

音声の構造的表象に基づく 英語発音分析結果の視覚化に対する一考察*

鎌田圭[†], 高澤真章[†], 朝川智[†], 峯松信明[†], 牧野武彦[‡], 広瀬啓吉[†] ([†] 東大, [‡] 中大)

1 はじめに

音声は、発声者の年齢や性別、収録機器の音響特性といった非言語的要因により不可避免的に歪み、同じ発話内容でも観測される音響事象は異なる。従来の音声工学システムでは、この音声の歪みを内包した特徴量を利用していため、話者の違いによる不安定性（不一致問題）を避けることは困難であった。

近年提案された「音声の構造的表象」では、このような音声の歪みを除去するために、音声事象の相対的な関係のみに着目している [1]。この表象は、ある音声事象は他の音声事象群との距離のみにによって規定され、個々の音声事象が持つ特性は一切参照しない。この表象を用いることで、英語学習者の発音状態を非言語的特徴の影響を受けずに記述可能である [2]。

例えば英語の母音発音を表象の対象とした場合、提案する発音の構造表象は、個々の英語学習者の発音状態を母音間距離行列の形で表現する。その結果、Fig.1 のような樹形図や、Fig.2 のような多次元尺度構成法による 2 次元母音図の形で表現することができる。しかし、発音状態を表す母音間距離行列（発音構造）は元来、高次元 ($n>2$) 空間における幾何学構造であるため、Fig.1 や Fig.2 のように 2 次元平面上で表示する際には、情報の欠落や情報の歪みが不可避である。この 2 次元化に伴う情報の劣化により、間違った情報を英語学習者に提示する可能性がある。

本稿では、英語発音構造を表す距離行列を 2 次元視覚化する方法として、計量的多次元尺度構成法を採用し、情報の歪みを抑えつつ、かつ、教育的に妥当な発音構造を記述する方法について考察した。

2 音声の構造的表象

音声に不可避に混入する非言語的特徴による音響歪みは、乗算性歪みと線形変換性歪みに大別される。乗算性歪みは、マイクなどの伝送特性や話者の声道形状の違いの一部に相当し、ケプストラムベクトル c に対するベクトル b の加算 $c'=c+b$ となる。線形変換性歪みは、声道長の差異や聴覚特性の差異が挙げられ、行列 A の乗算 $c'=Ac$ で精度良くモデル化される [3][4]。以上より、不可避かつ静的な非言語的音響歪みの一番簡素なモデルを考えれば、それは、アフィン変換となる。Fig.3 はこの 2 種類の歪み（乗算性歪み = 垂直、線形変換性歪み = 水平）を示している。

個々の音声事象を分布として捉え、全ての二事象間距離を計算すると（距離行列）、それは、事象群が成す幾何学構造を規定することになる。パタチャリヤ距

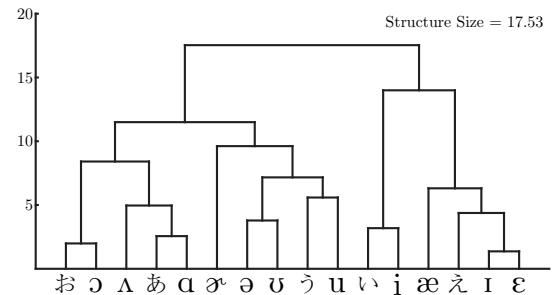


Fig. 1 発音構造の樹形図による視覚化

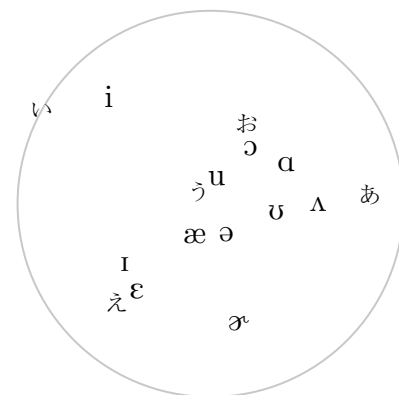


Fig. 2 発音構造の MDS による視覚化

離はアフィン変換のみならず（一対一対応かつ全空間において微分可能な）非線形変換も含め不変性を有するため [5]、これを用いることで、静的な非言語的音響歪みに一切不変な音声表象が得られる。この構造を「音響的普遍構造」と呼んでいる。

3 英語学習者の音声の収録

英語学習者（主に日本人）の音声として 564 名分の音声を収録した。収録内容は米語 11 単母音と日本語 5 母音である。収録では米語 /bVt/、日本語 /bVto/ を 1 回発声させ、母音部分のみを切り出して実験に用いた。米語 /bVt/ に有意味語が存在しない場合は、/b/ を /p/、/t/ を /d/ とするなどの処置を行った。

4 英語学習者の発音状態の視覚化

4.1 英語母音発音構造の抽出

得られた音声資料から母音区間を自動切り出しにより求め、各母音区間をガウス分布でモデル化する。用いた音響分析条件を表 1 に示す。日米 16 母音の母音間距離（パタチャリヤ距離）をすべて求め、 16×16

* Examination of visualizing English pronunciation analysis results based on the structural representation of speech. by K.Kamata, M.Takazawa, S.Asakawa, N.Minematsu, T.Makino, and K.Hirose

Table 2 母音群別の視覚化歪み

	母音群		表示 母音数	平均 相関係数	相関係数		
	日本語母音	英語母音			0.85 以上	0.90 以上	0.95 以上
全部	5 母音全部	11 母音全部	16	0.907	91.5 %	64.5 %	9.06 %
/あ/に関係 する母音群	/あ/, /お/	/ɑ/, /ʌ/, /æ/, /ə/, /ɜ/, /ɔ/	8	0.933	94.3 %	78.5 %	45.6 %
	/あ/, /お/	/ɑ/, /ʌ/, /æ/, /ə/, /ɔ/	7	0.948	97.2 %	88.3 %	58.1 %
	/あ/	/ɑ/, /ʌ/, /æ/, /ə/, /ɜ/	6	0.948	95.7 %	87.4 %	61.8 %
/い/に関係 する母音群	/い/, /え/	/æ/, /ə/, /ɪ/, /i/, /ɛ/	7	0.975	99.8 %	99.6 %	87.6 %
	/い/, /え/	/æ/, /ɪ/, /i/, /ɛ/	6	0.981	100.0 %	98.9 %	94.3 %
/う/に関係 する母音群	/う/, /お/	/ɑ/, /ə/, /u/, /u/, /ɔ/	7	0.949	97.3 %	89.9 %	60.9 %
	/う/, /お/	/ɑ/, /u/, /u/, /ɔ/	6	0.963	98.4 %	94.1 %	77.4 %

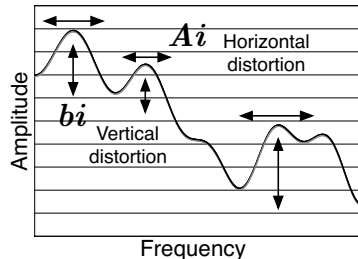


Fig. 3 スペクトルに対する水平/垂直方向の音響歪み

Table 1 音響分析条件

サンプリング	16bit/16kHz
窓長, シフト長	25 msec, 1 msec
特徴量	Mel Cepstrum (1-10 次元)
推定法と分布	MAP 推定によるガウス分布

の母音間距離行列で表現される英語発音構造（16 角形）を得る．この発音構造が，音声の音響量から非言語的要因に起因する歪みを排した発音表象である．

4.2 英語母音発音構造の視覚化

発音状態を示す距離行列を視覚化する方法として，多次元尺度構成法（以下 MDS）を用いた．MDS は計量的 MDS と非計量的 MDS に分けられるが，本稿では距離情報の再現を目指す計量的 MDS を用いた [6]．564 名の発音構造（母音間距離行列）に MDS を適用し，564 枚の 2 次元母音図を得た．

4.3 英語発音構造の視覚化に伴う歪みの評価

MDS によって作られた 2 次元母音図は母音間距離行列を完全には再現しておらず，低次元化に伴う歪みを内包している．本稿では，この視覚化による歪みを相関係数を用いて定量的に計測した． n 次元空間における母音間距離と 2 次元母音図上の母音間距離の相関を，564 枚の 2 次元母音図の各々について調べた．16 種類の母音があるため，2 次元母音図 1 枚につき，120 個の距離が存在する．これらの相関値を母音図に対する相関値とした場合，平均相関係数は 0.907 となった．相関係数 0.90 以上の 2 次元母音図は全体の 64.5 %，0.95 以上は 9.06 %であった．

4.4 母音群の分割による歪みの低減

一般的に，一枚の 2 次元母音図に表示する母音の数が少なければ，視覚化による歪みも小さくなる．そこで，16 母音を分割し，一枚の 2 次元母音図に表示する母音数を減らすことで，歪みを低減することを考

える．実際に 16 母音全てを同時に提示しても，どの部分に焦点を当てて分析結果を見るべきなのかが不明瞭であり，教育的な観点からも示すべき内容を絞った方が相応しい．しかし，英語教育的に意味のある母音グループにおいて，2 次元視覚化に伴う歪みが十分小さいことが必要条件となる．英語音声学の専門家である第五著者により，教育的に意味のある母音グループの候補を幾つか挙げてもらい，個々のグループに対して歪みの大きさを算出した．結果を Table 2 にまとめる．16 母音すべてを視覚化した場合に比べ，平均相関係数は高くなり，0.90 以上の相関係数を示す 2 次元母音図も多くなった．

5 まとめと今後の課題

音声の構造的表象に基づく英語母音発音構造に対して，計量的 MDS を用いることで，母音間の関係性が解りやすい 2 次元母音図の作成を試みた．しかし，この場合，2 次元化することにより高次元情報である母音間距離情報は歪んでしまう．そこで，母音群を分割し，一度に表示する母音数を減らすことにより，歪みを効果的に減少できることを確認した．教育上どの程度の相関係数を保持すべきなのかについては，実践の場を通して検討する予定である．

参考文献

- [1] 峯松信明, 西村多寿子, 西成活裕, 櫻庭京子, “構造不変の定理とそれに基づく音声ゲシュタルトの推定,” 信学技報, SP2005-12, pp.1-8, 2005.
- [2] 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,” 信学論 (D), vol.J90-D, no.5, pp.1249-1262, 2007.
- [3] M. Pitz and H. Ney, “Vocal tract normalization equals linear transformation in cepstral space,” IEEE Trans. Speech and Audio Processing, vol.13, pp.930-944, 2005.
- [4] 江森正, 篠田浩一, “音声認識のための高速最尤推定を用いた声道長正規化,” 信学論 (D-II), vol.J83-D-II, no.11, 2000.
- [5] 峯松信明, 朝川智, 広瀬啓吉, “線形・非線形変換不変の構造的表象とそれに基づく音響モデリングに関する理論的考察,” 音講論春, 1-P-12, pp.147-148, 2007.
- [6] 齋藤亮幸, “多次元尺度構成法,” 朝倉書店, 1980.