

☆常盤朔也, 竹本浩典 (千葉工大), 前川喜久雄 (国語研)

1 はじめに

われわれは調音音声学の精緻化のためにリアルタイム MRI (rtMRI) を用いて正中矢状断面における調音運動を記録した動画のデータベース (rtMRIDB) を構築し、公開している[1]。歯茎摩擦音などの分析をするためには歯列の位置が重要である。しかし rtMRI では、歯列は口腔と同様に輝度値が低く、画像上では黒く撮像されて区別できない。そこでわれわれは rtMRI 動画に歯列を補填する手法を提案してその精度を検討してきた[2]。この手法では、まず話者ごとに黒く撮像される歯列が白く撮像される舌や口唇などの軟組織で囲繞されて明瞭に確認できるフレーム (歯列抽出フレーム) を選択し、顎と歯列をトレースして抽出する。次に、歯列抽出フレームと歯列を補填する対象となる動画のフレーム (歯列補填フレーム) のそれぞれで顎と歯列を含む矩形領域を設定して両者を位置合わせする行列を MATLAB の `imregtform` 関数を用いて求める。最後に、この行列を用いて抽出した顎と歯列白黒反転させて補填する。補填精度は誤差面積率 (補填した歯列が周囲の軟組織にはみ出す割合) で評価した。その結果、精度は矩形領域の設定によって変化することが明らかになった。そこで本稿では、従来法を含めた複数の手法で矩形領域を設定し、精度を比較検討したので報告する。

2 材料と方法

2.1 材料

rtMRIDB に登録されている日本語標準話者の成人男性 4 名を用いた。話者と動画の本数は、s1:55 本, s4:53 本, s20:53 本, s21:58 本であった。各動画は 512 または 513 フレームで構成され、画像サイズは 256×256 ピクセル、空間解像度は $1 \times 1 \times 10$ mm である。また、撮像速度は約 14 フレーム毎秒の場合と約 27 フレーム毎秒の場合があった。なお、撮像は ATR-Promotions の Siemens 製 MAGNETON Prisma fit 3 で行った。

2.2 上顎の矩形領域 1 (提案法 1)

前報では、上顎での位置合わせに用いる矩形領域を鼻尖付近に設定しており (Fig. 1 (左)), 歯列補填の精度が低い場合には矩形領域を上下に 1.5 倍に拡大して補填をやり直した[2]。しかし、話者によっては拡大した矩形領域に口唇などフレームごとに形状が変化する部位が含まれてしまい、かえって精度が低下した。そこで、Fig. 1 (中) に示すように、矩形領域の下端は前鼻棘の高さ、右端は軟・硬口蓋のほぼ中央、上端は右端に沿って上行して前頭骨と脳との境界、左端は鼻尖よりやや前方として、唇を含まないように矩形領域を設定した。

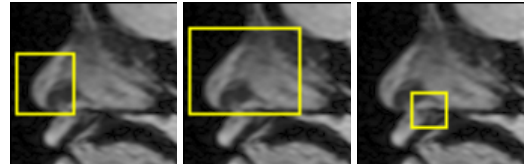


Fig. 1 上顎の矩形領域：従来法 (左), 矩形領域 1 (中), 矩形領域 2 (右)

2.3 上顎の矩形領域 2 (提案法 2)

矩形領域から口唇などフレームごとに形状が変化する部位を排除する別の方法は、フレームごとに特定しやすい参照点の周辺だけに矩形領域を縮小することである。前鼻棘は、rtMRI 動画から発話器官の輪郭を抽出する 2 段階抽出法[3]で精度よく特定できる。そこで、歯列を補填する全てのフレームで前鼻棘の座標を特定したのち、その周辺に Fig. 1 (右) で示す狭い矩形領域を設定した。

2.4 下顎の矩形領域 (提案法)

下顎の矩形領域から口唇や舌を排除するために、下顎骨の中心付近にある海綿骨の周囲のみに矩形領域を設定することとした。そこで、YOLOv8 [4]を用いて動画フレーム内から下顎の海綿骨を検出するモデルを作成した。なお、Table 1はこのモデルの学習に用いたフレームで、データセットの作成には `labelImg` [5]を用いた。作成したモデルを用いて、歯列補填フレームから下顎骨の海綿骨を抽出し、これを内包する矩形領域を設定した (Fig. 2)。

* Investigation of methods to improve the accuracy of tooth superimposition on rtMRI video frames, by TOKIWA, Sakuya, TAKEMOTO, Hironori (Chiba Institute of Technology), MAEKAWA, Kikuo (NINJAL)

Table 1 データセットに用いたフレーム

	各1枚	各2枚	総枚数
学習用データ	第1・最終フレーム	/a/,/i/,/u/,/e/,/o/,/k/, /s/,/t/,/r/,/m/,/p/,/N/, 無発話フレーム	28
検証用データ	/a/,/i/,/u/,/e/,/o/,/k/, /s/,/t/,/r/,/m/,/p/,/N/, 無発話フレーム		13

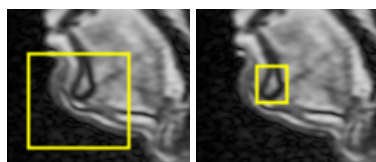


Fig. 2 下顎の矩形領域：
従来法（左），提案法（右）

2.5 補填法の改良

これまでの、フレームごとに補填した歯列が周囲の軟組織にはみ出したピクセルの割合である誤差面積率を計算するのみであった。しかし、本稿では誤差面積率が参考値を大幅に超えた場合、矩形領域の位置合わせに問題が発生したと判断し、矩形領域の高さと幅を1ピクセルずつ拡大して位置合わせと補填を再度実行することとした。

2.6 補填精度の評価

補填精度の評価は前報[2]と同様に誤差面積率と誤差面積率が参考値を超えたフレーム数の割合を表す参考値超過率で評価した。なお、参考値とは歯列抽出画像において、歯列を上下左右に1ピクセル平行移動するか±1度回転させたときの誤差面積率である。従来は画像の左上が回転中心であったが、本稿では重心位置とした。

3 結果と考察

Table 2 と Table 3 は従来法と提案法を用いて上顎と下顎の歯列を補填した場合の精度を話者ごとに示す。上顎でも下顎でも、提案法は従来法と比べて誤差の平均値と最大値、参考値超過率が低下した。この結果から、矩形領域に口唇など形状が変化する部位を含まないように設定することで補填の精度が向上したと考えられる。なお、上顎では提案法1に比べて提案法2の方が誤差の平均値は高く、最大値と参考値超過率は低い結果が得られたが、その要因は明らかにできなかった。

各手法を用いて1名の話者の全ての動画に歯列を補填にかかる目安時間は、従来法が2時間30分、上顎の提案法1が2時間30分、上顎の提案法2が8時間19分、下顎の提案

手法が3時間だった。なお、上顎の提案法2と下顎の提案法は一度モデルを作成すれば、補填にかかる時間は2時間ほどであった。

Table 2 従来法と提案法の
上顎の補填精度 (%)

	話者	s1	s4	s20	s21
	従来法	平均値	6.0	3.8	4.6
	最大値	35.1	27.3	28.2	25.0
	参考値	15.3	8.7	15.7	11.3
	参考値超過率	2.6	3.9	0.2	1.4
	話者	s1	s4	s20	s21
	提案法1	平均値	5.2	2.2	4.3
	最大値	30.1	16.6	16.4	12.9
	参考値	15.3	8.7	15.7	11.3
	参考値超過率	1.6	0.4	0.0	0.0
	話者	s1	s4	s20	s21
	提案法2	平均値	5.6	3.8	3.5
	最大値	28.9	10.4	15.3	11.7
	参考値	15.3	8.7	15.7	11.3
	参考値超過率	1.2	0.1	0.0	0.0

Table 3 従来法と提案法の
下顎の補填精度 (%)

	話者	s1	s4	s20	s21
	従来法	平均値	13.3	6.8	12.5
	最大値	33.9	21.1	37.1	25.1
	参考値	17.5	11.5	12.8	13.0
	参考値超過率	14.4	2.9	47.4	0.8
	話者	s1	s4	s20	s21
	提案法	平均値	8.0	5.5	2.4
	最大値	29.8	11.4	13.5	12.6
	参考値	17.5	11.5	12.8	13.0
	参考値超過率	0.8	0.0	0.0	0.0

4 まとめ

本研究ではrtMRI動画へ従来法を含めた複数の手法で矩形領域を設定し、日本人話者4名に歯列補填を行い、その精度を比較検討した。その結果、提案法は従来法と比べて誤差の各値が低下しており、精度を向上させることができた。このことからフレームごとに形状が変化する部位を含まないように矩形領域を設定することで、補填の精度を向上したと推定できる。

謝辞

本研究はJSPS科研費24K00071の支援を受けた。

参考文献

- [1] 前川ら, rtMRIDB_v1, <https://rtmridb.ninjal.ac.jp/>.
- [2] 常盤ら, 音講論 (秋), 961-962, 2023.
- [3] 藤澤ら, 音講論 (秋), 1015-1016, 2022.
- [4] Ultralytics, YOLOv8, <https://docs.ultralytics.com/>.
- [5] Tzutalin, labelImg, <https://github.com/tzutalin/labelImg/>.