発音クリニック
～音声の構造的表象を用いた外国語・方言発音分析～

峯松信明†，鈴木雅之‡，高澤真章‡，馬学彬*，中村綾乃★
†東大院・情報理工　‡東大院・工学系　*東大院・新領域　★東大・工学部

1　はじめに　～声の何を真似るべきなのか～

国際語としての英語オーラルコミュニケーション能力を付与すべく，各高校・大学にて，工夫を凝らしたカリキュラムが実施されている。また，2011 年からは全公立小学校に外国語活動が導入されるが，これは「聞く・話す」を中心とする「音の教育」である。これらの状況を鑑み筆者らは，音声情報処理技術を用いた発音評価技術（発音習熟度の推定，発音矯正部位の特定，発音に基づく学習者分類など）の構築を行っている。

教師の音声と，学習者の（同一内容の）音声を比較する場合，例えば動的計画法（DP）に基づいた比較を行えば，これは，発音の評価ではなく，声帯模写の評価を行う技術となるのは自明である。女性教師の発声を真似る場合に「女声」を出す男子学生は通常いない。教師発声の統計モデルとして HMM による音響モデルを準備した場合でも，子供用発音 CALL（Computer Aided Language Learning）教材は，母語話者の子供音声 DB から構築した HMM や，話者適応技術を用いて成人用 HMM を子供用 HMM へと変換する必要がある [1]。これらの事実は，従来の発音評価技術は，本来，声帯模写評価技術と呼ぶべき枠組みであることを示唆する。この枠組みの上に発音評価システムを実装するには，上記したような事前準備が，当然，必要となる。

相手の発した声が伝達する情報を，体格差を越えて，聞き手が模倣して発声する行為（音声模倣）は，ヒト特有の行為であり，動物では観測されない。動物の音声模倣は物理的模倣が基本である [2]。また，ヒトであっても音声模倣が物理的模倣となってしまう場合がある（自閉症者の中に見られる）が，この場合，音声言語の獲得に大きな障害を呈することになる [2]。

ヒトが音声模倣する場合，相手の声の何を真似ているのだろうか？発達心理学では，この声色の違いを越えた「言葉の骨格」を，語の音形，語ゲシュタルトなどの用語で参照している。本研究ではこの話者不変な音声パターンを，任意の写像に対する写像不変量（の一般解）を用いた音声表象として捉え，それに基づく発音 CALL システムを構築している [3, 4, 5, 6]。

Pronunciation Clinic,
N. Minematsu, M. Suzuki, M. Takazawa, X. Ma, and A. Nakamura,
The University of Tokyo

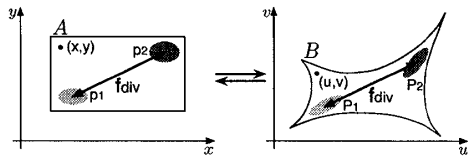

図 1: 任意の写像に対して不変な f-divergence

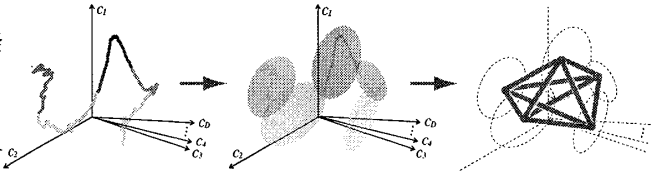
図 2: 一発声からの構造的表象の抽出

2　発声及び発音の構造的表象

声の音響的特徴は，話者の年齢，性別，更には，収録環境（マイクや部屋の音響特性）によって容易に変形する。ヒトの頑健な音声コミュニケーションは，物理的には非常に脆い現象の上で行われている。これらの静的な変形は，数学的には空間写像としてモデル化でき，入力音声に対して適切な写像を施すことで，別話者の音声に変換すること（音声変換）ができる。

筆者らによって，可逆かつ連続な任意写像に対する不変量の一般解が導出されている [7]。下記の f-divergence は任意写像に対して不変であり，任意写像に対する不変量は f-divergence のみである（**図 1** 参照）。

$$f_{div}(p_1, p_2) = \oint p_2(\boldsymbol{x}) g\left(\frac{p_1(\boldsymbol{x})}{p_2(\boldsymbol{x})}\right) d\boldsymbol{x}$$

p_i は注目する空間における事象 i である。各事象を分布として表現し，全ての分布間距離を f-div. で計測すれば，得られる距離行列は任意の写像に対して不変となる。この距離行列を構造的表象と呼ぶ（**図 2** 参照）。

教師音声中の音響事象群から構造を抽出し，同一内容を意図した学習者の発声から同様に構造を抽出して両者を比較すれば，年齢・性別などの要因がそぎ落とされた発音比較が可能となり，学習に有益な種々の情報の表示が可能となる。[6] では，音声変換技術を用いて巨人・小人の声を人工的に生成し，これらと，単独の母語話者英語教師との音声を比較するこで発音習熟度推定を行っており，体格には一切依存しない発音比較が可能となっていることを実験的に示している。

図 3: 教師を選ぶユーザ・インタフェース

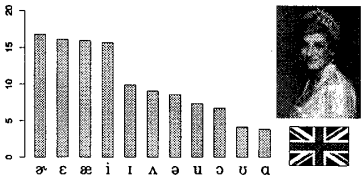

図 4: 矯正すべき母音とその優先度の表示

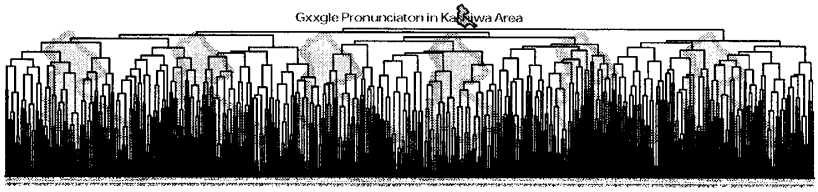

図 5: 大規模日本人学習者群の発音分類結果

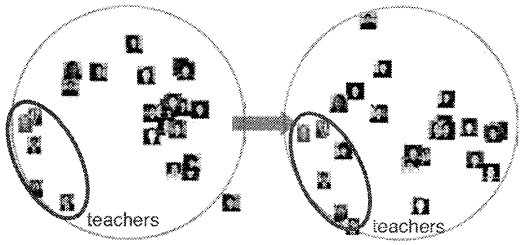

図 6: 発音学習前後におけるクラスマップの変化

3 発音クリニック

英語学習初心者を想定し，英語の単母音発音を評価するデモシステムを構築した。ここでは，beat, bit, bat, but など，11 単母音を含む英単語を発声させ，母音部位のみから分布を推定し，各学習者を 11 角形（母音構造）として表象する。これと，教師の（及び他学習者の）11 角形を比較し，様々な情報を学習者に提供する。下記に，本システムの機能及び今後の課題を示す。

3.1 モデル発音・教師の選択

多人数の母語話者音声データを用いて構築された不特定話者音響モデルを必要とする従来手法と大きく異なり，本手法では，比較対象とする英語教師一名の音声のみが必要になる。これは学習者に「自分と比較して欲しい発音・教師」を選ばせることが可能であることを意味する。既に，一部の著名人の英語音声を用いた「教師を選ぶ」インタフェースを構築している。発音学習の動機付けには，非常に効果的である（**図 3** 参照）。

3.2 発音変遷ログ

各学習者は，着目する発音部位（事象）群によって構成される距離行列として表象されるが，これを適切に視覚化することで，「発音カルテ」相当の情報を提供できる。学習によって発音の様態は変化する。その様子を逐一記録し（ログ化し），過去の自分との違いを意識した上で次のステップに進むことも可能となる。視覚化手法としては樹形図表記を用いている。

3.3 習熟度の推定と矯正部位の推定

選択された教師の発音構造と比較することで，学習者の発音習熟度を定量的に示すことができる（距離行列間差異の定量化）。距離行列には事象 i とそれ以外の事象間との距離情報が含まれているが，個々の事象間距離を教師・学生間で比較することで，その学習者が優先的に矯正すべき発音部位が推定可能である。言うなれば「教師発音へ至る近道の提示」である。No two students are the same. 同一教師を目標とした場合でも，異なる学習者には，異なる近道が提示される（**図 4** 参照）。

3.4 発音に基づく学習者分類

任意の距離行列に対して，その樹形図（分類木）が描画できる。11 角形間距離（構造間距離）を適切に定義できれば，学習者間の距離行列が得られる。これより，年齢・性別には依存しない，発音の様態のみに基づく学習者分類が可能である。[6] では，巨人や小人の発音を，体格に依らず分類できることを実験的に示している。既に，600 名に及ぶ日本人学習者の分類（**図 5** 参照）やクラスマップの作成（**図 6** 参照）を行っている。

3.5 今後の課題

現在，二重母音まで考慮した発音評価，任意の（同一）発話内容に対する教師・学習者間の発声比較のための技術構築を行っている。随時，デモシステムに反映し，より実践的なシステムへと繋げる予定である。

参考文献

[1] M. Russell *et al.*, Proc. SLaTE, CD-ROM (2007)
[2] 峯松，信学技報，SP2008-84, pp.31–36 (2008)
[3] 朝川他，電子情報通信学会論文誌，vol.J90-D, no.5, pp.1249–1262 (2007)
[4] 鎌田他，進学技報，SP2007-36, pp.73–78 (2007)
[5] 高澤，春季音講論，3-10-12, pp.489–492 (2008)
[6] M. Suzuki *et al.*, Proc. ASRU, pp.574–579 (2009)
[7] Y. Qiao *et al.*, Proc. INTERSPEECH, pp.1349–1352 (2008)