構造的表象を用いた話者間の発音距離行列の可視化に関する検討*

☆黒瀧 夏子, 鈴木 雅之, 峯松 信明, 広瀬 啓吉 (東京大学)

1 はじめに

近年、国際語としての英語を用いたコミュニケーションの重要性が高まっている。グローバル化が進む現在、世界の共通言語としての英語の需要はますます高まっている。実際に国際的に取引をしている日系企業が英語を社内公用語とした例などが挙げられる。このように、世界的に英語学習者が増加し、英語を母語としない英語話者が増加している。

第二言語習得において発音学習は非常に重要な位置を占めている [1][2]。これは外国語を習得する場合、母語干渉による訛りが頻繁に観測されるためである [3]。例えば学習者が日本語を母語とするとき、英語の発音に日本語の発音が干渉して英語母語話者とは違う発音になるという現象が起きる [4]。具体的には、/r/と/l/の発音の区別が困難な事例がよく見受けられる。このような母語干渉は時として他の英語話者とのコミュニケーションに困難をもたらすことがある [5]。このように、発音の母語干渉によって他の英語話者とのコミュニケーションがとれないことがある。逆にいうと、似た母語干渉を持つ話者同士は、コミュニケーションがとりやすいという事例も発生する [6]。日本での英語教育は近年、文法中心の教育法からコミュニケーション重視の教育法へ変化している。このような場合、似ている発音の学習者を探し自分に聞き取りやすい学習者を見つけることは、相手に話す・聞くというモチベーションを向上させる上で、有効であると考えられる。本研究では学習者が容易に他者との発音比較を行えるシステムの構築を目的として、発音比較の可視化手法について検討し、提案手法の有効性を示す。

2 発音の視覚化に関する先行研究

2.1 話者特性と発音特性の分離

発音された音声には様々な情報が含まれている。発音分析において重要なのは発音やイントネーションなどの言語的情報である。また音声には同時に学習者の声道長特性や録音機器の周波数特性などの発音とは無関係な非言語的情報も必ず含まれてしまっている。音声分析によく使用されるケプストラムには録音時のマイク特性、話者の声道長などの非言語的情報が内包されている。そのため、ケプストラムを直接参照して発音評価を行う場合、非言語的情報のミスマッチが発音評価に影響を及ぼす可能性がある。すなわち、この非言語的要因による音声の変動を何らかの方法で取り除いた上で教師と学習者間の発音比較をする必要がある。

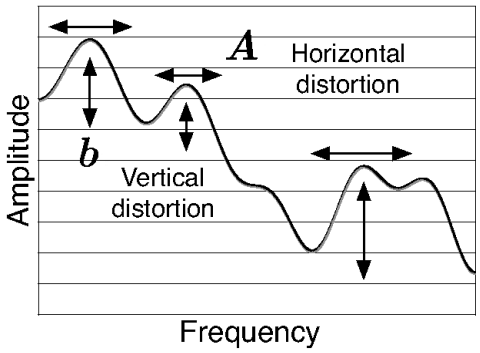


Fig. 1 行列 $\boldsymbol{A}$ とベクトル $\boldsymbol{b}$ によるスペクトル変形

さて、非言語的要因によるひずみにはいくつかの種類があり、主に加算性雑音、乗算性ひずみ、線形変換性ひずみの 3 種類に分類される [7]。加算性雑音はスペクトラム空間に対する加算として表され、背景雑音などが相当する。乗算性ひずみはスペクトルに対する乗算で表現されるひずみで、ケプストラム空間ではケプストラムベクトル $\boldsymbol{c}$ に対するベクトル $\boldsymbol{b}$ の加算 $\boldsymbol{c}' = \boldsymbol{c} + \boldsymbol{b}$ で与えられる。これは音響機器や伝送回路の周波数特性に相当する。音声を分析する際は必ず何らかの音響機器を通して音声を電気信号としなければならないため、不可避なひずみといえる。線形変換性ひずみは $\boldsymbol{c}$ に対する行列 $\boldsymbol{A}$ の乗算 $\boldsymbol{c}' = \boldsymbol{A}\boldsymbol{c}$ で表される。声道長の差異がこれに相当する。 声道長は話者の性別や年齢によって異なり、声道長の差異で話者性の違いが発生する。話者性の違いは非言語的特徴の中でも最も不可避なものである。以上により、加算性雑音を除く非言語的要因による音声変形はケプストラム空間においてアフィン変換 $\boldsymbol{c}' = \boldsymbol{A}\boldsymbol{c} + \boldsymbol{b}$ で近似される。

Fig.1 に非言語的特徴がスペクトルに及ぼす影響を示した。例えば、乗算性ひずみによる $\boldsymbol{b}$ の加算はスペクトルを垂直の方向に変化させる。また、話者の声道形状の一部も乗算性ひずみとして現れる。 線形変換性歪みによる $\boldsymbol{A}$ の乗算はスペクトルを水平の方向に変化させる。

2.2 構造的表象

以上のような非言語的特徴を取り除き、言語的特徴のみを評価する枠組みとして、構造的表象が提唱されている [8][9][10]。まず、構造的表象の不変性について述べる。二つの空間が可逆な空間写像で結びつけられ、それぞれの空間において対応する複数の分布が存在する場合、それぞれの空間における分布間の

* An experimental study of visualization of pronunciation structures by KUROTAKI Natsuko, SUZUKI Masayuki, MINEMATSU Nobuaki, HIROSE Keikichi (The University of Tokyo)

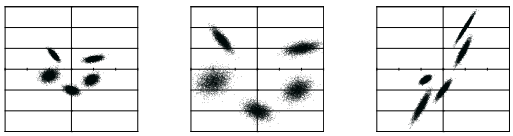


Fig. 2　音響的普遍構造

f-divergence は、常に不変となる [11]。*f*-divergence とは、分布間距離尺度の一種であり、二つの分布 p_1, p_2 間の *f*-divergence は以下の汎関数で表される。

$$f-\mathrm{div.}(p_1,p_2) = \oint p_2(x) g\left(\frac{p_1(x)}{p_2(x)}\right) dx \quad (1)$$

ただし、$g(t)$ は $t>0$ で定義する凸関数である。

ここでは *f*-divergence の一種として、ここではバタチャリヤ距離を用いる。バタチャリヤ距離は以下の式で表される。

$$BD(p_1,p_2) = -\ln \int_{-\infty}^{\infty} \sqrt{p_1(x)p_2(x)}dx \quad (2)$$
$$(p_1(x), p_2(x) : \text{確率密度関数})$$

ここで $p_1(x), p_2(x)$ はそれぞれの確率密度関数である。また、この分布がガウス分布であると仮定すると、

$$BD(p_1,p_2) = \frac{1}{8}\boldsymbol{\mu}_{12}^{\top}\left(\frac{\sum_1+\sum_2}{2}\right)^{-1}\boldsymbol{\mu}_{12} + \frac{1}{2}\ln\frac{|(\sum_1+\sum_2)/2|}{|\sum_1|^{\frac{1}{2}}|\sum_2|^{\frac{1}{2}}} \quad (3)$$

と表される。ここで、$\boldsymbol{\mu}_1(\boldsymbol{\mu}_2)$ と $\sum_1(\sum_2)$ はそれぞれ $p_1(x)(p_2(x))$ の平均ベクトルと共分散行列、$\boldsymbol{\mu}_{12}$ は $\boldsymbol{\mu}_1$ と $\boldsymbol{\mu}_2$ の差である。それぞれの分布にアフィン変換をかけても変換の前後でバタチャリヤ距離は不変となる。図 2 で表した 3 つの分布群を考える。これらの分布群はいずれもアフィン変換で結び付けられており、すべて等しいバタチャリヤ距離行列を持っている。ユークリッド平面上では異なる構造と捉えられるが、バタチャリヤ距離による距離行列は等しい。このことを用いて音声から非言語的特徴を除いた、言語的特徴のみを用いた構造を抽出できる。これらを発音分析に適用する事を考える。ケプストラム空間において同一話者の各母音間のすべての二分布間距離を求めることで、各母音間の距離を表す母音距離行列が得られる。この母音距離行列を教師と学習者で比較すれば、話者同士の母音の距離を比較できる。バタチャリヤ距離を使った場合に距離行列は対称行列になるため、その上三角成分のみを取り出し、それを発話者の発音の構造ベクトルと定義する。また、話者間の距離は、話者ごとに得られた構造ベクトル間のユークリッド距離を求めることで得られる。すべての二話者間の距離を計算することで、話者間の発音の距離行列が求められる。

2.3　多次元尺度構成法

求められた距離行列をそのまま学習者に提示しても、それを解釈するのは難しい。まずはこの結果を

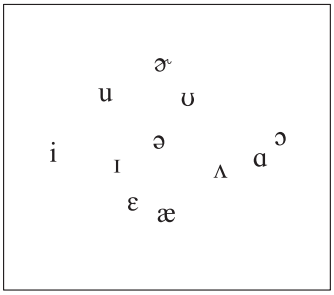


Fig. 3　母音間発音距離の MDS

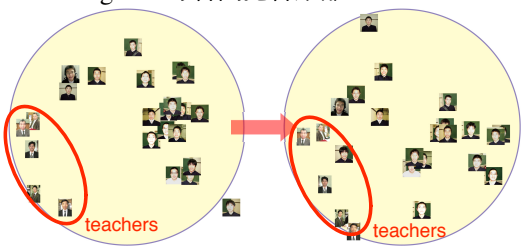


Fig. 4　話者間発音距離の MDS

人間が知覚できる 2 次元か 3 次元に変換する必要がある。多次元の距離行列の可視化手法として多次元尺度構成法 (Multi Dimensional Scaling; MDS) がある [12]。MDS は多変量解析の手法の一つであり、多次元空間を低次元に投影するとき、歪みを最小にするように投影する手法である。$m \times m$ の距離行列 $\boldsymbol{S}$ があるとき、b_{jk} を要素に持つ行列 $\boldsymbol{B}$ を求める。

$$b_{jk} = \frac{1}{2}\left(\frac{1}{m}\Sigma_j s_{jk}^2 + \frac{1}{m}\Sigma_k s_{jk}^2 - \frac{1}{m^2}\Sigma_j\Sigma_k s_{jk}^2 - s_{jk}^2\right) \quad (4)$$

このとき、s_{jk} は行列 $\boldsymbol{S}$ の各要素を表す。行列 $\boldsymbol{B}$ を求めたのち、行列 $\boldsymbol{B}$ に対する固有値問題をとき、固有値を大きい順に並べる。

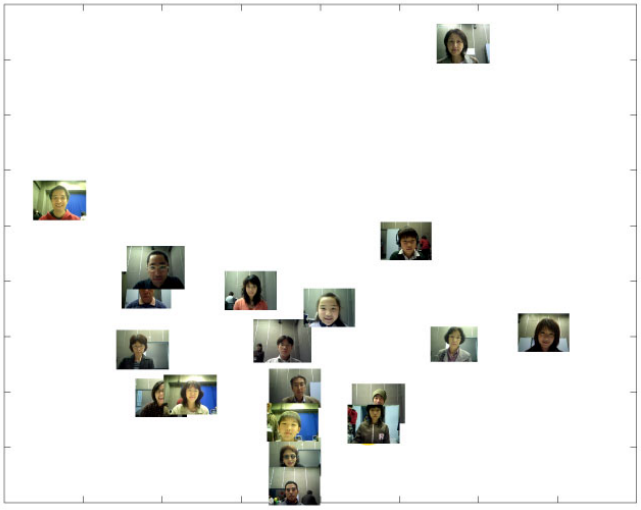
$$\boldsymbol{B}\boldsymbol{x}_t = \lambda_t \boldsymbol{x}_t (t=1,2,\cdots,m) \quad (5)$$
$$\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_m$$

得られた固有値 λ_t と固有ベクトル $\boldsymbol{x}_t$ より座標行列が求まる。

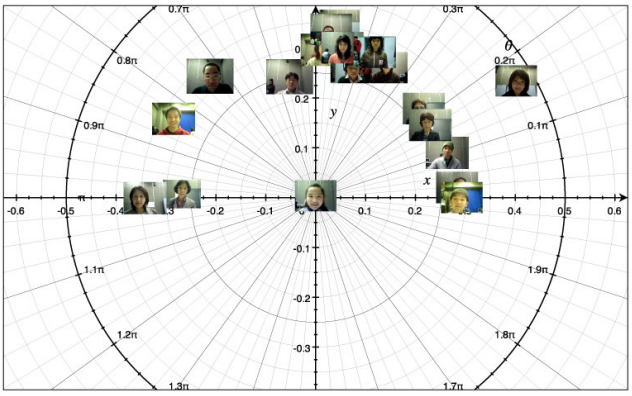
$$a_t = \sqrt{\lambda_t}\boldsymbol{x}_t (t=1,2,\cdots,r) \quad (6)$$
$$\boldsymbol{A} = (a_1, a_2, \cdots, a_r)$$

以上で r 次元空間における座標行列 $\boldsymbol{A}$ が求まる。2 次元空間への変換を行う場合 r は 2 となる。

母音間距離行列を MDS を用いて視覚化すれば、Fig.2.3 のように母音の分布の様子が 2 次元平面上に視覚化できる [8]。同様に、N 人の話者に対して話者間の発音距離行列を MDS を用いて視覚化すれば、Fig.3 のように N 人の話者の分布の様子が 2 次元平面上に視覚化される [13][14]。



(a) MDS



(b) 提案手法

Fig. 5　可視化結果

3　発音距離行列の視覚化に関する提案手法

英語学習システムを構築するにあたり、システムの利用者が直感的に容易に他者と発音比較が行えることは重要である。MDS を用いると、複数話者間の発音距離を表すことが可能である。しかしこの手法では距離行列の歪みを最小に投影するため、全体の位置関係はわかるが、一人の利用者に注目した際の各話者との距離は必ずしも適切ではない。そこで N 人の話者に対する発音間距離行列が与えられた場合に特定の話者一人（利用者）に着目し、その利用者と各話者との発音距離をわかりやすく可視化できないか検討した。また、可視化するときに発音の違いのみではなく、声色の個人性をもりこむと話者間の発音の違いををおもしろいのではないかと考えた。発音と声色の二種類のパラメタをプロットする時に、利用者を中心に据えて極座標表示をすることを考えた。よって半径に話者間の発音距離、角度に話者の個人性の違いを用いてプロットする。利用者と各話者との発音距離を構造ベクトル間のユークリッド距離により求める。利用者に直感的にわかりやすい提示の方法として、本手法では極座標表示をベースにして半径に話者間のユークリッド距離を取る。そして角度に話者の非言語的特徴を含めた特徴量を取り、極座標で表す。発音の類似度を利用者に提示することによって、学習者の学習意欲の向上を狙う。

Table 1　実験に用いた英単語

Words	bot	bat	but	bird	bit	beat
Vowels	ɑ	æ	ʌ	ɚ	ɪ	i
Words	boot	bet	bought	good		
Vowels	u	ɛ	ɔ	ʊ		

4　実験

19 名の話者に Table.1 の 10 単語を発音させ、強制アライメントで母音部分のみを切り取った母音セットを使用した。また、話者の顔写真も収録されている。母音セットからメルケプストラムを求め、各話者の

Table 2　構造を抽出するための音響分析条件

サンプリング	16bit / 16kHz
窓	窓長 25ms, シフト長 4ms
パラメータ	MFCC(12 次元)
音素間距離	対応するガウス分布間の $\sqrt{BD}$

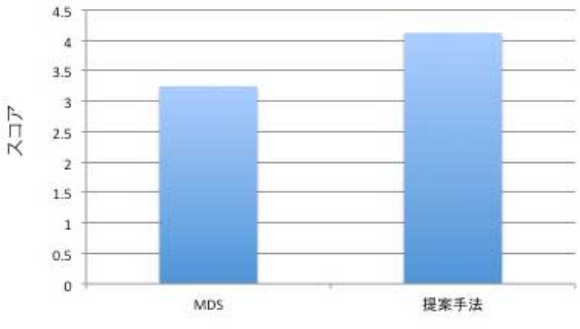


Fig. 6　実験結果

母音間距離行列を得た。得られた母音間距離行列から各話者の構造ベクトルを計算し、話者間の 19×19 の距離行列を求めた。提案手法の半径部分として得られた距離行列の利用者と各話者間との距離の行のみ取り出した。また利用者と各話者間で各母音のバタチャリヤ距離を求め、平均をとり利用者と各話者間の距離として提案手法の角度部分として利用した。こうして得られた学習者の距離行列を、従来手法である MDS 及び提案手法を用いて可視化した。そしてプロットしたグラフに、各点に対応した各話者の顔写真を使用する。結果を Fig.5(a) と Fig.5(b) に示す。

以上の方法で作成した視覚化結果について、評価実験を行った。評価実験に際して、12 名に Fig.5(a) と Fig.5(b) を見比べ、直感的にわかりやすさを 5 段階評価で点数をつけてもらった。

5 段階評価の平均点は Fig.6 のようになった。結果より、提案手法のほうがよりわかりやすいという結果になった。また、英語教育関係者からは会話相手を選ぶときに声色の好みがある人もいるので、視覚化のときにそのような情報があるのは便利だという意見もあった。直感的にわかりやすい情報提示を行うことによって、利用者の学習意欲向上が見込まれる。そのため、提案手法の方が MDS よりも学習システムへの応用に適していると言える。

5　まとめ

本論文では、英語学習システムのための発音評価の可視化手法の検討を行った。従来手法では言語的特徴のみを使って可視化を行なっていたのに対し、特定の利用者に関係する部分および非言語的特徴（声色の違い）を用いた可視化手法を提案した。どちらの手法が英語学習システムへの応用に適しているか評価するために比較実験を行った。その結果、提案手法の方が利用者にわかりやすい情報提示が行えることがわかった。ただし、どちらの方法が適しているかは状況に依存する。今後の課題としては極座標の角度の部分で表した声色の距離の計算方法が、発音の良し悪しと個人性の違い両方を含めた計算方法になっているので、個人性の違いをより表現したいのであれば使用する差の小さい母音または母語が一緒の場合は母語の母音を利用することが挙げられる。また、この視覚化が役立つシチュエーションとしては世界各地の英語学習者からの出身者が一同に介して英語を用いてコミュニケーションを図る場が考えられる。このような場で提案手法の可視化を示すことで声色の違いを用いた英語発音が似ている人を探しだし、お互いを知らない人同士の会話の手助けにするというものが考えられる。また、この学習システムをソーシャルネットワークサービス (Social Network Service; SNS) 上で動かせば、発音が似ている人や発音の目標とする人を見つけ、そのままテレビ電話などで会話することも可能になる。

参考文献

[1] 壇辻正剛, “共通教育における ICT 支援の外国語教育と発音指導”, 電子情報通信学会技術研究報告, SP2010-115, pp. 1–6, 2011

[2] J. Bernstein, “Objective measurement of intelligibility,” Proc. ICPhS, pp.1581–1584, 2003.

[3] 白井恭弘, “外国語学習の科学”, 岩波書店, 2008

[4] 竹蓋幸生, “日本人英語の科学”, 研究社, 1982

[5] 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, pp.1481–1484, 2011

[6] M.Pinet, P.Iverson and M.Huckvale, “Second-language experience and speech-in-noise recognition: the role of L2 experiencein the talker-listener accent interaction,” Proc. Workshop on Second Language Studies, CD-ROM, 2010

[7] 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,Vol.52,No.12,pp.3671–3681,2011

[8] 朝川智, 峯松信明, 広瀬啓吉, “音声の構造的表象に基づく英語学習者発音の音響的分析”, 電子情報通信学会論文誌, Vol. J90-D, No. 5, pp. 1249–1262, 2007

[9] 峯松信明, “音声の音響的普遍構造の歪みに着眼した外国語発音の自動評定”, 電子情報通信学会技術研究報告, SP2003-180, pp.31–36, 2004

[10] 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, Vol. 52, No. 5, pp.1899–1909, 2011

[11] Y. Qiao, N. Minematsu, “A study on invariance of f-divergence and its application to speech recognition,” IEEE Trans. on Signal Processing, Vol.58, No.7, pp.3884–3890, 2010

[12] 齋藤尭幸, “多次元尺度構成法”, 朝倉書店, 1980

[13] 高澤真章, 鎌田圭, 竹内京子, 朝川智, 峯松信明, 牧野武彦, 広瀬啓吉, “大規模英語学習者を対象とした音声の構造的表象に基づく発音評価とその応用”, 日本音響学会春季講演論文集,Vol.3,No.10–12,pp.489–492,2008

[14] 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, CD-ROM, 2008