

公募シンポジウム

公募シンポジウム7

歯科医療情報共有化と異分野融合によるデータ駆動型時代の歯科医療

2023年11月24日(金) 13:30 ~ 16:00 C会場 (EX1-B)

[3-C-3-05] 調音運動を記録したリアルタイム MRI動画への歯列補填

*竹本 浩典¹、常盤 朔也¹、前川 喜久雄²（1. 千葉工業大学、2. 国立国語研究所）

*Hironori Takemoto¹, Sakuya Tokiwa¹, Kikuo Maekawa²（1. Chiba Institute of Technology, 2. National Institute for Japanese Language and Linguistics）

キーワード：Tooth superimposition, real-time MRI, articulatory movements

われわれはリアルタイム MRIを用いて22名の話者の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）を公開している。これらの動画では歯列と口腔が同じ輝度値で黒となって区別がつかないため、歯茎摩擦音などの分析で問題となっていた。そこで、動画に歯列を補填する手法を開発し、精度を検討した。まず、話者ごとに1本の動画から上・下顎の歯列が舌や口唇などの軟組織で囲繞されて抽出可能なフレームを決定して歯列抽出フレームとした。そして、上・下顎の歯列をトレースして抽出し、それぞれ上・下顎歯列データとした。次に、歯列抽出フレームと歯列を補填する動画の第1フレームに上顎では鼻の先端、下顎では下顎骨下部を内包する矩形領域を設定した。そして、2つのフレームで対応する矩形領域を位置合わせする剛体変換行列を画像の強度に基づく勾配降下法で求めた。最後に、歯列データを剛体変換行列で写像することにより、第1フレームに歯列を補填した。そして、歯列抽出フレームの矩形領域の中心を補填する動画の第2フレームに剛体変換行列で写像し、この点を基準として第2フレームに矩形領域を設定した。この操作を繰り返すことにより、1本の動画に上下の歯列を補填した。この手法を用いて、現時点では22人中10人の話者の動画の計565本、289,304フレームに上・下顎の歯列補填を補填した。そして、その精度を補填した歯列が周囲の軟組織にはみ出す割合である誤差面積率で評価した。その結果、歯列抽出フレームで上・下顎の歯列を上下左右に1ピクセル平行移動あるいは±1度回転移動したときの最大の誤差面積率を基準とした場合、これを超過したフレームは上顎では2.6%、下顎では0.1%であった。よって、調音運動を記録したリアルタイム MRI動画に高い精度で歯列を補填することに成功したといえる。

調音運動を記録したリアルタイム MRI 動画への歯列補填

竹本浩典^{*1}、常盤朔也^{*1}、前川喜久雄^{*2}

*1 千葉工業大学、*2 国立国語研究所

Tooth superimposition on real-time MRI videos recoding articulatory movements

Hironori Takemoto^{*1}, Sakuya Tokiwa^{*1}, Kikuo Maekawa^{*2}

*1 Chiba Institute of Technology, *2 National Institute for Japanese Language and Linguistics

Real-time MRI can recode articulatory movements during speech to videos to help examine the movements quantitatively. However, consonants, including alveolar fricatives, in such videos are difficult to analyze since the teeth and oral cavity have the same low pixel value (black). We developed a method for superimposing the teeth on each frame of the videos in the midsagittal plane to overcome this limitation. First, a frame wherein the upper and lower jaws with teeth surrounded by the lips and tongue was selected and defined as the tooth-extraction frame. The upper and lower jaws with teeth were then extracted separately. Thereafter, two rectangular regions including the upper and lower jaws were set in the tooth-extraction and superimposed teeth target frames. Two rigid-body transformation matrices mapping the corresponding two rectangular regions between the tooth-extraction and target frames were then independently calculated using image registration. Finally, the upper and lower jaws with teeth were superimposed on the target frame using the two matrices. This method enabled the successful tooth superimposition of 565 videos including 289,304 frames from 10 speakers.

Keywords: Tooth superimposition, real-time MRI, articulatory movements.

1. 緒言

近年、MRI (Magnetic Resonance Imaging) を用いて実時間で任意の断面における体内運動を動画として記録する real-time MRI (rtMRI) の技術が発達した。そして、rtMRI をもちいた調音運動や歌唱技術の研究が行われるようになった¹⁻⁴⁾。そこでわれわれは、日本語の調音音声学を精緻化するために、rtMRI を用いて 22 名の話者の正中矢状断面における舌や喉頭などの調音運動を記録した動画のデータベース (rtMRIDB) を構築し、公開してきた⁵⁾。この動画を用いることにより、調音器官各部の運動タイミングなどを定量的に分析することが可能となった^{6,7)}。しかし、MRI 画像では歯列と口腔が共に低い輝度値 (黒) となり、画像上では区別できないため (図 1 左)、歯茎摩擦音などの分析で問題となっている。

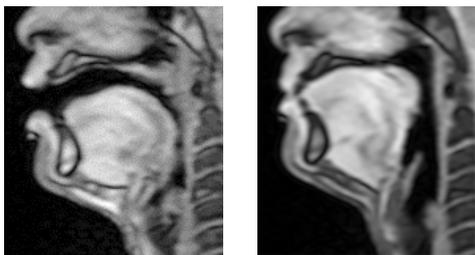


図 1 左: rtMRI 動画のフレーム、右: 歯列抽出フレーム

そこで、rtMRI 動画の各フレームに歯列を補填することができないか検討した。MRI 画像では、歯列や上・下顎の硬組織は黒く、唇や舌などの軟組織は白く撮像される。そのため、動画のフレームによっては、黒い上・下顎とその歯列が白い軟組織に囲まれて簡単に抽出することができる (図 1 右)。このフレームを歯列抽出フレームとする。他のフレームにおける上・下顎とその歯列の位置は、調音による頭部や下顎の運動によって歯列抽出フレームのものとは異なっている。しかし、これらの運動が矢状断面に平行であると仮定すれば、つまり

例えば口を開閉するとき下顎が右や左にずれなければ、歯列抽出フレームと歯列を補填したい他のフレームにおける上・下顎とその歯列の位置は、2 次元の剛体変換 (回転移動と平行移動) によって写像可能である。その剛体変換行列は、上・下顎とその歯列周辺を含む矩形領域を歯列抽出フレームと補填対象となるフレームとで位置合わせすれば得ることができる。この剛体変換行列を用いて、歯列抽出フレームから抽出した上・下顎とその歯列を白黒反転させ、補填対象となるフレームの上・下顎とその歯列の位置にオーバーレイすれば、そのフレームに歯列を補填できる。これを動画のすべてのフレームに適用すれば、rtMRI 動画に歯列を補填することができる。

2. 目的

本研究の目的は、調音運動を記録した rtMRI 動画の全てのフレームに歯列を補填する手法を提案し、その精度を検討することである。

3. 材料と方法

3.1 rtMRI 動画

rtMRIDB に登録されている話者 22 名から 10 名を選び、その動画の全てのフレームに歯列を補填した。話者と動画の本数は、s1:55 本、s4:53 本、s5:60 本、s7:60 本、s11:57 本、s12:58 本、s17:56 本、s18:59 本、s26:54 本、s27:53 本であった。各動画は 512 または 513 フレームで構成されており、補填対象となる動画は 565 本、合計 289,304 フレームであった。

動画にはキャリア文「これが〇〇型」による 2 モーラ語の約 20 発話などが記録されている⁵⁾。各動画の色深度は 16 bit、画像サイズは 256×256 ピクセル、空間解像度は 1×1 mm、スライス厚は 10 mm であった。また、撮像速度は約 14 フレーム毎秒または約 27 フレーム毎秒であった。なお、rtMRI 動画の撮像は ATR-Promotions に設置された Siemens 製 MAGNETOM Prisma fit 3 で行った。

3.2 歯列抽出フレームと歯列データ

話者ごとに 1 本の動画から、上・下顎およびその歯列が舌や口唇などの軟組織で囲まれて明瞭に観測されるフレームを決定し、歯列抽出フレームとした(図 1 右)。そして、上顎およびその歯列、下顎およびその歯列を半自動的に抽出し、白黒反転させて、それぞれ上顎歯列データ、下顎歯列データとした。なお、話者によっては、上顎と下顎で歯列抽出フレームが異なる場合もあった。

3.3 歯列補填

まず、下顎歯列データの補填の手順について述べる。

1. 歯列抽出フレームと歯列を補填する動画の第 1 フレームで下顎骨下部の緻密骨の中央にほぼ相同な点(基準点)をおき、その点を基準に下顎骨下部を内包する矩形領域を設定する(図 1 左・中)。
2. MATLAB の `imregtform` 関数による画像強度に基づく勾配降下法を用いて、歯列抽出フレームと歯列を補填する動画の第 1 フレームの矩形領域を位置合わせする剛体変換行列 M を求める。
3. 下顎歯列データを剛体変換行列 M で写像して第 1 フレームに歯列を補填する。
4. 歯列抽出フレームの基準点を剛体変換行列 M で移動した点を第 2 フレームの基準点とし、第 2 フレームの下顎骨を内包する矩形領域を設定する。
5. 上記 1~4 の手順を第 2 フレーム以下に順次適用して同じ動画の全てのフレームに下顎歯列データを補填する。

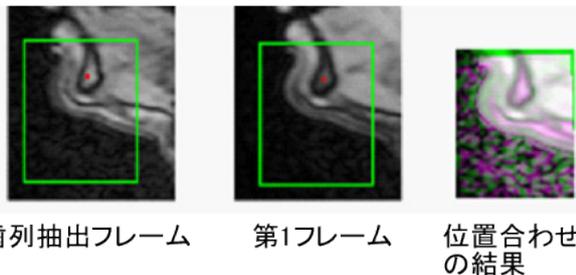


図 2 歯列抽出フレーム、第 1 フレームにおける下顎の参照点(赤)、矩形領域(緑)、矩形領域の位置合わせの結果(マゼンタ:歯列抽出フレーム、グリーン:第 1 フレーム)

上顎は頭部と一体になって運動するため、位置合わせのための基準点と矩形領域を鼻の先端付近に置いて、同様の処理を行った。

3.4 補填精度の評価

歯列を補填した精度を評価するために、先行研究⁸⁾と類似した方法を用いた。これは、補填した上・下顎歯列データを構成するピクセルがどの程度の割合で周囲の軟組織にはみ出したかを表す誤差面積率を用いる方法である。図 3 はその例である。図 3 左は上顎歯列データ、右は下顎歯列データを補填した例で、赤のピクセルは補填した上顎歯列データ、下顎歯列データのそれぞれ 18.4%と 12.7%であった。

しかし、誤差面積率だけでは、どの程度の回転移動・平行移動の誤差と対応するか判断できない。そこで本研究では、歯列抽出フレームで歯列データを上下左右に 1 ピクセル平行移動、あるいは±1 度回転移動させた際の誤差面積率を算出し、その最大値を誤差の参考値と定義する。そして、各話者で誤差面積率が参考値を超えたフレーム数を、歯列を補填した動画の総フレーム数で除した値を参考値超過率とし

て評価した。

なお、誤差面積率を計算するためには、歯列と周囲の軟組織を分離する閾値をフレームごとに設定する必要がある。これは、rtMRI 動画の特性として、後のフレームほど暗くなる、つまり平均の輝度値が低下するためである。そこで、歯列抽出フレームの平均輝度値が m_e 、歯列と周囲の軟組織を分離する閾値が t_e (この閾値より高ければ軟組織)、歯列を補填するフレームの平均輝度値が m_t とすると、そのフレームで歯列と周囲の軟組織を分離する閾値 t_t を $t_t = t_e + m_t - m_e$ で与えた。つまり、暗くなればそれだけ閾値も下げることにした。

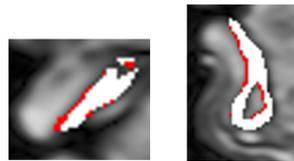


図 3 上・下顎歯列データを補填したフレームの例(左:上顎、右:下顎)。赤は周囲の軟組織にはみ出したピクセル。

4 結果

図 4 は s1、s4、s5 の動画に歯列を補填したフレームの例である。赤は周囲の軟組織にはみ出したピクセルである。いずれも目視では補填の結果は良好である。なお、上顎、下顎の誤差面積率はそれぞれ、s1 で 3.4%と 9.6%、s4 で 2.3%と 6.8%、s5 で 1.4%と 1.8%であった。

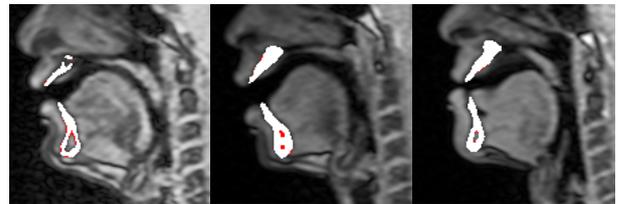


図 4 上・下顎歯列データを補填したフレームの例(左:s1、中:s4、右:s5)。赤は周囲の軟組織にはみ出したピクセル。

表 1、2 はそれぞれ 10 名の話者の全ての動画フレームに対する上顎・下顎の歯列補填の精度である。上段から、誤差面積率の平均値、最大値、参考値、歯列を補填した総フレーム数、参考値超過率である。

表 1 上顎の歯列補填の精度

話者	s1	s4	s5	s7	s11
平均値 (%)	5.9	3.8	2.5	9.2	5.1
最大値 (%)	30.4	27.3	16.5	50.4	26.2
参考値 (%)	15.7	8.9	10.3	18.5	11.9
総フレーム数	28,160	27,140	30,727	30,720	29,188
参考値超過率 (%)	2.0	3.3	1.4	4.0	4.7

話者	s12	s17	s18	s26	s27
平均値 (%)	9.8	3.4	10.7	4.4	10.1
最大値 (%)	33.5	40.9	65.8	26.2	40.9
参考値 (%)	21.8	13.5	21.9	14.0	26.2
総フレーム数	29,701	28,676	30,208	27,648	27,136
参考値超過率 (%)	2.9	2.8	3.6	0.8	0.7

上顎・下顎ともに、誤差面積率の平均値は参考値を下回り、参考値超過率は上顎では5%、下顎では0.5%を下回った。これは、全ての話者において上顎では95%以上のフレームで、下顎では99.5%以上のフレームで、位置合わせの誤差は平行移動・回転移動の成分としてそれぞれ1ピクセル程度・1度程度以内であることを示している。すなわち、定量的にも高い精度で歯列を補填できたといえる。

表2 下顎の歯列補填の精度

話者	s1	s4	s5	s7	s11
平均値 (%)	21.4	6.8	2.5	20.6	11.3
最大値 (%)	50.0	21.1	17.9	42.5	31.3
参考値 (%)	39.6	24.7	28.8	35.1	30.2
総フレーム数	28,160	27,140	30,727	30,720	29,188
参考値超過率 (%)	0.4	0.0	0.0	0.1	0.0

話者	s12	s17	s18	s26	s27
平均値 (%)	11.9	9.9	12.4	9.2	7.5
最大値 (%)	59.9	33.2	45.0	26.5	27.8
参考値 (%)	47.7	37.5	36.5	35.5	45.1
総フレーム数	29,701	28,676	30,208	27,648	27,136
参考値超過率 (%)	0.1	0.0	0.2	0.0	0.0

5 考察

下顎は上顎より発話中の動きが大きいかかわらず、上顎より参考値超過率が小さかった。しかし目視では、上顎と下顎で補填の精度に明確な差は見られなかった。そこで、その要因を検討した。

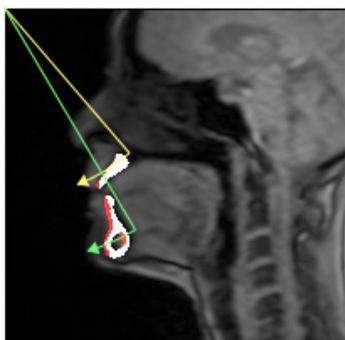


図5 s17の歯列抽出フレームで上・下顎歯列データをそれぞれ1度回転させたときに周囲の軟組織にはみ出したピクセル(赤)。回転中心は画像の左上。

誤差の参考値はどの話者でも上顎より下顎で大きかった。また、下顎の誤差の参考値は±1度の回転移動のいずれかであった。そこで、上顎と下顎を1度回転したときの誤差面積率を検討した。図5はs17の歯列抽出フレームで上・下顎歯列データをそれぞれ1度回転させた場合である。なお、MATLABでは画像の座標系の原点(回転中心)は左上である。黄色と緑色の線分はそれぞれ上・下顎歯列データのおよその半径を、矢印は回転方向を表す。上顎より下顎が下方にあるので、下顎の方が回転半径は大きく、移動量も大きい。また、上顎歯列データの長軸は全体として回転方向に沿った形状であるが、下顎歯列データの長軸は半径方向に沿った形状である。これらの要因により、1度の回転による誤差面積

率は上・下顎歯列データでそれぞれ9.9%と31.3%であり、下顎で著しく大きかった。そのため、目視では上顎と下顎で補填精度が同じ程度であったとしても、誤差の参考値が大きい下顎では参考値超過率が大幅に低下したと考えられる。

6 結論

本稿では、rtMRI動画の各フレームに歯列データを補填する手法を開発し、その精度を検証した。10名の話者の動画565本、合計289,304フレームに上下の歯列を補填した。参考値超過率に基づいて定量的に精度を検討した結果、全ての話者において上顎では95%以上、下顎では99.5%以上のフレームで、位置合わせの誤差は平行移動・回転移動の成分としてそれぞれ1ピクセル程度・1度程度以内であった。

しかし、目視では上顎と下顎の補填精度に明確な差はなかった。そこでその要因として誤差の参考値を検討したところ、回転移動の影響は上顎より下顎で大きく現れ、誤差の参考値が下顎で著しく大きくなったため、下顎では参考値超過率が大幅に低下したと考えられた。

なお、上顎でも下顎でも話者間で誤差面積率の平均値が大きくばらついた(表1, 2の平均値参照)。また、話者によって数は少ないが誤差面積率が大きなフレームがあった(表1, 2の最大値参照)。これらは、頭部や下顎の運動が矢状面に平行であるという仮定が必ずしも成り立たないためである可能性があるが、本研究では明らかにすることはできなかった。

謝辞

本研究はJSPS科研費20H01265の支援を受けた。

参考文献

- Engwall O. A revisit to the application of MRI to the analysis of speech production-testing our assumptions. In: *Proc of 6th International Seminar on Speech Production 2003*: 48.
- Narayanan S, Nayak K, Lee S, Sethy A, Byrd D. An approach to real-time magnetic resonance imaging for speech production. *J Acoust Soc Am* 2004; **115**: 1771-1776.
- Narayanan S, Toutios A, Ramanarayanan V, et al. Real-time magnetic resonance imaging and electromagnetic articulography database for speech production research (TC). *J Acoust Soc Am* 2014; **136**: 1307-1311.
- Echternach M, Birkholz P, Traser L et al. Articulation and vocal tract acoustics at soprano subject's high fundamental frequencies. *J Acoust Soc Am* 2015; **137**: 2586-2595.
- 前川 喜久雄. rtMRIDB.v1, 2022. [https://rtmridb.ninjal.ac.jp/ (cited 2023-Aug-30)].
- 天野沢海, 藤澤流以, 竹本浩典, 北村達也, 能田由紀子, 前川喜久雄. 日本語話者10名の正中面における/k/の声道形状の分析. 日本音響学会講演論文集(秋) 2022: 1013-1014.
- 藤澤流以, 天野沢海, 竹本浩典, 前川喜久雄. 日本語標準語話者の正中面における子音/s/の声道形状の分析. 日本音響学会講演論文集(春) 2023: 599-600.
- Takemoto H, Kitamura T, Nishimoto H, Honda K. A method of tooth superimposition on MRI data for accurate measurement of vocal tract shape and dimensions. *Acoust Sci & Tech* 2004; **25**: 468-474.