

## 舌断面形状のモデリングに関する予備的検討\*

北村達也 (甲南大) 蒔苗久則 (科警研) 伊藤仁 (東北工大)

## 1 はじめに

発話時の調音動態の3次元観測は音声の生成過程の研究にとって重要であるが、現時点ではリアルタイムかつ高精度にその観測を実現する手法は見られない。一方で、2次元の調音動態をリアルタイムかつ高精度に観測する手法としては、MRI動画、超音波断層装置、磁気センサシステムなどが実用化されている。そこで、本研究では、正中面における調音器官の断面形状(2次元)から外側の断面形状を推定することによって、3次元の調音動態を得る手法を検討する。

本研究では舌を対象に検討を行う。舌は弾性体であるが、正中面の断面形状とその外側の断面形状の間にはある程度の対応関係があると期待できる。そこで、MRI発話同期撮像法[1]により連続母音発話時の3次元舌形状を観測し、その正中面の断面形状から線形モデルに基づき外側の断面形状を推定することを試みた。その過程で、Maeda[2]に倣って主成分分析により舌断面形状を表すパラメータを削減した。本稿ではこの検討結果について報告する。

## 2 方法

## 2.1 MRI データ

成人男性1名の連続日本語母音/aieuo/発話時の頭頸部の運動を発話同期撮像法[1]により撮像した。撮像に用いたMRI装置は、ATR-Promotions 脳活動イメージングセンタに設置されているSiemens社製MAGNETOM Verioである。正中面を含む9枚の矢状面を30fpsで撮像し、59フレームのデータを得た。スライス厚とスライス間隔はともに3.0mmとした。

その他の撮像パラメータは以下の通りである。Echo time (TE)=1.62 ms, Repetition time (TR)=33 ms, Flip angle=15°, Field of view (FOV)=256 mm×256 mm, Pixel size=1 mm×1 mm, Number of averages=1。撮像には約4分間を要した。

この撮像に関しては「甲南大学におけるヒトを対象にした研究審査」により承認されている。

得られたMRIデータのうち、正中面とその右外側の2断面を使用した。そして、第1フレームから母音/u/の中心に対応する第37フレームまでを後述する線形モデルの学習用データとし、残りの第38から

第59フレームまでを評価用データとした。

## 2.2 舌断面形状の抽出と平滑化

最初に、各MR画像の舌領域を目視により決定し、その境界を舌断面形状とした。次に、舌断面形状を200点で等間隔に再サンプリングし、それらの点の $xy$ 座標を以下のように複素数で表現した[4]。

$$\mathbf{f}_i = \mathbf{X}_i + j\mathbf{Y}_i \quad (1)$$

ここで、 $\mathbf{X}_i$ ,  $\mathbf{Y}_i$ はフレーム(時刻) $i$ における舌断面形状の $x$ 座標、 $y$ 座標を表す。 $j$ は虚数単位である。

$\mathbf{f}_i$ にDFTを施すことによりHarmonic amplitude profiles (HAP)[4]を求めた。その $m$ 次以上を0で置換し、IDFTを施すことにより $\mathbf{f}'_i$ を求めた。この一連の処理は、形状のローパスフィルタリングに相当し、平滑化された舌断面形状が得られる。本研究では $m=10$ とした。

## 2.3 舌断面形状のモデリング

顔画像の主成分分析[3]と同様の方法で $M$ フレーム分の舌断面形状を表現するパラメータの数を削減した。まず、 $\mathbf{f}_i$  ( $i=1, 2, \dots, M$ )とそのフレーム平均 $\bar{\mathbf{f}}$ との差分(差分形状ベクトル)を求める。

$$\Psi_i = \mathbf{f}_i - \bar{\mathbf{f}} \quad (2)$$

次に、 $\Psi_i$ の分散行列の固有ベクトル $\mathbf{a}_{ki}$ を求める( $k$ は固有値を表すインデックス)。そして、この $\mathbf{a}_{ki}$ と $\Psi_i$ の線形結合により固有形状ベクトル $\mathbf{u}_k$ を得る。

$$\mathbf{u}_k = \sum_{i=1}^N \mathbf{a}_{ki} \Psi_i \quad (3)$$

ここで、 $N < M$ とすることによって、 $M$ 個より少ないベクトルで差分形状ベクトルを代表させることができる。 $k$ は1から $N$ までの値をとる。

次に、上記のパラメータを用いて、ある舌断面形状 $\mathbf{f}_t$ を再合成することを考える。 $\mathbf{f}_t$ と $\bar{\mathbf{f}}$ との差分形状ベクトル $\Psi_t$ と上で得られた $\mathbf{u}_k$ を用いて記述パラメータ $r_k$ を求める( $\cdot$ は内積を表す)。

$$r_k = \Psi_t \cdot \frac{\mathbf{u}_k}{\|\mathbf{u}_k\|} \quad (4)$$

そして、 $N$ 個の $\mathbf{u}_k$ ,  $r_k$ を用いて $\mathbf{f}_t$ を再合成する。

$$\mathbf{f}'_t = \sum_{k=1}^N \left( r_k \frac{\mathbf{u}_k}{\|\mathbf{u}_k\|} \right) + \bar{\mathbf{f}} \quad (5)$$

\*Preliminary study of modeling of tongue cross-sectional shape. by KITAMURA, Tatsuya (Konan Univ.), MAKINAE, Hiranori (NRIPS), and ITO, Masashi (Tohoku Inst. Tech.)

以上により, ある舌断面形状は, 平均形状  $\bar{f}$ , 固有形状ベクトル  $\mathbf{u}_k$ , 記述パラメータ  $r_k$  で表されることがわかる. なお, 本研究では舌断面形状の重心の正規化や下顎の開閉に伴う位置の補正は行わなかった.

## 2.4 舌断面形状の推定

本研究では, 断面  $p_1$  の記述パラメータ  $r_{(p_1,k)}$  および舌断面形状の重心の加速度  $\Delta^2 c$  から断面  $p_2$  の  $r_{(p_2,k)}$  を推定することを考える. 舌断面形状の重心の加速度を用いているのは, 筋力の作用方向を考慮に入れるためである. 本研究では以下の線形モデルにより推定を試みた.

$$\mathbf{y} = B\mathbf{x} \quad (6)$$

$\mathbf{x}$ ,  $\mathbf{y}$ ,  $B$  は以下の式で与えられる.

$$\mathbf{x} = (1, r_{(p_1,1)}, r_{(p_1,2)}, \dots, r_{(p_1,N)}, \Delta^2 c)' \quad (7)$$

$$\mathbf{y} = (r_{(p_2,1)}, r_{(p_2,2)}, \dots, r_{(p_2,N)})' \quad (8)$$

$$B = \begin{pmatrix} b_{(1,0)} & b_{(1,1)} & \cdots & b_{(1,N+2)} \\ b_{(2,0)} & b_{(2,1)} & \cdots & b_{(2,N+2)} \\ \vdots & \vdots & \ddots & \vdots \\ b_{(N,0)} & b_{(N,1)} & \cdots & b_{(N,N+2)} \end{pmatrix} \quad (9)$$

行列  $B$  の各要素は, 学習用データから得られたベクトル  $\mathbf{x}$ ,  $\mathbf{y}$  から重回帰分析により決定した. 本研究では  $N = 8$  とした.

本研究では, 正中面の舌断面形状から外側の舌断面形状を推定することを試みた. まず対象とするフレームの正中面からベクトル  $\mathbf{x}$  を求める. このとき,  $r_{(p_1,k)}$  は学習用データの  $\bar{f}$ ,  $\mathbf{u}_k$  を用いて算出している. 次に, 式 (6) により外側の断面の  $r_{(p_2,k)}$  を推定し, 最後に式 (5) により外側の舌断面形状を得る.

## 3 結果と考察

以下では, 正中面の1枚外側の断面を  $p_{m+1}$ , さらにその外側の断面を  $p_{m+2}$  と表す. これらの2断面の母音 /i/, 母音 /e/ の区間の中心に対応する舌断面形状の推定結果を図1, 図2に示す. 母音 /i/, 母音 /e/ の区間の中心はそれぞれ第26, 第42フレームに対応する. 前者は学習用データ, 後者は評価用データに含まれる.

母音 /i/ においては, 断面  $p_{m+1}$ ,  $p_{m+2}$  とともに, 舌断面の原形状 (図中青線) と推定形状 (図中赤線) の間の差異は小さい. 一方, 母音 /e/ においては, 最大で数 mm のオーダーの推定誤差が見られる. この要因として, 母音 /aiu/ のデータから作成した固有形状ベクトル  $\mathbf{u}_k$  が, 母音 /e/ の舌形状の表現に適していなかったことが考えられる. 従って, 学習データを増やすことによって, 推定精度の向上が期待できる.

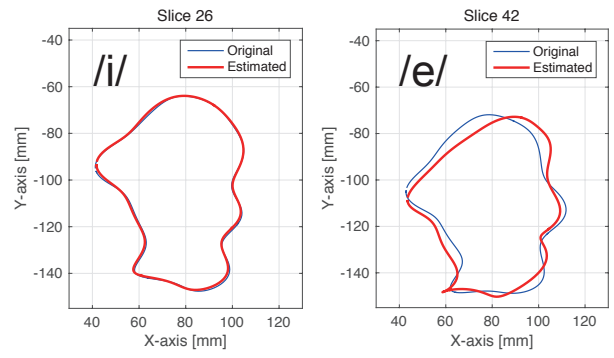


Fig. 1 Original and estimated tongue cross-sectional shapes for vowels /i/ (left) and /e/ (right) on plane  $p_{m+1}$ . Anterior side is left.

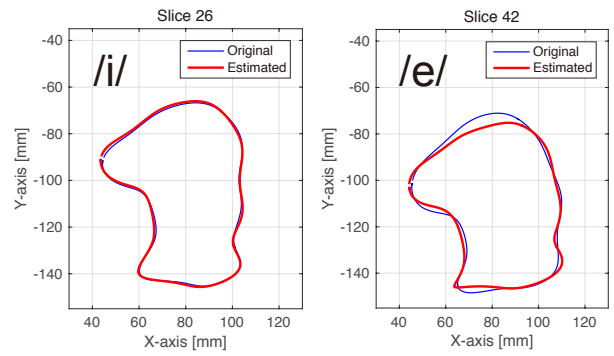


Fig. 2 Original and estimated tongue cross-sectional shapes for vowels /i/ (left) and /e/ (right) on plane  $p_{m+2}$ . Anterior side is left.

## 4 おわりに

本稿では, 3次元MRIにより得られた正中面の舌断面形状から外側の舌断面形状を線形モデルにより推定する手法の検討結果について報告した. 本研究では, 評価用データに対する推定精度は不十分であったが, 今後, 学習用データの追加, 非線形モデルの有効性の調査などを試みる.

**謝辞** 本研究の一部は平成27年度科研費(25280026, 25280066)の支援により行われた. 竹本浩典氏(NICT)のご助言に感謝する.

## 参考文献

- [1] Masaki *et al.*, *J. Acoust. Soc. Jpn.(E)*, 20 (1999).
- [2] Maeda, in *Speech production and speech modelling*, Hardcastle & Marchal (eds.), Kluwer (1990).
- [3] 永田ら, 信学論, J79-D-II (1996).
- [4] Park & Lee, *Comput. Biomed. Res.*, 20 (1987).