

科学研究費助成事業 研究成果報告書

平成 29 年 6 月 7 日現在

機関番号：52601

研究種目：基盤研究(C) (一般)

研究期間：2013～2016

課題番号：25370680

研究課題名(和文) 音声信号処理技術に基づく英語プレゼンテーションのための音声学習支援ソフトの開発

研究課題名(英文) Development of English pronunciation software for oral presentations

研究代表者

堀 智子 (HORI, Tomoko)

東京工業高等専門学校・一般教育科・教授

研究者番号：00269789

交付決定額(研究期間全体)：(直接経費) 3,700,000円

研究成果の概要(和文)：本研究は、英語プレゼンテーションにおける音声面の特徴を発話速度、ポーズ、ピッチ面から分析し、プレゼンテーションに特化した英語発音学習支援ソフトの開発を行った。学習ソフトに必要な音声の読み込み、再生、ピッチ変更、およびピッチ曲線の太さと色の変化による音の強さを表示する機能について検討・試行を行った。また、モデル音声を学習者の声質に近づける機能の試行や学習者とモデル音声を比較して評価するスコアリングシステムの作成を行った。今後はソフトの完成度向上に向けて、ユーザテストやユーザからの意見の反映、また各機能の改善を行う必要がある。

研究成果の概要(英文)：This study investigated the acoustic characteristics of English oral presentations and developed a computer-based learning system for practicing English pronunciation and prosody to deliver effective presentations. Regarding the acoustic characteristics, we analyzed presentations delivered by native speakers of English using Praat for speaking rate, pauses, and pitch. The system we proposed includes functions such as loading speech sounds, playing the sounds, displaying and altering pitch. We also proposed a voice conversion method based on a Gaussian mixture model, and speech scoring system using pitch and intensity. In addition, we proposed a function which changes the thickness and color of pitch contour to show intensity of the sounds as well.

研究分野：英語教育

キーワード：英語プレゼンテーション 英語音声特徴 英語発音学習ソフトウェア

1. 研究開始当初の背景

申請者らは、学生の英語によるプレゼンテーション能力向上をめざし 2006 年に校長裁量経費の支援の下、英語プレゼン研究会を結成し活動を始めた。2008 年度からは文部科学省の教育 G P 「国際通用力のある若き実践的エンジニア育成」プロジェクトに取組み、学生の指導にあたる中で、英語プレゼンテーション能力の向上には、相手に通じる英語を発音し内容を効果的に伝える力の養成がきわめて重要であることが明らかになった。それは、発表用英文原稿が用意できても実際発音などのデリバリーが不十分だと効果的なプレゼンにならない場合が少なくないからである。

効果的なプレゼンテーション時のピッチ、イントネーション、発話速度、リズム等の英語音声特徴に関する研究はまだ十分になされておらず、指導内容や方法も指導者の感覚と経験に大きく依存したままである。効果的なプレゼンテーション指導を進めるには、まずプレゼンテーション時の音声特徴を科学的に研究し、指導すべき項目を明確にすることが急務である。また、学生が発音面の自学自習を進めるためには、効果的なプレゼンテーションのモデルとなる英語を何度も聞いて繰り返す必要があるが、そうしたモデル音声が必要となる度に英語母語話者に録音を依頼せざるを得ず、個々の学生のニーズに十分応えきれないのが現状である。

そこで本研究では英語プレゼンテーション力向上に欠かせない英語発音面での指導項目を明らかにすると同時に、英語発音学習を支援するソフトウェアの開発を目指すに至った。

2. 研究の目的

本研究の目的は、まず効果的な英語プレゼンテーションに不可欠な英語音声面の特徴を音声分析によって明らかにし、日本人英語学習者への指導において重要な項目を明確にすることである。具体的にはプレゼンテーション時の発話流暢性に関係する発話速度、平均発話音節数 (mean length of runs)、ポーズやピッチ (平均ピッチ、ピッチ幅)、強勢といった観点からプレゼンテーション時の音声特徴を明らかにする。その上で、日本人学習者にとっての重要度や習得の難易度を考慮の上、効果的な英語プレゼンに欠かせない発音上のポイントを検討する。

また本研究の目的は、学習者のニーズに合わせてカスタマイズでき、音響学の知識がない人でも使える英語プレゼンテーションに特化した英語発音学習ソフトウェアの開発を行うことである。このソフトウェアは、プレゼンテーションにふさわしいモデル音声の提示、モデルと学習者の音声比較とフィードバック機能を持ち、どの部分でポーズを挿入し強勢を置くべきかなどをわかりやすく表示できる機能をそなえることを目指すも

のである。モデル音声の作成に既存の音声読み上げソフトを用いることや、モデル音声を学習者の声に近づける声質変換などの機能等についても検討する。

3. 研究の方法

1) プレゼンテーション時の英語音声特徴分析

プレゼンテーションの英語音声特徴を抽出するため、の英語母語話者を対象としたパイロットスタディを実施し、データ収集、録音および分析方法を検討することにした。10名の英語母語話者に英文 (10 センテンス) を観客に話すつもりで音読するよう依頼し、その音声を録音した。録音したセンテンス内の単語ごとのピッチ (F0)、長さ (duration)、強さ (intensity) を Praat を用いて測定した。その結果、この実施形態では、プレゼンテーション時の音声特徴が表れにくいことがわかり、言語材料と実施方法を改善することにした。また、プレゼンテーションの特徴を把握するには、ポーズの頻度、長さ、出現場所と回数も分析する必要があることわかった。そこで、これらの課題を検討した後、英語母語話者の数を 26 名 (女性 11 名、男性 15 名) に増やし本実験を行った。本実験では、共同研究者が学会で実際に使用した情報工学に関する研究発表の原稿とスライドを用い、英語母語話者に事前に練習をした上でプレゼンを行ってもらい、その音声を録音した。収集した音声データは発話速度 (speech rate, articulation rate)、平均発話音節数 (mean length of runs)、ポーズ (ポーズ長、生起場所、頻度)、ピッチ (平均ピッチ、ピッチ幅)、強勢といった観点から分析を行った。

2) 英語発音学習支援ソフト開発

ソフトウェアの学習機能開発として、学習ソフトに必要なモデル音声の読み込み、再生、ピッチ変更、学習者の音声取得とピッチ抽出および画面への表示などの機能について検討し、試行した。ユーザーインターフェイス作成では、従来の音声学習ソフトの多くが用いている音声波形表示は音響学の知識がない学習者にとってはわかりにくい点が多いため、直観的にピッチなどの変化が把握できるような表示形式を開発することを目指した。

また、学習者が真似しやすいモデル音声を提示する 1 つの方法として、モデル音声の声質を学習者の声質に近づけることを検討した。イントネーションは、効果的なプレゼンテーションの成否に大きく影響する要素であるが、その習得は個々の音素の習得より困難だと指摘する研究が多い。イントネーション上達には、まずモデル音声のピッチの動きを知覚・認識できるようになることが欠かせない。そこでモデル音声と学習者のピッチや声質を近づければ、モデル音声との違いが認識しやすく効果的ではないかという仮説を

たて、声質変換について検討・試行した。

4. 研究成果

1) プレゼンテーション時の英語音声特徴

英語母語話者の話速については、発話速度 (speech rate) と構音速度 (articulation rate) を測定した。発話速度とは、スピーチの音節数を発話時間 (ポーズ長も含む) で除したものであり、構音速度は発話速度と同様に計算するが発話時間にポーズ長を含まないものである。これらの指標は流暢性の指標としてよく用いられるものである。26名の平均発話速度は、4.28 syll/sec、または 162 word/min であった。英語母語話者が面談をした際の発話速度として報告されている 218 word/min (Götz, 2013) よりも遅い速度である。平均構音速度は、4.71 syll/sec であり、Anderson-Hsieh and Venkatagiri の研究 (1994) で報告されている 5.0 syll/sec より遅いことがわかった。発話平均音節数 (Mean Length of Runs) は 8.4 syll/utterance であり、他の研究 (Götz, 2013) で示された 12.68 syllables よりも少ないことがわかる。これらの結果より、プレゼンテーションを行った英語母語話者はゆっくりと、頻繁にポーズをいれ、発話を短く区切りながら話す傾向があることが明らかになった。

センテンス内に生起するポーズ (100 ms 以上の無音区間、ただし無声破裂音前は 150 ms) はその長さもポーズを用いる人の割合も様々であるが、“First of all,” “Secondly,” などの文頭において順序を示す副詞句の後には、ほぼ全員の英語母語話者が文全体の 10-12% の長さにあたるポーズを用いていた。また場所や時を表す副詞句の後では 4-5% の長さにあたるポーズを用いる傾向が見られた。

ピッチについては、プレゼンテーションの内容が聞き手に悲しみなどの感情を呼び起こす可能性がある部分で平均ピッチの低下とピッチ幅の縮小などが観察され、内容に対応したピッチ変動について今後詳細な研究が必要であることがわかった。

2) 英語発音学習支援機能開発

開発するソフトウェアではまず、学習を行う英文を事前に用意し、プレゼンテーション時の音声特徴を付加したモデル音声ファイルとともに保存しておき、このお手本音声とマイクから取得した学習者の音声を比較することで発音の改善等の指導を行うことを目指した。モデル音声と学習者音声の比較を行うための音声処理として母音を推定し、音節持続時間を計測することで音声中のポーズや強調部分の判定が可能となる。

① 音節の認識

音節の自動認識においては、音節核のフォルマントに着目し、ケプストラム法により母音のフォルマントの抽出を行った。ただし、

あいまい母音とよばれるシュア音については試行した自動認識法では認識が困難であることがわかり、シュア音に関するより詳細な音声特徴調査や検討の必要性が明らかになった。

② 音声表示形式

本研究では、音声の特徴や改善点が視覚的に理解できるような表示方法として、ピアノロール形式 (図 1) を採用した。人間の音声を扱うため、縦の升目表示はピアノの音階ではなく、人が発する音の周波数帯域に合わせるようにした。しかし、この方法ではモデル音声と学習者のピッチ変動差を直感的に認識するのが困難だと考え、モデル音声はピアノロール式表示、学習者音声はピッチ波形で表示する方式を考案した。ただし、この方法でも、音声の強弱を表示できないという問題点があった。そこで、図 2 のようにピッチ波形に太さ・色の変化を付加し、音の強さと高さを表現する方法を考案した。この方法ならば、学習者が自分の音声の大きさやピッチを直感的に把握しやすいが、発話速度を上げるとピッチ波形の動きが見えにくくなる点があることがわかり、今後改善が必要である。

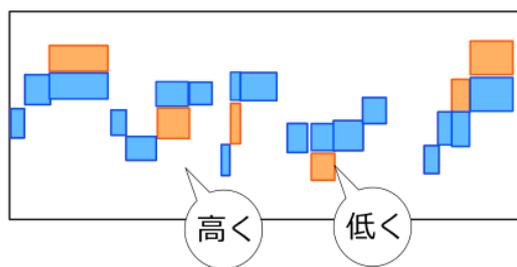


図 1 音声特徴のピアノロール表示

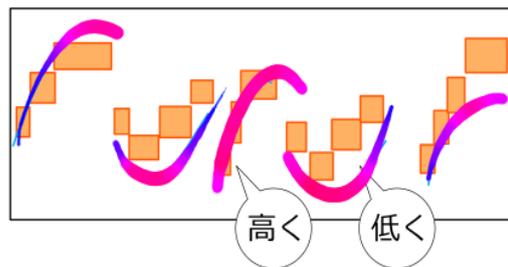


図 2 音声特徴のピッチ・音量波形表示

(後藤他,2015)

また、より視覚的にわかりやすい形式で学習できるようにするには、流れている音声に合わせて英文やコメントを表示することが有効であると考え、音声に合わせて英文を登録、再生が行えるシステムを開発した。英文の保存・再生に必要なファイルフォーマットは汎用性、可読性を考慮し、データ形式として広く用いられテキストファイルで構築できる Json 形式を用いることとした。試行の結果、音声の切れ目などを自動で検出しタイミングを登録できるような改良が必要であ

ることがわかった。

さらに、効果的なモデル音声を作成するため、既存のモデリング手法を利用してモデル音声の声質を学習者のものに近づけることを試みた。特に F0 の線形変換、混合正規分布モデルによるスペクトル包絡の統計的変換を行い、主観的評価を行った。結果としてある程度声質を近似させることができるものの、ノイズの発生や英語のプレゼンテーション時の音声特徴が声質変換によって失われたりすることが明らかになった。客観的な指標を用いた声質の変換精度評価や、変換によりモデル音声の重要な特徴が失われないような方法を検討する必要がある。最後に、学習者の発音をモデル音声と比較して評価するスコアリングシステムの採点アルゴリズムを検討した。ピッチによる採点を基礎点数とし、インテンシティの値を使い加点を行うことで評価を行う方法を作成した。

本研究では、以上のように効果的な英語プレゼンテーションに関係する英語音声面の特徴を抽出し、英語プレゼンテーションに特化した英語発音学習支援ソフトの開発を行った。今回明らかになった問題点を検討し、ソフトの完成度向上に向けて、ユーザテストやユーザからの意見の反映、また各機能の改善を進めていきたいと考える。

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)

[学会発表] (計 11 件)

- (1) Hori Tomoko, Yoshimoto Sadanobu, Kojima Tetsuya, and Noguchi Judy, “Acoustic Features of English Oral Presentations for Creating Speech Model Samples”, 外国語教育メディア学会, 2017 年 8 月 5-7 日, 名古屋学院大学 (愛知県、名古屋市)
- (2) 窪田智也、堀 智子、吉本定伸、小嶋徹也、英語プレゼン学習支援アプリのための採点アルゴリズムの改良、電子情報通信学会、2017 年 3 月 24 日、名城大学天白キャンパス (愛知県、名古屋市)
- (3) 窪田智也、堀 智子、吉本定伸、小嶋徹也、英語プレゼン学習支援アプリにおける採点機能の検討、教育システム情報学会、2017 年 1 月 7 日、愛媛大学 (愛媛県、松山市)
- (4) Hori Tomoko, “Exploring the use of Text-to-Speech software to create model speech sounds for English oral presentations”, Round Table Talk at Penn State University, 2016 年 11 月 11 日、ペンシルバニア州立大学、ステートカレッジ (アメリカ)
- (5) 後藤健太、西原悠貴、中平有樹、吉本定伸、小嶋徹也、堀 智子、鈴木幸一、英語プレゼンテーションに関する学習を支援するソフトウェアの開発、情報処理学会コンピュータと教育研究会、2016 年 2 月 13 日、東京農工大学 (東京都、小金井市)
- (6) 中平有樹、小嶋徹也、堀 智子、吉本定伸、鈴木幸一、声質変換技術に基づく英語プレゼンテーション支援ソフトウェアのためのモデル音声に関する検討、電子情報通信学会、2016 年 1 月 18 日、東北大学 (宮城県、仙台市)
- (7) 橋積裕紀、齋藤光、吉本定伸、堀智子、野口ジュディー・津多江、小嶋徹也、英語プレゼンテーション学習支援ソフトウェアの開発、電子情報通信学会総合大会、2015 年 3 月 15 日、立命館大学 (滋賀県、草津市)
- (8) 齋藤光、橋積裕紀、吉本定伸、堀智子、野口ジュディー・津多江、小嶋徹也、学生を対象とした英語プレゼンテーション学習支援ソフトウェアの開発、教育システム情報学会、2015 年 1 月 10 日、大阪産業大学 (大阪府、大東市)
- (9) 松永竜太郎、今井美和花、小嶋徹也・吉本定伸・堀 智子・野口ジュディー・津多江、英語プレゼンテーションのための音声学習支援ソフトウェアの開発、教育システム情報学会、2014 年 3 月 15 日、名古屋学院大学 (愛知県、名古屋市)
- (10) Hori Tomoko, Noguchi Judy, Yoshimoto Sadanobu, and Kojima Tetsuya, Acoustic Characteristics of English Oral Presentation Speech: Recorded vs. Synthesized sounds, 2014 年 2 月 22 日、電気通信大学 (東京都、調布市)
- (11) 松永竜太郎・小嶋徹也・吉本定伸・堀 智子・野口ジュディー・津多江、英語プレゼンテーション音声の特徴分析、電子情報通信学会、2014 年 1 月 27 日、東北大学 (宮城県、仙台市)

6. 研究組織

(1) 研究代表者

堀 智子 (HORI, Tomoko)
東京工業高等専門学校・一般教育科・教授
研究者番号：00269789

(2) 研究分担

小嶋 徹也 (KOJIMA, Tetsuya)
東京工業高等専門学校・情報工学科・教授
研究者番号：20293136

(3) 研究分担者

吉本 定伸 (YOSHIMOTO, Sadanobu)
東京工業高等専門学校・情報工学科・教授
研究者番号：00321406

(4) 研究分担者

野口ジュディー津多江
(NOGUCHI, Judy, Tsutae)
神戸学院大学・グローバル・コミュニケー
ション学部・教授
研究者番号：30351787