

調音運動動画アノテーションシステムの開発*

○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)

1 背景

リアルタイム MRI 撮像技術や磁気計測技術 (WAVE, EMA) 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することが可能になりつつある [1-3]. しかしながら収集されたデータに対するアノテーションや分析の環境は整っていない. 音声アノテーションツールとしては Praat [4] や ELAN [5] が有名であるが, 図 1 に示すように Praat は動画ファイルを扱えず, 図 2 に示すように ELAN は直接的に音声スペクトルの表示ができない. そのため, 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. 我々は, このような問題点に対処するために, HTML5 における Web Audio API [6] 及び HTML5Canvas [7] に注目し, ビデオファイルの音声及び画像データの同期的操作及び, 指定時刻画像のアノテーション機能を持つ Web アプリケーションを作成した.

2 目的

本アプリケーションの目的は, 調音運動計測データに対するアノテーションの環境を提供することである. 調音運動計測データに対するアノテーションを行うには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およびデータ管理
 - 調音運動画像の観測点記述およびデータ管理
 - 調音運動画像および音声スペクトルの同期的表示
- 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, 図 1 に示すように, 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. また, 本アプリケーションが対象とする MRI 画像は図 3 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 MRI 動画アノテーションシステムでは, 特定フレームの画像に対し, 任意の座標点を記述, 保存する必要がある. 加えて, 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.

更に, 作成するアプリケーションには以下の特性が望まれる.

- アプリケーション使用の簡便さ
- 共同作業性
- データの再利用性

本アプリケーションで対象とする音声, 画像アノテーションはいずれも解析者が手作業で行う必要がある. 膨大な時間を要する作業である. アノテーション作業を効率的にすすめるためには, 誰でも簡単に使用可能である必要がある. また, アノテーションを行う人間は作業場所を共有しているとは限らないため,

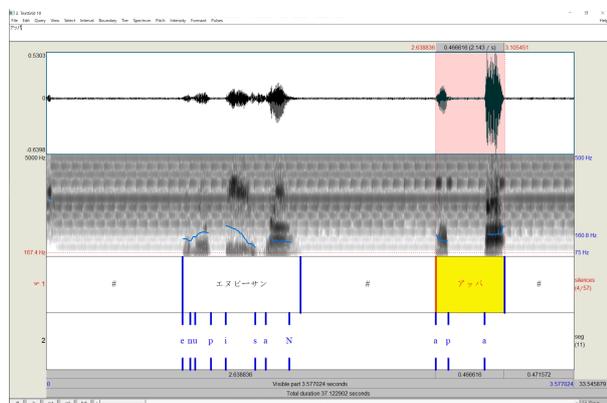


Fig. 1 Praat における音声アノテーション例

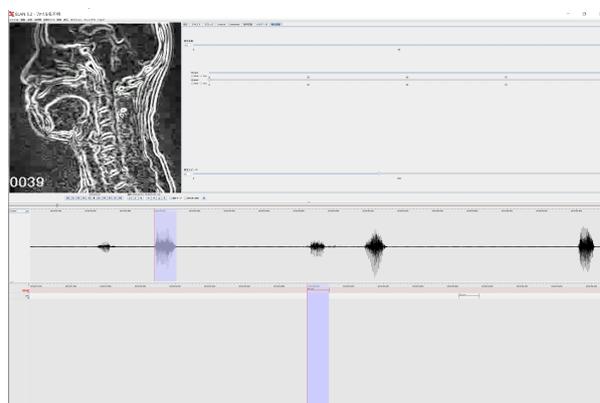


Fig. 2 ELAN における動画アノテーション例

*Development of an annotation system for the speech articulation movies. by ASAI, Takuya, KIKUCHI, Hideaki (Waseda University) and MAEKAWA, Kikuo (NINJAL)

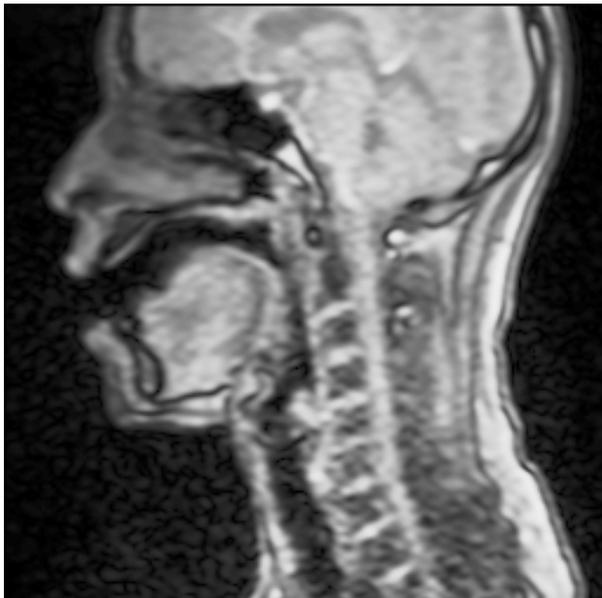


Fig. 3 MRI による調音運動画像例

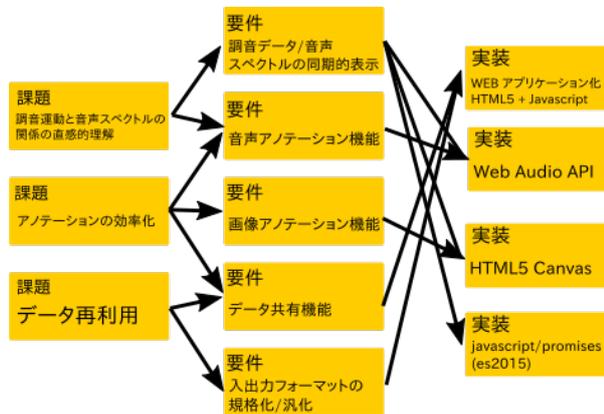


Fig. 4 調音データアノテーションシステムの課題、要件及び実装

データの円滑で安全なやり取りが行えることが望ましい。最後に、アノテーションデータを研究で使用するには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されることが想定される。そのため、アノテーションデータはなるべく汎用的な形式で出力されることが期待された。

3 提案手法

本アプリケーションにおける課題、要件及び、実装を図4に示す。上に挙げた課題の内、データの再利用性及び、アノテーションの効率化の観点から、ブラウザ上で動作するWebアプリケーションとして、MRI Viewerを作成した。Webアプリケーションとすることにより、データおよびアノテーション環境を同時に提供することや遠隔地における共同作業を行うことが可能になる。

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, Canvas API を最大限利用することにより、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 [8] 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 [9], promise オブジェクトの利用は vue.js [10] を利用した。

4 実装の現状

以下に作成された MRI Viewer の現時点での概要を示す。MRI Viewer 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。動画ファイルの受取りに成功すると図5に示す動画アノテーション画面に遷移する。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。

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。ここで、動画は、fps を明示することにより、前後1フレーム分まとめて再生することが可能である。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の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のみの動画が再生される。

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 + click することにより、ポイントコンポーネントの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

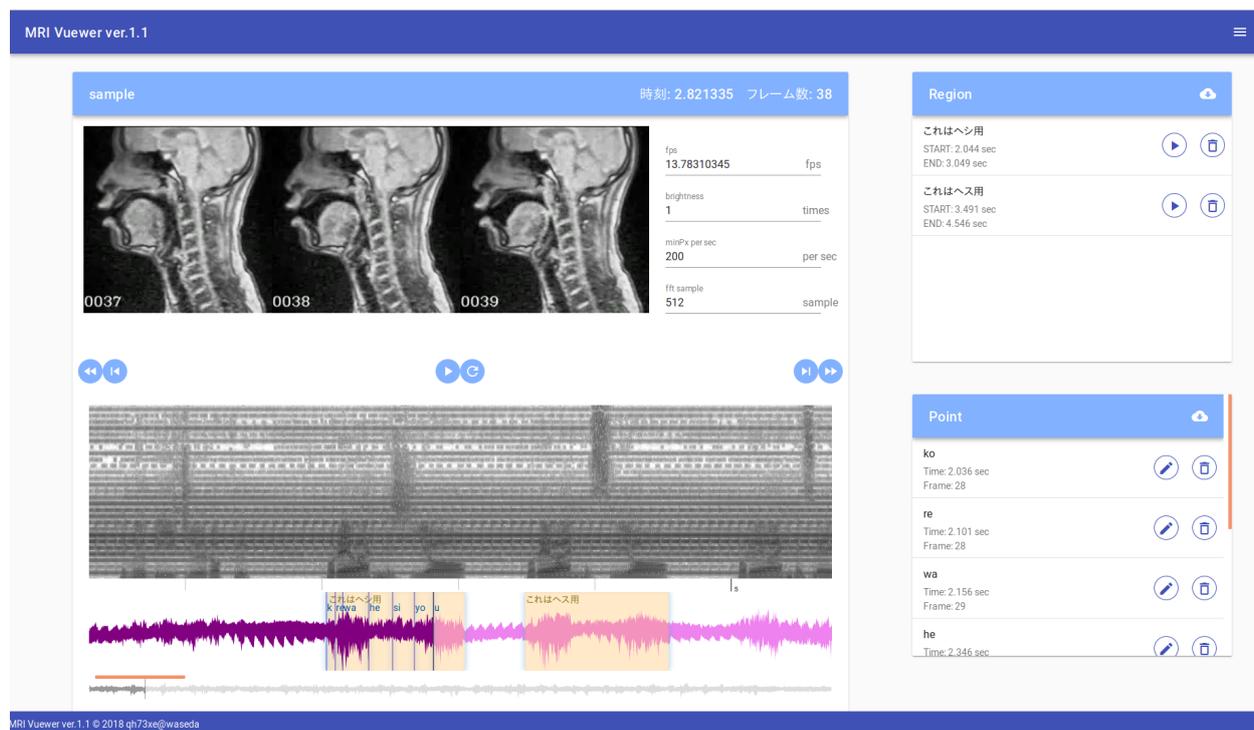


Fig. 5 動画アノテーション画面. 左半分がアノテーションコンポーネント, 右上部がリージョンコンポーネント, 右下部がポイントコンポーネント

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6に示す画像アノテーション画面に遷移する。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6右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ヤー構造になっており、図6左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ィ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

ションの半自動化を行うことも可能である。

5 まとめと今後の課題

本稿では、リアルタイムMRI撮像で収録された音声付き画像データの分析環境の拡充を目標に動画及び画像アノテーションツールを作成した。特に近年整備されたWeb Audio API, HTML5Canvas, promiseの機能を利用することにより、ビデオファイルの音声及び画像データの同期的操作が可能になった。これは、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に大きく貢献する機能である。また、Webアプリケーションとして作成することにより、ユーザーは特別なツールのインストールを必要とせず、調音動画データの解析を開始することが可能になる。ここで作成したアノテーションデータはCSVを始めとする単純な形式のファイルとしてダウンロード可能であり、アノテーションデータの公開や共有にも貢献すると期待される。

今後の課題としてはサーバーとの連携が挙げられる。今回は開発の仕様上できる限りサーバーとの通信を控え、ブラウザ側で必要な処理を行って来た。しかし、動画データそのものへのアクセス管理やアノテーション等の共同作業を安全に行うためにはサーバー側で中央管理を行うことが必要になる。また、音声のノイズ除去や画像認識を効率的に行うためには、動画データを音声ファイルや画像ファイルに変換し、それ

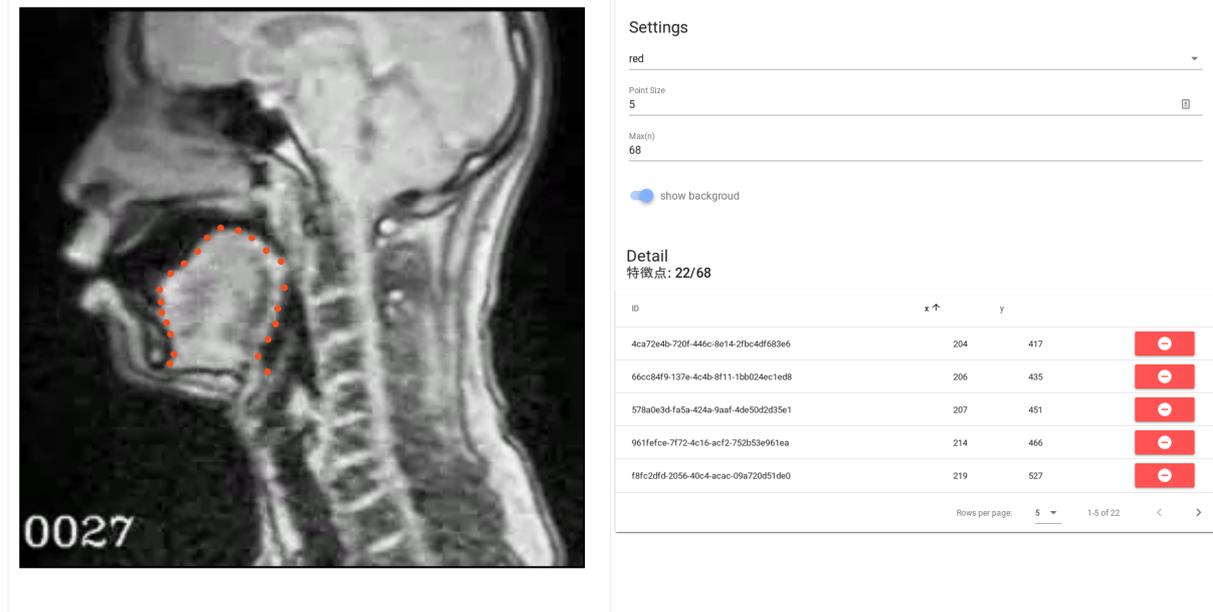


Fig. 6 画像アノテーション画面

を再構築することが必要になるが、現状の HTML5 や JavaScript の機能のみでこの作業を行うことは現実的ではない。そのため、種々の動画編集機能を有したサーバーを用意することも今後の課題となる。

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

- [1] Kawahara, S., Masuda, H., Erickson, D, et al. Quantifying the Effects of Vowel Quality and Preceding Consonants on Jaw Displacement: Japanese Data. *Journal of the Phonetic Society of Japan*, 2014, 18(2), 54-62.
- [2] KITAMURA, Tatsuya and HATANO, Hiroaki, Measurement of temporal change of vocal tract volume during production of plosive and fricative consonants IEICE technical report. *Speech*, 2012, 112(281), 19-23.
- [3] Vikram Ramanarayanan, Louis Goldstein, Dani Byrd, and Shrikanth S. Narayanan. An investigation of articulatory setting using real-time magnetic resonance imaging. *JASA*, 134 (1), 510-519, 2013.
- [4] Boersma, Paul & Weenink, David (2018). Praat: doing phonetics by computer [Computer program]. Version 6.0.40, retrieved 11 May 2018 from <http://www.praat.org/>
- [5] Wittenburg, Peter, Hennie Brugman, Albert Russel, Alex Klassmann & Han Sloetjes. 2006. ELAN: a Professional Framework for Multimodality Research. *Proceedings of the 5th International Conference on Language Resources and Evaluation (LREC 2006)*, 1556-1559.
- [6] MDN web docs. "Web Audio API" https://developer.mozilla.org/ja/docs/Web/API/Web_Audio_API (参照 2018-07-18)
- [7] MDN web docs. "< canvas >: グラフィックキャンバス要素" <https://developer.mozilla.org/ja/docs/Web/HTML/Element/canvas> (参照 2018-07-18)
- [8] MDN web docs. "Promise" https://developer.mozilla.org/ja/docs/Web/JavaScript/Reference/Global_Objects/Promise (参照 2018-07-18)
- [9] Guisch & thijstriemstra. "wavesurfer js" <https://wavesurfer-js.org/> (参照 2018-07-18)
- [10] Evan You. "The Progressive JavaScript Framework" <https://jp.vuejs.org/index.html> (参照 2018-07-18)