

超音波・MRI 調音データの輪郭自動抽出の試み—DeepLabCut を用いて

◎孫静（神戸大），北村達也（甲南大），林良子（神戸大）

1 はじめに

言語学習や発音矯正の研究において，舌の動きの解析は重要な要素である。特に，超音波と MRI による調音データは，舌の動きを詳細に捉えることができるため，広く利用されている。しかし，これらのデータから舌の輪郭を自動的に抽出することは技術的な課題が多く，手動での解析は時間と労力を要する。

従来の輪郭抽出手法（Edgetrak, GetContours, DeepEdge など）では，手動での輪郭描画が必要であったり，不明確な輪郭に対する精度が低下したりなどの制約があった [1][2][3]。機械学習ライブラリを用いた rtMRI 動画の輪郭抽出法[4]は高精度だが，使用法が難しい。

近年，機械学習技術の進展により，動画解析の自動化が進みつつあるが，超音波と rtMRI の両方を対象とした研究はいまだ限られている。本稿では，動物の行動解析のために開発された深層学習ベースのツールである DeepLabCut ver. 2.3.10 [5]を用いて，これらの動画から輪郭抽出を試みた結果を報告する。

2 DeepLabCut の概要

DeepLabCut (DLC) は，動物の行動解析のために開発された深層学習ベースのツールであり，その高い汎用性から人間の舌の動き解析にも応用されている[5]。DLC は，ユーザー定義の体部位のポーズ推定を行うために，畳み込みニューラルネットワーク（CNN）を用いて画像や動画中の特定のポイントを高精度に検出する。DLC の主な目的は，マーカーなしで動物や人間の行動を解析することである。

DLC は，ラベル付けした動画を使って CNN をトレーニングし，その後，新しいデータに対してモデルが自動的にラベルを付ける。これにより，高精度な動作解析が可能となる。DLC は GUI（Graphical User Interface）を提供しており，ユーザーは直感的に操作することができる。GUI によりラベル付けや学習の作業が容易になり，使用しやすいツールである。

3 方法

3.1 実験用データ

調音波ゼリーを塗布したマイクロコンベックスプローブ（8C-Rs）を実験参加者の下顎下にあて，超音波断層撮像装置（LOGIQ e Premium）によって正中矢状断面における舌の運動を B モードで撮像した（68 fps）。プローブは UltraFit ヘッドセットにて固定し，発話時にはなるべく頭を動かさないように指示した。rtMRI データとして，本研究では『リアルタイム MRI 日本語調音運動データベース』[6]を使用した。

3.2 輪郭抽出手法

DLC を用いて超音波動画と rtMRI 動画から舌や他の調音器官の輪郭の自動抽出を行った。学習データのトレースにおいては，まず，超音波動画や rtMRI 動画から，学習に適したキーフレームを選定する。キーフレームは，舌の位置や極端な形状をとるフレームを選ぶ [7]。次に，選定したキーフレームに対して，DLC を用いて舌などの輪郭や位置を手動でラベル付けする。動画の最適なサイズは 320 pixel×240 pixel で，このサイズを超えた動画を解析する場合，輪郭の自動抽出でエラーが出る。

4 実験結果

実験で得られた超音波と rtMRI データから DLC により舌の輪郭を抽出した。Fig. 1, Fig. 2 に，それぞれ DLC のデフォルトのモデルと筆者がトレーニングしたモデルで抽出した舌の輪郭で，Fig. 3 は筆者がトレーニングした MRI 用のモデルで抽出したものである。Fig. 1, Fig. 2 の左側は舌根，右側は舌尖の位置を示す。モデルの評価結果は Table 1 に示している。

超音波データにおいて訓練誤差は 3.21 pixel，テスト誤差は 3.34 pixel であった。p-cutoff を適用すると，訓練誤差は 3.04 pixel，テスト誤差は 3.22 pixel に減少した。p-cutoff

* Attempts to automatic extraction of contours of articulatory organs from ultrasound and MRI data using DeepLabCut, by SUN, Jing (Kobe Univ.), Kitamura Tatsuya (Konan Univ.), HAYASHI, Ryoko (Kobe Univ.).

は、予測ポイントの信頼度が一定の閾値（カ

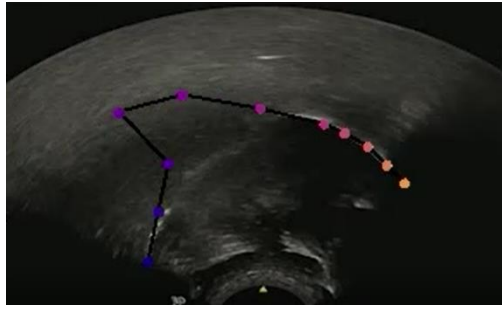


Fig. 1 Contours of the tongue extracted using the default model of DeepLabCut (DLC)

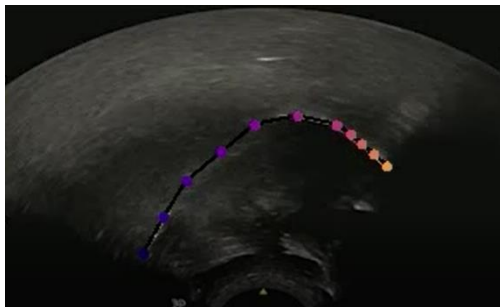


Fig. 2 Contours of the tongue extracted using the retrained model

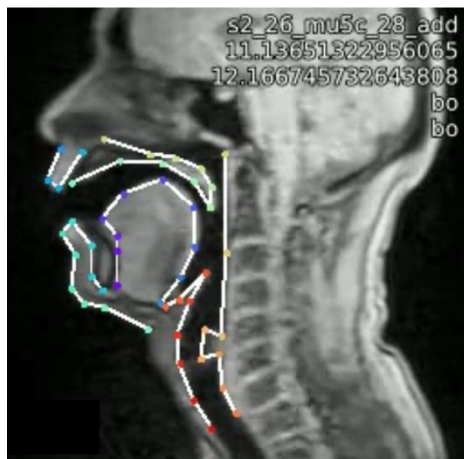


Fig. 3 Contours of speech organs extracted from rtMRI image

Table1 Evaluation results

	超音波	rtMRI
Training iterations	1030000	1030000
Train error (px)	3.21	1.71
Test error (px)	3.34	3.24
p-cutoff used	0.60	0.60
Train error with p-cutoff	3.04	1.71
Test error with p-cutoff	3.22	2.60

ットオフ値)を超えた場合にのみ、そのポイントを有効とする基準である。信頼度スコアは0から1の範囲で示され、値が高いほど予測の信頼性が高いことを意味する。p-cutoffを適用することで、信頼度が低い予測ポイントを除外し、予測の精度と信頼性が向上する。例えば、p-cutoffを0.6に設定すると、信頼度スコアが0.6以上のポイントのみが有効となり、それ以下のポイントは無視される。

rtMRI データの場合、訓練誤差は 1.71 pixel、テスト誤差は 3.24 pixel であり、p-cutoff を適用するとテスト誤差は 2.6 pixel にまで減少した。これらの結果から、DLC モデルが大量のデータで訓練された場合、非常に高い精度で舌の輪郭を抽出できることが示された。

5 まとめと考察

本研究では、DeepLabCut (DLC) を用いて超音波および rtMRI データから調音器官の輪郭を自動抽出する手法を検討した。結果として、DLC モデルを大量のデータで訓練すると、非常に高い精度で舌の輪郭を抽出できることが示された。DLC モデルの精度を高めるためには、数百枚の手動ラベル付き画像を使用することが推奨される。これらの成果は、音声生成のメカニズム理解や発音矯正において重要な貢献をもたらす可能性がある。また、深層学習技術の有効性が確認されたことから、他の応用分野への展開が期待される。

謝辞

本研究の一部は、JSPS 科研費 (Nos. 20H05630, 21H04348, 24K00071, 24K00067) の研究成果である。

参考文献

- [1] Li et al., *Clinical Linguistics and Phonetics*, 19(6-7), 545-554, 2005.
- [2] Tiede, GetContours Ver. 3.5, 2021. <https://github.com/mktiede/GetContours>
- [3] Chen et al., *Proc. ISSP*, 2020.
- [4] Takemoto et al., *Proc. INTERSPEECH 2019*, 904-908, 2019.
- [5] Mathis et al., *Nature Neuroscience*, 21(9), 1281-1289, 2018.
- [6] リアルタイム MRI 日本語調音運動データベース 第2版, 2024. <https://rtmridb.ninjal.ac.jp/>
- [7] 後藤ら, 音講論 (秋), 757-758, 2019.