リアルタイム MRI 動画を利用した パラ言語発話生成時調音運動の動体検出 *

○浅井拓也, 菊池英明 (早大), 前川喜久雄 (国語研)

1 はじめに

藤崎の分類によれば音声に含まれる情報は, 言語情報, パラ言語情報, 非言語情報の三種に分類される [1]. このうちパラ言語情報は, 話者が意図的に表出する情報でありながら文字には転写されない情報であり, 話し言葉の本質に深く関係している [2]. また, パラ言語情報は通常は言語情報と同時に生成, 伝達されるものであるため, 音声生成の側面から, パラ言語情報が言語情報の生成をどのように, 強調, あるいは妨げるのかが論じられてきた. 例えば前川らは音声発話時の調音運動を磁気計測技術 (EMA) を使用し観察し, 疑い発話時には感心発話時と比較して, 舌が声道の前寄りになっていることなどを報告している [3]. また Erickson らは, 英語のパラ言語発話においても, 前川らが報告した現象が確認されたことを報告し, 加えて怒りを表現する際に, 下顎の運動が大きくなることを示している [4].

これらの知見は非常に貴重なものであるが, 磁気計測技術は観測点が有限であり, 調音運動を網羅的に観察することは困難であるという問題を抱えていた. 一方で近年ではリアルタイム MRI 撮像技術 (以下 rtMRI) の発達によって, 調音運動の客観的な計測データを比較的容易に収集可能になりつつある [5]. この技術は音声発話時の正中断面上の運動を網羅的に観察できるものの, 各調音点の推定は容易ではなく, 解剖学的な知見に基づいた運動の観察はむずかしい.

我々のプロジェクトでは rtMRI 画像における各調音器官輪郭を教師あり学習に基づいて検出する試みも行っている [6] が, 本研究では画像を対象に調音器官を抽出するのではなく, 動画に含まれる運動そのものを直接的に解析することを目標とする. 本研究では動画解析の分野で一般的な動体検知技術である背景差分法を利用し, rtMRI 動画からの動体検出および, 検出された動体の観測を試みる.

2 方法

2.1 収録データ

我々は 2017 年度から 3 年計画で JSPS 科研費の補助をうけて日本語音声の rtMRI 動画データベースを構築中であり, 研究終了後には一般公開を予定している. このデータベースには男性 9 名女性 4 名分のパラ言語発話が含まれており, 中立, 疑い, 感心, 落胆の 4 種類のパラ言語表現を /山田さんが/ という文と共にそれぞれ 2 回以上発話している. 本研究には rtMRI 動画データベースに含まれるパラ言語発話の内, 比較的初期に収録され, 解析準備の整っている男女各 3 名を対象に解析を行った.

2.2 背景差分法

本研究では, 背景差分法の中でも, フレーム間差分法 (frame subtraction method) と呼ばれる手法を利用した. この手法は動画解析の領域において, 異なる時間に撮影した複数枚の画像から背景画像または移動物体の領域を直接取り出す場合によく使用される方法である [7]. 以下にフレーム間差分法による動体検出アルゴリズムの概略を示す.

まず時間的に連続する三枚の画像 I_1, I_2, I_3 があると仮定する. ここで, 画像 I_1, I_2 および I_2, I_3 に関して, それぞれの差分の絶対値を計算し, 差分画像 (I_{d1}, I_{d2}) を作成する.

$$\begin{aligned} I_{d1}(x,y) &= |I_1(x,y) - I_2(x,y)| \\ I_{d2}(x,y) &= |I_2(x,y) - I_3(x,y)| \end{aligned} \quad (1)$$

続いて, 差分画像 (I_{d1}, I_{d2}) の論理積を計算し, 論理積画像 I_a を生成する.

$$I_a(x,y) = I_{d1}(x,y) \wedge I_{d1}(x,y) \quad (2)$$

ここで作成された論理積画像 I_a を二値化処理することにより, 背景および前景を分離するマスク画像 I_m を得ることができる.

---

*Motion detection of articulatory movement with para-linguistic expression using real-time MRI animation.
by ASAI, Takuya, KIKUCHI, Hideaki (Waseda University) and MAEKAWA, Kikuo (NINJAL)

$$I_m(x,y) = \begin{cases} 255 & (I_a(x,y) > T) \\ 0 & (I_a(x,y) \leq T) \end{cases} \quad (3)$$

なお, rtMRI 動画は一般的な動画と比較すると, 多くのノイズを持ち, 輝度も高くはない. そのため, 本研究では, フレーム間差分法による動体検出の前処理として, Non-local Means Filter によるノイズ低減処理 [8] および, ガンマ補正による輝度の調整を行っている. これらの信号処理の実装に関しては画像処理の強力なライブラリ集である OpenCV [9] を使用した.

3 結果

3.1 背景差分法による発話領域抽出

背景差分法による rtMRI 画像からの動体検出例を図 1 に示す. 図 1a は検出された動体を可視化したものであり, 図 1b は検出された背景画像である. 各図では, 各画像フレームの差分値の総論理和をグレイスケールに変換し表現をしている. なお, 測定された動体画像を一定の閾値を定め輪郭抽出を行った結果を図中, 濃白線に示す. この箇所は, 発話中特に活発な運動が生じた領域であると解釈できる.

図 1a を確認すると, 舌の前後運動や, それに伴うと考えられる奥舌の上下運動が観察される. また, 下顎や唇, 咽頭付近に動体が検出されていることが分かる. これらの領域は, 従来の調音音声学において注目されてきた領域に一致するものであり, 背景差分法による動体検知が音声発話生成時の調音運動の特徴を捉えていることを示唆する結果である.

3.2 個人差の検討

続いて, 検出された動体の個人差を観察した. 図 2 に我々のデータの内, 男女 3 名ずつの動体検出結果を示す.

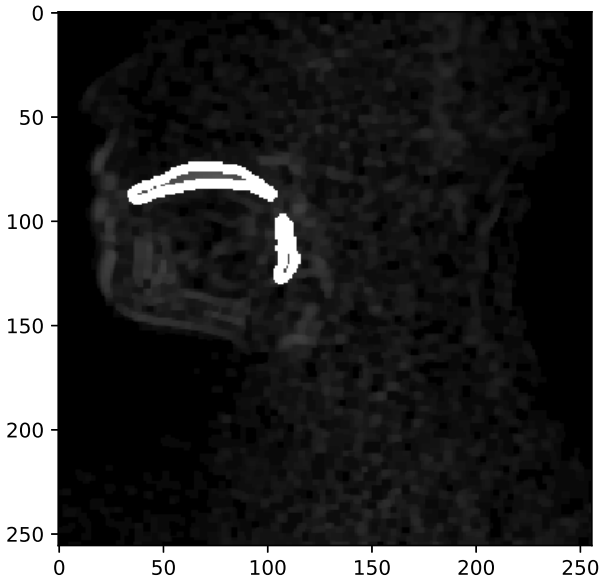
図 2 を確認すると, 全ての発話者において, 舌運動に当たる動体を検出することに成功した. 特に図 2a, 2b に関しては, 舌の前後運動や, それに伴うと考えられる奥舌の上下運動が観察された. 一方, 図 2c, 2e に関しては, 舌端の動きが特に強調されて表現された. 図 2f および 2d に関しては, 奥舌部分の運動が顕著に観察された.

また, 下顎の運動も, 全ての図において観察された. 特に図 2d においては他の発話者に比べ下顎の動きが顕著に観察された.

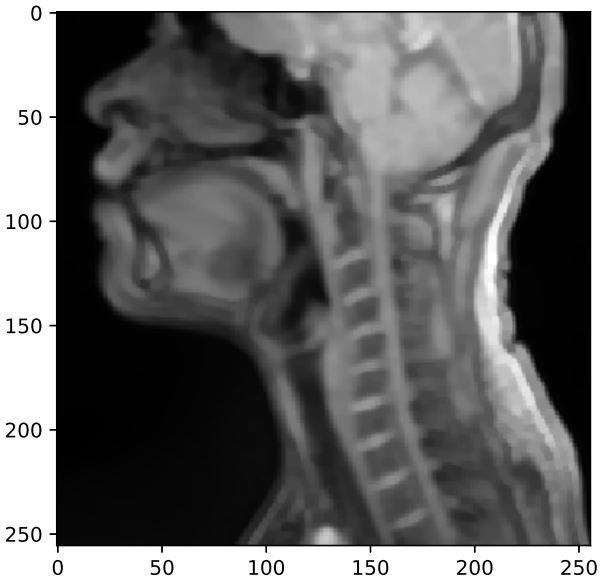
それぞれの発話者全体を比較すると, 上記の様に多少の差はあるものの, 同様の箇所に動体が検出されることが確認された.

3.3 パラ言語発話時の調音運動

最後にパラ言語初和別の調音運動の特性について述べる. ここでは上記, 個人差の解析結果を考慮して, 特に調音運動の特性の似ている図 2a, 2b の発話者を中心に議論を行う.

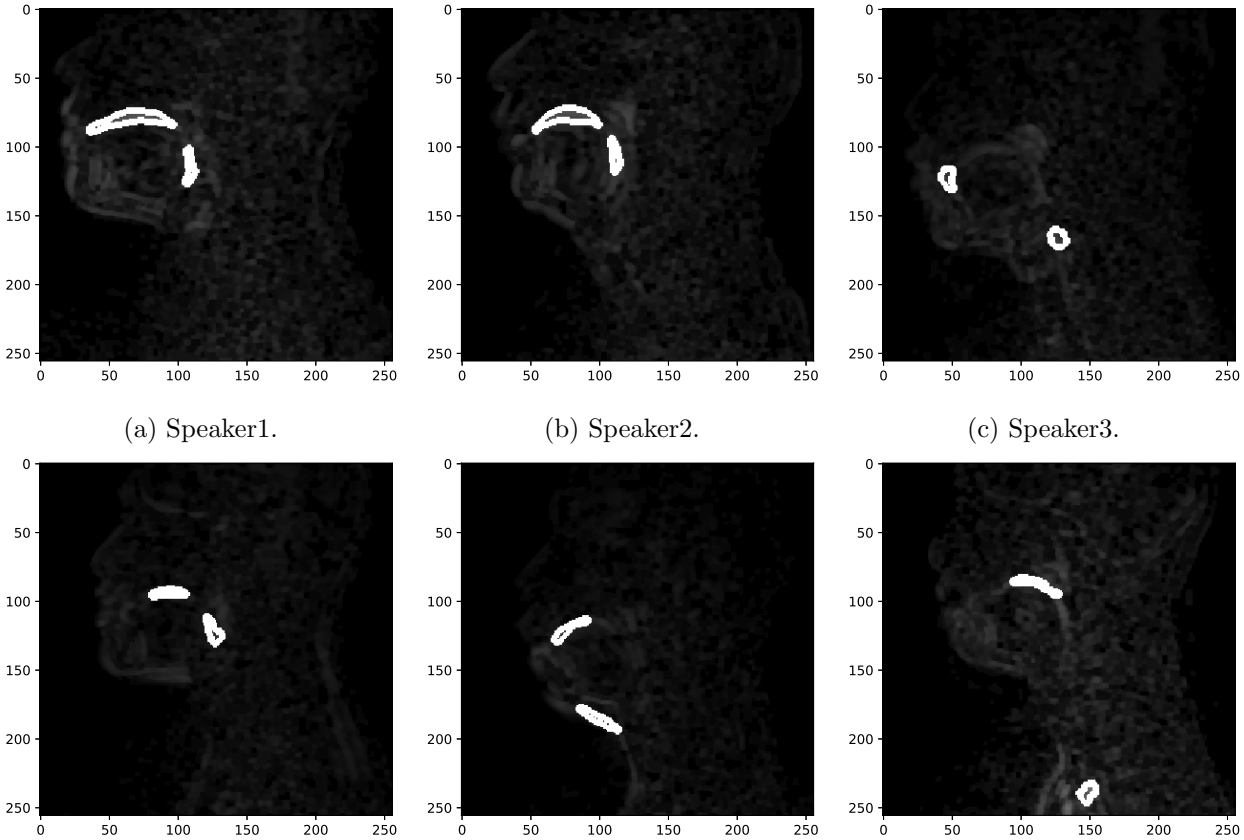


(a) Moving objects.



(b) Background image.

Fig. 1: Examples of motion detection by background subtraction method on a rtMRI animation.



(a) Speaker1. (b) Speaker2. (c) Speaker3.

(d) Speaker4. (e) Speaker5. (f) Speaker6.

Fig. 2: Examples of moving objects in each speaker.

図 3 に図 2a, 2b の発話者のパラ言語発話別の動体検出結果を示す. 図上段 3a, 3d が図 2a に, 図下段 3e, 3h が図 2b に対応している.

疑い発話時である図 3b, 3f と感心発話時である図 3c, 3g とを比較すると, 舌の前後運動を示す動体は両発話者ともに疑い発話時の方が声道前側に付置することが確認された. この結果は前川らの研究 [3] と同一の見解を示す結果である. 加えて, この領域における動体の長さを比較すると, 疑い発話時である図 3b, 3f は, 他のパラ言語表現と比較して前後方向に長く運動をしていることが判明した. 前川らは同研究において, 音声の音響特徴量を解析し, 疑い発話時には感心発話時と比較し第 2 フォルマント周波数が上昇することを報告している [3]. 今回発見された舌の前後方向の運動量の上昇は, この音響的効果を生成するために, 生じるものと考えられる.

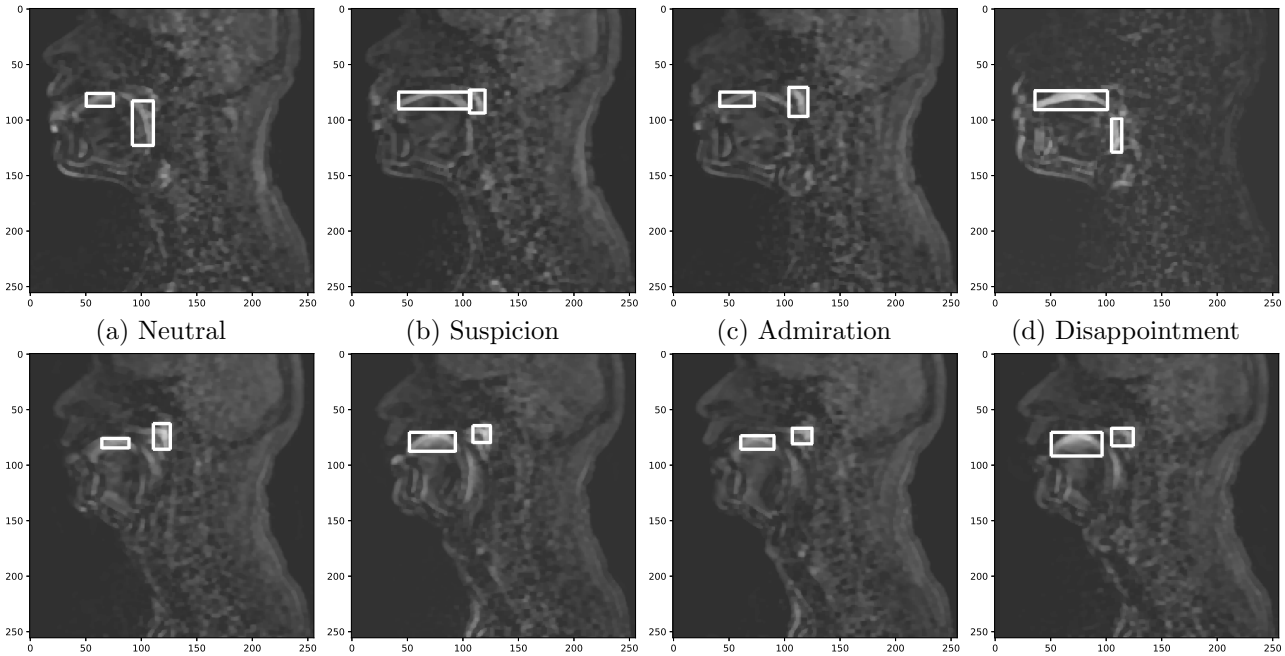
また, 落胆発話時である 3d, 3h を観察すると, 他の表現と比較すると舌動体位置は上下方向に高く付置しており, 領域の高さも高くなることが観測された. 加えて落胆発話時および疑い発話時には奥舌付近の動体 (図 3d) や, 軟口蓋付近の動体 (図 3h) と舌動体との距離が短くなることが観察された. これは落胆発話時および疑い発話時には中咽頭付近の空間を制御していることを示唆する結果である.

4 まとめと今後の課題

本研究では, 人手による各調音点の推定を必要としない rtMRI 動画による調音運動の解析方法の確立を目指し, 動画解析の分野で一般的な動体検知技術である背景差分法を利用し, rtMRI 動画からの動体検出および, その運動の観測を試みた.

まず, 背景差分法の中でも rtMRI 動画に応用しやすいフレーム間差分法を使い, rtMRI 動画の動体検出を試みた. その結果, 舌運動や, 唇, 下顎, 咽頭など調音音声学的に重要視される箇所に動体が検出された. また, 複数人を対象とした検出実験においては, 個人ごとに運動の特性は異なるものの, 基本的な調音器官の運動を検出しうるという知見を得た.

さらにパラ言語発話に動体検出を適応させたところ, 既存研究で指摘されている調音運動と同



(a) Neutral (b) Suspicion (c) Admiration (d) Disappointment

(e) Neutral (f) Suspicion (g) Admiration (h) Disappointment

Fig. 3: Examples of moving objects in each para-linguistic expression. Fig. 3a – 3d indicate moving objects of speaker1. Fig. 3e – 3h indicate moving objects of speaker2.

様の運動を観察できるとともに, 従来の磁気計測装置では計測困難であった声道奥の運動に関しても観察することに成功した.

今後の課題としては, 発話者ごとの運動特性の検討が挙げられる. 特に口腔内の形状と運動特性との関連を検討することにより, 各種運動の発話者間の比較や統計モデルの作成を行う.

謝辞 本研究は JSPS 科研費 JP17H02339 の助成を受けたものです.

参考文献

[1] Fujisaki, Hiroya. Prosody, models, and spontaneous speech. Computing prosody. Springer, New York, NY, 1997. 27–42.

[2] 前川喜久雄. パラ言語的情報–話しことばの本質 (日本語学のフォーカス). 別冊国文学 53 (2000): 172–175.

[3] Kikuo Maekawa and Takayuki Kagomiya. Influence of paralinguistic information on segmental articulation, Proceedings of ICSLP2000, Beijing, pp.349–352, 2000:10.

[4] Erickson, Donna, et al. Articulatory characteristics of emotional utterances in spoken English. Sixth International Conference on Spoken Language Processing. 2000.

[5] Vikram Ramanarayanan, et al. An investigation of articulatory setting using real-time magnetic resonance imaging. JASA, 134 (1), 510–519, 2013.

[6] 後藤翼 他. 機械学習による rtMRI 動画における発話器官の輪郭抽出方法の検討. 日本音響学会 2018 年秋季研究発表会講演論文集, 813–814, 2018.

[7] Medical Imaging Technology, ディジタル画像処理［改訂新版］／Digital Image Processing: An Algorithmic Introduction Using Java (Texts in Computer Science) Second Edition, 2017, 35(3), 171–172

[8] Buades, et al. Non-local means denoising. Image Processing On Line 1 (2011): 208–212.

[9] Bradski, G. The OpenCV Library. Dr. Dobb's Journal of Software Tools. 2000.