

発音の構造的表象に基づく母音矯正度推定の高精度化*

◎鎌田圭[†], 朝川智[†], 峯松信明[†], △牧野武彦[‡], 広瀬啓吉[†] ([†] 東大, [‡] 中央大)

1 はじめに

音声は、話者や音響機器といった不可避な非言語的特徴により歪み、同じ発話内容でも観測される音響事象は異なる。従来の音声工学システムでは、この音声の歪みを内包した特徴量を利用していたため、話者の違いによる不安定性を不可避であった。

近年提案された「音響的普遍構造」では、このような音声の歪みを避けるために音声事象の相対的な関係に着眼している [1]。この表象では、ある音声事象は他の音声事象との距離によって規定され、音声事象が持つ物理特性は参照しない。この新しい表象を応用することにより、英語学習者の発音状態を非言語的特徴の影響を受けずに記述可能である [2]。

この表象に基づいて、母音矯正順序の自動推定が提案されている [2]。この自動推定では、母音発音自体を評価した値を用いず、他の母音発音との距離を用いて母音矯正の必要度を推定するという特殊な手法が用いられている。この手法については母音図を用いた妥当性の検証がなされている [3]。

本稿では、この母音矯正度推定の妥当性を改善する母音矯正度推定手法について述べる。まず、先行研究である音声の構造的表象とそれに基づく母音矯正度推定、そして母音矯正度推定の妥当性の検証について紹介する。その後、この妥当性を改善する手法を提案する。

2 音声の構造的表象

音声に不可避に混入する非言語的特徴は、乗算性歪みと線形変換性歪みの2種類に分類される。乗算性歪みとしては、マイクロフォンなどの伝送特性や話者の声道形状の違いの一部が挙げられ、ケプストラムベクトル c に対するベクトル b の加算 $c' = c + b$ で表現される [4]。線形変換性歪みとしては、声道長の差異や聴覚特性の差異が挙げられ、 c に対する行列 A の乗算 $c' = Ac$ に相当する [5][6]。以上より、音声に不可避的に混入する非言語的特徴は、ケプストラムベクトル c に対して $c' = Ac + b$ としてアフィン変換でモデル化することができる。

このアフィン歪みを踏まえ、音声事象を分布として捉え、事象間の距離としてアフィン不変であるバタチャリヤ距離を採用すると、事象間距離は話者や音響

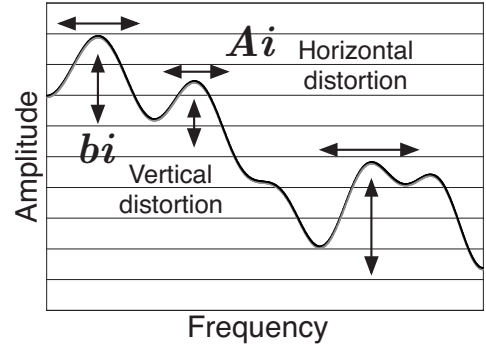


Fig. 1 スペクトルに対する水平/垂直方向の音響歪み

機器に対して不変量となる。音声事象群すべてに対して事象間距離をバタチャリヤ距離で取ることにより（即ち距離行列を求める）、話者不変の構造を抽出することが可能である。この構造を音響的普遍構造と呼ぶ。

3 構造的表象に基づく母音矯正度推定

構造的表象を用いて外国語学習者の母音矯正度推定をすることを考える。話者不変である構造的表象は、話者情報を音声から消去した上で、発音の正しさを評価することが可能である。

母音矯正度を推定するにあたって、学習者の手本となる教師の母音発音を用意する。学習者の母音群、教師の母音群、それぞれで話者不変の母音構造を抽出する。ここで問題となるのが、学習者の母音と教師の母音との距離を取って母音矯正度とすることが不可能ということである。学習者の母音は学習者自身の他の母音との距離のみによって表象されており、教師の母音に関する情報は一切持たないのである。

そこで、教師、学習者間における母音構造歪みへの各母音の影響度を持って、母音矯正度とする [2]。母音構造歪みは学習者が教師の母音発音群に到達するまでの距離と捉えられる。学習者の母音構造を S 、教師の母音構造を T としたとき (S と T は距離行列である)、 S と T の全体的な構造歪みは $D(S, T) = \sqrt{\frac{1}{m} \sum_{i < j} (s_{ij} - t_{ij})^2}$ で表される。 m は母音数であり、 s_{ij} , t_{ij} は距離行列 S , T の要素である。全体的な母音構造歪みに対する母音 v の影響は

$$d(S, T, v) = \sum_{j=1}^m |s_{vj} - t_{vj}| \quad (1)$$

*Improvement of automatic estimation of vowel correction priority for language learners based on the structural representation of pronunciation. by K.Kamata, S.Asakawa, N.Minematsu, T.Makino, and K.Hirose

[illegible]

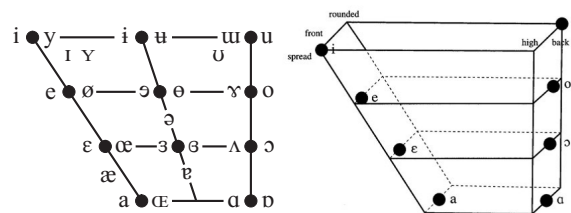
Table 2 日本語母音・米語母音の置換表

日本語母音	↔	米語母音
/あ/		/æ/, /ʌ/, /ɑ/, /ɔ/, /ə/
/い/		/i/, /ɪ/
/う/		/u/, /ʊ/
/え/		/ɛ/
/お/		/ɔ/

4 母音矯正度推定の妥当性の検証

日本語・英語双方が話せる日本人話者 12 名（男性 6 名，女性 6 名）の米語 11 単母音と日本語 5 母音を収録した。収録では米語 /bVt/, 日本語 /bVto/ を 5 回発声させ，母音部分のみを切り出して実験に用いた。米語 /bVt/ に有意味語が存在しない場合は，/b/ を /p/，/t/ を /d/ とするなどの処置を行った。

このようにして作成された 96 人 (12 話者×8 状態) の米語母音発音群を母音図を用いて評価した。1,056 個の母音を英語音声学の専門家 (第四著者) に聴取させ、母音図にプロットした。各母音は合計 4 つの評価軸で評価した。母音図にプロットすることにより、発音時の舌の前後・高低の 2 つの軸で評価し、この他に、円唇性と r 音性についても評価した。円唇性は無



円唇, 弱い円唇, 強い円唇の3段階で, r音声はr音声あり(+), r音声なし(-)の2段階で評価した。

4.2 絶対的評価と相対的評価の相関分析

一つは母音発音を表す座標値を用いて、教師の母音と学習者の母音の距離を直接測り母音矯正度とする。教師の母音図と学習者の母音図を母音図の軸（棒）を重ねて比較し、母音矯正度を推定することになる。本稿では、この方法で推定された母音矯正度を母音図の軸（棒）という絶対的な基準に基づいてることから、「絶対的評価」と称する。

もう一つは、構造的表象と同様の手法で得る母音矯正度である。これは母音図の軸（枠）が消失し、母音だけがプロットされた状態で、二つの母音群を比較し母音矯正度を推定する問題となる。この場合、Fig.4のように構造を回転&シフトさせ、二つの構造を重ね合わせた後に、母音矯正度を推定することとなる。母音図上の母音発音群から母音間距離行列を作成し、式(1)を適用して母音矯正度を得る。本稿では、こ

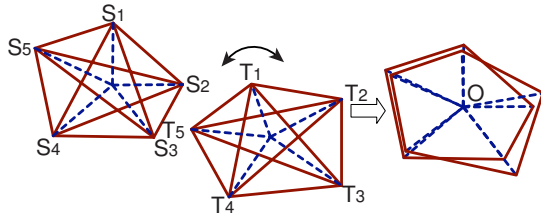


Fig. 4 回転とシフトによる構造の重ね合わせ

の母音矯正度を母音間の相対的な関係に基づいてることから、「相対的評価」と称する。

絶対的評価と相対的評価の相関分析を行うことで、構造的表象における母音矯正度の推定式 (1) の妥当性について検討する。絶対的評価を正解データとし、推定式 (1) から得られた相対的評価を相関係数によって評価する。相関分析を行う組み合わせは以下のとおりである。まず、ある話者の P8 (米語教師相当の母音発音) と、同一話者の P1 (日本人米語学習者の母音発音) を比較し、得られた二つの母音矯正度 (絶対的評価と相対的評価) を相関分析する (12 組 $\times$ 11 母音)。同様に、P8 と同一話者の P2, P3, ..., P7 を比較し、相関分析を行う。最後に、教師も学習者も限定せず、96 人すべてを教師・学習者として相関分析を行う (${}_{96}C_2 \times 11$ 母音)。

相関分析の結果をまとめたものが Fig.5 である。“96 $\times$ 96” は 96 人すべてを教師・学習者とした場合を表す。96 人総当たりの場合の相関係数は 0.772 であり、強い相関が確認された。この事実から、式 (1) は一定の信頼性があると考えられる。また、P8 を教師とした場合の相関係数について見ると、P2~P7 は 0.750 以上の高い水準を示しているが、P1 は 0.677 と他よりも弱い相関となった。式 (1) は母音 v 以外のすべての母音が信頼できるとの仮定を置いており、すべての母音が日本語母音である P1 はこの仮定から大きく逸脱するため、低い水準になったと考えられる。

5 母音矯正度推定の妥当性の改善

5.1 母音への重み付けによる母音矯正度の再推定

母音矯正度を測る式 (1) は、すべての母音を等価に扱っている。これは式 (1) を行列の演算に書き換えた式 (2) を見れば明らかである。

$$\begin{pmatrix} d_1 \\ d_2 \\ \vdots \\ d_m \end{pmatrix} = \begin{pmatrix} 0 & \delta_{12} & \dots & \delta_{1m} \\ \delta_{21} & 0 & & \delta_{2m} \\ \vdots & & \ddots & \vdots \\ \delta_{m1} & \delta_{m2} & \dots & 0 \end{pmatrix} \begin{pmatrix} 1 \\ 1 \\ \vdots \\ 1 \end{pmatrix} \quad (2)$$

ここで、 d_v は構造 S と構造 T を比較したときの要素 v の母音矯正度を表す。また、 δ_{ij} は、母音 i と母音 j の母音間距離の差 ($\delta_{ij} = |s_{ij} - t_{ij}|$) である。

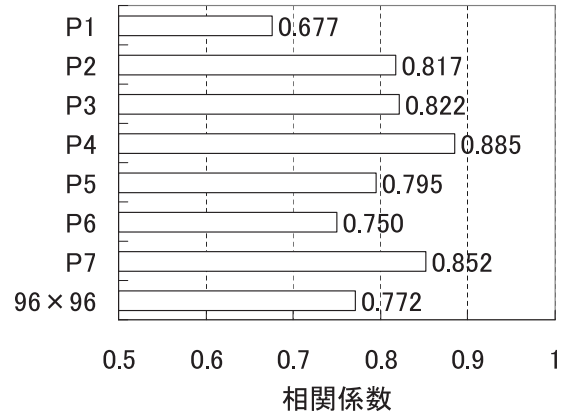


Fig. 5 絶対的評価と相対的評価の相関係数

式 (2) において、すべての要素が 1 からなる列が存在するが、この列は各母音の母音矯正度への影響度、すなわち重みと考えられる。式 (2) では、すべての母音の重みが 1 となっているが、母音発音の得手不得手に関わらず同じ重みを持たせることは不自然である。そこで、一度推定した母音矯正度を重みの導出に利用することを考える。重み w_v を求める式を次式のように定める。

$$w_v = c \times \frac{1}{d_v} \quad (3)$$

c は重み列正規化係数 ($\sum_{i=1}^m w_i = m$) である。式 (3) によって求められた重みを用いて、母音矯正度 d'_v を再推定するには次式を用いる。

$$\begin{pmatrix} d'_1 \\ d'_2 \\ \vdots \\ d'_m \end{pmatrix} = \begin{pmatrix} 0 & \delta_{12} & \dots & \delta_{1m} \\ \delta_{21} & 0 & & \delta_{2m} \\ \vdots & & \ddots & \vdots \\ \delta_{m1} & \delta_{m2} & \dots & 0 \end{pmatrix} \begin{pmatrix} w_1 \\ w_2 \\ \vdots \\ w_m \end{pmatrix} \quad (4)$$

この計算式の、 d_v の代わりに再推定された d'_v を用いて、重み w'_v を得ることにより、 d'_v を再々推定することも可能である。式 (2) で推定された要素 v の構造歪み d_v を相対的評価としたことに倣って、式 (4) で再推定された母音矯正度 d'_v を「重み付き相対的評価」と本稿では称する。

5.2 絶対的評価と重み付き相対的評価の相関分析

4.2 と同様に、P8 と P1~P7 の組み合わせでの相関分析と、96 人総当たりの場合での相関分析を行う。求められた相関係数をまとめたものが Fig.6 である。

96 人総当たりのときの絶対的評価と重み付き相対的評価の相関係数は 0.807 となり、相対的評価の相関係数 0.772 と比べ、3.5 ポイント改善した。また、P8 と P1, P2, ..., P7 を比較して相関分析をした場合でも、すべての場合で改善を示した。これにより式 (4) の有効性が示されたと言える。

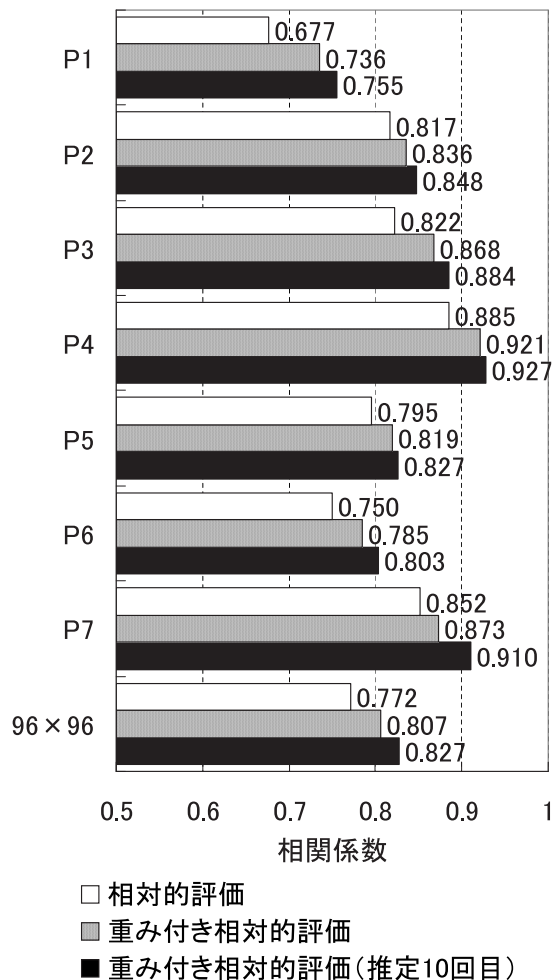


Fig. 6 絶対的評価と重み付き相対的評価の相関係数

次に、式 (3) の相対的評価 d_v の代わりに重み付き相対的評価 d'_v を適用して母音矯正度を再々推定した場合の相関係数を調べる。この重み付き相対的評価の再々推定を行った結果が Fig.7 である。ここでは、再々推定で得られた d'_v を利用して d''_v を推定する、というように再帰的に 10 回推定した。計算回数 1 回は従来手法による相対的評価、計算回数 2 回は前述の重み付き相対的評価、計算回数 3 回目以降が新たに推定された重み付き相対的評価である。結果はさらなる改善を示したものの相関係数の伸びは鈍った。そして、10 回ほどの計算で相関係数は収束した。10 回推定したときの相関係数についてまとめたものが、Fig.6 である。96 人総当たりのときの絶対的評価と重み付き相対的評価 (推定 10 回目) の相関係数は 0.827 となり、相対的評価と比較して 5.5 ポイント、重み付き相対的評価 (推定 2 回目) と比較しても 2.0 ポイントの改善を示した。また、P8 と P1, P2, ..., P7 を比較して相関分析をした場合でも、P1 が 0.750 を越えるなど、すべての場合で改善を示した。

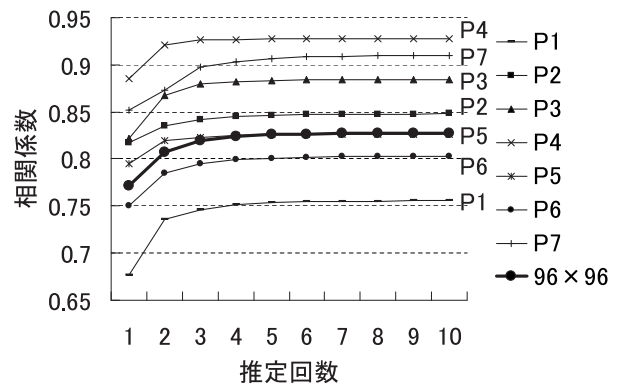


Fig. 7 推定に伴う相関係数の推移

6 おわりに

本稿では、提案されていた音声の構造的表象に基づく母音矯正度の推定手法について、示されていた妥当性を改善する手法を提案した。母音矯正度を考慮し、各母音に重み付けを行って母音矯正度を再推定した結果、妥当性を示す相関係数は 0.702 から 0.807 に上昇した。また、再推定された母音矯正度を用いて母音矯正度を再々推定する、といった再帰計算を行うことにより、相関が一層強くなることが確認された (相関係数 0.827)。

今後は、二重母音や子音、連続発声への拡張について検討する予定である。

参考文献

- [1] 峯松信明, 西村多寿子, 西成活裕, 櫻庭京子, “構造不変の定理とそれに基づく音声ゲシュタルトの推定,” 信学技報, SP2005-12, 1-8, 2005.
- [2] 朝川智, 峯松信明, 広瀬啓吉, “発音の構造的表象に基づく学習者分類と発音教生生成,” 信学技報, SP2005-156, 37-42, 2006.
- [3] 朝川智, 鎌田圭, 峯松信明, 牧野武彦, 広瀬啓吉, “米語母音発音に対する絶対的評価と相対的評価” 日本音響学会春季講演論文集, 2-2-1, 373-374, 2007.
- [4] D.Reynolds and L.P.Heck, “Speaker verification: from research to reality,” Proc. Int. Conf. Acoustics, Speech, and Signal Processing, tutorial session, 2001.
- [5] M. Pitz and H. Ney, “Vocal tract normalization equals linear transformation in cepstral space,” IEEE Trans. Speech and Audio Processing, vol.13, 930-944, 2005.
- [6] 江森正, 篠田浩一, “音声認識のための高速最尤推定を用いた声道長正規化,” 信学論 (D-II), vol.J83-D-II, no.11, 2000.