

母音発声時の声道形状における不変部位とその音響特性*

北村達也 本多清志 (ATR 人間情報科学研究所)

1. はじめに

音声中で個人性 (話者の特徴) を表す特徴量は、話者ごとに異なり、かつ音韻間で共通する成分であると定義することができる。もし、ある個人の声道において発声する音韻が変わっても形状が変化しない部位、すなわち不変部位が存在し、そのサイズや形状に個人差があるならば、そこが個人性の生成要因になっている可能性がある。そこで、本研究では声道における不変部位を探し出し、その音響的な影響を調査することを目的とする。

竹本ら^[1]は、3D MRI 動画撮像法を用いて、5 母音 /aiueo/ を連続発声する際の発話運動を記録している。彼らが示した声道断面積関数の時間パターンによると、声門から頭頂方向に約 25 mm の区間の断面積関数は、発話中にほとんど変化していない。

もしこの結果が一般的なものであれば、個人性の生成要因の 1 つである可能性がある。そこで、本研究では声門に近い部分に着目して不変部位を調査し、さらに有限要素法を用いてその音響的な影響を求める。

2. 撮像方法

2.1 撮像条件

本研究では静止画像を用いて分析を行う。撮像に用いた MRI 装置は、ATR 脳活動イメージングセンタに設置された島津 Marconi 社製 MAGNEX ECLIPSE 1.5T Power Drive 250 である。シーケンスは Fast Spin Echo 法、TE=11 ms、TR=3000 ms、FA=90 度。撮像領域は 256 × 256 mm (分解能 512 × 512 pixels)。スライスは矢状方向で、厚み 2 mm、間隔 2 mm、41 もしくは 51 スライス。加算回数は 1 回で、撮像時間は約 3 分である。

2.2 タスク

タスクは日本語 5 母音 /a/, /i/, /u/, /e/, /o/ の持続発声である。被験者は MRI 装置のガントリ内にて仰臥状態で発声した。

2.3 被験者

日本人成人男性 10 名について撮像を行った。しかし、撮像中の呼吸や唾液の嚥下などの影響で、喉頭部分の画像が乱れる被験者があった。そこで、喉頭部分も含めて画像の明瞭な 4 名のみを分析対象とした。

3. 横断面上における喉頭管の形状

正中矢状断面において声門の位置を決定し、その位置から頭頂方向 45 mm の面まで 5 mm ごとに横断面を抽出した。声門の位置は母音により上下するため母音ごとに決定した。そして、その MR 画像から視察により声道断面形状をトレースした。このようにして得られた声道断面形状のうち、1 名のものを図 1 に示す。

この図を見ると、声門から 30 mm 前後の区間において声道断面形状の概形が良く似ている。他の被験者でも同様の傾向が見られた。このことは、発声する母音が変わり、なおかつ声門が上下しても、こ

の区間の声道形状はあまり変化しないことを意味している。従って、この区間の声道形状の個人差が個人性生成に寄与している可能性がある。

声門付近に見られる左右の分岐管は梨状窩 (piriform fossa) である。梨状窩は音声の 4 kHz ~ 5 kHz 付近に零点を生じさせ、その周波数は発声中もほとんど変化しないことから、個人性への寄与が示唆されている^[2]。横断面の声道形状を見ても梨状窩の断面形状の変化は小さい。

4. 声道形状の類似度評価

図形の類似度の客観的な評価尺度の 1 つとして単純類似度法^[3]を導入し、類似度の評価を行った。

まず、前節の方法で抽出した横断面のトレースをもとにして、声道すなわち空気の部分が 1、それ以外が 0 の二値画像を作成する。次に、画像内での声道の位置を正規化し、Gaussian フィルタをかけ画像にぼけを加える。Gaussian フィルタの標準偏差は 0.5 mm とした。これは MR 画像における 1 pixel の 1 辺に対応する。そして、2 つの画像ベクトル g_1, g_2 の間の単純類似度 $S_s = (g_1, g_2)^2 / (\|g_1\|^2 \|g_2\|^2)$ を求める。ここで、 (g_1, g_1) は g_1 と g_1 の内積であり、 $\|g\|$ は g のノルム ($= \sqrt{(g, g)}$) である。単純類似度は $0 \leq S_s \leq 1$ の値をとる。

以上の計算を声門からの距離ごとに母音の全ての組合せ間で行い、平均値を求めた。その結果を図 2 に示す。この図において単純類似度が大きく減少する横断面は、被験者により異なり、声門から 30 mm ~ 40 mm の横断面である。従って、本研究で分析対象にした被験者に関して、単純類似度法にもとづく類似度評価を行うという条件下では、声門から 25 mm ~ 35 mm の区間を発声による声道断面形状の変化が少ない部位とみなすことができる。

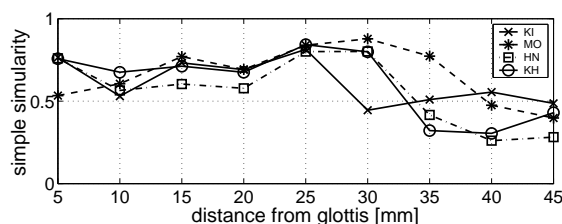


図 2. 声門からの距離と母音間の単純類似度との関係。

5. 梨状窩に起因する反共鳴の個人差

不変部位の音響的影響を明らかにするために、有限要素法 (FEM) を用いてシミュレーションを行った。特に、梨状窩による反共鳴について調査した。

5.1 有限要素モデル

声門から頭頂方向 70 mm までの区間の有限要素モデルを作成した。梨状窩の有るものと削除したものとを作成し、その影響を調査した^[4]。ただし、被験者の 1 人は喉頭管が非常に細いため、有限要素モデルを作成できなかった。

* An invariant section in the vocal tracts during vowel production and its acoustical feature.

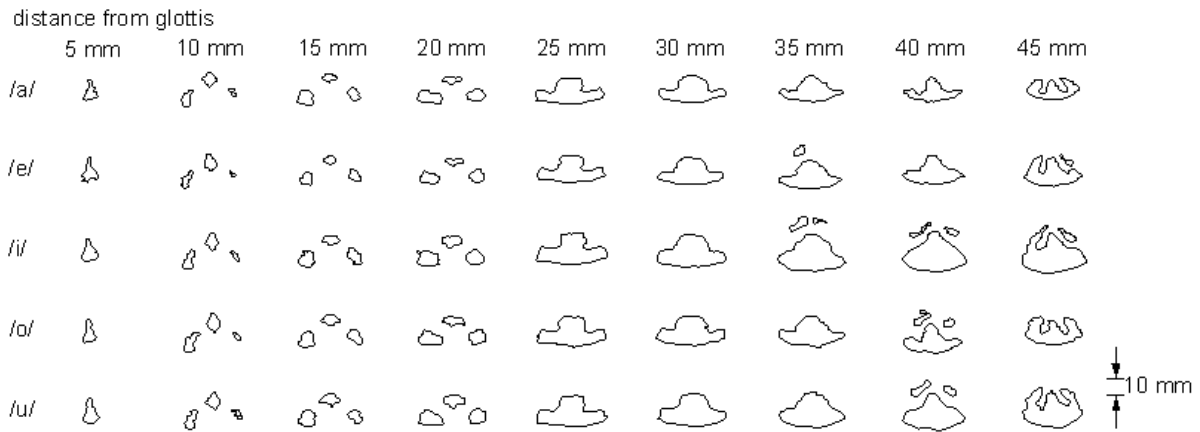


図 1. 被験者 KH の横断面の声道形状 .

まず、この区間の横断面画像をトレースした。用いた MR 画像は /i/ 発声時のものである。次に、これらの横断面画像から 3 次元モデルを作成し、Altair 社の HyperMesh を用いて有限要素モデルを作成した。有限要素は 1 次要素の 4 面体を用いた。有限要素モデルの一例を図 3 に示す。

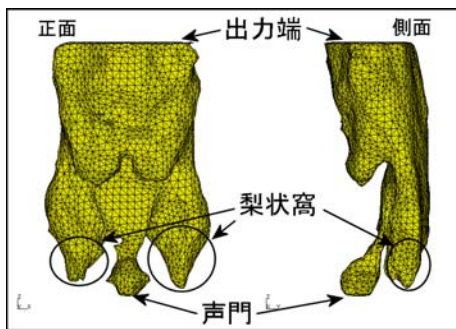


図 3. 有限要素モデルの例 .

5.2 シミュレーションの方法

FEM の解析プログラムとして LMS 社の SYS-NOISE を用いた。声門を入力端とし、正弦波振幅 1.0 m/sec の体積速度で励振させた。出力端は無限平面上の半径 a の円形ピストンとみなし、その放射インピーダンス $Z_r(\omega)$ を以下の式で与えた [5]。

$$Z_r(\omega) = \frac{j\omega L_r R_r}{R_r + j\omega L_r} \quad (1)$$

ここで、 $R_r = \frac{128}{9\pi^2}$ 、 $L_r = \frac{8a}{3\pi c}$ 、 c は音速である。

入力端と出力端以外の境界面は剛壁とした。この条件で音圧に関する波動方程式を解き、100 Hz から 5 kHz までを 20 Hz きざみで伝達関数を計算した。音速 c および空気密度 ρ はそれぞれ $c = 346.37 \text{ m/sec}$ 、 $\rho = 1.1733 \text{ Kg/m}^3$ を用いた。

5.3 実験結果と考察

FEM のシミュレーションにより求めた伝達関数を図 4 に示す。梨状窩が無いモデルの伝達関数には零点が見られないが、梨状窩が有るモデルの伝達関数には 4 kHz から 6 kHz の範囲に零点が生じることがわかる。

零点の周波数（反共鳴周波数）は被験者ごとに異なる。これは梨状窩の形状の個人差に起因するもの

である。被験者ごとの反共鳴周波数は、被験者 MO で 3900 Hz と 5100 Hz、被験者 HN で 5800 Hz、被験者 KH で 4100 Hz と 4780 Hz である。

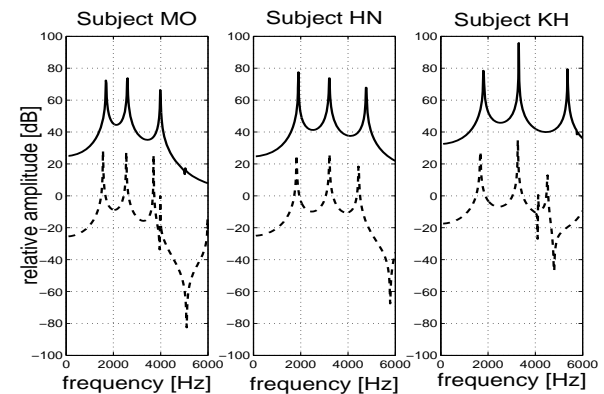


図 4. FEM により求めた伝達関数。実線：梨状窩無し。破線：梨状窩有り（左）被験者 MO（中）被験者 HN（右）被験者 KH。

6. おわりに

本研究では、声道における母音発声時の不変部位について調査した。被験者 4 名を対象にした調査の結果、声門から頭頂方向へ 25 mm ~ 35 mm の区間の断面形状は発声する母音が変わっても変化が少ないことを示した。さらに、この不変部位に含まれる梨状窩に起因する反共鳴周波数には個人差があることを示した。今後、個人性生成との関連をより詳細に調査する予定である。

謝辞 本研究は通信・放送機構の研究委託により実施したものである。

参考文献

- [1] 竹本ら, “3D MRI movie による声道断面積関数の時間パタンの抽出,” 信学技報 SP2001-24 (2001).
- [2] Dang et al., “Acoustic characteristics of the piriform fossa in models and humans,” JASA 101, 456-465 (1997).
- [3] 飯島, “パターン認識理論,” 森北出版, 東京 (1989).
- [4] 河西ら, “FEM モデルと実形モデルとの比較による梨状窩の音響特性の検討,” 信学技報 EA96-12 (1996).
- [5] Rabiner et al., “Digital processing of speech signals,” Prentice-Hall (1978).