

音声の構造的表象を用いた英語二重母音の発音評価に関する実験的検討*

中村綾乃, 鈴木雅之, 峯松信明, 広瀬啓吉 (東大), 牧野武彦 (中央大)

1 はじめに

近年国際化が進むのに従って英語教育の重要性が増してきており、2011年度からは全ての公立小学校にて、5・6年生の授業で外国語活動(主に英語)が必修科目になることも決まっている。しかし、実際に指導に当たるのはほとんどが英語教師の資格や授業経験のない小学校教員(学級担任)であり、補助となるALT(Assistant Language Teacher)などの人数も十分確保出来ていないのが現状で、多くの教員が授業を組むことに対して不安を抱えている[1]。

そういった現状の助けになるものの1つとして、計算機を使ったCALL(Computer Aided/Assisted Language Learning)システムがあげられる。学習者にあわせた学習を提供でき、自主的学習を促すCALLシステムは、自宅でも使えるため授業以外で英語に触れる機会を増やすという意味でも有効である[2]。

英語の発音評価もCALLシステムの1つとして実装されているが、従来のシステムは、学習者の性別や身長、年齢といった個人性によって判定結果が不安定になる不一致問題を抱えている。これまではこの問題に対処するために、大量の音声を集めてモデル発音を統計的に分布として表現し、声色の違いをばらつきとして扱う手法や、正解となるモデルに話者適応をかけて学習者の声色に合わせる手法などがとられていた。しかし、大量の音声を集めることは容易ではなく、また話者適応では肝心の発音が歪んでしまうことがあるなど、本質的な問題の解決には至っていない。

筆者らは、音響特徴量の相対的な距離に注目する音声の構造的表象という考えを発音評価に適用する方法を検討してきた[3]。そして構造表象を用いることで少量のデータで英語の発音評価が出来るシステムを実装した[4, 5]。このシステムでは、英語の短母音、長母音、弱母音、r色の母音と呼ばれる4種類、合計11母音について、1. 自分が発音を似せたいと思う教師の選択、2. 発音の習熟度の推定、3. 矯正すべき母音の特定、4. 複数の学習者の発音に基づく分類、などが可能である。

英語の母音には、上記の4種類以外に二重母音と呼ばれる母音がある。二重母音は日本語にはない独特の発音の仕方をする母音で、日本人が苦手としてい

る母音の1種であり、発音評価システムにおいて発音を誤りやすい音素として導入されることもある[6]。

また、二重母音の発音評価が出来るようになると、実際にアメリカの子供たちが言葉を学ぶときに用いられているライミング(語末の韻を踏む言葉遊び)[7, 8]をCALLシステムに取り入れることが出来るようになり、システム構築における実用的価値も大きい。

本稿では、すでに実装されている英語発音評価の手法を参考に、音声の構造的表象を用いて二重母音の発音評価をする手法を検討した。

2 音声の構造的表象に基づく発音評価

2.1 音声の構造的表象とは

音声には、1. どのような言葉を発したか(言語情報)、2. 誰が言ったか(話者性)、3. どのような音響機器を使ったか(音響機器特性)、4. どのような環境下で言ったか(収録環境)などといった情報が含まれている。言語情報以外をまとめて非言語情報と呼ぶ。発音評価をする場合には、必要とされるのは言語情報のみであり、非言語的要因による音響特徴の歪みをいかに無視して発音を評価するのが問題となる。

音声に混入する非言語的要因による雑音や歪みには、加算性雑音、線形性歪み、乗算性歪みの3種類がある。このうち、加算性雑音は話している環境の雑音などによるものであり、静かな部屋に移動して収録するといった手段で避けることが出来る。一方不可避免的に混入するのは、声道長などの話者性による線形変換性歪みと、音響機器などの特性による乗算性歪みである。ケプストラムを用いて音響事象を表現した場合、線形性歪みは $c' = Ac$ 、乗算性歪みは $c' = c + b$ と表現される。

ある空間内で n 点からなる幾何学的構造は、 ${}_nC_2$ 個の2点間距離を求めれば定義できる。よって、前述のような1次変換(アフィン変換) $c' = Ac + b$ に対して不変な距離尺度を用いることで、音響事象群を、不可避な非言語性歪みに不変な幾何学的構造として表現することが可能になる[9]。本稿では、二つの分布 $p_i(x)$, $p_j(x)$ に対して式1で表されるバタチャリヤ距離(BD)の平方根を用いて事象間距離を定義する。BDは、任意の連続かつ可逆な線形・非線形変換に対

* An experimental study on assessment of diphthongs of learners using pronunciation structure.

By NAKAMURA Ayano, SUZUKI Masayuki, MINEMATSU Nobuaki, HIROSE Keikichi (The Univ. of Tokyo), and MAKINO Takehiko (Chuo Univ.)

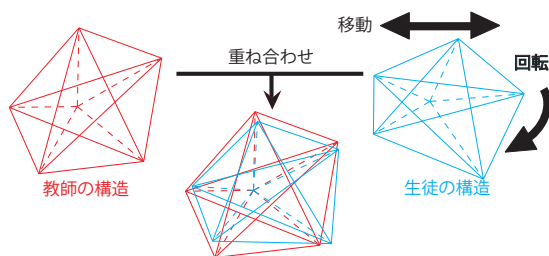


Fig. 1 5 母音で作った構造を回転・平行移動して重ね合わせる様子。

して不変であることが証明されている [10]。

$$BD(p_i, p_j) = -\ln \int \sqrt{p_i(x)p_j(x)} dx \quad (1)$$

2.2 構造間距離尺度

BD を用いて M 個の母音間の距離を全て求めると、 $M \times M$ の母音間距離行列で表される話者ごとの母音構造を定義することが出来る。

教師の母音構造 T と学習者の母音構造 S を作り、図 1 のように回転させたり移動させたりして近づけたとき、構造間距離は対応する 2 点間距離の総和の最小値として定義できる。本稿では式 2 で表される、構造の各エッジごとの正規化二乗誤差を構造間距離尺度として用いる [11]。 S_{ij} と T_{ij} は、各話者の構造の (i, j) 成分、つまり母音 i と母音 j 間の BD の平方根である。

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

教師との構造間距離が小さいほど学習者の発音が教師の発音に近いということになるので、構造間距離は発音評価の尺度として利用することが出来る。

3 構造的表象を用いた英語二重母音の評価

3.1 英語二重母音とは

5 つしか母音が無い日本語に比べ、英語は複雑な母音体系を持っている。英語の母音は、短母音、長母音、二重母音、弱母音、r 色の母音といった 5 種類に分けられ、全部で 22 母音になる [12]。このうち二重母音以外の母音を、音色が変わらないという意味で本稿では今後単母音と表記する。

二重母音には、 $/aɪ/$ 、 $/aʊ/$ 、 $/eɪ/$ 、 $/ɔɪ/$ 、 $/oʊ/$ の 5 種類がある。単母音とは違い、二重母音は 1 つの母音の中で音色が変わるという特殊な母音である。

日本語には二重母音に対応する母音は無いが、似ているものに連結母音というものがある。連結母音

Table 1 二重母音の評価において参照する単母音。

対象の二重母音	参照する単母音
$aɪ, aʊ$	$æ, ʌ, ɑ$
$ou, ɔɪ$	$ɔ$
$eɪ$	$ε$

は 2 つの単母音が続けて発音されたもので、日本語音声学的には 1 つの母音ではない。

二重母音を前半と後半の母音に分けて考えた場合、前半は明確なある特定の音色を有した発音になるが、後半は後続子音などの影響を受けやすく、あいまいな音になりがちである。一方、連結母音は単母音が続けて発音されるだけなので、前半と後半どちらも比較的明瞭に発音される。日本人が二重母音の発音を誤る場合、音素自体を置換してしまうという誤りに加え、日本語的な連結母音になってしまう、つまり後半母音を明瞭に発音しすぎてしまうという誤りもあると考えられる。

3.2 音声学者の評価方法

二重母音の評価の枠組みを考えるにあたって、音声学者 (第 5 著者) が考える、二重母音を評価する際の留意点を以下にまとめる。

まず二重母音の前半部分の音というのは、その他の英語のどの単母音とも異なっている。よってそれぞれの二重母音に対し表 1 にあるような混同しやすい単母音を参照し、各々を区別して発音できているかで評価が出来る。

一方後半部分に関しては、前半部分から音色が動いているということが重要であり、前節でも述べたようにあいまいな音になりがちなので、必ずしも決まった音色にならなければ不正解というわけではない。ただし、正しい音色の方向に向かって動いているかといったことは評価の対象になり、また少し厳しい評価をすれば音色の動きが足りていないか、もしくはは行き過ぎていないかといったことで減点になる。動きが足りないと単母音的に、行き過ぎると連結母音的になってしまう。

以上 2 点を総合したものが二重母音の評価となる。

3.3 構造的表象による評価法

英語二重母音の特徴を踏まえると、その発音評価には 1. 前半母音の音色に着目した評価、2. 音色の動きに着目した評価、を行うことが有効と考えられる。

1 点目については、これまでの単母音群の構造を使った発音評価とほとんど同じである。よって、2 重母音を切り分けることによって単母音の発音評価と同じ枠組みで母音の構造を作り、式 2 をスコアとして評価が出来ると考えられる。

Table 2 使用した単語

二重母音	kite(/ai/)	how(/au/)	sage(/ei/)
	oils(/ɔɪ/)	float(/ou/)	
参照単母音	hot(/ɑ/)	hatch(/æ/)	touch(/ʌ/)
	raw(/ɔ/)	test(/ɛ/)	

単母音の評価では、音色の特徴を表す特徴量としてメルケプストラムを用い、母音構造を作っていた。そのことを参考にすると、2点目に関しては、音色の動きを表す特徴量で母音構造を作り、同じように評価をすることが出来ると考えられる。音色の動きを捉える特徴量として、本稿ではメルケプストラムの変化量である Δ メルケプストラムを利用することを検討した。

なお、以降メルケプストラムを MCEP、 Δ メルケプストラムを Δ MCEP と略す。

4 実験

第 3.3 節で検討したように、音色に着目した評価と音色の動きに着目した評価に観点を分け、それぞれ音声の構造的表象を用いての評価が可能か検証した。

4.1 使用する音声

本実験では、ERJ データベースの音声を利用する [13]。このデータベースに収録されている音素バランス単語セットのセット 4 の中から、表 1 の母音を含む表 2 の単語を選んで用いた。セット 4 を読み上げている日本人学生 37 名と英語母語話者 8 名、計 45 名分の音声を学習者音声とした。その単語音声から母音部分のみを切り出し、二重母音はそのままのものと前半部分と後半部分に切り分けたものを利用した。切り分けた母音について、本稿では前半の母音には F(Former)、後半には L(Latter) を二重母音の記号の後ろに付与して表す。本来このような切り出しは何らかの方法で自動で行うべきであるが、今回は切り出しによるエラーの影響をなくすために手動で行った。

ERJ データベースでは、話者ごとに選択された一部の単語についてのみ英語教師による 5 段階の手動評価値がつけられており、その平均スコアが話者のスコアとなっている。今回の実験では、使用する全ての単語について母音の発音にのみ着目した評価が望ましいため、1 名の音声学学者 (第 5 著者) に手動評価値を付け直してもらった。評価に関しては、話者同士で発音を比較できるという構造の特徴を生かすため、評価者自身の発音にいかに近いかという観点からの 4 段階評価とし、その評価者の音声を収録して教師音声とした。二重母音に関しては、第 3.3 節の観点に基づき前半部分と後半部分別々に評価した。

以下の実験では、表 3 の条件下で構造を抽出した。

Table 3 音響分析条件

サンプリング	16bit/16kHz
窓長、シフト長	25 msec, 1 msec
特徴量	第 4.2 節 MCEP(1-10 次元) 第 4.3 節 MCEP(0-10 次元) Δ MCEP(0-10 次元)
推定量と分布	MAP 推定によるガウス分布

Table 4 話者 IWA-M02 の評価での各母音の評価の重み。

前半	重み	参照	重み	後半	重み
aiF	0.24	ɑ	0.34	aiL	0.13
auF	0.32	æ	0.23	auL	0.19
eiF	0.19	ʌ	0.35	eiL	0.14
ɔɪF	0.36	ɔ	0.35	ɔɪL	0.16
ouF	0.22	ɛ	0.26	ouL	0.21

4.2 音色に関する評価

(前半母音、後半母音、参照単母音)、(前半母音、参照単母音) というそれぞれ 15 角形、10 角形の母音構造を作り、教師音声との構造間距離を計算した。

母音構造を作るのに使った各母音につけられた手動評価値の平均をその話者の評価値とし、求めた構造間距離と相関をとると、後半母音を含めた 15 角形で -0.57、含めない 10 角形では -0.59 という相関が得られた。後半母音を含めないで構造を作った方が相関が高くなるという結果は、後半母音の音色があいまいであるという特徴や、評価のとき音色としては前半母音に着目するという音声学者の意見と一致する。

式 2 による構造間距離は母音 i と母音 j の正規化距離を加算した総和として計算されるが、重視すべき母音対、軽視すべき母音対が存在することは容易に推測される。そこで、各母音間距離に適切な重みをかけて構造間距離を再定義することを試みた。手動スコアとの相関を最大化するように重みを最適化する多段重回帰を行うと [14]、相関は -0.68 まであがった (評価はオープンテスト)。

今回行った 2 段重回帰では、1 段目ではある母音の評価をする時の、その母音と他の母音との距離にかかる重みが、2 段目は話者の評価を出す時の、各母音の評価にかかる重みが計算される。ある話者での後者の結果が表 4 のようになっており、やはり前半母音や参照単母音で重みが大きくなっている。

今後は、異なる音声学学者による評価の相関を調査し、0.7 弱という値の意味を検証することを考えている。

4.3 音色の動きに関する評価

音色の動きを捉える特徴量としての Δ MCEP が利用できるかを検討するため、まずは Ward 法によるボトムアップクラスタリング [15] を行って構造を可視化し、MCEP のみで構成した構造と Δ MCEP のみで構成した構造とで違う特徴を表すかを検証した。こ

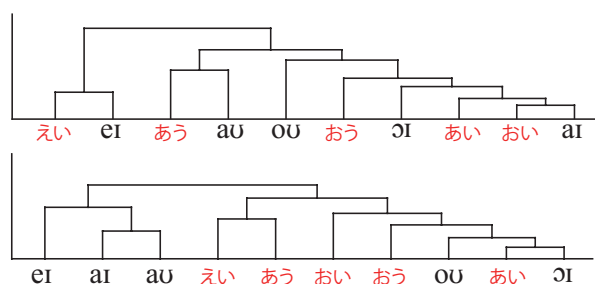


Fig. 2 日本人帰国子女の母音構造から作った樹形図。上が MCEP、下が ΔMCEP を使ったもの。

のような樹形図は、単母音の発音評価でも発音状態の記述として用いられている。

日本人帰国子女の男性に、英語二重母音を含む単語と、二重母音を日本語の単母音 2 つで置き換えたカタカナ語を読んでもらい、その母音部分を切り出して母音構造を作り、それから樹形図を作成した(母音部分全体で 1 つの分布を構成している)。カタカナ語の母音というのは、連結母音のように発音してしまった場合を想定しており、このような動きの違いを ΔMCEP によって捉えることが出来れば、音色の動きの評価に使えと考えられる。

作った樹形図を図 2 に示す。上の図が、これまで音色の評価に用いてきた MCEP を特徴量として作ったもの、下は ΔMCEP を特徴量として作ったものである。黒が二重母音、赤のひらがなが連結母音的な母音である(二重母音を切り分けずに用いたので、前後半を表す F、L は付与されていない)。

MCEP で樹形図を作った場合は、音色に着目してクラスタリングがなされるため、二重母音とその対応する連結母音が最初にクラスタを作ることが多いが、ΔMCEP ので作った構造木の方では、そのようなことは起こっていない。さらに、前者では二重母音とカタカナ語の母音がクラスタの中で混ざり合っているのに対し、後者では日本語と英語である程度クラスタが分かれていることが分かる。

このように、ΔMCEP は MCEP で母音構造を作ったときと違う特徴を示しており、音色の動きに基づく評価尺度として利用できると考えられる。

そこで次に、音色の動きとして評価に使えるかを検証するため、切り分けない二重母音で母音構造を作り、教師との構造間距離を出した。特徴量としては 0-10 次元の ΔMCEP を用い、手動評価値は二重母音の前半部分、後半部分にそれぞれつけられている評価値の平均とした。しかし、ここで出した構造間距離と手動評価値の間には十分な相関は見られなかった。

しかしながら、第 3.2 節でも述べたように二重母音を評価するには音色の評価だけでは不十分であると考えられるので、それ以外の要素も二重母音の発

音評価には関係していることは確かである。よって、ΔMCEP の構造での使い方や、さらにそのほかの特徴量を利用した評価を検討する必要がある。

5 まとめ

本稿では、音声の構造的表象の枠組みを用いて英語二重母音の発音評価をする手法を検討した。評価を二重母音の前半の音色に着目した評価、二重母音全体での音色の動きに着目した評価、という観点に分け、教師の母音構造との構造間距離をその学習者のスコアとして音声学者による手動評価との相関をとった結果、前者では良好な相関が見られた。

参考文献

- [1] 鳥飼玖美子, “危うし! 小学校英語”, 文春新書, 2006
- [2] JACET SLA 研究会, “文献から見る第二言語習得研究”, 開拓社, 2005
- [3] 朝川智 他, 電子情報通信学会論文誌 Vol.J90-D No.5, pp.1249-1262, 2007
- [4] 鎌田圭 他, 日本音響学会秋季講演論文集, 2-4-3, pp.327-330, 2007
- [5] 鎌田圭 他, 情報処理学会第 70 回全国大会講演集, 2ZL-6, pp.275-276, 2008
- [6] 久住大 他, 情報処理学会研究報告, Vol.2009-SLP-79 No.35, 2009
- [7] 足立恵子, “アメリカの子供のように英語を学ぶ本”, 中経出版, 2005
- [8] リーパー・すみ子, “アメリカの小学校ではこうやって英語を教えている”, 径書房, 2008
- [9] 峯松信明 他, 電子情報通信学会音声研究会 SP2005-12, pp.1-8, 2005
- [10] 峯松信明 他, 日本音響学会春季講演論文集, 1-P-12, pp.147-148, 2007
- [11] Masayuki Suzuki *et al.*, Proc.ASRU, pp.574-579, 2009
- [12] 川越いつえ, “新装版 英語の発音を科学する”, 大修館書店, 2007
- [13] 峯松信明 他, 日本教育工学会論文誌 vol.27 no.3, pp.259-272, 2004
- [14] 鈴木雅之 他, 日本音響学会春季講演論文集, 1-P-16, 2010
- [15] 岡太彬訓, 今泉忠, “パソコン多次元尺度構成法”, 共立出版株式会社, 1994