DNN 活動パターンの可視化による 発音評価フィードバックに関する基礎的検討

A study on feed-back of pronunciation evaluation by visualizing DNN activities

勝瀬 郁代
Ikuyo Masuda-Katsuse
近畿大学
Kindai University
*
Email: katsuse@fuk.kindai.ac.jp

あらまし：児童の発音評価を視覚的にフィードバックするために，DNN の活動パターンに着目した．まず予備研究として，DNN レイヤーの活動パターンを低次元多様体に布置した分布で，どの程度音素が識別できるかを調査した．はじめに多層パーセプトロンを構成し，音声特徴量を学習データ，音素ラベルを教師データとして学習した．そして parametric t-SNE 変換により，DNN レイヤーの活動パターンを 2 次元多様体上にマップした。多様体は局所的にはユークリッド空間とみなせるので、マージン最大化近傍法により識別テストを行った結果, 構音位置の平均正答率は 82.6%, 構音様式の平均正答率は 82.3%であった．
キーワード：発音評価，フィードバック，DNN，可視化，音素識別

1. はじめに

私たちはこれまで，言語通級指導教室の発音指導を支援するシステムを開発してきた．子どもたちが発音練習をするためのシステム(1)では，練習時の発音を，ターゲットである発音やその児童が誤りがちな発音と比較した結果をフィードバックしていたが，発音評価の精度があまりよくなく，改善を望まれていた．また，発音の正誤判定のみのフィードバックは歓迎されず，“どの程度正しいか”のフィードバックを望まれていた．むろん，構音上の問題点も指摘できればなおよい．

発音評価に関しては，Goodness of Pronunciation(2)に基づく方法など，これまで数多くの先行研究があるが，近年，DNN による音韻特徴表現に関する研究が報告されている．Nagamine らは，DNN の各ノードに見られる，音素の弁別素性の特徴を調べている(3)．また Sim は，DNN 中間層の活動パターンを低次元多様体に布置し，音素の分布を観察している(4)．

そこで本研究は，児童の発音評価を視覚的にフィードバックするために DNN の活動パターンを利用することを目的とする．本講では，そのための予備的研究として, DNN レイヤーの活動パターンを低次元多様体に布置したマップ上で，どの程度音素が識別できるかを調査した結果を報告する．

2. 音声特徴量

本研究では,「日本語話し言葉コーパス」(5)の音声から Kaldi ツールキット(6)を使って特徴ベクトルを抽出した．具体的には，13 次元 MFCC 特徴ベクトルを求め，±4 フレームのスプライシングを行った．LDA により 40 次元に次元圧縮し，MLLT により特徴ベクトルの相関を削減した．そして，fMLLR により話者の正規化を行った．このようにして得られた各フレームの特徴量を，学習データ（training data, validation data）と評価用のデータ(test data)に分けた．次に，HMM の遷移モデルの pdf のインデックスである 497 個の id との強制アライメントを行った．これらの id は音素と対応が取れているので, 最終的に，各フレームの属性として音素を割り当て，これを教師データとした．

3. ネットワークの構成

ネットワークは，入力層（40 次元），1 つの中間層（ノード数 256），4 つの中間層（ノード数 2048），出力層（25 次元）の多層パーセプトロンである．活性化関数として，出力層に対してのみ softmax 関数を用い，それ以外の層には ReLU 関数を用いた．このネットワークを用いて音素の識別テストを行ったところ，正答率は 85.0%であった．

学習データ(training, validation)，テストデータそれぞれについて, DNN の最終層の活動パターンを得た．

4. 低次元多様体への布置

3 章で得た活動パターンを 2 次元多様体に布置するために，parametric t-SNE(7)を用いた．parametric t-SNE は，training data のマッピング上に test data をマップすることができるため，評価したい音声と比較したい音声の関係を視覚的に確認できるのではないかと考えた．

parametric t-SNE では, 変換のためのモデルを DNN で学習する．本研究では，この DNN の中間層を 3 層（ノード数はそれぞれ 100, 100, 400）に設定した．学習とテスト用のデータとして, 3 章の DNN の学習と評価に用いたデータから，各音素につき 10,000 の

training data，2,000 の validation data，3,000 の test data をランダムに選んで使用した．

図 1，図 2 は，発音誤りがよく見られる音素対について学習データを布置したものである．図 1 は，構音位置は同じ歯茎硬口蓋で，構音様式が摩擦である/ɕ/（青）と破擦である/tɕ/（赤）の分布である．図 2 は，軟口蓋破裂音の/k, kʲ/（青）と硬口蓋摩擦音の/ ç/（赤）の分布である．どちらの図も，各音素独自の領域と，重複領域が確認される．分布が重複している領域に布置される音声は，聴感上もどちらの音素にも似た音が分布しているかどうかを確認する必要がある．実際に聴感上も区別しにくい音が分布しているのならば，ターゲットである音素の音声と誤りがちな音声を布置したマップ上に，評価したい音声をマップすることで，ターゲット発音や誤りがちな発音との関係を視覚的に知ることができる．

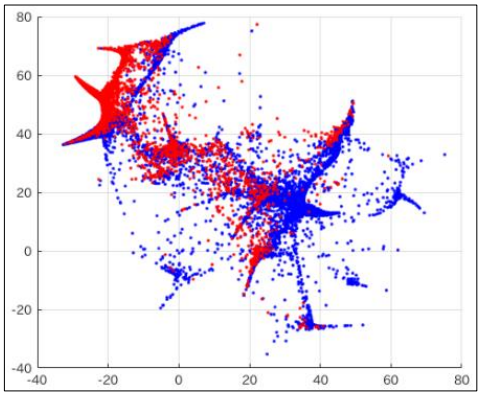

図 1　/ɕ/（青）と/tɕ/（赤）の分布

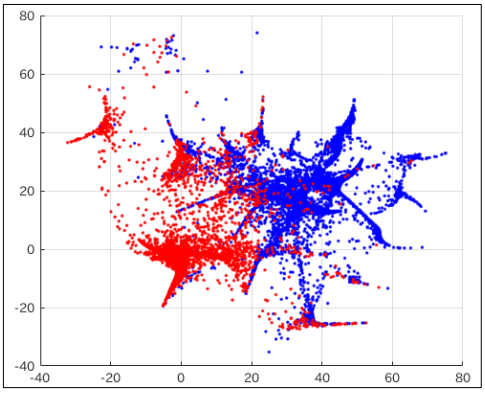

図 2　/k, kʲ /（青）と/ç/（赤）の分布

5.　識別テスト

多様体上の各点の近傍はユークリッド空間であるため，マージン最大化近傍法[8]により，音素認識テストを行った（k=5）．その結果，子音の認識率は 77.6%，母音の認識率は 80.0%であった．

先行研究[1]における発音評価のフィードバックでは, 起こりがちな発音誤りとの比較を提示するので，音素の認識よりも，構音の位置や様式の誤りの検出性能がより重要である．表 1 に，構音位置及び構音様式で音素を分別した場合の正答率を示す．構音位置の平均正答率は 82.6%，構音様式の平均正答率は 82.3%であった．

表 1　構音位置及び構音様式の正答率

構音位置の正答率		構音様式の正答率	
両唇音	0.830	鼻音	0.873
歯茎音	0.871	破裂音	0.848
歯茎硬口蓋音	0.820	摩擦音	0.843
硬口蓋音	0.792	破擦音	0.791
軟口蓋音	0.788	弾き音	0.813
声門音	0.747	接近音	0.770
平均	0.826	平均	0.823

6.　まとめ

本講では，音声特徴量を入力とし，音素ラベルを教師データとして学習した DNN のレイヤーの活動パターンを多様体上に布置し，そのマップ上に評価したい発音を布置することで，発音評価のフィードバックを行うことを目的とし，実際にどの程度音素を識別できるかを調査した．各音素は，一部領域が重複しつつも，異なる領域に分布した．今後，重複領域に布置された音声が，実際に聴感上も似ているかどうかを確認する必要がある．

7.　謝辞

本研究は，JSPS 科研費(16K00496)及び平成 29 年度産業理工学部プロジェクトによる助成を受けた．

参考文献

(1) 勝瀬郁代: “言語通級指導教室における発音指導を支援するシステム”，教育システム情報学会誌，第 34 巻，第 1 号, pp.7-19 （2017）
(2) Witt, S. and Young, S.: “Computer-assisted pronunciation teaching based on automatic speech recognition,” Language Teaching and Language Technology Groningen, The Netherlands (1997)
(3) Nagamine, T. et al.: “Exploring How Deep Neural Networks Form Phonemic Categories,” in INTERSPEECH 2015 (2015)
(4) Sim, K.C.: “On Constructing and Analysis an Interpretable Brain Model for the DNN based on Hidden Activity Patterns,” in ASRU (2015)
(5) 「日本語話し言葉コーパス」: http://pj.ninjal.ac.jp/corpus_center/csj/
(6) Povey, D. et al.: “The Kaldi Speech Recognition Toolkit,” IEEE 2011 Workshop on Automatic Speech Recognition and Understanding (2011)
(7) van der Maaten, L.: “Learning a Parametric Embedding by Preserving Local Structure”, PMLR 5:384-391 (2009)
(8) Weinberger, K.Q. and Saul, L.K.: “Distance Metric Learning for Large Margin Nearest Neighbor Classification,” J. Machine Learning Research 10 (2009)