

修士学位論文要約（令和4年3月）

文の発音難易度を考慮した英語自動発音評価に基づく発音学習システムに関する研究

内城 沙月

指導教員：伊藤 彰則

A Study on Pronunciation Learning System Based on Automatic English Pronunciation Evaluation Using Sentence Pronunciation Difficulty

Satsuki NAIJO

Supervisor: Akinori ITO

A system for learning pronunciation is called CAPT system. Among these systems, an interactive CAPT system, which allows learners to learn pronunciation interactively with a computer. In conventional interactive CAPT systems, there has been insufficient analysis of how the system evaluates the learner's proficiency by the system's evaluation scores. The accuracy of the automatic English pronunciation evaluation used in the system also remains an issue. In this study, we first investigate an automatic English pronunciation evaluation method. In addition to the machine scores obtained from speech recognition systems, the pronunciation difficulty score obtained from the pronunciation difficulty estimation system is also used to improve the evaluation accuracy. We found that the correlation coefficient improved by up to 0.11. Then, we constructed an interactive CAPT system and we conducted experiments by varying the pronunciation difficulty of the sentences and analyzed trends in evaluation scores by the pronunciation difficulty. We revealed difficult sentences tends to get the evaluation score lower.

1. はじめに

近年の国際化に伴い、英語能力を習得する必要性が拡大している。人間の教師による英語指導の体力的・時間的限界を解決する方法の1つとして、英語の発音学習を行うためのシステムである Computer-Assisted Pronunciation Training (CAPT) システムがある。その中でも学習者とコンピュータが対話形式で発音学習ができる対話型 CAPT システムは、実際の教師を用意する必要がないため低コストであることに加えて、英語でのコミュニケーションを効率的に習得できるという利点がある¹⁾²⁾。

従来の対話型 CAPT システムに関する研究では、自動発音評価システムにより学習者の習熟度を評価し、その評価値に合わせたフィードバック/発話選択を行う。しかし、システム内で用いられる英語自動発音評価システムの評価精度が特に短い発話で十分でない、またシステム内で学習者の発音の上手さである習熟度が適切に測られていないという課題がある。そこで、本研究では新たに文の発音難易度推定システムを用いた英語自動発音評価システムを提案し、それを応用した対話型 CAPT システムを構築を行い、習熟度評価方法について検討するための発音評価値の分析を行う。

2. 文の発音難易度推定システム

本研究では、任意の文を入力としてその文の発音難易度スコアを推定するための発音難易度推定システムを提案する。文の発音難易度推定システムの概要を図

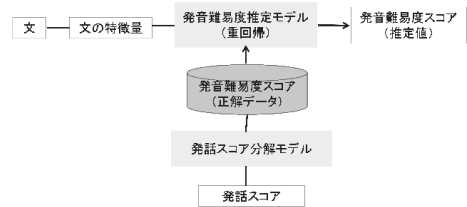


図1 文の発音難易度推定システム

1に示す。このシステムでは、文から文の特徴量を抽出し、その特徴量を入力とする発音難易度推定モデルにより発音難易度スコアを推定する。本研究では、この発音難易度推定モデルには重回帰分析を用いている。また、発音難易度スコアのデータベースを作成するために、発話スコア分解モデルを用いる。この発話スコア分解モデルは、人間により学習者の発話に付与されたスコアから発音難易度スコアを抽出する。また、文の特徴量としては音素の種類数・単語親密度・サプライヤル・調音特徴を用いる。

3. 文の発音難易度を考慮した英語自動発音評価システム

本章では、従来の自動発音評価システムを改良して、文の発音難易度を考慮した発音評価システムを提案する。提案法の概要を図2に示す。従来の英語自動発音評価システムとして Reference-free Error Rate (RER) を用いた発音評価システムを使用する³⁾。

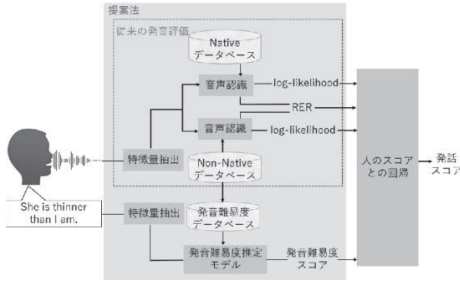


図2 文の発音難易度を考慮した英語自動発音評価システム

従来法では、nativeとnon-nativeの音声認識から得られた対数尤度と RER から人のスコアとの回帰を行い、機械スコアを得る。提案法では、この従来手法から得られる対数尤度・RER に加えて、学習者の発話する文から、前節で構築した発音難易度推定システムを使用して発音難易度スコアを推定し、これらを合わせて人のスコアとの回帰を行うことで最終的な機械スコアを得る。

実験では、発音難易度を考慮しない従来法と発音難易度を考慮した提案法で精度比較を行う。発音評価精度と発話の長さの関係を明らかにするために、

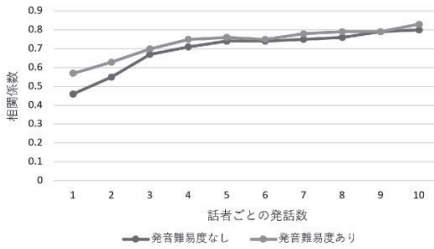


図3 機械と人間のスコアの相関係数

話者ごとに1発話から10発話までまとめて評価した場合の精度を測る。このときの人間のスコアと機械の

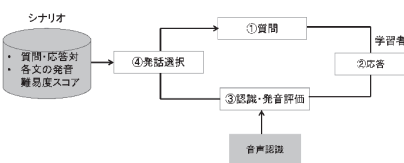


図4 学習者の習熟度を考慮した対話型CAPTシステム

スコアの相関係数を図3に示す。この結果から、短い発話のとき、発音難易度スコアを機械スコアに加えることで相関が高くなることが分かる。つまり、発音難易度を考慮することで機械スコアにおける評価の偏りを改善でき、結果として評価精度が向上することが明らかとなった。

4. 文の発音難易度を用いた対話型CAPTシステム
最後に、文の発音難易度を用いた対話型CAPT

システムを構築することで、習熟度評価方法について検討するための発音評価値(発話スコア)の分析を行う。構築した対話型CAPTシステムを図4に示す。このシステムでは、あらかじめいくつかのシナリオを用意し、各シナリオに対して、文の発音難易度推定システムを用いて発音難易度スコアを付与し、それを8つの難易度のクラスに分割する(シナリオスコア)。そして、特定の難易度のシナリオを提示し、システムと話者が交互に発話を行う。こうすることで、発音難易度ごとの発話スコアの傾向の分析を行う。

シナリオスコアごとの発話スコアの平均値を図5に示す。図5ではシナリオスコアが低く、発音が難しくなるほどすべての話者に共通して発話スコアの平均値が下がっていくことが確認できる。このことから、シナリオスコアから発話スコアの平均値が推定できれば、それらの推定値をさらに平均することで最終的に習熟度予測が可能であると考えられる。

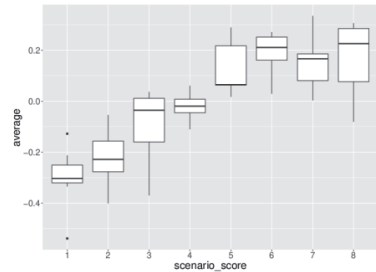


図5 発音何度ごとの発話スコア平均

5. まとめ

本研究では、対話型CAPTシステムに用いられる英語自動発音評価システムの精度向上と、システムにおける習熟度評価方法検討のためのシステム構築・発話スコア分析を行った。自動発音評価システムでは、文の発音難易度を考慮することで従来法よりも高い精度で評価が行えることがわかった。また、対話型CAPTシステムでは発音が難しくなると、すべての話者に共通して発音難易度が下がっていくことがわかった。

文献

- 1) F. Ehsani et al., "An interactive dialog system for Japanese," *Speech Communication*, 30(2-3), 167-177, 2000.
- 2) P. Su et al., "A recursive dialogue game for personalized computer-aided pronunciation training," *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, 23(1), 127-141, 2014.
- 3) J. Fu et al., "Automatic assessment of English proficiency for Japanese learners without reference sentence based on deep neural network acoustic models," *Speech Communication*, 116, 86-97, 2020.