大規模英語学習者を対象とした音声の構造的表象に基づく発音分類とその応用

鎌田 圭 † 高澤 真章 ††　竹内 京子 †††　朝川 智 † 峯松 信明 † 牧野 武彦 ‡　広瀬 啓吉 ‡‡
† 東京大学大学院新領域創成科学研究科　†† 東京大学工学部情報工学科　††† 東京大学大学院総合文化研究科
‡ 中央大学経済学部　‡‡ 東京大学大学院情報理工学研究科

1　はじめに

音声の音響的特徴は，話者や音響機器などの非言語的要因によって不可避的に変動し，同じ発話内容でも観測される音響事象は異なる．従来の音声システムでは，基本的に，大量の音声を用いて個々の言語事象を統計的にモデル化することで対処してきた．これは，不可避な音声歪みを内包した特徴量を利用していることに相当し，不一致問題の解決には至らず，話者を選ぶシステムとなっていた．音声認識の分野では，種々の適応技術（例えば話者適応）が用いられているが，本稿で検討するような発音学習に適用すると，下手な発音に対して高いスコアを与えるような適応が免れない．これは，話者の違いと発音の上手下手とがスペクトル包絡という同一物理現象に基づくため，両者を切り分ける術を持っていなかったことが根本原因である．

近年，非言語的要因に起因する静的な音響歪みを完全に排除して音声を表象する「音声の構造的表象」が提案された [1]．音声事象そのものではなく，事象間の相対的な関係のみを抽出／表象する．即ち，発話に含まれる有限個の事象から，全ての事象間差異（距離）を計算する．この際，一対一対応を満たす線形／非線形変換に対する変換不変量となる距離尺度 [2] を採択することで，非言語的変換に対して完全不変の音声表象となる．この音声表象は構造音韻論の数学的／物理的実装に相当し，外国語学習者の発音状態を非言語的要因の影響を受けずに記述することが可能である [3]．

提案表象を用いると，例えば子供の英語発音（細い声）と大人の英語発音（太い声）とを直接比較し，子供が母語話者であった場合，大人（学習者）の発音のどこを矯正すべきか，との教示を得ることができる [4]．また，複数の教師の中から，自分のお好みの教師を選ぶことが可能となる．図 1 は提案表象を用いて作成したデモシステム・インタフェースである．例えば，憧れの映画スターの音声データを使えば，その憧れに近づく最も効率的な発音矯正法を学習者別に提示することが可能である．更に本表象を用いることで，非言語要因に影響を受けずに，学習者群の発音（方言）分類が可能となる．地球上には 20 億の英語学習者がいると言われるが，学習過程における彼らの時間的変化を考慮すると，20 億 $\times N$ 個となる発音を分類し，英語発音の世界地図を作ることも不可能ではない [5]．このようなインタフェースや試みは「複数話者の音声混ぜ合わせ」を基本とする従来法では不可能であり，本表象によって初めて可能となるアプリケーションである．

本稿ではこの発音分類について，模擬英語学習者 96 名に対して行った発音分類の実効性検証，及び，3 年間に渡るオープンキャンパスを通して収集した柏市民 564 名の英語発音を分類することにより，柏に存在する英語方言の実験的定義を試みたので報告する．

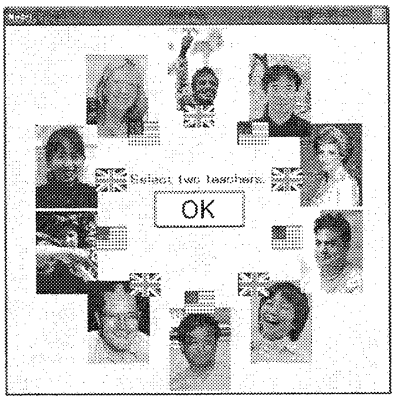

図 1: Select your favorite teachers!

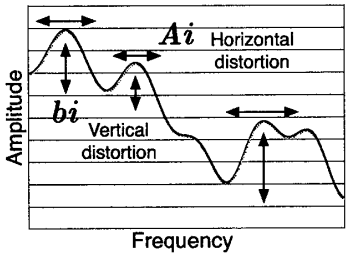

図 2: スペクトルに対する水平/垂直方向の音響歪み

2　音声（発音）の構造的表象

静的な非言語的音響歪みは，乗算性歪みと線形変換性歪みに大別される．乗算性歪みは，マイクなどの伝送特性や話者の声道形状の違いの一部に相当し，ケプストラムベクトル c に対するベクトル b の加算 $c'=c+b$ となる．線形変換性歪みは，声道長の差異や聴覚特性の差異が挙げられ，行列 A の乗算 $c'=Ac$ で精度良くモデル化される．以上より，不可避かつ静的な非言語的音響歪みは，アフィン変換でモデル化できる．

図 2 はこの 2 種類の歪み（乗算性歪み＝垂直，線形変換性歪み＝水平）を示している．個々の音声事象を分布として捉え，全ての二事象間距離を計算すると（距離行列），それは，事象群が成す幾何学構造を規定することになる．バタチャリヤ距離はアフィン変換のみならず，非線形変換も含め不変性を有するため，これを用いることで，静的な非言語的音響歪みに一切不変な音声表象が得られる．この構造を音響的普遍構造と呼んでいる．

3　日本人英語学習者の模擬音声を用いた発音分類

3.1　日本人英語学習者の模擬音声の作成

日本語・米語（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

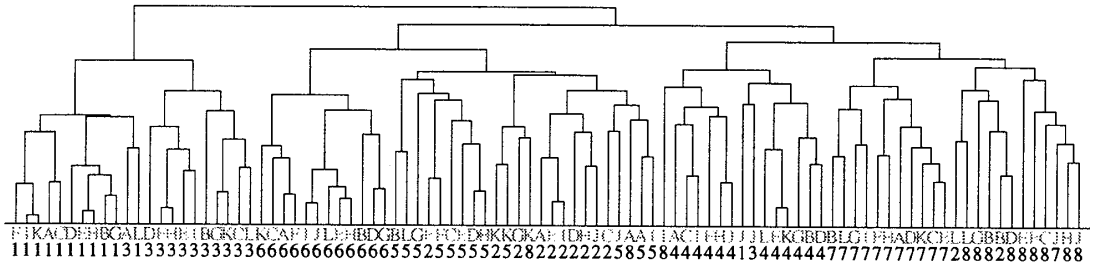

図 3: 構造的表象に基づく英語発音分類

表 1: 母音置換によって模擬された 8 種類の発音状態

	ɑ	æ	ʌ	ə	ɚ	ɪ	i	ʊ	u	ɛ	ɔ
P1											
P2	A	A	A	A	A						
P3						A	A	A	A	A	A
P4	A	A				A	A			A	A
P5			A	A	A			A	A		
P6	A		A		A					A	A
P7		A		A		A	A	A	A		
P8	A	A	A	A	A	A	A	A	A	A	A

A：米語母音を使用，　：日本語母音で置換

表 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/,/ɚ/,/ə/
/い/		/i/,/ɪ/
/う/		/u/,/ʊ/
/え/		/ɛ/
/お/		/ɔ/

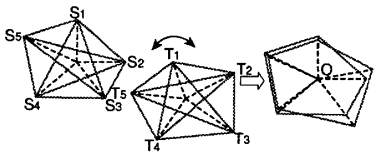

図 4: 回転とシフトによる構造群の接近と構造間距離

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．

収録した米語母音と日本語母音を同一話者内で置換し，米語教師や日本人米語学習者の音声を模擬した．置換パターンには表 1 を用い，1 話者の音声から 7 人の日本人米語学習者（P1～P7）と 1 人の米語教師（P8）相当の発音を用意した．なお，米語と日本語の母音置換には，表 2 に示す日本人学習者に頻出する置換母音を用いた．異なる米語母音が同一の日本語母音と置換される場合は，同一母音・異発声のサンプルを用いた．

3.2 学習者間距離の導出と学習者発音の自動分類

各学習者の 11 母音の母音間距離を全て求めることにより，母音構造を抽出する．そして，母音構造（発音訛り）間の距離を学習者間距離と定義する．構造間距離は，サイズ正規化後，一方の構造を回転及びシフトして他方に近づけた時に観測される，対応する二点間距離の総和の最小値として規定できる（図 4 参照）．この最小値は，正規化後の距離行列間のユークリッド距離として近似でき，この距離尺度を使用した [6]．

$$D(S,T) = \sqrt{\frac{1}{M}\sum_{i<j}(S_{ij} - T_{ij})^2} \qquad (1)$$

この距離尺度に基づいて，N 人の学習者から $N \times N$ の学習者間距離行列を求めれば，学習者群に対する樹形図が構成可能になる．前述の 96 名分の模擬音声に対して，Ward 法により作成した樹形図が図 3 である．A～L が話者を表し，1～8 が発音である．ほぼ発音状態ごとにクラスタを構成できていることが示されている．構造間距離を $D(S,T)$ とは異なり「二構造間の対応する母音実体間距離の総和」として定義すると図 3 は完全な話者分類となる [5]．実体を捉えることが話者が分類され，差異を捉えることで発音が分類できる．

表 3: 英語学習者クラスタの年齢・性別構成

	年齢				性別	
No.	0～19	20～39	40～59	60～	男	女
1	21	70	36	14	89	52
2	26	27	15	12	39	41
3	32	48	27	15	77	45
4	14	33	24	14	44	41
5	25	75	29	7	82	54

4 大規模英語学習者の発音分類

4.1 英語学習者の音声の収録

3 年に渡る東京大学柏オープンキャンパスを通して，英語学習者（主に日本人）の音声を延べ 697 名から収録した．収録内容は英語 11 単語を 1 回ずつである．収録した話者は 4 歳から 84 歳まで広い年齢層に渡る．録音環境が万全でなかったため，事前に聴取を行ない，録音状態が良好な音声として 564 名分を選別した．

4.2 大規模英語学習者の発音分類

$N \times N$ の母音構造間距離行列（N=564）に Ward 法によるボトムアップクラスタリングを適用し，学習者分類を行う．5 つのクラスタに分類したときの，各クラスタの年齢・性別構成を示したものが表 3 である．

話者性に束縛されることなく発音分類が行われていることがよく分かる．各クラスタに属する話者の音声を聴取したところ，「カタカナ発音（典型的な日本人英語）」や「/i/と/ɪ/を混同している発音」といった分類が行われていた．英語発音練度の低い学習者が多かったため，「アメリカ方言英語」や「イギリス方言英語」といったクラスタは得られず，「上手な英語発音」クラスタが 1 つ存在するのみであったが，5 クラスタ以上の分割を行うことで，クラスタに属する人数は少ないものの「アメリカ方言英語」や「イギリス方言英語」と呼べるクラスタを確認することが出来た．現在，本提案手法が算出する日本人英語分類の音声学的／教育学的妥当性について音声学者との議論を行なっている．

参考文献

[1] 峯松他, 信学技報, SP2005-12, 1-8, 2005.
[2] 峯松他, 春音講論, 1-P-12, 147-148, 2007
[3] 朝川他, 電子情報通信学会論文誌, J90-D, 5, 1249-1262, 2007.
[4] 鎌田他, 信学技報, SP2007-36, 73-78, 2007.
[5] 鎌田他, 信学技報, SP2006-77, 7-12, 2006.
[6] 峯松他, 信学技報, SP2005-13, 9-12, 2005.
[7] 峯松他, 信学技報, TL2006-39, 49-54, 2006.
[8] 峯松他, 信学技報, TL2006-40, 55-60, 2006.