社団法人　電子情報通信学会
THE INSTITUTE OF ELECTRONICS,
INFORMATION AND COMMUNICATION ENGINEERS

信学技報
TECHNICAL REPORT OF IEICE.

訛り・性別・年齢を考慮した自己視点からの世界諸英語発音の可視化

川瀬　佑司†　峯松　信明†　齋藤　大輔†　広瀬　啓吉†　沈　涵平††

† 東京大学 〒113–8656 東京都文京区本郷 7–3–1
†† 国立成功大学 〒70101 台湾台南市大学路 1 号
E-mail: {kawase,mine,dsk_saito,hirose,happy}@gavo.t.u-tokyo.ac.jp

あらまし　国際共通語である英語は、それぞれの国や地域の母語干渉により、多様な発音、所謂訛りが存在することが知られている。筆者らの先行研究では、様々な英語発音に対して話者を単位とした自動分類を検討している。ボトムアップ的な分類を行う場合、一般的には対象とする要素群に対して要素間距離行列が必要となる。先行研究では任意二話者間の発音距離の自動推定を行っている。さて距離行列を可視化する場合、多次元尺度法 (Multi-Dimensional Scaling, MDS) や樹形図を用いることが多い。本研究ではこれらに代わる、発音距離行列に対する新しい可視化手法を提案する。従来の可視化は、距離行列全体を表現することが狙いである。しかし可視化結果を渡される特定の学習者にとってみれば、知りたい主情報は自分とそれ以外の話者の関係性である。そこで本研究では、特定話者とそれ以外の話者の発音距離に着目し、さらには年齢や性別といった情報も含め、英語発音の自己視点からの可視化を提案する。提案手法では従来手法と異なり、可視化結果に歪みが全く生じないことが保証されている。
キーワード　世界諸英語、発音分類、可視化、歪み、自己視点、年齢差異、性別差異

Visualization of World Englishes pronunciations from a speaker's self-centered viewpoint using attributes of accent, gender, and age

Kawase YUJI†, Nobuaki MINEMATSU†, Daisuke SAITO†, Keikichi HIROSE†, and HanPing SHEN††

† The University of Tokyo, 7–3–1, Hongo, Bunkyo-ku, Tokyo, 113–8656 Japan
†† National Cheng Kung University, 1 University Road, Tainan City, 70101 Taiwan
E-mail: {kawase,mine,dsk_saito,hirose,happy}@gavo.t.u-tokyo.ac.jp

Abstract　English is the only language available for global communication and is known to have a large diversity of pronunciations due to the influence of speakers' mother tongue, called accents. Our previous studies made an attempt to do speaker-basis clustering of those pronunciations. The bottom-up clustering procedure requires a distance matrix only in terms of pronunciation differences among speakers and previous studies proposed a method to predict the pronunciation distance between any pair of the speakers. A distance matrix is often visualized on a two-dimensional plane by using the Multi-Dimensional Scaling (MDS) or drawing a dendrogram. In this study, a new method is proposed for visualization. When a visualization result is fed back to a learner, his main interest will be in the relations from himself to the others. Then, by using only a part of the distance matrix and other kinds of information such as age and gender, the proposed method can visualize multiple kinds of diversity found in English pronunciation from a speaker's self-centered viewpoint. Unlike the conventional methods, our proposal is guaranteed to cause no distortion or stress at all in results of visualization.
Key words　World Englishes, pronunciation clustering, visualization, stress, self-centered viewpoint, age difference, gender difference

1. はじめに

英語は世界共通語として、世界中の国や地域で話されている。通常学校教育における英語授業では米語（特に General American, GA）や英語（特に Received Pronunciation, RP）をモデル発音として採択することが多いが、国際的な環境で GA 話者や RA 話者と会話する機会はあまりない。通常は母語が異なる者同士が英語で意志疎通を図り、母語話者も地方訛りの英語を話してくることが多い。このような状況を鑑み、国際語（英語）には標準発音などなく、GA や RP も各々一つの Accented English として捉え、世界中には多様な英語が存在すると考える教育者も多い（World Englishes、世界諸英語 [1], [2]）。

このような立場に立てば、英語を使った他者とのコミュニケーション能力向上を図る場合、世界にはどのような多様な英語発音が存在し、自らはその中のどこに位置しているのかを把握することは重要であると考えられる。筆者らの先行研究では [3]〜[5]、世界中の英語利用者による特定パラグラフ読み上げ音声コーパス（Speech Accent Archive, SAA [6]）を用いて、個人を単位とした世界諸英語発音の自動分類が検討されている。これは、任意の二話者間の発音距離（発音の違いのみに着眼した発話距離）を入力音声のみから推定するタスクである。

学習者に対して分類結果（距離行列）を可視化して呈示する必要があるが、常套手段である多次元尺度法（Multi-Dimensional Scaling, MDS）（例えば図 1）や樹形図（例えば図 2）を用いた検討が行なわれている [7], [8]。しかし、これらの手法による二次元空間上の可視化は、距離行列全体を表現することが狙いである。一方、可視化結果を提示される特定の学習者（話者 n）にとってみれば、知りたい主情報は自分とそれ以外の話者との関係性であり、他話者 s と他話者 t が近いのか遠いのかは重要ではない。即ち MDS や樹形図は必要性の低い情報まで呈示している。更に上記の手法による可視化は、詳細は後述するが、不可避的に歪みが混入されてしまう一方、可視化結果を呈示された側には、どこが歪みなのか知る術がないという欠点がある。

英語は訛りによる多様性以外に、人の声である以上、年齢や性別などの要因によっても音響特性（声色）は当然変わってくる。外国語学習においては、自身の教師と類似した声色の発声の方が聞き取り易い（聞き慣れた声色の発声が聞き取り易い）などの報告もされている [9]。これらを鑑みると、発音の多様性と同様に、年齢や性別と言った非言語的な要因による音響的多様性も、可視化対象として考慮すべきだと考えられる。

本研究はこれらの先行研究を受け、発音距離行列に対する新しい可視化手法を提案する。従来手法の樹形図や MDS では話者群全体を可視化するが、提案手法では、可視化結果を提供する学習者を中心に据え、その学習者とそれ以外の学習者群との関係のみを可視化する。こうすることで、MDS や樹形図に見られる可視化情報の歪みを完全に取り去ることが可能になる。更に、外国語学習というコンテキストを鑑み、各話者の声色の違いも含めた可視化を考える。ここでは声色を特定する二要因である年齢と性別を用いた可視化を検討する。提案法に対して、従来法の比較に基づく客観評価と主観評価を行った。

図 1　MDS を用いた発音距離行列の可視化 [7]

Please call Stella. Ask her to bring these things with her from the store: Six spoons of fresh snow peas, five thick slabs of blue cheese, and maybe a snack for her brother Bob. We also need a small plastic snake and a big toy frog for the kids. She can scoop these things into three red bags, and we will go meet her Wednesday at the train station.

図 3　SAA 読み上げ文章

[pʰl̥iːz kʰɑːl stɛlʌ æsk hɚ ɾə bɹ̃ɪŋ ðiːz θ̃ɪŋz wɪθ hɛɪ
fɪʌ̃m ðə stoːɹ sɪks spu̟ːnz ə̯ fɪɛʃ snoʊ pʰiːʐ̟ faːɣ̟ θɪkˀ
slæːbz ə blu̟ː ʧiːʐ̟ ɛ̃n meɪbi ə snæ̰kˀ foɪ hɚ bɪʌðə bɑb
wi ɑlso n̪iːd ə smɑːl pʰl̥æstɪk sneɪkˀ ɛ̈n ə bɪgˀ tʰɔɪ fɪog fõɪ
ðə kʰɪɪdʐ̟ ʃi kʰĩn skuːp ðiːʐ̟ θ̃ɪŋz̃ ɪntʰu̟ θɪi ɹɛ̰d
bæːgz ɛ̃n wi wɪl miːt hɛɪ wɛ̃ntsdi æt ðə tɪeĩn steɪʃĩn]

図 4　米語母語話者サンプルに対する IPA 書き起こし

[pliz kol stɛlʌ æ̟̃sk ɜ̃ ɾə bɹ̃ɪŋ ðiːʐ̟ s̪̃ɪŋs wɪθ hɜɪ
fɪʌm ðə stoəɪ sɪks spõːnʐ̟ ʌf fɪɛʃ̟ snoʊ piːz
faɪɣ θɪk slæːbʐ̟ ə bluː ʧiːz ɛ̃n meɪbi ə snɛ̃kˀ fɜɪ hɜɪ bɪʌðɜɪ bɑːb
wi ɑlsə̃ nīd ə smɔ̃l plæstɪk sneɪk ə̃n ʌ bɪg tɔɪ fɪɑg fɜɪ
ðə kɪd̪s̪ ʃi kən skūːp ðiːz s̪̃ɪŋz mtu θɪiː ɹɛdˀ
bægz ɛ̃n wi wɪl goʊ̆ mit hɜɪ wɛ̃nsdeɪ ɛɻ ðə tɹeĩn steɪʃə̃n]

図 5　日本語母語話者サンプルに対する IPA 書き起こし

2. 発音距離推定

先行研究では SAA 音声コーパスから IPA 書き起こしに基づく発音距離の推定を検討している [3]〜[5]。本研究の発音距離行列はこれらの研究結果を用いた。

2.1 Speech Accent Archive

SAA コーパスは、米語音素の被覆率を考慮して設計された文章（図 3）の読み上げ音声を、任意の国、任意の話者から、Web を通して収集したコーパスであり、各音声サンプルに国際音声記号（International Phonetic Alphabet, IPA）による書き起こしが添付されている。例として、米語母語話者による読み上げ音声と、日本語母語話者による読み上げ音声に対する IPA 書き起しを図 4、図 5 に示す。修飾記号（diacritical mark）も使用されており、非常に詳細な書き起しが行なわれている。

2.2 発音距離推定

先行研究では、SAA コーパスに付属する IPA 書き起こしに対して動的時間伸縮法（Dynamic Time Warping, DTW）処理を行ない、任意の二話者間の発音のみに関する距離を算出している。IPA 書き起こしは音声学者によって、話者の性別や年齢などに左右されずに行なわれるため、DTW 距離は純粋に発音の差異を定量化していると考えられる。

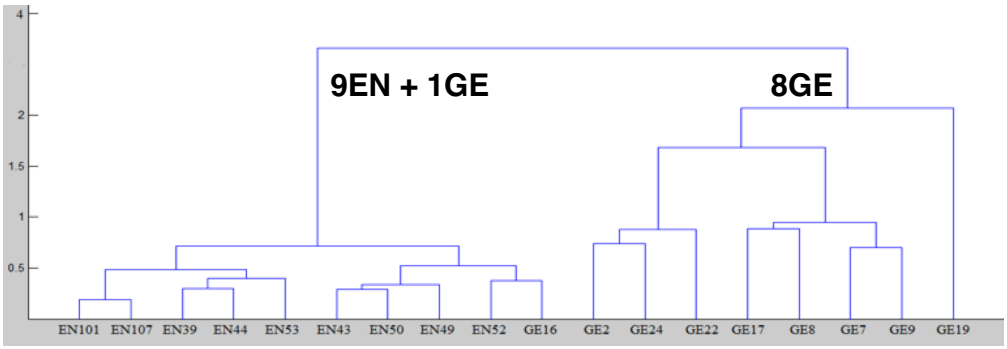

図 2　樹形図を用いたドイツ人話者（GE）とアメリカ人話者（EN）の可視化 [8]

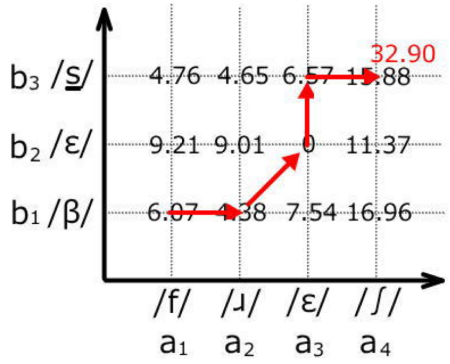

図 6　DTW による 2 つの IPA 間の比較

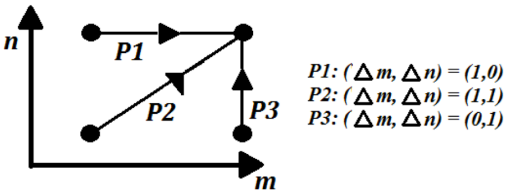

図 7　選択可能な局所的な経路

IPA 書き起こし間の DTW は単音間距離行列が必要である。SAA の IPA 書き起しに観測される異なり音声記号（単音）数は数百あるが、153 種類の記号により、書き起しの 95%を被覆できる。この 153 種類の単音を各々、音声学者に 20 回発声させて 3 状態の話者依存 Hidden Markov Model（HMM）を構築し、2HMM 間の状態間バタチャリヤ距離の平均を単音距離とすることで単音間距離行列を取得している。

この単音間距離行列を用いて書き起こし間の DTW を行ない、話者間の発音距離を算出する。訛りによって単音の挿入等が起こるため、書き起こし中の音声記号数は各話者で異なる。しかし、全員が同じ文章を読み上げ、また単語数は 69 単語で固定されているため、単語毎の対応を取ることで話者間の距離を算出できる。DTW は、2 つの時系列を時間的構造の相違を整合しながら比較するアルゴリズムである [10]。ある単語に対する二話者の音声記号列を以下で表現する。

$$A = a_1, a_2, \cdots, a_m, \cdots, a_M \quad (1)$$

$$B = b_1, b_2, \cdots, b_n, \cdots, b_N \quad (2)$$

a_m と b_n の距離を局所距離 $d(a_m, b_n)$ とし、これは単音間距離行列から得られる。図 6 のように、格子点 $(1,1)$ から (M,N) までの局所距離の合計が最小となる経路を選択し、この経路を最も二系列間の整合がとれた経路として採択する。ここで図 7 を局所的な経路選択の制約条件として導入している。$P1$・$P3$ の経路は IPA の挿入・削除を考慮した経路である。格子点 (m,n) までの部分累積距離は式（3）によって求まる。

$$\begin{aligned} DTW[m,n] = \min(&DTW[m-1,n] + d(a_m, b_n), \\ &DTW[m-1,n-1] + 2 \times d(a_m, b_n), \\ &DTW[m,n-1] + d(a_m, b_n)) \end{aligned} \quad (3)$$

$DTW[M,N]$ を、二話者に出現した音声記号数で正規化し、

$$\frac{DTW[M,N]}{M+N-1} \quad (4)$$

を当該単語における話者間距離とした。69 単語各々の話者間距離を求め、これらを合計することで SAA 文章（図 3）に対する話者 i と話者 j 間の発音距離とし、以下、d_{ij} とする。

3.　知覚的年齢推定

学習者にとっては、世界諸英語に観測される英語発音の多様性も厄介な問題であるが、年齢や性別に起因する音響的多様性も問題となることがある。即ち、ある特定話者の音声のみを集中的に聞くと、それと異なる声色の音声処理に困難を示すことがある [9]。そこで本研究では、英語発音の音響的多様性を、訛り・年齢・性別の三軸で捉え、これらを同時に可視化する手法を提案する。しかし年齢については、任意の話者から実年齢を知ることは社会通念上、難しいこともある。そこで知覚的年齢推定技術を利用し、音声から「聞こえ」としての年齢情報を自動推定し、それを用いることも検討する。

3.1　先 行 研 究

先行研究 [14] では 3 つの音声コーパス、JNAS（成人音声）[11]、S-JNAS（高齢者音声）[12]、CIAIR-VCV（子供音声）[13] の各話者の発声に対して聴取実験により、「聞こえ」としての年齢である知覚的年齢をラベリングさせ、それを用いることにより、未知音声に対する知覚的年齢を推定する技術を提案している。具体的には各話者毎に GMM を構築し、入力音声に対する GMM 尤度を計算し、各話者の知覚的年齢ラベル（分布）と尤度スコアの期待値操作によって知覚的年齢を計算する。知覚的年齢と推定年齢との相関は 0.88 であった。

3.2 知覚的年齢推定の再検討

先行研究は話者 GMM 尤度を用いているが、近年の話者認識研究では Supervector（SV）を特徴量とすることが多く [15]、本実験では GMM-SV と Support Vector Regression（SVR）を使用し、知覚的年齢推定タスクを再検討をした。

3.3 実験条件

精度向上を狙い、複数の点に着目し実験を行った。尚、ここでは男性話者についてのみ報告する。

3.3.1 使用した音声コーパスとデータ分割

使用した音声コーパスは、CIAIR-VCV（6〜12 歳）の男性話者 145 名、JNAS（20〜60 歳）の男性話者 153 名、S-JNAS（60〜90 歳）の男性話者 202 名である。各話者の音声データを分割し、同一知覚年齢に対応するサンプル数を増加させた。具体的には各話者の音声データを 60 秒で分割し（知覚的年齢ラベルは同じ）、区分化データを使って、SVR を学習、評価した。

3.3.2 使用した音声特徴量

先行研究では MFCC のみを用いて話者 GMM を構築している。本研究では MFCC と F0 を用い、各々に対して、混合数 64 の UBM-GMM からの MAP 適応により、話者依存の MFCC-GMM、F0-GMM を得た。これらより、二種類の SV が得られる。なお SV は UBM からも構成できるが、この UBM-SV は、話者依存 GMM から得られる SV に含まれるバイアス項として解釈できる。そこで、GMM-SV から UBM-SV を差し引いた MFCC-SV 差ベクトル、F0-SV 差ベクトルを結合した結果を、SVR の説明変数とした。実験条件を表 1 に示す。

3.4 実験手順

以下の手順で実験を行った。

（1） CIAIR-VCV、JNAS、S-JNAS 中の多数話者を 60 秒で分割した音声から UBM-GMM を、性別に依存させて学習した。MFCC、F0 に対して別個に構築する。

（2） MAP 適応を用いて、GMM-UBM から各話者の GMM を得る。MFCC、F0 に対して別個に構築する。

（3） 各話者 GMM の平均ベクトルを並べて連結し、SV を得る。MFCC、F0 に対して別個に構築する。

（4） MFCC から得られた SV と F0 から得られた SV を結合する。MFCC が 25 次元、F0 が 1 次元であり、各々の GMM の混合数は 64 であるため、次元数は 1664 となる。

（5） この SV を話者特徴量とし、知覚的年齢ラベルを用い、SVR の枠組みで知覚的年齢を予測した。

3.5 実験結果

知覚的年齢ラベルを横軸に、SVR により推定された知覚的年齢を縦軸にプロットすると図 8 となった。相関は 0.89 であり、先行研究よりも僅かに上昇した。

表 1 実験条件

使用コーパス	CIAIR-CVC, JNAS, S-JNAS
UBM	全発話
学習	偶数番号話者（60 秒分割）の UBM からの差分 GMM-SV
評価	奇数番号話者（60 秒分割）の UBM からの差分 GMM-SV
窓幅&シフト長	25ms length / 10ms shift
特徴量	12MFCC + 12ΔMFCC + ΔEnergy + 対数 F0
GMM 混合数	64

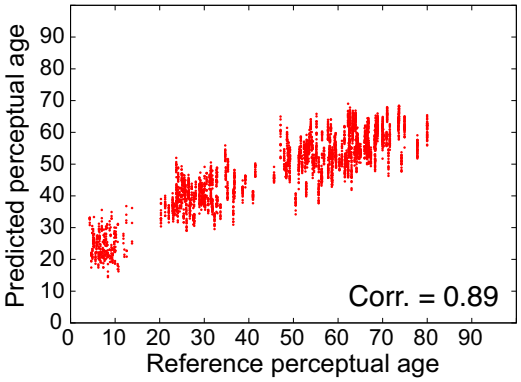

図 8 知覚的年齢と推定年齢との相関

4. 提案手法

4.1 MDS による可視化の問題点

m 次元空間の距離行列 $\{d_{ij}\}$ $(1 \leqq i, j \leqq N)$ は、N 個の点によって張られる幾何学的形状を一意に規定できる。MDS による可視化はその幾何学的形状を 2 次元空間に射影した結果である。射影後の二次元空間上の距離行列を $\{d'_{ij}\}$ とすると、MDS とは $\{d_{ij}\}$ と $\{d'_{ij}\}$ 間の差異を最小化するように変換する過程として捉えることができる。しかし、N>3 かつ m>2 の条件では、MDS の過程において歪み（ストレス）が生じる。

4.2 提案する可視化

距離行列を可視化する際、MDS や樹形図による可視化は距離行列全体を表現していたが、可視化結果を提示される特定の学習者にとってみれば、知りたい主情報は自分とそれ以外の話者との関係性である。つまり、話者 i と話者 j との発音距離行列 $\{d_{ij}\}$ において、話者 n が欲しいのは $\{d_{nj}\}$ $(j \neq n)$ の可視化であり、これは 1 次元空間（数直線）で実現できる。

提案手法では、特定話者（話者 n）に着目し、$\{d_{nj}\}$ と話者 j の年齢 $\{a_j\}(1 \leqq j \leqq N, j \neq n)$ の 2 つの量的情報を 2 次元空間で可視化し、更に、同性・異性という質的情報を 2 つの領域を別個用意して表現する。従来法（MDS）と提案法の比較を簡単にするために、上半円に n と同性話者を、下半円には異性話者を配置する。$\{d_{nj}\}$ と $\{a_j\}$ を用いた提案する可視化の概念図を図 9 に示す。話者 n は原点 **O** に、話者 J は **J** に配置される。a_J は角度で表され、d_{nJ} は **O** と **J** の距離で表される。

先行手法と提案手法の具体的な可視化結果の例を図 10 に示す。SAA 話者 370 名から無作為に抽出された男性 10 名・女性 10 名、計 20 名に対する結果である(注1)。左から MDS、青枠で囲まれた話者に対する提案手法、赤枠で囲まれた話者に対する提案手法、である。図 10 は実年齢を用いた可視化である。

(注1)：各話者のプロット位置に点ではなく顔写真を配置することで、より直感的な可視化とした。但し SAA 話者自身の顔写真は取得不可能であるため、SAA 話者情報から得られる性別・実年齢ラベルと、顔画像データベース [17] にある性別・実年齢ラベルを参照して、SAA の各話者に顔写真を付与した。

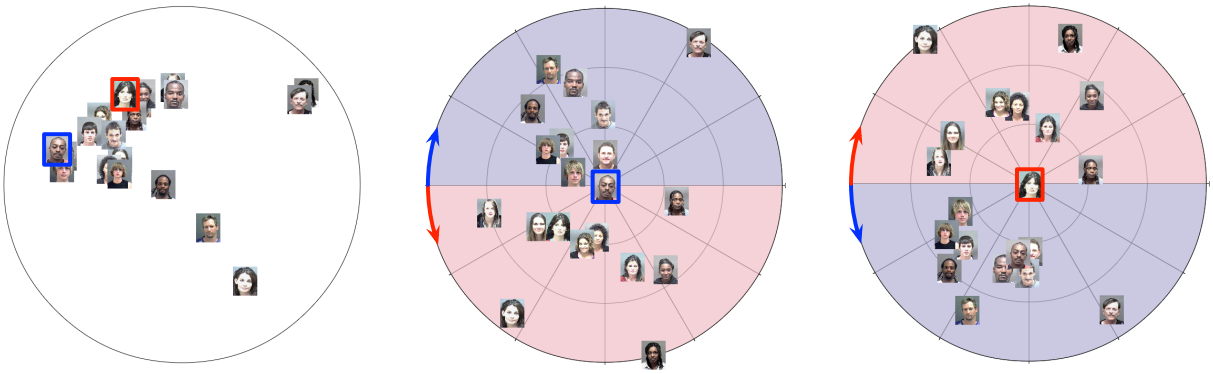

図 10　様々な英語発音に対する従来手法と提案手法の可視化結果

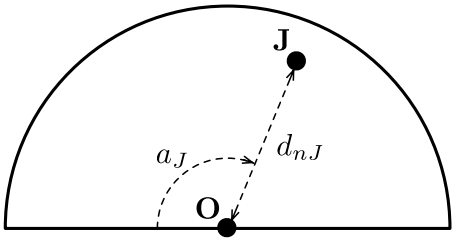

図 9　年齢と発音距離の可視化

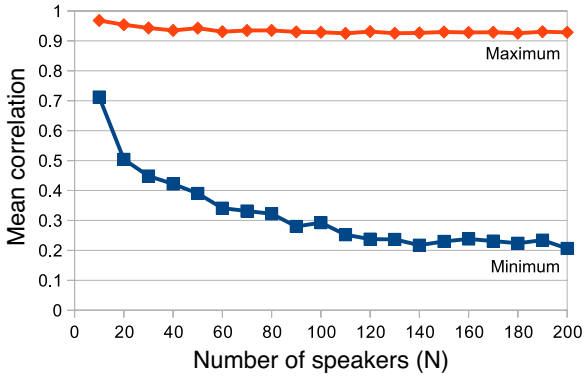

図 11　話者数の増加による相関値の減少

表 2　主観的比較結果

MDS による可視化	提案する可視化	p 値
6.0	6.3	15.7%

実年齢	推定年齢	p 値
5.9	5.0	4.8%

5.　提案手法についての評価

5.1　客観的比較

第 4.1 節で述べたように、MDS を用いた可視化は歪みを含むことがある。例えば図 10 においては、元の距離行列 $\{d_{ij}\}$ と MDS 処理後の距離行列 $\{d'_{ij}\}$ の間の相関は 0.87 である。各中心話者（話者 n）について $\{d_{nj}\}$ と $\{d'_{nj}\}$ の相関を求めると、その最大値と最小値は 0.98 と 0.73 であった。20 人の学習者の各々に対して MDS を用いた可視化結果を返す場合、学習者によっては歪んだ可視化結果が渡されることになる。しかし、その結果に歪みが含まれている事実をその学習者は知る術がない、という事実は教育的には非常に大きな問題である。

この歪みは、話者数 (N) が増加することでより大きくなり、相関の最小値が減少していくことが考えられる。N の増加によって相関の最小値が減少していく様子を図 11 に示す。図 11 では相関の最大値についても示した。ここでは、無作為に N 人の話者を抽出し相関の最大値と最小値を求め、これを 30 回繰り返しその平均を示している。相関の最大値については十分に保たれたが、最小値については単調減少が見られた。N が 100 の場合は最小値は約 0.3 となり、これは、ある学習者は大きく歪められた可視化結果を受けとることを意味する。

しかしながら提案手法では、図 9 に示すように距離 $\{d_{nj}\}$ をそのまま用いるため、全話者において可視化結果に歪みが全く生じない。この点は教育的に非常に重要である。

5.2　主観的比較

従来手法と提案手法について、アンケートによる主観評価実験を行った。第 4. 節で用いた 20 人の話者を、各手法を用いて可視化して被験者に呈示し、これらを評価させた。本評価実験は Web 上で行い、被験者が各手法を自由に選択し、中心話者を自由に選択できるようにした。顔写真をクリックすることで話者の音声が聞けるようにし、被験者が可視化に対する評価をより厳密に行えるよう配慮した。被験者は学生 17 名、社会人 13 名の合計 30 名である(注2)。評価項目は次の二点である。

1) 各手法の直感的な分かり易さについての評価

直感的分かり易さを 0〜10 の 11 段階で絶対評価させた。MDS の結果が歪みを含み、正確さに欠けることは教示していない。これは、可視化の正確さには着目させず、可視化の直感的な分かり易さのみに着目させるためである。

2) 実年齢と知覚的年齢の可視化における妥当性評価

実年齢と知覚的年齢それぞれの可視化を提示し、年齢と各話者の位置関係の妥当性を評価させた。尚、両手法は、年齢情報を異なる手法で推定していると説明し、一方が実年齢であることは伝えていない。1) 同様、11 段階の絶対評価である。

評価結果を表 2 に示す。直感的な分かり易さについては、従来手法に比べて提案手法が若干高い評価を得たが、分散分析を行ったところ、15.7%の有意水準のみしか得られなかった。こ

(注2)：日本語母語話者が多いが、それ以外にも英語・ポーランド語・ベトナム語・中国語・チェコ語・韓国語母語話者もいる。また語学教師も含まれる。

れは、提案手法と従来手法は、直感的な分かり易さに関してはほぼ同レベルであることを意味する。本主観評価実験の後に、両手法に対する自由記述も行わせた。それによると、発音距離に加え年齢・性別情報も含む提案手法の方が分かり易い、という意見が幾つか見られた。その一方、例えば教師の目線で学習者の発音多様性を見る場合は、MDS の俯瞰的な可視化が目的により合致しているとの意見も得られた。このことから、両手法ともに長所・短所があるように思われるが、従来手法は歪みを含み、正確性の問題があるという欠点は依然として残る。

教師が複数生徒の発音を可視化したい場合には、教師を中心に据えて提案手法を実施すれば良い。例えば、授業の前後で、教師に対して、生徒がどの程度中心に近づくか（授業を通して教師の英語発音にどの程度似ていくか）を見ることが可能となる。この教師目線の可視化は、学習者に対しても有益であろう。

実年齢及び知覚的年齢における話者の位置関係については、表 2 より実年齢の方が危険率 5%未満でより妥当であると判断され、知覚的年齢推定技術の不十分さが示される結果となった。この推定器の学習には日本語の音声コーパスを用いているが、評価の際には SAA 音声コーパス（即ち話者の母語は様々）を用いた点で、言語差が精度に影響したことも一要因であろう。実際にアプリケーションとして知覚的年齢推定技術が必要になる場合には、より目的に合致したコーパス・ラベル収集、及び、推定手法の調整・最適化が必要となるだろう。

6. ま と め

本研究では、特定話者を中心に据え、発音・年齢・性別を同時に可視化する手法を提案した。客観的な比較では従来手法と比べ、原理的に歪みを生じ得ない高い正確さが示された。しかし主観的な比較では、提案手法の長所・短所が指摘された。被験者のコメントの中には、提案手法を用いることで、世界諸英語学習の教育に効果的であるとの意見も見られた。

将来的には、TED(注3)アーカイブ中の様々な話者から、SAA 音声データを集める事を考えている。もし学習者を中心に TED 話者群を貼り付けた可視化が出来れば、その学習者に特化した TED talk ブラウザ、即ち、世界諸英語 talk ブラウザが実現できる。このような応用アプリケーションが、世界諸英語の効率的な学習にどのように貢献できるのかについても検討したい。

文 献

[1] B. Kachru, Y. Kachru, and C. Nelson, *The handbook of World Englishes*, Wiley-Blackwell, 2009.

[2] J. Jenkins, *World Englishes: a resource book for students*, Routledge, 2009.

[3] H.-P. Shen, N. Minematsu, T. Makino, S.H. Weinberger, T. Pongkittiphan, and C.-H. Wu, “Automatic pronunciation clustering using a world English archive and pronunciation structure analysis,” *Proc. Automatic Speech Recognition and Understanding*, pp.222–227, 2013.

[4] S. Kasahara, S. Kitahara, N. Minematsu, H.-P. Shen, T. Makino, D. Saito, K. Hirose, “Improved and robust prediction of pronunciation distance for individual-basis clustering of World Englishes pronunciation,” *Proc. ICASSP*, 2014 (to appear)

[5] N. Minematsu, S. Kasahara, T. Makino, D. Saito, K. Hirose, “Speaker-basis accent clustering using invariant structure analysis and the speech accent archive,” *Proc. Odyssey*, 2014 (to appear)

[6] Speech Accent Archive, `http://accent.gmu.edu/`

[7] N. Minematsu, “Training of pronunciation as learning of the sound system embedded in the target language,” *Proc. The Phonetic Conference of China and Int. Symposium on Phonetic Frontiers*, CD-ROM, 2008.

[8] H.-P. Shen, N. Minematsu, T. Makino, S. H. Weinberger, T. Pongkittiphan, C.-H. Wu, “Speaker-based accented English clustering using a world English archive”, *Proc. Speech and Language Technology in Education*, pp.184–188, 2013.

[9] D.B. Pisoni, “Some thoughts on normalization in speech perception,” in *Talker Variability in Speech Processing*, edited by K. Johnson and J.W. Mullennix, Academic Press, New York, pp.9–32, 1997.

[10] 迫江 博昭, 千葉 成美, “動的計画法を利用した音声の時間正規化に基づく連続単語認識,” 日本音響学会誌, vol.27, no.9, pp.483–490, 1971.

[11] 新聞記事読み上げ音声コーパス JNAS, `http://research.nii.ac.jp/src/JNAS.html`

[12] 高齢者話者データベース S-JNAS, `http://research.nii.ac.jp/src/S-JNAS.html`

[13] 子供の声データベース CIAIR-VCV, `http://research.nii.ac.jp/src/CIAIR-VCV.html`

[14] 山内 景太, 峯松 信明, 広瀬 啓吉, “話者認識技術を応用した知覚的年齢分布の自動推定,” 電子情報通信学会技術研究報告, SP2002-186, pp.43–48, 2003.

[15] T. Kinnunen and L. Haizhou, “An Overview of Text-Independent Speaker Recognition: from Features to Supervectors,” *Speech Communication*, vol.52, no.1, pp.12–40 (2010)

[16] Praat, `http://www.fon.hum.uva.nl/praat/`

[17] K. Ricanek and T. Tesafaye, “MORPH: A Longitudinal Image Database of Normal Adult Age-Progression,” *Proc. Int. Conf. Automatic Face and Gesture Recognition*, pp.341–345, 2006.

(注3)：Technology Entertainment Design。非英語母語話者も含む世界中の様々な分野の第一線で活躍する人物の英語による講演を開催しているグループ。インターネットを通じて講演動画を無料で提供している。