

調音特徴抽出に基づく IPA 母音チャートへの 英語発音リアルタイム表示

森拓郎[†] 入部百合絵[†] 桂田浩一[†] 新田恒雄[†]

豊橋技術科学大学 大学院工学研究科[†]

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。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 教材が開発されている[1]が、舌の位置や口の高低といった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ない[2]、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く、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。一方、英会話教室や教育機関では教師が調音動作を学習者にうまく伝えながら、正確な発音ができるように指導している。これら発音教育の場においては、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 母音チャートと呼ぶ）が用いられている[3][4]。そこで、学習者の発音も IPA 母音チャート上にマップし、正しい母音の調音位置と比較することで、調音の違いを容易に理解することができる。加えて、これらをリアルタイムに表示することで、学習者がチャート図より自分の口唇や舌の位置を確認しながら正しい位置へと近づけていく発音矯正を可能にする。そこで、本報告では調音特徴に基づき、学習者の調音ジェスチャを IPA 母音チャートへリアルタイムに表示する英語母音訓練システムについて詳述する。

2. 調音特徴に基づく発音訓練システム

我々は以前より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を開発している。本ソフトウェアは音声認識により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する機能を有している。本報告では、上記ソフトウェアにおいて学習者の音素誤りを指摘後にその発音を正しい発音へ導くため、発音の基本である調音動作を効果的に矯正するシステムを提案する。そのため、調音特徴を用いて IPA 母音チャート上に表示を行う機能（以下、英語母音発音マップと呼ぶ）を実装した。以下に調音特徴の抽出方法と IPA 母音チャートへの表示について述べる。

2.1. 調音特徴抽出部

調音特徴（Articulatory Feature; AF）は、単音分類に用いられる調音様式（発声に要する舌、口唇、口蓋などの調音器官の動作・様式、例：摩擦音、破裂音など）と調音位置（舌の位置や口唇の高低）の諸属性を指す特徴量である。従って、調音特徴が得られれば学習者の調音位置を推定することができ、IPA 母音チャートへのプロットが可能となる。今回は IPA から英語に関する部分を抽出して調音特徴を定義した。調音特徴は、音声スペクトル系列の時間微分と周波数微分から求めた局所特徴（Local Feature; LF）を多層ニューラルネット（Multi-Layer Neural network; MLN）に入力して得る。MLN はラベル付き音声データを用いて、時刻 $t-3$, t , $t+3$ の 3 フレームを入力し、調音特徴を教師信号として学習する。

2.2. IPA 母音チャート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 1 に示す。発音マップは IPA 母音チャート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（縦軸）、舌の盛り上がる位置を示すスケール（横軸）、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位置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、調音を徐々に修正しながら漸的に調音位置を矯正することができる。

図 2 に英語発音マップのシステム全体図を示す。システムは学習者の発声を検知すると調音特徴抽出

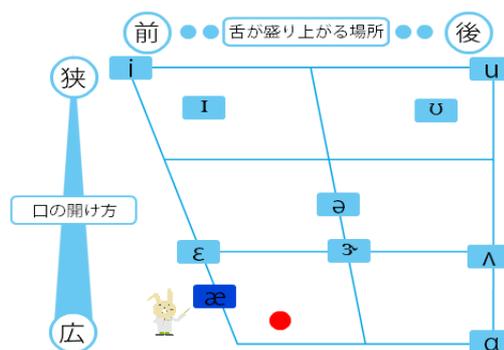


図 1 母音発音マップの画面例

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

[†]Graduate School of Engineering, Toyohashi Univ. of Tech

器により 10ms 毎に 48 次元の調音特徴を抽出する。抽出された調音特徴の母音に関する特徴列が座標変換器により 2 次元平面上の X, Y 座標に変換され、その座標に光点が移動する。特徴列から X 座標への変換は式 (3), Y 座標への変換は式 (6) によって行われる。

$$\alpha = \begin{cases} 1 & \text{if } AF_{\text{front}} \geq AF_{\text{back}} \\ 3 & \text{if } AF_{\text{front}} < AF_{\text{back}} \end{cases} \quad (1)$$

$$\Delta_x = \begin{cases} AF_{\text{central}} - AF_{\text{front}} & \text{if } AF_{\text{front}} \geq AF_{\text{back}} \\ AF_{\text{back}} - AF_{\text{central}} & \text{if } AF_{\text{front}} < AF_{\text{back}} \end{cases} \quad (2)$$

$$X = D_{\text{width}} * (\alpha + \Delta_x) / 4 \quad (3)$$

$AF_{\text{front}}, AF_{\text{central}}, AF_{\text{back}}$ はそれぞれ調音特徴列の口唇の開き具合を示す「前舌」「中舌」「後舌」に対する特徴量を指している。 D_{width} は発音マップの水平方向の長さである。

$$\beta = \begin{cases} 1 & \text{if } AF_{\text{close}} = \max \\ 1 & \text{if } AF_{\text{close}} \geq AF_{\text{open_mid}} & \text{if } AF_{\text{close_mid}} = \max \\ 3 & \text{if } AF_{\text{close}} < AF_{\text{open_mid}} & \text{if } AF_{\text{close_mid}} = \max \\ 3 & \text{if } AF_{\text{close_mid}} \geq AF_{\text{open}} & \text{if } AF_{\text{close_mid}} = \max \\ 5 & \text{if } AF_{\text{close_mid}} < AF_{\text{open}} & \text{if } AF_{\text{close_mid}} = \max \\ 5 & \text{if } AF_{\text{open}} = \max \end{cases} \quad (4)$$

$$\Delta_y = \begin{cases} AF_{\text{close_mid}} - AF_{\text{close}} & \text{if } AF_{\text{close}} = \max \\ AF_{\text{close_mid}} - AF_{\text{close}} & \text{if } AF_{\text{close}} \geq AF_{\text{open_mid}} \\ AF_{\text{open_mid}} - AF_{\text{close_mid}} & \text{if } AF_{\text{close}} < AF_{\text{open_mid}} \\ & \text{if } AF_{\text{close_mid}} = \max \\ AF_{\text{open_mid}} - AF_{\text{close_mid}} & \text{if } AF_{\text{close_mid}} \geq AF_{\text{open}} \\ AF_{\text{open}} - AF_{\text{open_mid}} & \text{if } AF_{\text{close_mid}} < AF_{\text{open}} \\ & \text{if } AF_{\text{open_mid}} = \max \\ AF_{\text{open}} - AF_{\text{open_mid}} & \text{if } AF_{\text{open}} = \max \end{cases} \quad (5)$$

$$Y = D_{\text{Height}} * (\beta + \Delta_y) / 6 \quad (6)$$

$AF_{\text{close}}, AF_{\text{close_mid}}, AF_{\text{open_mid}}, AF_{\text{open}}$ はそれぞれ調音特徴列中の舌の盛り上がる位置を示す「狭」「半狭」「半広」「広」に対する特徴量を指している。 D_{height} は発音マップの垂直方向の長さである。発音マップは台形状であるため座標をプロットする際は、適宜台形のスケールに合わせて座標を変換し、マップ上へプロットする。

3. 評価実験

提案する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 母音チャート上にプロットし、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。そこで、英語発音マップの性能を評価するため、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た。ここで、正解座標とは、国際音声学会が定義した IPA 母音チャート上の各母音の座標値である。MLN の学習には、TIMIT [5] の

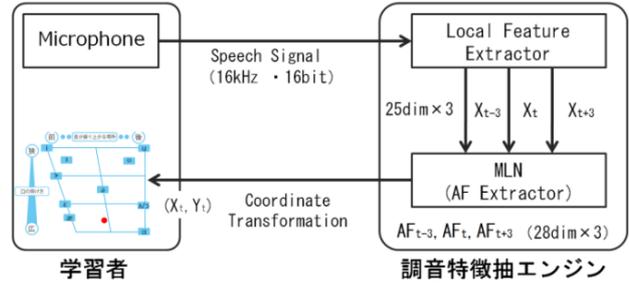


図 2 発音マップシステムの基本構成

2,600 文 (男性話者 325 名) を使用した。テストデータは TIMIT の 896 文 (男性話者 112 名) である。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。

実験結果を図 3 に示す。全ての英語母音において、本システムが音声から算出したプロット座標と正解座標の一致率は 70% 以上であり、全母音の平均一致率は 85.5% であった。今回の実験から、英語母語話者の母音発音に対して高い精度で発音マップ上にプロットできていることが確認された。

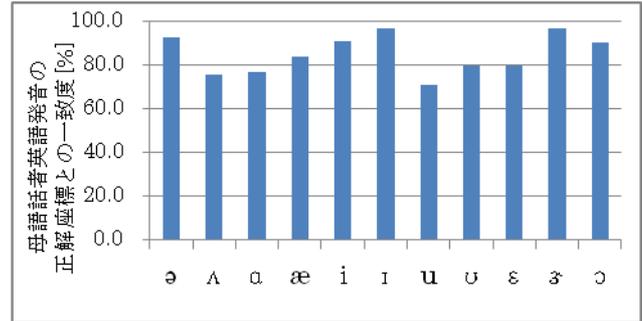


図 3 発音マップのプロット精度

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、評価実験により全ての母音に対して 70% 以上のプロット精度を得られることが確認できた。今後は更なるプロット精度の向上を図るとともに、子音の調音動作を矯正可能な発音マップの開発を進めていく。

参考文献

- [1] Sonic Print - 株式会社アルカディア <http://www.arcadia.co.jp/SP/index.html>
- [2] 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)
- [3] 佐伯拓郎, 中貴俊, ヤーココラ伊勢井敏子他: 3D フォルマント母音図における発音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
- [4] IPA vowel chart: <http://www.arts.gla.ac.uk/ipa/vowels.html>
- [5] Garofolo, J.S. et al.: TIMIT Acoustic Phonetic Continuous Speech Corpus, Linguistic Data Consortium (1993)