

1 はじめに

話者や収録環境に依存しない音声の音響的普遍構造が存在することが示されている [1]。音響的普遍構造表現は、音声に対して不可避免的に混入する乗算性及び線形変換性歪みに対して不変であり、発声者及び聴取者の生理学的条件や収録・伝送・再生環境の音響的条件に依存しない形で音声の物理表象である。

本報では、英語における強勢・弱勢母音に対して音響的普遍構造の抽出を行い、構造のサイズが発話者の調音努力を記述することを示す。

2 音声の音響的普遍構造 [1]

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、音素間の距離をバタチャリヤ距離の平方根によって定義する。ここで、分布 u と v 間のバタチャリヤ距離は、以下の式によって与えられる。

$$BD(u, v) = -\ln \int_{-\infty}^{\infty} \sqrt{P_u(x)P_v(x)} dx \quad (1)$$

但し、 P_u, P_v はそれぞれ u, v の確率密度関数である。

与えられた (ある空間内の) n 点に対して、 nC_2 個存在する全対角線の長さを考慮する (即ち距離行列を求める) ことは、 n 点で構成される構造の情報を考えることに等しい。即ち、全ての音素分布間距離を上式により求めることは、音素群がなす構造を情報理論的に規定することである。

上記により求められる音素間距離行列は、各分布を構成するデータに如何なる共通の一次変換を施しても不変である、という性質を持つ。ケプストラムベクトル c に対する一次変換 $Ac + b$ において、行列 A による変形は構造の回転として観測され、 b による変形は構造のシフトとして観測されることとなる。ここで、周波数ウォーピングがケプストラムベクトルに対する A の演算、フィルタリングが b の演算に相当し、前者は声道長の差異やバーク尺度などの聴覚特性、後者はマイク・伝送路の特性・録音室の音響特性などを表していることを考えると、この構造は、発声者及び聴取者の生理学的条件、及び収録・伝送・再生環境の音響的条件に依存せず普遍的に観測される構造であるといえる。以上のことからこの構造を、音声の音響的普遍構造と定義している。

このように音響的普遍構造は、そのシフトと回転に明確な物理的意味を持つ。本研究は、構造のサイズに対する物理的解釈を実験的に検討する。

3 英語母音における弱勢母音の位置づけ

英語は日本語と異なり、強勢弱勢の属性と母音のカテゴリとが独立ではなく、弱勢としてのみ出現する母音、ほぼ強勢となる母音などが存在する。弱勢母音 (schwa) は弱勢としてのみ現れる母音である。schwa は母音図中、全ての母音の中心に位置し [2]、また、母音生成を音響管近似した場合、声道を断面積一定の管で近似した際に得られる音質である [3]。即ち、schwa は最も基本的な、その発声に大きなエネルギー用いない母音であると考えられることができる。

英語音声学によると、母音は弱勢になると schwa に近づくといわれている [2]。中心母音である schwa に近づくということは、弱勢化された母音は互いの間隔が短くなることを意味し、これは上記した音響的普遍構造が小さくなることを示唆している。本研究では、上記の考察を定量的に検討することを目的とする。

4 英語強勢・弱勢母音の分析

4.1 分析手法

音声データ中のすべての母音に対して強勢弱勢ラベリングを行う。そして、そのラベルを用いて強勢母音・弱勢母音を別々に学習し音素 HMM を構築する。

次に、第 2 節で説明した音響的普遍構造の抽出を、強勢母音のみ・弱勢母音のみの場合それぞれに対して行う。具体的には、強勢母音のみ・弱勢母音のみ、それぞれの音素セットに対して音素 HMM 間の距離行列を求め、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として視覚化する。

Ward 法は、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要素群を一点で代表させた時の歪みとなる。これは VQ 歪みに相当し、また構造全体のサイズとして解釈できる。

4.2 分析条件

米語母国語話者 2 名 (男性 1 名、女性 1 名) に対して、強勢母音・弱勢母音のラベリングを聴取により行い、各話者の音響モデルを表 1 の条件に従って作成した。母音の種類は、i, I, e, E, @, a, c, o, U, u, Y, O, W, R, A, x である。

次に、得られた強弱別の HMM より強勢母音・弱勢母音それぞれの音素間距離行列の算出を行うが、その際、二重母音 (o, e, Y, W, O) と、強勢と弱勢とでデータ量に大きな偏りがみられた母音については分析の対象から除外し、表 2 に示される音素を対象と

して音素間距離行列を求めた。また、バタチャリや距離の計算の際には、ケプストラムパラメータの0次元の項(パワー項)は除いて計算を行っている。

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	改良ケプストラム (0~12 次元)
話者	米語母国語話者 男性 1 名 女性 1 名
学習データ	男性 746 文, 女性 709 文
HMM	1 混合 monophone (全角分散行列)
トポロジー	3 状態 1 分布

表 2. 音素セット

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

さらに比較のため、強勢母音・弱勢母音を区別しないで学習した HMM(通常の HMM) も作成した。

4.3 分析結果

女性話者の場合の強勢母音・弱勢母音における音素樹形図をそれぞれ図 1, 図 2 に示す。また、通常の HMM での全母音の音素樹形図を図 3 に示す。各図、樹形図の頂点のノードにおける数値が構造のサイズを表している。

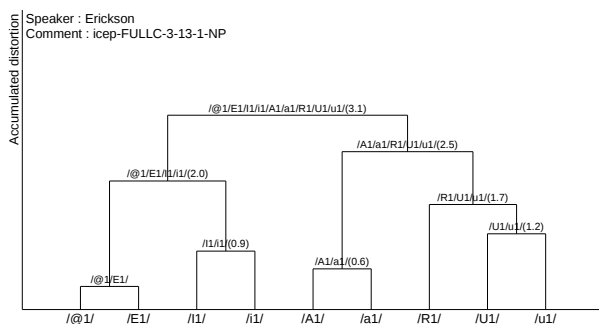


図 1. 強勢母音の音素樹形図

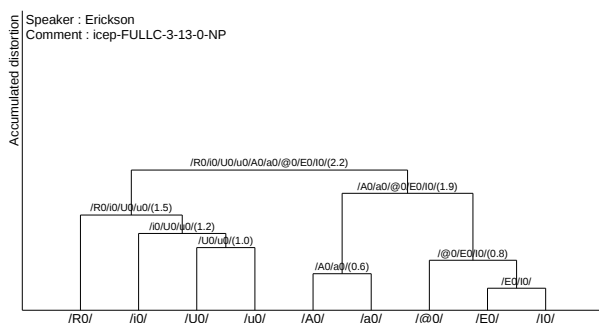


図 2. 弱勢母音の音素樹形図

構造のサイズについて見てみると、強勢母音での構造サイズは弱勢母音のサイズに比べて約 1.4 倍になっていることがわかる。全母音の場合と比較しても、強勢母音はより大きく、弱勢母音はより小さくなってい

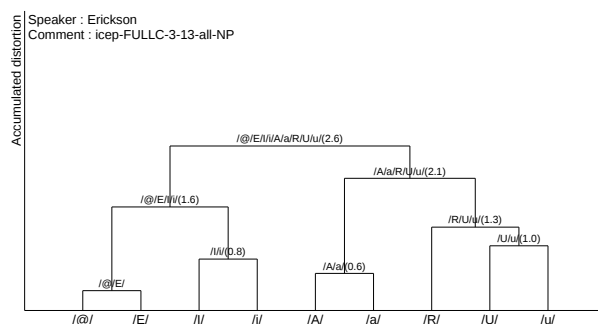


図 3. 全母音の音素樹形図

る。男性話者の場合でも同様の傾向が見られ、強勢サイズは弱勢サイズの約 1.2 倍という結果が得られた。また、樹形図の形について見てみると、強勢母音の樹形図は、全母音の場合と比べてサイズのみが大きくなっており形は変わらないことが確認できる。これは男性話者の場合でも同様であった。

以上の結果より、英語母音の強勢弱勢が音響的普遍構造のサイズとして表現されることが示された。母音の弱勢化が schwa 化であることを考えると、第 3 節に示したように、弱勢化は「より楽な母音発声」化であると解釈できる。言い換えれば、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のが強勢母音であり、不明瞭化したのが弱勢母音である。その意味において、音響的普遍構造のサイズは調音努力を指すと解釈できる。

5 まとめ

英語における強勢母音・弱勢母音に対して、音声の音響的普遍構造によって分析を行った。音響的普遍構造のサイズを分析した結果、強勢母音の場合は弱勢母音に比べて約 1.4 倍のサイズになっており、この構造のサイズが発話者の調音努力を表していることが確認できた。

今後は、種々の特徴パラメータや HMM の学習条件で構造の抽出を行い、構造抽出により適した分析条件を検討する必要があると思われる。

また、調音努力から一歩話を進め、構造のサイズを用いた非言語情報の推定に関しての実験的検討が行われている [4] ので、興味のある読者は参照していただきたい。

参考文献

- [1] 峯松: “音声に内在する音響的普遍構造とそれに基づく語学学習者モデリング”, 電子情報通信学会音声研究会資料, SP2003-179, 2004
- [2] 川越: 英語の音声科学, 大修館書店, 1996
- [3] 中川他: 音声・聴覚と神経回路網モデル, オーム社, 1990
- [4] 浜野他: “音声の音響的普遍構造に着目した非言語情報推定に関する実験的検討”, 日本音響学会春季講演論文集, 2004.