

言語研究における調音データの活用例 —超音波を用いた舌形状分析の実践—*

○溝口愛 (前橋工科大)

1 はじめに

近年、言語研究において超音波舌画像化 (Ultrasound Tongue Imaging: UTI) 技術の応用が進んでいる[1]。

本講演では、UTI による舌形状分析の実例を示し、分析方法について解説する。

2 活用例 1 舌輪郭比較

中国語を母語とする日本語学習者の日本語音節末鼻音/N/発話時の舌の調音位置を観察した研究[2]では、舌輪郭の比較により、学習者の調音における特徴が明らかにされた。

2.1 超音波画像データの収集

超音波診断装置 (MicrUs, EXT-1H) を使用し、口腔矢状断面のリアルタイム画像を録画した。プローブは、実験参加者が装着したヘッドギアに固定した。音声と超音波画像の録音・録画、および音声と動画の同期には、専用ソフトウェア AAA[3]を使用した。

分析対象語は、中国語の鼻音/n/, /ŋ/および比較対象となる非鼻音/t/を含む 4 単語、および日本語/N/を含む 1 単語とした。単語内の母音はすべて/a/であった。言語ごとにセッションを分け、分析対象外の単語 (中国語 25 単語、日本語 14 単語) を含め疑似ランダム順に各単語 10 回の発話を行った。

2.2 超音波画像データの分析

対象音素の音響的中間点に対応する舌動画フレームに対し、Matlab 上で動作するスクリプト GetContours[4]を使用して、舌輪郭をトレースした。各舌輪郭上の 100 点が、x/y 座標平面上の数値として抽出、保存された。

一般化加法モデル (GAM)[5]で、x における y の値を、話者ごと音素ごとに予測するモデルを作成した。モデルの舌輪郭を、x 軸上の観察された最大点と最小点でトリミングし、95%信頼区間と共にプロットした。

2.3 結果

Fig.1 は、話者 ACF03 の中国語発話時 (上図)、および中国語音節末/n/と日本語/N/発話時 (下図) の舌輪郭モデルである。

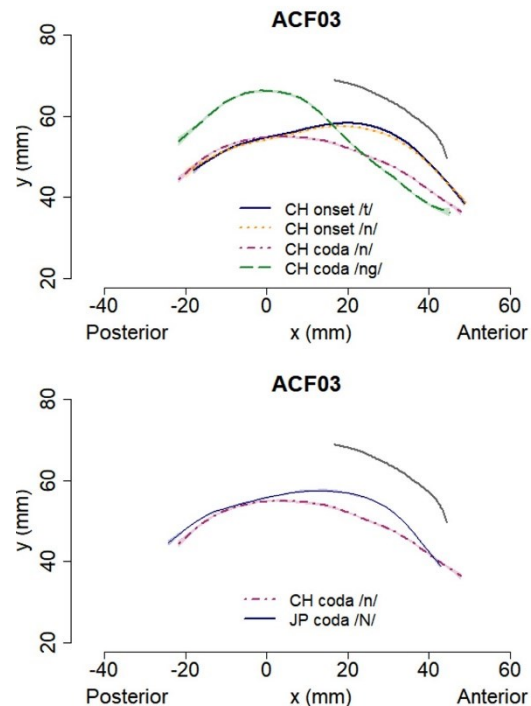


Fig.1 Tongue Contours Excerpt from [2].

中国語音節頭の/t/と/n/はほぼ同じ舌形状で発話しているのに対し、音節末の/n/はそれらよりも低い位置で調音していることがわかる。また、日本語の/N/は、中国語音節末/n/とは異なる舌形状で発音している。この話者は初級学習者であるが、日本語/N/が、中国語/n/とは違うことを認識しつつも、奥舌を使った調音ができていることが明らかとなった。

このように、舌輪郭をモデル化しプロットすることで、音素ごとの舌形状の比較が容易となる。

3 活用例 2 舌の最高点分析

日本語鼻音の音響分析と調音分析を行い、舌形状が鼻子音のフォルマントに及ぼす影響

* Utilizing articulatory data in linguistic research: Practice of tongue shape analysis using ultrasound, by MIZOGUCHI, Ai (Maebashi Institute of Technology).

を考察した研究[6]では、舌輪郭の最高点が分析に使用された。

3.1 超音波画像データの収集

超音波診断装置 Ultrasonix SonixTouch を使用し、口腔矢状断面のリアルタイム画像を録画した。同時に、モーションキャプチャーシステム Optotrak Certus Motion Capture System (NDI) を用いて、話者の頭の動きとプローブの動きをトラッキングした。実験終了後、超音波画像は HOCUS[7]により、頭の位置の補正が行われた。

3.2 超音波画像データの分析

GetContours[4]で、分析対象音素/m, n, ŋ, N/の音響的中間点における舌輪郭を抽出後、最高点を計測した。

3.3 音響分析

分析対象音素の第一～第三フォルマント (N1-N3) およびスペクトルの重心 (CoG) を計測した。

3.4 結果

計測した音響・調音情報を用いて多項ロジスティック回帰分析 (nnet[8]) を実施。話者および音素を予測するモデルを作成した。

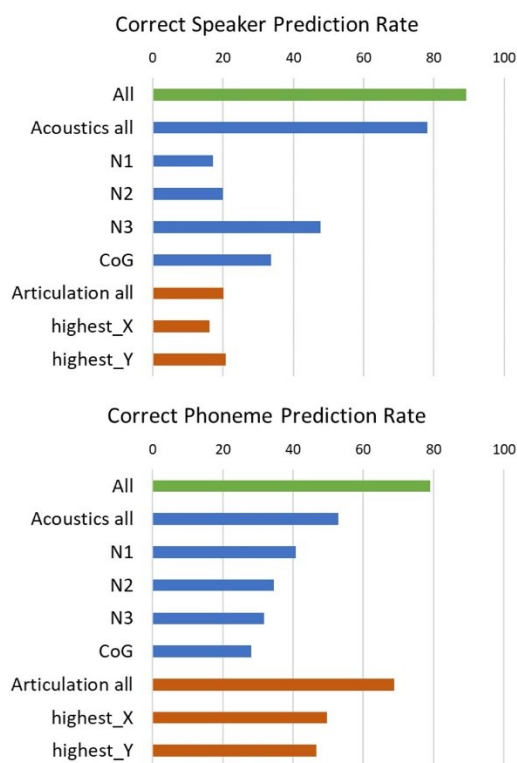


Fig.2 Correct Prediction Rate. [6]

Fig.2 は、音響・調音情報から発話者 (男性 4 名, 女性 6 名) を推定する各モデルの予測正解率 (上図), および音素 (/m, n, ŋ, N/) を推定する各モデルの予測正解率 (下図) である。話者の予測精度は、音響情報を用いたモデルの方が調音情報を用いたモデルより高く (上図), 音素の予測精度は、調音情報を用いたモデルの方が音響情報を用いたモデルより高い結果となった (下図)。N1 のみを用いたモデルでも、音素を 40.8% 正しく予測していることから、N1 にはある程度の調音位置情報が含まれていると推察される。

調音と音響の関係については、さまざまな議論があるが、調音データが不足しがちである。UTI による調音研究の寄与が期待される。

4 おわりに

超音波舌画像化で収集した調音データは、舌輪郭の視覚化や、舌の最高点を用いた予測モデルの作成など、様々な分析に使用可能であり、今後さらなる活用が見込まれる。

謝辞

本研究は JSPS 科研費 JP19K13254, JP23K12212 の助成を受けた。

参考文献

- [1] イアン・ウィルソン, (全訳) 荒井隆行, 石田真子, 溝口愛, 日本音響学会誌, vol. 70, no. 10, pp. 560–564, 2014.
- [2] Mizoguchi, Morimoto, Li, Arai, Proceedings of the 20th ICPhS, pp. 2622–2626, 2023.
- [3] Articulate Instruments Ltd., Articulate Assistant Advanced (AAA), 2022.
- [4] Tiede, GetContours. GitHub repository, <https://github.com/mktiede/GetContours>, 2022.
- [5] Wood, *Generalized additive models: an introduction with R*. Chapman and Hall/CRC, 2006.
- [6] Mizoguchi, Tiede, Whalen, HISPhonCog, 2023, pp. 68–69.
- [7] Whalen *et al.*, Journal of Speech Language and Hearing Research, vol. 48, no. 3, pp. 543–553, 2005.
- [8] Venables, Ripley, *Modern Applied Statistics with S, 4th ed.* New York: Springer, 2002