2段階モデルによるrtMRI動画からの発話器官の輪郭抽出

☆藤澤流以，△堀井千陽，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 調音運動データベース（rtMRIDB）」を構築中で，現在 22 名の話者のデータを試験公開している[1]。

発話器官各部の運動速度や接触タイミングなどを解析するためには，動画の各フレームから発話器官の輪郭を点群として抽出する必要がある[2]。先行研究では，18 名の話者から 8 話者を選出して発話器官の輪郭を学習させることで[3]，多話者に適用可能な汎用学習器を生成した[4]。しかし，汎用学習器では必ずしも正しく輪郭が抽出されない問題があった。

そこで本研究では，まず，学習させる話者の選出方法を再検討して汎用学習器を生成した。次に，公開中の話者から 10 名を選んで汎用学習器を適用し，抽出した輪郭を微調整して話者ごとの学習器を生成する 2 段階モデルを考案した。そして，この学習器を用いて 10 名のすべての動画（計 289,280 フレーム）から発話器官の輪郭を抽出したので報告する。

2　材料と方法

2.1 材料

日本人成人 18 名（男性 12 名：M1～M12，女性 6 名：F1～F6）の rtMRI 動画を用いて多話者から発話器官の輪郭を抽出できる汎用学習器を生成した。そして，データベースで公開中の話者から 10 名（F2，F3，F5，F6，F8，M1，M2，M4，M8，M14）を選び，汎用学習器で抽出した輪郭を微調整した個人ごとの学習器を生成して輪郭を抽出した。

rtMRI 動画は各話者につき約 50 本撮像された。各動画は 512 フレームで構成され，撮像速度は約 14 フレーム毎秒，空間解像度は 1×1×10 mm で，キャリア文「これは○○型」による 2 モーラ語の約 20 発話が記録されている。なお，撮像は Siemens 製 MAGNETOM Prisma fit 3（ATR-BAIC）で行った。

2.2 輪郭抽出の対象部位とトレース

輪郭抽出の対象とした部位は Fig.1 の舌(赤)，口唇・下顎(黄)，軟・硬口蓋(緑)，咽頭後壁(桃)，喉頭蓋・声帯(青)の 5 部位とした。また，学習器を生成するために，Table 1 に示す音素を含むフレームをトレースした。

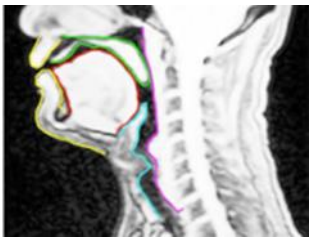

Fig. 1 抽出対象となる発話器官の 5 部位

Table 1 トレースしたフレームの音素

部位/使用枚数	各1枚	各2枚	各3枚	総枚数
口唇・下顎	/a/,/u/,/e/,/m/ 無発話,第1・512フレーム	/i/,/p/,/o/		13
舌	/a/,/e/,/k/,/t/,/r/ /n/,無発話,第1・512フレーム	/i/,/o/,/s/		15
軟・硬口蓋	/i/,/k/(u),/k/(e),/t/ /m/,無発話 ,第1・512フレーム	/a/		10
咽頭後壁	/g/,/m/,第1・512フレーム	/k/,/a/ 無発話		10
喉頭蓋・声帯	/i/,/e/,/o/,第1・512フレーム	/a/,/k/,無発話	/u/	14

2.3 輪郭の抽出精度の評価

輪郭の抽出精度は，評価する部位をトレースした者が目視で評価した。評価は 1 本の動画すべてのフレームを確認し，5 を評価者がトレースしたと同等，3 は解剖学的な知識を持たない者がトレースしたと同等として，5 段階で評価した。

2.4 汎用学習器に用いる話者の選出

汎用学習器を生成するためには，発話器官の部位ごとに様々な形態をバランスよく学習させる必要がある。18 名の話者で発話器官ごとに対して輪郭形状のクラスタ分析を行った。

Fig. 2 は，舌のクラスタ分析の例を示す。まず，デンドログラムの枝の高さに基づいて 8 つのグループに分割する。次に，各グループ

*Speech organ contour extraction from rtMRI using two-step model, by FUJISAWA, Rui, HORII, Chiharu, AMANO, Takumi, TAKEMOTO, Hironori (Chiba Institute of Technology), KITAMURA, Tatsuya (Konan Univ.), NOTA, Yukiko, MAEKAWA, Kikuo(NINJAL)。

からランダムに 1 話者を選出する。そして，これらの話者を用いて学習器を生成して試験的に輪郭抽出を行う。その結果，精度評価が 3 を下回る話者がいれば，同じグループの別の話者と置換して学習器を再生成する。この過程を平均 4 回繰り返すことで，全話者から学習器を部位ごとに生成した。

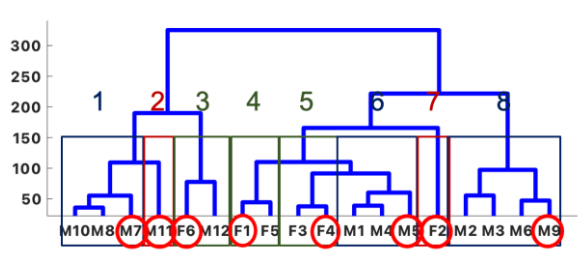

Fig. 2 舌のクラスタ分析の結果

2.5 学習器の個人化

データベースで公開中の 10 名に対して，汎用学習器で抽出した輪郭を微調整して個人ごと，部位ごとに学習器を生成した。この学習器を用いて抽出した輪郭の精度を評価した。

3 結果と考察

3.1 汎用学習器の輪郭抽出精度

Table 2 は汎用学習器による輪郭の抽出精度の評価結果である。－は学習器に含まれており，評価が高くなるため，評価していない話者と部位を示す。なお，F8 と M14 はどの部位も形態のクラスタ分析に含まれていない未知話者である。ほぼすべての話者と部位で評価値が 3 を上回ったが，未知話者である F8 の咽頭後壁，F8 と M14 の喉頭蓋・声帯で 3 を下回った。これは，今回生成した汎用学習器は必ずしも未知の話者から精度よく輪郭を抽出できるわけではないことを示している。

Table 2 汎用学習器の輪郭抽出結果

精度評価	F1	F2	F3	F4	F5	F6	M1	M2	M3	M4
口唇・下顎	3	-	4	3	-	-	5	5	-	4
舌	-	-	4	-	3	4	4	4	-	-
軟・硬口蓋	-	-	4	-	-	4	4	-	3	3
咽頭後壁	3	4	4	-	3	-	3	-	-	3
喉頭蓋・声帯	-	4	-	4	-	-	4	3	-	-
	M5	M6	M7	M8	M9	M10	M11	M12	F8	M14
口唇・下顎	-	5	-	4	-	5	4	-	4	4
舌	3	5	-	4	4	4	-	-	3	4
軟・硬口蓋	4	5	4	3	-	5	-	-	3	4
咽頭後壁	-	3	-	-	4	5	3	-	2	3
喉頭蓋・声帯	3	4	4	-	4	5	-	3	2	2

3.2 個人化した学習器の輪郭抽出精度

Table 3 は個人化した学習器の精度評価の結果で，括弧内の数値は汎用学習器からの向上の程度を示す。全ての話者の精度は 3 を上回り汎用学習器より精度が向上した。また，未知の話者に対しても評価は 3 を上回った。これは，個人化した学習器は未知の話者にも有効なことを示す。しかし，軟・硬口蓋，喉頭蓋・声帯が他の部位より精度が低い傾向が見られた。これは輪郭が不鮮明になりやすくトレース自体が困難であったことが原因と考えられる。なお，最初からトレース作業を行うことに比べて，汎用学習器が抽出した輪郭を微調整することで作業時間を約 50%削減し，効率的に学習器を生成することができた。

Table 3 個人化した学習器の精度

精度評価	F2	F3	F5	F6	F8	M1	M2	M4	M8	M14
口唇・下顎	5	4	5	4	4	5	5	4	4	4
舌	5	5(+1)	5(+2)	5(+2)	5(+2)	5(+1)	5(+1)	5	5(+1)	4
軟・硬口蓋	4	5(+1)	4	4	3	4	4	5(+2)	3	4
咽頭後壁	5(+1)	5(+1)	5(+2)	5	5(+3)	5(+2)	5	5(+2)	5	5
喉頭蓋・声帯	4	4	4	3	4(+2)	4	5(+1)	4	4	5(+3)

3.3 後処理

抽出した輪郭の点群を等間隔化し，部位間のオーバーラップを修正する後処理を行った。この輪郭をデータベースに登録して公開する。

4 まとめ

本研究では，まず 18 名の発話器官形状の分析結果に基づいて汎用学習器を生成した。次に，汎用学習器を用いて 10 名の rtMRI 動画から輪郭抽出を行った。そして，これを微調整して個人ごとに学習器を生成し，輪郭抽出を行った。目視による精度評価の結果，個人化した学習器は汎用学習器より高い精度で輪郭を抽出できることを確認し，学習器を 2 回に分けて生成する 2 段階モデルが有用であることを示した。なお，輪郭を最初からトレースするのではなく，汎用学習器で抽出した輪郭を微調整することで，作業時間を約 50%削減できた。

謝辞

本研究は JSPS 科研費 20H01265 の助成を受けて実施した。

参考文献

[1] 前川ら，rtMRIDB_v1, https://rtmridb.ninjal.ac.jp/.
[2] 後藤ら，音講論（春），779-780，2020.
[3] King,Mach, Learn. Res.，1755-1758, 2009.
[4] 天野ら，音講論（春），745-746，2021.