IS RHOTICITY ON THE TIP OF YOUR TONGUE? INVESTIGATING TONGUE SHAPES FOR ENGLISH /r/ IN FRENCH LEARNERS WITH ULTRASOUND

Alice Léger, Hannah King, Emmanuel Ferragne

Université Paris Cité, CLILLAC-ARP, F-75013 Paris, France
alice.leger@etu.u-paris.fr, hannah.king@u-paris.fr, emmanuel.ferragne@u-paris.fr

ABSTRACT

English /r/ takes different tongue shapes from one speaker to another. It is well-established that *tip-down* and *tip-up* shapes produce perceptually similar outputs. However, it remains unclear why speakers intuitively acquire one type or another. The present study considers the hypothesis that rhotic and non-rhotic varieties of English may influence the acquisition of different tongue shapes. We provide articulatory data on the pronunciation of English /r/ by 19 French learners of English, 10 with rhotic and 9 with non-rhotic accents. Ultrasound tongue images were recorded for onset and coda /r/ in various vocalic contexts and were classified as either tip-up or tip-down. Although rhoticity as a predictor of tongue shape does not reach statistical significance, we found a tendency for rhotic speakers to use a higher proportion of tip-down shapes. We conclude that while rhoticity may partly influence tongue shape, other factors are also at play, including co-articulatory constraints.

Keywords: rhotic, articulation, English, French learners

1. INTRODUCTION

The English post-alveolar approximant /r/ has intrigued researchers for its unusual articulatory variability. Native speakers produce equivalent acoustic outputs, at least up to the third formant [1], with various degrees of retroflex and bunched tongue shapes [2, 3]. Retroflex configurations are *tip-up* and raised towards the hard palate, while bunched configurations are *tip-down*, with the dorsum retracted. However, it remains unclear why individuals develop tip-down and/or tip-up shapes. Across studies, some consistent coarticulatory patterns of tongue shape distribution emerge, with tip-down shapes typically occurring with high front vowels [3, 4, 5] and in coda position [6]. Physiological differences such as oral cavity size or palate domedness have also been suggested to affect articulatory variability [7, 8, 9]. Sociolinguistic factors are reported in Scottish English, with tip-down and tip-up variants prevailing in middle-class and working-class speakers, respectively [10]. However, given the lack of perceptual distinction between tongue configurations in American English [11], the social distribution of /r/ may rather be due to a perceptible temporal contrast between tongue shapes specific to Scottish English [12].

The present study examines Heyne et al.'s [13] hypothesis that accent rhoticity may influence tongue shape acquisition. Most research on /r/ has focused on rhotic Englishes [2, 3, 10], but more recent studies on non-rhotic Englishes [13, 4] indicate a stronger preference for tip-down patterns in rhotic than in non-rhotic accents. As tip-down shapes are preferred in postvocalic position, non-rhotic varieties, which do not have /r/ in that context, may therefore discourage the development of a tip-down variant [13].

To investigate the influence of rhoticity on tongue shape acquisition, we looked at the articulation of French learners of English. Despite non-rhotic Standard Southern British English being traditionally taught in French schools [14], we observe both rhotic and non-rhotic accents among students. Learners' articulations of English /r/ are largely understudied, except for Polish [15] and Mandarin learners [16]. And yet, with its complex set of three constrictions at the lips, palate and pharynx [2, 17, 18], English /r/ is likely a challenging new sound for learners. Indeed, /r/ is one of the latest sounds to be acquired in native English-speaking children [19, 20]. As suggested by the Speech Learning Model (SLM) [21, 22] and the Perceptual Assimilation Model (PAM) [23, 24], influence from the L1 system is expected when learning a non-native sound. Nevertheless, the French rhotic /ʁ/ appears phonetically distinct enough from English /r/ to prevent the assimilation of the two sounds in French learners. Substitutions for [w], closer in articulation, also remain rare [25].

As language classes in France rarely include pronunciation training, the acquisition of lingual gestures for /r/ in French learners appears rather implicit. Interestingly though, King and Ferragne [4] were not able to find mentions of a tip-down variant in pronunciation manuals. Thorough descriptions of tip-up constrictions alone are provided, with Ashton and Shepherd [26] notably considering it the 'correct position' relative to the tip-down alternative.

We explored the tongue shape patterns acquired by French learners of English in order to determine whether rhoticity contributes to the acquisition of a specific lingual gesture. If rhoticity indeed influences acquisition, we predict to find more tip-down patterns in learners with a rhotic accent and more tip-up patterns in learners with a non-rhotic accent. In addition, we expect learners to develop similar preferences to native speakers in the distribution of tongue shapes. That is, a stronger preference for tip-down shapes (i) surrounding high-front vowels, and (ii) in coda position.

2. METHODOLOGY

2.1. Participants

Data were collected from 20 French learners of English (13F, 7M), aged between 18 and 32. All were students at Université Paris Cité and had a rhotic (R) or non-rhotic (NR) accent. Rhoticity was perceptually assessed prior to the experiment based on audio recordings of 12 English sentences containing 22 possible productions of coda /r/. One subject's data was excluded due to ultrasound visualisation issues. We provide results from 10 rhotic and 9 non-rhotic learners. The study was granted ethical approval by the Université Paris Cité's ethics board.

2.2. Procedure

Participants read a randomised list of 14 monosyllabic words in isolation, each one repeated three times. /r/ occurred in word-initial and word-final position in various vocalic contexts, before /iː, ɪ, e, æ, ɑː, ɒ, ʌ, ɔː, uː/ and after /ɪ, e, ɑː, ɜː, ɔː/. Ultrasound tongue images were collected using an Echo Blaster 128 unit with a 5-8MHz transducer along with audio data in the Articulate Assistant Advanced (AAA) software [27]. The probe was stabilised relative to the head using an UltraFit Headset [28]. Before recording, a bite plate was used to adjust the probe-to-chin angle to each speaker's occlusal plane to improve interpretation of tongue position [29]. A Focusrite Scarlett 6i6 sound card and a Sennheiser t.bone Earmic500 condenser microphone were used for audio recording.

2.3. Data processing

For each /r/ produced, ultrasound frames were manually extracted at the point of maximal constriction and were visually classified as tip-up (TU) or tip-down (TD). Sequential frames were holistically examined to better assess the position of the tongue tip, as in [3, 4]. Tongue contours in the frames depicting maximal constriction were tracked automatically in AAA and manually corrected when necessary. The resulting splines were extracted in Cartesian coordinates, yielding 42 data points per token. Missing values at either end of a tongue contour were removed whereas missing values inside contours were reconstructed based on a moving average filter with a span of 4 data points. We present tongue contour data from 17 speakers as technical issues prevented us from tracking the contours in 2 speakers. Among the remaining 714 contours, 119 were discarded as no data point could be extracted. The size of our final 595 tokens ranged from 12 to 32 data points ($\mu = 24.16, \sigma = 3.43$).

One of the statistical models we used (Section 2.4) included discrete cosine transform (DCT) coefficients (estimated from the Y coordinates of each contour) as predictors. We computed these in an attempt to summarise overall tongue shapes with as few parameters as possible, which might constitute an alternative characterization of tongue shapes when data points at the tip are not available. Earlier uses of the DCT in phonetics have applied to the modelling of formant trajectories [30]. Here, based on data from speaker 128, we observed that an inverse transform with the first four coefficients offered a good approximation of the original contours and therefore kept these four coefficients for our model.

2.4. Statistical analysis

A binary decision tree was fitted to the data with speaker, speaker rhoticity, word, and syllable position (onset vs coda) as predictors and tongue shape (TU vs TD) as response. The accuracy of the model was computed using 10-fold cross-validation. The split criterion was Gini's diversity index. The maximum number of splits was constrained to 4 in order to make for optimal interpretability.

Tongue shape classification was further assessed using binomial generalised linear mixed-effects models (GLMM). The maximal set of successfully converging random slopes and intercepts were included for speakers and where possible, words. The significance of main effects was tested using likelihood ratio tests. Indications of significance in the final models were calculated with Satterthwaite's approximations for degrees of freedom.

3. RESULTS & DISCUSSION

Figure 1 displays the tongue contours in 17 speakers. TU shapes appear in blue and TD in yellow. Speakers are sorted according to rhoticity (R vs NR). We observed more TU than TD shapes overall. Just two speakers (151, 214) systematically used TD shapes, both of whom belonged to the rhotic group.

In total, 666 ultrasound images of /r/ (252 in non-rhotic speakers; 414 in rhotic) were classified as either TU or TD. 3.2 % and 34.1 % of the tokens were classified as TD in the non-rhotic and rhotic group, respectively. As non-rhotic speakers rarely produced coda /r/ (11 tokens), the proportion of TD shapes in onset /r/ was also evaluated. Onset /r/ was classified as TD in 3.3 % of the tokens in non-rhotic and 24.4 % in rhotic speakers. These results follow our predictions that TD shapes are more common in rhotic than in non-rhotic speakers and that within rhotic speakers, there is a stronger preference for TD shapes in coda than in onset position.

The cross-validated binary decision tree model reached 96 % accuracy. As Figure 2 shows, the model picked speaker and word as relevant predictors and left out speaker rhoticity and syllable position. Although the model did not use rhoticity as a predictor, we note that all the speakers predicted to use a TD shape (128, 151, 214, 308) were in the rhotic group. This suggests that while having a rhotic accent does not mean that a speaker will systematically use a TD shape, TD shapes are much more typical of rhotic than of non-rhotic speakers. In the two speakers with variable tongue shapes (128, 308), TD shapes are predicted to occur in coda and only with high front vowels (/iː, ɪ/) in onset.

As rhotic and non-rhotic speakers all produced onset /r/, we ran a GLMM with tongue shape as the response variable on the onset data, regressed against speaker rhoticity and vocalic context. Random intercepts for subjects were included. While the main effect of rhoticity was not significant [$\chi^2(1) = 0.12$, $p = 0.729$], vocalic context was significant [$\chi^2(8) = 89.93$, $p < .001$]. The final model indicated that for an average speaker, a TD tongue shape is significantly more likely to occur in the context of /iː/ than it is for /e, æ, ɑː, uː, ʌ, ɔː, ɒ/. No significant difference was observed between /iː/ and /ɪ/. Although rhoticity failed to reach significance, these results follow our prediction that TD shapes are more likely to occur with high-front vowels, as observed in previous studies.

To assess the impact of syllable position, a further GLMM was run with tongue shape as the response variable on the data from rhotic speakers only. Main effects included syllable position and vocalic context. Random intercepts were included for subjects and words. Both main effects were significant [position: $\chi^2(1) = 11.84$, $p < .001$; vowel: $\chi^2(8) = 15.65$, $p < .05$]. The resulting model showed that for the average rhotic speaker, TD tongue shapes are significantly more likely to occur in coda than in onset, following our initial prediction. Regarding the effect of vocalic context, the same results as in the previous onset model were observed, again following our intial prediction that surrounding high-front vowels (/iː/ and /ɪ/) are more likely to result in TD shapes than the other vowels in the dataset.

As discussed in Section 2.3, DCT coefficients were computed in an attempt to summarise overall tongue shapes. Tongue tip height was also calculated by extracting the Y value of the rightmost Cartesian coordinate in each tongue contour. We then explored how these five numerical variables (DCTs 1-4 & tip height) may predict our manual coding of tongue shape in a final GLMM. Tongue shape as the response variable was regressed against the five numerical variables, which were converted to z-scores. Random intercepts were included for speakers. The addition of random intercepts for words failed to converge. The following main effects were significant: DCT1 [$\chi^2(1) = 20.99$, $p < .001$]; DCT3 $\chi^2(1) = 47.29$, $p < .001$]; DCT4 [$\chi^2(1) = 17.02$, $p < .001$] and tip height [$\chi^2(1) = 28.52$, $p < .001$]. While DCT2 failed to reach significance [$\chi^2(1) = 0.43$, $p = 0.514$]. Tokens with higher DCT1, higher DCT3, lower DCT4 and lower tongue tip values were significantly more likely to be coded as TD. With regards to tongue tip height, these results are coherent, as TD tongue shapes should naturally have a lower tongue tip than TU shapes. Higher DCT1 - sometimes referred to as the zeroth coefficient [30] - values suggest that the mean Y value in TD is higher. The absence of a difference in DCT2 tells us there is no evidence supporting a potential distinction in terms of magnitude and direction of change from the mean. The difference in DCT3 indicates that the degree of U-shaped curvature is greater in TD contours. And lower DCT4 values in TD show that TU contours contain energy in the higher frequencies, i.e. potential abrupt changes in trajectories. While our use of the

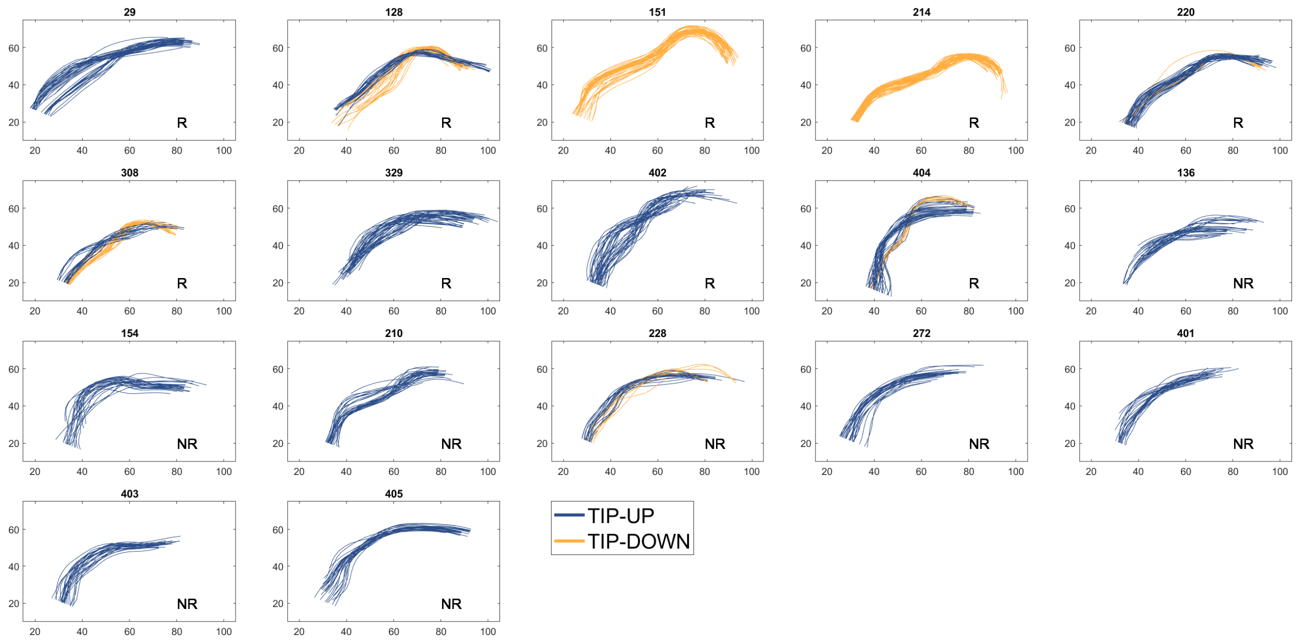

Figure 1: Tongue shapes in 17 participants. R: rhotic; NR: non-rhotic.

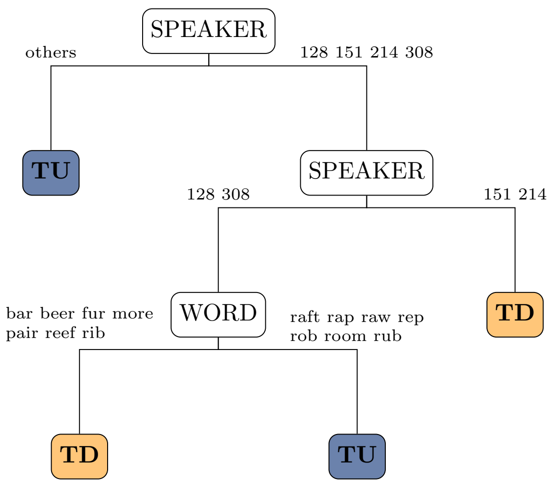

Figure 2: Classification tree.

DCT here remains exploratory, and more evidence has yet to be provided as to the benefits of such an approach, we must note that the DCT produced a compact and potentially interpretable representation of our contours.

4. CONCLUSION

This study aimed to evaluate if rhotic and non-rhotic accents constrain lingual gesture acquisition of /r/ in French learners of English. Although our results predict no significant effect of rhoticity, rhotic learners used proportionally more tip-down configurations than non-rhotic learners. In addition, speakers who used a fully tip-down pattern all belonged to the rhotic group. Tip-down shapes were significantly more frequent in coarticulation with high front vowels and in coda position, thus mirroring tendencies found in native speakers. We conclude that while having a rhotic accent does not mean that a speaker will systematically use tip-down shapes, tip-down shapes occur more frequently in rhotic than in non-rhotic speakers. This native-like production of a non-native sound suggests that the acquisition of /r/ lingual gestures is highly constrained by mechanical or physiological factors. Given that tip-up shapes prevailed across groups regardless of these preferences, we suggest, following Mielke et al. [3], that tip-up shapes may be the default variant, and tip-down gestures context-specific.

5. REFERENCES

[1] X. Zhou, C. Y. Espy-Wilson, S. Boyce, M. Tiede, C. Holland, and A. Choe, "A magnetic resonance imaging-based articulatory and acoustic study of retroflex and bunched American English /r/," *The Journal of the Acoustical Society of America*, vol. 123, no. 6, pp. 4466–4481, 2008.

[2] P. Delattre and D. C. Freeman, "A dialect study of American r's by X-ray motion picture," *Linguistics*, vol. 6, no. 44, pp. 29–68, 1968.

[3] J. Mielke, A. Baker, and D. Archangeli, "Individual-level contact limits phonological complexity: Evidence from bunched and retroflex

/ɹ/," *Language*, vol. 92, no. 1, pp. 101–140, 2016.

[4] H. King and E. Ferragne, "Loose lips and tongue tips: The central role of the /r/-typical labial gesture in Anglo-English," *Journal of Phonetics*, vol. 80, p. 100978, 2020.

[5] D. Ong and M. Stone, "Three dimensional vocal tract shapes in /r/ and /l/: A Study of MRI, Ultrasound, Electropalatography, and Acoustics," *Phonoscope*, vol. 1, no. 1, pp. 1–13, 1998.

[6] J. M. Scobbie, E. Lawson, S. Nakai, J. Cleland, and J. Stuart-Smith, "Onset vs. coda asymmetry in the articulation of English /r/," in *Proceedings of the 18th International Congress of Phonetic Sciences*, Glasgow, UK, 2015.

[7] J. Brunner, S. Fuchs, and P. Perrier, "On the relationship between palate shape and articulatory behavior," *The Journal of the Acoustical Society of America*, vol. 125, no. 6, pp. 3936–3949, 2009.

[8] S. Bakst, "Palate shape influence depends on the segment: Articulatory and acoustic variability in American English /ɹ/ and /s/," *The Journal of the Acoustical Society of America*, vol. 149, no. 2, pp. 960–971, 2021.

[9] D. Dediu and S. R. Moisik, "Pushes and pulls from below: Anatomical variation, articulation and sound change," *Glossa: a journal of general linguistics*, vol. 4, no. 1, 2019.

[10] E. Lawson, J. M. Scobbie, and J. Stuart-Smith, "The social stratification of tongue shape for postvocalic /r/ in Scottish English," *Journal of Sociolinguistics*, vol. 15, no. 2, pp. 256–268, 2011.

[11] A. Twist, A. Baker, J. Mielke, and D. Archangeli, "Are "covert" /ɹ/ allophones really indistinguishable?" *Penn Working Papers in Linguistics*, vol. 13, no. 2, p. article 16, 2007.

[12] E. Lawson, J. M. Scobbie, and J. Stuart-Smith, "Bunched /r/ promotes vowel merger to schwar: An ultrasound tongue imaging study of scottish sociophonetic variation," *Journal of Phonetics*, vol. 41, no. 3-4, pp. 198–210, 2013.

[13] M. Heyne, X. Wang, D. Derrick, K. Dorreen, and K. Watson, "The articulation of /ɹ/ in New Zealand English," *Journal of the International Phonetic Association*, vol. 50, no. 3, pp. 366–388, 2020.

[14] A. Przewozny-Desriaux, "Research Environments For Pedagogical Tools: Teaching and Learning with Authentic Oral English," *Recherche et pratiques pédagogiques en langues de spécialité*, no. Vol. 37 N°2, 2018.

[15] P. Lyskawa, "The ultrasound study of /ɹ/ in non-native speakers," in *Proceedings of International Congress of Phonetic Sciences*. London, UK: International Phonetic Association, 2015.

[16] S. Chen and P. P. K. Mok, "Speech production of rhotics in highly proficient bilinguals: Acoustic and articulatory measures," in *Proceedings of the 19th International Congress of Phonetic Sciences*, Melbourne, Australia, 2019.

[17] A. Alwan, S. Narayanan, and K. Haker, "Toward articulatory-acoustic models for liquid approximants based on MRI and EPG data. Part II. The rhotics," *The Journal of the Acoustical Society of America*, vol. 101, no. 2, pp. 1078–1089, 1997.

[18] M. K. Tiede, S. E. Boyce, C. K. Holland, and K. A. Choe, "A new taxonomy of American English /r/ using MRI and ultrasound," *The Journal of the Acoustical Society of America*, vol. 115, no. 5, pp. 2633–2634, 2004.

[19] A. B. Smit and L. Hand, "The Iowa Articulation Norms Project and its Nebraska Replication," *Journal of Speech and Hearing Disorder*, vol. 55, pp. 779–798, 1990.

[20] L. Cialdella, H. Kabakoff, J. L. Preston, S. Dugan, C. Spencer, S. Boyce, M. Tiede, D. H. Whalen, and T. McAllister, "Auditory-perceptual acuity in rhotic misarticulation: Baseline characteristics and treatment response," *Clinical Linguistics & Phonetics*, vol. 35, no. 1, pp. 19–42, 2021.

[21] J. E. Flege, "Second language speech learning: Theory, findings, and problems," *Speech perception and linguistic experience: Issues in cross-language research*, vol. 92, pp. 233–277, 1995.

[22] J. E. Flege and O.-S. Bohn, "The revised speech learning model (slm-r)," *Second language speech learning: Theoretical and empirical progress*, pp. 3–83, 2021.

[23] C. T. Best and M. D. Tyler, "Nonnative and second-language speech perception, commonalities and complementarities," in *Language Experience in Second Language Speech Learning: In Honor of James Emil Flege*. John Benjamins Publishing, 2007, pp. 13–34.

[24] C. T. Best, "A direct realist view of cross-language speech perception," *Speech perception and linguistic experience*, pp. 171–206, 1995.

[25] I. Trifu-Dejeu, N. Herry-Bénit, and S. Lopez, "Le /r/ de l'anglais dans le corpus IPCE-IPAC : Évaluation quantitative et qualitative des productions orales des apprenants francophones," *Anglophonia*, no. 32, 2021.

[26] H. Ashton and S. Shepherd, *Work on your accent*, 2nd ed. Glasgow: Collins, 2020.

[27] Articulate Instruments Ltd., *Articulate Assistant Advanced User Guide: Version 2.19*. Edinburgh, UK: Articulate Instruments Ltd., 2012.

[28] L. Spreafico, M. Pucher, and A. Matosova, "UltraFit: A Speaker-friendly Headset for Ultrasound Recordings in Speech Science," in *Proceedings of Interspeech*, Hyderbad, India, 2018, pp. 1517–1520.

[29] E. Lawson, J. Stuart-Smith, and L. Rodger, "A comparison of acoustic and articulatory parameters for the GOOSE vowel across British Isles Englishes," *The Journal of the Acoustical Society of America*, vol. 146, no. 6, pp. 4363–4381, 2019.

[30] J. Elvin, D. Williams, and P. Escudero, "Dynamic acoustic properties of monophthongs and diphthongs in Western Sydney Australian English," *The Journal of the Acoustical Society of America*, vol. 140, no. 1, pp. 576–581, 2016.