志, 太田健吾
2. 発表標題 プレゼンテーションスキル訓練システムのための冗長表現の自動検出
3. 学会等名 令和元年度計測自動制御学会四国支部学術講演会
4. 発表年 2019年

1. 発表者名 太田健吾, 市原大和, 増田晃平, 元木洋志
2. 発表標題 プレゼンテーションスキルの向上を支援する模範音声の自動生成システム
3. 学会等名 電子情報通信学会総合大会
4. 発表年 2019年

〔図書〕 計0件

〔産業財産権〕

〔その他〕

-

6. 研究組織

	氏名 (ローマ字氏名) (研究者番号)	所属研究機関・部局・職 (機関番号)	備考
--	---------------------------	-----------------------	----

7. 科研費を使用して開催した国際研究集会

〔国際研究集会〕 計0件

8. 本研究に関連して実施した国際共同研究の実施状況

共同研究相手国	相手方研究機関
---------	---------