Influences of tone on vowel articulation in Mandarin Chinese

Jason A. Shaw
University of Western Sydney, Sydney, Australia

Wei-rong Chen
National Tsing Hua University, Hsinchu City, Taiwan

Michael I. Proctor
Macquarie University, Sydney, Australia

Donald Derrick
University of Canterbury, Christchurch, New Zealand

Corresponding author:
Jason A. Shaw
MARCS Institute | School of Humanities and Communication Arts,
University of Western Sydney, Locked Bag 1797, Penrith, NSW 2751, Australia
Email: J.Shaw@uws.edu.au
Abstract

**Purpose:** The aim of this paper is to evaluate the correspondence between units of speech motor control and units of phonological contrast. Lexical tone languages offer a valuable window on this issue. Because vowels and tones are phonologically contrastive, temporally co-extensive and (to a large degree) physiologically independent, patterns of articulatory (in)stability in vowels across tones reveal the nature of articulatory control structures and their relation to phonological organization.

**Methods:** Electromagnetic Articulography (EMA) was used to track fleshpoints on the tongue tip (TT), tongue body (TB), tongue dorsum (TD), lips, and jaw, while native Mandarin speakers (n = 6) produced three vowels, /a/, /i/, /u/, combined with four Mandarin tones: high, rising (low-high), low, and falling (high-low).

**Results:** Low and rising tones patterned together in conditioning lower TB targets for /a/ and higher TB targets for /i/. Other aspects of vowel articulation were stable across tones, including TD position and the relative spatial positioning of the tongue and jaw.

**Conclusions:** Results are consistent with models of speech production in which vowels are specified by a small number of parameters with direct mappings to articulatory control parameters but also suggest a role for larger organizational units, e.g., words, in articulatory organization.

**KEYWORDS (5 allowed):** tone, vowels, Articulatory kinematics, Mandarin Chinese
Introduction

Models of speech production generally assume that the glottal source and the supra-glottal vocal tract filter are independent (e.g., Fant 1960) – an assumption implicit as well in modern phonological theories that routinely treat vowels and tones as independent primitives (e.g., Duanmu 2007; Gao 2009). In contrast, traditional Chinese phonology divided the syllable into two non-decomposable parts: an ‘initial’ (shēngmŭ) and a ‘final’ (yûnmŭ) (e.g., Chao 1968). The ‘initial’ is the first consonant of a syllable. The ‘final’ includes the nuclear vowel, tone and optional coda into a single unit (Chao 1968: 19). Holistic supra-phonemic units, such as the finals of traditional Chinese phonology, are consistent as well with more contemporary exemplar-based models that posit word-specific phonetics or online abstraction over exemplars of various-sized units (e.g., Pierrehumbert 2002).

As tones and vowels are temporally co-extensive and (to a large degree) physiologically independent, articulatory kinematic data can provide a clear window on their phonological independence. If vowels are independent from tones, vowel articulation should remain relatively constant across tonal environments. On the other hand, if units of speech production are larger, more holistic complexes, such as phonetically detailed targets at the level of words (or ‘finals’), we would expect each tone-vowel combination to have a unique articulation, which may in turn have perceptual benefits. In languages that have complex tonal contrasts, such as the contour tones of Mandarin, the time course of tone production unfolds more slowly than vowel articulation. A tone-specific vowel movement could provide earlier information about tone than $f_0$. Such a result could explain why, in Chinese spoken word recognition, vowel identity conditions the time course of tone recognition such that, for example, the rising tone is recognized faster when it is temporally co-extensive with /i/ than when it is co-produced with /u/ (Shaw et al. 2013).
In this study, we addressed the question of tone-vowel independence by conducting an EMA experiment of natural variation in tongue displacement across tones. Previous data suggest that tongue position varies to some degree with tone height (Erickson et al. 2004; Hoole & Hu 2004; Hu 2004). These studies report EMA data from one or two speakers with a limited number of contexts, tones, and repetitions. More data is needed to evaluate the stability of vowel targets across tones, and, in doing so, address the broader theoretical question: whether units of phonological contrast map cleanly onto units of speech motor control (e.g., Saltzman and Munhall 1989).

Methodology

Speakers
Six native speakers of Mandarin Chinese (3 male) 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: 1 ‘high’, 2 ‘low-high’, 3 ‘low’, and 4 ‘high-low’. Each syllable was produced 12 times by each speaker, generating a corpus of 864 tokens (12 repetitions x 3 vowels x 4 tones x 6 speakers). Syllables were presented in Pinyin and randomized with fillers and words included for other experiments.


Equipment

We used an NDI Wave electromagnetic articulograph system sampling at 100Hz to capture articulatory movement. The NDI wave supports 5D sensors and 6D sensors, which can be used for automatic head correction by a proprietary algorithm. We used only 5D sensors for this experiment, attached to the tongue tip (TT), body (TB), dorsum (TD), lips, jaw, nasion and mastoids and corrected for head movement computationally, after data collection. Acoustic data were recorded simultaneously at 22KHz 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 outside of the NDI Wave magnetic field 45cm from participants. 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 stimulus.

Post-processing

Head movements were corrected computationally after data collection with reference to the mastoid and nasion sensors. The post-processed data was rotated so that the origin of the spatial coordinates corresponds to the bite plane (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 boot-strapping algorithm (automatically estimating as well as visual/manual inspecting the minimal and maximal f0 boundaries for each subject). This allowed for
interpolation over areas of creaky voice, which were present for some tokens, particularly for
Tone 3.

Articulatory analysis
The articulatory data analysis focuses on the position of EMA sensor 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. The landmark-based analysis of the data is motivated in part by the assumption
that articulatory gestures pass through a finite number of states as they unfold over time
(Gafos 2002). Using the velocity profile to determine vowel targets has the advantage of
constraining landmark determination both spatially and temporally and can be contrasted with
landmark-based approaches employing temporally arbitrary heuristics, such as the maximum
spatial displacement of the tongue, or spatially arbitrary heuristics, such as the temporal
midpoint of the vowel.
We also explored the velocity minimum as a possible articulatory landmark for analysis but
found that the threshold of peak velocity provided more reliable measurements across tokens.
The monophthongs in the experiment tended to have relatively long periods of low velocity
around the point of maximum opening corresponding to the vowels. Although the NDI Wave
system produced high spatial resolution recordings, even a small degree of measurement error
(≈.6mm), makes picking out the true velocity minima from the wide basin of low velocity
movement, subject to sizeable temporal variation. Using the threshold of peak velocity
mitigates the effect of measurement noise, providing a reliable vowel target landmark across
tokens.
One drawback of the landmark-based analysis is that it may not capture effects of tone on vowel dynamics that occur earlier or later in time than the movement towards target. We are particularly interested in the effects of tone on vowel articulation that occur early in the time course of the vowel, as these can potentially explain why vowel identity conditions early recognition of certain tones (Shaw et al. 2013). In addition to the landmark-based analysis, we have also explored time series analysis of the opening phase of vowel movements. This analysis produced converging evidence for the effects of tone on vowel articulation that we report below. For reasons of space, we focus in this paper on the landmark-based analysis, which is sufficient to capture the effect of tone on the opening phase of the vowel, and leave analysis of the time series data—including effects of tone that occur later in the vowel—to future work.

**Statistical analysis**

Before conducting statistical analysis or averaging across speakers, we first computed the z-score of positional coordinates within speaker, articulator (EMA coil), 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. All statistical analyses were conducted on these normalized values. In order to visually represent the effect of tone in mm units, we converted average z-scores back into mm values. This was done by “reverse z-score”, i.e., 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 mm in this way preserves the structure of the differences present in z-scores but expresses them in mm units. To statistically analyze the effect of tone
on vowel position, we conducted separate repeated measures ANOVAs on z-scores for each
sensor and vowel in vertical and longitudinal (anterior-posterior) dimensions. To visualize the
effects we also show central tendency and 95% confidence interval estimates for each
comparison.

Results

We report the results in three sections. First we report $f_0$ across the tones and vowels. These
results indicate consistent $f_0$ patterns across vowels. We then analyze tone-conditioned spatial
variation found at the vowel target landmark, focusing on the mid-sagittal plane. After
considering the vertical and longitudinal dimensions separately, we introduce another
measure that quantifies displacement in these dimensions relative to the jaw sensor.

Time course of $f_0$

Figure 1 shows the average $f_0$ contour for each vowel across tones. $F_0$ was sampled at regular
intervals based on percentage of total vowel duration. The raw $f_0$ samples were converted to
$z$-scores within speaker, an effective normalization procedure for tone, before averaging.
Figure 1 demonstrates that our speakers produced $f_0$ trajectories across vowels that are highly
consistent with previous findings for Mandarin Chinese, including a small effect of intrinsic $f_0$
on static high (tone 1) and low (tone 3) tones (c.f., Shi and Zhang 1987).

Figure 1: Average normalized $f_0$, y-axis, plotted by time expressed as a percentages of total
vowel duration, x-axis, for each combination of tone (T1-T4) and vowel (/i/-/a/-/u/) in the
corpus.
The effect of tone on vowel targets

Figure 2 provides a summary of tongue and jaw position within the mid-sagittal 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 Tongue Tip (TT), Tongue Body (TB) and Tongue Dorsum (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/. Significant effects of tone on vowel position are indicated by dotted red rectangles drawn at TB for /i/ and at TB, TT, and Jaw for /a/.

Figure 2: Mean midsagittal tongue position (12 repetitions, 6 speakers) for Mandarin vowels /i/-/a/-/u/ produced with Tones 1 'high', 2 'low-high', 3 'low', 4 'high-low'. 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). Dotted line rectangles indicate where mean lingual sensor positions differed significantly with tone.

Figure 3 compares vowel target 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/ \(F(3,15)=4.58, p < .05\). Post-hoc tests indicate that the significant effect of tone is attributable to tones 2 (rising) and 3 (low). These tones – those that begin low – were significantly different from tones 1 (high) and 4 (falling) – those that begin high – for both /a/ (TB, TT, and jaw receiver coils) and /i/ (TB only). The difference between tone 2 and tone 3 was not significant, nor was the difference between tone 1 and tone 4. Effects of tone on /u/ were not significant.
Figure 3: Mean height (12 repetitions, 6 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 Tones 1 'high', 2 'low-high', 3 'low', 4 'high-low'. Error bars represent 95% confidence intervals.

The structure of Figure 4 parallels that of Figure 3. Figure 4 shows the longitudinal dimension of vowel targets for TD, TB, TT, and Jaw receiver coils by tone. The only significant effect of tone on longitudinal position was found for /a/ at the TT \(F(3,15)=4.35, p < .05\). Post-hoc tests indicated that this difference is attributable to tone 3 (low tone), which was significantly different from tone 1 (high), tone 2 (rising), and tone 4 (falling).

Figure 4: Mean longitudinal position of Tongue Dorsum (TD), Tongue Body (TB), and Tongue Tip (TT), and jaw sensors for Mandarin vowels /a/ (left), /i/ (middle), and /u/ (right) produced with Tones 1 'high', 2 'low-high', 3 'low', 4 'high-low'. Error bars represent 95% confidence intervals.

**Jaw results**

To further investigate the relationship between jaw and TB movement, we computed the Euclidean distance between the TB sensor, where opposite effects of tone were found for /a/ and /i/, and the jaw sensor. Unlike the analysis reported above, this measure takes into account changes across tones in both vertical and longitudinal dimensions. Figure 5 summarizes the measurements across speakers. As expected, the TB sensor is farthest away from the jaw sensor for the /u/ target (42.3mm), followed by the /a/ target (40.9mm), and then the /i/ target (38.9mm), 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. A repeated measures ANOVA on TB-to-jaw distance with tone and vowel as independent factors showed a marginal effect of vowel \([F(3,15)=3.83, p = .058]\) but not tone \([F(3,15) < 1]\) and no interaction between tone and vowel \([F(3,15) < 1]\). The null effect of tone on TB-to-jaw distance for /u/ can be expected, since neither the TB sensor nor the jaw sensor was individually influenced by tone. For /a/, we have already seen that both the jaw and the TB sensor were influenced by tone in the vertical dimension. We now see that these parallel movements maintain a fixed distance between TB and jaw sensors. This indicates that the magnitude of jaw displacement is comparable to the magnitude of TB displacement. The result reinforces our view, expressed above in the discussion of /a/, that the effect of tone on vowel targets is mediated by jaw movement. The stable TB-to-Jaw distance for /i/ indicates that, here also, the significant effect of tone on vowel height can be attributed to the jaw. This was not apparent from analyses in 3.2 in part because the contribution of the jaw to TB position is divided across vertical and longitudinal dimensions. The TB-to-jaw distance incorporates these into a single measure, bringing out an invariance masked by single dimensional analyses.

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.

Discussion

Effect of the Jaw on vowel position

Effects of tone on vowel target position, summarized in Figure 2, were observed in two of the three vowel contexts examined: tones that start low (tone 2 and tone 3) pattern together in
their influence on tongue position. The results for /a/ production are consistent with the findings of Erikson et al. (2004), who observed that the tongue body is lower (and F1 higher) for /a/ produced with a low tone. The broader constellation of effects found for /a/ with tones that start low (lower jaw, lower TB, lower and more posterior TT) points to a common physiological explanation. Reduction of vocal fold tension for low tones can be achieved by lowering the larynx (Honda et al. 1999; Moisik et al. 2014), which could pull the jaw down (Honda 1995). Jaw movement in the opening phase of vowels involves a rotational component, whereby the jaw rotates around a terminal hinge, the temporomandibular joint, and vertical and horizontal translations of that axis (Edwards and Harris 1990). The pattern of effects for /a/ is as expected if the effect of tone on vowel position is mediated by the rotational component of jaw movement. Since the jaw lowers in an arc-like motion, greater lingual displacement as a function of jaw movement is expected for sensors distal to the temporomandibal joint. On this account, the relative stability of the TD sensor follows from its posterior position. Given a degree difference in jaw rotation, the magnitude of the effect on sensor displacement is proportional to the distance from the terminal hinge. Thus, the more anterior lingual sensors, TT and TB, show larger effects of tone, as expected if driven by rotational jaw movement. Moreover, for the most anterior lingual sensor, TT, lowering goes hand in hand with retraction. This also is expected if the arc-like motion of the jaw is driving the effect of tone on tongue position.

Differential effect of tone on /a/ and /i/

Accounting for the effect of tone on /i/ production is less straightforward. The only significant difference was found at the TB sensor in the vertical dimension. As with /a/, tones that start low, tone 2 and tone 3, pattern together. However, unlike for /a/, tones 2 and 3 influence /i/ in the opposite direction. For /i/, the TB was higher for tone 2 and tone 3 than for tones that start
high, tone 1 and tone 4. Figure 3 zooms in on these differences comparing the effect for /i/ (middle panels) with that found for /a/ (left panels). At the TB, tones 2 and 3 pattern together but they influence /a/ and /i/ in different directions.

The physiological explanation we offered for the effect of tone 2 and 3 on lingual position for /a/ does not generalize straight-forwardly to /i/. However, it may be the case that TB is raised for /i/ with low tones to keep the acoustics of /i/ stable across the four tones by countering mechanistic factors. In other words, the same pull of low tones on the jaw may receive lingual compensation for /i/ but not /a/. This may seem odd from the standpoint of articulatory-acoustic dynamics, where the formant values of /i/ are relatively robust to articulatory variation (Stevens 1989), but the compensation account is reasonable given the vowel space of Mandarin Chinese. Mandarin has only one low monophthong, /a/, but is comparatively crowded in the upper regions of the vowel space, containing /y/, /i/, and /u/ monophthongs (Duanmu 2007). Lingual compensation for tone-induced pull on the jaw for /i/ and /u/ may be driven by increased articulatory precision required to maintain contrast between non-low vowels.

Alternatively, it is possible that low tone production with /i/ involves a different laryngeal mechanism than low tone production with /a/. Recent work on Chinese tones has established two mechanisms of laryngeal articulation engaged in low tone production, one which involves larynx lowering and one which involves larynx raising together with laryngeal constriction (Moisik et al., 2014). Of particular interest is that the stimuli in Moisik et al. (2014) 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/. A complication in applying this mechanistic account to our data involves the
fact that tones 2 and 3, the tones that start low, pattern together in their influence on TB height whereas Moisik et al. (2014) observe larynx raising only for tone 3.

**Stability of /u/ across tones**

In contrast to /a/ and /i/, which both showed tone-conditioned variation, lingual position for /u/ was stable at all three lingual sensors (Fig. 3 and Fig. 4), and in the relative position of the tongue and jaw (Fig. 5). At least in the opening phase of the vowel – the focus of our analysis here – /u/ is resistant to the coarticulatory influence of tone. This stability is potentially attributable to active control of the jaw, part of the coordinative structure supporting rounding. If so, we predict similar coarticulatory resistance for other rounded vowels, e.g., /y/ and /o/.

**Tone-vowel (in)dependence**

Significant effects of tone on lingual position measured at the vowel target were observed for two of the three vowels examined. This result may appear at first blush to support the hypothesis that tones and vowels are inter-dependent, as in the ‘finals’ of traditional Chinese phonology, more holistic accounts of lexical representation, such as exemplar theory (Pierrehumbert 2002), or other theories that advocate speech production units larger than the vowel, e.g., Fujimura's (1986) icebergs. We offered a partial physiological explanation for why /a/ is lowered for tones that start low, tone 2 and 3. The effect of these tones on /i/ was in the opposite direction. For /i/, TB was higher for tones 2 and 3. We speculated on possible mechanistic (Moisik et al. 2014) and functional accounts for this pattern. However, with respect to the question of tone-vowel independence, maintenance of small but systematic differences in vowel target as a function of tone supports an integrated representational hypothesis. In the absence of a model that can account for why low tones lead to higher TB for /i/ and lower TB for /a/, the pattern appears arbitrary and, as such, supports the inter-dependence view of tone-vowel relations. TB height may vary across /i1/, /i2/, /i3/, and /i4/ in
our data because each of these ‘finals’ are independent units of speech production, or because
/pi1/, /pi2/, /pi3/, and /pi4/ are all different words of Chinese.

Tone-specific vowel variation may contribute to rapid recognition of tones. There is recent
evidence that the time course of tone perception is influenced by vowel quality such that, in
particular, the rising tone (tone 2) is recognized faster when it is produced with /a/ and /i/ than
with /u/ (Shaw et al. 2013). These are the vowel contexts in which the rising tone exerted the
greatest coarticulatory influence on TB position in our data. Early recognition of the rising
tone (which has a later f0 inflection point than the other Mandarin tones) for just these vowels
might be attributable to tone-specific variation in tongue height or to the relationship between
f0 and lingual position. Lingual targets unique to tone-vowel combinations may therefore
function to enhance spoken word recognition for words minimally differentiated by tone.

In contrast to the case for tone-vowel inter-dependence exposed above, we believe that the
data also can be viewed as unequivocally supporting tone-vowel independence. However, this
interpretation of the data requires that we either focus on certain areas of the tongue, e.g., the
stable TD sensor, or that we pursue an alternative expression of vowel targets, the position of
the tongue relative to the jaw.

Vowel targets are typically considered to be spatial positions corresponding, for example, to
the static images of x-rays (e.g., Stevens and House, 1955). Although details of specific
models vary, we take the standard view of vowel targets to involve the position of the tongue
relative to the palate. This can be expressed in terms of constriction location and degree, as in
Task Dynamics (Saltzman and Munhall, 1989), or as a fixed target in space with quantifiable
dimensions (Guenther, 1995). Although there are important differences between these models
of vowel targets, they have in common that the production goal is not expressed in terms of
the relation between active articulators. Rather, the production goals are expressed
independently of the coordinative structures that may achieve them. On this view, to see the
data as supporting tone-vowel independence requires focusing on a specific portion of the
tongue. Only the TD sensor remains stable across tones. On the view that the entire surface of
the tongue contributes to the vowel target, our data instead indicates that vowel targets in
Chinese vary with tones in ways that are not fully predictable from physiological constraints
on coarticulation, at least not according to our current understanding. This conclusion comes
with the important caveat that our data, sourced from lingual fleshpoints, underdetermine the
entire shape of the tongue. Additionally, we have focused on the midsagittal plane and on the
spatial position of the receiver coils at the vowel target. Complimentary data from other
aspects of vowel articulation may reveal other areas of relative stability across tones (or other
systematic influences on tone). The view that emerges from three key lingual articulatory
markers, is that only the TD is stable for each vowel across variation in tone. The isolated
stability at this area of the tongue is consistent with the TD as the low dimensional articulator
under control in vowel production.

Besides restricting our definition of vowel target to the TD sensor, there is an alternative
expression of vowel targets that permits an unequivocal interpretation of the data in terms of
tone-vowel independence. This alternative is that it is the relation between articulators that
serves to dictate production goals for these vowels. Seen through the lens of a relative notion
of vowel target, our data provide strong support for tone-vowel independence. The
relationship between tongue position and jaw position remained stable across tones, even as
sensor position varied, e.g. at the TB sensor for /a/ and /i/. As a consequence, the Euclidean
distance between the TB sensor and the jaw differed for phonologically distinct vowels, /a/,
/i/, and /u/, but remained constant across tones. It is therefore possible to characterize the three
Chinese vowels in this study in terms of a single dimension, the relationship between lingual
and jaw position. On this view, spatial variation in lingual position need not disturb
achievement of vowel targets, as long as the jaw is free to co-vary in accordance with a vowel-specific tongue-jaw relation.

Of the two perspectives on the data that permit interpretations in terms of tone-vowel independence, it is not yet clear which (TB-to-jaw distance or TD position) is more important for vowel targets.

**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 influence of tone and vowel production on jaw position. Others, i.e., the higher position of TB for /i/ when produced with a rising (low-high) tone, do not have a clear physiological basis, but may function to enhance tone perception, as evidenced in spoken word recognition tasks that evaluate the timecourse 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 offer a nuanced picture of speech production goals consistent at once with low-dimensional vowel specification (in terms of TD constriction or tongue-jaw relations), as well as a richer phonetic specification of particular tone-vowel combinations or possibly words.

**Acknowledgements**

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 also 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. Research was funded by a grant from the MARCS Institute to Jason A. Shaw, Michael Tyler, Michael Proctor, Donald Derrick and Chong Han.

References


Figure 1: Average normalized f0, y-axis, plotted by time expressed as a percentages of total vowel duration, x-axis, for each combination of tone (T1-T4) and vowel (/i/-/a/-/u/) in the corpus.
Figure 2: Mean midsagittal tongue position (12 repetitions, 6 speakers) for Mandarin vowels /i/-/a/-/u/ produced with Tones 1 'high', 2 'low-high', 3 'low', 4 'high-low'. Tongue edges are represented as quadratic fits between Tongue Tip (TT), Tongue Blade (TB) and Tongue Dorsum (TD) sensors (within vowel and within tone). Dotted line rectangles indicate where mean lingual sensor positions differed significantly with tone.
Figure 3: Mean height (12 repetitions, 6 speakers) of tongue dorsum (TD), tongue blade (TB), and tongue tip (TT) sensors for Mandarin vowels /a/ (left), /i/ (middle), and /u/ (right) produced with Tones 1 'high', 2 'low-high', 3 'low', 4 'high-low'. Error bars represent 95% confidence intervals.
Figure 4: Mean longitudinal position of tongue dorsum (TD), tongue blade (TB), and tongue tip (TT) sensors for Mandarin vowels /a/ (left), /i/ (middle), and /u/ (right) produced with Tones 1 'high', 2 'low-high', 3 'low', 4 'high-low'. Error bars represent 95% confidence intervals.
Figure 5: Mean Euclidean distance between the Tongue Blade (TB) sensor and the Jaw sensor for vowels /a/-/i/-/u/, produced with each tone (1-4). Error bars indicate 95% confidence intervals.