

## 英語発音評価における各種正規化の検討

久住 大<sup>†1</sup> 加藤 正 治<sup>†1</sup>  
小坂 哲 夫<sup>†1</sup> 松 永 到<sup>†1,\*1</sup>

日本人学習者が発声した英語発音の自動評価における、性能改善について検討を行った。本研究では、河合らが提案した、2言語の音響モデルを用いて非母語発音の誤りを検出する方法に基づいた手法を用いる。音声の話者性または収録環境の違いにより、音響モデルと入力音声 mismatches を生ずる。そこで本研究では、ケプストラム平均正規化やヒストグラム同等化を用いることにより音響的特徴の mismatches を低減し、さらに音素に対する重み付けを行うことで日本人、米国人の相対的差異を保ちながら発音誤りの誤検出を低減する。

### Investigation of the various normalization methods in English pronunciation evaluation

MASARU KUSUMI,<sup>†1</sup> MASAHARU KATO,<sup>†1</sup>  
TETSUO KOSAKA<sup>†1</sup> and ITARU MATUNAGA<sup>†1,\*1</sup>

We investigated the performance improvement in automatic evaluation of English pronunciation uttered by Japanese students. The system is based on the error detection method of non-native pronunciation proposed by Kawai et al. in which the acoustic models of two languages are used. In this case, there is a problem of the acoustic mismatch between acoustic models and input speech because of speaker characteristics and/or recording environment. Then we employ the cepstral mean normalization (CMN) or the histogram equalization (HEQ) in order to minimize the acoustic mismatch. In addition, we try to reduce the miss detection of pronunciation error by using the phoneme weighting method while keeping the relative differences between two languages.

### 1. はじめに

我々は、日本人学習者による英語の発音に対して、音素ごと発音の良否を自動的に判断し、評価するシステムを開発している。これまで発音自動評価の研究は種々行われているが、河合らは2言語の音響モデルを用いた音声認識による発音誤りの検出法の提案を行っている<sup>1)</sup>。ここで、2言語の音響モデルとは、非母語話者と母語話者の音響モデルのことであり、日本人英語の発音評価の場合、日本人の日本語と英米人の英語音響モデルとなる。この場合、話者性や収録環境の違い等により入力とモデル間で mismatches が生ずる、対策として話者適応の利用が考えられるが、不正確な発音の英語音声で話者適応した場合、英語の音響モデルが不正確な発音を表現するモデルとなり、発音評価に悪影響を与える。話者適応として小笠原らによってパイリンガル発音データを用いた2言語混合音響モデルの話者適応方法の提案されている<sup>2)</sup>。両言語を正確に発音することが出来るパイリンガル話者の音声は話者適応に有用であるが、一般的に入手は困難である。そこで、本研究では、発音評価に悪影響を与えない正規化法について検討する。正規化法としては、音声認識で有効性が示されているケプストラム平均正規化法 (CMN) とヒストグラム同等化 (HEQ) を使用する。HEQ はもともと画像処理でよく使われている手法であり、近年、特徴正規化技術として音声処理に適応された<sup>3)4)</sup>。これにより、雑音下における音声認識の性能向上などが達成された。本研究ではこのヒストグラム同等化を用いることで、各種音響モデル間の分布の差を一致させることで性能向上を図る。また、CMN や HEQ を比較することで発音評価にどのような正規化が適切かを検討する。さらに、これまでの発音評価システムでは、英語話者でも日本語誤りが多く検出されてしまう問題があったので、本研究では、重み付け法を併用する。重み付け法では、日本語音素へ低い値の重みを与えることにより誤る確率を低減させる。この場合、過度な重み付けは、英語と日本語の評価に悪影響を与える危険性があるが、上記の正規化法を併用することにより、この悪影響の低減を図る。また、本評価システムでは、音素レベルでの間違いを検出するために、発音誤り規則を用いている。これまでの研究<sup>5)</sup>では、4つのカテゴリの誤り規則を用いてきたが、本研究では、8カテゴリに拡張し検討を行う。

†1 山形大学大学院理工学研究科

Graduate School of Science and Engineering, Yamagata University

\*1 現在、ビック東海

Presently with VIC TOKAI

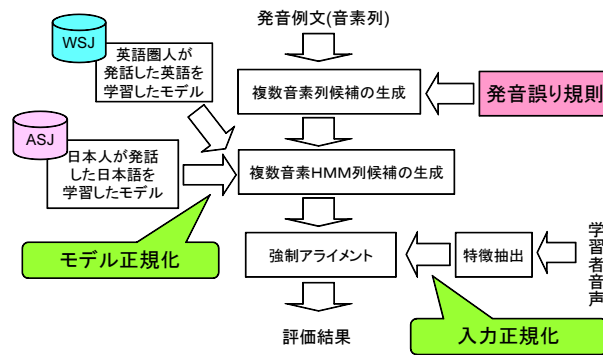


図 1 発音学習システムの基本構成

Fig. 1 Basic configuration of pronunciation learning system

## 2. 発音自動評定法

### 2.1 概要

本研究では、自動英語発音評定システムの開発を行っている。具体的なシステムのイメージは以下の通りである。まず、学習者に英語の文が提示され、それによって発声する。システムは発音全体の評価点、及び音素レベルでの発音がどの位置でどのように間違っているかを提示する。以上を実現するためには、学習者の発声に対し音素レベルでのアライメントを取ることが必要になる。アライメントをとる場合は英語圏人が発話した英語音響モデルと日本人が発話した日本語音響モデルを利用する。この時、音素レベルでの間違いを検出するために発音誤り規則を用いる。

システムの構成を図 1 に示す。学習者が英語を発声すると、システムは発音誤り規則に基づいて複数音素列候補の生成を行う。次にそれらと英語圏人が発話した英語と日本人が発話した日本語の 2 つの音響モデルから複数音素 HMM 列候補の生成を行う。そして、この生成した音素 HMM と学習者音声とを照らし合わせることで発音評定を行う。

本研究では、性能向上を目指し以下の検討を行う。各種正規化法及び重み付け法を併用することで、日本語音響モデルと英語音響モデルの mismatches の低減及び発音誤り割合の低減を目指す。また、発音誤り規則の詳細な検討を行う。従来 4 カテゴリーの誤りについて検討を行ってきたが、本研究では、さらに 4 カテゴリーを加えた計 8 カテゴリーについて、それぞれが発音評定とどのような相関を持つか検討を行う。

### 2.2 誤り規則

発音自動評定システムの基本構成として、非母国話者音素 HMM と母国話者音素 HMM から発音誤り規則に従って生成した複数音素 HMM 列候補と、評価音声とを強制アライメントをとることにより発音を評価する<sup>1)</sup>。このシステムでは、日本人学習者が犯しやすい発音誤りを規則化し、発声例文から発音誤りを含む複数音素列を生成する。これまで、以下の 4 規則の検討を行ってきた<sup>5)</sup>。以下は、挿入誤りの場合、太字の部分が入り込まれた日本語音素で、置換誤りの場合は、「英語音素 → 日本語音素」の置き換えがあった音素を太字で、脱落誤りは () を付けて太字で示す。

- 母音挿入 (単語末)・・・単語末の英語子音の後ろに日本語母音が入り込まれる規則  
 例 sing (s ih ng **u**)
- 母音置換・・・英語の母音が日本語の母音に置き換えられる規則  
 例 the (dh **ah** → dh **a**)
- 母音挿入 (子音間)・・・連続した子音間に日本語母音が入り込まれる規則  
 例 study (s **u** t ah d iy)
- 子音置換・・・英語の子音が日本語の子音に置き換えられる規則  
 例 child (ch ay **l** d → ch ay **r** d)

本研究では、さらに 4 規則の検討を行い、それらの妥当性の検討を行うことで発音評定精度の向上を目指す。その 4 規則を以下に示す。

- 子音脱落 (単語末)・・・単語末で母音の後ろの英語子音 /r/ が脱落する規則  
 例 far (f aa (**r**))
- 母音置換 (二重母音)・・・英語の二重母音 /ay/, /aw/, /oy/ が日本語の母音に置き換えられる規則  
 例 final (f **ay** n ah l → f **a** i n ah l)
- 子音挿入 (19 世紀借用語)・・・19 世紀までの借用語において発生する挿入誤り規則  
 例 extra (eh k i|**u** s t ah)
- 脱落誤り (単語始め)・・・単語の始まりで /w/, /y/ が脱落する規則  
 例 would ((**w**) uh d)

例として、図 2、図 3 に単語末の母音挿入誤り規則、二重母音置換誤り規則について図示した。○印はそれぞれ音素モデルを表しており、音素名の後に”\_J”と付加しているものが日本語音素である。本システムではこの日本語音素を検出した場合、発音誤りと判断する。

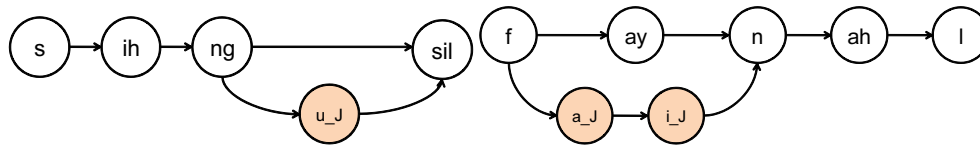


図 2 単語末母音挿入誤り:sing

Fig. 2 The vowel insertion error in the end of a word: sing

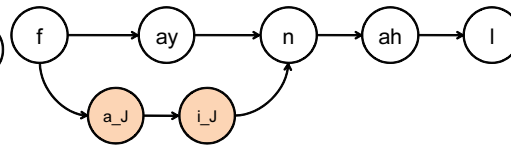


図 3 二重母音置換誤り:final

Fig. 3 Diphthong substitution error: final

### 3. 各種正規化及び出力尤度の重み付けの併用

上記発音評定システムでは、英語話者でも日本語誤りが多く検出されてしまうといった問題点があったが、日本語音素の出力尤度に低い値の重みを与えることで英語話者の日本語音素への誤りを減少させることができる。しかしながら、重み付けを行うことで日本語と英語の音響モデルの尤度差が減少し、発音評定性能が劣化する可能性がある。この問題を解決するために本研究では正規化法と重み付け法の併用について検討する。

#### 3.1 正規化法

音響モデルのミスマッチを改善するためには言語に影響されない話者性や収録環境の違いのみの差を適応する必要がある。しかし、一般的に話者性や収録環境の違いのみを抽出することは困難である。ここで各話者について、その音声のケプストラムの分布は言語によらず類似したものとなると仮定する。実際には、言語による差もあると考えられるが、収録環境や話者による差の方が、言語間の差よりも大きいことが観測される。そこで、モデル間の距離を測ることで両音響モデルがどのような関係になっているのかを検討した。モデル間の距離にはバタチャリア距離 (B 距離) を用いた。B 距離による音響モデル間の相互距離を表 1 の上段に記載する。表中の WSJ は Wall Street Journal (WSJ) データベースから学習した米国人の英語モデル、ASJ は日本音響学会の新聞記事読み上げコーパス (ASJ-JNAS) から学習した日本人の日本語モデル、J-E は日本人学生による読み上げ英語音声 (ERJ)<sup>8)</sup> データベースから学習した日本人英語のモデル、E-E は ERJ から学習した米国人英語のモデルをそれぞれ表す。表より正規化を行わない場合、米国人英語 (E-E) モデルは、WSJ から学習した米国人英語モデル (WSJ) よりも ASJ から学習した日本人の日本語モデル (ASJ) に近いことが分かる。また、日本人英語 (J-E) も同様である。すなわち、モデル間距離は言語の差よりもデータベースの差の影響を強く受けることが分かる。よりモデル間の様子を分かり易く示すため COSMOS 法<sup>6)</sup> を用いて可視化を行った。オリジナルの COSMOS 法とは

表 1 B 距離による各モデル間の相互距離

Table 1 B distance among acoustic models

正規化前	WSJ	ASJ	J-E	E-E
WSJ	0	24.21	12.11	25.80
ASJ		0	2.86	3.21
J-E			0	5.60
E-E				0
正規化後	WSJ	ASJ(DB)	J-E(DB)	E-E(DB)
WSJ	0	1.70	1.66	1.33
ASJ (DB)		0	0.29	0.19
J-E (DB)			0	0.34
E-E (DB)				0

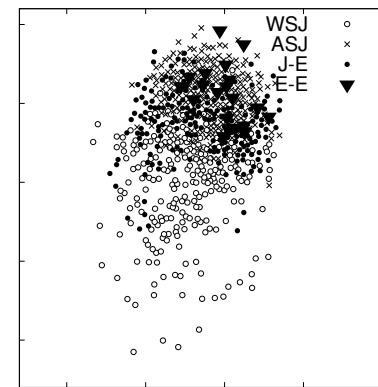


図 4 正規前のモデル間距離の可視化

Fig. 4 Visualization of the distance between models before normalization

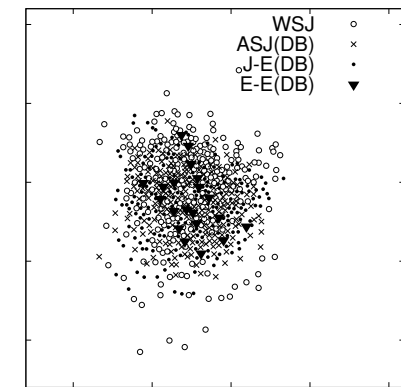


図 5 HEQ 正規後のモデル間距離の可視化

Fig. 5 Visualization of the distance between models after HEQ

異なり、話者間の距離には B 距離を用いた。この結果を図 4 に示す。図中の 1 点は、それぞれの話者を表す。図 4 の正規化前のプロットを見ると、WSJ と ASJ が大きく分かっているが、J-E、E-E とともに英語にも関わらず ASJ の分布に近い分布となっている。

以上の問題を解決するために各種正規化法について検討する。正規化法としては、音声データの特徴ベクトルを平均正規化するケプストラム平均正規化法 (CMN) や、ケプストラ

ムの分散が1になるように正規化を行うケプストラム平均分散正規化法 (CVN)<sup>9)</sup> があげられる。さらに、ヒストグラム同等化 (HEQ) があげられる。HEQ では評価データの累積密度関数 (CDF:Cumulative Density Function) が学習データの CDF と一致するように評価データのケプストラムを変換する。C<sub>E</sub> を評価データの CDF, C<sub>T</sub> を学習データの CDF とすると、入力 o<sub>t</sub> は以下のように変換される。

$$o'_t = HEQ(o_t) = C_T^{-1}(C_E(o_t)) \quad (1)$$

図5及び表1の下段に、CMN及びHEQを用いデータベース間の正規化を行った後の各モデル間の距離を示す。この図より、ASJの分布及びERJデータの分布がWSJに重なっていることが分かる。つまり、HEQにより収録環境の差がなくなったため、各モデルがWSJに似た分布になったと考えられる。以上により、正規化を用いることにより音響モデル間のミスマッチが大幅に低減できることが分かる。しかし、これらの正規化の言語差に対する影響は未知である。これを明らかにするため、本研究では、各種正規化のうち、CMN及びHEQについて発音評価に与える影響の検討を行う。

3.2 重み付け法

これまでの検討で英語話者でも日本語誤りが多く検出されてしまう問題があった。このため本研究では、重み付け法を併用し、誤りの低減を図る。出力尤度に対する重み付けは以下のように行う。b<sub>i</sub>(o<sub>t</sub>)を状態S<sub>i</sub>における出力確率、λを日本語音素に対する重みとすると、λ・b<sub>i</sub>(o<sub>t</sub>)を計算することにより重み付けする。λとして1以下の値を用いることにより、日本語音素の出力を抑制する。重み付けを行うことにより、英語と日本語のモデルの尤度差が減少し、評価性能の劣化が予想される。上記正規化法との併用により、この劣化が十分に抑えられるか検討する。

4. 実験条件

4.1 学習、評価データ

音声分析条件としては、窓幅/周期が32ms/8ms、特徴ベクトルが1~12次のMFCCと対数パワー、及びその1次と2次の回帰係数の計39次元とする。英語モデル用学習データにはWall Street Journal(WSJ)の米国人話者238名(男性119名、女性119名)の69,094文、日本語モデル用学習データには日本音響学会の新聞記事読み上げコーパス(ASJ-JNAS)の日本人話者204名(男女各102名)の31,511文を用いる。評価データには日本人学生による読み上げ英語音声データベース(ERJ)の日本人話者190名(男女各95名)の英語1900文、英語教師8名(男女各4名)の英語3215文、一般米国人12名(男性4名、女性8名)

表2 音素リスト(無音を除く)  
Table 2 Phoneme list (silence excepted)

日本語 34音素	a i u e o aa ii uu ee oo ei ou w y xy r h f z j s sh ch ts p t k b d g m n N cl
英語 39音素	aa ae ah ao ih iy uh uw ey eh er aw ay ow oy ch l m n ng b d dh f g hh p r s sh jh k t th v w y z zh

表3 英語教師間の発音評価の相関

Table 3 Correlation of the pronunciation evaluation between English teachers

教師の組み合わせ	相関
R1/R2	0.783
R1/R3	0.845
R1/R4	0.766
R2/R3	0.796
R2/R4	0.841
R3/R4	0.750
平均	0.797

の英語4827文を用いる。音響モデルは英語と日本語の2種類の音響モデルを併用し、それぞれ3状態16混合monophone HMMを用いる。この場合の音素の定義を表2に記載する。英語音素についてはCMU音素辞書を参考にした。<sup>7)</sup>

4.2 システムの評価法

日本人の英語発声には英語教師による5段階の評定ラベルが1文ごとに付与されている。このラベルと本研究における発音評価システムで検出された日本語音素への発音誤りの割合を比較して、システム評価を行う。本来であれば音素誤りの主観評価を用いてシステムの音素誤り検出の性能評価をするべきであるが、音素レベルの主観評価を行うのは困難である。このため、文ごとの評価結果との比較を行った。各文ごと4人の評定値が付与されているため、これを平均した値をその文の評定値として使用する。また、このラベルの精度を検討するため、教師間の評定にどの程度の相関があるかを調査した。4名の各教師が各日本人話者の発話10文に付与した評定を話者ごとに平均する。日本人話者は190名なので、190個の評定平均が各教師ごとに求まる。この評定平均について評価教師1対1で相関を取る。評価者をそれぞれR1~R4としたときの結果を表3に示す。表より、0.797の相関が得られることが分かる。

### 5. 実験結果および考察

提案する手法を用いた本発音自動評価システムの精度を評価するため、誤り検出割合と英語教師による評定平均の比較を行った。評価対象の日本人の発話数は1900文であるが、これを評定値順にソートし、評定の低い話者から50文ごと1クラスとし、この英語教師による評定平均とシステムにより検出された発音誤りの相関関係をみる。本研究では、正規化法として、ケプストラム平均正規化 (CMN) とヒストグラム同等化 (HEQ) を用いて実験を行った。CMN は、学習データ及び評価データに対して1文ごと正規化を行う。また、HEQ は、評価データ (ERJ)、日本語学習データ (ASJ-JNAS) 及び英語学習データ (WSJ) 全体のヒストグラムをそれぞれ作成し、評価データと日本語学習データの分布が英語学習データの分布に近づくよう正規化した。

表4に母音挿入・置換、子音置換誤り規則における各種正規化別の相関係数を示す。実験では、非正規化、CMN 及び CMN+HEQ について重み付けを行った場合と行わなかった場合の検討を行った。重み付けなしで正規化を行った場合、英語話者誤り平均が低下し、且つ相関の絶対値が低下しない結果となった。すなわち、発音評定の性能低下なしに、英語話者の日本語音素への誤りを減少できた。相関係数の絶対値が一番高くなったのは CMN+HEQ を行った時で-0.709であった。システムの評価としては米国人の誤り割合が低い方が良いので、重み付けを併用した実験を各種正規化において行った。重みをつけずに実験を行った場合、非正規化において米国人の誤り割合の平均は36.34%のため、重み付けによる誤り割合の減少が可能であることが考えられる。ここで、重みは米国人の誤り割合が5%程度になるように調整した。非正規化、CMN のみでは重み  $10^{-6}$  (母音置換のみ  $10^{-4}$ )、CMN+HEQ では重み  $10^{-2}$  を付加させた時、英語話者誤り平均が5%以下に抑えることができた。この場合、非正規化あるいは CMN のみでは相関の絶対値が低下しており、システムの性能が劣化することが分かる。一方、HEQ を併用する場合は相関の絶対値の低下は見られなかった。よって CMN+HEQ と重み付けを併用することにより、システムの性能を保ったまま英語話者の誤り平均を低下させることが可能であることが分かった。

各規則における結果の例を図6、図7に示す。図中では CMN を行った場合 (□)、CMN+HEQ を行った場合 (◆) の結果を示す。なお、両方とも日本語音素への重み付けを併用している。横軸が英語教師による評定平均で縦軸がシステムにより検出された発音誤りの割合である。両者の相関の絶対値が大きいほど、システムの性能が良いということになる。評定の低い話者から50文ごと1クラスとし、この評定平均を求めプロットした。

表4 母音挿入・置換、子音置換誤り規則における各種正規化別の相関係数

Table 4 The correlation coefficients according to various normalization in vowel insertion and substitution, and consonant substitution

実験条件	重み付け	非正規化		CMN		CMN+HEQ	
		○	○	○	○	○	○
相関係数	母音挿入 (単語末)	-0.791	-0.553	-0.779	-0.557	-0.802	-0.861
	母音置換	-0.058	0.134	-0.448	-0.281	-0.375	-0.335
	母音挿入 (子音間)	-0.667	-0.363	-0.646	-0.468	-0.804	-0.721
	子音置換	-0.787	-0.713	-0.846	-0.702	-0.857	-0.822
	平均	-0.576	-0.374	-0.680	-0.502	-0.709	-0.685
日本人話者誤り平均 (%)		63.19	16.20	61.69	8.83	50.69	18.38
米国人話者誤り平均 (%)		36.34	5.65	32.32	2.47	18.50	3.35

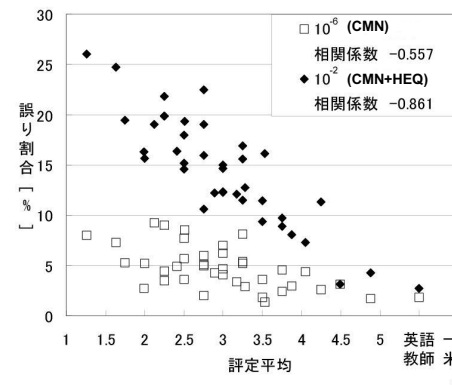


図6 単語末の母音挿入

Fig. 6 Vowel insertion in the end of a word

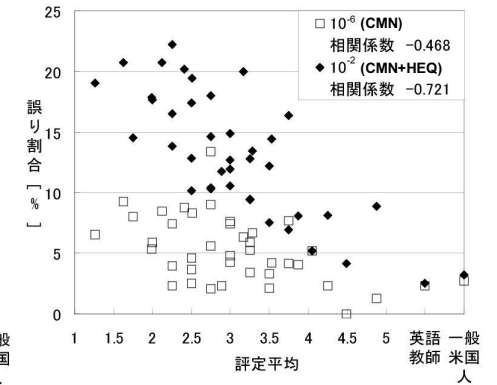


図7 子音間の母音挿入

Fig. 7 Vowel insertion between consonants

よって  $1900/50 = 38$  点がプロットされている。縦軸については、50文中実際に検出された日本語音素への発音誤り数を、発音誤りが起こる可能性のある数で割ることにより割合を算出した。また参考として英語教師8名および、一般米国人12名についてもプロットした。CMNのみを用いた場合 (□)、評定平均と誤り割合の相関が低下しており、発音評定性能が劣化することがわかる。さらに、図7のように日本人の方が米国人話者よりも誤り割合が低い場合も発生し、問題がある。一方 HEQ を用いた場合 (◆)、相関の低下は見られず、HEQ が有効であることが分かる。以上のように CMN に重み付けを併用する手法では、米

表 5 母音置換, 子音挿入, 脱落誤り規則における各種正規化別の誤り割合及び日英語者間の相対的差異 (%)  
Table 5 The error rates of vowel substitution, consonant insertion, and consonant deletion in various normalization, and the relative differences between Japanese and English speakers(%)

話者	規則名\重み付け	非正規化		CMN		CMN+HEQ	
		○	○	○	○	○	○
日本人話者	脱落誤り (単語末)	67.33	29.50	57.62	27.13	63.56	18.61
	母音置換 (二重母音)	60.12	36.77	54.63	28.51	47.49	8.99
	子音挿入 (借用語)	69.23	35.00	65.22	33.33	65.38	15.00
	脱落誤り (単語始め)	31.95	10.12	26.56	9.21	24.23	7.97
	平均	50.86	17.63	44.75	21.17	41.50	10.15
米国人話者	脱落誤り (単語末)	43.75	19.87	38.89	17.76	43.57	13.58
	母音置換 (二重母音)	42.68	18.20	38.46	20.92	34.62	5.50
	子音挿入 (借用語)	47.59	14.97	36.36	13.90	26.74	5.35
	脱落誤り (単語始め)	25.00	8.64	20.67	7.63	19.08	5.77
	平均	36.69	13.15	32.22	15.59	30.53	6.95
日英語者間の相対的差異 (%)		27.87	25.44	28.01	26.33	26.44	31.52

国人の誤り割合を抑えることができても, 日本人学習者の相関までは維持できず, HEQ を行う必要があることがわかる。

本研究ではさらに, 二重母音置換, 19 世紀借用語での子音挿入, 単語始め及び単語末の脱落誤り規則の計 4 カテゴリーの誤り規則についての検討を行った。表 5 に新たに追加した 4 カテゴリーでの各種正規化, 重み付け法の有無による日本人話者及び米国人話者の誤り割合を示す。また, 日本人話者, 米国人話者それぞれの 4 規則の誤り割合の平均を算出し, 両者の平均の誤り割合の相対的差異を示している。これらの規則は, 評価データ内で出現する回数が少なく, 50 文ごとのクラスで相関を検討する方法では信頼性の高い評価ができない。このため, 日本人話者全体と米国人話者全体の誤り割合の相対的差異に着目した。非正規化, CMN, CMN+HEQ はそれぞれ 26%~ 28%程度 of 相関的差異がある。しかし, これらの規則でも米国人話者の誤り割合は 30%を超えるため, 重み付け法による誤り割合の減少が必要であると考えられる。そこで, 相関的差異をある程度保ちながら重み付けを行う。表 5 より, 非正規化, CMN のみでは, 日英語者間の相関的差異を保ちながら重み付けを行ってもそれぞれ米国人話者で 13.15%, 15.59%の誤り割合が生じてしまう。一方, CMN+HEQ を行うことで, 米国人話者で 6.95%の誤り割合, 日英語者間の相関的差異は 31.52%と重みを付ける前よりも両者の差を付けつつ誤り割合を抑えることができた。この点からも CMN+HEQ と重み付けの併用が, CMN と重み付けの併用よりも性能が高いことが分かる。

## 6. ま と め

本研究では, 音声の話者性や収録環境の違いによる音響モデルや入力音声のミスマッチを改善するために, 各種正規化の検討を行った。正規化には, ケプストラム平均正規化やヒストグラム同等化を用い, 評価音声と音響モデルの音響的特徴を近づけることで, 発音評定性能を維持しながら音素誤りの誤検出を低減することを検討した。また, 出力尤度の重み付けを併用することで米国人の日本語発音への誤りについての更なる削減を検討した。これまで検討されてきた 4 規則については, CMN+HEQ に重み付けを行うことが, 発音評定性能を維持しながら誤り割合を低減するのに最も効果があった。本研究では, さらに 4 カテゴリーを拡張し, 計 8 カテゴリーについて検討した。拡張した 4 規則においても, 各種正規化に重み付け法を併用した実験を行い, 同様の検討を行った。この 4 規則についても, CMN と重み付けの併用よりも CMN+HEQ と重み付けの併用を行う方が性能が高く, 日米話者間の相対的差異を維持しながら誤り割合を抑えることに成功した。以上より, CMN+HEQ と重み付け法を併用することで発音評価システムの発音評価が高精度に行えることが分かった。

## 参 考 文 献

- 1) 河合剛, 石田朗, 広瀬啓吉: 2 言語の音響モデルを用いた音声認識による非母語発音誤りの検出と発音評価, 音響学会誌, Vol.57, No.9, pp. 569-580 (2001).
- 2) 小笠原洋一, 大河雄一, 鈴木基之, 伊藤彰則, 牧野正三: バイリンガル話者音声に基づく二言語混合音響モデルの話者適応法の検討, 信学技報, SP2003-127(2003)
- 3) A. Torre, J. C. Segura: non-linear transformations of the feature space for robust speech recognition, Proc. of ICASSP 2002, pp.401-404 (2002).
- 4) Yasunari Obuchi: 頑健な音声認識のためのデルタケプストラム正規化, Advanced Research Laboratory, Proceeding of ICA2004, pp.2587-2590 (2004).
- 5) 小坂哲夫, 松永到, 加藤正治: ヒストグラム同等化を用いた英語発音自動評定の性能改善, 日本音響学会講演論文集, 1-R-11, pp.405-408 (2009).
- 6) 庄境, 奈木野: 多次元尺度構成法による音響空間の 2 次元可視化, 情処研究報告, 2004-SLP-052-23, pp. 129-136 (2004).
- 7) The CMU Pronouncing Dictionary, <http://www.speech.cs.cmu.edu/cgi-bin/cmudict?in=C+M+U+Dictionary>
- 8) 峯松 信明: 外国語学習用読み上げ音声データベース, 日本音響学会誌 59 巻 6 号, pp.345-350 (2003).
- 9) Olli Viikki, Kari Laurila: Cepstral domain segmental feature vector normalization for noise robust speech recognition. Speech Communication, 25:133-147, (1998).