

7N-1 パルスエコー法による 舌像からの音韻抽出

岡田謙一, 横山光男, 北川 節, 山本高司
(慶応義塾大学理工学部計測工学科) (キャノン)

1. はじめに

最も重要な調音器官である舌は、その形状、動きが複雑で、正確な調音モデルを設定することはなかなか困難である。しかし、調音器官の有する構造的制約や対象言語の音韻体系により、発声時における各部の変形は相関性が強く、少数の主要な変動要素で表現することができ、これを調音パラメタとして処理することができると思われる。

本研究は、舌の調音モデルを設定し、そのパラメタを音声波からの推定ではなく、パルスエコー法による超音波診断装置を用いて、調音運動中から直接抽出しようとするものである。

2. 舌像の取り込みと前処理

超音波パルスを生体内部に向けて発射すると生体組織の不連続面で反射が起き、反射時間は距離を、反射波の強度は不連続面の媒質を反映する。そこで、超音波パルスの発射方向を一平面内に限り順に走査し反射波の強度を輝度変調すれば、走査面によって作られる断層面像を映像化することができる。いわゆる高速走査方式のパルスエコー像のBモード法を用いる。本実験で使用した超音波診断装置は、3.5MHzの機械式セクタ走査方式のプローブを持ち、視野角は90°である。このプローブを下顎の下から上向きにあてると、図1に示すような舌の正中断面像が実時間で得ら

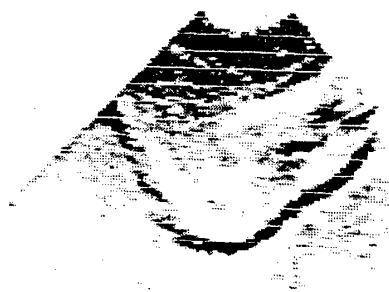


図1 超音波診断装置のエコー像

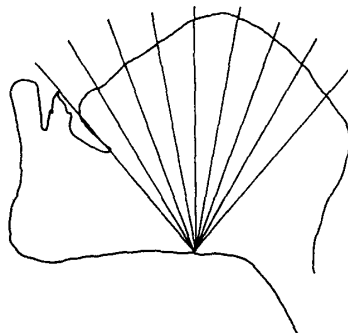


図2 舌モデルの座標系

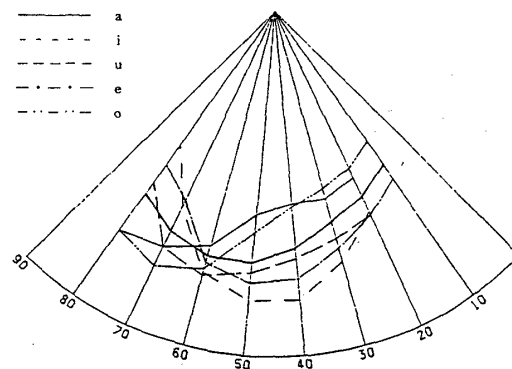


図3 母音調音時の舌面

れ、調音運動による舌の動きをほぼカバーすることができる。

エコー像には舌像と共に、プローブ表面のインシヤルエコー、下顎皮膚によるエコー、舌骨や舌の筋肉によるエコー、雑音などが存在するが、舌像は比較的高輝度であるためある閾値により2値化する。また舌像の動く範囲は限定されているものとし、画像にマスクをかけ余分なエコー像を取り除く。これらの処理を行ったエコー像に対し、ある定められた複数の始点から輪郭追跡を行い、これを舌像とする。

3. 調音パラメタの抽出

舌の形状を定量的に扱うために、断面像に図2で示すような極座標系を設定し、舌上面までの距離を要素とする9次元ベクトルとして舌形状を表す。座標系は下顎に固定され下顎の開閉にともなって回転するが、舌の中立的状態が下顎の開度によらず一定であるとすれば、舌の能動的な運動を表現できると考えられる。

調音パラメタ $X(x_0, \dots, x_8)$ には、いくつかの欠落部分が存在するが(図3参照)、この原因としては次のものが考えられる。(1)舌の前後方向の移動により、座標軸上に舌像が存在しない

Phoneme recognition by the tongue shape by the pulse-echo methods

Ken-ichi Okada¹, Teruo Yokoyama¹, Misao Kitagawa¹, Takashi Yamamoto²

1 KEIO University, 2 Canon, Ltd.

場合があり、主に舌先のデータが欠ける。(2)舌の表面と口蓋との接触により反射エコーが検出されない場合、主に舌の奥のデータが欠ける。(3)前処理により誤ってデータが捨てられる場合、主に舌の中央部が欠ける。(3)は補間により処理するが、舌先、舌奥のデータ欠落はそれ自身を特徴量と考え、舌先の欠落数 i 、舌奥の欠落数 j により (i, j) 欠落空間に属するデータとする。

4. 音韻情報の抽出

舌形状ベクトルの個々の要素は、舌の構造上の制約から互いに相関をもって変化する。そこで舌の形状ベクトルを、欠落空間毎に主成分分析し適当な主成分空間に投影することにより、調音状態に対応した領域の設定を行う。

舌の運動は、主に外舌筋と内舌筋の相互作用と下顎などの運動の影響により決定され、近似的には外舌筋による舌全体の上下および前後方向への移動と、内舌筋による舌尖部における形の変化の合成と考えられる。また、外舌筋の動作は主に母音の調音に、内舌筋の動作は主に子音の調音に関係する。そこで2つの筋群の運動を分析するために、データを母音部と子音部に分け、別々に主成分分析を行う。

これらの処理の結果、欠落空間毎に図4に示すような音韻に対応する領域の設定(領域のラベル付け)が試みられた。

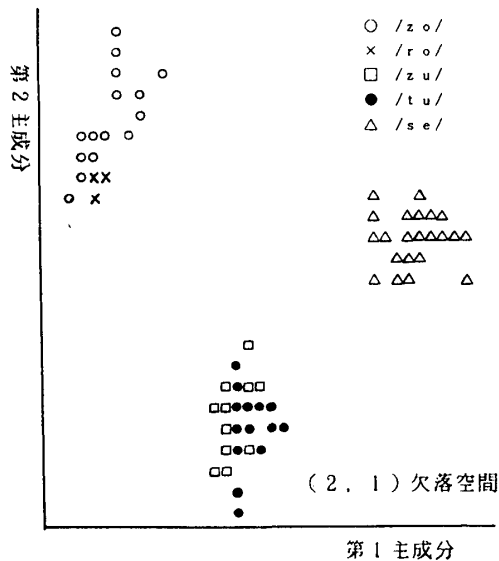


図4 主成分分析による音韻認識空間

5. 音韻の認識実験

前述の調音パラメタの有効性を調べるため、簡単な音韻認識実験を行った。舌が調音運動に直接関与する子音として /k/, /g/, /t/, /s/, /z/, /r/ の6つの音韻を選び、5母音と組み合わせた30単音について、成人男子1名の発音による30フレーム/秒、2171枚のエコー像を入力データとした。これらのデータから舌の形状ベクトルを抽出し、全データについて欠落空間毎に主成分分析を行い、この結果より形状ベクトルを時系列の順序に主成分空間へ投影し、対応する領域のラベルを認識結果とした(図5参照)。

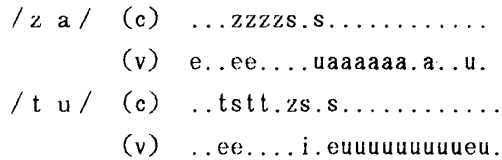


図5 認識実験の例

6. おわりに

まだサンプル数が少ない実験であるが、舌の調音モデルは、母音とある種の子音の音韻認識に有効であると思われる。特に破裂音では、破裂直前の無音区間に準定常的な状態が存在することが確認された。現在、対象の拡大、調音パラメタの洗練化を検討中である。

参考文献

- (1) 橋本, 谷本: 舌の側面輪郭の2次曲線による近似, 音響学会誌, vol.34, no.3, pp.10-20, 1978.
- (2) 石崎: 調音結合の動的処理について—子音に挟まれた母音への応用, 電子通信学会論文集, 1979.
- (3) 白井: 調音モデルの音声認識への応用, 音響学会誌, vol.37, no.5, pp.216-222, 1981.
- (4) 中島真人, 渡辺央: 新しい超音波診断技術, 南江堂, 1984