

音声の構造的表象に基づく非母語話者の英語発音分析

朝川 智[†] 峯松 信明^{††} 村上 隆夫[†] 伊勢井敏子・ヤーッコラ[†] 広瀬 啓吉[†]

[†] 東京大学大学院情報理工学系研究科

^{††} 東京大学大学院新領域創成科学研究科

〒 113-0033 東京都文京区本郷 7-3-1

E-mail: †{asakawa,mine,murakami,tijaakkola,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．本研究では，構造的表象に関して2つの観点から行った分析を報告する．一つ目の分析では，データ量が限られている場合においても構造的表象により学習者の発音状態を記述することが可能であり，さらに学習者の発音が上達する様子を追跡する事が可能となることを実験的に示す．二つ目は，韻律的特徴，特に強勢弱勢母音に着眼し，構造のサイズとの関連について分析する．構造サイズと学習者の発音習熟度との間に高い相関があることを実験的に示し，自動発音評価への応用の可能性について考察する．

キーワード 音声の構造的表象，日本人英語，分布推定，強勢弱勢母音，発音自動評価

Analysis of the non-native English pronunciation based on structural representation of speech

Satoshi ASAKAWA[†], Nobuaki MINEMATSU^{††}, Takao MURAKAMI[†], Toshiko

ISEI-JAACKOLA[†], and Keikichi HIROSE[†]

[†] Graduate School of Information Science and Technology,

^{††} Graduate School of Frontier Sciences, University of Tokyo

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

E-mail: †{asakawa,mine,murakami,tijaakkola,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 new analyses using the proposed method. The first analysis is done to examine whether the method can trace the development of a student's pronunciation appropriately using only a limited amount of speech. The second one focuses on the prosodic aspect of the pronunciation, especially stressed and unstressed vowels. The former indicates that the proposed method can show history of the student's development adequately and the latter clarifies that size of the pronunciation structure is highly correlated with the pronunciation proficiency.

Key words structural representation of speech, Japanese English, estimation of distributions, stressed and unstressed vowels, automatic evaluation of the pronunciation

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、純粹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完全に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]。ただ、先行研究における分析手法では一話者当たり 60 文程度の発話が必要であり、よりインタラクティブなアプリケーションを考えた場合には、より少ない発話で構造的表象を抽出する必要がある。また先行研究では、分節的特徴のみに着眼しており、韻律的特徴に関しては全く考慮されていない。

本研究では上記の問題点を解決すべく、音声の構造的表象に関する 2つの分析を行った。一つ目の分析では、データ量が限られている場合においても構造的表象により学習者の発音状態を安定に記述することが可能であり、さらに学習者の発音が上達する様子を追跡する事が可能となることを実験的に示す。二つ目は、韻律的特徴、特に英語の強勢弱勢母音に着目し、構造のサイズとの関連について分析する。強勢母音群・弱勢母音群が張る構造のサイズと学習者の発音習熟度との間の相関を分析し、韻律を含めた自動発音評価への応用の可能性について検討する。

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 [7]。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、

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

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

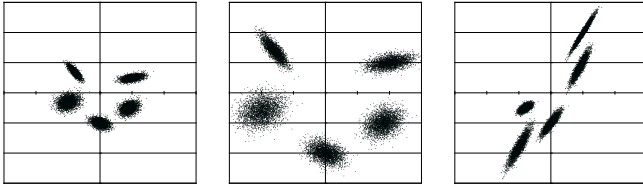


図 1 構造不変の定理

Fig. 1 Theorem of the invariant structure

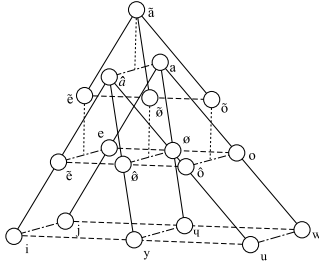


図 2 ヤコブソンによる幾何学的音韻構造

Fig. 2 Jakobson's geometrical structure

よるフランス語における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. 少量データによる発音構造化

3.1 最大事後確率推定による音声事象分布推定の安定化

データ量が限られている場合に問題となるのは、分布推定における安定性である。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る。従って、本分析での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 [8]。

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 1 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 (140ms) 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。一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チャリヤ距離を計算し、距離行列 (即ち 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 5 つの構造より任意の 2 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。構造間差異 D の算出には下記の近似式を用いている [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 は母音種類, M は母音数を表す。分析結果を図 3 に示す。構造間差異の平均は ML, MAP

表 1 分析条件 (Section 3.1)

Table 1 Acoustic conditions of the analysis (Section 3.1)

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

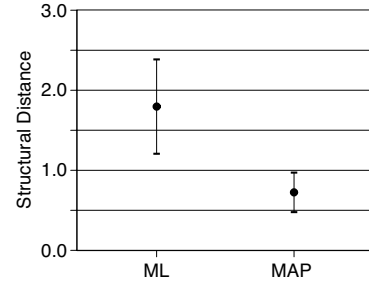
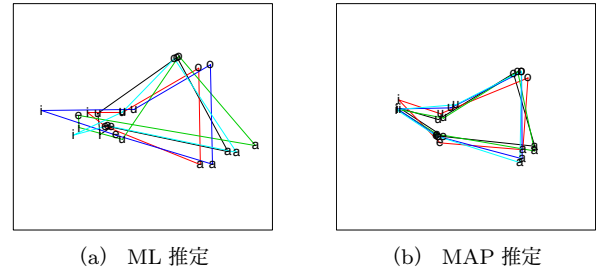


図 3 各分析手法における話者内構造間差異

Fig. 3 Intra-speaker structural distance



(a) ML 推定

(b) MAP 推定

図 4 多次元尺度構成法による母音構造の視覚化

Fig. 4 Visualization of the vowel structures with MDS

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た結果の一例を図 4 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 5 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、発声毎のばらつきが抑えられているのがわかる。以上の結果より、MAP 推定を用いることにより、少量データの場合でもより安定した発音構造化が可能となることが確認された。次節の分析では MAP 推定を用いることとする。

3.2 少量データによる発音構造化と発音状態変化の追跡

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、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持ち、口・腹周りの肉体的訓練から英語発音に取り組んだ経験を持つ日本人成人男性より、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 (/i, ɪ, e, æ, ʌ, ɑ, ɔ, u, ʊ, ɜ, ə/) と、日本語母音 5 種類 (/a, i, u, e, o/) をそれぞれ収録した。収録

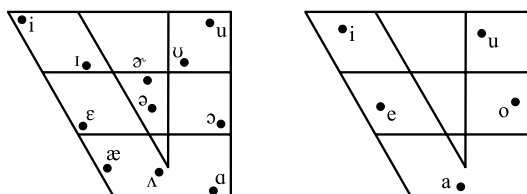


図 5 英語と日本語における母音図

Fig. 5 Vowel charts of American English and Japanese

表 2 母音置き換えの組み合わせ

Table 2 Vowel substitution table

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

表 3 分析条件 (Section 3.2)

Table 3 Acoustic conditions of the analysis (Section 3.2)

sampling	16bit / 16kHz
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)

話者は上手な英語を発音することができると同時に、日本人である以上故意に日本語訛りの英語を発音することもできる。以下の分析では、両者の英語を混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。

図 5 に英語と日本語における母音図を示す。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 2 に示す。この母音置換の組み合わせに基づいて、下記のように発音状態を定義する。

S1: 全ての英語母音を日本語母音に置き換えて発音する

S2: /ɑ, ʌ, æ, ɜ, ɔ/ の発音が矯正される

S3: S2 の状態から /i, ɪ/ の発音が矯正される

S4: S3 の状態から /u, ʊ/ の発音が矯正される

S5: S4 の状態から /ɛ/ の発音が矯正される

S6: S5 の状態から /ɔ/ の発音が矯正される

英語母音の発音が上達していく過程を状態 **S1** から状態 **S6** への遷移によりシミュレートする。

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 発音のみを用いている。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音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 /bat/ の発音を用いてそれぞれの発音を置き換える。

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として視覚化を行った。図 6 に 6 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

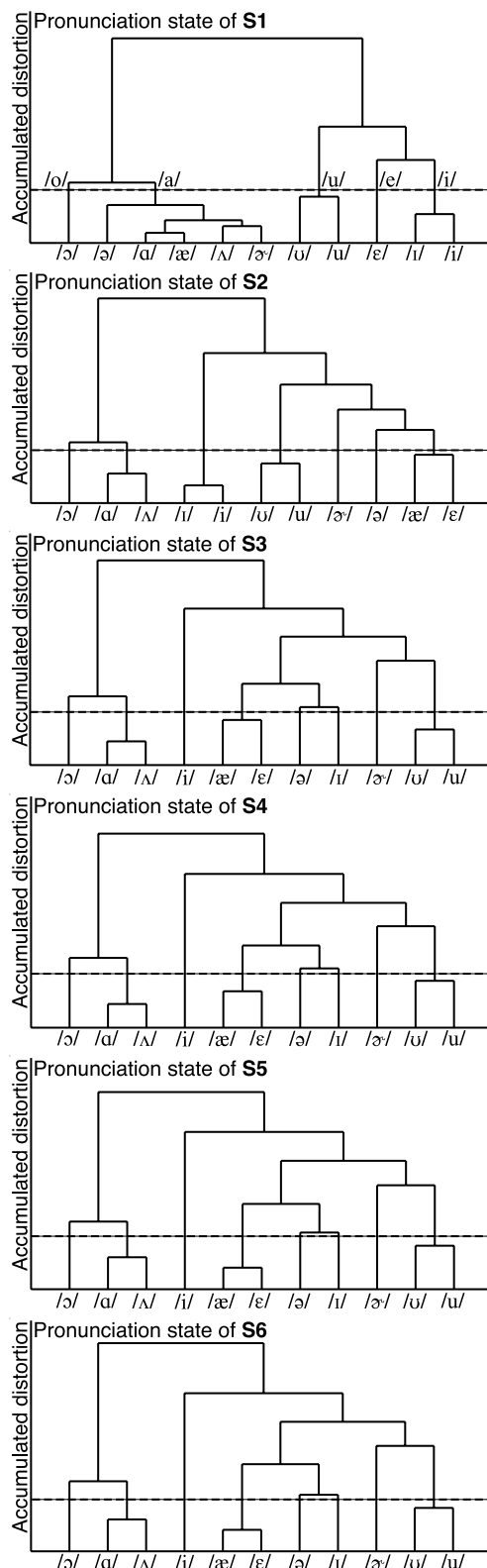


図 6 各状態における発音木

Fig. 6 Six tree diagrams of the pronunciation

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 5 つのクラスに明確に分かれており、日本人発音の癖が表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、英語音声学的にも美しい様子を呈していることがわかる。この両者の樹形図は、図 5 の英語・日本語の母音図とも対応がとれている。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, ʌ, æ, ɤ, ə/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、/i, ɪ/のクラスが分かれてくる。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/ɛ/が近づいているのがわかる。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、一母音に対して一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、[8] では一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。

4. 英語強勢弱勢母音の分析

4.1 構造のサイズの物理的解釈

先行研究 [5], [6] において、我々は構造的表象を用いた発音の分節の特徴に対する評価について検討を行ってきた。また、前節では母音生成に関してその評価を行った。本節では、発音の韻律的特徴、特に強勢・弱勢に着眼し、構造的表象との関係を明らかにする。

英語には弱勢としてのみ出現する母音、弱母音 (schwa) がある。英語音声学によると、schwa は母音図中において全ての母音の中心に位置し (図 5)、その他の母音は弱勢になると schwa に近づくとされている [10]。このことを構造的表象に当てはめて考えると、弱勢化された母音は互いの距離が短くなり、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。強勢母音はより大きな調音努力を払って母音間の差異を明確化したものであり、弱勢母音は不明瞭化したものである事を考えると、音響的普遍構造のサイズが発話者の調音努力を記述していると考えられる。

4.2 母語話者音声の強勢弱勢母音分析

上記の考察を定量的に検討するため、既に [11] において、母語話者音声データを用いて強勢母音群・弱勢母音群の普遍構造サイズの分析を行った。米国人英語教師の音声データ (709 文) に対して、データ中の全ての母音に強勢・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音素 HMM を構築する。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る。Ward 法とは、統合により生じる累積歪みが最小となる二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歪みとなる。これは VQ 歪みに相当し、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見なし定量的な検討を行った。強勢・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、/i, ɪ, u, ʊ, ɛ, æ, ʌ, ɑ, ɔ, ɔ̃/を対象として、表 4 の条件にて分析を行った。

表 4 分析条件 (Section 4.)

Table 4 Acoustic conditions of the analysis (Section 4.)

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

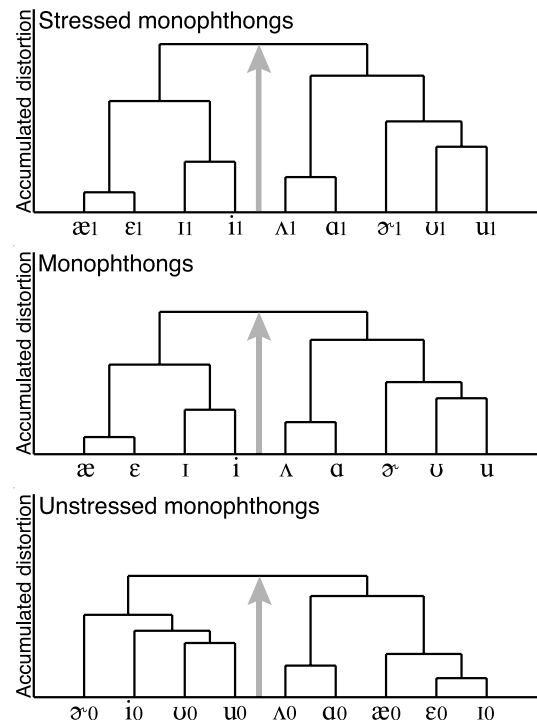


図 7 英語母音における樹形図

Fig. 7 Tree diagrams of American English vowels

結果を図 7 に示す。上から、強勢母音・全ての母音・弱勢母音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。強勢母音構造のサイズは、弱勢母音構造のサイズの約 1.4 倍という結果になった。樹形図の形に関しては、全母音の場合と比較して、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。形の違いに関しては、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の理由が考えられる。以上の結果より、音響的普遍構造のサイズの大小が発話者の調音努力の大小を表していると解釈することができる。

4.3 学習者音声の構造サイズと発音習熟度との相関分析

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。一方で日本語の母音体系においては、図 5 を見てもわかるように、中心に位置する母音が存在しない。それ故、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く、英語特有のリズム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。本分析では、強勢・弱勢構造サイズの比の値を基に、強勢となるべき母音がいかに強勢らしく、弱勢となるべき母音がいかに弱勢らし

く発声されているかを定量化し、各話者の発音習熟度との相関について分析する。

音声データとして、ERJ (English Read by Japanese) データベース [12] 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 4.2 節 と同様である。発音習熟度に関しては, ERJ データベースには分節的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み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ものの, その等時性については明確な接点を持たない。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, ここでは分節的特徴に基づく評定値を用いた。

結果を図 8 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 (弱勢サイズに対する強勢サイズの比の値) である。サイズ比が 1.0 より大きいほど, 強勢は強勢らしく, 弱勢は弱勢らしく発声されている, ということになる。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。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事が確かめられた。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国語話者 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, 強弱の区別をつけて発音ができており, 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, 機能語のような弱く発音すべき単語を, 逆に強く発音してしまう, といった原因が考えられる。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うるが, 今回用いたデータでは発話者の意図, 即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, 日本人英語によくみ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。

5. ま と め

本報では, 音声の構造的表象に基づく非母語話者英語音声に対して行った二つの分析について報告した。一つ目の分析では, データ量が限られている場合においても構造的表象により学習者の発音状態を安定に記述することが可能であり, さらに学習者の発音が上達する様子を追跡する事が可能となることを実験

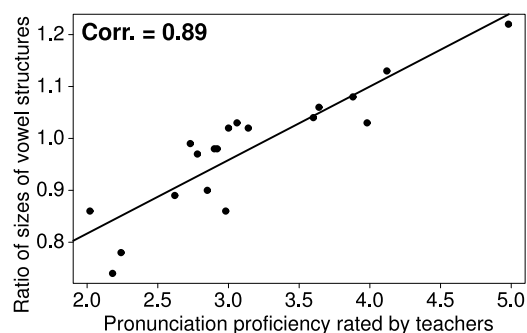


図 8 構造サイズ比と英語教師による評定との相関

Fig. 8 Correlation between structure sizes and human scores

的に示した。二つ目では, 韻律的特徴, 特に英語の強勢弱勢母音に着目し, 構造サイズとの関連について分析を行った。強勢母音群が張る構造と強勢母音群が張る構造のサイズ比により, 強勢・弱勢らしさを定量化し, 学習者の発音習熟度との間に高い相関が見られることを確認した。以上の結果より, 音声の構造的表象を用いた発音評価アプリケーションの可能性は大きく広がったといえる。

今後の課題として, 前者の分析に関しては, 分析対象とする音素の拡張とより多くの話者の音声データでの分析を行い, 本手法の有効性のさらなる検討が必要であると考えられる。また後者の分析に関しては, 日本人英語によくみ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挙げられる。さらに, 本報告で行った二つの分析を統合し, 少量データによる韻律的特徴を含めた発音習熟度自動評定について検討をしていきたいと考えている。

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. ICASSP*, pp.889-892 (2005)
- [3] 峯松信明他 "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", *信学技報*, SP2004-27, pp.47-52 (2004)
- [4] 峯松信明他 "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", *信学技報*, SP2005-12, pp.1-8 (2005)
- [5] N. Minematsu, "Pronunciation assessment based upon the phonological distortions observed in language learners' utterances," *Proc. ICSLP*, 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. ICSLP*, pp.1317-1320 (2004)
- [7] M. Pitz et al., "Vocal tract normalization as linear transformation of MFCC," *Proc. EUROSPEECH*, pp.1445-1448 (2003)
- [8] 村上隆夫他, "音声の構造的表象を用いた日本語母音系列の自動認識", *信学技報*, SP2005-14, pp.13-18 (2005)
- [9] 峯松信明他 "音声の構造的表象とその距離尺度", *信学技報*, SP2005-13, pp.9-12 (2005)
- [10] J. Clark and C. Yallop, *An introduction of phonetics and phonology*, 2nd edition, Blackwell Publishers Inc. (1995)
- [11] 朝川智他, "音響的普遍構造のサイズと単語境界における音響的分離度に着目した米語音声の音響分析", *信学技報*, SP2004-28, pp.53-58 (2004)
- [12] N. Minematsu et al., "Development of English speech database read by Japanese to support CALL research," *Proc. ICA*, pp.557-560 (2004)