Vol.2011-SLP-89 No.15
2011/12/19

発音訓練のための調音特徴に基づく IPA 母音図へのリアルタイム表示

森拓郎† 入部百合絵 桂田浩一 新田恒雄

近年，外国語を学ぶ日本人学生を対象に CALL 教材の開発が盛んに行われているが，学習者の調音動作の誤りを正確に指摘できる教材はまだ開発されていない．我々は，教師と学習者の調音の違いを分かり易く理解でき，さらに正しい調音動作への矯正方法を直感的に読み取ることのできる，英語発音訓練システムの開発を進めている．本報告では，母音発音に焦点を当て，学習者音声の調音点を IPA 母音図上に表示することができる，英語発音マップシステムとその評価について報告する．提案するシステムは，学習者の音声から多層ニューラルネット(MLN)を用いて調音特徴を抽出し，二次元平面上へ座標変換することにより，英語発話母音をリアルタイムに IPA 母音図上へプロットする．英語母語話者に対する評価実験を行った結果，評価対象の全ての英語母音について，良好なプロット精度が得られた．

Real-time Visualization of English Pronunciation on an IPA Vowel-Chart Based on Articulatory Feature Extraction

Takurou Mori† , Yurie Iribe† , Kouichi Katsurada† , and Tsuneo Nitta†

CALL systems that can support Japanese students to study foreign languages have been developed in recent years. We have been developing an English pronunciation training system that enables learners to evaluate their pronunciation at an articulation level by using an articulatory-feature (AF) extractor. In this paper, firstly, an English pronunciation training system that can plots articulatory manner/place of a learner's English pronunciation on an International Phonetic Alphabet (IPA) vowel-chart in real time is described. The system converts an input utterance into an AF-sequence by using multi-layer neural networks (MLNs) of the AF extractor, then the AF-sequence is converted to the coordinate on an X-Y dimensional surface and plotted on an IPA vowel-chart that can shows the correctness of his/her articulation. In the experiments of native speakers on a TIMIT corpus, satisfactory accuracy of plotting was achieved.

1. はじめに

近年，大学や高校などの教育機関では外国語発音の自学自習用に様々な CALL 教材が導入されている．また一般の学習者がインターネット上で外国語発音の自主学習を行える音声認識技術を用いたオンライン CALL 教材も開発されている[1]．一方，英会話教室や教育機関では教師がモデルの音を学習者に聞かせるとともに，舌や口唇など調音器官の発話時の動作イメージ（以下，調音動作と呼ぶ）を学習者にうまく伝えながら，正確な発話ができるように指導している．正しい発音を身につけるためには学習者の調音動作に対して的確な指示を与えることが重要である．

これまでに学習者の発音と正しい発音の調音点を視覚的に評価する機能を備えた CALL 教材として Sonic Print[2]が開発されている．Sonic Print は学習者音声のフォルマント周波数をリアルタイムに分析し F1-F2 平面上にプロットするため，学習者は自分の発音が教師の正しい母音を示す領域に収まるよう繰り返し発音を練習する[3]．F1-F2 平面上の各母音領域は口腔断面図に描かれる調音位置と対応しているが，具体的な調音位置の情報（舌の位置や口の高低）はフォルマント周波数とプロットの位置から読み取らなければならず，音声学の知識を持たない学習者が調音動作を改善するためには非常に分かりづらい．そのため，母語に存在しない調音動作を獲得することに多大な時間を要する．

我々は教師と学習者の調音点の違いが一目で理解でき，さらに正しい調音動作への矯正法を直感的に読み取ることのできる英語母音訓練システムについて検討している．大学の発音教育においては、口唇の開き具合や舌の位置を模式的に表した国際音声記号（International Phonetic Alphabet: IPA）による図表（以下，IPA 母音図と呼ぶ）がよく用いられる [4]. IPA 母音図を活用することで、正しい母音の調音位置と学習者の発音の調音位置の違いを容易に確認することができる．そこで，本論文では調音特徴に基づき学習者の調音点を IPA 母音図上にリアルタイムに表示する英語発音訓練ソフトについて述べる．

† 豊橋技術科学大学 大学院工学研究科
Graduate School of Engineering, Toyohashi University of Technology

Vol.2011-SLP-89 No.15
2011/12/19

2. 調音特徴に基づく発音訓練システム

本稿では，調音特徴とその抽出手順について述べた後，我々が開発している調音特徴を用いた日本人向け英語発音訓練システムと IPA 母音図へのリアルタイム表示について詳述する．

2.1 調音特徴

調音特徴（Articulatory Feature; AF）は，単音分類に用いられる調音様式（母音，子音，有声，無声など）と調音位置（前舌，半狭，半広，など）の諸属性を指す．表 1 に示すように，AF では，あらゆる音素は調音特徴の有無(+/ -) を示すベクトルで表現できる．AF を音声認識で利用する際の利点は，調音的に近い音素同士を距離の近いベクトルとして表現できることである．

今回用いた調音特徴セットは，国際音声記号（International Phonetic Alphabet: IPA）から英語に関する部分を取り出したものであり，次元数 28 次元，音素数 44 (sil を含む)から構成される．ここに述べた調音特徴セットは，後述するニューラルネットワークの学習において教師信号として用いられる．

2.2 調音特徴の抽出

図 1 に調音特徴抽出器の構成を示す．まず，AF 抽出器に入力された音声は局所特徴（Local Feature, LF）に変換される．LF の抽出手順を図 2 に示す．入力音声は，16kHz でサンプリングされた後, 25ms のハミング窓で 10ms 毎に 512 点の FFT 処理を受ける．

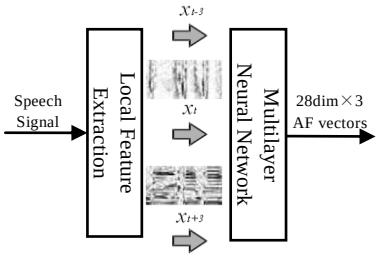

図 1 調音特徴抽出器の構成

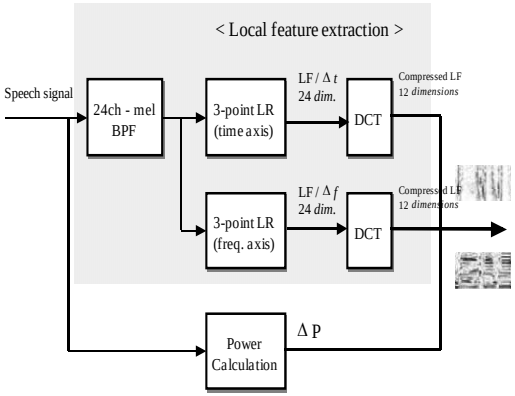

図 2 局所特徴抽出過程

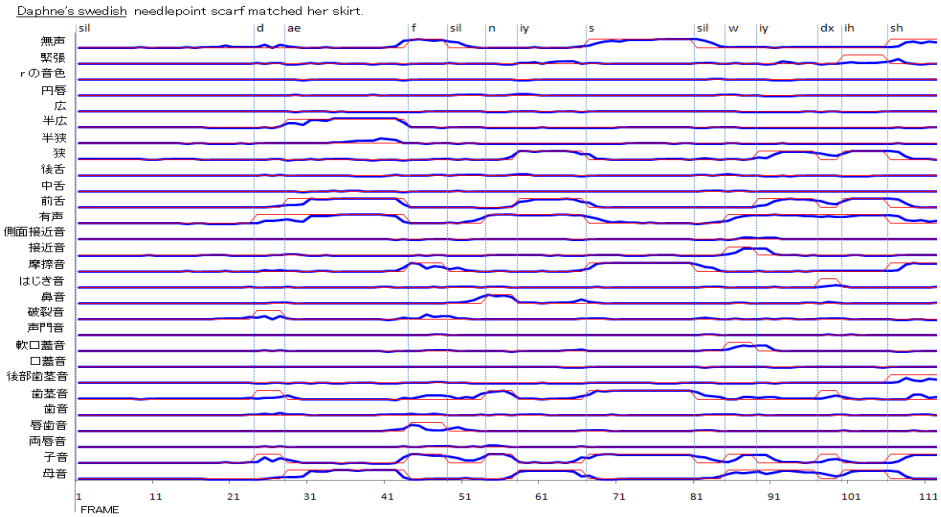

図 3 調音特徴系列の例(細線：教師信号，太線：MLN 出力)

この結果はパワースペクトルの形で積分され，中心周波数を（聴覚に近似した）メル尺度間隔で設計した 24-ch の BPF (Band Pass Filter) 出力にまとめられる．ここまでが分析処理である．続いてパワースペクトル系列上の音響特徴抽出が行われるパワースペクトル系列が構成する曲面は，多様体として見ると時間と周波数方向の局所的な微分要素で表現できる（微分多様体）．

そこで，BPF 出力を 3×3 の局所特徴に変換するため，時間軸と周波数軸上に各々3 点の線形回帰 (Linear Regression; LR)演算を行い，微分特徴としての LF を抽出する．二つの局所特徴は各 24 次元であるが, 続いて離散余弦変換 (Discrete Cosine Transform; DCT)処理によって半分の 12 次元に圧縮される。これに対数パワー成分の微分要素を加えた 25 次元の特徴が LF である．

LF は，多層ニューラルネットワーク(Multilayer Neural Network; MLN)によって AF へ変換される．入力の LF と出力の AF には，ともに注目フレーム$\mathbf{x_t}$ と前後 3 点離れたフレーム（$\mathbf{x_{t-3}}$, $\mathbf{x_{t+3}}$）を用いた． すなわち，入力は 75 次元（25×3）の LF，出力は 84 次元（28 × 3）の AF である．図 3 に MLN の出力例を示す.注目フレーム$\mathbf{x_t}$ だけでなく，前後 3 点離れたフレーム($\mathbf{x_{t-3}}$, $\mathbf{x_{t+3}}$) のスペクトル情報を含むことで，AF への変換精度が向上する. 学習はラベル付き音声データを用いて行い, ＋の属性を 0.9, －の属性を 0.1 とし，誤差逆伝播法を用いた．

表 1 英語母音調音特徴セット

	iy	ih	ey	eh	ae	aa	ay	aw	ao	ow	oy	uh	uw	ah	er	ax
母音	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
子音																
有声	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
前舌	+	+	+	+	+		+	+			+					
中舌															+	+
後舌						+		+	+	+	+	+	+	+		
狭	+	+	+				+	+		+	+	+	+			
半狭			+	+						+						+
半広				+	+				+		+			+	+	
広						+	+	+								
円唇									+				+			
r の音色															+	
緊張		+	+				+	+		+	+	+				

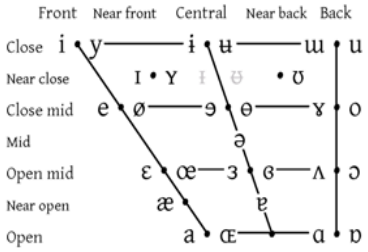

図 4 IPA 母音図[5]

2.3 英語発音訓練ソフトウェア

我々は，調音特徴に基づく音声認識技術を用いた英語発音訓練ソフトウェア「日本人のための発音先生 –英語編-」の開発を行なっている．本ソフトウェアは正しい英語発音を身につけるために，調音特徴を用いて音素単位で学習者の発音を評価し，誤りがある場合はその原因となる調音動作の違いを指摘し正しい発音へと導く．本ソフトウェアを用いた学習の手順を以下に示す．

(1) 日本語の音と混同しやすい発音（例えば，日本語の「ア」に対する英語の /ə/,/ʌ/,/ɑ/,/æ/）がまとめられたメニューの中から学習したい音素を選択する．
(2) (1)で選択した音素を含む練習単語の一覧が表示されるので，練習する単語を選択する．
(3) 学習者が単語を発声すると音声認識器により音素毎にその正しさが評価され，誤った音素に対してはその誤り内容が示される．
(4) 誤った調音動作を矯正するために発音マップを起動する．

(4)で学習者に提示される発音評価は，図 に示すようにネットワーク文法の形態で表示され，学習者の発音のどの部分が誤っているのかを音素単位で示す．音声認識器により音素単位での発音誤りを示すことはできるが，その誤りをどのように矯正するかが重要である．

そこで，学習者の発音について調音動作レベルで誤りを指摘し，矯正を行うため，2.4 に述べる英語母音の調音動作を IPA 母音図上にリアルタイムで表示する機能（以下，英語母音発音マップと呼ぶ）を実装した．

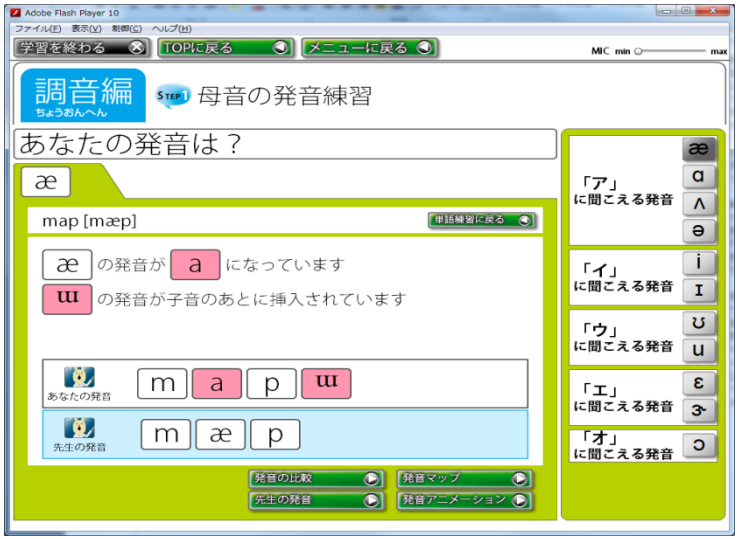

図 5 発音訓練ソフト発音評価画面

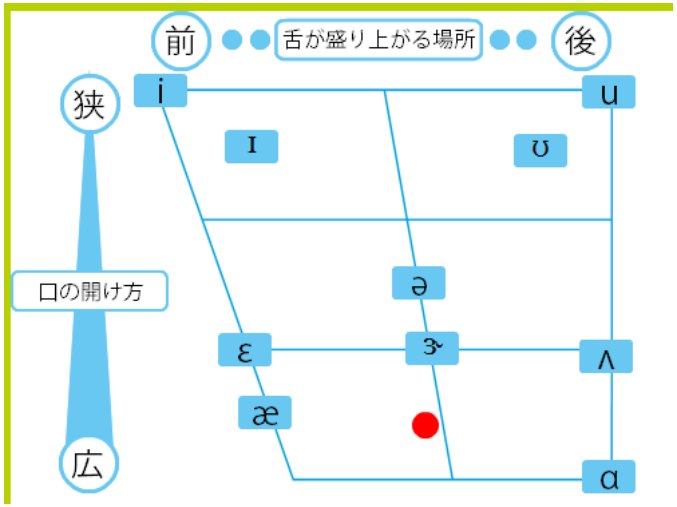

図 6 発音マップ画面例

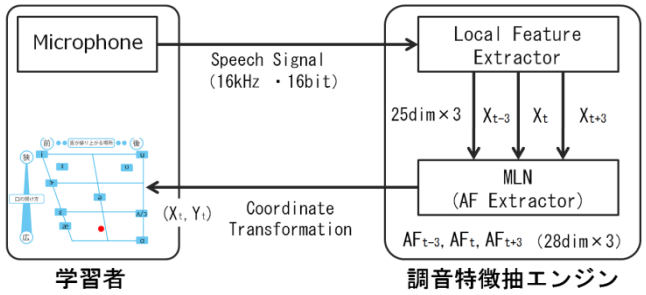

図 7 発音マップシステムの基本構成

2.4 調音特徴による IPA 母音図上へのリアルタイム表示

英語母音発音マップの画面例を図 6 に示す．発音マップは図 4 の IPA 母音図([5]より引用)を模した梯形図に発音記号が配置され，口唇の開き具合を示すスケール，舌の盛り上がる位置を示すスケール，そしてユーザの発音位置を示す赤い光点からなる．

光点は学習者の発音の調音位置に対応した座標点にプロットされるため，学習者は発音しながら自分の調音動作を確認することが可能である．光点が目標とする発音記号に近いほど正しく発音できていることを示しているため，学習者は口唇の開き具合と舌の盛り上がる位置のスケールを参考にして，調音を徐々に修正しながら漸近的に調音動作を矯正することができる．

英語母音発音マップシステムの構成を図 7 に示す．具体的なシステムのイメージは以下の通りである．

1）システムが学習者の発声を検知する
2）調音特徴抽出器により 10ms 毎に 48 次元の調音特徴を抽出する．
3)抽出された調音特徴系列の母音に関係する特徴列を座標変換器に入力して 2 次元平面上の X,Y 座標に変換する．
4）発音マップ上の光点を変換後の座標へ移動する．

次に座標変換器内で実行される調音特徴から X,Y 座標への変換アルゴリズムを以下に示す．

1）X 座標ついては「前舌音(AF_{front})」「中舌音($AF_{central}$)」「後舌音(AF_{back})」の特徴量をもとに図 8 図 に示す手順で変換される
2）Y 座標については「狭母音(AF_{close})」「半狭母音(AF_{close_mid})」「半広母音(AF_{open_mid})」「広母音(AF_{open})」の特徴量に基づき図 9 に示す手順で変換する

なお, D_{width}は発音マップの X 方向の長さ, D_{height}は発音マップの Y 方向の長さを示す．

図 8,図 9 に示すアルゴリズムは，最大値を取る特徴量のみではなく，調音特徴系列上で隣接する特徴量も用いて座標を決定している．/ə/のように調音位置の変動が大きい音素は条件によって隣接する特徴にも+が付与されるため，その特徴も座標変換に用いることで，「/ɪ/ に近い/ə/」といった微妙な発音に対しても安定してプロットを行うことができる．

座標変換器から得られる座標値を発音マップ上にプロットすると，マップが台形であるためマップ下部では正確にプロットすることができないため，適宜変換を行った後にマップ上へプロットする．ただし，後述する評価実験では正しいマッピングとの距離の差を簡単化するため，座標変換器から得られた座標に基づいて評価した．

図 8 調音特徴から X 座標への変換アルゴリズム

if AF_{front} が最大
　　$X = (AF_{central} / AF_{front}) * (D_{width} / 4)$
else if $AF_{central}$ が最大
　　if $AF_{front} > AF_{back}$
　　$X = D_{width} * (1/2) – (AF_{front} / AF_{central}) * (D_{width} / 4)$
　　else
　　$X = D_{width} * (1/2) + (AF_{back} / AF_{central}) * (D_{width} / 4)$
else
　　$X = D_{width} - (AF_{central} / AF_{front}) * (D_{width} / 4)$

図 9 調音特徴から Y 座標への変換アルゴリズム

if AF_{close} が最大
　　$Y = (AF_{close\_mid} / AF_{close}) * (D_{width} / 6)$
else if AF_{close_mid} が最大
　　if $AF_{close} > AF_{open\_mid}$
　　$Y = D_{height} * (1/3) - (AF_{close} / AF_{close\_mid}) * (D_{width} / 6)$
　　else
　　$Y = D_{height} * (1/3) + (AF_{open\_mid} / AF_{close\_mid}) * (D_{width} / 6)$
else if AF_{open_mid} が最大
　　if $AF_{close\_mid} > AF_{open}$
　　$Y = D_{height} * (2/3) - (AF_{close\_mid} / AF_{open\_mid}) * (D_{width} / 6)$
　　else
　　$Y = D_{height} * (2/3) + (AF_{open} / AF_{open\_mid}) * (D_{width} / 6)$
else
　　$Y = D_{height} - (AF_{open\_mid} / AF_{open}) * (D_{width} / 6)$

3. 評価実験

英語母音発音マップは学習者の発音を 2 次元平面である IPA 母音図上にプロットし，マップ上の発音記号との相対的な位置から発音動作の違いを視覚的に教示するものである．従って，ネイティブ英語発音に近い発音がなされた場合は，発音記号と同じ座標上にプロットされることが理想である．そのため，今回は英語母語話者の発話から抽出した調音特徴をマップへ変換した座標値と各英語母音の正解座標を比較し，開発した英語母音発音マップの精度を評価する．本実験に用いた音声資料は TIMIT[6]であり，詳細を以下に示す．

D1：学習セット（MLN 学習用）
TIMIT 2600 文，男性 325 名（16kHz, 16bit）
D2：評価セット
TIMIT 896 文，男性 112 名（16kHz, 16bit）

3.1 調音特徴の抽出精度

後述するプロット精度評価の前に，プロット精度に影響を及ぼす可能性の高い調音特徴の抽出精度を算出した．学習セットにより学習済みの MLN を用いて評価セットの音声データから抽出した AF28 次元に対して，次式に示す抽出精度(AF-Correct Rate; AFCR)を計算した．

AFCR = (正しく抽出できたフレーム数 / フレーム数) × 100 [%]

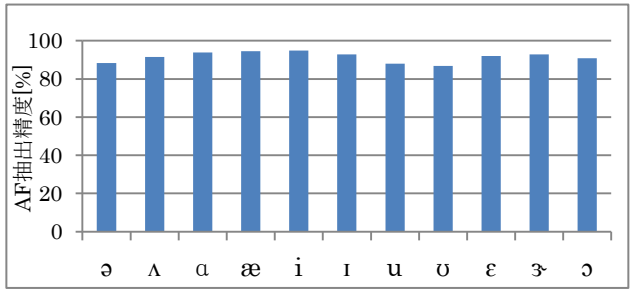

図 10 英語母音に対する調音特徴抽出精度

表 2 英語母音音素の発音マップにおける座標値

	ə	ʌ	ɑ	æ	i	ɪ	u	ʊ	ɛ	ɝ	ɔ
X	50	100	100	0	0	25	100	80	0	50	100
Y	50	66	100	82	0	25	0	20	66	66	66

図 10 に音素毎の調音特徴の抽出精度を示す．全ての音素に対して 85%を超える抽出率が得られた．ただし，/ə/の抽出率が比較的低い理由として，英語の/ə/の調音位置が発話のスタイルや前後の音などの条件によって，大きく変動することが考えられる．また，/u/，/ʊ/については，学習データ中の/ʊ/の出現数が他の母音に比べ非常に少ないため MLN の学習が不十分であり，調音特徴が比較的近い/u/の学習に悪影響を及ぼした可能性がある．

3.1 実験結果

評価の基準となる英語母音の正解座標を表 2 に示す．正解座標は MLN に与えた教師データを座標変換器に通して得た値である．表 2 に従い各母音を発音マップ上に配置したものを図 11 に示す．プロットの正確さを評価する尺度には，発話から得られた座標と正解座標との一致率を用いた．一致率は発話から抽出された座標と正解座標との距離に反比例し，話者の発音の座標と正解座標の距離が 0 の場合は 100%となる．

各英語母音に対するプロット一致率を図 12 に示す．図 12 では，全ての音素において 70%以上の一致率が得られた．特に/ə/は予備実験において調音特徴の抽出率が低かったにも関わらず高い一致率が得られた．これは 3.1 で述べた調音位置の変動が図 8，図 9 の変換アルゴリズムによって吸収されたためであると考えられる．

さらに本実験では話者によるプロットのばらつきを確認するために，TIMIT の発話データに付与されている各話者の方言情報をもとに話者を 8 つのグループに分け，それぞれのグループに対してプロットを行った．話者グループ毎の平均座標を台形のマップに適応させる変換処理を行った後にプロットした結果を図 13 に示す．図中の破線で示す領域が IPA の母音図を模した領域である．幾つかの母音は正解座標から離れた位置にプロットされているが，どの音素も話者グループ間のばらつきが小さいことが分かる．

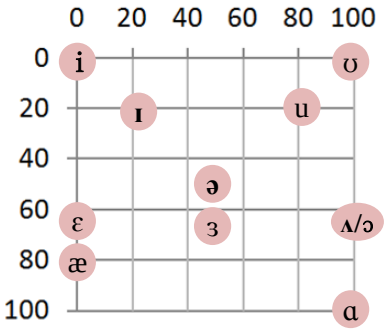

図 11 英語母音の正しい調音位置の座標

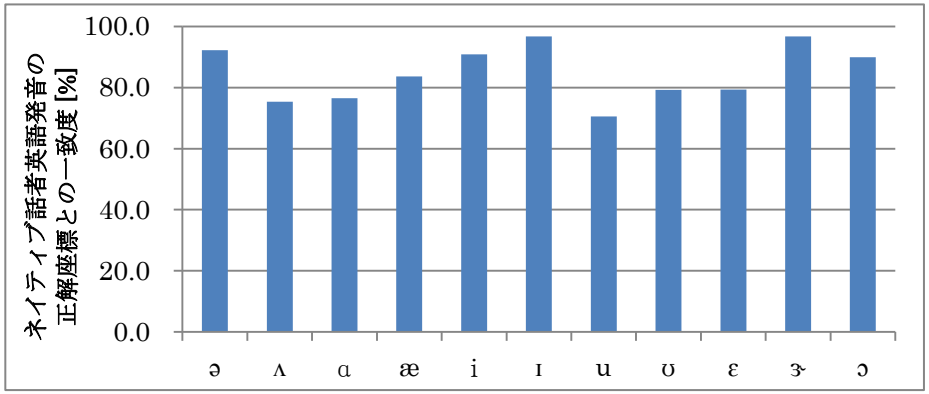

図 12 ネイティブ話者英語発話によるプロットと正解座標との一致度

4. まとめ

英語発音訓練ソフトにおける調音動作の教示および矯正機能として，調音特徴に基づき学習者の調音動作を IPA 母音図上にリアルタイムにプロットする英語母音発音マップを開発し，評価実験によりそのプロット精度を評価した．英語母語話者音声を用いた評価実験の結果，全ての音素において 70%以上の一致率が得られた．さらに，発話スタイルの異なる話者グループ間でもプロットのばらつきが小さいことが確認できた．今後，発音マップのプロット精度をより高めるべく，MLN の調整や座標変換アルゴリズムの改良を検討したい．なお，今後子音の調音動作を教示可能な発音マップも開発する予定である．子音の発音については，調音様式の違いも重要であるため調音位置と調音様式を表示することのできる仕組みが必要となる．

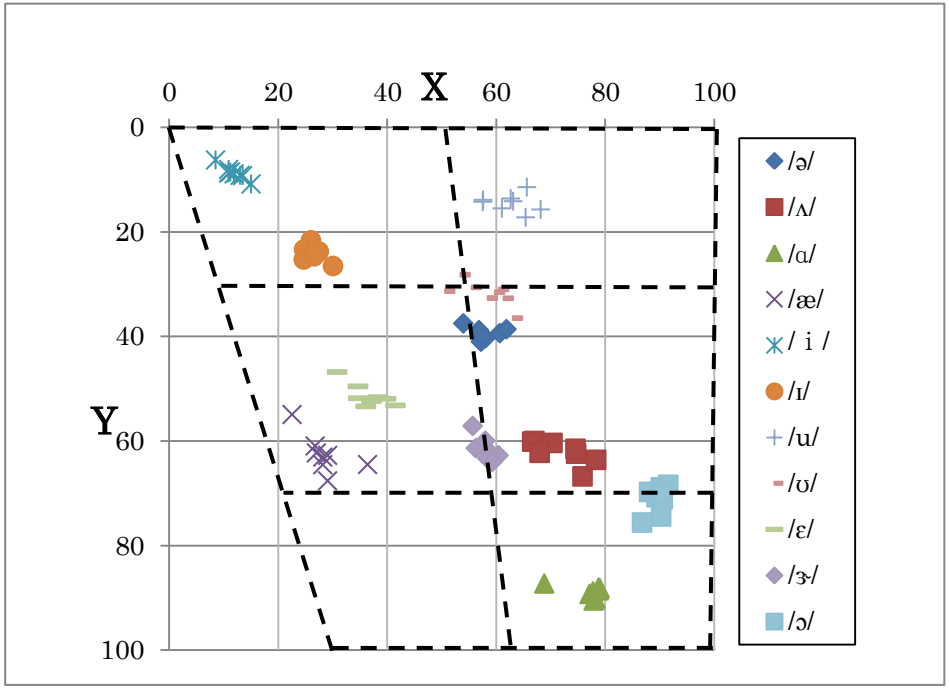

図 13 英語母音に対する話者グループ毎のプロット例

参考文献

1 EnglishEntral - 株式会社 EnglishCentral
http://www.englishcentral.com
2 Sonic Print - 株式会社アルカディア
http://www.arcadia.co.jp/SP/index.html
3 菊地歌子，島崎のぞみ，境一三： 日本人フランス語学習者のための発音学習教材，電子情報通信学会技術研究報告 SP, Vol.110(452), pp.25-29(2010)
4 佐伯拓郎，中貴俊，ヤーッコラ伊勢井敏子他： 3D フォルマント母音図における発声母音のリアルタイム可視化，電子情報通信学会総合大会講演論文集 2009 年_情報・システム(1), pp.169, 2009
5 IPA vowel chart: http://www.arts.gla.ac.uk/ipa/vowels.html
6 Garofolo , J.S. et al.: TIMIT Acoustic Phonetic Continuous Speech Corpus, Linguistic Data Consortium (1993)