

## 構造的特徴量を用いた CALL システムの開発\*

☆鈴木雅之, 峯松 信明, 広瀬 啓吉 (東大)

## 1 はじめに

2011 年度より, 小学 5・6 年生に対する英語の「話す／聞く」活動が必修化される. また近年, 日本の企業において英語を公用語化とする動きも始まった. このような時代背景から, Computer Aided Language Learning (CALL) システムのニーズは今後高まっていくと考えられる.

CALL システムのうち, 発音学習を行うシステムを考える. このようなシステムを構築するためには, 学習者が目標とすべき発音を多く集め, 音響モデルを作成する必要がある. また, できるだけ声質の近い話者の音声を集めることが望ましい [1]. しかしながら, 学習者によって目標とすべき発音は異なっており, 学習者のニーズに合うように音響モデル作成用のデータを用意するのは非常に難しい. また, 声質の近い話者を集めることは, 子供を対象とする場合に特に難しい [2].

筆者らはこの問題を, 話者の違いに近似的に不変な音声の構造的表象を利用することで解決を図っている [3]. 構造的表象を用いることで, ある学習者とある教師の発音を, 一対一で比較することが可能になる. 既に, 構造を用いた自動発音評定や話者分類に関する検討を行ない, その有効性を示してきた [4].

本論文では, 音素ごとに評定値を推定する手法について, 手法の改善及び評価を行ったことを報告する. さらに, これらの成果をデモンストレーションシステムとして実装したことも報告する.

## 2 音声の構造的表象

音声の構造的表象の概念図を Fig. 1 に示す. ある二つの分布に任意の一対一対応変換を施しても, その分布間の  $f$ -divergence は常に一定となることが示されている [5]. 音声の構造的表象とは, ある話者の発声を音素などの単位でケプストラム空間上で分布化し, それらすべての分布間の  $f$ -divergence を計算することで得られる距離行列 (幾何学的には構造) のことである. ケプストラム空間上で, 話者の違いはある一対一対応変換でおよそ近似できるため, 構造的表象は話者の違いに近似的に不変となる.

構造的表象を用いて発音評価を行なうには, 学習者の構造と教師の構造間の差異を見ればよい. この手法として我々は, 多段階重回帰分析を提案している [4].

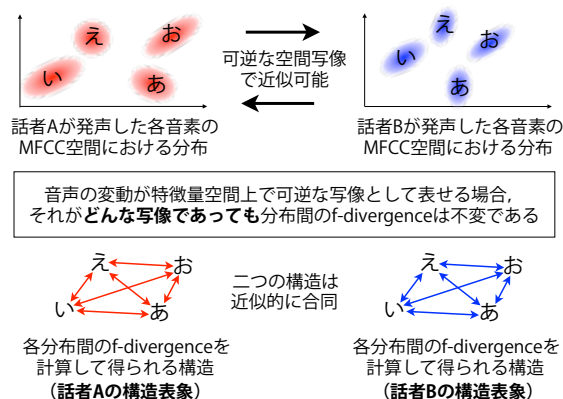


Fig. 1 静的で可逆な変換に不変な音声の構造的表象

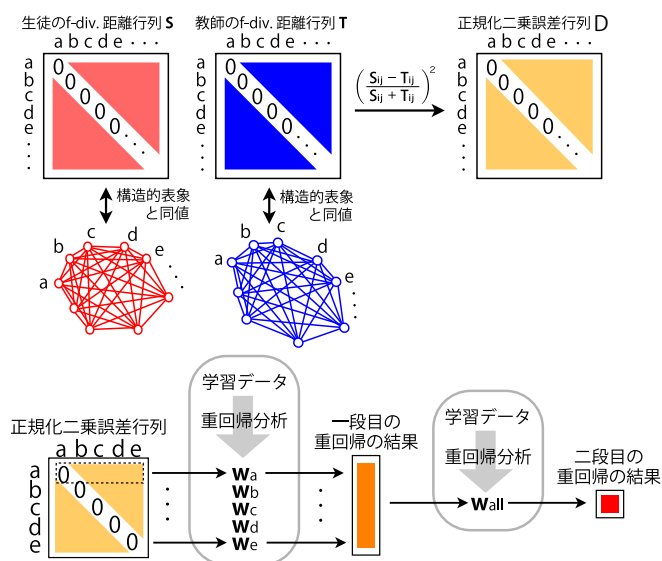


Fig. 2 構造的表象を用いた発音評価の枠組み

多段階重回帰分析の枠組みを Fig. 2 に示す. まず, 学習者の構造 ( $f$ -divergence 距離行列  $S$ ) と教師の構造 ( $f$ -divergence 距離行列  $T$ ) の, 距離行列の各要素の差異を計算する.

$$D_{ij}(S, T) = \left( \frac{S_{ij} - T_{ij}}{S_{ij} + T_{ij}} \right)^2 \quad (1)$$

このような要素  $D_{ij}$  を持つ行列  $D$  を, 正規化二乗誤差行列と呼ぶ.

次に, この正規化二乗誤差行列を, 二段階にわけて重回帰分析を行う. 重回帰分析の目的変数は, 手動で付けられた発音評定スコアである. これにより, 一段

\* A CALL system using structural features. by M.Suzuki, N.Minematsu, and K.Hirose (The University of Tokyo)

Table 1 実験結果

| ɑ:   | æ    | ʌ    | ɔ:   | ɛ    | ə:   | ɪ    | i:   | ʊ    | u:   |
|------|------|------|------|------|------|------|------|------|------|
| 0.57 | 0.55 | 0.74 | 0.03 | 0.04 | 0.73 | 0.82 | 0.86 | 0.73 | 0.55 |

Table 2 手動評価値間の相関の平均値

| ɑ:   | æ    | ʌ    | ɔ:   | ɛ    | ə:   | ɪ    | i:   | ʊ    | u:   |
|------|------|------|------|------|------|------|------|------|------|
| 0.71 | 0.68 | 0.52 | 0.47 | 0.16 | 0.82 | 0.66 | 0.68 | 0.47 | 0.48 |

Table 3 手動評価値間の標準偏差

| ɑ:   | æ    | ʌ    | ɔ:   | ɛ    | ə:   | ɪ    | i:   | ʊ    | u:   |
|------|------|------|------|------|------|------|------|------|------|
| 0.97 | 1.08 | 0.84 | 0.79 | 0.66 | 1.15 | 1.17 | 1.19 | 0.88 | 0.61 |

目の重回帰分析の結果から音素ごとのスコアを、二段目の重回帰分析の結果から学習者のスコアをそれぞれ得ることができる。さらに、重回帰分析で学習するパラメータの数を減らし、過学習を防ぐ利点もある。

### 3 音素ごとのスコアの推定

構造的表象と多段階重回帰分析を用いた発音評価の枠組みを、音素ごとのスコアを推定するタスクにおいて適用する。音声データとして、日本人中学生 36 名と 5 名の英語教師が発声した、米語単母音を網羅する 10 単語セット (pot[ɑ:], bat[æ], but[ʌ], bought[ɔ:], bet[ɛ], bird[ɜ:], bit[i], beat[i:], put[ʊ], boot[u:]) の読み上げ音声を用いる。この音声データには、音声学者 4 名が、各母音それぞれのスコアを 4 段階で評価している。

この単語音声データから、HMM の強制アライメントにより母音部分のみを切り出し、その平均と分散を計算することで各母音を分布化 [6] し、構造的表象を抽出した。これを用いて多段階重回帰分析による発音評価を行い、音素ごとの発音評価を行う。ここで、重回帰分析では、過学習を防ぐため二次の正則化項も導入したリッジ回帰を行う。さらに、正規化二乗誤差行列の要素は非負であり、かつ 0 に近いほど教師と学習者差が小さいということができるので、評価スコアを目的変数とする重回帰分析で学習すべき重みパラメータは、非正となることが望ましい。そこで、重回帰分析において、非正の制約条件を導入する [7]。これにより、より高い精度の分析が実現できる。

提案手法を用いて音素ごとの評価を行った結果と、4 名の手動評価値の平均との相関値を Table 1 に示す。結果、多くの母音において十分に高い相関が得られていることがわかる。また、Table 2 に 4 名の手動評価値間の相関の平均値を示す。これがおおよそ音声学者のパフォーマンスということができ、提案手法は母音によっては音声学者と同程度の精度で発音分析を行

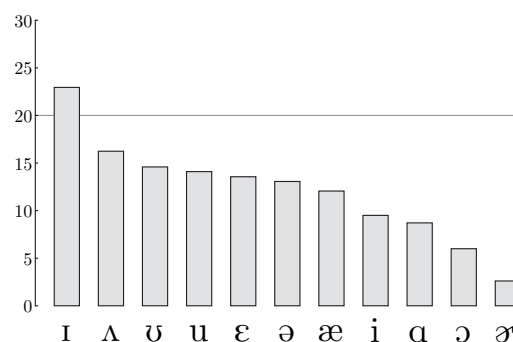


Fig. 3 どの音素から直していくべきかの表示例

えている。なお、Table 1, 2 共に相関が低い母音 (ɔ:, ɛ, u など) は、日本語の母音がほぼそのまま代用できる母音である。そのため、そもそもスコアの分散が低くなっており (Table 3), 相関が低くてもスコアの予測性能としては十分であるといえる。

### 4 CALL システムの実装

上記の実験で取り扱ったタスクを、実際に web 上で動作する CALL システムとして実装した。音声の録音には w3voice skeleton [8] を利用し、先のタスクで扱った英単語 10 単語を発音するだけで、音素ごとのスコアや、既に登録されている話者を含めた話者分類の結果、音声学者のコメントなどを見るように実装した。Fig. 3 にシステムの出力の一部分を示す。

### 5 結論

本論文では、構造的表象と多段階の重回帰分析を用いた発音評価の枠組みにおいて、重回帰分析に非正の制約条件を導入することでさらなる精度の向上が得られることを示した。さらにそれを CALL システムとして実装した。

### 参考文献

- [1] 羅 他, 信学技報, SP2009-32, pp.51–56, 2009.
- [2] M. Russell *et al.*, *Proc. SLATE*, CD-ROM, 2007.
- [3] 峯松, 信学技報, SP2003-180, pp.31–36, 2004.
- [4] 鈴木 他, 信学技報, SP2010-37, pp.13–18, 2010.
- [5] Y. Qiao *et al.*, *IEEE Trans. on Signal Processing*, Vol. 58, 7, pp.3884–3890, 2010.
- [6] 朝川 他, 電子情報通信学会論文誌, vol.J90-D, no.5, pp.1249–1262, 2007.
- [7] Lawson, C.L *et al.*, Prentice-Hall, Chapter 23, pp. 161, 1974.
- [8] 西村 他, 情報処理学会研究報告, 2007-SLP-68-3, 2007.