

## 日本語話者10名の正中面における母音声道形状の分析

☆天野沢海, 竹本浩典 (千葉工大)

北村達也 (甲南大), 能田由紀子, 前川喜久雄 (国語研)

## 1 はじめに

われわれはリアルタイム MRI 動画による日本語調音音声学のデータベースを構築してきた[1]。そして、動画の各フレームから機械学習を用いて半自動的に発話器官の輪郭を抽出する研究も行ってきた[2,3]。現在は20名の話者の動画から、フレームごとに発話器官5部位の輪郭を、トレースと同程度の精度で抽出することが可能である[4]。

各部位の輪郭は、話者間で解剖学的に相同な始点と終点の間を等しい数の点群として抽出している。したがって、点群を輪郭に沿って等間隔に再配置してセミランドマークとすれば、話者間やフレーム間で発話器官の形態を比較・分析することが可能となる[5]。

先行研究では、安静呼吸時の発話器官の形態を話者間で主成分分析した[5]。本研究では、安静呼吸時から母音発声時への各点の移動ベクトルを10名の話者間で主成分分析し、各母音の口の構えをとる運動の共通性と個人性を検討したので報告する。

## 2 材料と方法

## 2.1 輪郭データ

まず、データベースから標準語話者男女6名(M1, M2, M3, F1, F2, F3)と近畿方言話者4名(M4, M5, F4, F5)の安静呼吸時と母音/a/, /i/, /u/, /e/, /o/を発声しているフレームを選択した。次に、各フレームから輪郭を抽出した。そして、発話器官の5部位の輪郭を始点から終点まで等間隔の点群に変換した。なお、各部の輪郭は、舌を40点、上唇を15点、下唇・下顎を25点、軟・硬口蓋を30点、咽頭後壁を28点、喉頭蓋を30点で抽出した。よって、各フレームの輪郭は合計168点のxy座標で構成される。言い換えれば、 $168 \times 2 = 336$ 成分のベクトルである。

## 2.2 輪郭の分析

$m$ は話者を表す番号で $m = 1, 2, \dots, 10$ とする。

安静呼吸時の輪郭を表すベクトルを $r_m$ 、5母音発声時の輪郭を表すベクトルをそれぞれ $a_m, i_m, u_m, e_m, o_m$ とする。

安静呼吸時から各母音発声時の輪郭への移動ベクトルを $ra_m, ri_m, ru_m, re_m, ro_m$ とする(例: $ra_m = a_m - r_m$ )。ここで、例えば $ra_m$ を話者間で平均したベクトル $\bar{ra}$ は、安静呼吸時から母音/a/を発声するための発話器官の平均的な移動・変形を表している。すなわち、/a/の口の構えをとる運動の共通性である。よって、 $r_m$ と $a_m$ を話者間で平均したベクトルをそれぞれ $\bar{r}, \bar{a}$ とすると、 $\bar{a} = \bar{r} + \bar{ra}$ で表すことができる(図1)。

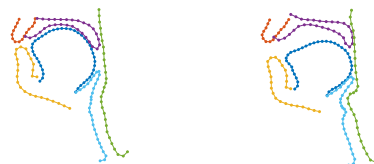


図1 左： $\bar{r}$ ，右： $\bar{r} + \bar{ra}$

また、 $r_m + \bar{ra}$ は、話者 $m$ の安静呼吸時の発話器官の形状を/a/の口の構えをとる運動の共通性によって変形させたものとなる[6]。しかし、口の構えをとる運動には発話器官の形状などに起因する個人性が含まれている。これは、話者間で $a_m, i_m, u_m, e_m, o_m$ を主成分分析することにより、個人差が大きな運動の成分として抽出することができる。

## 3 結果と考察

図2は、/a/の口の構えをとる運動の個人性の第1・第2主成分(PC1, PC2)である。赤と青はそれぞれの運動の $\pm 1$ 標準偏差( $\pm 1SD$ )を示す。PC1は主に舌背と口蓋との距離、PC2は主に口の開きと喉頭の上下位置に関する運動である。なお、他の母音でも、PC1, PC2のうちどちらかが舌背と口蓋との距離に関する運動であった。

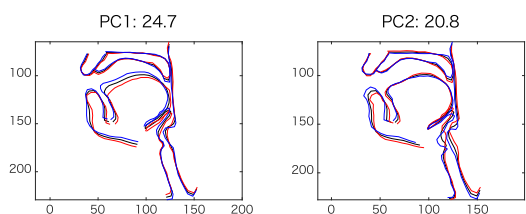


図2. /a/の口の構えをとる運動の個人性のPC1 (左)・PC2 (右)。黒: 平均, 赤: +1SD, 青: -1SD。上部の数値は寄与率。

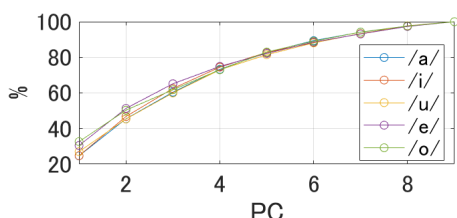


図3. 主成分の累積寄与率

図3は5母音の主成分の累積寄与率を示す。全ての母音でPC1が30%程度、PC2が20%程度と寄与率が低い。これは口の構えをとる運動の個人性にはばらつきが大きいことを示していると考えられる。よって、少数の主成分で個人差を表現することは困難である。

しかし、安静呼吸時の発話器官の形状を、運動の共通性のみによって変形させるよりも、個人性を加えて変形させた方が、実際に発声する際の発話器官の形状に近いことが予想される。そこで、以下の手順で定量的に評価した。まず、安静呼吸時の発話器官の形状を運動の共通性のみで変形させた形状を「共通性による形状」とする。次に、これにPC4までの個人性を表す成分を加えて変形させた形状を「共通性・個人性による形状」とする。さらに、実際に母音を発声した際の発話器官の形状を「実形状」とする。そして、「共通性による形状」と「実形状」との形態距離[7]から、「共通性・個人性による形状」と「実形状」との形態距離の差を求めた。この値が大きいほど、個人性を加えることで実形状に近づいたことを示す。なお、形態距離とは対応する点の距離の平均値で、単位はpixelである。

表1はその結果である。全て正の値であることから、個人性を加えることで発話器官の形状がより「実形状」に近づいたといえる。しかし、話者間で値に大きなばらつきが見られたことから、個人性を加える効果に個人差が大きいことを示す。

図4は、F1の母音/a/の「共通性による形状」、「共通性・個人性による形状」、「実形状」である。個人性を加えることで明確に「実形状」に近い形状が得られた。

表1 運動の個人性を付加する効果

	/a/	/i/	/u/	/e/	/o/
M1	2.81	1.74	1.60	1.73	3.19
M2	0.80	2.10	1.14	1.34	0.31
M3	0.67	1.19	0.30	1.03	0.37
M4	1.91	3.26	2.60	2.28	1.03
M5	0.84	1.86	1.32	2.32	1.29
F1	4.23	1.67	1.75	1.79	2.86
F2	4.07	0.78	2.70	1.45	3.85
F3	1.13	2.25	1.61	2.21	1.69
F4	2.40	3.46	3.83	4.88	3.67
F5	4.48	4.44	5.92	5.42	5.80



図4 F1の/a/の共通性による形状(赤), 共通性・個人性による形状(青), 実形状(黒)

#### 4 まとめ

本研究では、10名の話者の安静呼吸時と5母音発声時の発話器官の輪郭データを用いて各母音の口の構えをとる運動の共通性と個人性を分析した。そして、安静呼吸時の発話器官の形状を共通性だけで変形させるより、個人性も加えて変形させることでより実際の母音発声時に近い形状が得られることを示した。安静呼吸時の発話器官の形状の分析によって運動の個人性の主成分スコアを予測できれば、未知話者の安静呼吸時の発話器官の形状から発声中の発話器官の形状を予測することが可能になると思われる。

#### 謝辞

本研究はJSPS 科研費 20H01265の助成を受けた。

#### 参考文献

- [1] 前川ら, 音講論(春), 1247-1248, 2018.
- [2] Takemoto et al., Proc. of Interspeech 2019, 904-908, 2019.
- [3] 後藤ら, 音講論(春), 779-780, 2020.
- [4] 天野ら, 音講論(春), 745-746, 2021.
- [5] 後藤ら, 音講論(秋), 655-656, 2020.
- [6] 三輪ら, ISS-P-017, 74, 2018
- [7] 天野ら, 音講論(秋), 781-782, 2021