

音声の構造的表象に基づく英語学習者発音の音響的分析

朝川 智^{†a)} 峯松 信明[†] 広瀬 啓吉^{††}

Acoustic Analysis of the Pronunciation of Japanese Learners of English
Based on Structural Representation of Speech

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

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的かつ静的な非言語性ひず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．音声の絶対的な物理実体を捨象し，音声事象間の相対的な関係（差異）のみに基づいて音声を構造的に表象することにより，音響の実体に対して不可避的に混入される非言語的特徴を表現する次元を消し去り，従来とは全く異なる新しい音声アプリケーションが可能となる．本論文では，発音学習応用を目的として構造的表象に対する種々の基礎検討を行う．英語強勢・弱勢母音の分析，少量データ時における発音の構造化，構造的表象に基づく学習者分類と発音教示の自動生成を通して，発音学習応用における構造的表象の有用性を実験的に検討する．

キーワード 音声の構造的表象，日本人英語，強勢・弱勢母音，学習者分類，発音教示生成

1. ま え が き

音声には，その生成の際に話者の声道形状の特性，収録や伝送の際には音響機器の特性，更には聴取の際には聴取者の聴覚特性，といった非言語的特徴が不可避的に混入する．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が，スペクトルは上記の非言語的特徴によって不可避的にひずんだものであり，このひずんだ物理表象に基づいて構築されたシステムは，不可避的に不一致問題（mismatch problem）を抱えることになる．その結果，「システムがユーザを選ぶ」という状況が発生する．応用分野として医学や教育を考えた場合，これはまず第一に避けるべき状況であることは自明である．これまで，この不一致問題を回避するために，大規模データベースによる不特定話者音響モデルに対して話者適応技術を導入することで解決を図ってきたが，発音評価

という応用を考えると，話者（学習者）への適応は粗悪な発音の評価点を相対的に上げる方向へ働く問題も生じる．これは，話者の違いと発音の善しあしがスペクトル包絡という同一の物理現象によって表象されること，すなわち，両者を分離する術を音声工学が持ち合わせていないことが根本的な原因である．音声認識技術を応用した CALL（Computer Aided Language Learning）システム構築に関する研究もこれまで広く行われてきたが [1]～[3]，その一方で，CALL システムの不安定性も指摘されており，教育への導入に対して懐疑的な報告も聞かれている [4]．

音素の定義を，語の意味から切り離して言語学に求めた場合，2 種類の定義が導かれる [5]．1) A phoneme is a class of phonetically-similar sounds and show certain characteristic patterns of distribution in the language or dialect under consideration. 2) A phoneme is one element in the sound system of a language having a characteristic set of inter-relations with each of the other elements in that system. 第一の定義が不特定話者音響モデルを導い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，不可避的に混入する非言語的特徴を消し去ることはできない．一方，第二の定義は「音素は他の音素と

[†] 東京大学大学院新領域創成科学研究科，柏市
Graduate School of Frontier Sciences, The University of
Tokyo, 5-1-5 Kashiwanoha, Kashiwa-shi, 277-8561 Japan

^{††} 東京大学大学院情報理工学系研究科，東京都
Graduate School of Information Science and Technology,
The University of Tokyo, 7-3-1 Hongo, Bunkyo-ku, Tokyo,
113-0033 Japan

a) E-mail: asakawa@gavo.t.u-tokyo.ac.jp

の関係によって初めて定義される対象」であるとしており、言語学においてはこの音素の相対的定義こそが第一義であるといわれている [6], [7]。近年、不可避免地に混入する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 [8]。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、音素の相対的定義に基づいた音響モデリングである。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 [9]。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、認知心理学的には音声ゲシュタルトとして解釈される [7]。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 [10], [11]。

本論文では、上記の先行研究を踏まえ、発音学習応用を目的として構造的表象に対する種々の基礎検討を行う。まず 2. において、先行研究として音声の構造的表象について述べる。3. では、英語における強勢弱勢母音の生成に着目して、構造のサイズに対する物理的解釈を検討する。4. では、少量の発声データからの発音構造化の際に問題となる分布推定の不安定性を解消し、データ量が限られている場合においても学習者発音状態の記述と発音状態変化の追跡が可能となることを実験的に示す。5. では、複数の話者・複数の発音状態における発音構造に対してクラスタリングを行い、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の分類が可能となることを述べる。更にそれを踏まえて 6. では、学習者の発音構造と目標とする発音構造との間の構造ひずみに基づいて、目標に近づくために最も効果的な音素を次に矯正すべき対象として学習者に提示する、自動発音教示生成について述べる。最後に、7. にて結論と今後の課題を述べる。

2. 音声の構造的表象

2.1 音声に不可避免地に混入する非言語的特徴

音声の物理的実体に混入するひずみとしては、主に加算性雑音、乗算性ひずみ、線形変換性ひずみの 3 種類に分類される。このうち、音声に「不可避免地」混入するものは乗算性ひずみ、線形変換性ひずみの二つである。加算性雑音は、時間軸上の加算、すなわ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得る雑音であり、テレビ・ラジオなどの背景雑音がその典型例といえる。これらは場所を移動するなどの対策

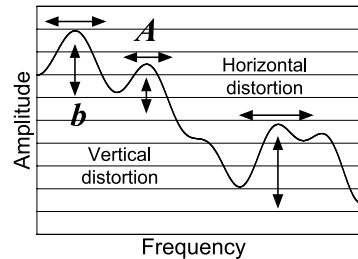


図 1 A, b に起因する 2 種類のスペクトルひずみ
Fig. 1 Two types of spectrum distortion caused by A and b .

をとることで物理的に抹消することが可能であり、不可避免的な雑音ではないため、ここでは扱う対象とはしない。

乗算性ひずみは、スペクトルに対する乗算で表現されるひずみであり、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。マイクロホ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ひずみとして扱われる [12]。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ひずみであるといえる。

線形変換性ひずみは、 c に対する行列 A の乗算 $c' = Ac$ で表現されるひずみである。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 [13], [14]。すなわち、声道長の差異、聴覚特性の差異は近似的に線形変換性ひずみとして扱うことができる。これらも不可避免的なひずみである。

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地に混入し、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。図 1 は、アフィン変換 $c' = Ac + b$ が対数スペクトルに与える影響を示したものである。対数スペクトルの水平変化は A の乗算として、垂直変化は b の加算として記述される。これら A, b が話者や収録環境ごとに変化し、様々なひずみを生むことになる。

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、 ${}_NC_2$ 個のすべての 2 点間距離を求めると、その N 点で張られる構造は一意的に規定される（ただし、鏡像としてのあいまい性は残る）。すなわち事象群に対して、すべての 2 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ひずむ。なぜなら、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。この不可避免的にひずむ構造は、事象群が観測される空間をひず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、すべ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、ヘリング距離などを用いた場合、各分布に対して単一の任意のアフィン変換を施しても、二分布間距離は不変である。すなわち距離行列は不変であり、その結果、構造も不変となる

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。

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

$0 \leq \int_{-\infty}^{\infty} \sqrt{p_1(x)p_2(x)} dx \leq 1$ 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。また上式は、二つの分布の相関値に対して対数をとった値としても解釈できる。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、

$$BD(p_1(x), p_2(x)) = \frac{1}{8} \mu_{12}^T \left(\frac{\Sigma_1 + \Sigma_2}{2} \right)^{-1} \mu_{12} + \frac{1}{2} \ln \frac{|(\Sigma_1 + \Sigma_2)/2|}{|\Sigma_1|^{\frac{1}{2}} |\Sigma_2|^{\frac{1}{2}}}$$

となる。 T は転置を、 $|\cdot|$ は行列式を表す。 μ_1, Σ_1 (μ_2, Σ_2) は $p_1(x)$ ($p_2(x)$) の平均ベクトル及び分散共分散行列であり、 μ_{12} は $\mu_1 - \mu_2$ を表す。このとき、二つの分布に対して共通のアフィン変換 $Ac + b$ を掛けた場合、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ゆがめる距離尺度であることに起因する [7]。この構造はアフィン変換に対して不変であり、およそ非言語的特徴に依存しない、どの話者どのマイクを用いたとしても同様に観測される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。構造音韻論は同一の幾何学的構造が話者を問わず普遍的に存在すると主張する。例えば、ヤコブソンはフランス語の母音と準母音に対して、図 2

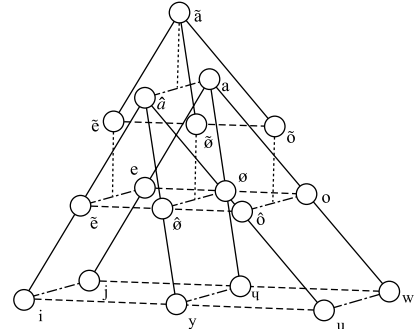


図 2 ヤコブソンによる幾何学的音韻構造
Fig.2 Jakobson's geometrical structure of French vowels and semi-vowels.

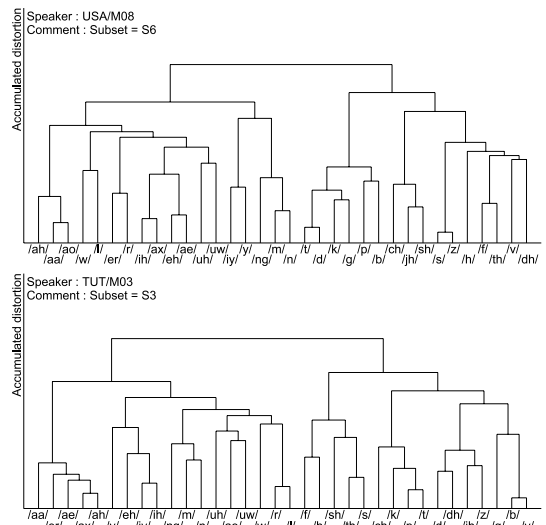


図 3 母語話者と日本人話者による英語音素樹形図
Fig.3 Tree diagrams of an American English speaker and a Japanese English speaker.

のように幾何学的に構造化を行っている [15]。構造的表象は、このような幾何学的音韻構造を非ユークリッド空間において実装することに相当する。

2.3 構造的表象に基づく学習者モデリング

上述の音響的普遍構造を語学学習者音声に適用した場合、年齢・性別などの話者性、マイク・伝送路特性などの音響特性を消し去ることができ、学習者の発音状態を記述することが可能となる。[10] では、日本人英語読上げ音声データベース (ERJ, English Read by Japanese) [16] を用いて、全 222 名の話者の英語発音データに対し、提案手法により音素間距離行列を算出し、ボトムアップクラスタリングにより樹形図として視覚化を行っている。図 3 は母語話者 (上) と日本人

学生（下）の英語音素樹形図の一例である．より低い位置で統合しているノード同士が音響的に近いと判定された音素対である．母語話者の発音樹形図を見ると，子音はほぼ有声/無声の組を作っており，それらは閉鎖音・破擦音・摩擦音の三つの部分木におよそ分かれているのが確認できる．更に，全体を大きく二つに分けると，左側には母音及び鼻音・流音・渡り音，右側には閉鎖音・破擦音・摩擦音が属しているのが分かる．鼻音・流音・渡り音は母音と非常に多くの音声学の共通点をもつことが示されており [17]，上記の分析結果は音声学的に見ても正しい構造であるといえる．それに対して，日本人話者の樹形図では，日本語のアに似た音である Low vowel (/aa, ah, ae/) や schwa (/ax/) などが固まっている様子や， /r/ と /l/， /v/ と /b/， /f/ と /h/， /s/ と /th/ など，日本人英語に見られる癖を見事に表現しているのが分かる．

3. 構造的表象に基づく英語強勢・弱勢分析

3.1 構造サイズの物理的解釈

本節では，構造のサイズの物理的解釈を検討するために，英語母音の強勢弱勢生成に着目する．英語は日本語と異なり，強勢弱勢の属性と母音のカテゴリーとが独立ではなく，弱勢としてのみ出現する母音，ほぼ強勢となる母音などが存在する．弱母音 (schwa, /ə/) は弱勢としてのみ現れる母音である．図 4 に，英語における母音図 (左) と，母語話者音声より抽出した母音間距離行列を多次元尺度構成法 (MDS, Multi Dimensional Scaling) により視覚化したもの (右) を示す．以下では，母音は単母音 (monophthong) のみを扱う．schwa は母音図中，すべての母音の中心に位置している，すなわち，舌の位置としては全母音の中心に位置していることが分かる．また，schwa は断面積一定の音響管で近似される母音である [18]．英語音

声学によると，母音は弱勢になると schwa に近づくとされている [19]．これらの事実は，schwa が最も基本的，かつ，その発声に最も労力を要しない母音であること，そして，弱勢化された母音は互いの距離が短くなることを意味し，これらを構造的表象に当てはめて考えると，弱勢母音群によって張られる構造のサイズはより小さくなることが考えられる．強勢母音はより明りょうに発声して母音間の差異を明確化したものであり，弱勢母音は不明りょう化したものであると考えれば，構造のサイズが発話者の調音努力 (いかに明りょうに，他と区別して発声しようとしたか) を記述していると考えられる．

3.2 母語話者における強勢・弱勢母音構造分析

上記の考察を定量的に検討するため，母語話者音声データを用いて強勢母音群・弱勢母音群の構造サイズ分析を行った．米国母語話者 2 名 (男性 1 名，女性 1 名) の音声データ (男性 746 文，女性 709 文) に対して，データ中のすべての母音に強勢・弱勢ラベリングを施し，そのラベルを用いて強勢母音・弱勢母音を別々に学習して表 1 の分析条件において音素 HMM を構築した．強勢と弱勢でデータ量に大きな偏りが見られた母音と二重母音については分析対象から除外し， /ɑ, æ, ʌ, ɜ, ɪ, i, ʊ, u, ε/ を対象として分析を行った．得られた音素 HMM より，強勢母音・弱勢母音それぞれの音素セットに対して音素間距離行列を求め，Ward 法を用いたボトムアップクラスタリングにより樹形図を求めた．Ward 法によるボトムアップクラスタリングは距離行列を視覚化するための一手法であるが，統合により生じる累積ひずみが最小となる 2 要素を順次統合していく方法であり，最終的に得られる樹形図の高さは全要素を一点で代表させたときのひずみとなる．これはコードブックサイズが 1 のときの VQ ひずみに相当し，音素群を構造としてみた場合の構造の半径に相当する量として解釈することができる．これを構造のサイズとみなして定量的な検討を行った．また比較のため，強勢・弱勢の区別をせずに学習した通常の

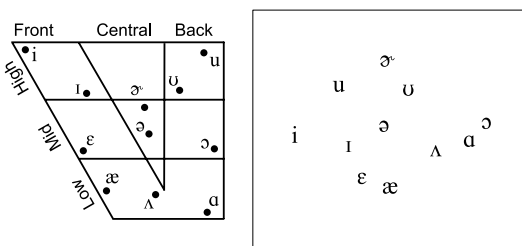


図 4 英語母音図と MDS による母音構造の視覚化
Fig. 4 The vowel chart of American English and an MDS chart obtained from vowel examples.

表 1 音響分析条件 (3.2)
Table 1 Acoustic analysis condition (3.2).

sampling	16 bit / 16 kHz
window	25 ms length and 10 ms shift
parameters	Mel cepstrum (0~12)
HMMs	1-mixture monophones with full matrices
topology	3 states and 1 distribution per HMM (GM)

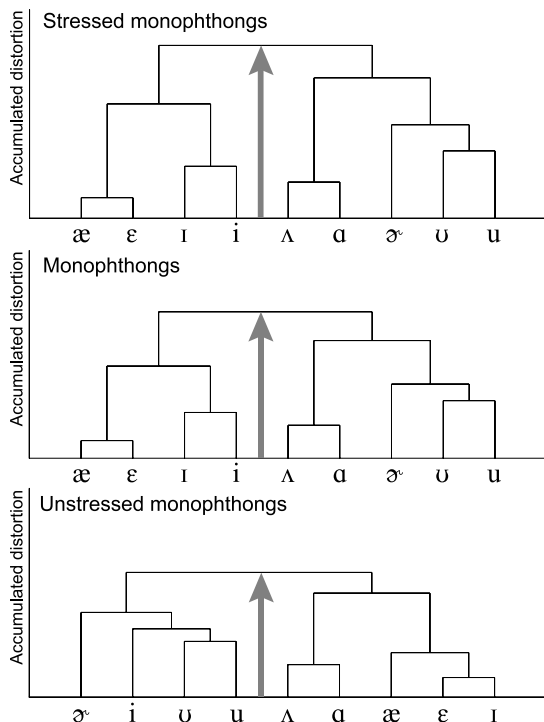


図 5 英語母音における樹形図

Fig. 5 Tree diagrams of the American English vowels.

HMM も作成して同様に構造サイズの算出を行った。

3.3 分析結果と考察

女性話者における結果を図 5 に示す。上から、強勢母音群・すべての母音・弱勢母音群における樹形図である。全母音で構築した樹形図に比べて、強勢母音のみの樹形図はより大きく、弱勢母音ではより小さくなっていることが確認された。強勢母音ではほぼ同じ形をしているのに対し、弱勢母音の場合では若干異なる形となっている。強勢と弱勢とで形が異なることに関しては、強勢母音と弱勢母音における前後の音素環境の違いなどが理由として挙げられる。また、二人の話者における強勢・弱勢母音群でのそれぞれの構造サイズを表 2 に示す。強勢母音での構造サイズは弱勢母音のサイズに比べて男性話者では約 1.2 倍、女性話者においては約 1.4 倍になっていることが分かる。また、全母音の場合の構造と比較しても、強勢母音群構造はより大きく、弱勢母音群構造はより小さくなっているのが分かる。

以上の結果より、英語母音の強勢弱勢が構造のサイズとして表現されることが示された。母音の弱勢化が

表 2 強勢・弱勢母音群構造における構造サイズ
Table 2 Size of the stressed and unstressed structures of the vowels.

	強勢母音	弱勢母音	全母音
男性話者	2.7	2.2	2.3
女性話者	3.1	2.2	2.6

schwa 化であることを考えると、3.1 で推測したように、弱勢化は「より楽な母音発声」化であると解釈できる。言い換えれば、より大きな調音努力をはらって母音間の差異を明確化したのが強勢母音であり、不明りょう化したのが弱勢母音である。その意味において、構造のサイズは「調音努力」を指すと解釈できる。

3.4 構造サイズをもとにした強勢・弱勢生成の自動評価

英語の弱勢母音とは、より小さな調音努力によって発声された母音であり、schwa により近づき、構造サイズがより小さくなるような母音であることが前節の分析により確認された。一方で日本語の母音体系においては、中心母音が存在せず英語らしく強弱の区別をつけて発音することは日本人の英語学習者にとっては非常に難しい。英語発音において、強勢・弱勢の適切な生成は英語特有のリズムを表現する上で非常に重要であり、英語らしい発音の習得が高いハードルとなっている一因であるともいえる [20], [21]。

前節にて、英語における強勢・弱勢母音群における構造サイズが発話者の調音努力を記述していることを実験的に示した。学習者発音に対して構造サイズを算出すれば、これらサイズが「強勢となるべき母音をいかに強勢らしく発声できているか」「弱勢となるべき母音をいかに弱勢らしく発声できているか」を表す指標になると考えられる。次節において、強勢・弱勢構造サイズの比の値を用いて強勢らしさ・弱勢らしさを定量化し、学習者発音における構造サイズ比と各話者の発音習熟度との相関について分析する。

3.5 学習者の構造サイズと発音習熟度との相関分析

音声データとして、ERJ データベース中の文セット 6 (60 文) を読み上げた話者 19 名 (男性 10 名、女性 9 名) のデータを用いた。強勢弱勢母音の構造サイズ算出方法、分析条件については 3.2 と同様である。強勢らしさ・弱勢らしさの定量化には、弱勢サイズに対する強勢サイズの比の値 (強勢サイズ ÷ 弱勢サイズ) を用いる。この値が 1.0 より大きければ大きいほど、強勢は強勢らしく、弱勢は弱勢らしく区別をつけて発

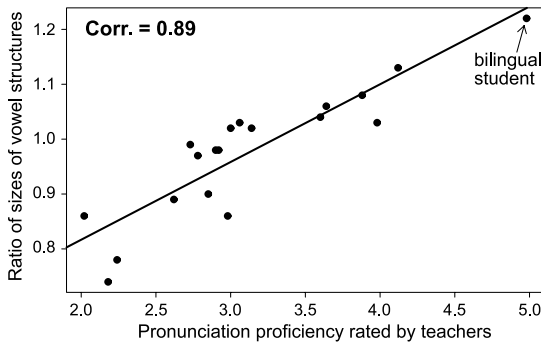


図 6 構造サイズ比と英語教師による評定との相関
Fig. 6 Correlation between structure sizes and human scores.

音できている，ということになる．発音習熟度に関しては，ERJ データベースには分節の特徴に着眼した評定スコア（5 点が最高，5 人の米国人英語教師によるスコアの平均値）が付けられており，このスコアを発音習熟度として用いることとした．なお，ERJ データベースにはリズム文読上げに対する評定スコアもあるが，この評定値は，強勢シラブル出現に関する等時性への着眼も含まれる．本分析手法では母音の継続長は反映される（長母音ではスペクトルが安定し，分散が小さくなるため，結果として分布間距離が大きくなる．よって，短母音構造に比べて長母音構造はそのサイズが大きくなる）ことが確認されている [22] もの，その等時性については明確な接点をもたない．また，母音の schwa 化というのは明確な音色の変化を伴うものであり，ここでは分節の特徴に基づく評定値を用いた．

3.6 分析結果と考察

結果を図 6 に示す．横軸は英語教師による評定スコア，縦軸が強勢弱勢構造サイズ比である．図より，発音習熟度が高い話者ほどサイズ比が大きくなっている傾向が確認できる．分析した 19 名の話者中で，強弱構造サイズ比が最も高い約 1.2 という値の話者がいるが，ERJ データベース 202 人の中にいた唯一のバイリンガル話者である．発音習熟度とサイズ比の相関係数は 0.89 となり，比較的高い相関があることが確かめられた．習熟度の低い話者においてはサイズ比が 1.0 を下回っており，比較的低いばらつきが大きい．また，習熟度が低いにもかかわらずサイズ比が大きくなるという逆転が一部見受けられる．参考までに，ERJ データベース中の母語話者 4 名の音声データに対して同一条件下で分析したところ，構造サイズ比は平均で 1.17

となった．

上記の結果より，発音習熟度の高い話者ほど，強勢を強勢らしく・弱勢を弱勢らしく，強弱の区別を付けて発音ができている，習熟度とサイズ比の間に高い相関があることが確認できた．低習熟度話者において，サイズ比が 1.0 を下回っている話者が見られるが，機能語（前置詞や冠詞など）のような弱く発音すべき単語は，よく知っている単語ということもあり逆に強く発音してしまう，といった原因が考えられる．また，低習熟度話者におけるばらつきや習熟度とサイズ比の逆転現象については，現時点の分析方法では低習熟度話者において頻発すると考えられる母音挿入に対応できていないことが原因の一つではないかと考えられる．また，文単位の発声における強勢の位置は発話者の意図によって変わり得るが，今回用いたデータでは発話者の意図，すなわち文として読んだときの強勢位置までは統制されていないことも原因として挙げられる．今後，日本人英語によく見られる母音挿入誤りなどに対応した分析手法の確立と，発話意図の統制された音声データによる分析が必要であると考えられる．

4. 少量データによる発音構造化

4.1 少量データ時の問題とその解決策

先行研究 [10], [11] においては，英語学習者の発音状態の記述に対して，約 60 文程度の音声データを必要としていた．しかし，発音状態を見るたびにこのような多量の音声データを収録することは学習者の負担も大きく，学習者のモチベーション低下を招く可能性がある．より少ないデータ量で発音の構造化が可能となれば，学習者の負担の軽減だけでなく，数単語の発声から即座に構造化，視覚化を行うことができるようになり，よりインタラクティブな発音学習アプリケーションが可能となる．更には，発音学習以外への応用においても有用であると考えられる．

データ量が限られている場合には，音響事象分布推定における安定性が問題となる．最尤（ML, Maximum Likelihood）推定は分布の推定手法として広く用いられているが，用いるデータ量が少ない場合には，不適切な分布を推定する可能性があり，その安定性に問題が生じる．したがって，少量データ時には音響事象分布の推定に ML 推定を用いることは適切ではない．この問題に対処するため，最大事後確率（MAP, Maximum a Posteriori）推定による音響事象分布推定が検討されている [23]．本章では，MAP

表 3 音響分析条件 (4.2)

Table 3 Acoustic analysis condition (4.2).

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

推定の安定性を検証し、少量データ時における発音構造化の有用性を検討する。MAP 推定に基づく構造化の詳細に関しては [23] の手法を踏襲する。

4.2 最大事後確率推定による音声事象分布推定の安定化

発音状態構造化の分析に先立ち、ML 推定と MAP 推定における安定性を比較するため、日本語 5 母音の孤立発声を用いて予備分析を行った。音声資料として、話者 8 名 (男性 4 名、女性 4 名) より孤立的に発声した日本語 5 母音系列 /a/-/i/-/u/-/e/-/o/ をそれぞれ 5 回収録したものをを用いた。表 3 に示す分析条件にてケプストラムを抽出し、各母音に対して中心前後 14 フレーム (140 ms) を用いて分布化した。分布推定手法として ML 推定と MAP 推定とを用いる。1 回の発声より得られた 5 母音の各分布に対して分布間のパターン距離を計算し、距離行列 (すなわち 5 母音によって張られる構造) を求め、多次元尺度構成法により視覚化した。また、両分析方法の違いを定量的に評価するため、話者 8 名に対して同一話者内における五つの構造より任意の二つの構造間差異を算出し、それぞれの分析手法での構造間差異に対して、分散分析により有意差検定を行った。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 (A の乗算)・シフト (b の加算) させて他方の構造と合わせたときの対応する 2 点間距離の最小値に近似的に比例する [24]。

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

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。

分散分析の結果を図 7 に示す。構造間差異の平均は ML, MAP においてそれぞれ 1.79, 0.724 であった。また分散分析の結果, $F_{(1,158)} = 225, p < 0.001$ となり, ML 推定と比べて MAP 推定では構造間差異が有意に小さいことが確認された。構造の視覚化を行っ

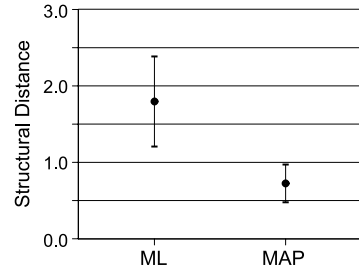
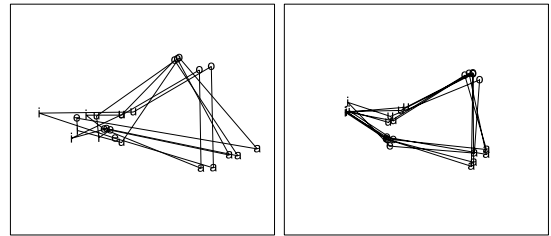


図 7 ML 推定・MAP 推定における話者内構造間差異

Fig. 7 ML-based and MAP-based intra-speaker structural distances.



(a) ML 推定

(b) MAP 推定

図 8 多次元尺度構成法による母音構造の視覚化

Fig. 8 Visualization of the vowel structures with MDS.

た結果の一例を図 8 に示す。各分布推定手法に対して、同一話者の 5 回の発声より得られた五つの距離行列を多次元尺度構成法により二次元平面に写像したものである。MAP 推定を用いることで、発声ごとのばらつきが抑えられているのが分かる。以上の結果より、MAP 推定を用いることで、少量データの場合でも安定した発音構造化が可能となることが確認された。

4.3 少量データによる発音構造化と発音状態遷移の追跡

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、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ひずみが観測されることが考えられ、外国語学習とはひず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

本分析では、英語母音を対象として、少量データにおいて学習者の発音が上達する様子、すなわちひずんだ音韻構造から正常な音韻構造への遷移の様子を記述・追跡することが可能となることを実験的に示す。本来であれば、発音上達の様子を実際の英語学習者から収録するのが理想ではあるが、発音状態を適切に記述・

表 4 収録に用いた英単語
Table 4 Words used for recording.

Words	vowels
bot	ɑ
bat	æ
but	ʌ
botanical	ə
bird	ɜ
bit	ɪ
beat	i
good	ʊ
boot	u
bet	ε
bought	ɔ

表 5 母音置換の組合せ
Table 5 Vowel substitution table.

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

表 6 発音状態の定義 (4.3)
Table 6 Pronunciation states (4.3).

	ɑ	æ	ʌ	ə	ɜ	ɪ	i	ʊ	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	E	E	E	E	E	E	E	J	J	J	J
S4	E	E	E	E	E	E	E	E	E	J	J
S5	E	E	E	E	E	E	E	E	E	E	J
S6	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

追跡できているかどうかを検討するためにはそれぞれの発音状態が既知である必要がある。本分析では技術的な有効性を確かめることに主眼を置き、音声データとしては発音上達の様子を下記に示すような形で模擬したデータを用いることとした。

英語劇の舞台役者としての経験をもつ日本人成人男性より、11 種類の英語母音 (/ɑ, æ, ʌ, ə, ɜ, ɪ, i, ʊ, u, ε, ɔ/) に対して表 4 に示す単語発声を収録し、5 種類の日本語母音 V (/a, i, u, e, o/) に対しては /bVto/ の発声をそれぞれ収録した。以下の分析では、両者の英語を交ぜ合わせることで英語学習者の様々な発音状態をシミュレートする。日本人学習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。その典型的な例を表 5 に示す。この母音置換の組合せに基づいて、表 6 のように発音状態を定義し、状態 S1 から状態 S6 への遷移により英語母音の発音が上達していく過程をシミュ

表 7 音響分析条件 (4.3)

Table 7 Acoustic analysis condition (4.3).

sampling	16 bit / 16 kHz
window	25 ms length and 4 ms shift
parameters	FFT cepstrum (1~10)
HMMs	1-mixture monophones with diagonal matrices
topology	3 states and 1 distribution per HMM (GM)

レートした。

収録した音声から目視により母音部分を切り出し、表 7 に示す音響分析条件のもと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、/ʌ/と/æ/の二つを置き換える場合は、2 回の日本語としての/a/の発声を用いてそれぞれの発音を置き換える。各発音状態において、得られた音素分布に対してパタチャリヤ距離により音素間距離行列を算出した。距離行列の視覚化には、Ward 法によるボトムアップクラスタリングを用いて樹形図を求めた。

4.4 分析結果と考察

図 9 に六つの状態における発音状態を表す樹形図を示す。S1 の樹形図は日本人的な発音によって構築した木であるが、確かに日本語 5 母音の五つのクラスターに明確に分かれており、日本人発音の癖が現れている様子が確認できる。それに対して S6 の発音木では、図 4 の母音図との整合性もとれており、英語音声学的にも正しい樹形図となっていることが分かる。また、S1 から S6 への変化に着眼すると、確かに樹形図が徐々に変化している様子が見て取れる。例えば、S1 から S2 への変化 (/ɑ, æ, ʌ, ə, ɜ/の矯正) では、日本語の/a/に似た音が固まってしまう日本人の癖が矯正されていることが確認される。S2 から S3 の変化では、/ɪ, i/のクラスターが分かれてくる。S3 から S4 では/u/と/ʊ/の音の違いがより大きくなり、S5 では/æ/と/ε/が近づいているのが分かる。S5 と S6 についてはほとんど違いは見られなかった。

以上の結果より、1 母音に対して 1 発声という限られたデータ量でも発音状態を適切に記述し、学習者の上達の様子を追跡・記録することが可能となることが確認できた。本分析は単一話者のみに対して行われたものではあるが、[23] では 1 発話を単位とする構造化においても話者の違いを表現する次元を消去すること

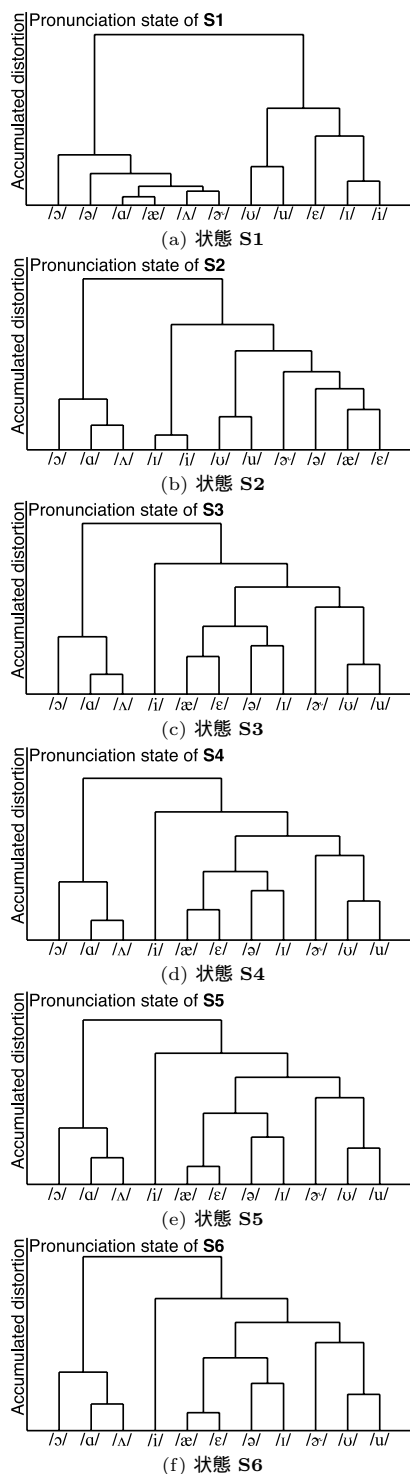


図 9 各状態における発音木

Fig. 9 Six tree diagrams for each pronunciation state.

に成功しており、本手法は他の話者に対しても適用可能であると考えられる。

5. 構造的表象に基づく学習者分類

5.1 学習者の記述・比較・分類

前章では、英語の上達する様子を模擬した音声を用いて、ひず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た。学習者を構造として記述することが可能となれば、二つの構造間の距離尺度を定義することにより 2 学習者間の差異をスカラ量として定義でき、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。2 学習者間距離がスカラ量で表現されると、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。

構造的表象に基づく学習者の分類は、先行研究 [25] において既に試みられている。しかし、読上げ文セットの不均一性や読上げ方などの収録条件による影響を排除できておらず、また、数十文の読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。すなわち、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交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もと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。

5.2 構造的表象に基づく学習者分類

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、11 種類の英語母音（/ɑ, æ, ʌ, ə, ɜ, ɪ, i, u, ɛ, ɔ, ʊ/）に対して表 4 に示す単語発声を収録し、5 種類の日本語母音 V（/a, i, u, e, o/）に対しては /bVto/ の発声をそれぞれ収録した。4.3 と同様に、英語母音と日本語母音を同一話者内で交ぜ合わせることで英語学習者における様々な発音状態を模

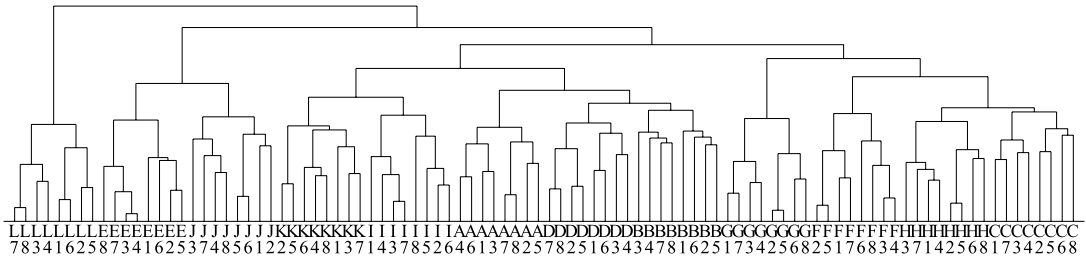


図 10 音響の実体に基づく学習者発音構造のクラスタリング

Fig. 10 Substance-based clustering of the 96 pronunciation structures.

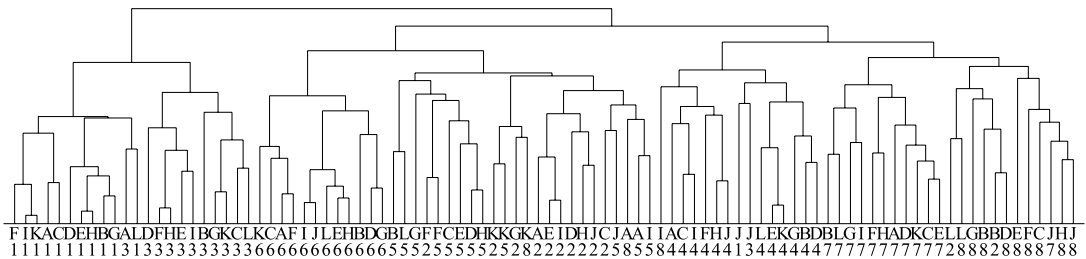


図 11 構造に基づく学習者発音構造のクラスタリング

Fig. 11 Structure-based clustering of the 96 pronunciation structures.

表 8 発音状態の定義 (5.2)

Table 8 Pronunciation states (5.2).

	ɑ	æ	ʌ	ə	ɔ	ɪ	ɪ	u	u	e	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	J	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 9 音響分析条件 (5.2)

Table 9 Acoustic analysis condition (5.2).

sampling	16 bit / 16 kHz
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)

擬する．表 5 の母音置換の組合せを用いて，表 8 に示すような八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 9

に示す音響分析条件のもと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．話者 12 名のそれぞれ八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 (1) 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．

比較のために，各母音の音響の実体同士，すなわちケプストラム分布そのものを直接比較することにより学習者間距離行列を算出し，同様のクラスタリングを行った．これは，DP や HMM で使われる実体の直接比較による距離定義である．音響の実体に基づく学習者間距離は以下のように定義する．

$$D' = \sqrt{\frac{1}{M} \sum_i BD(v_i^P, v_i^Q)} \quad (2)$$

ここで， v_i^P は学習者 P における母音 i のケプストラム分布を表す．

5.3 分析結果と考察

2 種類の学習者間距離に基づくクラスタリング結果を，それぞれ図 10，図 11 に示す．樹形図のリーフノードにおける数字（1～8）が表 8 における各状態番

号を表し、A~L が話者を表している。これら二つの樹形図は、学習者間距離の定義が違うのみで、全く同一の音声データが用いられているが、音響の実体同士を比較した場合（図 10）では完全に話者によるクラスが形成されているのに対して、構造に基づくクラスタリング（図 11）では、およそ同一の発音状態によって各クラスが形成されており、発音状態による分類が行われていることが確認できる。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。これは例えば、英語を学ぶ数億人の学習者を（性別や年齢といった要因を切り離して）話者を単位として分類できることを意味する。このような分析は方言研究に対して応用することも可能である [26]。

図 11 において、発音状態が混合したクラスが一部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なまりが異なることが原因の一つとして考えられる。英語の母音構造は地域方言に強く影響を受ける [27] ため、厳密には、同じ状態番号でも話者ごとに発音状態は異なる。例えば、すべて英語発音である S8 が、S2 や S5 のクラスに紛れ込んだ話者については、表 8 から考えると、/i, i, e, 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更なる分析が必要である。

6. 構造的表象に基づく発音教示生成

6.1 学習者分類と発音教示生成

前章にて、構造的表象を用いることにより話者性をそぎ落とし、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できれば、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ひずみに基づいて、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。前章で行った発音分類をもとに、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。

6.2 構造間の要素差異に基づく構造ひずみの導出

二つの構造間において、ある音素に起因する構造ひずみを下記のように定義する。2 名の話者の発音構造（すなわち音素間距離行列）をそれぞれ P, Q とし、構造間の各要素の差異に基づいて、母音 v に起因する 2 構造間の構造ひずみ $d(v)$ を下記の式により導出する。

$$d(v) = \sum_{i=1}^M |P_{vi} - Q_{vi}| \quad (3)$$

学習者発音構造と目標とする発音構造の間の構造ひずみを各音素に対して算出した場合、より大きな構造ひず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。

6.3 構造的表象に基づく発音教示生成

本節では、ある発音状態における典型的な構造を定義し、これらの間の構造ひずみを求めることで、状態遷移に対する発音教示を実験的に検討する。5.2 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、11 種類の英語母音（/ɑ, æ, ʌ, ə, ɜ, ɪ, i, ʊ, u, e, ɔ/）に対して表 4 に示す単語発声を収録し、5 種類の日本語母音 V（/a, i, u, e, o/）に対しては /bVto/ の発声をそれぞれ収録した。発音状態定義については 5.2 と同様である（表 8）。5.2 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ひずみを式 (3) により各母音ごとに算出することで、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二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。

6.4 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 12 の (a)~(g) に示す。縦軸は各母音に起因する構造ひずみを表す。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、すべてが日本語母音にて置換されている状態である S1 状態における構造ひずみについて見てみると、11

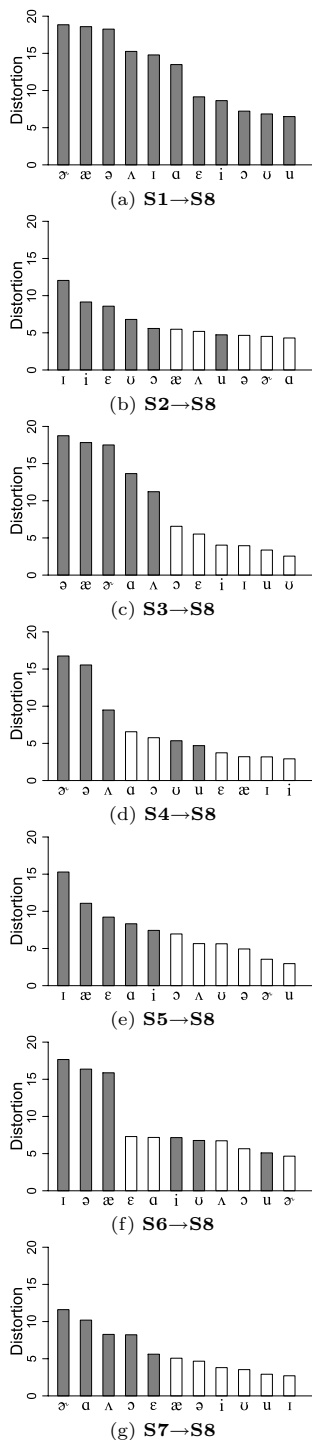


図 12 自動推定された母音矯正順序

Fig. 12 Automatically estimated order of vowel correction

種類の母音を構造ひずみの大きさによって大きく三つのグループに分けることができる。/ɤ, æ, ə/の母音は構造ひずみが大きく、学習の優先度が高いと判定されたグループである。そして、構造ひずみが中程度のグループとして/ʌ, ɪ, ɑ/が、構造ひずみの小さなグループとして/ε, ɨ, ɔ, ʊ, u/が並んでいる。高い優先度を示しているグループに日本人が発音を苦手とする/ɤ, æ, ə/が並んでいるのが確認できる。一方で、構造ひずみの小さなグループである/ε, ɨ, ɔ, ʊ, u/の母音は、それぞれ日本語の/e, i, o, u, u:/と音色的に似ており、これらと置き換えて発音してもコミュニケーションに支障は来さないといわれている[28]。S1 に対する教示結果は音声学的に見ても妥当性の高いものであるといえる。

S2～S7の結果については、日本語母音にて置き換えられているものが大きな構造ひずみの値を示しており、学習優先度が高いと判定されていることが明確に確認できる。S2, S4, S6では、日本語で置換した母音の中でも優先度が低く見積もられている母音も見られるが、/i, ʊ, u/に関しては日本語の/i, u, u:/に音色的に似ており、これらの母音では日本語的な音色による発音でも、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、日本語母音の置換えに関する情報を全く用いることなく、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 構造間の構造ひず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、構造ひずみとして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た、より発音教示に適した構造ひずみ算出手法も可能であろう。

7. む す び

本論文では、近年新しく提案されている音声の構造的表象に対して、発音学習応用を目的として種々の基礎検討を行った。3. では、構造のサイズが調音努力に対応していることを確認した。調音努力の定量化が可能となることにより、3.4 での分析のように学習者発音に対する強勢弱勢生成の自動評定が可能となるだけでなく、構造サイズを正規化することによる「調音努力の正規化」という従来とは全く異なる音声の正規化が可能となると考えられる。4. では、データ量が限られている場合においても安定した発音構造化が可能で

あることを示し、学習者発音状態の記述と発音状態変化の追跡が可能であることを確認した。原理的には、一発声（例えば一単語）に対して構造化を行うことも可能である。既に、構造的表象を音声認識に応用する実験的検討を始めており、高い頑健性を実験的に示している [23]。5. 及び 6. では、非言語的特徴に影響されることなく、発音状態による学習者の分類と、目標に近づくための発音教示生成が可能となることを実験的に示した。

本論文での分析は、構造的表象の技術的な側面に主眼をおいた上で、その有効性の検討を行ったものであり、発音学習における教育効果に関しては今後の検討課題である。本論文での分析結果を踏まえ、構造的表象を用いた発音教育システムを構築し、実際の学習者を対象として運用を行うことで、発音学習教材としての有効性を検討するとともに、各手法の更なる高精度化を図ることが今後行うべきこととして考えられる。

文 献

- [1] 中川聖一, “語学学習における音声言語処理技術の利用,” 信学誌, vol.85, no.12, pp.942–943, Dec. 2002.
- [2] 山本秀樹, “音声メディアを利用した教育システム,” 人工知能誌, vol.17, no.4, pp.452–457, July 2002.
- [3] M.L. Swartz and M. Yazdani, *Intelligent Tutoring Systems for Foreign Language Learning*, Springer-Verlag, New York, 1992.
- [4] A. Neri, C. Cucchiari, and H. Strik, “Automatic speech recognition for second language learning: How and why it actually works,” *Proc. Int. Congress of Phonetic Sciences*, pp.1157–1160, Aug. 2003.
- [5] H.A. Gleason, *An introduction to descriptive linguistics*, Holt, Rinehart & Winston, New York, 1961.
- [6] フェルディナン・ド・ソシュール, 一般言語学講義, 岩波書店, 東京, 1940.
- [7] N. Minematsu, T. Nishimura, K. Nishinari, and K. Sakuraba, “Theorem of the invariant structure and its derivation of speech Gestalt,” *Proc. ISCA Tutorial and Research Workshop on Speech Recognition and Intrinsic Variation*, pp.47–52, May 2006.
- [8] N. Minematsu, “Mathematical evidence of the acoustic universal structure in speech,” *Proc. Int. Conf. Acoustics, Speech, and Signal Processing*, pp.889–892, March 2005.
- [9] ローマン・ヤコブソン, 構造的音韻論, 岩波書店, 東京, 1996.
- [10] N. Minematsu, “Pronunciation assessment based upon the phonological distortions observed in language learners’ utterances,” *Proc. Int. Conf. Spoken Language Processing*, pp.1669–1672, Oct. 2004.
- [11] 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*, pp.1317–1320, Oct. 2004.
- [12] D. Reynolds and L.P. Heck, “Speaker verification: From research to reality,” *Tutorial of Int. Conf. Acoustics, Speech, and Signal Processing*, May 2001.
- [13] M. Pitz and H. Ney, “Vocal tract normalization equals linear transformation in cepstral space,” *IEEE Trans. Speech Audio Process.*, vol.13, no.5, pp.930–944, Sept. 2005.
- [14] 江森 正, 篠田浩一, “音声認識のための高速最尤推定を用いた声道長正規化,” 信学論 (D-II), vol.J83-D-II, no.11, pp.2108–2117, Nov. 2000.
- [15] R. Jakobson and J. Lutz, “Notes on the French phonemic pattern,” *Word*, vol.5, no.2, pp.151–158, 1949.
- [16] 峯松信明, 富山義弘, 吉本 啓, 清水克正, 中川聖一, 壇辻正剛, 牧野正三, “英語 CALL 構築を目的とした日本人及び米国人による読み上げ英語音声データベースの構築,” 日本教育工学会論文誌, vol.27, no.3, pp.259–272, March 2004.
- [17] A.C. Gimson, *An introduction to the pronunciation of English*, Edward Arnold, London, 1980.
- [18] 千葉 勉, 梶山正登, 母音 — その性質と構造, 岩波書店, 東京, 2003.
- [19] 川越いつえ, 英語の音声科学, 大修館書店, 東京, 1996.
- [20] 渡辺和幸, 英語のリズム・イントネーションの指導, 大修館書店, 東京, 1994.
- [21] 竹藪幸生, 日本人英語の科学, 研究社, 東京, 1982.
- [22] N. Minematsu, S. Asakawa, and K. Hirose, “Paralinguistic information represented as distortion of the acoustic universal structure in speech,” *Proc. Int. Conf. Acoustics, Speech, and Signal Processing*, pp.261–264, May 2006.
- [23] T. Murakami, K. Maruyama, N. Minematsu, and K. Hirose, “Japanese vowel recognition using external structure of speech,” *Proc. IEEE Automatic Speech Recognition and Understanding Workshop*, pp.203–208, Nov. 2005.
- [24] 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,” 信学技報, SP2005-13, May 2005.
- [25] 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判定,” 信学技報, SP2003-180, Jan. 2004.
- [26] W. Labov, S. Ash, and C. Boberg, *Atlas of North American English: Phonetics, Phonology and Sound Change*, Walter De Gruyter, Berlin, 2001.
- [27] M. Huckvale, “ACCDIST: A metric for comparing speakers’ accents,” *Proc. Int. Conf. Spoken Language Processing*, pp.29–32, Oct. 2004.
- [28] 英語音声学研究会, 大人の英語発音講座, NHK 出版, 東京, 2003.

(平成 18 年 8 月 30 日受付, 11 月 15 日再受付)



朝川 智 (学生員)

2002 東大・工・電子情報卒．2005 同大大学院情報理工学系研究科修士課程了．現在，同大学院新領域創成科学研究科博士課程在学中．日本学術振興会特別研究員 DC1．日本音響学会会員．



峯松 信明 (正員)

1990 東大・工・電気卒．1995 同大大学院工学系研究科博士課程了．博士(工学)．同年豊橋技術科学大学情報工学系助手．2000 東大大学院工学系研究科助教授．2004 同新領域創成科学研究科助教授．2002 瑞国 KTH 客員研究員．音声分析・認識・合成・応用，音声知覚，音声学・音韻論と，幅広い観点から音声コミュニケーションを研究．日本音響学会，情報処理学会，人工知能学会，日本音声学会，ISCA，IPA，CALICO，EuroCALL 各会員．



広瀬 啓吉 (正員)

1972 東大・工・電気卒．1977 同大大学院工学系研究科博士課程了．工博．同年東大工学部電気工学科講師．1994 同電子工学科教授．1999 同大大学院新領域創成科学研究科教授．2004 年 10 月より同情報理工学系研究科教授．1987 米国 MIT 客員研究員．音声言語情報処理分野一般，特に韻律に着目した研究．IEEE，米国音響学会，ISCA，日本音響学会，情報処理学会，人工知能学会，言語処理学会等各会員．