JSLHR

Supplement Article

Influences of Tone on Vowel Articulation in Mandarin Chinese

Jason A. Shaw,[a] Wei-rong Chen,[b]
Michael I. Proctor,[c] and Donald Derrick[d]

Purpose: Models of speech production often abstract away from shared physiology in pitch control and lingual articulation, positing independent control of tone and vowel units. We assess the validity of this assumption in Mandarin Chinese by evaluating the stability of lingual articulation for vowels across variation in tone.
Method: Electromagnetic articulography was used to track flesh points on the tongue (tip, body, dorsum), lips, and jaw while native Mandarin speakers (*n* = 6) produced 3 vowels, /p**a**/, /p**i**/, /p**u**/, combined with 4 Mandarin tones: T1 "high," T2 "rising," T3 "low," and T4 "falling."
Results: Consistent with physiological expectations, tones that begin low, T2 and T3, conditioned a lower position of the tongue body for the vowel /a/. For the vowel /i/, we found the opposite effect, whereby tones that begin low, T2 and T3, conditioned a higher tongue body position.
Conclusions: The physiology of pitch control exerts systematic variation on the lingual articulation of /a/ across tones. The effects of tone on /i/ articulation are in the opposite direction predicted by physiological considerations. Physiologically arbitrary variation of the type observed for /i/ challenges the assumption that phonetic patterns can be determined by independent control of tone (source) and vowel (filter) production units.

Models of speech production generally assume that the glottal source and the supraglottal vocal tract filter are independent (e.g., Fant, 1960)—an assumption implicit as well in modern phonological theories that treat vowels and lexical tones as independent primitives (Duanmu, 2002; Gao, 2009). The assumption of source–filter (and tone–vowel) independence abstracts away physiological linkage between pitch control and lingual articulation. The biomechanics of pitch control (e.g., Honda, 1983) suggest that lexical tone production could possibly influence the lingual articulation of vowels. Coarticulatory influences of tone on vowel production may explain some otherwise peculiar facts about lexical tone perception. In Mandarin spoken word recognition, the time required to recognize a lexical tone depends on the vowel with which it is coarticulated (Shaw et al., 2013). Tone-conditioned vowel variation may potentially provide early cues to word identity in a manner similar to anticipatory coarticulation of segments (e.g., Beddor, McGowan, Boland, Coetzee, & Brasher, 2013). When the F0 contour of lexical tones unfolds more slowly than the lingual targets of vowels, tone-conditioned variation on vowel targets could inform tone identity before other informative phonetic cues, such as the F0 turning point (i.e., the inflection point of the tone), become available in the signal. To investigate the possibility of tone-conditioned variation in Chinese vowel targets, we conducted an electromagnetic articulography (EMA) experiment recording the articulatory kinematics of vowels produced with the four lexical tones of Mandarin Chinese.

Although there is a sizable body of research aimed at investigating how lingual articulation influences pitch, including debates about the physiological basis (Honda, 1995; Kingston, 1992; Whalen, Gick, Kumada, & Honda, 1999) and language specificity (Connell, 2002; Whalen & Levitt, 1995) of the effects, there is less work on the influences of tone on vowel articulation. A handful of previous studies indicate that tongue position does indeed vary to some degree with tone (Erickson, Iwata, Endo, & Fujino, 2004; Hoole & Hu, 2004; Hu, 2004). A commonly observed effect is that production of the vowel /a/ has lower tongue and jaw positions when produced with a low tone than when produced with a high tone (Erickson et al., 2004;

---

[a]University of Western Sydney, Australia
[b]National Tsing Hua University, Hsinchu City, Taiwan
[c]Macquarie University, Sydney, Australia
[d]University of Canterbury, Christchurch, New Zealand

Correspondence to Jason A. Shaw: J.Shaw@uws.edu.au

Editor: Jody Kreiman
Associate Editor: Susanne Fuchs

Received January 28, 2015
Revision received August 26, 2015
Accepted October 7, 2015

DOI: 10.1044/2015_JSLHR-S-15-0031

Disclosure: *The authors have declared that no competing interests existed at the time of publication.*

Hoole & Hu, 2004). Two other relevant findings are that the effect of tone on vowel is not uniform across vowels /a/ and /i/ (Hoole & Hu, 2004; Hu, 2004) and that tone may influence vowels at different points on the tongue depending on the vowel (Hu, 2004). It is not entirely clear whether these effects of tone on lingual articulation of vowels have a physiological basis, a cognitive–representational basis, or both.

One mechanism for reduction of vocal fold tension, as required for low tones, involves lowering the larynx (Honda, Hirai, Masaki, & Shimada, 1999), which could pull the jaw down via the posterior displacement of the hyoid (Honda, 1995). Unless the tongue is raised to compensate, a lower jaw would result in lower tongue height for /a/. Other reported effects of tone on vowel production may have similar physiological explanations. For example, the differential effects of tone on /a/ and /i/ may follow from the different ways that the jaw and hyoid bone support lingual articulation of these vowels. For /i/, activation of the genioglossus required to advance the tongue dorsum may pull the hyoid bone forward, countering the impact of a lowered larynx on hyoid position (cf. Honda, 1983, p. 274). Physiological effects predicted to affect /a/ may thus be absent or negligible for /i/.

Another attested mechanism for low tone production involves a constriction at the larynx accompanied by larynx raising (Moisik, Lin, & Esling, 2014). Because this mechanism does not involve lowering the larynx, it would not have the same impact on the hyoid and jaw described above. Thus, it is alternatively possible that some (or all) of the tone-conditioned vowel variation observed in past studies of Chinese is arbitrary, language-specific phonetic variation. Such a finding would challenge speech production models that derive phonetic patterns from the simple combinatorics of independent tone and vowel units.

In contrast to modern phonological theories, traditional Chinese phonology divided the syllable into two nondecomposable parts: an initial (*shēngmŭ*) and a final (*yùnmŭ*; e.g., Chao, 1968). The *initial* is the first consonant of a syllable. The *final* includes the nuclear vowel, tone, and optional coda in a single unit (Chao, 1968, p. 19). More holistic phonological labels, such as the finals of traditional Chinese phonology, could potentially capture tone-conditioned variation on vowels of both the physiologically motivated and arbitrary varieties. Also relevant in this regard are theories of word-specific phonetics (e.g., Pierrehumbert, 2002). In lexical tone languages, vowels may show subtle (and possibly arbitrary) variation across tones because different tones constitute different words, and words may exert a bias on vowel articulation.

The kinematic studies on Chinese tone–vowel interaction reviewed above are too limited in data to address this question conclusively. They report EMA data from just one or two speakers with a small number of contexts, tones, and repetitions. More data are needed to evaluate the stability of lingual articulation across tonal environments and to explore the physiological basis of tone-conditioned lingual variation. We take up these issues here, replicating past work with a larger number of speakers and repetitions and assessing the physiological basis of tone-conditioned lingual articulation.

To assess the effect of tone on lingual articulation, we analyzed the spatial position of three EMA receiver coils on the tongue during articulation of three vowels, /a/, i/, and /u/, each produced with four Mandarin tones: T1 “high,” T2 “rising,” T3 “low,” and T4 “falling.” The physiological explanation suggested above for how pitch control may affect lingual articulation involves pull on the jaw (due to larynx lowering for low tones). To test this, we analyzed the Euclidean distance between jaw and tongue for each vowel across tones. Stability in tongue-to-jaw distance across tone-conditioned variation in tongue position would support the physiological hypothesis. This pattern of results would be consistent with the phonological independence of tones and vowels. As an alternative, arbitrary tone-conditioned variation in the lingual position of vowels would support integrated tone–vowel representations, such as finals, or biases in production exerted by phonetically detailed representations of words (as in the proposal of Pierrehumbert, 2002). On this account, the position of the tongue or the relative position of the jaw and tongue could vary for the same vowel across tone–vowel combinations as each combination represents a different word. The clearest case of arbitrary variation would involve tone-conditioned lingual variation that goes in the opposite direction of the biomechanical influences as in, for example, if /a/ produced with low tones showed a higher tongue position than /a/ produced with high tones or, likewise, if /i/ produced with high tones showed a lower tongue position than /i/ produced with low tones.

Methodology

Speakers

Six native speakers of Mandarin Chinese (three men, three women) participated. Participants were aged between 21 and 25 years ($M = 23.7$, $SD = 1.5$) at the time of the study. They were all born in Northern China (Beijing and surrounding areas) and lived there until at least 18 years of age. All participants were screened to ensure that they spoke standard Mandarin Chinese. Procedures were explained to participants in Mandarin by the second author. Participants were compensated for their time and local travel expenses.

Materials

Each speaker produced multiple repetitions of three maximally dispersed vowels (/i/-/a/-/u/) in labial-initial syllables (/pV/) with each of the four Mandarin tones: T1 high, T2 rising (low-high), T3 low, and T4 falling (high-low). Each syllable was produced 12 times by each speaker, generating a corpus of 864 tokens (12 repetitions × 3 vowels × 4 tones × 6 speakers). Syllables were presented in Pinyin and randomized with fillers and words included for other experiments.

Equipment

We used an NDI Wave EMA system sampling at 100 Hz to capture articulatory movement. NDI Wave 5DoF sensors (receiver coils) were attached to three locations on the sagittal midline of the tongue and on the lips, jaw (below the lower incisor), nasion, and left/right mastoids. The most anterior sensor on the tongue, henceforth tongue tip (TT), was attached less than 1 cm from the TT. The most posterior sensor, henceforth tongue dorsum (TD), was attached as far back as was comfortable for the participant. A third sensor, henceforth tongue body (TB), was placed on the TB roughly equidistant between the TT and TD sensors. Acoustic data were recorded simultaneously at 22 KHz with a Schoeps MK 41S supercardioid microphone (with Schoeps CMC 6 Ug power module).

Stimulus Display

Syllables were displayed in pinyin on a monitor positioned 45 cm outside of the NDI Wave magnetic field. Stimulus display was controlled manually using a visual basic script in Excel. This allowed for online monitoring of hesitations, mispronunciations, and disfluencies. These were rare, but when they occurred, participants were asked to repeat syllables. This method ensured that we recorded 12 fluent tokens of each target item.

Postprocessing

Following the main recording session, we also recorded the bite plane of each participant by having him or her hold a rigid object with three 5DoF sensors attached to it between their teeth. Head movements were corrected computationally after data collection with reference to three sensors on the head, the left/right mastoid and nasion sensors, and the three sensors on the bite plane. The head-corrected data was rotated so that the origin of the spatial coordinates corresponded to the occlusal plane at the front teeth.

Acoustic Analysis

Pitch tracking was conducted in MATLAB using the YAAPT algorithm (available at http://ws2.binghamton.edu/zahorian/yaapt.htm). Parameters were optimized for each speaker using a bootstrapping algorithm (automatically estimating as well as visually/manually inspecting the minimal and maximal F0 boundaries for each subject). This allowed for interpolation over intervals of creaky voice, which were present for some tokens, particularly for T3. Raw F0 samples were then converted to *z*-scores within speaker.

Articulatory Analysis

The articulatory data analysis focuses on the position of EMA receiver coils at the achievement of lingual targets for vowels. Vowel targets were determined by a 20% threshold of peak velocity of the TD sensor in the opening movement of the vowel. Labeling was done using the *findgest* algorithm in MVIEW, a program developed by Mark Tiede at Haskins Laboratories. Vertical and horizontal positional coordinates were extracted at the vowel target for the jaw and lingual sensors (TT, TB, TD). Positional coordinates were then *z*-scored within speaker, articulator (EMA sensor), and dimension (vertical, longitudinal) and across items. This transformation facilitates comparison across speakers (of the effect of tone on vowel position) by normalizing for speaker differences in both mean sensor position and sensor position variance. From the standpoint of assessing the effect of tone on vowel position, both of these are unwanted sources of variability. Statistical analyses were conducted on normalized values. In order to visually represent the effect of tone in millimeter units, we converted average *z*-scores back into millimeter values. This was done by reverse *z*-score—that is, multiplying the average (across subjects) *z*-score by the average (across subjects) standard deviation and adding the average (across subjects) sensor position. Converting *z*-scores into millimeters in this way preserves the structure of the differences present in *z*-scores but expresses them in millimeter units. To statistically analyze the effect of tone on vowel position, we conducted separate repeated-measures analyses of variance on *z*-scores for each sensor and vowel in vertical and longitudinal (anterior–posterior) dimensions. To visualize the effects, we also show the means and 95% confidence interval estimates for each comparison.

To assess the contribution of the jaw on differences in tongue position across tones, we computed the Euclidean distance between jaw and lingual sensors and analyzed the effect of tone on these distances. If the effect of tone on vowel articulation is driven by the jaw, then we would expect to find tone-conditioned differences in lingual and jaw position but stable distances between the tongue and the jaw.

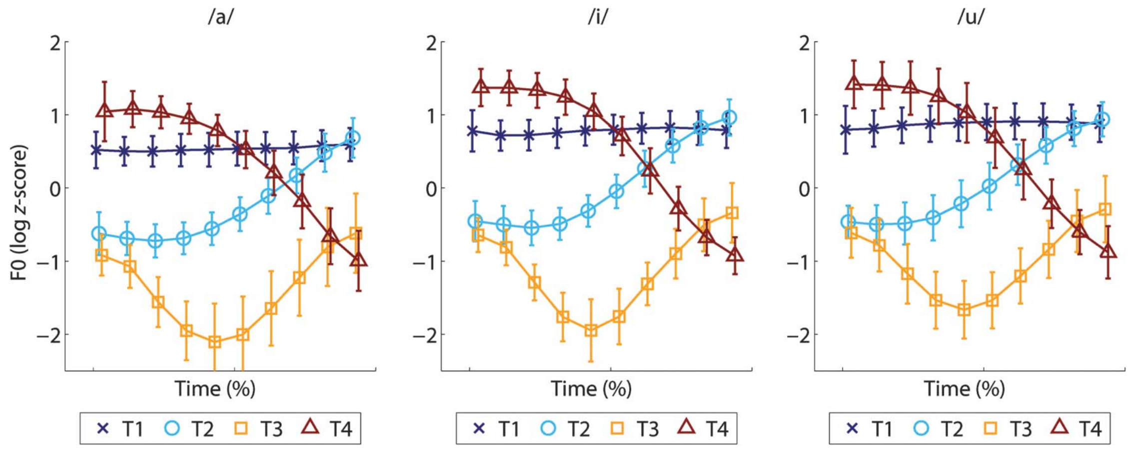
Results

We report the results in three sections. First, we report F0 across tones and vowels. These results indicate consistent F0 patterns across vowels. We then analyze tone-conditioned spatial variation found at the vowel target landmark, focusing on the midsagittal plane. After considering the vertical and longitudinal dimensions separately, we report displacement in these dimensions relative to the jaw sensor.

Time Course of F0

Figure 1 shows the average F0 contour for each vowel across tones. F0 was sampled at regular intervals on the basis of percentage of total vowel duration. Figure 1 demonstrates that the speakers in this study produced F0 trajectories across vowels that are highly consistent with previous findings for Mandarin Chinese, including a small

Figure 1. Average normalized F0, *y*-axis, plotted by time expressed as a percentage of total vowel duration, *x*-axis, for each combination of tone (T1–T4) and vowel (/i/-/a/-/u/) in the corpus.



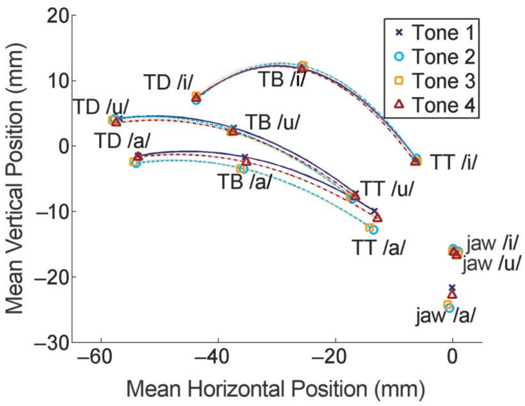
effect of intrinsic F0 on T1 high and T3 low tones (cf. Shi & Zhang, 1987).

The Effect of Tone on Vowel Targets

Figure 2 provides a summary of tongue and jaw positions within the midsagittal plane across tones and vowels. Each point represents the average spatial position across speakers at the vowel target landmark. Tongue edges are represented as quadratic fits between TT, TB, and TD sensors (within vowel and within tone). Similar tongue and jaw position across tone can be seen for /u/, small differences for /i/, and larger differences for /a/.

Figure 3 compares vowel height (i.e., position in the vertical dimension) for TD, TB, TT, and jaw receiver coils across tones. Significant effects of tone, indicated by shading, were found in the vertical dimension at the TB for both /i/, $F(3, 15) = 5.95$, $p < .01$, and /a/, $F(3, 15) = 11.55$, $p < .001$; at the TT sensor for /a/, $F(3, 15) = 6.87$, $p < .01$; and at the jaw for /a/, $F(3, 15) = 4.58$, $p < .05$. The direction of the effects, which can also be seen in Figure 3, are as follows: For /i/, TB is higher for T2 and T3 than it is for T1 and T4. For /a/, TT, TB, and jaw are all lower for T2 and T3 than they are for T1 and T4. Tukey's post hoc tests indicate that the significant effects of tone for both /a/ and /i/ are attributable to T2 and T3. These tones—those that begin low—were significantly different from T1 and T4—those that begin high—for both /a/ (TB, TT, and jaw receiver coils) and /i/ (TB only). The difference between T2 and T3 was not significant, nor was the difference between T1 and T4. Effects of tone on /u/ were not significant, nor was the effect of tone on TD position for any of the vowels. Figure 4 shows the longitudinal dimension of vowel targets for TD, TB, TT, and jaw receiver coils by tone. The structure of Figure 4 parallels that of Figure 3. Again, significant effects are shaded in gray. The only significant effect of tone on longitudinal position was found for /a/ at the TT sensor, $F(3, 15) = 4.35$, $p < .05$. Tukey's post hoc test indicated that this difference is attributable to T3, which was significantly more posterior than the other three tones.

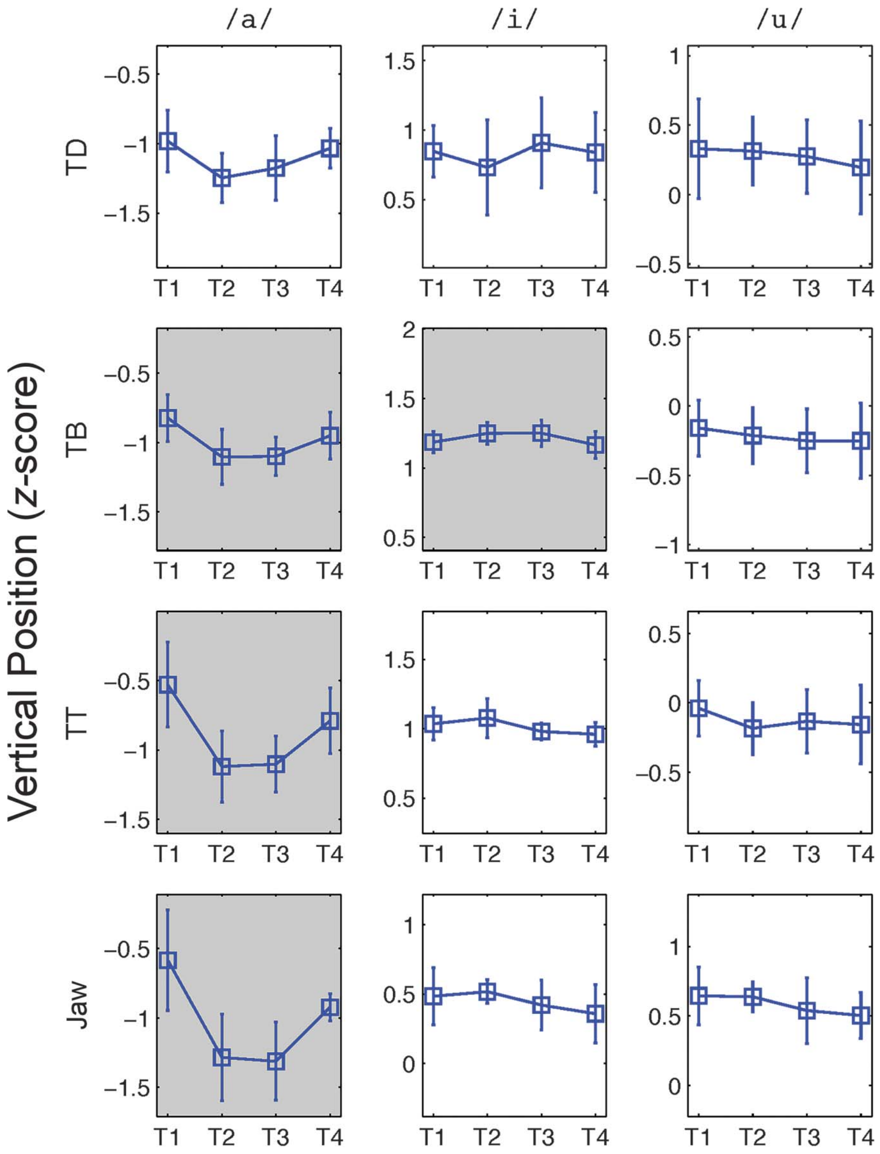
Figure 2. Mean midsagittal tongue position (12 repetitions, six speakers) for Mandarin vowels /i/-/a/-/u/ produced with T1 "high," T2 "rising," T3 "low," and T4 "falling." Tongue edges are represented as quadratic fits between tongue tip (TT), tongue body (TB), and tongue dorsum (TD) receiver coils (within vowel and within tone).



Distance Between Jaw and Tongue Position

In addition to analyzing the effect of tone on sensor position individually, we analyzed the position of lingual sensors relative to the jaw. Figure 5 summarizes the Euclidean distance between jaw and TB sensor, the location on the tongue where most of the significant effects of tone were found. As expected, the TB sensor is farthest away from the jaw sensor for the /u/ target (42.3 mm), followed by the /a/ target (40.9 mm), and then the /i/ target (38.9 mm), which is closest to the jaw. However, in contrast to the analyses of vertical TB displacement reported above, there is a negligible influence of tone on TB-to-jaw distance.

Figure 3. Mean height (12 repetitions, six speakers) of tongue dorsum (TD), tongue body (TB), tongue tip (TT), and jaw sensors for Mandarin vowels /a/ (left), /i/ (middle), and /u/ (right) produced with T1, T2, T3, and T4. Error bars represent 95% confidence intervals.

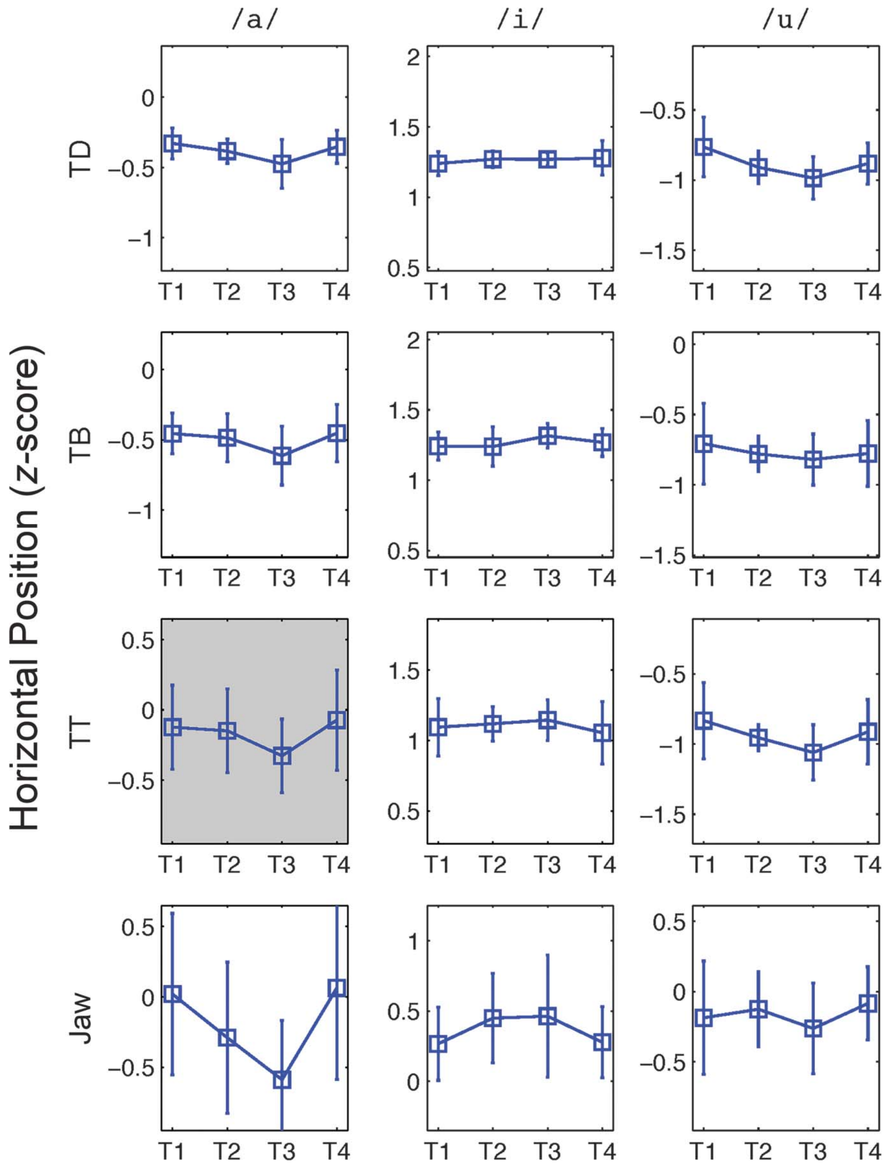


A repeated-measures analysis of variance on TB-to-jaw distance with tone and vowel as independent factors showed a significant effect of vowel, $F(3, 15) = 3.83$, $p = .058$, but not tone, $F(3, 15) < 1$, and no interaction between tone and vowel, $F(3, 15) < 1$, indicating that the position of the TB relative to the jaw is stable across tones.

Discussion

Tones, as well as vowels, are contrastive in Mandarin Chinese, so they must be represented in the lexicon. At issue in our assessment of the effects of tone on vowel articulation is the nature of such representations, their relationship to the phonetic signal, and, in particular, if phonological combinatorics made on the basis of tones and vowels can represent Chinese word forms without information loss. Because changes in tone result in changes to the identity of Chinese words, phonetically detailed representations of words, as proposed in exemplar theory, permit arbitrary phonetic variation in vowels across tones. In contrast, theories deriving phonetic patterns from the coordination of independent vowel and tone units (e.g., Gao, 2009) are more restrictive. Combinatorial theories predict only the physiological effects of pitch control on lingual articulation to surface in the phonetics. As a possible physiological effect, we hypothesized that pull on the jaw from the pitch control mechanism could lead to tone-conditioned variation in lingual position.

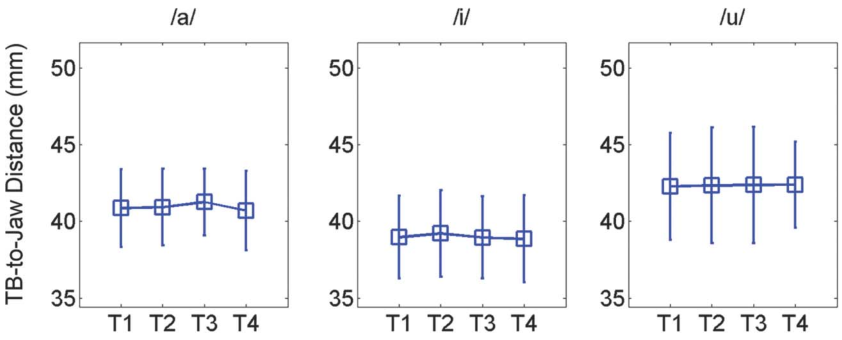
Figure 4. Mean longitudinal position of tongue dorsum (TD), tongue body (TB), and tongue tip (TT) sensors for Mandarin vowels /a/ (left), /i/ (middle), and /u/ (right) produced with T1 “high,” T2 “rising,” T3 “low,” and T4 “falling.” Error bars represent 95% confidence intervals.



Results showed that some aspects of lingual articulation were stable across tones. The position of the most posterior receiver coil on the tongue, the TD sensor, was stable across tones for all vowels. Likewise, the distance between the jaw and the TB remained stable across tones, also for all vowels. Last, for the vowel /u/, the jaw and all three lingual sensors remained stable, showing no significant differences across tones. These areas of articulatory stability are consistent with tone–vowel independence, both phonologically and physiologically.

The spatial position of the two more anterior lingual sensors, TT and TB, was influenced by tone for either one (/a/) or two (/a/ and /i/) of the three vowels examined (see Figure 3). The results for /a/ production are consistent with the findings of Erickson et al. (2004) and Hoole and Hu (2004), who observed that the TB is lower for /a/ produced with a low tone. In our data, the two tones that start low (T2 and T3) patterned together in their influence on tongue and jaw position. For /a/ produced with T2 and T3, the TB, the jaw, and the TT were lower (and the TT was more posterior) than when produced with T1 and T4. For /i/, the only significant effect of tone was found at the TB sensor in the vertical dimension. As with /a/, tones that start low, T2 and T3, patterned together in influencing /i/. However, unlike for /a/, T2 and T3 influenced /i/ in the opposite direction. For /i/, the TB was higher for T2 and T3 than for

Figure 5. Mean Euclidean distance between the tongue body (TB) sensor and the jaw sensor for vowels /a/-/i/-/u/, produced with each tone (1–4). Error bars indicate 95% confidence intervals.



T1 and T4. Thus, at the TB, T2 and T3 pattern together, but they influenced /a/ and /i/ in different directions (see Figure 3).

The constellation of effects found for /a/ occurring with low tones (stable TD but lower jaw, lower TB, lower and more posterior TT, and stable tongue-to-jaw distance) can be explained in terms of shared physiology between pitch control and lingual articulation. One mechanism for slackening vocal fold tension, as required for low tones, involves lowering the larynx (Honda et al., 1999; Moisik et al., 2014). Lowering the larynx can exert a pull downward on the jaw through tissue connecting the laryngeal complex to the hyoid bone (Honda, 1995). Vowels also exert an influence on jaw position. Jaw movement in the opening phase of vowels involves a rotational component (i.e., pitching), whereby the jaw rotates around a terminal hinge, the temporomandibular joint, and vertical and horizontal translations of that axis (Edwards & Harris, 1990). The pattern for /a/ is as expected if the effect of tone on vowel position is mediated by the rotational component of jaw movement. When the jaw lowers due to rotation (decrease in pitch), the impact of jaw lowering on tongue height is greatest for sensors distal to the temporomandibular joint. Given a degree difference in pitch, the magnitude of the effect on lingual sensor displacement is proportional to the distance of the sensor from the terminal hinge. On this account, the relative stability of the TD sensor follows from its posterior position (close to the hinge). The more anterior lingual sensors, TT and TB, show larger effects of tone as expected if driven by jaw pitch. Moreover, for the most anterior lingual sensor, TT, lowering goes hand in hand with retraction (more posterior position for T3). This also is expected if the arc-like motion of jaw rotation around the condyle is driving the effect of tone on tongue position. Because the tongue can move somewhat independently of the jaw, it could in principle rise to counter any effect of tone production pull on the hyoid bone/jaw, maintaining a consistent aperture even as the jaw lowers for low tones. The absence of any such compensatory movement, evidenced most directly in the stable jaw-to-TB distance across tones, is consistent with any account of speech production generating phonetic patterns through control of independent tone and vowel primitives.

The other vowel that showed tone-conditioned variation in lingual position was /i/. The small but reliable effects of tone on TB height for /i/ are in the opposite direction predicted by the physiological account. Shared activation of the genioglossus posterior muscle in the advancement of the TD for /i/ and in pitch raising should result in a higher, more advanced tongue position for /i/ when produced with high pitch than with low pitch (Honda, 1995, p. 228). Contrary to this prediction, our data show the opposite effect. The TB was higher for /i/ when produced with low tones than with high tones. Just as for /a/, significant variation in TB height across tones co-occurred with relatively stable distances between the TB and jaw. For /i/, the distance remained stable across tones (despite the significant increase in TB height for T2 and T3) because the jaw had a more advanced and raised position for /i/ with low tones. Although the effects of tone on jaw position for /i/ were not significant, the direction of the variation—raised and advanced jaw for low tones—is consistent with larynx raising, an alternative mechanism for lowering pitch. Using concurrent ultrasound and nasal endoscopy methods, Moisik et al. (2014) observed two mechanisms of laryngeal articulation engaged in Chinese low tone production: one which involves larynx lowering and one which involves larynx raising together with laryngeal constriction. Of particular interest is that the stimuli in Moisik et al. included only words containing the vowel /i/ due in part to methodological considerations associated with nasal endoscopy. It is possible that preferences for different mechanisms of low tone production vary with vocalic context and that larynx raising (for low tones) is more likely in /i/ than in /a/. If so, then the raised and advanced jaw position contributing to the significant increase in TB height for /i/ with low tones may also have a physiological basis. Complications in applying this mechanistic account to our data include the fact that T2 and T3, the tones that start low, pattern together in their influence on TB height, whereas Moisik et al. observed larynx raising only for T3. In addition, laryngeal constriction tends to initiate creaky voice. Although many of the T3 tokens in our data did have creaky voice, creaky voice was rare for T2. Thus, it is unlikely that laryngeal raising alone accounts for the higher TB position for both T2 and T3 in our data.

The anatomy and physiology of the larynx is remarkably complex. Further research may reveal a viable physiological account for why /i/ has a higher TB when produced with low tones than with high tones. However, in our current understanding, the raising of TB for /i/ with tones that start low is physiologically arbitrary. In fact, it runs perfectly contrary to the direction of the expected physiological effect. Effects of tone on lingual articulation were greatest for /a/, presumably because mechanisms for low tone and /a/ production have shared musculature working in the same direction in this combination. For /i/ with low tones, the muscular forces are antagonistic—the genioglossus pulls the hyoid forward to advance the TD while the anterior cricoid pulls it backward to facilitate lowered pitch production. In Honda's (1983) hyoid data, these types of antagonistic forces from pitch control and vowel articulation cancel each other out, resulting in intermediate hyoid position. In our Chinese data, they create the environment for seemingly arbitrary phonetic variation.

How can we model the reliable variation in TB height produced across Chinese? From the standpoint of physiology, this variation is arbitrary. Arbitrary phonetic variation of this type can be explained with reference to word-based phonetic distributions as in exemplar theory (Pierrehumbert, 2002) or theories that advocate speech production units larger than a vowel, including the finals of traditional Chinese phonology or Fujimura's (1986) icebergs, but they are a challenge for more restrictive speech production theories generating phonetic patterns from the coordination of independent tone and vowel primitives. Thus, the maintenance of small but systematic differences in vowel target as a function of tone for /i/ supports an integrated representational hypothesis. TB height may vary across /i1/, /i2/, /i3/, and /i4/ in our data because each of these finals is an independent unit of speech production, each with its own lingual targets, or because /pi1/, /pi2/, /pi3/, and /pi4/ are all different words of Chinese with word-specific phonetic targets.

We close by returning to the issue of tone perception in light of our articulatory results. The pattern of tone-conditioned variation found in articulation in this study roughly parallels patterns of early tone perception in spoken word recognition (Shaw et al., 2013). Shaw et al. (2013) found that Tone 2 was recognized fastest when it was coarticulated with /a/. This could be because the coarticulatory influence of Tone 2 on /a/ provides early information about the identity of tone. If so, it is interesting to note that T2 was recognized comparatively quickly (faster than T2 with /u/) when it was coarticulated with /i/ even though the effect of tone on lingual articulation in this study went in the opposite direction as would be predicted by physiological factors alone. Lingual targets unique to tone–vowel combinations may therefore function to enhance spoken word recognition for words minimally differentiated by tone. Moreover, this appears to be the case regardless of whether tone-conditioned phonetic variation is of the physiologically conditioned or arbitrary variety.

Conclusion

Articulatory kinematic data offers a novel empirical approach to investigating the form of phonological representations. Tones exert a systematic influence on vowel articulation in Mandarin Chinese. Some of the effects can be attributed to the shared physiological connections between the larynx and the tongue. Others—that is, the higher position of TB for /i/ when produced with T2—do not have a clear physiological basis but may nonetheless function to enhance tone perception as evidenced in spoken word recognition tasks that evaluate the time course of tone perception (Shaw et al., 2013). Alongside systematic influences of tones on TB height for /i/ and a constellation of effects on /a/, there were also aspects of vowel articulation that remained stable across tones. These included TD position and the relative position of the jaw and tongue (all lingual sensors). Taken together, the results support a model of speech production that captures both the stability of some aspects of vowel articulation across tones as well as those aspects of phonetic variation that are localized in particular tone–vowel combinations.

Acknowledgments

Research was funded by an internal research grant from the MARCS Institute to Jason A. Shaw, Michael Tyler, Michael Proctor, Donald Derrick, and Chong Han. We would like to thank our six Mandarin participants as well as Chong Han, Jia Ying, and Yuan Ma for help recruiting them. We are grateful for preliminary discussions of the data with Allard Jongman, Joan Sereno, San Duanmu, Cathi Best, and Denis Burnham and for feedback received at the 10th International Seminar on Speech Production. The comments of editor Jody Kreiman, associate editor Susanne Fuchs, and two anonymous reviewers greatly improved the manuscript.

References

Beddor, P. S., McGowan, K. B., Boland, J. E., Coetzee, A. W., & Brasher, A. (2013). The time course of perception of coarticulation. *The Journal of the Acoustical Society of America, 133,* 2350–2366.

Chao, Y.-R. (1968). *A grammar of spoken Chinese*. Berkeley: University of California Press.

Connell, B. (2002). Tone languages and the universality of intrinsic F0: Evidence from Africa. *Journal of Phonetics, 30,* 101–129.

Duanmu, S. (2002). *The phonology of standard Chinese*. Oxford, United Kingdom: Oxford University Press.

Edwards, J., & Harris, K. S. (1990). Rotation and translation of the jaw during speech. *Journal of Speech and Hearing Research, 33,* 550–562.

Erickson, D., Iwata, R., Endo, M., & Fujino, A. (2004, March). *Effect of tone height on jaw and tongue articulation in Mandarin Chinese*. Paper presented at the International Symposium on

Tonal Aspects of Languages: With Emphasis on Tone Languages, Beijing, China.

Fant, G. (1960). *Acoustic theory of speech production*. The Hague, the Netherlands: Mouton.

Fujimura, O. (1986). Relative invariance of articulatory movements: An iceberg model. In J. S. Perkell & D. Klatt (Eds.), *Invariance and variability in speech processes* (pp. 226–242). Hillsdale, NJ: Erlbaum.

Gao, M. (2009). Gestural coordination among vowel, consonant and tone gestures in Mandarin Chinese. *Chinese Journal of Phonetics, 2,* 43–50.

Honda, K. (1983). *Relationship between pitch control and vowel articulation* (Haskins Laboratories Status Report on Speech Research SR-73, pp. 269–282). New Haven, CT: Haskins Laboratories.

Honda, K. (1995). Laryngeal and extra-laryngeal mechanisms of F0 control. In F. Bell-Berti & L. J. Raphael (Eds.), *Producing speech: contemporary issues for Katherine Safford Harris* (pp. 215–232). Woodbury, NY: AIP Press.

Honda, K., Hirai, H., Masaki, S., & Shimada, Y. (1999). Role of vertical larynx movement and cervical lordosis in F0 control. *Language and Speech, 42,* 401–411.

Hoole, P., & Hu, F. (2004, March). *Tone-vowel interaction in standard Chinese*. Paper presented at the International Symposium on Tonal Aspects of Languages: With Emphasis on Tone Languages, Beijing, China.

Hu, F. (2004, March). *Tonal effect on vowel articulation in a tone language*. Paper presented at the International Symposium on Tonal Aspects of Languages: With Emphasis on Tone Languages, Beijing, China.

Kingston, J. (1992). The phonetics and phonology of perceptually motivated articulatory covariation. *Language and Speech, 35,* 99–113.

Moisik, S. R., Lin, H., & Esling, J. H. (2014). A study of laryngeal gestures in Mandarin citation tones using simultaneous laryngoscopy and laryngeal ultrasound (SLLUS). *Journal of the International Phonetic Association, 44,* 21–58.

Pierrehumbert, J. (2002). Word-specific phonetics. In C. Gussenhoven & N. Warner (Eds.), *Laboratory Phonology 7* (pp. 101–139). New York, NY: Mouton de Gruyter.

Shaw, J. A., Tyler, M. D., Kasisopa, B., Ma, Y., Proctor, M. I., Han, C., . . . Burnham, D. K. (2013, August). *Vowel identity conditions the time course of tone recognition*. Paper presented at the 14th Annual Conference of the International Speech Communication Association, Lyon, France.

Shi, B., & Zhang, J. (1987). Vowel intrinsic pitch in Standard Chinese. *Proceedings of the 11th International Congress of Phonetic Sciences, Tallinn, Estonia: Academy of Sciences of the Estonian SSR, 1,* 142–145.

Whalen, D. H., Gick, B., Kumada, M., & Honda, K. (1999). Cricothyroid activity in high and low vowels: Exploring the automaticity of intrinsic F0. *Journal of Phonetics, 27,* 125–142.

Whalen, D. H., & Levitt, A. G. (1995). The universality of intrinsic F0 of vowels. *Journal of Phonetics, 23,* 349–366.