

## 連続母音発話 MRI データに基づく舌断面形状モデルの評価\*

○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 次元的観測は音声生成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れを観測する手法は見られない。一方で、2次元の調音動態であれば、リアルタイム MRI、超音波断層装置、磁気センサシステムなどによりリアルタイムかつ高精度に観測可能である [1]。従って、2次元調音動態から 3次元調音動態を推定できれば有益である。

そこで、前報 [2] 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手法を提案した。この手法では、まず正中面およびその外側の MRI 動画データから得られた舌断面形状を主成分分析によって少数のパラメータで表現する。そして、正中面の舌断面形状から外側の舌断面形状を推定する線形モデルを学習する。その線形モデルに正中面の舌断面形状を入力すれば、外側の舌断面形状が得られる。この手法は学習用データを用いた評価 (クローズドテスト) では良好な結果が得られたが、学習用データ以外を用いた評価 (オープンテスト) は十分に実施できなかった。そこで、本研究では後者も含めた評価を行う。

## 2 方法

## 2.1 MRI データ

20代の成人男性 1名の連続日本語母音 /aiueo/, /oẽuia/ 発話時の頭頸部の運動を発話同期撮像法 [3] 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 5枚の矢状面を 60 fps で撮像し、113 フレームのデータを得た。スライス厚は 5 mm、スライス間隔は 1 mm 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)=1.28 ms, Repetition time (TR)=857.3 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。

本研究では、得られた MRI データのうち正中面 ( $S_m$ ) とその左外側の断面 ( $S_{m+1}$ )、さらにその左外側の断面 ( $S_{m+2}$ ) の計 3 断面を使用した。

## 2.2 舌断面形状の抽出と平滑化 [2]

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  $xy$  座標を複素数で  $f_i = X_i + jY_i$  と表現した [5]。ここで、 $X_i$ ,  $Y_i$  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。

$f_i$  に DFT を施すことにより Harmonic amplitude profiles [5] を求めた。その  $m$  次以上 ( $m = 10$ ) を 0 で置換し、IDFT を施すことにより形状のローパスフィルタリングを行った。

以下では、/aiueo/ 系列, /oẽuia/ 系列の舌断面形状セットをそれぞれ  $D_t$ ,  $D_e$  と呼ぶ。 $D_t$ ,  $D_e$  はフレームごとに 3 断面の舌断面形状を含む。

## 2.3 舌断面形状のモデリングと推定モデル [2]

顔画像の主成分分析 [4] と同様の方法で舌断面形状を表現するパラメータの数を削減した。この処理により、ある舌断面形状は、平均形状、固有形状ベクトル、記述パラメータで表される。文献 [2] と同様に、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。

これらのパラメータを 3 断面 ( $S_m$ ,  $S_{m+1}$ ,  $S_{m+2}$ ) について求めた。そして、 $S_m$  の記述パラメータおよび舌断面形状の重心の加速度を入力として与えると外側の断面 ( $S_{m+1}$ ,  $S_{m+2}$ ) の記述パラメータを推定する線形モデルを求めた。得られた記述パラメータから舌断面形状を得ることができる。

## 2.4 評価実験

まず、 $D_t$  の平均形状、固有形状ベクトル、記述パラメータを求めた。次に、 $D_t$  の平均形状、固有形状ベクトルを用いて  $D_e$  の記述パラメータを計算した。そして、 $D_t$  のデータから上述の線形

\*Evaluation of a tongue cross-sectional shape model using MRI data of continuous vowel production. by KITAMURA, Tatsuya (Konan Univ.), MAKINAE, Hisanori (NRIPS), ITOH, Masashi (Tohoku Inst. Tech.)

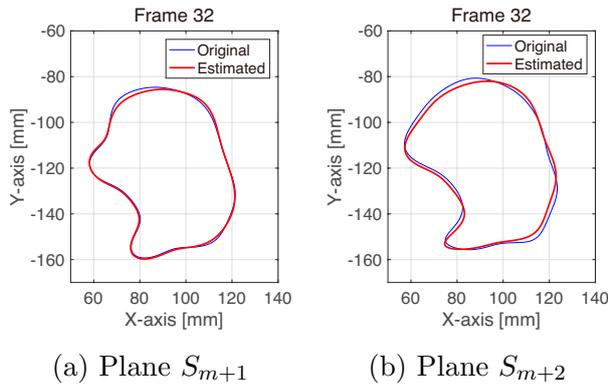


Fig. 1 Original and estimated tongue shapes on the planes  $S_{m+1}$  and  $S_{m+2}$  for the closed test.

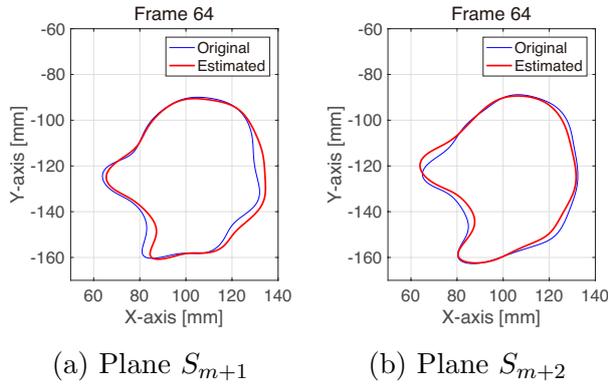


Fig. 2 Original and estimated tongue shapes on the planes  $S_{m+1}$  and  $S_{m+2}$  for the open test.

モデルのパラメータを決定した。クローズドテストでは、その線形モデルを用いて  $D_t$  の舌断面形状を推定し、オープンテストでは  $D_e$  の舌断面形状を推定した。

推定精度の評価は以下の式で得られる面積一致率  $r_a$  にて行った。

$$r_a = \frac{A_u}{A_t} \quad (1)$$

ここで、 $A_u$  は推定対象と推定結果の共通領域の面積、 $A_t$  は推定対象の舌断面形状の面積である。

### 3 結果と考察

クローズドテスト、オープンテストにおける推定結果の例を図1, 2に示す。前者では舌断面形状をほぼ正確に推定できている。後者でも概形は一致しているものの、図2 (a)の咽頭部、下顎部、図2 (b)の舌尖部などに局所的に3 mm程度の差異が見られる。

次に、クローズドテスト、オープンテストにおける断面  $S_{m+1}$ ,  $S_{m+2}$  の面積一致率  $r_a$  のフレー

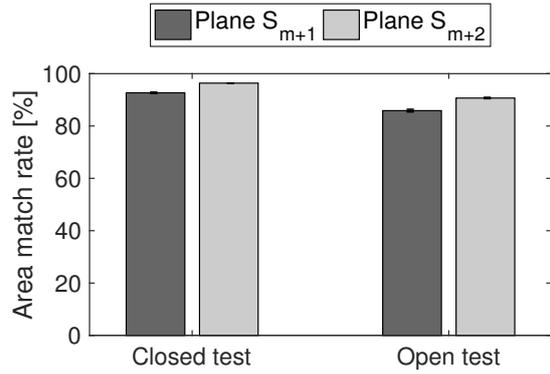


Fig. 3 Area match rates of the tongue model for the closed and open tests. The error bars represent the standard error.

ム間平均値および標準誤差を図3に示す。前者では  $r_a$  が約95%と高い一致率を示した。一方、後者では前者と比較して低下するものの、いずれの断面でも  $r_a$  が90%程度であった。この結果は、提案法により比較的高精度での推定が可能であることを示している。なお、上記の推定例における  $r_a$  は、図1 (a): 96.6%, 同 (b): 92.4%, 図2 (a): 89.8%, 同 (b): 91.3%であった。

本研究では全体的な誤差の把握のため面積による評価を行った。しかし、声道の変形がその音響特性に及ぼす影響は均一ではないため、推定誤差による音響的影響を考慮した評価法を検討する余地がある。

### 4 おわりに

本研究では、先に提案した舌断面形状モデル [2] の評価を行った。モデルの性能は学習用データの追加や統計モデル等の導入により向上すると期待できる。次の段階では、他の話者のデータを用いた評価を実施する。

謝辞 本研究は科研費 (25280066, 25280026) の支援により行われた。

### 参考文献

- [1] 北村, 技術情報協会 (2016).
- [2] 北村ら, 音講論 (春) (2016).
- [3] Masaki *et al.*, *J. Acoust. Soc. Jpn. (E)*, 20 (1999).
- [4] 永田ら, 信学論, J79-D-II (1996).
- [5] Park & Lee, *Comput. Biomed. Res.*, 20 (1987).