

The Articulation of the Palatal Gesture in American English [ɹ]

Khalil Iskarous*

¹Haskins Laboratories, 300 George St., New Haven, Connecticut 06511, U.S.A.

iskarous@haskins.yale.edu

***Abstract.** This study investigates the variability in the location of the palatal constriction in the variants of the American English [ɹ]. X-ray Microbeam data from 10 subjects were analyzed. The location of the palatal constriction was determined by first computing the oral portion of the midsagittal distance function. The basic result is that there is indeed variability in the placement of the palatal constriction. It is shown that this variability can be derived and understood within the framework of Articulatory Phonology.*

1. Introduction

A long-standing debate in the speech production literature concerns the nature of the control variable in the speech motor control process (Löfqvist, 1997). Some researchers hold that the goal in speech production is to control an acoustic pattern (Guenther et al. 1998), while others hold that the goal is to control the parameters of a constriction in the vocal tract (Saltzman and Munhall, 1989; Browman and Goldstein, 1990). This is a very important theoretical question, since it concerns the essence of how speech production is planned. One segment that has been consistently used to illustrate the acoustic target theory is the American English [ɹ]. It has been known for a long time that several configurations of the tongue can be used to produce this segment (Delattre and Freeman, 1968), the two most well known of which are “retroflex” and “bunched”. But these different varieties of articulation all seem to result in a low F3. The American English [ɹ] is used as the example par excellence of acoustic target theories since it seems to exhibit a many-to-one relation between articulation and acoustics. The argument used is that the target of speech production would likely to be set at a level that exhibits less variability. The articulator level exhibits great variability, as shown by the many positions possible of the front of the tongue, while less variability is exhibited at the acoustic level, as shown by the consistently low F3. Therefore, the argument goes, the acoustic stage is more likely to be the target level on which speech motor control is planned.

It has also been known, however, at least since Delattre and Freeman (1968), that the articulatory maneuvers of the tongue in American English [ɹ] result in three

*Supported by NIH grants NIDCD DC-02717 and 006705.

constrictions: pharyngeal, palatal, and labial. The palatal constriction is the one that can be accomplished using many different articulatory maneuvers, using different regions of the front portion of the tongue. It therefore seems that there are two levels on which we can analyze the speech production process. The first level relates the movement and orientations of particular organs or parts of organs, and the second level concerns the locations in the vocal tract where constrictions are made (Saltzman and Munhall, 1989; Guenther, 1998). This distinction between levels is not only a convenient analytic tool. Linguistic patterning in the world's languages reveals that there is indeed a distinction between two levels in speech production. Most linguistic contrasts can be expressed in terms of the location and degree of constriction, and it does not matter which articulator accomplishes the constriction. For instance, in English, some speakers produce [s] with a laminal articulation, whereas others produce it with an apical articulation (Ladefoged & Maddieson, 1996). But in some languages, two segments can contrast based on which articulator is used, even though the place of articulation and degree of the constriction are the same. For instance, in Eastern Arrente, there is a contrast between apical and laminal articulations of the same fricative with the same constriction location and degree (Ladefoged & Maddieson, 1996). There is therefore linguistic support for distinguishing two levels in the speech production process. There maybe other levels as well like muscle length level, but such levels do not seem to be required through linguistic evidence.

To understand the target of speech production, therefore, we must analyze variability at both levels of speech production, and compare this variability with the variability at the acoustic level. The main goal of this paper is to analyze variability in constriction location for the palatal constriction of American English [ɹ] at the constriction level. We expect to find variability at this level, since it is known that there exists acoustic variability in the different variants of [ɹ], and an acoustic difference has to be based on some articulatory difference at the constriction level, since the acoustic pattern generated by the vocal tract is simply an acoustic interpretation of the vocal tract configuration. Examples of the acoustic differences found include the following: Hagiwara (1995) showed that tip-up [ɹ] has higher F3 than tip-down [ɹ], Epsy-Wilson & Boyce (1999) showed differences in F4 and F5 for the different [ɹ] shapes, and Hashi et al. (2003) showed that parameters of the F3 transitions were different for the different [ɹ] shapes. Another goal of this paper is to present a theoretical framework that makes sense of constraints on variability at both the level of articulator used and at the constriction level.

In Section 2, the X-ray Microbeam data used in this study will be described. Section 3 will show that there are small, but consistent, differences in the locations of the palatal constriction, depending on which articulator is used. In section 4, it will be argued that the difference in palatal constriction location is a coarticulatory effect similar to the fact that the velar constriction for [k], for instance, is more interior if followed by [i] than by [u]. Moreover, it will be shown that the Articulatory Phonology Framework (Browman and Goldstein, 1990) allows us to predict the constraints on variability at the articulatory and constrictional levels.

2. Methods

X-ray Microbeam data are used in this study to analyze tongue motion during American English [ɹ]. The data were extracted from the X-ray Microbeam (XRMB) Speech Production Database (Westbury, 1994) and consist of 5 repetitions of the words *row* and *across* from 10 speakers, 5 male and 5 female. These utterances were chosen, since retroflex articulations are most likely to surface in the onset position of *row*, while bunched articulations are most likely to be used after the [k] in *across*. But there is also subject-to-subject and token-to-token variability to consider. The male utterances are a subset of the data analyzed by Hashi et al. (2003). The data consist of movement of 4 markers on the Tongue (T1-T4) at a sampling rate of 145 Hz. Details on the placement of the pellets can be found in Westbury (1994). Of the 100 tokens (10 subjects x 10 tokens), 17 tokens were missing due to the mistracking of one of the tongue markers.

In this study, we are interested in the location of the palatal constriction, which is hard to determine from the location of the markers alone. A cubic spline was therefore fit to the markers in each frame. Figure 1 shows two examples of the transition from [ɹ] to the following vowel by same subject. The markers are plotted along with the cubic spline fits to the markers. Spline fits have been shown to provide a good estimate of the real outline of the tongue (Kaburagi & Honda, 1994).

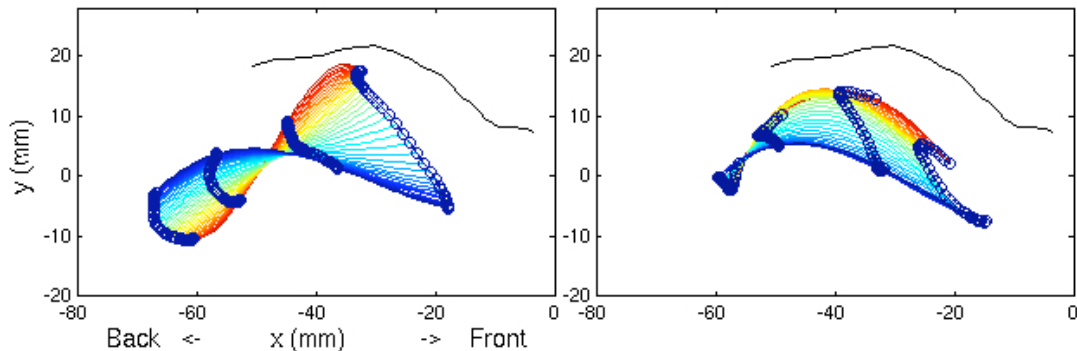


Figure 1. (Left) Retroflex articulation of [ɹ] in *row* by Subject 24 (male). (Right) Bunched articulation of [ɹ] in *across* by same subject. The first frame is red and the last is blue.

To determine the location of the palatal constriction, the midsagittal distance function was computed. The midline of the oral cavity was determined by using a bisection algorithm (Story et al., 1996) starting with the hard palate trace and a neutral tongue edge, selected from a frame before the subject starts speaking. A grid is then formed by generating lines perpendicular to the midline. The grid was determined only in the

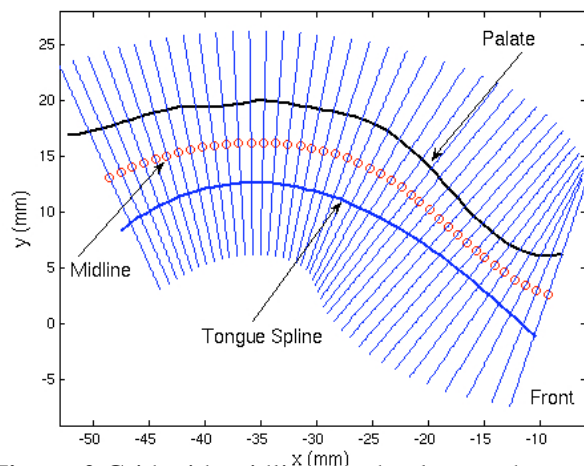


Figure 2 Grid with midline, hard palate, and tongue spline superimposed. Subject 28 (male).

area of interest, which is the oral cavity. Figure 2 shows an example grid from a male subject.

The midsagittal distance function and its change in time during the articulation were then determined by measuring the distance from the tongue surface to the hard palate at each gridline. One problem that often occurs is that the tongue tip is drawn into the vocal tract so that there is no intersection between the tongue and the most anterior gridlines. To be able to find the midsagittal distance at these locations, it is necessary to know the location of the floor of the mouth, which is not available. In this study, the midsagittal distance function in such cases is simply left undefined. Figure 3 shows two examples of the change in the midsagittal distance function as a function of time in the oral cavity for the transitions in Figure 1. The horizontal axis represents time and the vertical axis represents the location in the vocal tract, where gridline 1 is most anterior. The focus of this paper is on the location of the palatal constriction at the moment of greatest excursion of the tongue in the palatal region, which was determined by finding the frame at which zero upward velocity is achieved. The location of the palatal constriction was then found by automatically determining the minimum in the midsagittal distance function in the oral region and picking the gridline at which the minimum occurs.

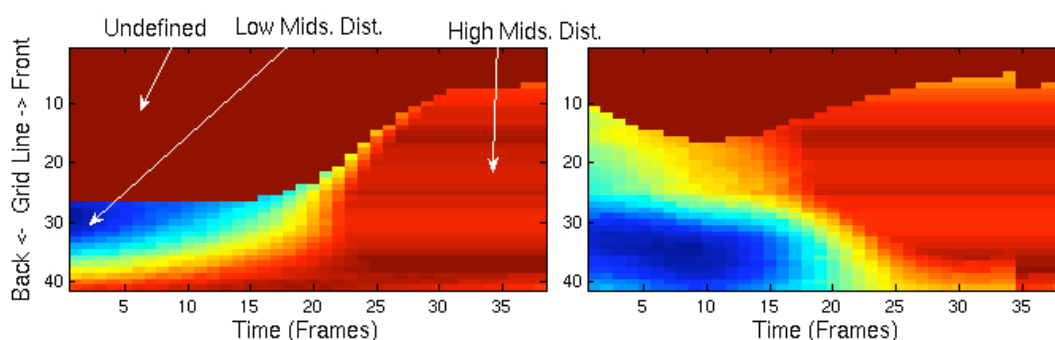


Figure 3. Change in Midsagittal Distance Function for data in Figure 1. Dark blue signifies low value for the Midsagittal Distance Function, whereas red signifies a high value.

3. Results

The goal of the analysis is to investigate if, on average, the palatal constriction for the [ɹ] in *row* was more forward in the oral cavity than the [ɹ] in *across*, since the first is likely to be made with the tongue tip and the anterior portion of the tongue body. Based on visual inspection, eight of the ten subjects included in this study used a retroflex articulation in *row* and a bunched articulation in *across*. Subject 33 (female) used bunched articulations in both, but even in this case, visual inspection reveals that there is a small difference in the placement of the palatal constriction when the [ɹ] is after [k] and when it is not. Figure 4 presents the data for the female (left) and male (right) subjects. The basic result is that there is indeed variability in the location of the constriction of the palatal constriction between the different articulatory variants. A single factor ANOVA was run on the male and female data separately. There is

significant difference between the two variants analyzed for both females ($p < .05$) and males ($p < .05$). The average difference in location of the constrictions is about .65 cm across all subjects. Variation in the degree of constriction degree variation was not measured in this study, but needs investigation.

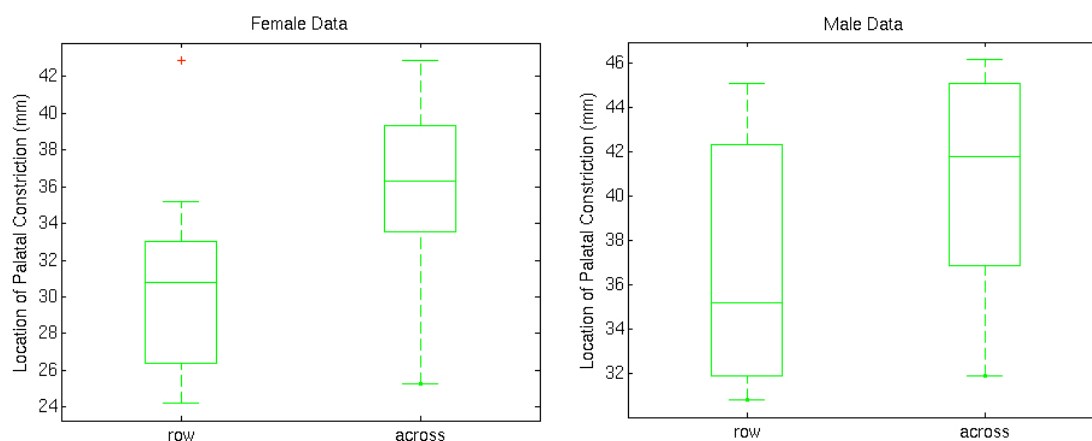


Figure 4. Distribution of palatal constriction location for [ɹ] in the words *row* and *across* for 5 female subjects (left) and five male subjects (right).

4. Discussion

The presence of variability in the placement of the palatal constriction for the different articulatory [ɹ] variants is not in itself surprising. First of all, as has been mentioned, several studies have presented significant differences in the acoustic patterns of the [ɹ] variants. Hagiwara (1995) showed differences in F3, where the “tip-up” (roughly retroflex) variants had higher F3 than the tip-down variants. Epsy-Wilson & Boyce (1999) found differences in F4 and F5 among the variants. And Hashi et al. (2003) found the transition in F3 to be significantly steeper in the retroflex cases than the bunched cases.¹ Presence of acoustic differences would allow us to expect differences in constriction locations of one or more of the [ɹ] constrictions. Moreover, the fact that different portions of the tongue are used to achieve the palatal constriction in the different variants would allow us to predict small differences in the placement of the constriction due to different mechanical constraints on different portions of the tongue. Therefore the empirical variability at the constriction level is predictable due to the presence of variability at the lower articulatory level and the higher acoustic level. An important question that remains, however, is how the variabilities at different levels are constrained and related.

Articulatory Phonology and Task Dynamics are parts of a framework that answers these questions (Browman and Goldstein, 1990; Saltzman and Munhall 1989). In this framework, there are two levels in speech production: an articulatory level and a gestural level. The latter specifies the contrasts present in a language, whereas the

¹ Significant differences have not always been found. Westbury et al. (1998) did not find reliable acoustic differences.

former specifies the articulators that can be used to realize each gestural contrast. Articulatory variability is a result of there being several articulators that conspire in a synergy to achieve a given contrast. For instance, the tongue tip and Jaw work together to achieve a gesture whose constriction location is in the alveolar region. Experiments have shown that if one articulator is perturbed, the other is able to compensate so that the desired contrast is achieved (Kelso et al., 1986). It is also known that in some languages like American English, the alveolar gesture for [s] can be achieved with the tongue tip or the blade in one position. The theory predicts this type of variability since the articulator set for the alveolar gesture contains both tongue and blade. It is also important to note that there are mechanical differences between the articulators in a given gesture's articulator set. Some variability at the articulatory level can therefore be due to different mechanical constraints on articulators in a set. Variability at the constriction level is also predicted within this framework, since gestures from surrounding linguistic units overlap in time. How the gestures within a linguistic unit and across units are phased with each other will determine how the constrictions within each unit are physically realized. If two gestures overlap in time and they both control the same articulator, the motion of the articulator will depend on both gestures and the influence of the two gestures will result in what is called "blending" (Saltzman and Munhall, 1989). For instance, in American English, the velar constriction for [k] will emerge more forward in the oral cavity if followed by [i] than if followed by [u]. The reason is that the velar gesture for [k] overlaps in time with the palatal or uvular gestures in [i] and [u], respectively. The tongue body is in the articulator set for both the palatal and velar gestures in [ki], and therefore its motion will be influenced by both gestures, resulting in a constriction between the two regions. The same applies for [ku]. Therefore both articulatory and constrictional variability are predictable in the theory. Acoustic variability is also predictable, as a result of the fact that the acoustic signal is physically derived from the locations and degrees of constrictions in the vocal tract. An understanding of context variability in a language is achieved by answering several questions: 1) What are the gestures? 2) How are the gestures phased with each other as a function of the prosodic structure of the language, speech rate, and other conditioning factors? 3) Which articulators are in the Articulator set of each gesture? 4) How is the synergy of articulators for each gesture structured? 5) What are mechanical constraints on each articulator?

Within the framework of Articulatory Phonology and Task Dynamics, [ɹ] ceases to be a strange object. The articulatory variants of [ɹ] are predictable, since both the tongue tip and the anterior portion of the tongue body are able to achieve the narrow palatal constriction for [ɹ].² The articulatory variants for [ɹ] are therefore analogous to the laminal and apical variants of [s] in English. The high mobility of the tongue tip is probably an important contributing factor to its involvement in several segments with articulatory variants. Variability at the constriction level in *across* is derived from the phasing of the velar gesture for [k] with the palatal gesture of [ɹ]. Both gestures use the anterior of the tongue body and therefore influence its motion and the constrictions it

² /i/ also contains a palatal constriction, but the tongue tip cannot be used to achieve that constriction, both since the palatal constriction is further back than that for [ɹ] and since it is a broader constriction (longitudinally) than that for [ɹ].

can achieve. The data presented in this study, which shows that the palatal constriction for the bunched case is further back than the retroflex case, is predicted based on the blending of the influence of the gestures. When a bunched articulation is used for [ɹ], even when there is no nearby velar segment, we find that the palatal constriction is further back than if the same subject were to use a retroflex articulation. From a speech production perspective, this does not seem surprising since there are different mechanical constraints on the tongue tip than on the anterior portion of the tongue body which are in the palatal gesture's constriction set. A difference of a few millimeters between bunched and a retroflex articulations of the same gesture is predictable from the fact that the massive tongue body has a greater pull on the anterior portion of the tongue body than on the tongue tip.

5. Conclusion

It has often been said that American English [ɹ] presents evidence for acoustic control of speech production due to a many-to-one-relation between the articulatory and acoustic levels. The goal of this paper was to show that there are actually three levels involved, that there is variability at all three levels, and that the variability at the levels is constrained by the relation between these levels. American English [ɹ] in this respect is not special. All contrastive units when sequenced together to form words will exhibit variability at the articulatory, constriction, and acoustic levels. And this variability can only be understood when the intergestural and intragestural aspects of speech production are investigated.

References

- Browman, C. and Goldstein, L. Tiers in articulatory phonology with some implications for casual speech. In J. Kingston and M. Beckman, editors, *Papers in Laboratory Phonology I: between grammar and physics of speech*, pages 321-376, Cambridge University Press, 1990.
- Delattre, P. and Freeman, D. A dialect study of American R's by X-ray motion picture. *Linguistics*, 44: 29-68, 1968.
- Espy-Wilson, C. Y., & Boyce, S. E. A simple tube model for American English /r/. In J. Ohala, Y. Hasegawa, M. Ohala, D. Granville, & A. C. Bailey, editors, *Proceedings of the 14th International Congress of Phonetic Sciences*, pages 2137-2140, Berkeley, CA: University of California, Berkeley, 1999.
- Guenther, F.H., Hampson, M., & Johnson, D. A theoretical investigation of reference frames for the planning of speech movements. *Psychological Review*, 105, 611-633, 1998.
- Hagiwara, R. Acoustic realizations of American /R/ as produced by women and men. *UCLA Working Papers in Phonetics*, 90, 1-187, 1995.

- Hashi, M., Honda, K., & Westbury, J., Time-varying acoustic and articulatory characteristics of American English [ɪ]: a cross-speaker study. *Journal of Phonetics*, 31, 3-22, 2003.
- Kaburagi, M. & Honda, K., Determination of the sagittal tongue shape from the positions of points on the tongue surface. *Journal of the Acoustical Society of America*, 96, 1356-1366, 1994.
- Kelso, J. A. S., Saltzman, E. L., & Tuller, B., The dynamical perspective on speech production: Data and theory. *Journal of Phonetics*, 14, 29-60, 1986.
- Ladefoged, P. & Maddieson, I., *The sounds of the world's languages*. Oxford: Blackwell, 1996.
- Löfqvist, A. Theories and models of speech production. In W. Hardcastle and J. Laver, editors, *The Handbook of Phonetic Sciences*, pages 404-426. Oxford: Blackwell, 1997.
- Saltzman, E. L., & Munhall, K. G., A dynamical approach to gestural patterning in speech production. *Ecological Psychology*, 1, 333-382, 1989.
- Story, B.H., Titze, I.R., and Hoffman, E.A., Vocal tract area functions from magnetic resonance imaging, *Journal of the Acoustical Society of America*, 100(1), 537-554, 1996.
- Westbury, J. R., Hashi, M., & Lindstrom, M. J., Differences among speakers in lingual articulation for American English /ɪ/. *Speech Communication*, 26, 203-226, 1998.