

科学研究費補助金研究成果報告書

平成 24 年 6 月 12 日現在

機関番号：12601

研究種目：基盤研究(B)

研究期間：2008～2010

課題番号：20300265

研究課題名(和文) 音声の分節的・韻律的特徴を包含する発音の構造的表象に基づく外国語教育・学習支援

研究課題名(英文) Technical development of environments for learning and teaching foreign languages by using structural representation of pronunciation

研究代表者

峯松 信明(MINEMATSU NOBUAKI)

東京大学・大学院工学系研究科・教授

研究者番号：90273333

研究成果の概要(和文)：外国語発音の分析精度の向上を目的として、音声の構造的表象に基づいた発音スコアの算出、発音エラーの検出、及び、発音の様態に基づく学習者分類について検討した。構造表象に基づいて発音を表象すると、音と音の関係性のみを捉えることになるが、この関係性と従来の音の絶対性に基づく特徴とを組み合わせることでスコア算出・エラー検出の精度、向上を測ることができた。学習者分類の妥当性も実験的に示すことができた。

研究成果の概要(英文)：To improve the performance of analyzing foreign language pronunciations, structural representation of speech was introduced to the task of scoring the pronunciation, detecting pronunciation errors, and clustering students' pronunciations. The structural representation allows us only to represent the pronunciation only based on sound-to-sound relations. By using the relations as well as absolute sound features, which have been conventionally used, the performance of scoring and error detection was improved. High validity of students' clustering was also shown.

交付決定額

(金額単位：円)

	直接経費	間接経費	合計
2008年度	6,500,000	1,950,000	8,450,000
2009年度	4,800,000	1,440,000	6,240,000
2010年度	3,400,000	1,020,000	4,420,000
年度			
年度			
総計	14,700,000	4,410,000	19,110,000

研究分野：音声工学

科研費の分科・細目：科学教育・教育工学

キーワード：構造的表象、発音学習、分節的／韻律的特徴、調音音声学、英語教育

1. 研究開始当初の背景

音声の構造的表象は、ある話者の音声から、その話者の体格や性別に起因する音響的なバイアス項(即ち非言語的な特徴)を取り除くことを目的として導出される。まず音声ストリームを音イベントの系列として捉える。音イベントは分布でモデル化され、全ての音と音の関係性(距離)を f -divergence で計測すれば、全音イベント群は距離行列として

表現できる。この距離行列は全ての音に、共通の写像を施しても不変である。非言語的要因によるバイアス項は写像として表現できるため、この行列表現は不変項となる。

しかしながら、全ての音イベントに対して共通の写像を仮定するのは強い仮定であり、 f -divergence を計測すべきは共通の写像が適用される音イベント群に限定すべきである。例えば母音群のみを対象とすれば、共通

写像は妥当な仮定であるが、子音までを考慮するとこの仮定は妥当ではない。従来の構造表象に基づく発音分析は、対象を母音に限定したものが多く、共通写像の仮定が比較的当てはまりやすい音を対象としていた。

2. 研究の目的

従来の構造表象を用いた発音分析・評価方法の検討に加えて、以下の項目を研究目的として定めた。

- 1) 子音までを対象として扱う方法を検討する。この場合、どの音と音が共通の写像で結ばれるかは自明では無いため、学習データを用いた自動選択手法、あるいは、関係性を利用する際の信頼度（重み）推定の自動化を検討する。
- 2) これまでの構造表象の利用は各学習者の発音スコアの推定が主目的であり、発声のどこに不適切な発音があるのかを示す誤り検出には適用されてこなかった。ここでは母音を対象として誤り検出についても検討する。
- 3) 従来研究においても学習者分類については予備的検討を行ってきたが、結果の妥当性については十分な検討ができていなかった。本研究では音声学者による分類と比較することでその妥当性も検討する。
- 4) また、構造表象は音声の音色の動き・うねりの様子を話者不変に捉える方法と解釈でき、これは韻律的特徴の一つとして位置づけることが適当であると思われる。この解釈の理論的考察を行なう。
- 5) 構造表象理論の精緻化を発音分析のみならず音声認識の観点からも行なう。
- 6) 構造表象は、言語学的には古典的な考え方であるが、技術的には新しい考え方である。これに基づく発音評価の教育的妥当性について、関係者に対して分かり易く情報発信することも検討する。

3. 研究の方法

- 1) 母音・子音が合計で N 種類あった場合、 $N \times N$ の距離行列が構成され、ここには $N(N-1)/2$ 種類の距離（音素間距離）が定義される。これが不変量となる訳だが、いずれの不変量が信頼性の高い不変量なのかは未知である。
従来、学習者と教師の距離行列間の差分を「差分行列」として定義し、差分行列の要素の二乗和でもって学習者・教師の発音差の定量化としていた。母音のみを対象とした場合は十分な精度を示したが、子音までを考慮すると問題が生じる。
ここでは、回帰分析の考え方を導入し、単純な二乗和ではなく、重み付き二乗和を考え、それを用いて学習者の習熟度を推定する。しかし、 $N(N-1)/2$ 種類の距離

値から直接的に習熟度を推定するのはパラメータ数が多いために過学習が起こる可能性が高い。そこで、段階的に回帰を行なう（各音素スコアの推定を行い、それらとまとめる形で話者スコアを推定する）方法を検討する。

- 2) 上記で示した音素スコアの予測は、そのまま発音誤り検出にも使える技術となる。ここでは、母音を対象として発音スコアを母音毎に自動推定することを検討し、音声学者による（母音ごとの）評定スコアと比較することで、その精度を検討する。またこれとは別に、与えられた任意の二母音距離行列に対して、差分行列の各要素の信頼度を繰り返し演算によって、自動推定することについても検討する。
- 3) 学習者分類については、既に従来研究において予備的検討を行なってきたが、得られた結果の科学的根拠、妥当性については十分な検討が行なえていない。そこで本研究では、学習者音声を音声学者に母音図化させ、その母音図に基づいた話者分類と、構造表象による話者分類とを比較することで、後者の妥当性を検証する。母音図化された各話者の母音発音を二話者間で比較する場合、母音図の台形枠をその二話者で重ねればよい。構造表象は、台形枠がない母音図と解釈できるため、母音図による分類と構造表象による分類とは直接比較対象として妥当であると考えられる。
- 4) 構造表象を使ってある一発声を表象する場合、音色の絶対的特性ではなく、音色の動きを個々のイベント間（時間的に離れたイベント同士も含む）の距離として計測する。これは音色の動きを話者不変に表象することと同値である。一般に韻律的特徴と呼ばれる音声特徴は、音の高さや音の大きさの時間変化の相対的パターンで定義されることが多い。構造表象は音色の時間変化の相対的パターンであり、これは韻律的特徴の一つとして定義可能である。ここでは、このように定義することの理論的妥当性について論じ、言語障害学的考察（自閉症）や進化人類学的な考察（サルとヒトの情報処理能力の差）も行なう。
- 5) 構造表象を用いた発音分析とは別に、構造表象を用いた音声認識の精緻化についても検討する。これは音声認識を目的として行なわれた検討の多くは、発音分析にも応用可能だからである。
- 6) 本プロジェクト成果の教育的妥当性について、は各種学会の講演会などを利用し、広く関係者に問いかける活動を通して現場の教師からの理解を得る努力を継続して行なう。

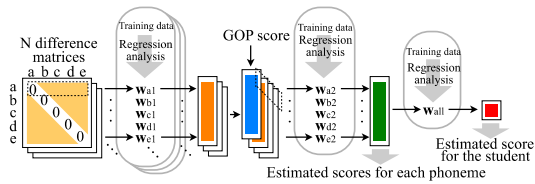


図1 多段の重回帰分析

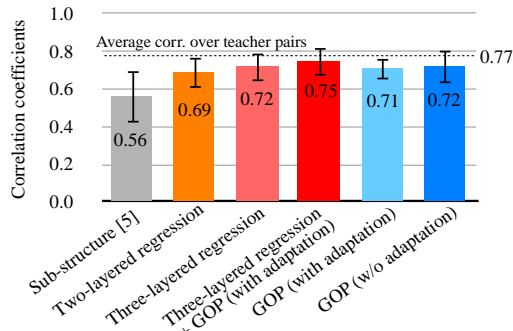


図2 自動スコアと教師スコアの相関

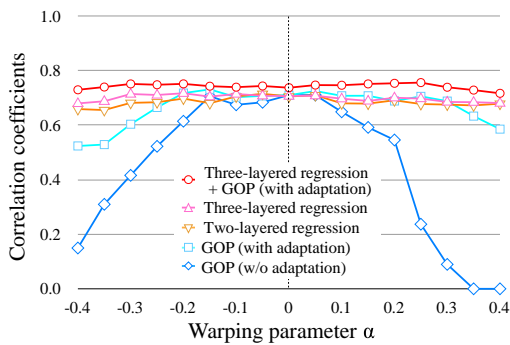


図3 様々な声色の音声に対する性能

4. 研究成果

- 1) 学習者行列と教師行列との差分を表現する差分行列に対して、図1に示すような多段階の回帰分析をかけ、最終的に学習者の発音習熟度を推定する方式を提案した。この場合、差分行列以外のスコア（従来から使われている GOP スコアなど）もスコアの一つとして導入できる。図2に自動スコアと教師スコアとの相関を示す。提案手法（赤）は従来手法（適応ありの GOP）より高い相関を示した。更に、学習者音声をその声道長を変形させることで声色を変形させ、その音声を入力として用いたところ、従来手法より高い頑健性を示すこともできた（図3参照）。
- 2) 上記の成果を（英語母音発音を対象として）そのままエラー検出に応用したところ、良好な結果を得ることができた。図4に各母音毎に自動推定される母音スコアと自動スコアとの相関を示す。また参考値として異なる二教師間での相関、あるいは、同一教師内での相関についても示す。

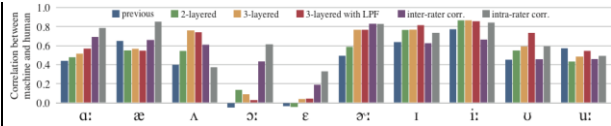


図4 各母音に対する自動スコアと教師スコアの相関

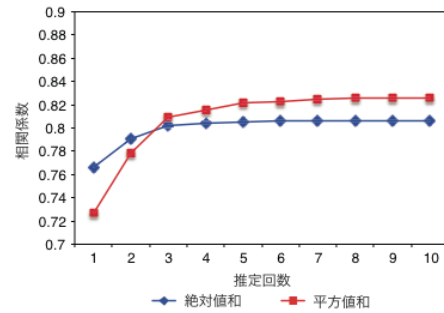


図5 重みの再推定による相関係数の向上

学習者行列と教師行列との差分を表現する差分行列に対して、図1に示すような

図より、自動スコアは教師間相関とほぼ同等の相関値を示しており、十分な精度が出ていることが分かる。これとは別に、教師行列、学生行列を各行で分割し、その行の差分行列要素の二乗和が大きいものほど、対応する音の生成において問題が生じていると判定する手法についても検討した。この場合、検査対象となっていない音が正しく発声できる訳では無いため、差分要素の中で信頼できる要素とそうでない要素とがある。ここでは、従来手法で行なっていた各行の差分要素の二乗和がより大きなものほど発音に不備があると仮定し、不備がある音との距離情報については重みを軽くすることで、音素スコア計算式を更新する方法をとった。図5に示す通り、重み計算を更新することで、より精度の高い発音誤り検出が可能になった。但し、これらの研究で検討したのは母音のみであり、子音が混入した場合の対処は今後の課題となった。

- 3) 学習者分類の妥当性に関しては、まず、音声学者に学習者音声を母音図化してもらった。その際に、通常舌の上下前後関係のみならず、円唇性も記述できるように、三次元母音図を用いた（図6参照）。母音図を用いた学習者分類と、構造表象に基づく学習者分類とを図7に示す。リーフからルートまで辿って行くと、両者の高い類似性を見ることができる。
- 4) 構造表象は、言語学的には構造的音韻論を数学的・物理的に解釈し、定式化したものと考えることができる。これは音の関係性を捉えることで不変性を担保する枠組みであるが、同様の枠組みは音高、即ちメロディーにも当てはまる。

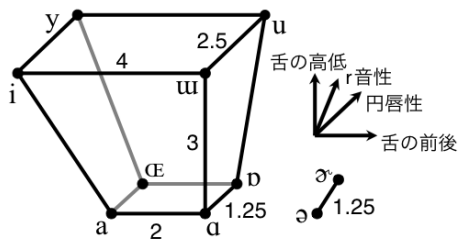


図 6 三次元母音図

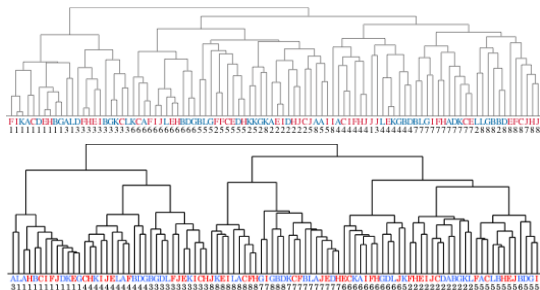


図 7 自動分類と音声学者による分類

音楽の場合、各音の絶対的音高に対する命名が音名であり、各音が他の音とどのような関係にあるのか（どれだけの音高差、音程を持っているのか）により決まる音の機能に基づいた命名が階名である。後者は、音の相対的な特性のみに基づいて求まっており、調不変となる。音声の場合も同様、音色の相対音感という形で音色間距離を適切に定めれば距離だけによって定まる距離行列が変換不変となる。このような類似性より、音声は音色を使ったメロディーとして解釈することも可能である。また、動物は基本的に音高の相対音感を持っておらず、また他固体の声を真似る音声模倣行為も、音真似（声帯模倣的）となる。人間の場合、相対音感を生後獲得し、また、他個体とは音的にずれた声で模倣する。これを音色の相対性より定まる構造を真似ていると解釈すると、音声言語の獲得が困難となる重度自閉症者（刺激そのものを記憶する戦略に長けている）の情報処理モデルの構築が可能となる。これは、動物と共通した情報処理戦略であると言える。これらの理論的考察は、通信学会の招待論文としてまとめることができた。

- 5) 構造表象の持つ話者不変性は、発音評価のみならず、音声認識においても歓迎されるべき特性である。そこで、音声認識の音響特徴量として構造を用いることを検討した。従来の特徴量と異なり、時間的に離れた二イベント間の距離を特徴量としているため、従来の音声認識の枠組み（特にデコーダ）を直接使うことが困難となる。そこで、構造表象を特徴量と

して利用可能なデコーダを試験的に開発し、人工的なタスクにおいてその有効性を検証した。

- 6) 本研究で検討した構造表象は、言語学的には構造的音韻論という古典的な議論と等価であるが、言語教育という観点から見た場合、必ずしも教育者の周知の知識となっている訳ではない。そこで各種の英語教育の研究発表会にて講演（招待講演を含む）を行い、研究成果の発表に努めた。田辺英語教育学研究会、大学英語教育学会・リーディング研究会、早稲田大学大学院言語学専攻科ゼミ、大学英語教育学会関東支部大会、外国語教育メディア学会などの各種研究会、学会で招待講演、基調講演をさせて戴く機会を得、研究成果を広く告知することができた。

5. 主な発表論文等

（研究代表者、研究分担者及び連携研究者には下線）

〔雑誌論文〕（計 8 件）

1. 峯松信明, 鎌田圭, 朝川哲, 牧野武彦, 西村多寿子, 広瀬啓吉, “音声の構造的表象に基づく学習者分類の検証と発音矯正度推定の高精度化”, 情報処理学会論文誌, 査読有, vol. 52, no. 12, pp. 3671-3681 (2011)
2. 鈴木雅之, 峯松信明, 広瀬啓吉, “音声の構造的表象と多段階の重回帰を用いた外国語発音評価”, 情報処理学会論文誌, 査読有, vol. 52, no. 5, pp. 1899-1909 (2011)
3. 峯松信明, “グローバル時代における英語発音とその科学的な分析方法”, 大学英語教育学会関東支部学会誌, 査読有, No. 7, pp. 5-14 (2011)
4. 峯松信明, 櫻庭京子, 西村多寿子, 喬宇, 朝川智, 鈴木雅之, 齋藤大輔, “音声に含まれる言語的情報を非言語的情報から音響的に分離して抽出する手法の提案 - 人間らしい音声情報処理の実現に向けた一検討 -”, 電子情報通信学会論文誌, 査読有, vol. J94-D, no. 1, pp. 12-26 (2011)
5. N. Minematsu, Y. Qiao, S. Asakawa, M. Suzuki, “Speech structure and its application to robust speech processing,” Journal of New Generation Computing, 査読有, vol. 28, no. 3, pp. 299-319 (2010)
6. Y. Qiao, N. Minematsu, “A study on invariance of f-divergence and its application to speech recognition,” IEEE Trans. on Signal Processing, 査読有, vol. 58, no. 7, pp. 3884-3890 (2010-7)
7. 村上隆夫, 峯松信明, 広瀬啓吉, “音声の構造的表象に基づく日本語孤立母音系列を対象とした音声認識”, 電子情報通信学会論文誌, 査読有, vol. J91-A, no. 2, pp. 181-191

(2008)

〔学会発表〕(計 27 件)

1. N. Minematsu, K. Okabe, K. Ogaki, K. Hirose, "Measurement of objective intelligibility of Japanese accented English using ERJ (English Read by Japanese) database," Proc. INTERSPEECH, 査読有, pp.1481-1484 (2011)
2. N. Minematsu, "Human speech model based on information separation and its application to speech processing," Proc. Int. Symposium on Chinese Spoken Language Processing, 査読有, pp.477-482 (2010)
3. D. Luo, Y. Qiao, N. Minematsu, Y. Yamauchi, and K. Hirose, "Regularized-MLLR speaker adaptation for computer-aided language learning system," Proc. INTERSPEECH, 査読有, pp.594-597 (2010)
4. M. Suzuki, Y. Qiao, N. Minematsu, and K. Hirose, "Integration of multilayer regression with structure-based pronunciation assessment," Proc. INTERSPEECH, 査読有, pp.586-589 (2010)
5. M. Suzuki, Y. Qiao, N. Minematsu, and K. Hirose, "Pronunciation proficiency estimation based on multilayer regression analysis using speaker-independent structural features," Proc. Int. Workshop on Second Language Studies, 査読有(2010)
6. D. Luo, Y. Yamauchi, and N. Minematsu, "Speech analysis for automatic evaluation of shadowing," Proc. Int. Workshop on Second Language Studies, 査読有(2010)
7. M. Suzuki, N. Minematsu, D. Luo, and K. Hirose, "Sub-structure-based estimation of pronunciation proficiency and classification of learners," Proc. Int. Workshop on Automatic Speech Recognition and Understanding, 査読有, pp.574-579 (2009)
8. D. Luo, Y. Qiao, N. Minematsu, Y. Yamauchi, K. Hirose, "Analysis and utilization of MLLR speaker adaptation technique for learners' pronunciation evaluation," Proc. INTERSPEECH, 査読有, pp.608-611 (2009)
9. M. Suzuki, L. Dean, N. Minematsu, K. Hirose, "Improved structure-based automatic estimation of pronunciation proficiency," Proc. ISCA Tutorial and Research Workshop on Speech and Language Technology in Education, 査読有(2009)
10. D. Luo, N. Minematsu, Y. Yamauchi, K. Hirose, "Analysis and comparison of automatic language proficiency assessment

between shadowed sentences and read sentences," Proc. ISCA Tutorial and Research Workshop on Speech and Language Technology in Education, 査読有(2009)

11. N. Minematsu, S. Asakawa, Y. Qiao, D. Saito, and T. Nishimura, "Implementation of robust speech recognition by simulating infants' speech perception based on the invariant sound shape embedded in utterances," Proc. Speech and Computer, 査読有, pp.35-40 (2009)
12. D. Luo, N. Minematsu, Y. Yamauchi, and K. Hirose, "Automatic assessment of non-native shadowed utterances," Proc. Int. Symposium on Chinese Spoken Language Processing, 査読有, pp.41-44 (2008)
13. D. Luo, N. Shimomura, N. Minematsu, Y. Yamauchi, and K. Hirose, "Automatic pronunciation evaluation of language learners' utterances generated through shadowing," Proc. INTERSPEECH, 査読有, pp.2807-2810 (2008)
14. N. Minematsu, "Are language learners myna birds? -- a note of warning from a serious speech engineer --," WorldCALL, Poster Session, 査読有(2008)
15. N. Minematsu and T. Nishimura, "Speech as timbre-based melody -- What in parents' voices do infants imitate acoustically? --," Proc. Int. Conf. Language, Music, and the Mind, Paper Session, 査読有(2008)
16. N. Minematsu, T. Nishimura, D. Saito, S. Asakawa, Y. Qiao, "Holistic and prosodic representation of the segmental aspect of speech," Proc. Int. Conf. Speech Prosody, 査読有, pp.169-172 (2008)
17. N. Minematsu, "Training of pronunciation as learning of the sound system embedded in the target language," Proc. The 8th Phonetic Conference of China and Int. Symposium on Phonetic Frontiers, 査読有(2008)

〔図書〕(計 2 件)

1. "人間に近づく音声認識", 日経サイエンス 6月号, pp.94-99, 日経サイエンス社 (2009)
2. "Development of ERJ database for CALL research", in Computer processing of Asian spoken languages, edited by S. Itahashi and C. Tseng, pp.151-154, Consideration Books (2010)

〔産業財産権〕

- 出願状況 (計 0 件)
- 取得状況 (計 0 件)

[その他]

ホームページ等

<http://www.gavo.t.u-tokyo.ac.jp/~mine/japanese/index.html>

6. 研究組織

(1) 研究代表者

峯松 信明 (MINEMATSU NOBUAKI)

東京大学・大学院工学系研究科・教授

研究者番号：90273333

(2) 研究分担者

山内 豊 (YAMAUCHI YUTAKA)

東京国際大学・商学部・教授

研究者番号：30306245

牧野 武彦 (MAKINO TAKEHIKO)

中央大学・経済学部・准教授

研究者番号：00269482