

発音の構造的表象に基づく学習者分類と発音教示生成

朝川 智[†] 峯松 信明[†] 広瀬 啓吉^{††}

[†] 東京大学大学院新領域創成科学研究科

^{††} 東京大学大学院情報理工学系研究科

〒 113-0033 東京都文京区本郷 7-3-1

E-mail: †{asakawa,mine,hirose}@gavo.t.u-tokyo.ac.jp

あらまし 音声の生成, 伝達, 聴取の過程において混入する不可避免的な静的歪みに影響を受けない新しい音声の表象手法として, 音声の音響的普遍構造が提案されている. この構造的表象によって英語学習者の発音状態を描くことが可能であり, さらに英語学習者の発音自動評価が可能となることが示されている. 本報では, 構造的表象を用いた学習者の分類と学習者に対する発音教示に関する2つの分析について報告する. 前者の分析では, 英語学習者における様々な発音状態を模擬した音声を用いて, 話者・性別といった非言語的特徴に影響されることなく, 学習者群を発音状態によって分類可能であることを示す. さらに後者の分析では, その分類を基に, 各状態に属する学習者の発音構造と学習者が目標とする発音状態における構造との比較を行い, 2構造間において最も大きい構造歪みを呈する音素を求めることで, 目標に近づくために最も効果的である音素, 即ち次に学習すべき音素の提示が可能となることを実験的に示す.

キーワード 音声の構造的表象, 日本人英語, 学習者クラスタリング, 発音教示生成

Clustering language learners and generating pronunciation instructions based on the structural representation of the pronunciation

Satoshi ASAKAWA[†], Nobuaki MINEMATSU[†], and Keikichi HIROSE^{††}

[†] Graduate School of Frontier Sciences,

^{††} Graduate School of Information Science and Technology, University of Tokyo

7-3-1, Hongo, Bunkyo-ku, Tokyo 113-0033 Japan

E-mail: †{asakawa,mine,hirose}@gavo.t.u-tokyo.ac.jp

Abstract Recently, a novel acoustic representation of speech was proposed, where dimensions of the non-linguistic factors can hardly be seen. Using this structural representation, individual learners were described as distorted phonemic structures and automatic evaluation of the pronunciation was investigated. This paper describes two analyses using the proposed method. The first analysis is done to examine whether the method can acoustically classify language students purely according to their pronunciation, i.e., independently of difference in gender, size, age, etc. The second one shows that the proposed method can indicate production of which phoneme should be corrected at first for individual students. This indication is done by comparing two pronunciation structures, one extracted from utterances of a student and the other extracted from those of his/her teacher. By detecting the phoneme which produces the largest structural distortion, the phoneme is shown to the student as the most emergent phoneme to correct.

Key words structural representation of speech, Japanese English, clustering language learners, instruction generation

1. はじめに

従来の音声工学では, 音響音声学に基づきスペクトルをその

物理表象として用いてきた. しかしスペクトルには性別・年齢などの生理学的特性や収録機器などの音響的特性の違いといった非言語的特徴が音響的な歪みとして不可避免的に含まれており,

この歪んだ物理表象を用いて構築された CALL システムは、結果として不安定なシステムとなる場合がある。これら非言語的特徴に影響されることなく、純粋に発音状態のみを記述することのできる方法は無いのだろうか。

音素の定義を、意味論を切り離して言語学に求めた場合、2つの定義が導かれる [1]. 1) a phoneme is a class of phonetically-similar sounds and 2) a phoneme is one element in the sound system of a language having a characteristic set of interrelations with each of the other elements in that system. 前者の定義により音声認識で用いられる不特定話者音響モデルが導かれたことは自明である。しかし、どれだけ大量の/a/を集めたとしても、あくまで平均的な分布を求めたにすぎず、不可避免的に混入する非言語的特徴を消し去ることはできない。近年、上記の非言語的特徴を表現する次元そのものを保有しない「音響的普遍構造」が提案された [2]。本表象は、複数の音声事象の関係のみに着眼する物理表象であり、後者の音素定義に基づいた音響モデリングである。即ち、この構造的表象は、言語学的には構造音韻論の物理実装と解釈される。一方、人間が音声コミュニケーションを行う際にもこの構造的表象を利用していることが示唆されており [3]、認知心理学的には音声ゲシュタルトとして解釈される [4]。筆者らは既に、この新しい音声の構造的表象を発音学習に応用し、英語学習者の発音状態の記述と、より頑健で安定した発音評価が可能となることを報告した [5], [6]。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する様子の追跡が可能となることを示した [7]。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。そして、発音状態による分類が可能となれば、日本人英語学習者の典型的な発音状況をいくつかの状態として定義することが可能となり、各状態に属する学習者に対してどの音素の発音を矯正するのが目標に到達するためには効果的であるのか、という発音教示が可能となる。

本報では、下記に示すような形で種々の学習者の様々な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、学習者群を発音状態によって分類可能であることを実験的に示す。さらにその分類を基に、各状態に属する学習者の発音構造と目標とする発音構造とを比較して最も構造歪みの大きい音素を求めることで、目標に到達するために最も効果的である音素の提示、即ち発音教示が可能となることを実験的に示す。

2. 音声の構造的表象

2.1 音声に不可避免的に混入する非言語的特徴

音声の物理的実体に混入する歪みとしては、主に加算性雑音、乗算性歪み、線形変換性歪みの三種類に分類される。このうち、音声に「不可避免的に」混入するものは乗算性歪み、線形変換性歪みの二つである。加算性雑音とは、時間軸上の加算、即ち近

似的にはスペクトルに対する加算としても表現されうる雑音であり、テレビ・ラジオなどの背景雑音がその典型例と言える。これらは場所を移動するなどの対策を練ることで、物理的に抹消することができるので、不可避免的な雑音ではない。本研究ではこの加算性雑音は扱う対象とはしない。

乗算性歪みは、スペクトルに対する乗算で表現される歪みであり、ケプストラムベクトル c に対するベクトル b の加算 $c' = c + b$ に相当する。マイクロフォンなどの伝送特性がその典型例である。また、話者の声道形状の違いの一部も近似的に乗算性歪みとして扱われる。音声は必ずある話者によって発声され、ある音響機器によって収録されるので、これらは不可避免的な歪みであるといえる。線形変換性歪みは、 c に対する行列 A の乗算 $c' = Ac$ で表現される歪みである。話者の声道長の差異、聴取者の聴覚特性の差異を表すために、対数スペクトルに対して周波数ウォーピングが施されるが、単調増加かつ連続である周波数ウォーピングは、 c に対する A の乗算で表されることが示されている [8]。即ち、声道長の差異、聴覚特性の差異は近似的に線形変換性歪みとして扱うことができる。これらも不可避免的な歪みである。以上をまとめると、音声の物理的実体には非言語的特徴が不可避免的に混入し、これらはケプストラムベクトル c に対するアフィン変換 $c' = Ac + b$ で表現される。

2.2 音声に内在する音響的普遍構造

空間内に存在する N 点に対して、 ${}_NC_2$ 個の全ての二分布間距離を求めると、その N 点で張られる構造は一意に規定される。即ち事象群に対して、全ての二事象間距離を求めると、その事象群を構造的に表象することになる。ケプストラム空間内の N 点に対して構造を考えた場合、その構造は非言語的特徴によって不可避免的に歪む。何故なら、非言語的特徴はアフィン変換としてモデル化されるからである。この不可避免的に歪む構造は、空間を歪ませることによって不変な構造として定義可能となる。

構造不変の定理：意味のある記述が分布としてのみ可能な物理現象を考える。分布群に対して、全ての二分布間距離を求める（距離行列）。二分布間距離として、バタチャリヤ距離、カルバック・ライブラ距離、ヘリングー距離などを用いた場合、各分布に対して単一の任意一次変換を施しても、二分布間距離は不変である。即ち距離行列は不変であり、その結果、構造も不変となる（図 1 参照）。

以下、バタチャリヤ距離を用いて話を進める。二つの分布の確率密度関数をそれぞれ $p_1(x)$, $p_2(x)$ とすると、バタチャリヤ距離は以下の式で表される。

$$BD(p_1(x), p_2(x)) = -\ln \int_{-\infty}^{\infty} \sqrt{p_1(x)p_2(x)} dx \quad (1)$$

$0 \leq \int_{-\infty}^{\infty} \sqrt{p_1(x)p_2(x)} dx \leq 1$ を確率として解釈すれば、これは自己情報量となり、単位は [bit] となる。二つの分布がガウス分布で表現されているとき、バタチャリヤ距離は、

$$BD(p_1(x), p_2(x)) = \frac{1}{8} \mu_{12}^T \left(\frac{\sum_1 + \sum_2}{2} \right)^{-1} \mu_{12} + \frac{1}{2} \ln \frac{|(\sum_1 + \sum_2)/2|}{|\sum_1|^{\frac{1}{2}} |\sum_2|^{\frac{1}{2}}} \quad (2)$$

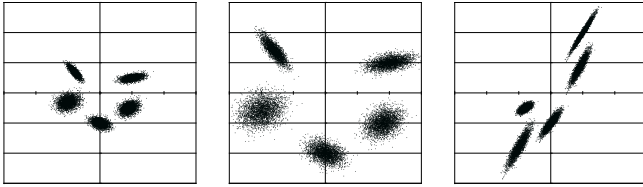


図 1 構造不変の定理

Fig. 1 Theorem of the invariant structure

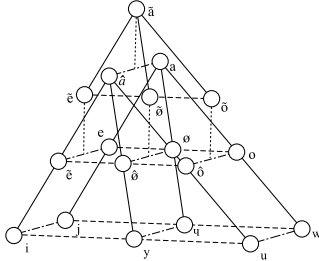


図 2 ヤコブソンによる幾何学的音韻構造

Fig. 2 Jakobson's geometrical structure

となる． μ_{12} は $\mu_1 - \mu_2$ である．このとき，二つの分布に対して共通のアフィン変換 $Ac + b$ をかけた場合，バタチャリヤ距離はその前後で不変である．これは，バタチャリヤ距離が空間を歪める距離尺度であることに起因する．この構造はアフィン変換に対して不変であり，凡そ非言語的特徴に依存しない普遍的な構造となる（音響的普遍構造）．この構造的表象は，構造音韻論の物理実装として位置づけられる．図 2 はヤコブソンによるフランス語における母音及び準母音の幾何学的構造を表したものであり，この構造は話者に対して不変である．

3. 構造的表象に基づく学習者分類

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，音声物理に不可避免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，純粋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 [7]．学習者を構造として記述すること可能となれば，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ができ，学習者間や学習者と教師の間での構造の比較を行うことが可能となる．構造間距離 D は下記の式に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 [9]．

$$D = \sqrt{\frac{1}{M} \sum_{i < j} (P_{ij} - Q_{ij})^2} \quad (3)$$

P_{ij}, Q_{ij} は距離行列の要素， i, j は音素の種類， M は音素数を表す．2 学習者間距離がスカラー量で表現されると，任意の

表 1 母音置き換えの組み合わせ

Table 1 Vowel substitution table

Japanese vowels	English vowels
a	ɑ, ʌ, æ, ɜ, ə
i	i, ɪ
u	u, ʊ
e	ɛ
o	ɔ

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．

構造的表象に基づく学習者の分類は，先行研究 [10] において既に試みられている．しかし，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ず，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価スコアのみを用いた検討にとどまっている．即ち，先行研究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，同一話者の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により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 ($/ɑ, æ, ʌ, ə, ɜ, i, ɪ, u, ʊ, ɛ, ɔ/$) と，日本語母音 5 種類 ($/a, i, u, e, o/$) をそれぞれ収録した．両者の発声を同一話者内で混ぜ合わせることで英語学習者のさまざま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．その典型的な組み合わせ例を表 1 に示す．この組み合わせを用いて，表 2 に示す 8 つの発音状態を定義した．**S1** は全部日本語で置換，**S8** を全部英語発音の状態とし，**S2**～**S7** については，全音素 11 種類のうちの半分（5 個か 6 個）の音素が置換されるような状態を定義した．

収録音声から目視により母音部分を切り出し，表 3 に示す音響分析条件の下でケプストラムパラメータを求め，MAP 推定を用いて分布化した．各母音に対してそれぞれ 1 発声のみを用いている．複数の英語母音を 1 種類の日本語母音で置き換える場合は，日本語母音は異なる発声を用いるものとする．例えば， $/bat/$ と $/bæt/$ を置き換える場合は，2 回の日本語発音

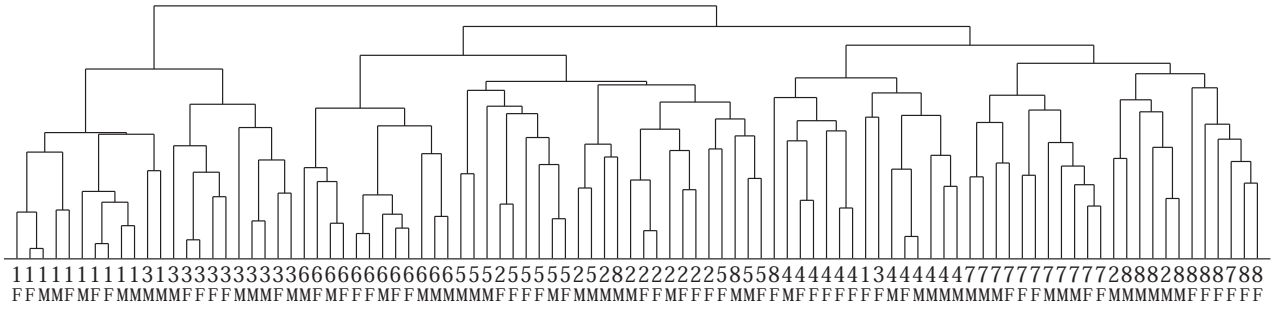


図3 学習者発音構造のクラスタリング
Fig.3 Result of clustering language learners

表2 発音状態の定義

Table 2 Pronunciation states

	α	æ	Λ	ə	σ	ɪ	i	u	u	ε	ɔ
S1	J	J	J	J	J	J	J	J	J	J	J
S2	E	E	E	E	E	J	J	J	J	J	J
S3	J	J	J	J	J	E	E	E	E	E	E
S4	E	E	J	J	J	E	E	J	J	E	E
S5	J	J	E	E	E	J	J	E	E	J	J
S6	E	J	E	J	E	J	J	J	J	E	E
S7	J	E	J	E	J	E	E	E	E	J	J
S8	E	E	E	E	E	E	E	E	E	E	E

E:英語の母音発声を使用, J:日本語の母音発声で置換

表3 音響分析条件

Table 3 Acoustic analysis condition

sampling	16bit / 16kHz
window	25 ms length and 1 ms shift
parameters	FFT cepstrum (1~10)
HMMs	1-mixture monophones with diagonal matrices
topology	3 states and 1 distribution per HMM (GM)

の/bat/を用いてそれぞれの発音を置き換える。話者12名のそれぞれ8つの発音状態において、得られた音素分布に対してバタチャリヤ距離により音素間距離行列を算出し、合計96個の学習者発音構造を得た。これらの発音構造に対して、任意の異なる2構造間の距離を式(3)により算出して学習者距離行列を求め、Ward法によるボトムアップクラスタリングを行うことで学習者発音構造の分類を行った。

3.3 分析結果と考察

クラスタリングによる分類結果を図3に示す。樹形図のリーフノードにおける数字が表2における各状態番号を表しており、M・Fがそれぞれ男性・女性を表している。多くの状態において同一の状態が固まってクラスタを構成していることが確認できる。異なる状態が混合した部分（例えばS2とS5とS8）もいくつか見受けられるが、全体的にはおよそ発音状態による分類が行われていることが確認できる。男女差については、同一状態のクラスタ内では性別によって固まっている部分が見受けられるものの、男女が混合したクラスタも存在するこ

とから、男女の違いによる影響をおよそ排除できていると考えられる。

上記の結果より、学習者構造のクラスタリングにより、発音状態によって分類できていることが確認できた。発音状態の定義や数についてはより詳細な検討が必要であると考えられるが、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、発音状態によって学習者を分類することが可能となることが実験的に示された。一部、状態が混合したクラスタが見られたことについては、各話者で英語を身につけた過程が異なり、それぞれ発音の癖や訛りが異なるため、同じ状態番号でも厳密には同一の発音状態ではないことが原因として考えられる。例えば、全て英語発音であるS8が、S2やS5のクラスタに紛れ込んだ話者については、表2から考えると/i, i, ε, ɔ/の発音が日本語的な発音になっている、といった推測が立てられるが、詳細に関しては当該話者に対するさらなる分析が必要である。

4. 構造的表象に基づく発音教示生成

4.1 学習者分類と発音教示生成

前節にて構造的表象を用いることにより話者や性別などの非言語的特徴をそぎ落とし、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した。実際の英語学習者音声データを用いてこのような発音状態による分類を行うことが出来れば、日本人英語学習者の典型的な発音状況をいくつかの状態として定義することが可能となり、更には、各状態に属する学習者に対してどの音素の発音を矯正するのが目標に到達するためには効果的であるのかという発音教示が可能となる。

本分析では、学習者の発音構造と目標とする発音状態における発音構造との間の構造歪みに基づいて、目標に近づくために最も効果的であるような音素を次に矯正すべき対象として学習者に提示する。前節で行った発音分類を基に、ある状態における典型的な発音構造を定義し、それらの比較を行うことで発音状態遷移に対する教示生成を実験的に検討し、その有効性を確かめることを目的とする。

4.2 構造間の要素差異に基づく構造歪みの導出

2つの構造間において、ある音素に起因する構造歪みを下記のように定義する。2名の話者 S, T の発音構造（即ち音素間距離行列）をそれぞれ $C^S = \{c_{ij}^S\}, C^T = \{c_{ij}^T\}$ とする。 i, j

($1 \leq i, j \leq M$, M は音素数) は音素種類を表し, 行列の要素 c_{ij}^S は話者 S における音素 i と音素 j との音素間距離を表す. 2 つの構造間の各要素の差異に基づいて, 2 構造間の音素 p に起因する構造歪み $d(p)$ を下記の式により導出する.

$$d(p) = \sum_{j=1}^M |c_{pj}^S - c_{pj}^T| \quad (4)$$

学習者発音構造と目標とする発音構造の間の構造歪みを各音素に対して算出した場合, より大きな構造歪みを呈する音素が矯正すべき対象であるとみなすことができる.

4.3 分析手法

本節では, ある発音状態における典型的な構造を定義し, これらの間の構造歪みを求めることで, 状態遷移に対する発音指示を実験的に検討する. 第 3. 節と同様に発音状態を既知とするため, 英語母音と日本語母音を混合して学習者の発音状態を模擬した. 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, 英語母音 11 種類 ($/a, \ae, \Lambda, \ə, \varepsilon, i, i:, u, u:, \epsilon, \upsilon/$) と, 日本語母音 5 種類 ($/a, i, u, e, o/$) について, $/bVt/$ の発声を収録した. 発音状態定義については 3.2 節と同様である (表 2). 3.2 節と同様の手順により, 12 名の各 8 状態に対して発音構造を抽出し, 同一状態における 12 話者の平均をとった構造を各状態における典型的な構造として用いた. **S1**~**S7** の状態から **S8** 状態への遷移に対して, 各状態における平均構造間の構造歪みを式 (4) により各母音毎に算出することで, 矯正すべき母音の順序を自動推定した. なお, 自動推定に用いた情報は 2 つの距離行列のみであり, どのような母音置換が行われているかという情報に関しては一切用いていない.

4.4 分析結果と考察

状態 **S1**~**S7** から状態 **S8** への遷移の分析結果を図 4 の (a)~(g) に示す. 縦軸は各母音に起因する構造歪みを表す. 灰色で色づけされているものは日本語の母音にて置き換えられている母音である. まず, (a) の **S1**~**S8** 遷移について見てみると, **S1** 状態は全て日本語母音にて置換されている状態であるが, 母音の種類によって構造歪みに差があるのが分かる. 特に日本語の母音とは音色が異なり, 日本人にとって発音が苦手な母音である $/\varepsilon, \ae, \ə/$ や $/i/$ などが高く見積もられている. 逆に $/\epsilon, i:, \upsilon, u:/$ などは, 日本語母音で代用することの代償は比較的小さいといえる. (b)~(g) については, 日本語母音にて置き換えられているものが大きな構造歪みの値を示していることが分かる. 一方で, $/u, \upsilon/$ の 2 つに関しては全体的に低い値となっており, (b), (d), (f) のような日本語母音と置き換えられた場合でもそれほど高い値とならない. これは (a) に見られたように $/u, \upsilon/$ は日本語的な音色による発音でもそれほど致命的な発音誤りにはならないと解釈することができる.

上記の結果から, 日本語母音の置き換えに関する情報をまったく用いることなく, 純粋に学習者の発音構造と目標とする構造のみに基づいて, 2 構造間の構造歪みを算出することで, 音色的に致命的な発音誤りを提示することが可能であることが実験的に示された. 本分析では, 構造歪みとして 2 構造間の要素差異の総和を用いたが, 単なる総和ではなく分散なども考慮し

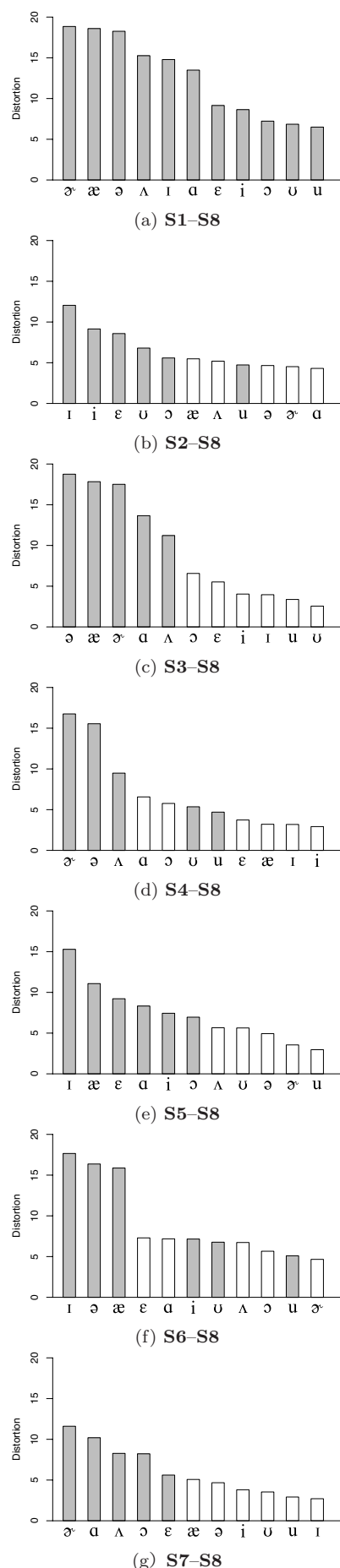


図 4 自動推定された母音矯正順序

Fig. 4 Automatically estimated order of vowel correction

た, より発音教示に適した構造歪み算出手法を検討する必要があると考えられる。

5. ま と め

本報では, 構造的表象を用いた学習者の分類と, その分類に基づいて学習者に対して発音教示の自動生成を行った。前者の分析では, 同一話者における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った。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した。さらに後者の分析では, ある発音状態における構造と目標とする発音状態との構造との間の構造歪みに基づいて, 目標と比較して最も大きな構造歪みを呈する母音を次に矯正すべき母音として学習者に提示することで, 学習者に対する発音教示が可能となることを示した。

今後, 実際の英語学習者音声を多量に収集し, 本手法により発音状態による分類を行うことが出来れば, 日本人英語学習者に対して, いくつかの典型的な発音状況を定義することが可能となると考えられる。既に, 本学柏キャンパスにおけるキャンパス公開などの場を利用し, 様々な学習暦を持つ日本人 200 名以上の音声収録を行っている。これらのデータの分析結果は別の機会に報告する予定である。なお, 英語発音分類は日本人学習者に限った話ではなく, 世界中に存在する全ての英語発音(ただし, 同一の方言を学習している学習者が対象となる。たとえば米語の学習者の発音)を分類することも原理的には可能であるといえる。典型的な発音状態を定義することができれば, 学習者が目標へ到達するまでの状態遷移の道筋と, 具体的にどの音素を矯正すべきかという学習計画の提示などが可能になる。さらには, 実際のトレーニング(ビデオ教材や英語教師による訓練など)と発音状態繊維との対応データを収集しデータベース化することで, 学習者に対して目標への道筋とそのレッスンメニューを提案することも可能になると考えられる。今後, 実際の英語学習者による発音を用いて分析を行い, 英語教師との議論を通して本手法の有効性を更に検討していく必要があると考えられる。

文 献

- [1] H. A. Gleason, *An introduction to descriptive linguistics*, New York: Holt, Rinehart & Winston (1961)
- [2] N. Minematsu, "Mathematical evidence of the acoustic universal structure in speech," Proc. Int. Conf. Acoustics, Speech, & Signal Processing (ICASSP'2005), pp.889–892 (2005)
- [3] 峯松信明, 松井健, 広瀬啓吉, "構造音韻論の物理実装に基づく新しい音声の音響的表象", 電子情報通信学会技術研究報告, SP2004-27, pp.47–52 (2004)
- [4] 峯松信明, 西村多寿子, 西成活裕, 櫻庭京子, "構造不変の定理とそれに基づく音声ゲシュタルトの導出", 電子情報通信学会技術研究報告, SP2005-12, pp.1–8 (2005)
- [5] N. Minematsu, "Pronunciation assessment based upon the phonological distortions observed in language learners' utterances," Proc. Int. Conf. Spoken Language Processing (ICSLP'2004), pp.1669–1672 (2004)

- [6] N. Minematsu, "Pronunciation assessment based upon the compatibility between a learner's pronunciation structure and the target language's lexical structure," Proc. Int. Conf. Spoken Language Processing (ICSLP'2004), pp.1317–1320 (2004)
- [7] S. Asakawa, N. Minematsu, T. I. Jaakkola, and K. Hiorse, "Structural representation of the non-native pronunciations," Proc. European Conf. Speech Communication and Technology (EUROSPEECH'2005), pp.165-168 (2005)
- [8] M. Pitz and H. Ney, "Vocal tract normalization as linear transformation of MFCC," Proc. European Conf. Speech Communication and Technology (EUROSPEECH'2005), pp.1445–1448 (2003)
- [9] 峯松信明, 志甫淳, 村上隆夫, 丸山和孝, 広瀬啓吉, "音声の構造的表象とその距離尺度", 電子情報通信学会技術研究報告, SP2005-13, pp.9–12 (2005)
- [10] 峯松信明, "音声の音響的普遍構造の歪みに着眼した外国語発音の自動評定", 電子情報通信学会技術研究報告, SP2003-180, pp.31–36 (2004)