INTERSPEECH 2016
September 8–12, 2016, San Francisco, USA

L2 Acquisition and Production of the English Rhotic Pharyngeal Gesture

Sarah Harper[1], *Louis Goldstein*[1], *Shrikanth Narayanan*[2]

[1] Department of Linguistics, University of Southern California, USA
[2] Department of Electrical Engineering, University of Southern California, USA
skharper@usc.edu, louisgol@usc.edu, shri@sipi.usc.edu

Abstract

This study is an investigation of L2 speakers' production of the pharyngeal gesture in the English /ɹ/. Real-time MRI recordings from one L1 French/L2 English and one L1 Greek/L2 English speaker were analyzed and compared with recordings from a native English speaker to examine whether the gestural composition of the rhotic consonant(s) in a speaker's L1, particularly the presence and location of a pharyngeal gesture, influences their production of English /ɹ/. While the L1 French speaker produced the expected high pharyngeal constriction in their production of the French rhotic, he did not appear to consistently produce an English-like low pharyngeal constriction in his production of English /ɹ/. Similarly, the native Greek speaker did not consistently produce a pharyngeal constriction of any kind in either his L1 rhotic (as expected) or in English /ɹ/. These results suggest that the acquisition and production of the pharyngeal gesture in the English rhotic approximant is particularly difficult for learners whose L1 rhotics lack an identical constriction, potentially due to a general difficulty of acquiring pharyngeal gestures that are not in the L1, the similarity of the acoustic consequences of the different components of a rhotic, or L1 transfer into the L2.

Index Terms: phonetics, speech production, articulation, second language acquisition

1. Introduction

The present study is an investigation of the production of the pharyngeal component of the English rhotic approximant by second language (L2) speakers, examining whether the inclusion of a pharyngeal constriction is part of the gestural constellation in their L1 rhotic and, if is present, whether the location of this constriction within the pharynx may influence their production of English /ɹ/.

Although the exact tongue posture observed in the formation of English /ɹ/ is highly variable between speakers and contexts, with over a dozen possible tongue postures ranging from "retroflex" to "bunched" attested among native English speakers [1], its gestural composition is stable across these different tongue postures. In most dialects of American English, the approximant /ɹ/ is produced by the coordination of three supralaryngeal constrictions: a labial constriction, a lingual constriction in the palatal region, and a pharyngeal constriction in which the tongue root retracts towards the pharyngeal wall [2]. According to perturbation theory, these three constrictions conspire to generate the most salient acoustic correlate of English /ɹ/, its low F3 value. This can be explained as a consequence of the fact that the location of each of its characteristic supralaryngeal constrictions (the lips, the palatal region and the pharynx) corresponds to an F3 velocity maximum in a neutral tube (cf. [3]). Considering that all three of these constrictions do, to some extent, generate similar acoustic consequences in the production of English /ɹ/, it raises the question as to whether second language (L2) English speakers may fail to acquire one of these gestures due to misperception of the underlying gestural representation of /ɹ/. Furthermore, since all of these gestures have a similar acoustic consequence, the question emerges of whether a failure to produce one of them would have a noticeable impact on the perceived authenticity of their production of the English rhotic.

In addition, the acquisition of the pharyngeal gesture by an L2 speaker may be influenced by the speaker's L1 production experience. This may be particularly relevant in acquiring the pharyngeal gesture in English /ɹ/, as sounds involving gestures in this region of the vocal tract tend to pose significant perceptual and/or control challenges for L2 learners. For example, the claim that the failure of many non-native Arabic speakers to accurately produce pharyngeal consonants arises from learners' difficulties in both correctly perceiving and reproducing these sounds is supported by cases in which the sound used as a substitute for a pharyngeal consonant is acoustically distinct (i.e., the tendency for learners to produce a glottal stop instead of the voiced pharyngeal fricative [ʕ]) [4]. Considering this, it seems likely that the acquisition of the pharyngeal component of English /ɹ/ would pose a particular challenge to adult L2 learners, depending on their L1 experience, regardless of whether or not they accurately perceive its presence as a gestural component of English /ɹ/ [5, 6].

One of the principal components that characterizes rhotics is the presence of a pharyngeal gesture [7]; this parametric specification of that gesture can vary, including a high ("dorsopalatal") pharyngeal constriction, such as the canonical French rhotic and many of its allophones, and a low ("epiglotto-pharyngeal") constriction [8] like English /ɹ/. As such, learners' success in acquiring the pharyngeal gesture in English /ɹ/ may differ depending on whether their native language is like Greek, which has a rhotic consonant (an alveolar tap) without a pharyngeal constriction [9], or like French, which has a rhotic consonant with a constriction at a different location in the pharynx than the English rhotic.

Given both the aforementioned difficulty for L2 speakers to learn to produce consonants involving pharyngeal gestures and the observation that L1 articulatory patterns may influence the production of similar L2 sounds [10], we may predict that one or both of the non-native English speakers may fail to produce a low pharyngeal constriction in their production of /ɹ/. This could indicate that the speakers have failed to accurately acquire this gesture as part of their mental representation of the L2 sound, either because of a failure to

http://dx.doi.org/10.21437/Interspeech.2016-658

perceive its presence or due to the influence of their L1 gestural patterns, or that speakers have acquired the mental representation of this gesture but have failed to produce it due to the influence of L1 articulatory patterns. Alternatively, if both the French and Greek participants produce English /ɹ/ with a low pharyngeal constriction, this would indicate that they have accurately acquired this component of the segment, although they may conceivably still differ from native English speakers with respect to the relative timing or magnitude of the gesture.

2. Methods

2.1. Materials and Participants

Recordings from three male speakers were used in this study: one L1 English speaker (ENG), one L1 French/L2 English speaker (FR), and one L1 Greek/L2 English speaker (GK). All three were recorded reading the fable "The North Wind and the Sun" in English and, when applicable, in their native language. All speakers were living in the United States at the time they were recorded, and the two non-native speakers were both highly proficient in English. The recordings were made using a rtMRI protocol developed for the analysis of speech production [11]. Participants were recorded speaking while lying on their back, and were imaged in the mid-sagittal plane with a resolution of 68 x 68 pixels and a reconstructed frame rate of 23.18 frames per second. A total of 230 tokens of rhotic consonants were collected from these recordings, with 30 to 56 collected for each speaker in each language.

2.2. Analysis

A region of interest technique was used to estimate tongue movement within different regions of the pharynx during rhotic production [12]. Two pseudo-circular regions with a radius of three pixels were manually defined along an automatically derived vocal tract midline: one ('DOR') high in the pharynx, to measure movement in the region of the pharynx associated with the dorsopalatal constriction in the French rhotic, and the other ('PHAR') in the lower half of the pharynx to measure movement associated with the pharyngeal constriction in the English rhotic approximant (Figure 1).

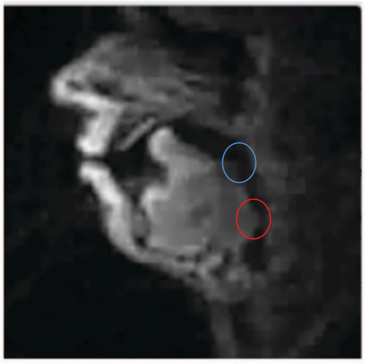

Figure 1: *Image of vocal tract during English /ɹ/ constriction (speaker ENG) with regions of interest overlaid (DOR = blue, PHAR = red).*

The average pixel intensity inside each region was calculated for each frame, with higher pixel intensity indicating a greater amount of tissue in the region [13]. In order to facilitate the comparison of pixel intensity values across different speakers and across different recordings of the same speaker, each measurement was scaled relative to the minimum and maximum pixel intensity values recorded for that ROI within the recording the measurement was taken from. The pixel intensity time series were smoothed using a locally weighted linear regression technique with a relatively tight kernel width of h=.9 [14]. The onset and maximum constriction of the relevant lingual gestures were found using an algorithm developed in [15], with movement onset defined as the frame where the articulator velocity reached 10% of its maximum velocity. For tokens where temporal landmarks were hard to identify due to the reduced magnitude of the gesture, DOR and PHAR measurements were taken from the points in time that corresponded to the onset and maximum constriction of the tongue tip constriction gesture.

In addition to measuring the pixel intensity within DOR and PHAR at the frame where the maximum constriction of each gesture was identified, a measure indexing movement displacement of the tongue root was also calculated by finding the mean pixel intensity difference between the gestural onset and the point of maximum constriction.

3. Results

A series of linear mixed effects models (LMEMs) were run independently to statistically evaluate whether there were differences in the realization of the low pharyngeal gesture in English /ɹ/ production for each speaker (a between-speakers comparison). The dependent variable in all cases was pixel intensity within either the DOR or the PHAR region, while the variable SPEAKER (ENG, FR and GK) was included as a fixed factor. Additionally, to account for the influence of vocalic context on rhotic production, tokens were coded for preceding vowel context; the decision to code for vowel context based on the identity of the preceding vowel was made in anticipation of greater perservative coarticulatory effects on rhotic production than anticipatory effects. VOWEL (front or back) was included as a random effect in all statistical models.

3.1. Production Differences in L1 Rhotics

The results of a between-languages comparison of mean pixel intensity values in the pharyngeal ROI during rhotic production for each speaker in their L1 are shown in Figure 2. As mentioned in section 3.2, higher values for this measurement (closer to 1) indicate that more tongue tissue is present within this region at a specific time point, suggesting a greater degree of constriction. As we can see, the English speaker (ENG) has noticeably higher pixel intensity values within this region as compared to the French and Greek speakers in their first languages, as was expected considering only English has a low pharyngeal constriction out of these languages. Similarly, the low pixel intensity values observed for GK speaking in his native Greek are also expected, as the Greek rhotic alveolar tap is not characterized by any pharyngeal constriction.

An LMEM run to examine between-speakers comparisons of L1 data returned significant effects of both SPEAKER and VOWEL (both $p < .001$), with a post hoc pairwise comparisons of means revealing significant differences between all three speakers for VOWEL, and significant differences between the L1 English speaker ENG and each of the other speakers (GK and FR) along the dimension SPEAKER.

The SPEAKER x VOWEL interaction was not found to be significant.

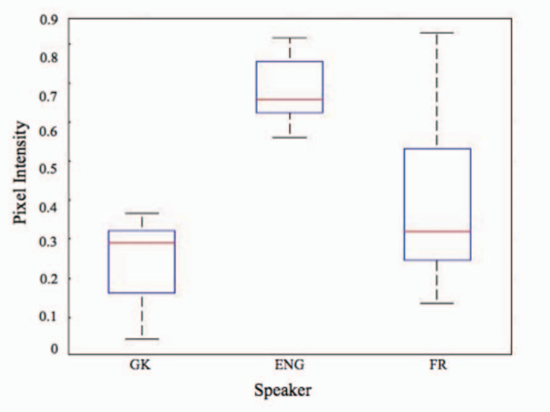

Figure 2: *PHAR pixel intensity values for each speaker at the point of maximum constriction in their L1 rhotic*

Although the mean pixel intensity value for the French speaker, FR, is relatively similar to the Greek speaker GK and lower than the English speaker ENG, the range of values measured in this region is noticeably wider for him than for the other speakers. This is most likely a consequence of the lingual motion necessary to achieve the expected high pharyngeal constriction in this language, as movement of the lower tongue root may be required to achieve the desired higher constriction. Additionally, the observed variability is likely due at least in part to interaction with flanking vowels, especially given the significant effect of VOWEL.

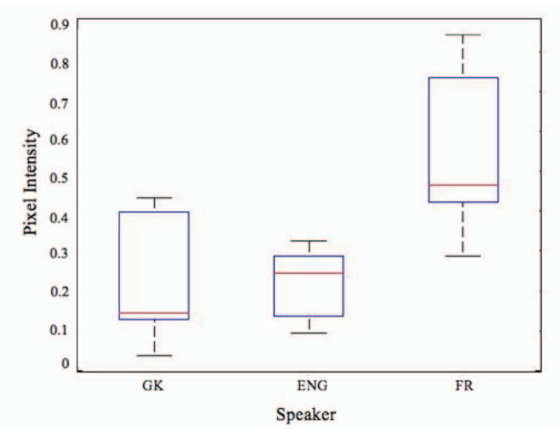

Figure 3: *DOR pixel intensity values for each speaker at the point of maximum constriction in their L1 rhotic*

In Figure 3 we see the results of a between-speakers comparison of mean pixel intensity values in the dorsal ROI during rhotic production for each speaker in their L1. Although not as robust a contrast as that observed in the Pharyngeal ROI, the LMEM for this comparison returned a significant effect of both SPEAKER and VOWEL (both $p < .001$), with no significant interaction between SPEAKER and VOWEL. Post-hoc pairwise comparisons of means revealed that along the dimension SPEAKER, the L1 Greek speaker GK and the L1 English speaker ENG did not differ significantly, although significant differences were found when comparing both speakers to the L1 French speaker FR. As can be seen in Figure 3, pixel intensity values for the French speaker are noticeably higher in this region than for either of the other speakers, which follows from the fact that of these languages only the French rhotic utilizes a high pharyngeal constriction.

3.2. Production of the English Rhotic

The results of a between-speakers comparison of PHAR pixel intensity values during English /ɹ/ production are shown in Figure 4. As we can see, there are robust differences between the measurements recorded for each speaker, with the native English speaker ENG displaying the overall highest pixel intensity values in this ROI, the L1 Greek speaker GK displaying the lowest values, and the L1 French speaker FR displaying values somewhere in the middle. The LMEM for between-speaker comparisons of English data returned significant effects of both SPEAKER and VOWEL (both $p < .001$), with significant differences found between all speakers in post hoc pairwise comparisons of means. The SPEAKER x VOWEL interaction was not found to be significant.

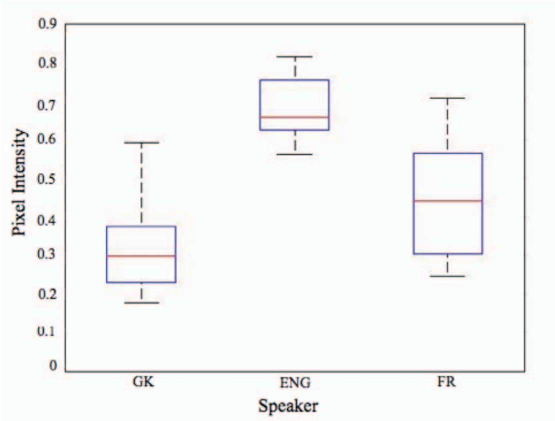

Figure 4: *PHAR pixel intensity values for each speaker at the point of maximum constriction in English /ɹ/*

Crucially, this mirrors the pattern observed for each speaker's pixel intensity measurements in this ROI in their native language (as discussed in section 3.1). In particular, the continued observation of relatively low pixel intensity in this region for GK, the native Greek speaker, seems to suggest that he does not include a low pharyngeal gesture in his production of English /ɹ/. Additionally, while the mean pixel intensity observed for the native French speaker FR is noticeably higher than in was in his L1, it is still significantly lower than the measurements obtained for the native English speaker ENG in this region.

The results of a between-speakers comparison of pixel intensity values in the DOR ROI during English /ɹ/ production, although not as robust as the differences observed for the PHAR ROI, still allow the observation that FR exhibits a much larger range of constriction degrees in his production of English /ɹ/ than either the native English speaker or the L1 Greek speaker (Figure 5). This seems to suggest that, for at least some tokens, a dorsopalatal constriction similar to that observed in the French rhotic is being created. An LMEM run to examine between-speaker comparisons of L1 data returned a significant effect of SPEAKER ($p = .0035$), but not of VOWEL ($p = .2414$). Post-hoc pairwise comparisons of means revealed that, as was observed for the L1 data, there was a significant difference in the mean values along this dimension for the L1 French speaker when compared to both of the other speakers, while the mean values recorded for the L1 Greek and L1 English speakers did not differ significantly from one another. The SPEAKER x VOWEL interaction was not found to be significant.

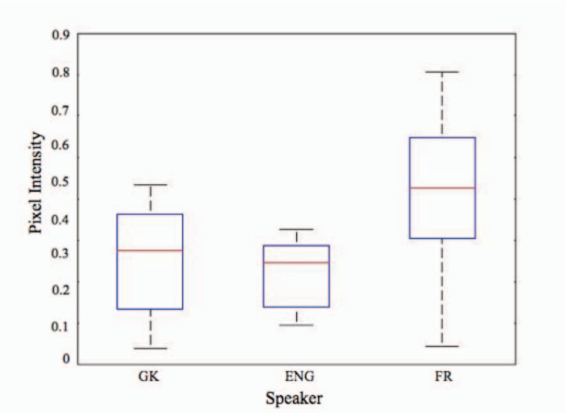

Figure 5: *DOR pixel intensity values for each speaker at the point of maximum constriction in English /ɹ/ (L to R: L1 Greek, L1 English, L1 French)*

4. Discussion

The results of this investigation provide preliminary evidence that, at least for some L2 English speakers, the pharyngeal gesture in /ɹ/ has not been acquired, or at the very least is not being produced. In theory, this could be due to the learners' failure to adequately perceive this gesture, leading to its absence from their mental representation of English /ɹ/. Alternatively, the absence of a pharyngeal gesture in non-native speakers' production of English /ɹ/ may be due to difficulty in the production of a pharyngeal gesture in an L2 environment that is absent from the L1 (i.e., implementing a pharyngeal gesture in the production of an L2 category when the corresponding L1 category does not include such a gesture).

While neither of the L2 English speakers in this study seem to produce English /ɹ/ in a native-like manner, as judged by the lack of a low pharyngeal gesture in their production, it is less clear whether there is direct influence from the production of pharyngeal gestures in their L1 rhotic on their production of English /ɹ/. For example, the measurements obtained for FR in the DOR ROI while speaking English were variable enough that a clear pattern of transfer from the L1 to the L2 failed to emerge. Although the mean for FR's DOR pixel intensity values is roughly equivalent in both English and their L1, French, the overall distribution of values for this speaker in English extends far lower than in French, suggesting the occasional absence of a high pharyngeal gesture in his production of the English rhotic. However, the observation that this range of values is still rather large and overlaps considerably with his French DOR measurements seems to indicate that FR is, at least for some tokens, producing a higher pharyngeal constriction similar to that incorporated in his L1 rhotic. Additionally, while GK's failure to produce any consistent pharyngeal gesture in either Greek or English may indicate L1 transfer, it may also be the consequence of a general difficulty acquiring new pharyngeal gestures in an L2.

Interestingly, although the evidence as a whole seems to point to neither of the non-native English speakers producing an English-like pharyngeal gesture in their production of English /ɹ/, it is worth noting that the mean pixel intensity value measured within the PHAR region for FR was higher when he was speaking in English than when he was speaking in his native language, French (Figures 2 and 4). While this may be due, in whole or in part, to differences in the stimuli speakers were presented with in each language (as the read speech from which the data was drawn was not controlled in such a way to guarantee similar vocalic contexts for tokens in different languages), it also presents the interesting possibility that these speakers are at least attempting to approximate the appropriate pharyngeal gesture in their production of English /ɹ/. As demonstrated in previous work on L2 speech production, the influence of the L1 on the production of L2 sounds is pervasive even at high proficiency levels, with learners apt to produce L2 categories as intermediate to their own L1 production and native speaker norms for the target language. A more in-depth analysis, with more carefully controlled stimuli, will be necessary to gain a fuller understanding of the acquisition process for the pharyngeal gesture and reveal whether or not an intermediate degree of constriction is truly observed for speakers attempting to approximate this gesture.

One final question that has not yet been addressed in this investigation, but will be evaluated perceptually in an extension of this project, is whether the absence of a pharyngeal gesture in the production of English /ɹ/ has a noticeable effect on the perception of this sound as "native-like" in its production. Impressionistically, the English rhotic produced by the L1 Greek speaker, GK, did not sound native-like in the least. However, the L1 French speaker, FR, does have a more native-sounding English /ɹ/ despite the evidence that he is not producing, or at least not consistently producing, the low pharyngeal gesture that characterizes this sound. This seems to indicate that the pharyngeal gesture is necessary for a native-like English /ɹ/; however, it may be that the presence of *any* pharyngeal gesture, whether it be low or high, may suffice in combination with the other gestures comprising the English /ɹ/ constellation to produce the acoustic effects of a native-like segment.

5. Conclusion

This study examined the potential for the presence or location of a pharyngeal constriction in L2 English speakers' L1 rhotics to influence their production of English /ɹ/. Neither the L1 French speaker, whose native language contains a high pharyngeal gesture, nor the L1 Greek speaker appeared to consistently produce an English-like low pharyngeal constriction in their production the English rhotic. These results suggest that the acquisition and production of the pharyngeal gesture in the English rhotic approximant is particularly difficult for learners whose L1 rhotics lack an identical constriction, although further research will be necessary to gain a fuller understanding of the factors that influence and impede the acquisition process for the pharyngeal gesture.

6. Acknowledgements

This research was supported by NIH grant DC007124 and a USC Provost's Ph.D. Fellowship to the first author.

7. References

[1] M. Tiede, S. Boyce, C. Holland and K. Choe, "A new taxonomy of American English /r/ using MRI and ultrasound", *Journal of the Acoustical Society of America*, vol. 115, pp. 2633, 2004.

[2] A. Alwan, S. Narayanan and K. Haker, "Toward articulatory-acoustic models for liquid approximants based on MRI and EPG data. Part II: The rhotics", *Journal of the Acoustical Society of America*, vol. 101, pp. 1078-1089, 1997.

[3] C. Espy-Wilson, S. Boyce, M. Jackson, S. Narayanan, and A. Alwan, "Acoustic modeling of American English /r/", *Journal of the Acoustical Society of America,* vol. 108, pp. 343-356, 2000.

[4] A. M. Elgendy, *Aspects of Pharyngeal Coarticulation*. Ph.D. dissertation, University of Amsterdam, 2001.

[5] J. E. Flege, "The production of 'new' and 'similar' phones in a foreign language: Evidence for the effect of equivalence classification", *Journal of Phonetics*, vol. 15, pp. 47-65, 1987.

[6] J. E. Flege, "Second language speech learning: Theory, findings and problems". In *Speech Perception and Linguistic Experience: Theoretical and Methodological Issues*, W. Strange (ed), pp. 233-277. Baltimore: York Press, 1995.

[7] T. J. Magnuson, "The Story of /r/ in Two Vocal Tracts". In J. Trouvain and W. J. Barry (eds.), *Proceedings of the 16th International Congress of Phonetic Sciences*, pp. 1193-6, 2007.

[8] P. Delattre, "Pharyngeal features in the consonants of Arabic, German, Spanish, French and American English", *Phonetica*, vol. 23, pp. 129-155, 1971.

[9] A. Arvaniti, "Illustrations of the IPA: Modern Greek", *Journal of the International Phonetic Association,* vol. 19, pp. 167-172, 1999.

[10] M. Olsen, "The L2 Acquisition of Spanish Rhotics by L1 English Speakers: The Effect of L1 Articulatory Routines and Phonetic Context for Allophonic Variation", *Hispania*, vol. 95, 1, pp. 65-82, 2012.

[11] S. Narayanan, K. Nayak, S. Lee, A. Sethy, and D. Byrd, "An approach to real-time magnetic resonance imaging for speech production", *Journal of the Acoustical Society of America*, vol. 115, 4, pp. 1771–1776, 2004.

[12] M. Proctor, A. Lammert, A. Katsamanis, L. Goldstein, C. Hagedorn and S. Narayanan, "Direct Estimation of Articulatory Kinematics from Real-time Magnetic Resonance Image Sequences". *Proc. Interspeech* Florence, pp. 281–284, 2011.

[13] A. Lammert, V. Ramanarayanan, M. Proctor, and S. Narayanan, "Vocal tract cross-distance estimation from real-time MRI using region-of interest analysis". *Proc. Interspeech* Lyon, pp. 959-962, 2013.

[14] A. Lammert, L. Goldstein, and K. Iskarous, "Locally-weighted regression for estimating the forward kinematics of a geometric vocal tract model", *Proc. Interspeech* Makuhari, pp. 1604-1607, 2010.

[15] M. Tiede, "MVIEW: Multi-channel visualization application for displaying dynamic sensor movements", 2010.