

フォルマント構造推定による日本人用英語発音教示システム

坪田 康 † 壇辻正剛 ‡ 河原 達也 †

† 京都大学大学院 情報学研究科 知能情報学専攻

‡ 京都大学 総合情報メディアセンター/情報学研究科

〒 606-8501 京都市 左京区 吉田本町

日本人の英語母音発音改善のための教示システムの実現および、そのための学習者適応の方法について提案する。本システムでは、日本人学習者の日本語の母音に対する各学習者のフォルマントを計測することで学習者固有のフォルマントの構造とした。この学習者固有のフォルマント、日本語と英語の調音チャート上の関係及び、調音チャートとフォルマントチャートの相関関係を利用して学習者の理想的な英語発音のフォルマント構造を推定する。このフォルマント構造と実際のフォルマント構造を比較し、調音教示情報を提示するシステムを構築した。学習効果確認リスニング試験を行った結果、それぞれに成績の向上が見られた。

キーワード 日本人の英語、フォルマント構造推定、調音教示、学習者適応

English Pronunciation Instruction System for Japanese using Formant Structure Estimation

Yasushi Tsubota † Masatake Dantsuji ‡ Tatsuya Kawahara †

† Graduate School of Informatics

Kyoto University, Kyoto 606-8501, Japan

‡ Center for Information and Multimedia Studies

Kyoto University, Kyoto 606-8501, Japan

We aim to construct a new system to instruct English vowel pronunciation for Japanese learners. We regard formant structure of Japanese vowels pronounced by Japanese learners of English as their own formant structure. We propose an adaptation of the formant structure for English learners. This structure is transformed to learner's ideal English formant structure following the relationship between English and Japanese articulation charts which are mutually concerned with formant structure. After the learners' initial English pronunciation is input, it is compared with the ideal one, and articulatory instructions are output. To assess the effect of this system, some students took listening tests before and after learning, and improved their scores.

key words Japanese English, formant structure estimation, articulatory instruction, learner adaptation

1 研究の背景

近年の音声分析技術の発展に伴い、計算機支援型発音学習システムの開発が活発に行われている [1][2][3]。しかし、従来の発音学習システムの主流は学習者が自分で発音を比較・評価したりするものや、計算機による評価は可能であっても教示は与えられないものである。

本研究では、日本人用英語発音教示システムを構築することにより、上記の問題を改善しようとするものである。今回は母音を対象とした。理由は、日本語と比較して英語母音の音素の数が格段に多く、学習者にとって判別が困難であり、子音より学習が困難であることが挙げられる。

また、母音の重要な音響特徴であるフォルマントには個人差が存在するので補正する必要がある。しかし、学習者自身の発音する英語発音はそもそも話者情報として信頼性に欠けるため、そこから目標とすべき英語発音を推定することは困難である。そこで本研究では、学習者の母国語である日本語の母音を分析し、それを英語に写像することによってフォルマント構造推定を行い、学習者適応とした。

本システムでは、音素認識とフォルマント分析による評価から調音教示を行う。このシステムを利用した学習者は、調音特徴に対して直接教示を得られるので比較的容易に発音を習得することが可能であると思われる。また、母音の調音上の特徴と音響上の特徴の関連付けが強化されるためリスニング能力の向上にも役立つであろう。

以下、2章で本システムで考慮した日本人の英語母音発音の誤り傾向について述べた後、3章で音素認識器の構成、4章でフォルマント分析、5章で全体の構成について述べ、6章で評価実験について述べる。

2 日本人の誤り傾向

日本語と英語の差異は多く存在する。例えば、日本語は一般に開音節言語でピッチアクセントの言語であるのに対して、英語は閉音節で強さアクセント(強勢)の言語である。実際の学習においてはこのような差異を考慮した学習が望まれる。本システムを構築する際に考慮した誤りを整理する [4][5]。

2.1 母音の挿入

日本語は開音節言語であるために、日本人は子音で終る単語の発音になれていない。そのため、“beat”/bit/という発音を/bitto/と誤りやすい。これは、注意を喚起すれば矯正可能な誤りである。本システムはこの誤りに対して挿入が確認されたと提示する。

2.2 単母音の誤り傾向

日本人は英語の音を日本語の似た音で発音する傾向がある。また、日本語と英語では同じ調音チャート上の領域でも同じ数の音素で被覆しているとは限らない。例えば、英語では、/i/と/e/と/i:/で覆われている領域は日本語では/i/と/ε/で覆われている。また、/a/と/Λ/と/æ/で覆われる領域は日本語の/a/一つでほぼ覆われる。このため、発音の混同がよく生じる。¹

日本人にとってなじみの深い音響的な特徴で区別する傾向の強い英語の母音の組み合わせも存在する。音響特徴や音色も異なるにも拘らず、/i:/と/i/とは母音の持続時間の長短で区別をしようとする傾向が強い。例えば、短く発音しても弛緩母音を緊張母音で発音してしまうことにより、“live”と“leave”の混同が起こり、意味上の誤解を与えてしまうこともある。

2.3 二重母音の誤り傾向

二重母音は、日本語と英語では発音の仕組みが異なる。日本語の連鎖母音は、音素が二つ続く連母音であるのに対して、英語の二重母音は一音節中に二種類の母音が存在し、一つの要素が他の要素よりも顕著に現れる。これに対して、日本語は単母音が二つ続くだけである。日本語の連母音のように英語の二重母音を発音してしまう傾向が強い。例えば、日本語の「愛」のように英語の“eye”を発音することが多いが異なる音になることに注意しなければならない。

二重母音と長母音の区別は日本語では重要でないのに対して、英語では重要である。例えば、日本語では、「きょう」と「きょおと」はどちらも「京都」として扱うことが多い。それに対して、英語では“boat”

¹ / / は音素記号, [] は音声記号を表す。ただし、本稿では論旨の展開上英語との区別を分かりやすくするために /あ/, /い/, /う/, /え/, /お/ の表記を用いた。

と“bought”は異なる母音音素であるため日本人が混同しやすい。

以上の傾向をふまえて、日本人が誤りやすい母音の最小対 (minimal pair, minimum pair)² の単語が研究されている。これを表 1 に示す。本システムでは、このような誤りに対応するようにシステムを設計した。

表 1: 日本人が誤りやすい最小対語

1	bott[bat]	bat[bæt]	but[bat]
2	beat[birt]	bit[bít]	
3	boot[burt]	book[bʊk]	
4	bit[bít]	bet[bet]	
5	bought[bɔ:t]	bott[bat]	
6	bait[bet]	bet[bet]	
7	bought[bɔ:t]	boat[bɔ:t]	

3 音素認識による評価・教示

3.1 音素認識器による評価の問題点

英語の音素認識器は英語の母音の音素の中から、学習者の発音に最も近いものを出力する。この音素認識器による評価と選択された単語の母音を比較して調音上の特徴の誤っている部分を教示する。

しかし、この方法には幾つかの問題点がある。一つは、音素音素認識器により、/bat/が/bat/に誤ったという教示は、それぞれの特性を知っている専門家にとっては意味があるが、初学者にとっては分かりにくい教示となる。従って、初学者にとっても理解しやすい教示を考慮しなければならない。

また、英語音素認識器の尤度は相対的なので、実際に有効に区別して発音できているかどうかを判定する基準としては必ずしも信頼できない。

本システムでは、学習者に分かりやすい教示とするために、認識音素とターゲットの音素に対して調音特徴を描いた図を提示する。また、どの程度近づいているかの指標は次章のフォルマント分析により行った。

² 最小対とは、2つの語において、ある単音もしくは2つの単音の複合を除いて他の音声部分と同じである場合、こうしたペアを最小対 (minimal pair, minimum pair) と呼ぶ。最小対の単語を見つけ出すことにより、音素目録を作成することが可能となる。

3.2 英語音素認識器の構成

単文読み上げ音声データベースである TIMIT データベース (男性話者 438 名のべ合計 6,300 文のアメリカ人のネイティブスピーカーによる発話を収録) を用いて英語音素モデルの構築を行った。詳細は図 3 に示す通りである。

この音響モデルを評価する実験を行った。評価の際の文法は全ての音素が自由に生起し得る文法とし、テストデータとして、TIMIT データベース付属のテストセットを利用した。結果を表 4 に示す。

実際のシステム使用時には、音素環境が同じで母音のみが異なる単語を学習単語として利用しているため、図 5 に示すような文法を用いた。また、音素認識アルゴリズムとしては Viterbi アルゴリズムを用いた。

表 2: 英語音声データベース TIMIT

地域	男性話者数
New England	31
Northern	71
North Midland	79
South Midland	69
Southern	62
New York City	30
Western	74
Army Brat	22
(moved around)	
total	438

表 3: 英語音響モデルの仕様

音響特徴ベクトル	MFCC(13次元) +ΔMFCC +ΔΔMFCC(計 39次元)
音素の種類数	63

表 4: 英語音響モデルの精度

混合分布数	correctness	accuracy
16	53.52	50.33

表 5: 使用文法

評価用文法	<i>sil</i> < 音素 > <i>sil</i>
システム用文法	<i>sil</i> b(母音)t(<i>sil</i> 母音)

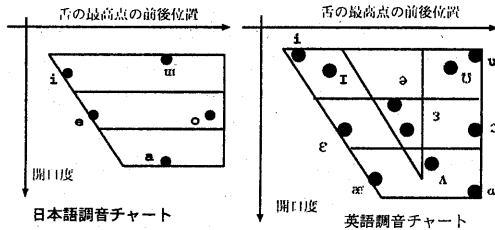


図 1: 母音調音チャート

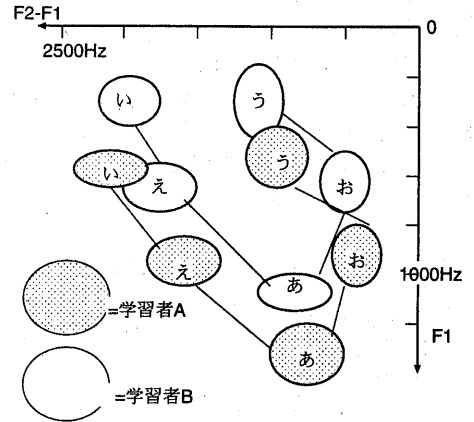


図 2: フォルマントチャートの例

3.3 調音図とテキストによる教示

母音は3つの調音特徴、開口度、舌の最高点の前後位置、唇の形状（円唇対非円唇）により記述される。この知識を用いて各言語の母音調音を記したのが1に示したような母音調音チャートである。この3つの特徴を正しく調音することができれば、理論的には正しい発音が可能である。実際には、その調音を身につけるために試行錯誤が必要である。その過程において、調音図による教示は有効であろう。さらに調音図の理解を助けるためテキストによる教示も行う。

4 母音のフォルマント分析による評価・教示

4.1 母音フォルマント分析の問題点

前章で述べたように、母音は3つの調音上の特徴で記述できる。音声の共鳴周波数の倍音であるフォルマントは、この調音上の特徴と相関関係があることが知られている。それぞれ、開口度は第一フォルマント（以下F1）、舌の最高点の前後位置は第二フォルマント（以下F2）と第一フォルマントの差と、唇の円唇化は第三フォルマント（以下F3）等の高次フォルマントの低下と対応していることが知られている[7]。この関係を用いて、本システムは、学習者の母音の調音を評価し教示情報を与える。

フォルマント分析を用いる際に、幾つかの問題点が存在する。母音のフォルマントは前後の子音の調音位置の影響を受けてフォルマントが変動する。これは、子音の調音位置によって母音がその前の状態を残したまま母音の調音をするためである。この前後の部分はフォルマント遷移と呼ばれている[8]。このフォルマント遷移の影響は母音によっても影響の受け方が異なるので、この影響の排除が課題となってくる。

また、母音のフォルマントの値は、個人によって声道の長さが異なるため話者毎に異なる値を持つ。そのため、学習をする際には何らかの形で学習者にシステムを適応させる必要がある。以下でフォルマント分析と調音特徴について述べた後、それぞれのアルゴリズムについて説明する。

4.2 フォルマント遷移の影響の排除

母音のフォルマントの値は、前の子音の調音点に近いところから始まり、中央部は比較的定常で、後部は後ろの子音の調音点へと向っていく。この影響をできる限り排除していくために本システムでは音素認識器が母音と判定した部分のさらに中央部8割部分のフォルマントの平均値をフォルマントの値としている。

4.3 学習者適応

母音のフォルマントは話者毎に異なる。例えば、図2に示すように学習者Aの/い/の音と学習者Bの/え/の音が重なる場合もある。これは、教示システムを構築する上で非常に重要な問題で、いかに話者にシステムを適応させ教示を出すかが重要となっている。

従来の話者適応技術では、5母音の相対関係は常に成り立つことを用いて幾つかの母音を要素とする母音系列を相対関係に基づいて評価し、認識する手法や幾つかの母音標準パターンを予め用意しておき、最も近いと判定された標準パターンに修正を加えて行く方法などが用いられた [9][10]。

また、最近では話者の声道長に着目した話者正規化手法が提案されその有効性が報告されている。これを利用した話者適応法も幾つか出されている [11][12]。

しかし、学習者の英語母音発音は安定でなく、また信頼性に欠けるため、そこから学習者の理想的な発音のフォルマント構造を推定することは困難である。したがって、本システムでは学習者の母国語である日本語の母音のフォルマント構造から英語のフォルマント構造へと写像する。

日本語の5母音の内F1とF2-F1の値の中で最小値と最大値をとるものを大小関係の比較により決定する。この組み合わせによる4点で囲まれた領域は言語にかかわらず話者特有の調音領域であるとみなして、その領域中に英語の調音を反映したフォルマントチャートを構築した。

また、唇の形状に対する教示に関しては行っていない。その理由として、日本語では唇の形状、特に円唇化を伴わずに発音する場合もあるので、各話者毎の特徴を掴むのが困難であるためである。

4.4 フォルマントによる調音上の特徴の評価

本システムでは、学習者の理想フォルマントの分布と実際のフォルマントの値をF1, F2-F1, F3の値でそれぞれ比較し調音の評価としている。例えば、英語の理想の第一フォルマントの値より実際の第一フォルマントの値が小さければ、学習者の発音は口の開き具合が小さいと評価する。実際のフォルマントチャートを図4に示す。

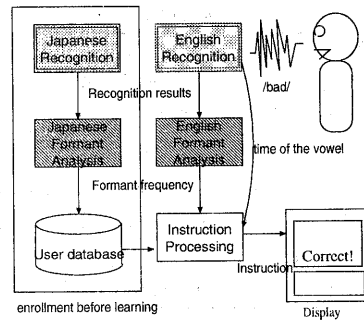


図3: システム構成図

4.5 グラフとテキストによる教示

前節で述べたように、低次フォルマント周波数の値は変換することによって調音上の特徴の評価として用いることができるので、教示方法としては、二次元グラフに口の開き具合いと舌の最高点の前後位置をプロットし、唇の形状は値を残すことで比較できるようにしている。また、グラフの読み取りを容易にするために、下方のテキストボックスに文字による解説を提示した。

5 システムの構成

システムの構成を図3に示す。

5.1 音素認識器の構成

本システムでは、日本語音素認識と英語音素認識の両方を用いる。英語音素認識器の構成は、前に述べたのでここでは割愛する。音響モデルとしては、IPA98年のモデルを使用し [6]、言語モデルは、音素環境を整えるため、/k-V-r-V/とした。システムの使用時には、「狩る」、「着る」、「来る」、「蹴る」、「凝る」の単語を順に発声してもらっている。Viterbiアルゴリズムにより認識を実行した。日本語の音響モデル、言語モデルの仕様をそれぞれ表6、表7に示す。

5.2 フォルマント分析

日本語発音と英語発音に対して、母音部分のフォルマントを計測するために、フォルマント分析を行

表 6: 日本語音響モデルの仕様

学習セット	男性 130 名による 読み上げ音声
音響特徴ベクトル	MFCC(12 次元) + Δ MFCC + Δ Pow(計 25 次元)
CMS	発声単位で実行
音素の種類数	43
出力分布	16 連続混合

表 7: 使用文法

システム用文法	<i>sil - k - (母音) - r - (母音)</i>
---------	----------------------------------

う. 音素認識器より受取った母音の時間情報から, 前後の子音の影響部分を排除して, 中央のフォルマント周波数の平均と分散を計算する.

5.3 教示処理

音素認識結果と選択単語に対応した調音図を表示, 同時にテキストによる調音図の解説も下のテキストボックスに表示する. また, F1 と F2-F1 の値をグラフ上に示す. また, その値の大小比較によりグラフの解説をテキストで表示する. 実際の教示情報例を図 4 および図 5 に示す.

6 実験・評価

本システムの実験の評価を行った. まず, 英語と日本語のバイリンガルスピーカーによってフォルマント構造推定手法の妥当性を検証し, その上で実際の学習効果を調べた.

6.1 フォルマント構造推定の検証

日本語フォルマント構造から推定された英語フォルマント構造が正しく英語調音を反映しているかを検証する実験を行う. そのために, 英語のネイティブスピーカーの協力を仰ぎ, 推定英語母音フォルマントと実際の英語母音フォルマントの値を比較する. 具体的には, 日本語と英語のバイリンガルであるアメリカ人に日本語母音サンプルを発音してもらい,

その日本語フォルマントから英語母音フォルマントを推定した.

その結果, おおむね正しい位置に分布していることが明らかになった. おおむね, 推定位置と一致しているが "bought" と "bott", "book", "boot" が内部にない. また, 母音の分布に重なりがある場合がある. 実際の分布を 6 に示す.

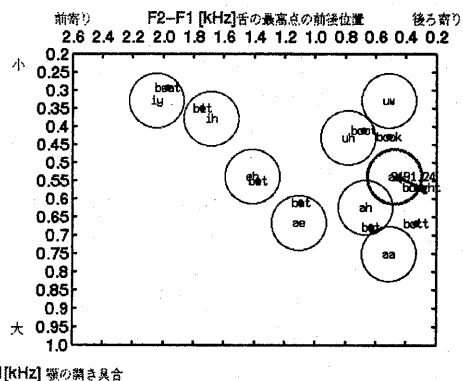


図 6: 推定フォルマント構造分布と実際のフォルマント構造分布

図 6: 推定フォルマント構造分布と実際のフォルマント構造分布

6.2 学習効果の検証

本システムを利用して, 実際に教育効果を確認する実験をおこなった. ATR のヒアリングソフト "HearingSchool" の母音弁別試験を 5 名の成人男性学習者に本システムの利用前と利用後に受けてもらい, 成績の変化を調べた. また, 学習者自身の感じる学習効果についてアンケートを行った.

リスニング能力を効果確認とした理由としては, 調音と知覚は密接な関係にあることや, 発音の評価は一般に難しいことが挙げられる. 調音と知覚が密接な関係を示す理論としては, 「音声知覚の調音参照説」や「音声知覚のモータ理論」等の仮説がある [13].

この仮説を基に, 5 名の学習者に対して実験を行った結果, おおむね学習効果があることが確認された. 実際の試験の成績の変化を表 8 に示す. また, システム使用後には, 注意すべき音響特徴が理解できたと全ての学習者がコメントしていた. また, 調音の評価による教示の効果により, 音響特徴と調音特徴との対応をとることができるようになったと全ての

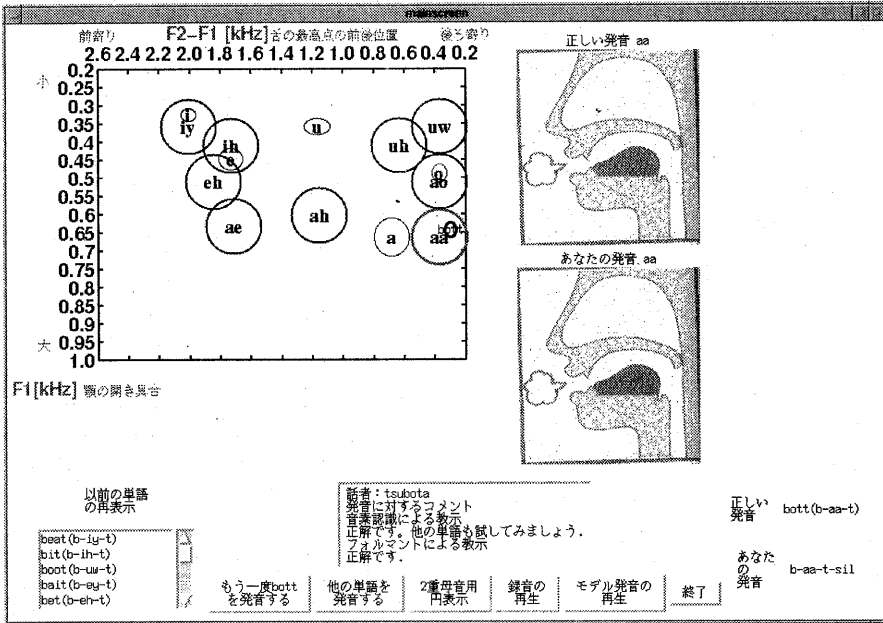


図 4: 教示情報例 (正解)

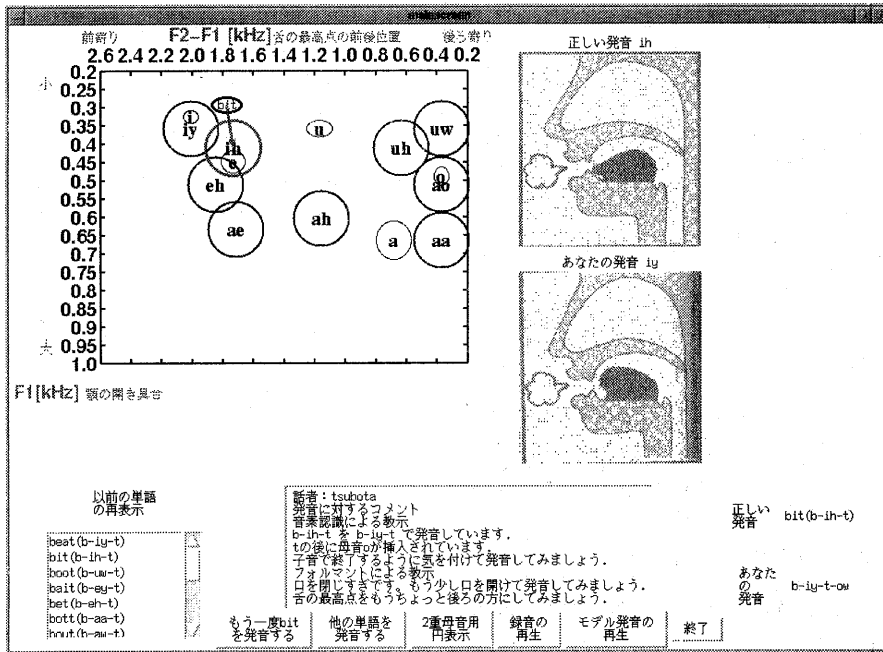


図 5: 教示情報例 (不正解)

学習者がコメントしていた。

この結果より、本システムの学習効果を確認することができた。

表 8: システム利用による成績の変化

学習者	A	B	C	D	E
学習前	60	80	85	75	85
学習後	85	95	90	75	90

7 まとめ

調音指導による母音習得支援システムの実現及びその評価を行った。本システムでは、学習者の英語フォルマント構造推定を学習者の母国語である日本語のフォルマント構造からの写像により行い、実際の英語発音フォルマント構造との比較・評価により調音への指導を実現した。

また、実験により本システムにおける学習者適応の正当性の一端が証明されることとなった。また、学習効果もおおむね確認することができた。

今後の課題として、音素認識精度の向上が挙げられる。その上で、多くの被験者による学習者適応の評価を試みる予定である。

謝辞 龍谷大学理工学部教授堂下修司先生のご助言と関西大学文学部教授 Scott Johnson 先生、京都大学総合情報メディアセンター助手清水政明先生のご協力に深謝致します。
当研究の一部は文部省科学研究費課題番号 20011819 によるものである。

参考文献

- [1] Jo, C.-H., Kawahara, T., Doshita, S. and Dantsuji, M.: Computer-Aided Pronunciation Learning System based on Classification of the Place and Manner of Articulation, Technical report, 電子情報通信学会 (1998).
- [2] 河合剛, 広瀬啓吉: 二言語間の音響モデルを用いた音声認識による非母語話者の発音誤り検出, 日本音響学会講演論文集, pp. 221-222 (1998).
- [3] 中川聖一, Reyes, A. A., 鈴木英之, 谷口泰広: 音声認識技術を利用した英会話 CAI システム, 情報処理学会論文誌, Vol. 38, No. 8 (1997).
- [4] 巽一朗: 英語の発音がよくなる本, 中経出版 (1998).
- [5] Averil, P. and Ehrlich, S.: *Teaching American English Pronunciation*, Oxford University Press, chapter 8 (1997).
- [6] 河原達也 李晃伸 小林哲則 武田一哉 峰松信明, 伊藤克亘, 伊藤彰則, 山本幹雄, 山田篤, 宇津呂武仁, 鹿野清宏: 日本語ディクテーション基本ソフトウェア (97年度版), 日本音響学会誌, pp. 175-180 (1999).
- [7] Ladefoged, P.: *A Course in Phonetics*, Harcourt Brace College Publishers, chapter 8 (1993).
- [8] D.Kent, R., Charles Read (荒井 隆行 菅原 勉訳): *The Acoustic Analysis of speech* (音声の音響分析), Singular Publishing Group (海文堂), chapter 6 (1996).
- [9] 杉山雅英, 好田正紀: 認識結果を利用した母音標準パターンの教師無し話者適応化法, 電子通信学会論文誌, Vol. 69, No. 8, pp. 1197-1204 (1986).
- [10] YU, R., 木村正行: 母音間の相対関係モデルに基づく不特定話者母音認識, 電子通信学会論文誌, Vol. 69, No. 9, pp. 1320-1327 (1986).
- [11] 阿部俊朗, 鈴木基之, 牧野正三, 阿曾弘具: 音素毎の話者クラスターリングに基づく話者適応法, 情報処理学会研究報告, 97-SLP-19-7 (1997).
- [12] 内藤正樹, 匂坂芳典, Deng, L.: 声道の特徴量を用いた話者クラスターリング手法の検討, 情報処理学会研究報告, 97-SLP-19-6 (1997).
- [13] 山田恒夫, 足立隆弘: 英語リスニング科学的上達法, ATR 人間情報通信研究所 (1998).