

英語強勢・弱勢母音における音響的普遍構造のサイズと 発音習熟度に関する分析*

◎ 朝川 智 (東大・情報理工) 峯松 信明 (東大・新領域) 広瀬 啓吉 (東大・情報理工)

1 はじめに

音声生成、伝達、聴取の過程に混入する不可避免的な静的歪みに全く影響を受けない新しい音声の表象手法として音響的普遍構造が提案されている。音響的普遍構造のサイズに着眼し、英語強勢・弱勢母音に対する構造分析を行い、構造サイズが発話者の調音努力を記述していることを既に報告した [1]。また一方、音響的普遍構造を用いて英語学習者の発音習熟度の自動評定を行う試みも行われている [2]。

本報では、日本語母語話者の英語発話に対して、英語強勢・弱勢母音群の構造サイズと各話者の発音習熟度を分析し、強勢・弱勢構造サイズ比と発音習熟度との間に高い相関があることを実験的に示す。

2 音声の音響的普遍構造

2.1 音声の音響的普遍構造

ある言語に含まれる各音素を、ケプストラムベクトルによって構成される多次元ガウス分布であると仮定し、 n 個の音素に対する nC_2 個の音素間距離行列を求めることで音声を構造化することを考える。音素間距離としてバタチャリヤ距離の平方根を用いる。二つの確率密度分布 $p_i(x)$ と $p_j(x)$ 間のバタチャリヤ距離は、以下の式によって与えられる。

$$BD(p_i(x), p_j(x)) = -\ln \int_{-\infty}^{\infty} \sqrt{p_i(x)p_j(x)} dx \quad (1)$$

バタチャリヤ距離は、任意の二分布間に対して如何なる共通のアフィン変換を施しても距離は不変である、という性質を持つ。つまり、得られる音素間距離行列もアフィン変換前後で不変であるといえる。ケプストラムベクトル c に対するアフィン変換 $Ax+b$ では、行列 A の乗算は声道長特性、聴覚特性など話者・聴取者の生理学的な差異を表現し、ベクトル b の加算は音響機器などの差異を表現する事を考えると、音響的普遍構造はこれらの差異によって一切歪まない物理表象であると言える。

2.2 音響的普遍構造のサイズに対する物理的解釈

英語には弱勢としてのみ出現する母音、弱母音 (schwa) がある。英語音声学によると、schwa は母音図中において全ての母音の中心に位置し、その他の母音は弱勢になると schwa に近づくと言われている [3]。これを音響的普遍構造という物理表象の上で考えると、弱勢化された母音は互いの距離が短くなり、普遍構造のサイズが小さくなることが考えられる。上記の考察を定量的に検討するため、既に [1] において、母語話者音声データを用いて強勢母音群・弱勢母音

群の普遍構造サイズを分析し、強勢・弱勢が構造のサイズの大小として表現されることを実験的に示した。強勢母音はより大きな調音努力を払って母音間の差異を明確化したものであり、弱勢母音は不明瞭化したものである事を考えると、音響的普遍構造のサイズは発話者の調音努力を指すと解釈できる。なお、[1] でも述べているように、音響的普遍構造はスペクトルのみならず、パワー・継続長などの韻律的特徴についても考慮された音声の物理表象となっている。

本研究では、強勢・弱勢母音の構造サイズ分析を、日本人の英語音声データを対象として行う。日本語と英語とは非常に大きな差異を持つ言語対であり、特に英語特有のリズムの習得は日本人学習者にとって非常に高いハードルとなっている [4] 一方で、日本人英語が英語母語話者に通じない最大の要因はリズムである、との分析も報告されている [5]。そして、英語リズム習得には、どこを強く発音しどこを弱く発音するのかしっかりと区別することが重要であると言われる。本分析では、強勢・弱勢構造サイズを基に、強勢となるべき母音がいかに強勢らしく、弱勢となるべき母音がいかに弱勢らしく発声されているかを定量化し、各話者の発音習熟度との相関について分析することを目的とする。

3 普遍構造サイズと発音習熟度の分析

3.1 分析手法

音声データ中の全ての母音に対して強勢・弱勢ラベリングを施し、そのラベルを用いて強勢母音・弱勢母音を別々に学習して音素 HMM を構築する。得られた音素 HMM より、強勢母音・弱勢母音それぞれの音素セットに対して音素間距離行列を求め、Ward 法を用いたボトムアップクラスタリングにより樹形図を求める。Ward 法とは、統合により生じる累積歪みが最小となる二要素を順次統合していく方法であり、最終的に得られる樹形図の高さは全要素を一点で代表させたときの歪みとなる。これは VQ 歪みに相当し、音素群を構造としてみた場合の構造の半径に相当する量として解釈することができる。これを構造のサイズと見なして定量的な検討を行う。

音声データとしては、ERJ (English Read by Japanese) データベース [6] 中の文セット 6 (60 文) を読み上げた話者 19 名 (男性 10 名、女性 9 名) のデータを用いて、表 1 の条件に従って HMM を作成した。強勢・弱勢のラベルは、収録時に使われた読み上げリストに付与されている強弱ラベルに対し、機

* Estimation of the pronunciation proficiency based on ratio of sizes of the acoustic universal structures of stressed and unstressed vowels.

By Satoshi Asakawa, Nobuaki Minematsu and Keikichi Hirose (University of Tokyo)

表 1. HMM 学習条件

サンプリング	16bit / 16kHz
窓	窓長 25 ms, シフト長 10ms
パラメータ	メルケプストラム (0~10 次元)
話者	日本語母国話者 男性 10 名 女性 9 名
学習データ	60 文 (音素バランス文の一部)
HMM	1 混合 monophone (全角分散行列)
トポロジー	3 状態 1 分布

表 2. 実験で用いた母音セット

強勢母音セット	a1,i1,u1,A1,E1,I1,U1,R1,@1
弱勢母音セット	a0,i0,u0,A0,E0,I0,U0,R0,@0

能語の母音を全て弱勢とする変更を施したものを使用した。分析対象とした母音は、二重母音と強勢・弱勢でデータ量に大きな偏りが見られた母音については分析対象から除外した。最終的に表 2 に示す母音を対象として音素間距離行列を求め、強勢・弱勢それぞれの構造サイズを算出した。

発音習熟度に関しては、ERJ データベースには分節的特徴に着眼した評定スコア (5 点が最高、5 人の米国人英語教師によるスコアの平均値) が付けられており、このスコアを発音習熟度として用いることとした。なお、ERJ データベースにはリズム文読み上げに対する評定スコアもあるが、この評定値は、強勢シラブル出現に関する等時性への着眼も含まれる。本分析手法では母音の継続長は反映されるものの、その等時性については明確な接点を持たない。また、母音の schwa 化というのは明確な音色の変化を伴うものであり、ここでは分節的特徴に基づく評定値を用いた。

3.2 分析結果

結果を図 1 に示す。横軸は英語教師による評定スコア、縦軸が強勢弱勢構造サイズ比 (弱勢サイズに対する強勢サイズの比の値) である。サイズ比が 1.0 より大きいほど、強勢は強勢らしく、弱勢は弱勢らしく発声されている、ということになる。

図より、発音習熟度が高い話者ほどサイズ比が大きくなっている傾向が確認できる。発音習熟度とサイズ比の相関係数は 0.89 となり、比較的高い相関がある事が確かめられた。習熟度の低い話者においてはサイズ比が 1.0 を下回っており、比較的ばらつきが大きい。また、習熟度が低いにもかかわらずサイズ比が大きくなるという逆転が一部見受けられる。

参考までに、ERJ データベース中の母国語話者 4 名の音声データに対して同一条件下で分析したところ、構造サイズ比は平均で 1.17 となった。

3.3 考察

上記の結果より、発音習熟度の高い話者ほど、強勢を強勢らしく、弱勢を弱勢らしく、強弱の区別をつけて発音ができている事が分かる。そして、習熟度とサイズ比の間に高い相関があることが確認できた。低習熟度話者において、サイズ比が 1.0 を下回っている話者が見られるが、機能語のような弱く発音

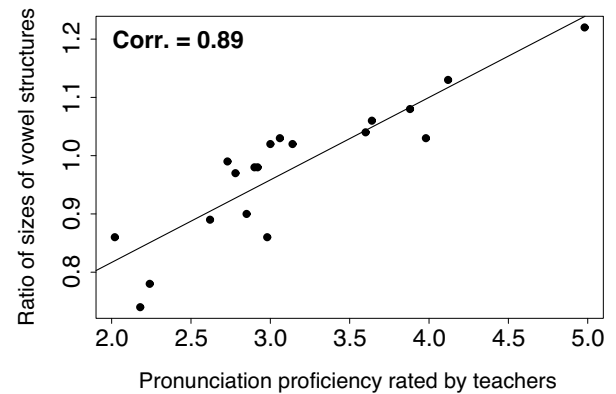


図 1. 強勢弱勢構造サイズ比と発音習熟度との相関

すべき単語を、逆に強く発音してしまう、といった原因が考えられる。低習熟度話者におけるばらつきや習熟度とサイズ比の逆転現象については、低習熟度話者において頻発すると考えられる母音挿入に対して現時点の分析方法では対応できていないことが原因の一つではないかと考えられる。また、文単位の発声における強勢の位置は発話者の意図によって変わりうるが、今回用いたデータでは発話者の意図、即ち文として読んだときの強勢位置までは統制されていないことも原因として挙げられる。今後、日本人英語によくみられる母音挿入誤りなどに対応した分析手法の確立と、発話意図の統制された音声データの収集が必要である。

4 まとめ

日本人の英語音声データを対象として、強勢となるべき母音がいかに強勢らしく、弱勢となるべき母音がいかに弱勢らしく発声されているかを、強勢・弱勢母音群に対する音響の普遍構造のサイズ比により定量化し、発音習熟度と構造サイズ比との間に高い相関があることを分析的に確認した。

今後の課題として、日本人英語によくみられる母音挿入誤りなどに対応した分析手法の確立が挙げられる。また現在、発話意図の統制された音声データの収集を行っており、それを用いてより精度の高い分析を今後行っていく予定である。

参考文献

- [1] 朝川他, “音響の普遍構造のサイズと単語境界における音響的分離度に着眼した米語音声の音響分析”, 信学技報, SP2004-28, pp53-58 (2004)
- [2] 峯松, “構造音韻論の物理実装に基づく学習者表象と発音学習支援”, 信学技報, TL2004-47, pp47-52 (2004)
- [3] 川越いつえ, 英語の音声科学, 大修館書店 (1996)
- [4] A.Cutler, “Segmentation problems, rhythmic solutions,” *Lingua*, 92, pp.81-104 (1994)
- [5] N.Minematsu, et al., “CART-based factor analysis of intelligibility reduction in Japanese English,” *Proc. EUROSPEECH*, pp.2069-2072 (2003)
- [6] 峯松他, “英語 CALL 構築を目的とした日本人及び米国人による読み上げ英語音声データベースの構築”, 日本教育工学会論文誌, vol.27, no.3, pp259-272 (2004)