動画生成AIによるrtMRI動画のアップスケールの試み*

☆藤木 祐輔，北村 達也（甲南大），前川 喜久雄（国語研）

1 はじめに

MRI は発話時の音声器官の形態を観測する手法として音声生成過程の研究に貢献してきた．かつての MRI は時間分解能が低く，調音運動を観測するには同じ発話を繰り返すなどの特殊な方法[1]が必要であったが，リアルタイム MRI (rtMRI) により発話動態の観測が可能になった．その時間分解能としては 100 fps を超える報告[2,3]もあるが，日本国内の MRI 装置による rtMRI 動画では 27 fps にとどまっている．これは子音など速い調音運動の観測には十分とは言えない．

一方で，生成 AI の飛躍的な発展により，画像や動画の解像度や時間分解能を大幅に向上させる（アップスケール）ことができるようになってきた．もしこの技術を用いて rtMRI 動画の解像度や時間分解能を向上させることができれば，MRI 装置の性能の制約を超えることができる．そこで，本研究では AI を活用した画像・動画のアップスケーラーや画像補間 AI (Frame Interpolation for Large Motion, FILM[4]) を用いて rtMRI 動画の時間補間を試みた．

2 データ

国立国語研究所が公開している「リアルタイム MRI 日本語調音運動データベース」第 2.1 版[5]のうち，男性話者 S1 が/ara/と単独発話した rtMRI 動画を使用した．この動画では舌尖が挙上して歯茎に向かって動き，歯茎に接触し，その後離れるという一連の動きが記録されている．rtMRI 動画の時間分解能は 27.17 fps（以下，27 fps と表記），画像サイズは 256×256 ピクセル，撮像領域は 256×256 mm，スライス厚は 10 mm であり，MP4 形式で保存されている．上記の/ara/発話動画の時間長は約 0.9 s，フレーム数は 24 である．

3 実験 1 アップスケーラーによる時間補間

3.1 方法

ffmpeg（バージョン 8.0.1）を用いて，27 fps の rtMRI 動画M_{org}のフレームレートを 10 fps に変換した．これを動画M_{10}と表す．次に，Topaz Labs 社の画像・動画アップスケーラー Topaz Studio（バージョン 1.0.4）の frame interpolation 機能を用いて，動画M_{10}を 27 fps の動画M_{27}に再変換した．そして，動画M_{27}において動画M_{org}のフレームが再現されているかを目視により確認した．

3.2 結果と考察

動画M_{org}において/r/の調音時に舌尖が歯茎に接触した状態から/a/の調音に向かう 4 フレームを時間軸上に配置して Fig. 1(a)に示す．また，動画M_{10}と動画M_{27}においてこの区間に対応するフレームを Fig. 1(b)および(c)に示す．なお，この図では rtMRI 動画から口腔領域のみを切り出している．

Fig. 1(a)および(c)では，フレームが 1/27 s ごとに並べられている．Fig. 1(a)の第 1 フレームでは舌尖が歯茎に接触しており，続く第 2 フレームにて舌尖が歯茎から離れている．その後，第 3，第 4 フレームでは舌尖の挙上は見られない．Fig. 1(b)ではフレームが 1/10 s ごとに並べられている．Fig. 1(c)では第 1 フレームでは舌尖が歯茎に接触しているが（ただし，舌尖がぼやけている），第 2 フレーム以降では舌の挙上は見られず，動画M_{org}に存在した Fig. 1(a)第 2 フレームの舌形状が消失している．この結果から，現段階の Topaz Studio の frame interpolation 機能では物理学的法則に従った運動に関する時間補間は困難であり，それゆえに調音運動の rtMRI 動画を正確に時間補間することは難しいといえる．

---

* A study on real-time MRI movie upscaling using image processing artificial intelligence, by FUJIKI, Yusuke, KITAMURA, Tatsuya (Konan Univ.), and MAEKAWA, Kikuo (NINJAL).

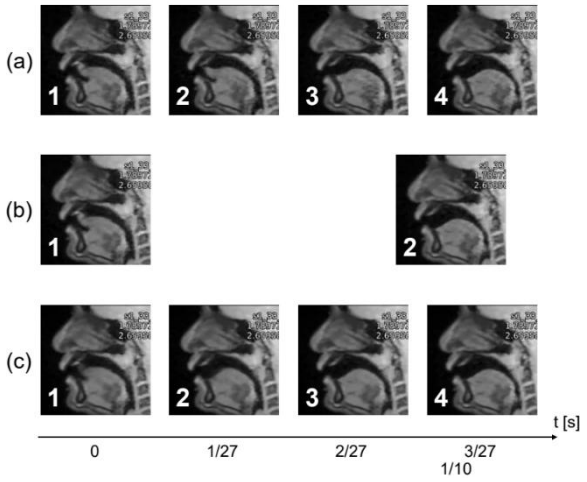

Fig. 1 Successive frames of movies (a) M_{org}, (b) M_{10}, and (c) M_{27}.

4　実験 2 FILM による時間補間

4.1 方法

動画M_{org}の舌尖が歯茎に接触したフレーム（Fig. 1(a)の第 1 フレーム）と第 3 フレームを入力とし，その中間フレーム（第 2 フレーム）を Reda ら[4]による FILM にて生成した．

FILM は深層学習ベースの画像補間手法で，2 つの画像を入力としてその中間時点の画像を生成する．FILM のモデルは，(1) 各入力データの特徴抽出器，(2) 画素単位の動きを求める双方向 (bi-directional) 運動推定器，(3) 補間画像を出力する統合モジュールの 3 つから成る．

4.2 結果と考察

補間対象である動画M_{org}の第 2 フレーム，およびそのフレームを FILM にて生成させた画像をそれぞれ Fig. 2(a)，(b)に示す．Fig. 2(b)では舌尖がやや薄くなっているものの，舌輪郭は Fig. 2(a)のものに近い形状が得られている．今後，多量のデータを用いた詳細な評価が必要ではあるが, FILM は調音運動の rtMRI 動画の時間分解能の向上に利用できる可能性がある．

なお, Fig. 2(b)の左下には数字の 4 桁目に“0”と“2”を補間した図形（文字）が現れている．この図形は FILM の補間の性質を示している．

Fig. 2 (a) Original frame and (b) frame interpolated by FILM.

5　おわりに

本研究では，画像処理 AI (Topaz Studio)および画像補間 AI (FILM)を用いて rtMRI の時間補間を試みた．その結果，前者では本研究で対象とした舌運動の補間は困難であったが，後者では比較的妥当な補間が可能であることを示唆する結果が得られた．

謝辞

本研究は JSPS 科研費（No. 24K00071）および甲南デジタルツイン研究所の支援を受けた．

参考文献

[1] Masaki et al., *J. Acoust. Soc. Jpn.*(*E*), 20(5), 375‐379 (1990).
[2] Fu et al., *Magn. Reson. Med.*, 73, 1820‐1832 (2015).
[3] Ilts et al., *Quant. Imaging Med. Surg.*, 5, 374‐381 (2015).
[4] Reda et al., *ECCV2022* (2022).
[5] Maekawa, *AST*, 46(1), 45‐54 (2025).