

発話のしにくさを自覚する 話者の調音運動の観察

北村達也（甲南大学知能情報学部） t-kitamu@konan-u.ac.jp

能田由紀子（国立国語研究所）

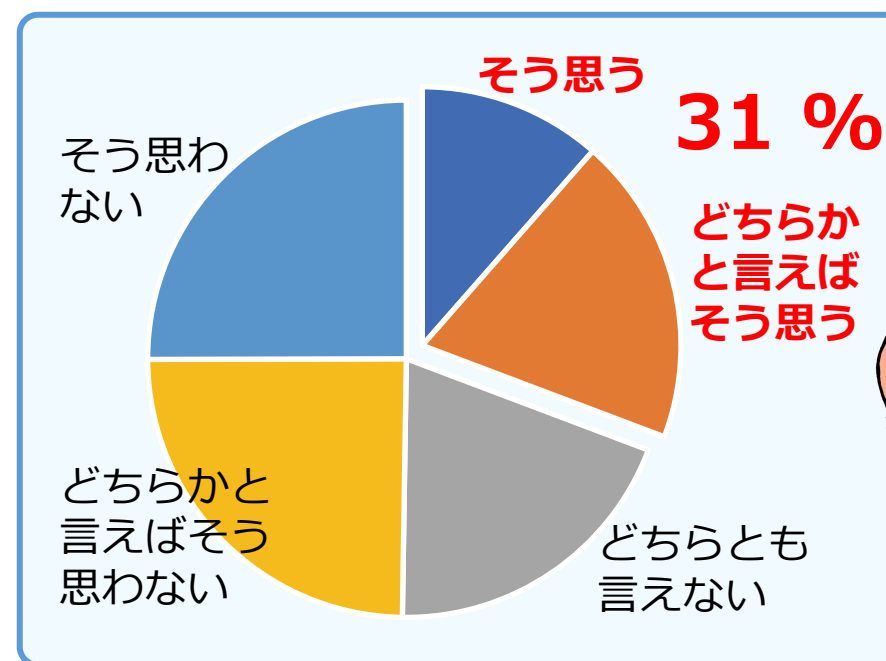
吐師道子（県立広島大学保健福祉学部）

発話のしにくさの自覚に関する調査

北村+ (2019)

- 対象：全国の大学生・院生約2,000名
- 緊張でうまく話せないケース，医学的な問題があるケースは除外
- 約3割がある程度以上の発話のしにくさを自覚
- 女性より男性，文系より理系の学生が自覚しやすい
- 発話のしにくさを自覚する人の9割が発話の改善を希望

発音がうまくいかないことがありますか？

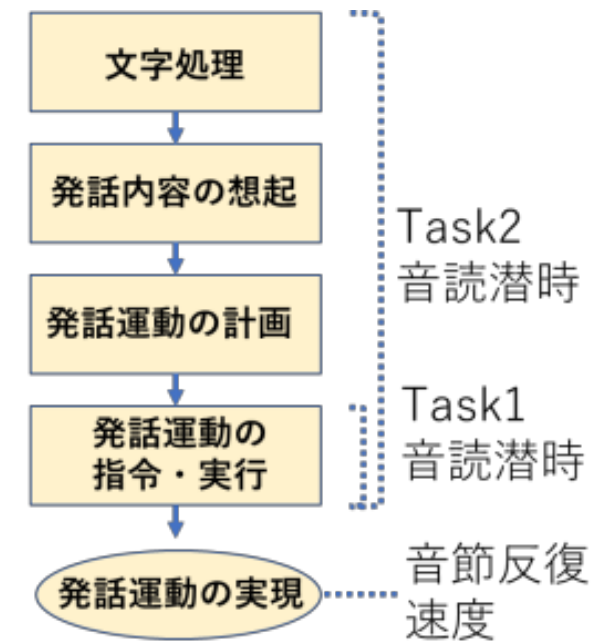


発話のしにくさの要因の調査

古田+ (2021)

- 発話のしにくさには発話過程のいずれかの段階における時間的な滞りが関与しているのでは？
- 音読潜時を調査
 - (1) 文字列を見て，一定時間経過後に発話
 - (2) 文字列を見た直後に発話
- 発話のしにくさの自覚の程度と音読潜時には有意な相関なし
- 発話のしにくさの要因は様々なのでは？

本研究では調音器官の構造や運動に着目

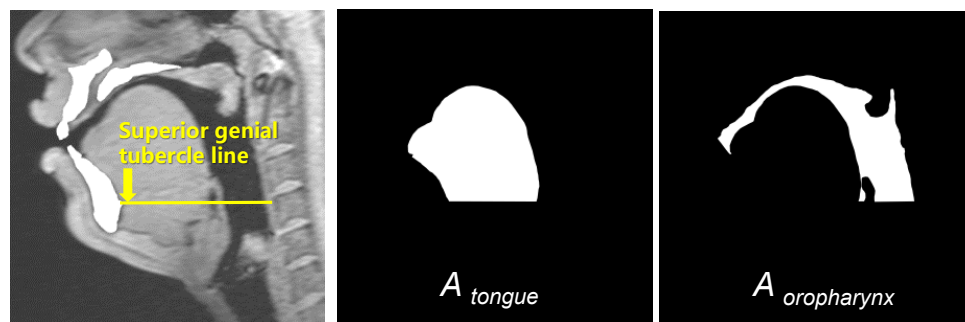


古田+ (2021)

Relative tongue size

Feng+ (2017)

- RTS: 母音/i/発話時の正中面上の舌と声道の面積の比



(a) Lower boundary

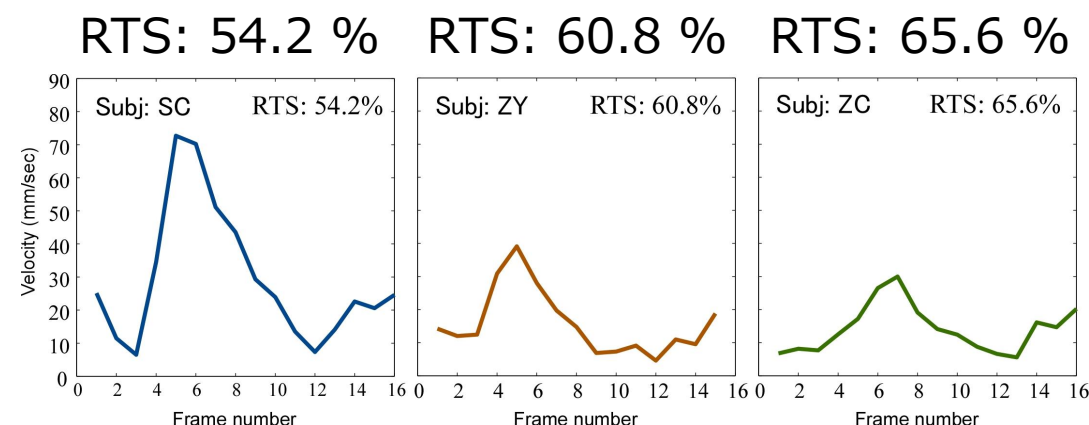
(b) tongue area

(c) oropharynx area

$$\text{Relative tongue size (RTS)} = \frac{A_{\text{tongue}}}{A_{\text{tongue}} + A_{\text{oropharynx}}}$$

RTSの定義 (Feng+, 2017)

- tagged-MRIを利用して舌の運動速度を計測
- RTSと速度に負の相関あり ($N=5$)

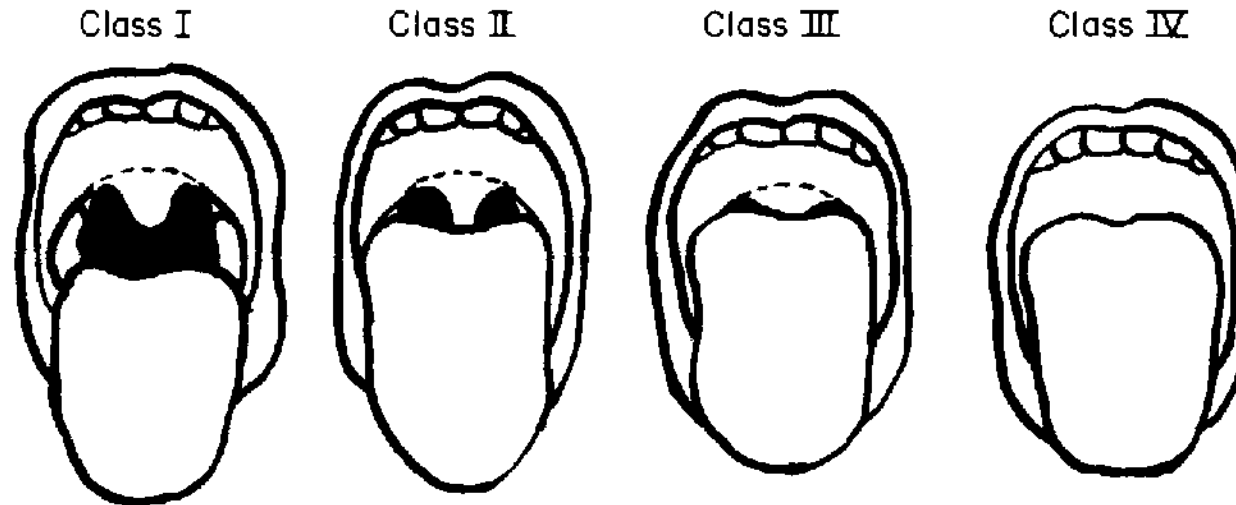


各フレームの舌運動速度 (Feng+, 2017)

Mallampati score

Samsoon & Young (1987)

- 気管内挿管のしやすさの目安
- RTSの視覚的評価に利用可では (Feng+, 2017)



Samsoon & Young (1987)

調音運動の周期性

立川+ (2015)

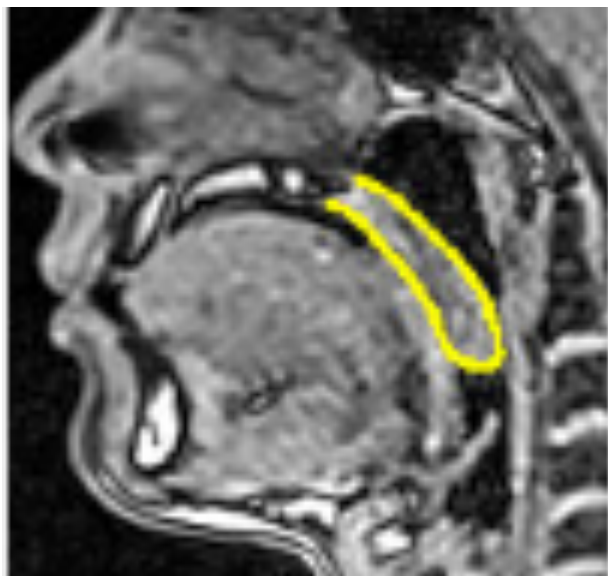
- 対象：発話のしにくさを自覚する成人4名(NS), 自覚しない成人3名(S)
- EMAにて音節反復課題(/ra/)などの舌運動を計測
 - NS群は反復回数が少ない傾向
 - NS群は舌尖の速さが小さい傾向
 - NS群は舌尖の速さの経時変化パターンが安定しない(不安定な)傾向

実験参加者

- 男子大学生計3名
 - 健常者 (音声や聞こえに関して医師に指摘されたことはない)
 - 3名とも課外活動を通して第1著者と数年間に渡る付き合いあり
- 発話のしにくさの自覚がない：1名 (S)
- 発話のしにくさの自覚がある：2名 (NS1, NS2)
 - NS1はかなり音声聞き取りにくい
 - NS1は口蓋帆が長い
 - NS2は音声はやや聞き取りにくい

参考：NS1の正中面のMRIデータ

立川+ (2017)



口蓋帆が長い



MRI撮像, 音声収録条件

- リアルタイムMRI撮像条件
 - ATR BAICのSiemens製MAGNETOM Prisma fit 3T
 - フレームレート : 14 fps
 - 撮像領域 : 256 mm x 256 mm
 - 画像分解能 : 1 pixel/mm
- 音声収録条件
 - MRI撮像と同時
 - 光マイクロフォン
 - レコーダ : Marantz PMD-671
 - 標本化周波数44.1 kHz, 量子化16 bit

タスク

- 通常の話速を指示
- VCVの繰り返し発話（一息で続くだけ）
 - あがあが…, あかあか…
 - あだあだ…, あたあた…
 - その他（本発表では未使用）
- 文音声
 - 「ぼくの体を支えられるもの」
 - 「パパとママとぼくとで豆まきをします」
 - 「骨粗しょう症の座席調査票を複写します」

分析項目

1. 話速

- 1秒あたりのVCVの回数

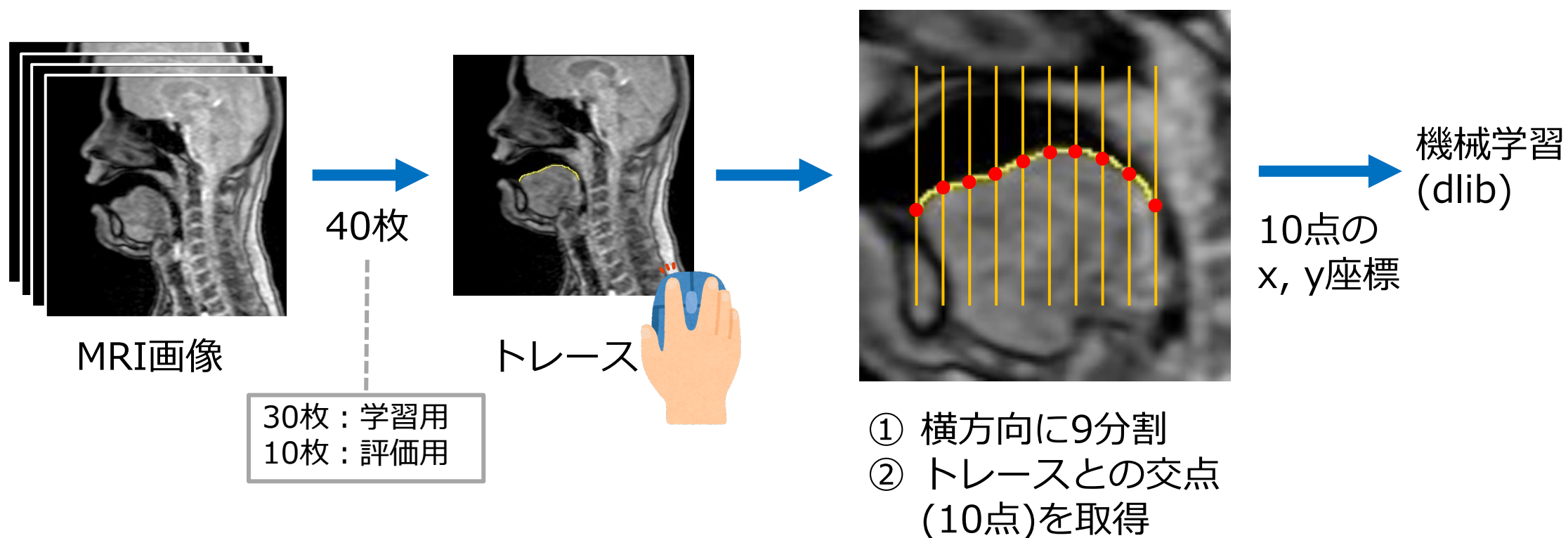
2. Relative tongue size

- 「パパとママとぼくとで豆まきをします」の「し」の/i/区間を利用(無声化しているが)

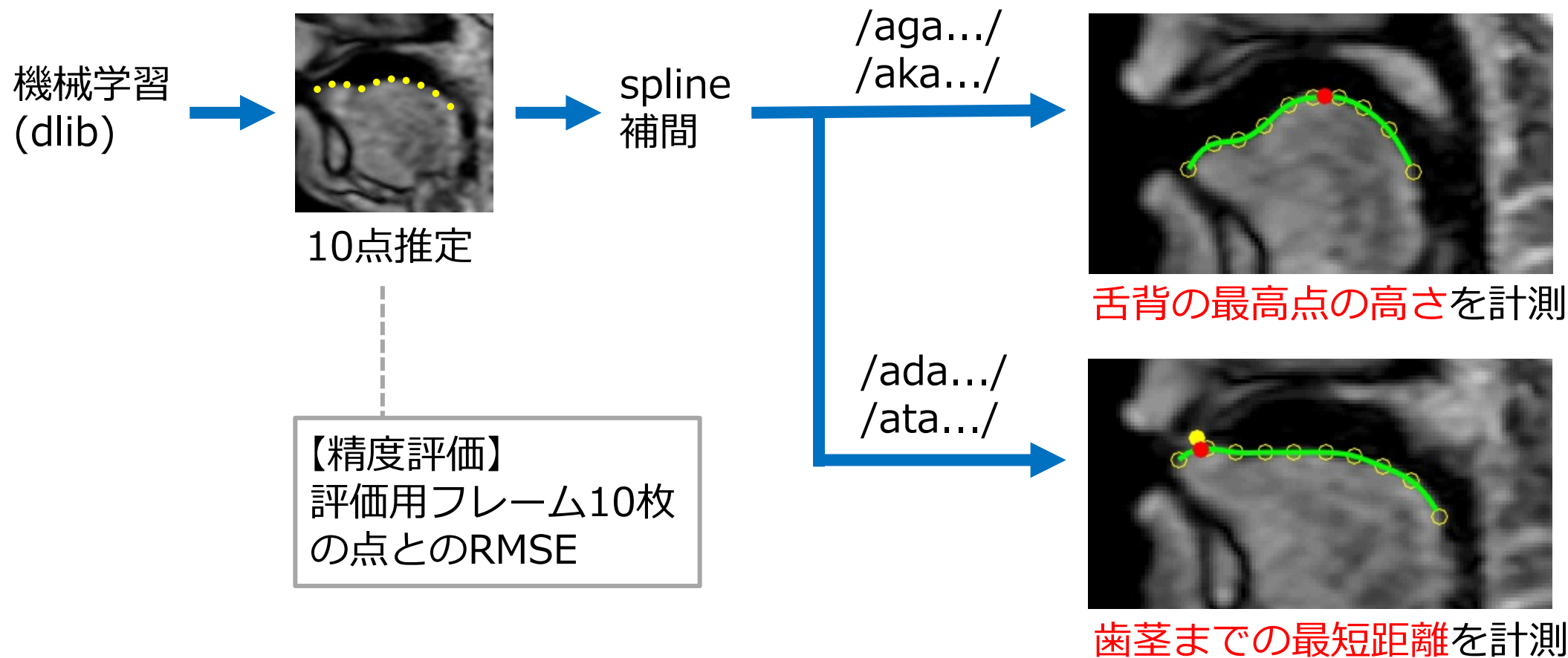
3. 舌の変位

- 運動の周期性を見るため
- 発話区間の舌の輪郭をトレースし, その上の10点を機械学習ライブラリdlib (King, 2009)で学習し, 舌の輪郭を抽出
- (/aga.../, /aka.../) 舌背の最高点の変化を観察
- (/ada.../, /ata.../) 歯茎から舌までの最短距離の変化を観察

舌の変位の計測(1)



舌の変位の計測(2)



結果(1) 話速

表：1 s あたりのVCVの回数 (回/s)

	S	NS1	NS2
あが	3.13	5.51	2.57
あか	3.57	4.59	2.86
あだ	5.00	5.03	3.48
あた	4.40	4.74	3.35



音声に撮像ノイズが重畳しています

結果(2)

Relative tongue size

「パパとママとぼくとで豆まきをします」の/i/区間を利用

表 : Relative tongue size

	RTS
S	60.7 %
NS1	73.8 %
NS2	70.1 %



S



NS1



NS2

Feng+ (2017) では54.17 %~67.57 %

結果(3)

dlibによる輪郭抽出精度

- 30フレームのデータで学習, 別の10フレームのデータで評価
- 舌上面輪郭上の10点のRMSE (10フレーム分を平均)

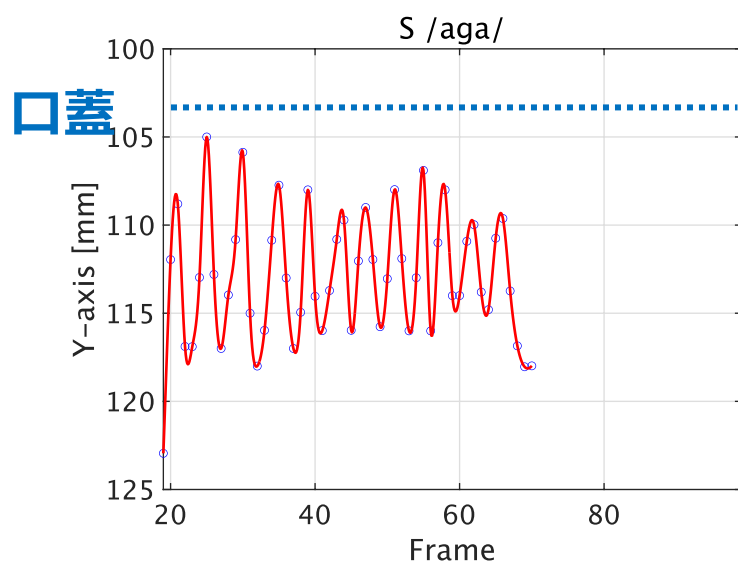
表 : RMSE (mm)

	S	NS1	NS2
あが・あか	1.63	1.48	1.57
あだ・あた	1.40	2.88	1.57

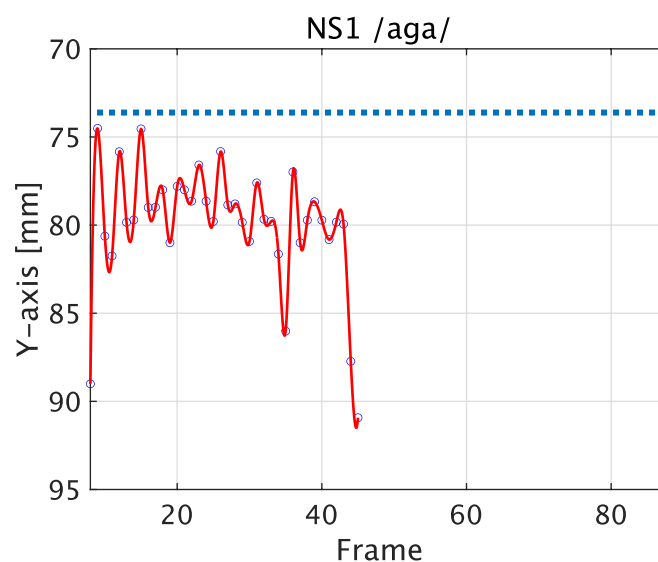
結果(4)

舌背の最高点の高さの変化(/aga.../)

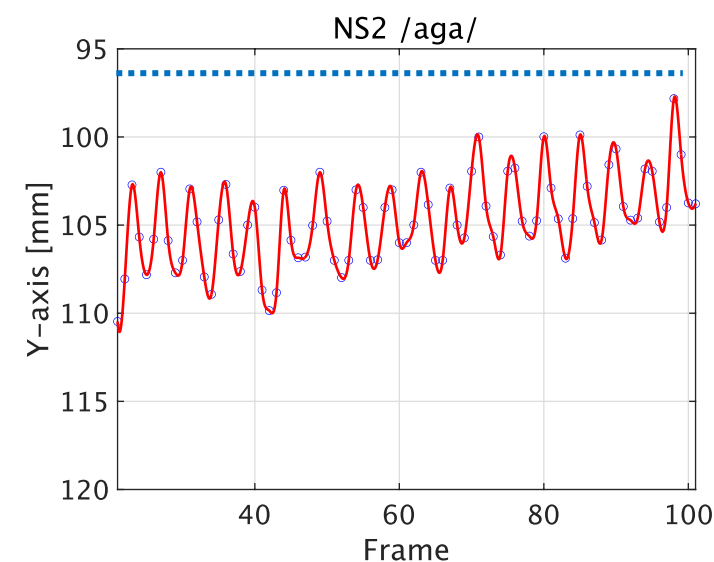
赤い線は各フレームの計測値をspline補間したもの



話者S



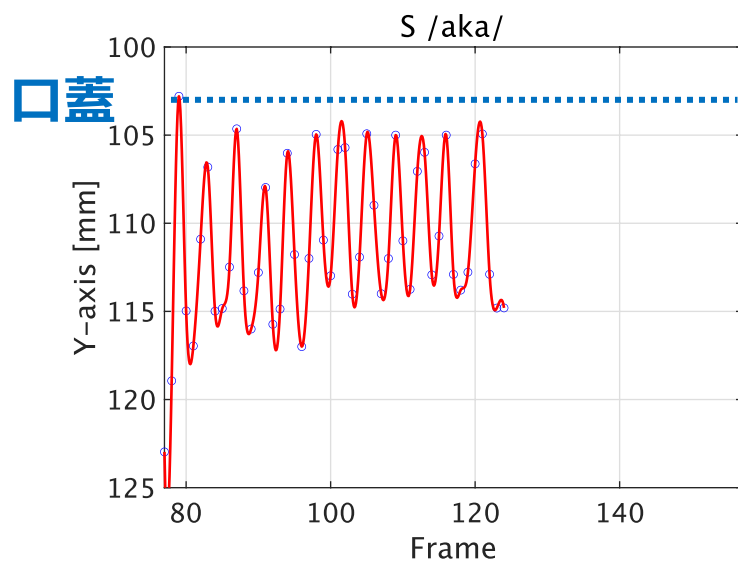
話者NS1



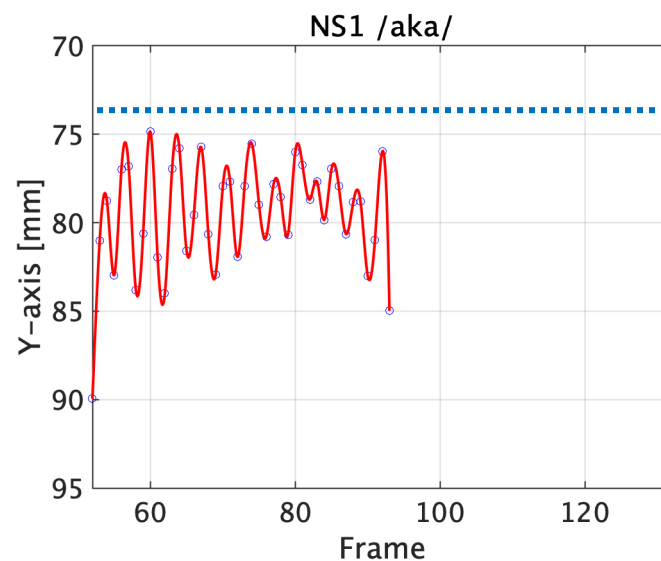
話者NS2

結果(5)

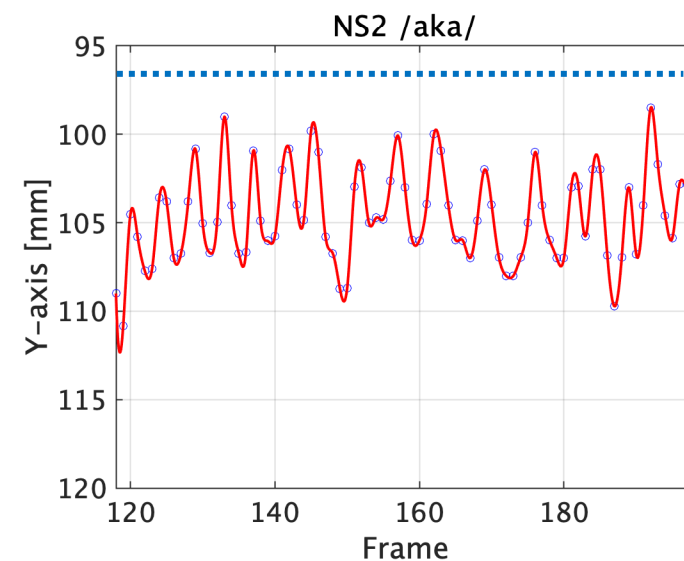
舌背の最高点の高さの変化(/aka.../)



話者S



話者NS1



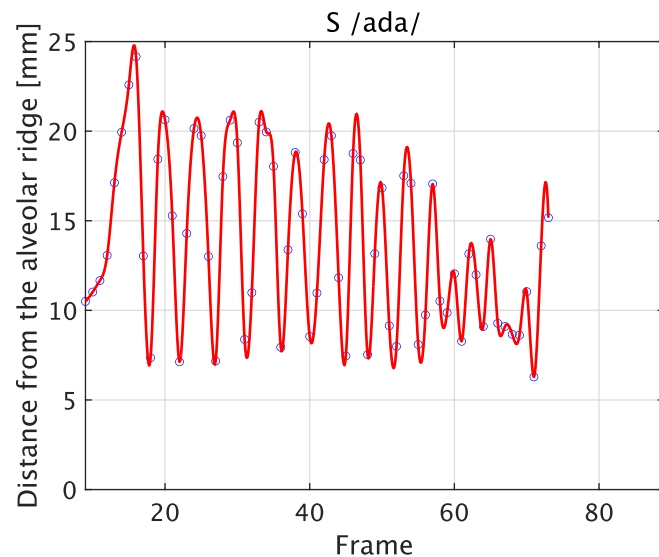
話者NS2



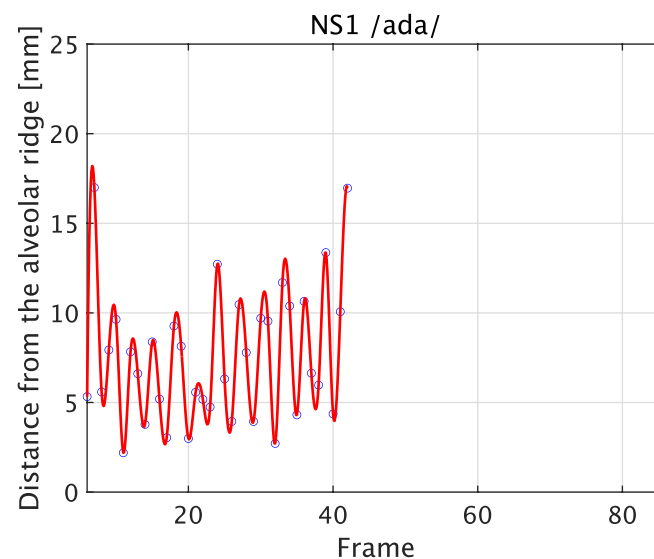
結果(6)

舌尖と歯茎の距離(/ada.../)

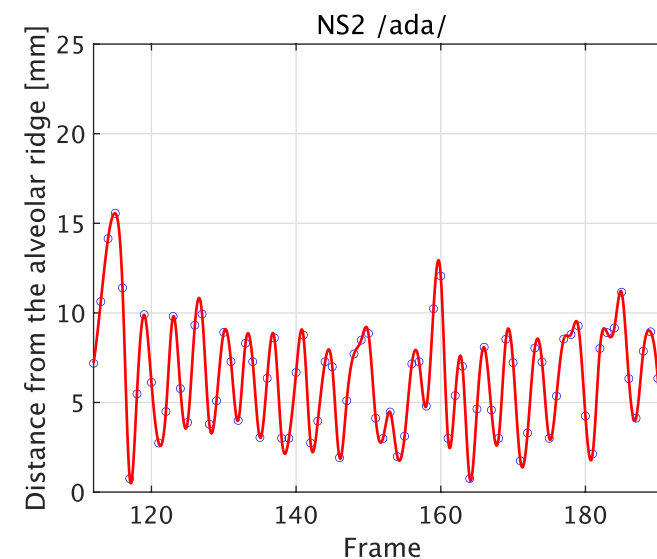
0 mmは接触に対応



話者S



話者NS1

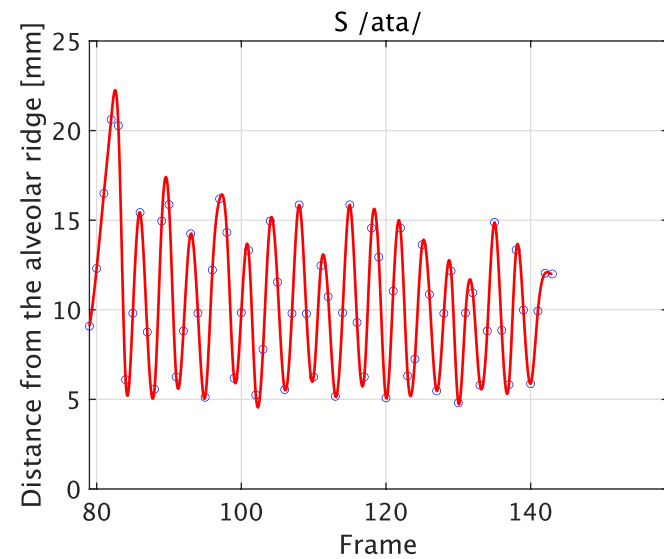


話者NS2

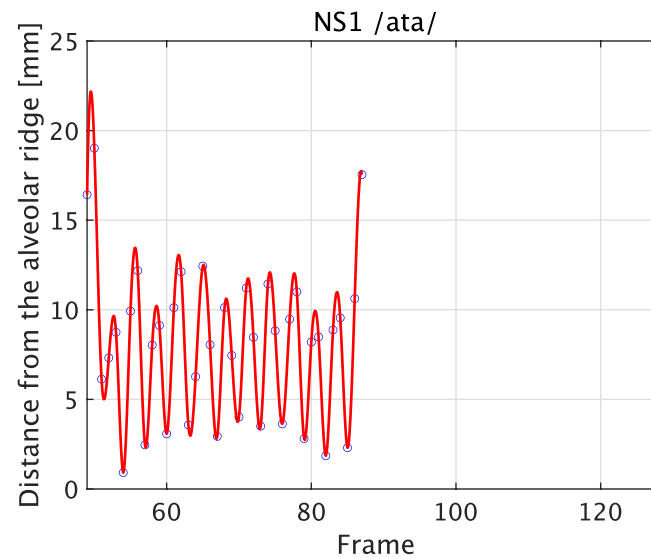


結果(7)

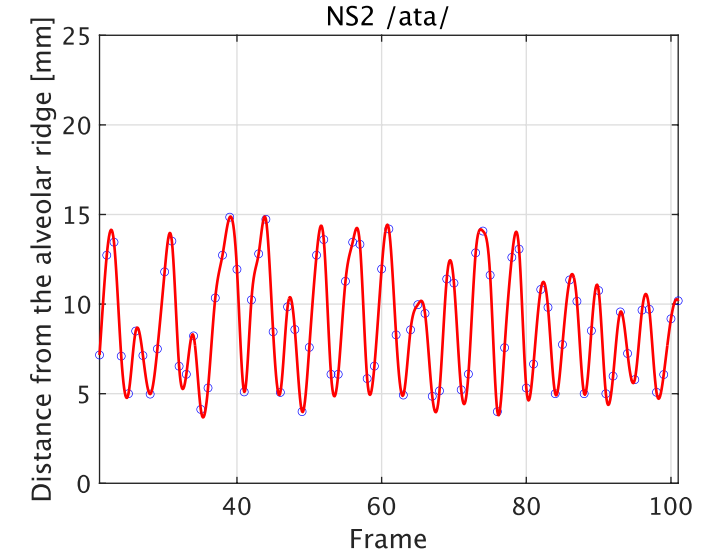
舌尖と歯茎の距離(/ata.../)



話者S 



話者NS1 



話者NS2 

考察

- Relative tongue sizeがS/NSの違いに関連する可能性がある。
 - Feng+ (2017)はRTSと舌の運動速度との相関を指摘している。
 - MRIを使わない方法として, Mallampati scoreも使える可能性がある。
- SとNSでは舌の運動の周期性に違いがある。
 - 立川+ (2015)の結果を支持
 - 調音器官の(リズムミクな)制御のスキルの差が現れているのでは？
 - 音声の明瞭性にも関連しそう
 - 発話のしにくさを自覚する人ほど聞き返されることが多いと感じている(北村+, 2019)ことと対応か
 - 自己相関関数やCepstral peak prominence (CPP)などで定量的に評価できそう
- 同様な情報を得る低コストな方法があるといい。
 - 音声？超音波？

まとめ

- 発話のしにくさを自覚する話者2名(NS1, NS2), 自覚しない話者1名(S)を対象に, MRI動画を用いて発話器官の静的, 動的な特徴を分析した.
- 話者数が少ないが, RTSや発話運動の周期性, 安定性が発話のしにくさの自覚に関連する可能性を指摘した.

本研究の一部はJSPS科研費 (20H00291, 16K13226) により行われた.