音声の構造的表象に基づく学習者分類の教育的妥当性に関する実験的検討*

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

1 はじめに

年齢，性別，個人性，音響機器など，音声には不可避的に非言語的情報が付随し，これが音声の音響的特徴を変形させ，音声記号としては同一音であったとしても，物理的には異なる特性を持った音響現象として観測されることになる。従来の音声工学では，数千・数万という話者の音声から同一記号音を収集して統計的にモデル化し，発音学習応用においては，この統計モデルと学習者音声とを直接的に比較する方法論をとってきた。しかし，言語獲得や外国語学習において，親や教師の音声の物理的特性を直接的に真似ようとする幼児や生徒は皆無である。女性教師の発音を真似ようと女声の獲得を目指す男子学生はいない。音韻意識が希薄な幼児とて同様，父親の太い声を出そうと努力する訳では無い。一方，九官鳥は音を真似るため，優秀な九官鳥は聞けば飼い主が分かる [1]。幼児は音の何を真似ているのだろうか [2]？

筆者らは，(線形，非線形を問わず) 変換不変なる音響特徴量が存在することを数学的に示し，それに基づいて，話者不変の音声の全体的・構造的表象を導出し，音声認識，発音学習などに応用してきた。この表象は，音声を，(音高ではなく) 音色が変化する音楽として捉えることに相当する [2]。相対音感者が，移調前後の物理的には異なる音ストリームを，同一のドレミ列として記述するように，異なる話者による音ストリームを同一のモーラ列として記述するのが音声の書き起こしと考える。筆者らは，幼児の音声模倣は，この音楽的な音パターンを基盤とする，と考えている。この時，個々の音韻の意識は不要である [2]。

従来筆者らは，日本人英語学習者の母音発声 (孤立単語音声から切り出した母音区間) に対して，構造解析を行ない，凡そ話者性に依存しない形で，(歪んだ) 母音構造を抽出し，それに基づいて発音評定，即ち，矯正すべき母音を自動推定する枠組みについて検討してきた。更には，話者不変なる構造表象であることに着目し，性別／年齢を越えた学習者群の自動分類についても検討してきた。本稿では学習者群の自動分類結果を，調音音声学／英語音声学を専門とする音声学者の聴取による学習者分類結果と比較し，自動分類結果の教育的妥当性について検討する [3]。

2 話者不変量としての音声の構造的表象

音声に混入する不可避的な非言語的特徴は，乗算性歪みと線形変換性歪みに大別され，対数パワースペクトルに対して，前者は垂直方向の変形として，後者は水平方向の変形として現れる。これらの歪みはケプストラムベクトル c に対して $c' = Ac + b$ として線形変換でモデル化することができ，このモデルに対する不変物理量を用いることで話者不変の音声表象が構成される。筆者らは，分布間距離として知られるバタチャリヤ距離がこの性質を満たすことに着目し，音声事象を全て分布として推定し，全ての分布間距離を求め，距離行列として音声事象群を表象することを提案している。距離行列は幾何学的構造を規定するため，これは音声の構造的表象となる。なお，この構造的表象は線形のみならず，非線形変換に対しても不変性を満たすことが示されている [2]。

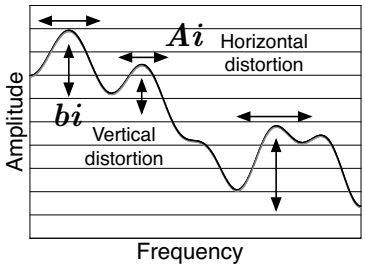

Fig. 1 スペクトルに対する水平/垂直方向の音響歪み

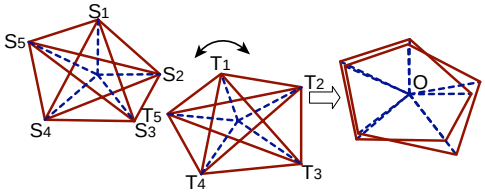

Fig. 2 回転＆シフトによる構造間距離の計算

3 学習者間距離の導出と学習者自動分類

図 1 に示した線形変換モデルで話者の違いを考えた場合，行列 A の乗算は構造の回転に，ベクトル b の加算は構造のシフトに対応する。即ち，異なる二話者の母音構造を比較する場合，**図 2** に示すように，回転＆シフト後の，両構造の対応する頂点間距離の総和の最小値を求めることになる。この最小値は，2 つの距離行列をベクトルとして見なした場合のユークリッド距離に近似的に比例することが示されており，本稿でもこれを学習者間距離として採用する (下式。M は母音数)。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，学習者群に対する分類木／樹形図が構成可能となる。

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

*An experimental study on pedagogical evaluation of the learners' classification based on the structural representation of pronunciation. by S. Asakawa, K. Kamata, N. Minematsu, T. Makino, and K. Hirose

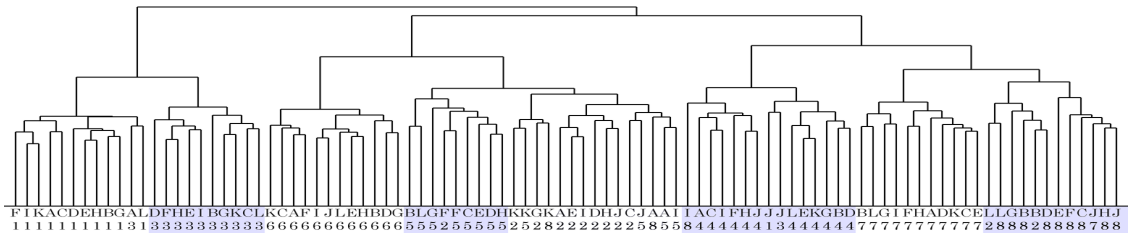

Fig. 3　構造的表象（自動音響分析）に基づく 96 発音構造（[12 話者 A〜L] × [8 発音状態 1〜8]）の分類

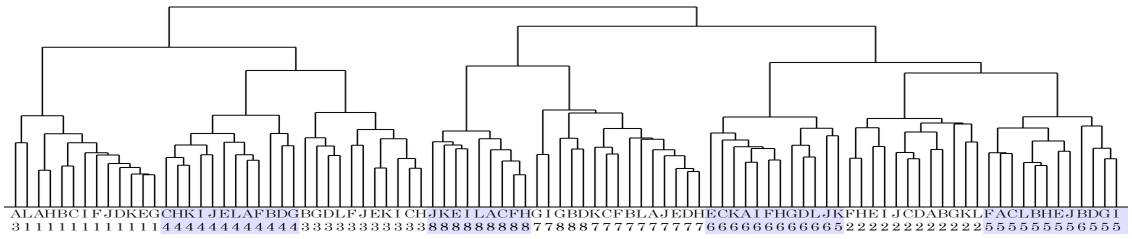

Fig. 4　母音図（聴取）に基づく 96 発音構造（[12 話者 A〜L] × [8 発音状態 1〜8]）の分類

Table 1　日本語母音・米語母音の対応表

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/

Table 2　母音置換に基づく 8 種類の母音発音

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

A：米語母音を使用，J：日本語母音で置換

4　日本人英語模擬音声の収集

帰国子女の日本人 12 名（男女 6 名ずつ，A〜L）に米語 11 母音を/b V t/の形で，及び，日本語母音を/b V t o/の形で発声させた。米語は母音当たり 1 発声，日本語は母音当たり 5 発声収録した。この米語母音の幾つかを日本語母音と置換する形で日本人英語音声を模擬した。日米母音の対応を**表 1** に，この表に基づいて定義した 8 種類の日本人英語母音発音を**表 2** に示す。各話者毎に 8 種類の発音が定義されるため，合計 96 種類の母音群構造が得られる。下記では，この 96 種類の母音構造を 2 種類の方法で分類する。

5　聴取による距離行列（母音図）の作成

96 種類の母音構造中に存在する全ての母音（1,056 個）を英語音声学／調音音声学を専門とする音声学者（第四著者）に，各構造（11 母音）毎に聴取させ，母音図としてプロットさせた。通常，2 次元の母音図は舌の位置を指定するために使われるが，本研究ではこれに加え，円唇性を 3 段階，r 音性を 2 段階で評定した。得られた結果（2 次元母音図＋円唇性＋ r 音性，即ち 4 つの数値の組）から，次のようにして距離行列を計算した。まず，2 次元母音図については，広く用いられている 4:3:2 の台形表示を踏襲した。円唇性及び r 音性に対する次元の重みについては，参考数値を調音音声学の文献に見つけることが困難であったため[1]，母音間のケプストラム距離，及び 4:3:2 の台形表示を考慮して，**図 5** に示す 4 次元母音図を用い，そのユークリッド距離で距離行列を構成した。

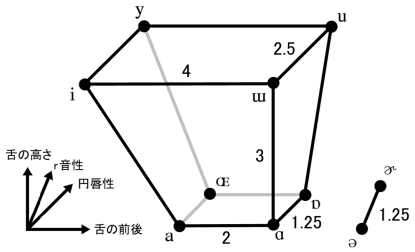

Fig. 5　4 次元母音図と各軸の重み

6　学習者群の自動分類と聴取による分類

自動音響分析から得られた距離行列と，音声学者の聴取による距離行列に対して，**式 (1)** による構造間距離尺度を用いて二種類の 96×96 距離行列を作り，樹形図を構成した。結果を**図 3** と**図 4** に示す。P4 の発音の位置が両者で異なるものの，良好な対応関係が見られる。なお，樹形図は多次元空間に存在する幾何学構造に対する 2 次元視覚化であるため，樹形図化することによる歪みは免れない。そこで，96 種類の 11×11 距離行列（5,280 個の母音間距離）に対する，音響分析及び聴取間の相関係数を求めたところ，0.72 であり，良好な対応関係を得ることができた。

参考文献

[1] 宮本, “音を作る・音を見る,” 森北出版, 1995.
[2] 峯松他, 人工知能学会, SIG-Challenge-0624-6, 35–42, 2006.
[3] 鎌田他, 電子情報通信学会, SP2006-77, 7–12, 2006.

[1]母音位置＋円唇性による 3 次元母音図などが提案されているが，円唇性の軸（奥行き）の長さに対する定量的な議論は無い。