E4

調音運動動画アノテーションシステムの開発と応用

浅井拓也, 菊池英明 (早稲田大学), 前川喜久雄 (国立国語研究所)
takuya.waseda.1119@gmail.com, kikuchi@waseda.jp,
kikuo@ninjal.ac.jp

1 はじめに

調音音声学には多くの未解明問題が存在し, その多くは調音運動に関する客観的なデータの不足ないし欠如に起因している. 調音運動の客観的観測手法としては, X 線マイクロビーム・EMA・WAVE などを用いた研究があるが (例えば, Westbury, Milenkovic, Weismer, & Kent (1990); Yehia, Rubin, & Vatikiotis-Bateson (1998); Kitamura & Hatano (2012) 等), これらの計測装置は数個のセンサーの位置情報だけを提供するものであり, 音声発話時の喉頭・咽頭を含めた声道の正中矢状面全体の計測は困難な課題であった. それに対し, MRI 装置を使用した計測では, 喉頭・咽頭を含めた声道の正中矢状面全体の情報が含まれる. 特に近年では, MRI 装置の性能向上および高度なサンプリング技術の適用によって, リアルタイムでの MRI 動画撮像が可能になってきており (Ramanarayanan, Goldstein, Byrd, & Narayanan, 2013), 日本では ATR Promotion の脳活動イメージングセンタが, 正中矢状断面に限定した動画を毎秒 14 フレームで撮像するサービスを提供している. リアルタイム MRI（以下 rtMRI）データは, 従来手法に比して情報量が圧倒的に豊富であり, 調音音声学の再構築を促す可能性を秘めている.

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. しかし rtMRI 動画はデータに対するアノテーションや分析の環境が整備されているとは言いづらく, 単にデータを公開するだけでは rtMRI 動画に基づいた調音音声学研究は普及しにくいと予想される. そのため上記科研費研究では, rtMRI データの解析環境の整備も進めている. 本稿では動画データビューワーである MRI Vuewer[1] の設計と実装を報告し, その応用例として日本語子音の硬口蓋化現象の簡単な分析結果を報告する.

2 MRI Vuewer

2.1 設計

MRI Vuewer は, 調音運動計測データに対するデータビューワーであり, アノテーション環境である. 類似の機能をもったツールとしては ELAN (Wittenburg, Brugman, Russel, Klassmann, & Sloetjes, 2006) が有名であるが, ELAN には動画データに同期した音声スペクトルを直接的に表示する機能が欠けており, 調音運動とそれが生成する音声スペクトルとの関連を直感的に理解することが困難である.

rtMRI 動画データの観察, アノテーションには最低限, 以下の機能を実装する必要がある.

- 音響的イベントの記述
- 調音運動画像および音声スペクトルの同期的表示
- 前後調音運動画像の表示
- 調音運動画像の観測点記述

一般に音声言語資源に対して研究目的でアノテーションを行う際には, 発話, 単語, 音素列等のイベントの時間的な境界を記述していく. このような作業を行うには, 収録された音声のスペクトログラムを目視で確認し, その境界認定を行う必要がある. そのため, 音声のスペクトログラムの表示が必要である. さらに, 特定の調音運動と, その結果生成される音声のスペクトルとの関連を直感的に把握するためには, 調音運動の画像データと音声のスペクトログラムは時間的に同期された形で表示を行う必要がある. 本アプリケーションが対象とする rtMRI 画像は調音運動を撮影したものである. 運動の

[1] Vuewer の表記はアプリケーション作成時に使用したライブラリに由来する.

観察を行うには, ある時刻の画像のみに注目するのではなく, 前後フレームの画像を考慮する必要がある. このような需要を満たすためには, 指定時刻から任意フレーム分前後にずらした画像を表示する機能が必要となる. 加えて rtMRI 動画データは, 図 1 に示すように, 観測点が明示的に表現されておらず, 手動もしくは, 物体検知等の画像認識技術を利用する必要がある. 後者のアプローチによる調音運動の切り出しも現在検討中であるが, この場合においても機械学習用の特別な教師データを作成する必要がある. このような需要から, 特定フレームの画像に対し, 任意の座標点を記述, 保存する機能が必要となる.

また, 作成するアプリケーションは rtMRI 動画データベースとともに公開される予定である. そのため, データベースとの連携をしやすい形でアプリケーションを開発すること, 研究で使用する際には, その観察結果を何らかの統計解析アプリケーションや機械学習用アプリケーションで利用しやすい形式で取り出し可能であることが望まれた.

図 1: MRI による調音運動画像例

rtMRI 動画データベースとの連携の観点から, MRI Vuewer はブラウザ上で動作する Web アプリケーションとして作成された. Web アプリケーションとすることにより, データベース公開後には, 誰でもすぐに rtMRI 動画の観察が行えるような環境を提供することが狙いである.

Web アプリケーションは一般的に, サーバー, クライアント間での通信が必要不可欠であり, 特に動画や音声といった容量の大きなデータのやり取りは不得手である. そのため, 本アプリケーションでは, HTML5 で制定された, Web Audio API (MDN web docs, 2018b), Canvas API (MDN web docs, 2018c) に注目し, サーバー側ではなく, ブラウザ側で必要な音響解析および画像処理を行うことを試みた. このようにすることにより容量の大きなデータのやり取りを最小限に留めることが目的である. また, 近年 JavaScript で採択された promise オブジェクト (MDN web docs, 2018a) を利用することにより, 調音運動動画と音声スペクトル画像を同時に再生することを試みた. なお, アプリケーション作成の効率化のため, Web Audio API の利用は wavesufer.js (Guisch & thijstriemstra, 2018), promise オブジェクトの利用は vue.js (You, 2018) を利用した. なお, 本アプリケーション名はこのライブラリに由来する.

2.2 実装

以下に作成された MRI Vuewer の現時点での概要を示す[2]. MRI Vuewer は一つ以上の動画ファイルをサーバーまたはローカル PC より受け取る. 受取り可能な動画ファイル形式は, MP4, WebM, Ogg 等に対応している. つまり, MRI Vuewer そのものは rtMRI 動画以外の動画ファイルに対しても, データビューワーおよびアノテーション環境として機能する.

動画ファイルの受取りに成功すると図 2 に示す動画アノテーション画面に遷移する. この画面は上に上げた必要要件のうち, 上 3 つを満たす画面である. この画面は大きく, アノテーションコンポーネント, リージョンコンポーネント, ポイントコンポーネントの 3 つのコンポーネントに分かれている.

[2] ブラウザから https://kikuchiken-waseda.github.io/MRIVuewer/ にアクセスすることで試用版をみることが可能. ただし rtMRI 動画データに関しては現在整備中であるため, 公開していない.

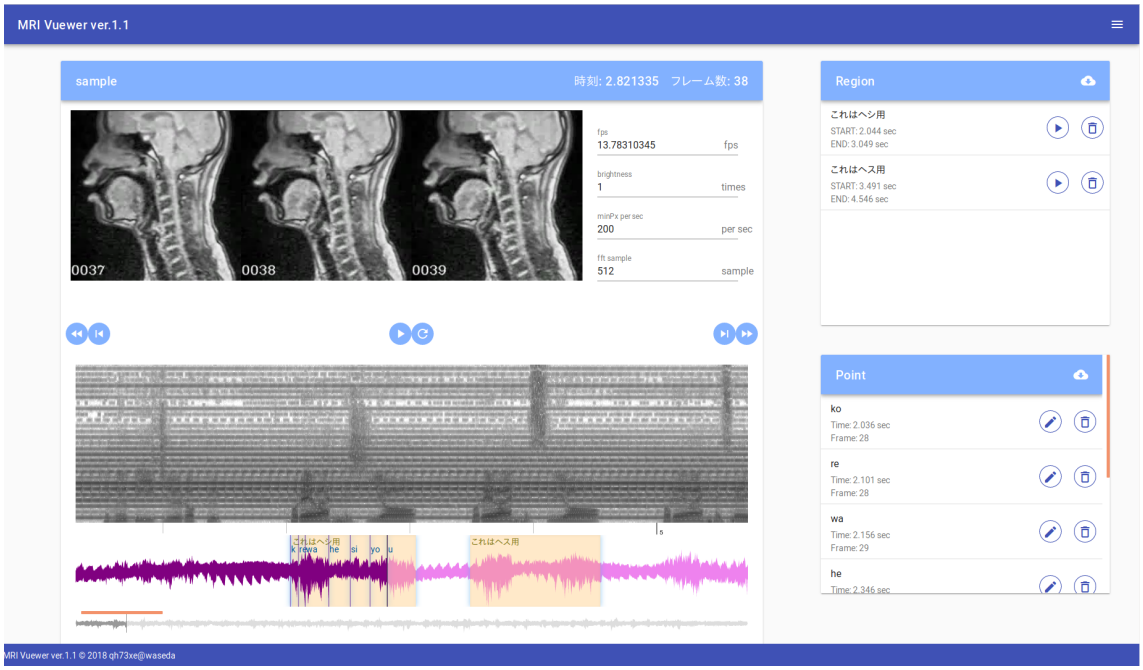

図 2: 動画アノテーション画面. 左半分がアノテーションコンポーネント, 右上部がリージョンコンポーネント, 右下部がポイントコンポーネント

アノテーションコンポーネントには調音運動の動画データ及び, 音声のスペクトログラムが表示されており, 動画の再生および音声波形へのアノテーションが可能である. アノテーションコンポーネントにある音声波形部分をクリックするとその時刻に撮影された rtMRI 動画が表示される. rtMRI 画像は 3 つ表示されているが, これは中心が現在時刻, 左右がそれぞれ前後 1 フレーム分の画像である. アノテーションコンポーネント中央にある再生ボタンをクリックすることで現在時刻から動画およびスペクトログラム画像を同時に再生することができる. また, 通常再生の他に動画 1 フレームずつのコマ送りを行うことも可能である. アノテーションコンポーネントの音声波形表示箇所をドラックアンドドロップをすると, リージョンコンポーネントに一つの要素が生成される. これは発話や単語, 音素区間といった開始, 終了点を持つイベントを記述するために使用される. また, リージョンコンポーネントにある再生ボタンをクリックすると, 登録された時間区間分の動画のみが再生される. アノテーションコンポーネントの音声波形表示箇所を ctrl キーを押しながらクリックすることにより, ポイントコンポーネントに一つの要素が生成される. これは, 単一時刻に対するイベントを記述するために使用される. リージョンコンポーネント及びポイントコンポーネントの内容は, 動的にブラウザのローカルストレージに保存され, ユーザーは特に明示的な保存操作をしなくともアノテーションの記録を残すことが可能である. また, それぞれのコンポーネント上部にあるダウンロードボタンをクリックすることで, CSV 形式でアノテーションデータを取得することができる.

ポイントコンポーネントにある編集ボタンをクリックすることで図 3 に示す画像アノテーション画面に遷移する. この画面はアプリケーション必要要件のうち, 調音運動画像の観測点記述機能を満たす. この画面では, ポイントコンポーネントで指定された時刻の調音データをキャプチャーし, 静止画として編集することが可能である. 図 3 左に示しているように調音画像の任意の点をクリックすることで円形のマークが描画される. この描画領域は複数の画像がレイアー構造になっており, 図 3 右にあるスイッチを押すことで, 調音画像の表示, 非表示の切り替えを行うことが可能である. 画面上部にあるダウンロードボタンをクリックすることにより, CSV 形式の座標点データを取得可能である. ただし, この画像表示は画面表示時のウインドウサイズに依存するため, X, Y 座標点以外にアノテー

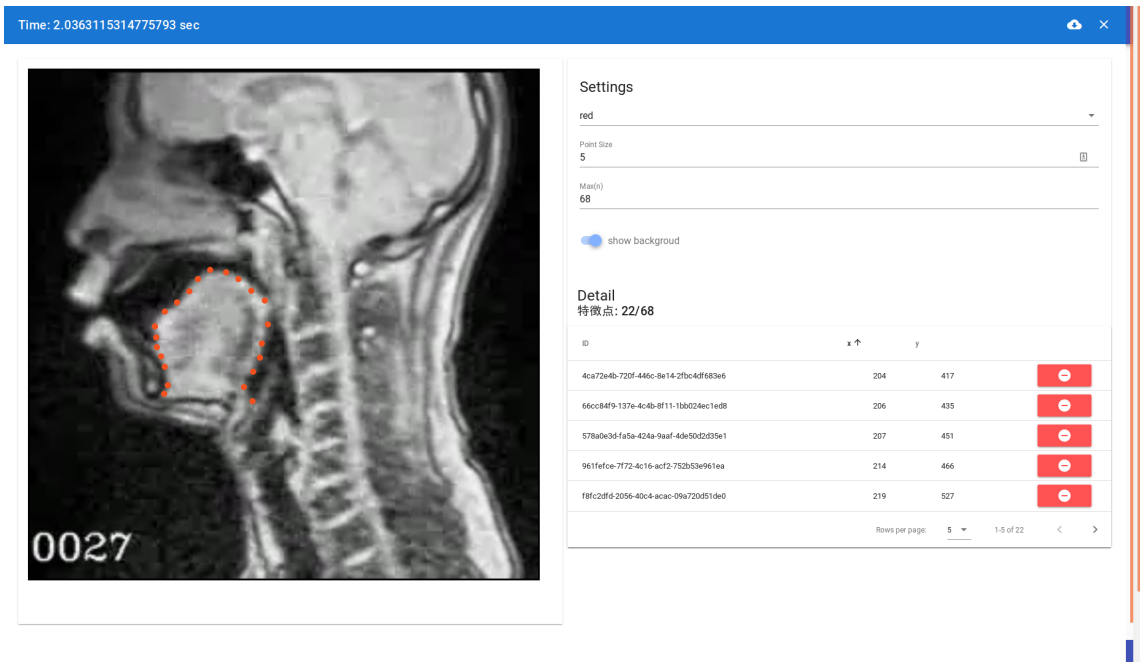

図 3: 画像アノテーション画面

ション時の画像の縦横幅も記録している. なお, ページ上部のクリアボタンをクリックすることで動画アノテーション画面に遷移することができる. この画面上で, 調音画像のアノテーションを行うことで, 観測点が明示的に表現されていない動画データに対して, 事後的かつ手動による観測点を追加することが可能である. また, 既存の物体検知およびランドマーク検知システムの学習データとして, このアノテーションデータを利用することで, 画像アノテーションの半自動化を行うことも可能である.

3　応用例

3.1　データ

MRI Vuewer を用いた調音音声学的分析の試行例として, 日本語カ行子音の調音位置を分析した. 標準語のカ行子音は「キ」以外のモーラの子音が [k], 拗音と「キ」の子音が硬口蓋化した [kʲ] で表記されるのが普通である (例えば 斎藤純男 (2006) 参照). この表記の妥当性を rtMRI データベースを用いて検証する. 検証のポイントは, /k/の調音点が上述のように二分されるかどうかである. 現在構築を進めているリアルタイム MRI 動画データベースには日本語モーラリストの読み上げ課題が含まれており, カ行については直音の「カキクケコ」, 拗音の「キャキュキョ」 に「キェ」を加えた 9 モーラが対象となっている. 各モーラとも発話回数は 1 回である. [k] ないし [kʲ] の調音では, 舌と口蓋による声道の閉鎖が形成されるが, 一般に閉鎖は声道のかなり長い区間にわたって形成されるため, 単一の調音点を決定することに困難がある. そこで, 閉鎖された声道区間のうち最も声門に近い部位 (右端) を調音点に認定することとした. また子音の閉鎖は一定時間持続し, その間も連続的に変化するので, 測定のタイミングを決める基準も必要である. これについては, rtMRI 動画の視察で, [k] ないし [kʲ] の閉鎖の開放が明瞭に確認できるフレームを決定し, そこから 2 フレーム遡ったフレームを調音点の測定対象とした. 図 4 に同一話者による /ka/, /ki/, /ke/, /kja/ の測定例を示す. 各図は rtMRI の連続する 3 フレームであり, 右端が声道の開放が認められるフレーム, 左端が子音の調音点を測定したフレーム, 左端フレーム中の丸印が決定された調音点である. 調音点は, 座標の原点をフレームの左上隅に設定して測定しており, 単位はミリメートルである.

3.2　分析結果

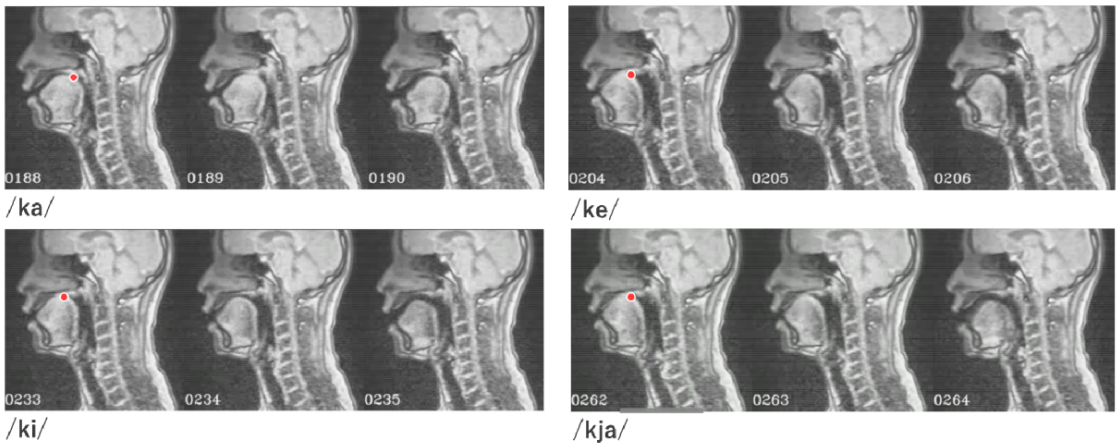

図 4: rtMRI データからの子音調音点の測定例

図 5 に, 現在構築中のデータベースに含まれる標準語話者 6 名 (いずれも男性, 30 60 代) の測定結果をまとめる. 横軸は調音点の X 座標、縦軸が Y 座標である. 黒丸が平均値, エラーバーは標準誤差であり, Y 軸エラーバーの先端にモーラの別を示した. 声道サイズの正規化処理等は施していない生データの分析である.

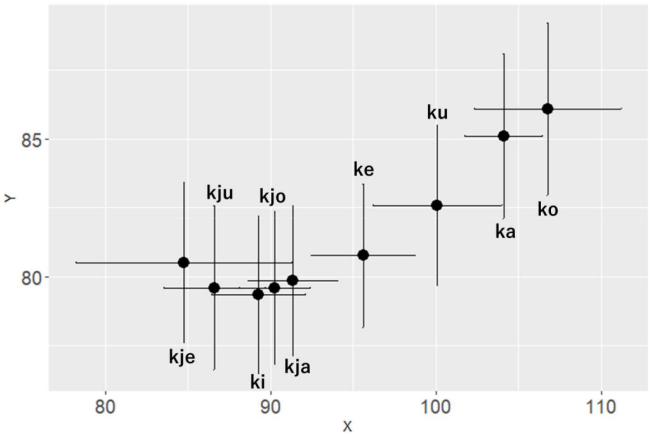

図 5: 子音/k/の調音位置の分布. 単位は [mm]

図 5 では拗音子音 /kj/ が左下に, 直音子音/k/が右上にまとまっている. つまり拗音の調音点は直音に較べて口唇よりに分布している. これは硬口蓋化の効果として従来から想定されてきた調音上の相違点である. また直音のうち/ki/だけは拗音と同じクラスターに属しているが, これも母音/i/による硬口蓋化の効果として想定されてきたものである.

図 5 における新しい発見は, 拗音子音に較べて直音子音の分布がまとまりに欠けていることである. 特に/ke/の直音子音は拗音子音のクラスターに隣接する位置に分布しており, 硬口蓋化に類した副次調音の影響を被っている可能性をうかがわせる. データベースに実際の言語行動ではほとんど用いられることのない「キェ」が入っているのは, この問題を検討するためであったが, 図 5 を見ると, /kje/ は拗音子音のなかでも X 座標値が最も小さい値をとっている. これは/ke/との調音上の距離を保つための調整である可能性がある.

4 まとめ

以上のように rtMRI データは調音音声学における従来の想定の妥当性を確認するためにも, 新たな問題を発見するためにも有益である. 今後はデータベースを拡充するとともに, 調音音声学上の基本問題を順次とりあげて分析する予定であるが, そのためにも MRI Vuewer をはじめとする分析環境の充実が急がれる.

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

Guisch, & thijstriemstra (2018, July) *wavesurfer js*. https://wavesurfer-js.org.

Kitamura, T., & Hatano, H.(2012)"Measurement of temporal cange of vocal tract volume during production of plosive and fricative consonants." *IEICE technical report. Speech*, 112, 19-23.

MDN web docs (2018a, July) *Promise*. https://developer.mozilla.org/ja/docs/Web/JavaScript/Reference/Global_Objects/Promise.

MDN web docs (2018b, July) *Web audio api*. https://developer.mozilla.org/ja/docs/Web/API/Web_Audio_API.

MDN web docs (2018c, July) *canvas*. https://developer.mozilla.org/ja/docs/Web/HTML/Element/canvas.

Ramanarayanan, V., Goldstein, L., Byrd, D., & Narayanan, S. S. (2013) "An investigation of articulatory setting using real-time magnetic resonance imaging." *The Journal of the Acoustical Society of America*, 134, 510–519.

Westbury, J., Milenkovic, P., Weismer, G., & Kent, R. (1990) "X-ray microbeam speech production database." *The Journal of the Acoustical Society of America*, 88, S56-S56.

Wittenburg, P., Brugman, H., Russel, A., Klassmann, A., & Sloetjes, H. (2006) "Elan: a professional framework for multimodality research." *LREC*, 1556–1559.

Yehia, H., Rubin, P., & Vatikiotis-Bateson, E. (1998) "Quantitative association of vocal-tract and facial behavior." *Speech Communication*, 26, 23–43.

You, E. (2018, July) *The progressive javascript framework*. https://jp.vuejs.org/index.html.

純男 斎藤 (2006) 『日本語音声学入門』三省堂.