Tongue-Jaw Synergy in Vowel Height Production: Evidence from American English

Aude Noiray, Khalil Iskarous, Leandro Bolanos, D. H. Whalen
Haskins Laboratories
Aude.Noiray@haskins.yale.edu

Abstract

Vowels are traditionally described according to three articulatory dimensions: height, frontness and rounding. Despite the linguistic importance of vowel height in many languages, there is still disagreement about its physiological implementation and its acoustic consequences. One area of controversy is whether the jaw or the tongue dorsum is the main contributor to this linguistic distinction. In American English, height is highly contrastive, distinguishing five front and five back vowels in most dialects. The present paper investigates jaw/tongue synergy in distinguishing /i/ from /ɪ/ as a function of vowel and speaker using two methods of articulatory data collection: x-ray Microbeam (XRMB) and digital ultrasound imaging (HOCUS). For most subjects, both tongue and jaw do contribute to the linguistic distinction between /i/ and /ɪ/. However, the tongue's contribution is greater than previously assumed.

1 Introduction

The features of tongue height and frontness are standard ways to distinguish vowels in the phonological literature. However, despite their wide use, these features have been controversial at least since Russell's x-ray study [7], where he showed high variability in the configuration of the tongue, illustrating, for instance, that the tongue needs not to be higher in /I/ than in /e/ [1]. Based on high variability, in the articulatory specification of vowels, Laedefoged has argued that height and frontness are not articulatory features *per se*, but refer to vowel acoustics, where high vowels are those with high F1 and front vowels are those with high F2 [2,3]. However, research on vowel acoustics since [6] has found a large intra-speaker and inter-speaker variability in the acoustic specification of vowels, even for steady-state vowels. Therefore, it is not clear, as commonly assumed, that articulatory variability in vowels is necessarily accompanied with acoustic invariability. A related element of the debate is whether the jaw or the tongue is the main contributor to the height distinction between vowels such as /i/, and /ɪ/. Lindblom and Sundberg posited that the height dimension would mainly be due to jaw opening [4], while Woods suggested that vocal tract narrowing is more appropriate than height [11]. In agreement with Woods, some researchers within the framework of Articulatory Phonology suggest that vowels, like consonants, can be classified according to constriction location and constriction degree, with the latter task distinguishing between vowels of various heights.

The aim of the present paper is to study the contribution of the tongue and jaw to the distinction between vowels like /i/ and. The current investigation is part of a larger study that aims to further analyze the acoustic and area function variability for the same subjects. Two issues are specifically addressed: 1) the distinction between passive motion of the tongue, where jaw position directly determines tongue position, as opposed to active involvement of the tongue for vocalic height contrast and 2) the possibility of attributing different functions to the tongue subparts for achieving vowel height. These questions can only be addressed by separately analyzing the motion of both articulators. Tongue configurations can be expressed in a coordinate system relative to the head or one relative to the jaw [5]. In order to dissociate tongue and jaw contributions to tongue configurations, we used two methods of tongue data collection and compared the results. Vowel productions collected with Wisconsin x-ray Microbeam Database (XRMB) [8] and initially expressed relative to the head were then recalculated in a jaw framework using Westbury's *translation-rotation* method [9]. Results obtained in both head and jaw frames were compared with recent ultrasound data collected in jaw frame and reoriented to the head frame using HOCUS (Haskins Optically Corrected Ultrasound System, [10]).

2 Experimental procedure

The XRMB database consists in the two-dimensional tracking of lips, tongue and jaw motion (145 Hz) with simultaneous recording of the speech wave (21700 Hz). Four pellets were glued on the tongue at regular distances, to get an estimate of its functional divisions (the tongue tip (T1), tongue blade (T2), tongue dorsum (T3) and root (T4)). Two pellets were placed on the vermillion border of upper and lower lip as well as 2 additional reference markers on the jaw (at the central incisor and on a molar tooth). Details on the methods and subjects can be found in [8]. This setup allowed for measurement of the pitch angle of the jaw, simultaneous with measurement of the tongue in the head-based occlusal framework, using the *translation-rotation* method. In this method, the pitch angle (midsagittal rotation) of the jaw is calculated by estimating the angle of a line that rotates with the jaw, obtained as the extension from the jaw pellet to the projection of the molar pellet onto the midsagittal plane. Further details are available in [9]. We used Task 14 from the XRMB, with single repetitions of the vowels [i, ɪ, ɛ ɑ, æ, ʊ, u] by 18 American-English subjects (10 females and 8 males). Only /i/ and /ɪ/ were used. T3 was used as the indicator of the highest point of the tongue.

The second method consisted in an ultrasound recording of tongue motion in 6 American English speakers (4 females, 2 males) producing 6 repetitions of 12 vowels embedded in an *hVd* context. Ultrasound imaging has been increasingly used for investigating complex tongue motions. This technique is noninvasive and allows dynamic observations of most of the tongue with immediate feedback on the motions recorded. It is ideal for the investigation of vowels, since the tongue edge in the midsagittal plane can be detected from the blade to the hyoid bone, allowing us to investigate the tongue back and root configuration. In contrast, flesh-point tracking techniques like XRMB are unable to tract the tongue back or root, since more subjects are not able to keep a pellet very far back on the tongue. However, a limitation of ultrasound technique is the slow 30 Hz rate of analog acquisition, which fails in tracking fast tongue motions (e.g. rhotics). The HOCUS system (Haskins Optically Corrected Ultrasound System), which allows for tracking tongue motion with respect to the head [10], was augmented to include digital acquisition of ultrasound data, combining high temporal rate (127 Hz) and optical tracking of the probe and head motion to allow natural body movement during speech. The set-up was composed of six infrared tracking markers (Optotrak) glued on a helmet and five on the ultrasound probe to allow rigid body reconstruction and allow us to re-express tongue motions in head coordinates. Correction from jaw-based to head-based coordinates is made possible by correcting the tongue edge position with a rotation and a vertical translation in the direction of the motion of the probe, which is optically detected. Details can be found in [10].

The ultrasound probe was placed under subject's chin and held by a spring-loaded probe holder fixed to a weighted stand. The stand was designed with a pivot to adjust probe angle to subject's position. An adjustable chair allowed control over the position of the speaker. This custom integrated set-up allowed the probe to move vertically along with the subject's jaw opening but restrained lateral and anterior-posterior motion to obtain steady edges on the midsagittal plane and a relatively constant portion of the oral cavity across sequences. Ultrasound images were synchronized with the audio and Optotrak data using a trigger pulse. A probe-based grid was overlaid on the image. The center of the grid was chosen as the point of intersection between the lines defining the imaging sector. The highest point of the tongue and posterior-most point on the root of the tongue were obtained as the intersection of the image of the tongue with that grid.

3 Results

The first aim was to compare the effect of the jaw and tongue on the position of the highest point of the tongue in /i/ and /ɪ/. This was done by comparing the position of the highest point of the tongue in the head-based vs. jaw-based coordinate frames. In the latter frame, change in tongue height can only be attributed to tongue muscle activity, whereas in the head-based frame, the jaw is also a contributor to height. For the XRMB data, the comparison consisted in subtracting the vertical position of T3 in /ɪ/ from that for /i/, both in head-based frame, and in the jaw-based frame (after the jaw component is eliminated using the *translate-and-rotate* method). The black crosses in Figure 1 show the /i/-/ɪ/ vertical distance in the jaw-based frame (horizontal axis) as compared to the head-based frame (vertical axis) for the 18 subjects. Also shown in the figure is a line of slope 1. If a point lies on this line, it means that the jaw is effectively not making a contribution to the distinction between /i/ and /ɪ/, since the distance between the highest point on the tongue for the two vowels is the same in both coordinate frames. For points on the slope-line, there can be jaw rotation, but such rotation does not affect the effective distance between the vowels, since it would be in equal positive or negative amounts for the two vowels. Points above the line indicate that the head-based vertical distance is greater than the jaw-based difference, signaling a positive contribution of the jaw to the distance. Points below the slope-1 line indicate negative jaw contribution to the vertical distance between /i/ and /ɪ/. Except for one subject, all the data lies above the slope-1 line. Therefore for 17 out of the 18 XRMB subjects, there is a positive contribution of the jaw to the vertical distance between /i/ and /ɪ/. However the jaw contribution is greater than the contribution of the tongue only for 4 subjects. This was determined by calculating the percentage contribution of the jaw to the total distance. Also shown in the figure is the regression line.

The HOCUS based points are shown in gray, also in Figure 1. Each point represents one of 6 tokens for each of 6 subjects. . There are two main differences between the two imaging modalities. The first is that for HOCUS, we see many more points below the slope-1 line, indicating jaw action that is *opposed* to the tongue action.

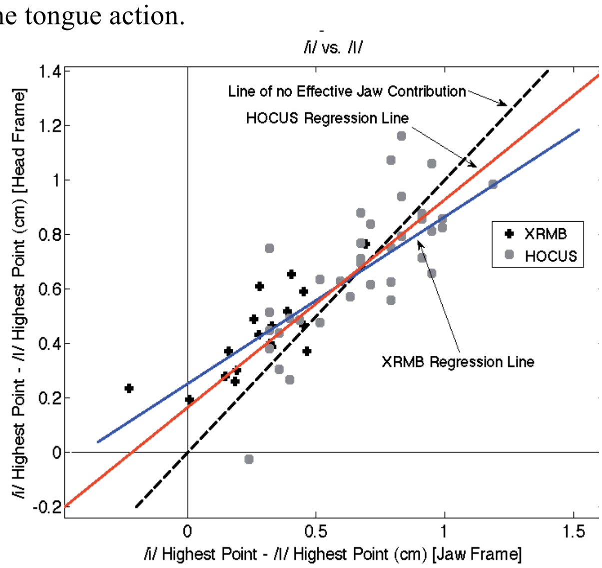

Figure 1: *Difference in highest point ([i] - [ɪ]) in the jaw frame (tongue motion only) and head frame (jaw and tongue motion combined) for XRMB data (circles) and Ultrasound data (squares). A diagonal line was manually added indicating instances where the contribution of the jaw is null.*

For these tokens, the jaw lessens the vertical distance between the vowels, instead of raising it. 17 out of the 36 tokens had jaw pitch difference less than 0, with mean -.82 and standard deviation of .7. In contrast, 19 tokens had positive pitch difference with a mean of 1.4 and standard deviation of 1.3. Second, there is a greater vertical difference between /i/ and /ɪ/ measured in HOCUS as opposed to XRMB. This can be seen in the fact that there more gray points the farther to the right. This second effect can be explained by the fact that the entire tongue dorsum can be seen in the ultrasound data, whereas T3 samples only one point on the dorsum. Therefore, the highest point on the tongue is more easily detectable using HOCUS than XRMB. We measured T4, instead of T3, but this made no difference in the results. The regression line for the HOCUS data is also shown. Statistically, as evident form the regression lines, the two modalities are quite similar, with the slope and offset for XRMB being .61 and .24, and .76 and .17 for HOCUS. And for both modalities, the average vertical difference between /i/ and /ɪ/ is about 5mm. Therefore despite the many differences, between the modalities, similar results are obtained.

The second aim was to explore which active muscular system in the tongue contributes to distinguishing between /i/ and /ɪ/. Ladefoged and Declerk [3] identified two possible systems: Posterior Genioglossus (PGG) contraction pushing the tongue forward and upward and mylohyoid (MH) contraction pushing the tongue front upward. In this study, muscle contraction was not directly measured, however we attempted to infer muscle system activation based on the kinematic evidence. To infer which of these systems is the main actor in the tongue's active system for distinguishing /i/ form /ɪ/, we measured the horizontal displacement of the backmost point on the tongue root and the vertical probe coordinate, one of the rigid body coordinates measured by HOCUS, which measures the vertical displacement of the chin. Initially we though that this latter measure would be highly correlated with the pitch angle of the probe, but a correlation analysis revealed a correlation coefficient of .20, meaning that the two measures are relatively independent, one measuring primarily rotational motion of the jaw, while the other measures chin vertical displacement. The horizontal displacement in the pharynx was within 0-2 mm, making it unlikely that the PGG is the main contributor, unless it has extremely high leverage. Correlation of the vertical rigid body coordinate of the probe, which we take to be an estimate of chin stiffening, most probably due to MH contraction, with the vertical displacement of the tongue, yielded a moderate correlation of .46. We therefore posit that both systems are active in raising the tongue front, with MH being the primary contributor. However, in this system vertical motion of the probe can be a result of vertical motion of the jaw or jaw floor, perhaps confounding the results.

4 Discussion and Conclusion

The aim of the present paper was to determine the contribution of jaw and tongue for the phonological contrast of height and more specifically for [i] versus [I]. Our results from both XRMB and HOCUS show that the tongue and jaw interact in distinguishing /i/ from /ɪ/, but that the tongue, especially the jaw floor system, has a much higher effect than previously assumed. This work also indirectly validates XRMB and HOCUS with each other, showing that two modalities using very different physical imaging methods, and which use different analysis algorithms and estimation methods yield statistically similar data related to a small physical distinction, such as the /i/-/ɪ/ distinction. In further work we will use the subject differences found in this study to trace the effect of the tongue-jaw variability found here with area-function and F0, F1, F2, and F3 variability.

This work was supported by NIH DC-02717.

5 References

[1] E.Fischer-Jørgensen, Some vowel features, their articulatory correlates, and their explanatory power in phonology. In V. Fromkin (ed.): *Phonetic Linguistics: Essays in honour of Peter Ladefoged*. Academic Press, New York, 1985.

[2] J.K. Johnson, P. Ladefoged, M. Lindau. Individual differences in vowel production. *Journal of Acoustical Society of America*, 94:701-714, 1993.

[3] P. Ladefoged, J. DeClerk, R. Harshman. Control of the tongue in vowels. In A. Rigault & R. Charbonneau (Eds.), *Proceedings of the Seventh International Congress of Phonetic Sciences*, 349-354, The Hague: Mouton, 1972.

[4] B. Lindblom, J. Sundberg. Acoustical consequences of lip, tongue, jaw and larynx movement. *Journal of Acoustical Society of America*, 50: 1166–1179, 1971.

[5] D.J. Ostry, K.G. Munhall. Control of jaw orientation and position in mastication and speech. *Journal of Neurophysioly,* 71: 1528-1545, 1994.

[6] G.E. Peterson, H.L. Barney. Control methods used in a study of the vowels. *Journal of Acoustical Society of America*, 24:175-184, 1952.

[7] G.O. Russell, *The Vowel*: *Its Physiological Mechanism as Shown by X-ray*. Ohio State Univ. Press, 1928.

[8] J.R. Westbury, *X-ray microbeam speech production database user's handbook*. Waisman Center, Madison: University of Wisconsin at Madison, 1994.

[9] J.R. Westbury, M.J. Lindstrom. Tongues and lips without jaws: A comparison of methods for decoupling speech movements. *Journal of Speech, Language, and Hearing Research*, 45:651-662, 2002.

[10] D.H. Whalen, K. Iskarous, M.K. Tiede, D.J. Ostry, H. Lehnert-LeHouillier, E. Vatikiotis-Bateson, E. Hailey. The Haskins Optically Corrected Ultrasound System (HOCUS). *Journal of Speech, Language, and Hearing Research*, 48: 543-553, 2005.

[11] S. Wood. A radiographic analysis of constriction location for vowels. *Journal of Phonetics,* 7: 25-43,1979.