発音の構造的表象に基づく学習者群のクラスタリング*

◎朝川 智, 峯松 信明, 広瀬 啓吉 (東大)

1 はじめに

音声生成、伝達、聴取の過程に混入する不可避的な静的歪みにおよそ影響を受けない新しい音声の表象手法として音響的普遍構造が提案されている [1]。この音声の構造的表象によって英語学習者の発音状態の記述と発音習熟度の自動評定が可能となることが示されている [2,3]。さらに、構造的表象を用いることで、少量データ時において英語母音の発音状態が変化の様子の追跡が可能となることを示した [4]。これは発音カルテが作成可能であることを意味する。即ち各学習者の発音状態を、性別や年齢に影響されることなく「音響的に」表現でき、かつ、発音学習の発達状況を詳細に追跡できることを意味する。個々の学習者の記述が可能となると、次に求められるのが学習者群を分類する技術である。本報では、下記に示すような形で種々の発音状態を模擬的に作成し、話者・性別といった非言語的特徴に影響されることなく学習者群の分類が可能となることを実験的に示す。

2 音声の構造的表象

ある話者によって発声された各音素を、ケプストラムベクトルによって構成される多次元ガウス分布であると仮定し、 n 個の音素に対する nC_2 個の音素間距離行列を求めることで音声を構造化することを考える。音素間距離としてバタチャリヤ距離の平方根を用いる。二つの確率密度分布 $p_i(\mathbf{x})$ と $p_j(\mathbf{x})$ 間のバタチャリヤ距離は、以下の式によって与えられる。

$$BD(p_i(\mathbf{x}), p_j(\mathbf{x})) = -\ln \int_{-\infty}^{\infty} \sqrt{p_i(\mathbf{x})p_j(\mathbf{x})} d\mathbf{x} \quad (1)$$

バタチャリヤ距離は、任意の二分布間に対して如何なる共通のアフィン変換を施しても距離は不変である、という性質を持つ。つまり、得られる音素間距離行列もアフィン変換前後で不変であるといえる。ケプストラムベクトル $\mathbf{c}$ に対するアフィン変換 $\mathbf{A}\mathbf{c} + \mathbf{b}$ では、行列 $\mathbf{A}$ の乗算は声道長特性、聴覚特性など話者・聴取者の生理学的な差異を表現し、ベクトル $\mathbf{b}$ の加算は音響機器などの差異を表現する事を考えると、音響的普遍構造はこれらの差異によって歪まない物理表象であると言える。

3 クラスタリングによる発音状態の分類

3.1 学習者の記述・比較・分類

音韻群を構造化することにより、音声物理に不可避的に混入する非言語的特徴がそぎ落とされる。学習者音声に対して音韻構造を抽出すると、純粹に学習者の発音状態を記述することが可能となる。そこには主に母国語に起因する構造的歪みが観測されることが考えられ、外国語学習とは歪んだ音韻構造から正常な

音韻構造への遷移としてモデル化することができる。既に、英語の上達する様子を模擬した音声を用いて、歪んだ構造から正常な構造への遷移に対する構造分析を行っている [4]。学習者を構造として記述することが出来れば、2つの構造間の距離尺度を定義することで、2学習者間の差異をスカラー量として定義することができる。構造間距離 D は下記の式より近似的に求められることが示されており、一方の構造を回転・シフトさせて他方の構造と合わせたときの対応する二点間距離の最小値に近似的に比例する [3]。

$$D = \sqrt{\frac{1}{M} \sum_{i < j} (P_{ij} - Q_{ij})^2} \quad (2)$$

P_{ij} , Q_{ij} は距離行列の要素、 i, j は母音種類、 M は母音数を表す。2学習者間距離がスカラー量で表現されると、任意の学習者集団に対して学習者距離行列を定義することができ、学習者全体のボトムアップクラスタリングにより学習者発音の分類が可能となる。既に [3] において学習者発音構造の分類は試みられているが、読み上げ文セットの不均一性や読み上げ方などの収録条件による影響を排除できていない。また、数十文の読み上げによる分析であり実際の応用では困難となる可能性がある。更には、各話者の「発音構造」としての発音状態が原則的に未知なため、各学習者に対する評定スコアのみを用いた検討になっている。即ち、[3] における学習者クラスタリングは、クラスタリング結果の妥当性についての検討は今後の課題となっていた。

本分析では、英語と日本語の母音発声を用いて、両者を同一話者内で混ぜ合わせることで様々な学習者の発音状態を模擬し、複数の話者・複数の状態における発音構造を対象としてボトムアップクラスタリングを行う。模擬的な音声を用いることで各発音構造における発音状態が既知となり、発音構造クラスタリングに対して発音状態を基にした検討が可能となる。構造的表象を用いることで、話者や性別などの非言語的特徴に影響されることなく、学習者の発音状態によって分類することが可能となることを実験的に示す。

3.2 分析手法

帰国子女あるいは英語劇経験者の日本人 12 名 (男性 6 名、女性 6 名) より、種々の母音 V に対して /bVt/ の発声を収録した。母音 V としては、英語母音 11 種類 (/a, æ, ʌ, ə, ɜ, ɪ, i, u, ʊ, e, ɔ/) と、日本語母音 5 種類 (/a, i, u, e, o/) をそれぞれ収録した。両者の発音を同一話者内で混ぜ合わせることで英語学習者の様々な発音状態を模擬する。日本人学習者 (特に低習熟度話者) が英語を発音する際、英語の母音を日本語の母音に置き換えて発音してしまうことがしばしばある。その典型的な置換の組み合わせ例を Table 1 に示す。この組み合わせを用いて、Table 2 に示す 8 つ

* Clustering language learners using the structural representation of the pronunciation.
by Satoshi Asakawa, Nobuaki Minematsu and Keikichi Hirose (University of Tokyo)