

日本人の英語発表論文の発音評価に関する研究

中川 聖一 (豊橋技科大), 壇辻 正剛 (京大)

1 はじめに

国際化と高度情報化の到来を間近に控えた現代社会において、次代の我が国の発展を担うべき大学生に対する国民の要請は、国際社会での活躍である。しかしながら、日本人の英語力の低迷は国民の関心を引く大きな問題となっている。マスコミ紙上を賑わした P I S A (The Programme for International Student Assessment、学習到達度調査) では、日本はもはや最上位グループではないとの否定的な見解が展開されたのが記憶に新しい。P I S A は OECD (経済協力開発機構) が 2000 年から 3 年ごとに行っている調査であるが、日本では学力低下の象徴としてセンセーショナルに取り上げられている。しかし、国際間比較では、むしろ英語力の比較によく用いられる TOEFLR の平均スコアのほうがよりショッキングな結果を示している。

このような英語力に危機感を抱いた文部科学省も特色 G P や現代 G P、「英語が使える日本人」の育成のための行動計画、大学入学者選抜大学入試センター試験へのリスニング問題の導入など、いろいろな取組を通じて矢継ぎ早に手を打つようになった。また、大学においても、学部から大学院へのステップワイズな英語力の養成が一つの課題として浮上するようになった。そのような状況の中で、大学院レベルでの専門的な英語力の養成として、近年、注目を集めているのが、英語による国際会議での研究発表に対するコンピュータやマルチメディアなどの情報通信技術、すなわち I C T を利用した外国語習得の支援の試みである。

京都大学学術情報メディアセンター語学教育システム研究分野と豊橋技術科学大学メディア科学リサーチセンターでは、協力して、日本人の英語論文発表の発音評価に関する研究を共同研究として遂行している。

2 京都大学の試み

京都大学側では壇辻研究室の修士論文として提出された長瀬沙耶香「プレゼンテーションにおける発音音声の特徴と印象評価 - 生命科学専攻の日本人学生を対象とする E S P - 」[1] で得られた研究成果を基盤にして研究の展開をはかっている。文献 [1] では、生命科学に関する模擬国際会議を開催し、日本人学生のプレゼンテーション発表のデータ収集を行ない、日本人学生と英語教師の主観に基づく印象評価と、実

際の発話の韻律情報の特徴との関連を分析し、発話の韻律が印象評価に与える影響を検討した。

3 豊橋技科大の試み

3.1 英語教師による発音評価

音声データは Translanguage English Database (TED)[2] から日本人男性話者 16 名、ネイティブ 5 名の全 289 発話の音声を用いた。このデータベースには英語文の発音を、英語教師 5 人が 5 段階 (5 が最も良い) で評価したスコアが 1 話者ごとに付与されている [3]。英語教師間の評価の相関は 0.691 である。英語教師 4 人の平均スコアと残る 1 人のスコアとの相関は 0.791 である。

3.2 音響モデル

使用した音響モデルは、英語音素 HMM と日本人適応化英語音素 HMM、日本語音節 HMM の 3 種類である。英語音素 HMM は TIMIT コーパス、WSJ コーパスを用いて学習した 41 種類の HMM である。また、日本人適応化音素 HMM は、英語音素 HMM を日本人の英語音声データを用いて MAP 適応により追加学習したものである。日本語音節 HMM は ASJ データベースの音声試料と新聞記事読み上げ音声 JNAS を用いて学習した 116 種類の HMM である。

3.3 特徴量と英語教師の発音評価との相関

対数尤度: 英語音素 HMM による尤度は英語の発音が上手いほど、日本人適応化 HMM は典型的な日本人の発音に近いほど尤度が高くなると予想されるので対数尤度を特徴量として使用する。発声された英語文に対応する英語音素モデルを連結して求めた対数尤度をフレーム数で正規化したもの (LL_{native}) と英語教師の発音評価との相関、日本人適応化音素モデルを連結して求めた対数尤度をフレーム数で正規化したもの ($LL_{non-native}$) と英語教師の発音評価との相関を求めた。

尤度比: 対数尤度で述べた 2 つの正規化対数尤度を用い尤度比 R (実際は対数尤度の差) を求めた。

最適な音素列とのマッチング結果の尤度: 英語音素 HMM を用い音声を認識し、最適な音素列とのマッチング結果の尤度をフレーム数で正規化した尤度 (LL_{best}) を求め、英語教師の発音評価との相関を求めた。

事後確率: 英語音素モデルを使用した正規化した対数尤度 (LL_{native}) と最適な音素列とのマッチング

結果 (LL_{best}) の正規化した尤度の差で事後確率 L' を求めた。

音声認識結果: 最適な音素列とのマッチング結果を用いて置換率 ($Sub.$), 挿入率 ($Ins.$), 脱落率 ($Del.$), 正解率 ($Cor. = 1.0 - Sub. - Del.$), 精度 ($Acc. = 1.0 - Sub. - Ins. - Del.$) を求めた。

パワーとピッチの標準偏差: アクセントのつけ方が上手いほうがパワーやピッチの標準偏差の値が大きくなると考えられるので, パワーとピッチの標準偏差を特徴量として使用した。

話速: 音声の無音部分を取り除き, フレーム数を N_{frame} , 音素数を $N_{phoneme}$ とし, 話速 S を $N_{phoneme}/N_{frame}$ で定義した。

連続単語認識結果: 連続単語認識を行い, 最適な単語列とのマッチング結果を用いて正解率を求めた。音響モデルには英語音素 HMM, 言語モデルには TED コーパス, WSJ コーパスで学習したバイグラムモデル, 認識デコーダには本研究室で開発した SPOJUS を用いた。

日本語音節 HMM を用いた尤度比: 日本語音節 HMM を用い音声を認識し, 最適な音素列とのマッチング結果の尤度をフレーム数で正規化した尤度 (LL_{jap}) を求めたものと, 英語音素モデルを使用した正規化した対数尤度と最適な音素列とのマッチング結果 (LL_{best}) の正規化した尤度を用いた尤度比 (R_{jap}) を求めた ($LL_{best}-LL_{jap}$)。

日本人適応化 HMM を用いた尤度比: 日本語音節 HMM を用いた尤度比 (R_{jap}) 同様, 日本人適応化 HMM を用い, 最適な音素列とのマッチング結果の尤度 ($LL_{adapt-best}$) を求め, 尤度比 (R_{adapt}) を求めた ($LL_{best}-LL_{adapt-best}$)。

3.4 統計的発音評価法と結果

統計的な発音評価のために, 様々な音響的特徴量を組み合わせる方法として重回帰分析を用いた

特徴量の組み合わせは 1 文ごとでは LL_{native} , R , R_{jap} , パワーの標準偏差, 音素認識 (脱落率), 5 文ごとでは R , パワーの標準偏差, 単語認識率 (正解率, TED), 単語認識率 (正解率, WSJ), 10 文ごとでは LL_{native} , $LL_{non-native}$, R_{jap} , パワーの標準偏差, 音素認識 (脱落率) を用いてスコアを推定した。5 文ごと, 10 文ごとは, 例えば 5 文ごとの場合, 1~5, 2~6 のように 1 文ずつスライドしながらその文の特徴量の平均を取る手法を取った。評価結果を表 1 と表 2 に示す。

Table 1 特徴量と教師の評価との相関

特徴量	1 文ごと	5 文ごと	10 文ごと
LLbest	-0.473	-0.613	-0.660
LLnative	-0.466	-0.625	-0.669
LLnon-native	-0.638	-0.771	-0.804
R:尤度比	0.800	0.859	0.880
Radap	0.772	0.827	0.822
Rjap	0.719	0.804	0.811
音素認識 (正解率)	0.299	0.461	0.483
音素認識 (脱落率)	0.056	0.116	0.220
音素認識 (置換率)	-0.298	-0.567	-0.662
事後確率	0.214	0.273	0.349
単語認識率 (正解率, TED)	0.113	0.256	0.281
単語認識率 (正解率, WSJ)	0.102	0.163	0.261
パワーの標準偏差	-0.066	-0.057	-0.002
ピッチの標準偏差	0.495	0.638	0.691
話速	0.523	0.692	0.773

Table 2 推定スコアと教師の評価点の相関

	1 文ごと	5 文ごと	10 文ごと
CLOSED	0.851	0.902	0.918
話者 OPEN	0.804	0.866	0.887

4 おわりに

日本人の英語による国際会議発表の発音評価を, 英語教師の発音評価と相関が高い特徴量の組み合わせを統計的な手法を用いることで実現した。英語教師との相関の高かった特徴量を重回帰分析により統計的に組み合わせることで英語教師による発音の評価と相関が高いスコアを算出することができ, 英語教師に近い発音評価ができた。

参考文献

- [1] 長瀬沙耶香, “プレゼンテーションにおける発話音声の特徴と印象評価 - 生命科学専攻の日本人学生を対象とする E S P -”, 京都大学・人間・環境学研究所・修士論文, 2007
- [2] <http://www.elda.org/catalogue/en/speech/S0031.html>
- [3] 太田, 中川, “日本人の英語文発声の発音評価法”, 日本音響学会春季研究発表会, 3-8-12, 2007.3