Tone-Vowel Interaction in Standard Chinese

Philip Hoole[1] *& Fang Hu*[2]

[1]Phonetics Institute, Munich University
[2]Phonetics Lab, Institute of Linguistics, Chinese Academy of Social Sciences
hoole@phonetik.uni-muenchen.de | hufang@cass.org.cn

Abstract

The effect of tone production on the position of the articulators (tongue, jaw, lower lip) was examined by means of three-dimensional electromagnetic articulography. Most differences found involved Tone 3 relative to the other tones. While /u, a/ showed retraction and lowering this pattern did not extend to all vowels. Head position during vowel production was also analyzed. Even more consistently than for the articulatory data, Tone 3 tended to differ from the other tones, showing a lower and more posterior head position. Finally, tongue and jaw influences on intrinsic pitch in Mandarin are briefly considered.

1. Introduction

In this paper we present a pilot study using the new 3D articulograph to investigate the relationship between tone production and supraglottal articulations. There are perhaps three main angles from which this area has been approached: Firstly, there is the question of whether the tongue (and also lip and jaw) position for a given vowel varies systematically with the tone. Acoustic analysis by Zee [10] indicated that statistically significant differences are readily found, but that the patterns of formant change differed over speakers. Torng [7] found no statistically reliable pattern over a group of six speakers, but occasional instances of consistent differences in jaw height for specific vowels and speakers (combined with some cases of lingual compensation for greater jaw opening). Erickson et al. [2] found retraction for Tone 3 vs. Tone 1 for the vowel /a/.

Thus the present evidence indicates that some articulatory differences can be expected, but it is not yet clear if they have any functional relevance or simply reflect articulatory idiosyncracies of individual speakers.

The second area is related to the observation that perception of tone may be possible to some extent based on visual information alone (Burnham et al., [1]). The articulatory substrate to this is not yet clear, but one possibility that has been raised is that F0 and head movement may be systematically related (Yehia et al., [9]). We here present some preliminary data on this issue, by exploiting the fact that the 3D articulograph can monitor head motion during speech.

The third area linking articulation and tone production involves the question of intrinsic pitch differences. Data in e.g Zee [10] and Torng et al. [8] indicates that the well-known intrinsic pitch differences between high and low vowels certainly occur in tone languages. One specific area of discussion with regard to intrinsic pitch is whether it is more closely related to tongue or to jaw height. We will touch briefly on this issue below.

2. Methods

The 3D articulograph (AG500, Carstens Medizinelektronik) was used to monitor the movement of tongue, jaw, lower-lip and head. Three sensors were mounted on the tongue, at distances of approximately 1, 3 and 5cm from the tongue-tip respectively. The 3D articulograph is actually better regarded as a 5-dimensional system, since each sensor provides not only three spatial coordinates, but also two angular coordinates. In practice, this means that it is possible to monitor not only the position of tongue fleshpoints, but also the angle of a tangent to the surface of the tongue at each fleshpoint. The information on sensor orientation has a further important advantage since it makes it possible to recover the 6 degrees of freedom of head movement using only two sensors (located on upper incisors and nose). For articulatory analysis it is essential to recover head motion in the 3D system, since the head is free to move within the transmitter assembly. Head movements have an undoubted communicative function and are thus always available as a byproduct of this normalization procedure. See Hoole et al., [5] for further background and assessment of this new articulographic system.

The articulatory data was recorded at a sample rate of 200Hz and low-passed filtered at 25Hz. The data was oriented such that the principal component of jaw movement was vertical. The audio signal was recorded together with synchronization information on DAT tape. In addition, front-view and profile video recordings were made.

3. Speech Material

The core material consisted of the five Mandarin vowels /i/, /y/, /u/, /ɤ/ and /a/. These were spoken as isolated vowels on each of Mandarin tones 1 to 4 (high, high rising, low falling rising, and high falling). Each group of vowels was spoken first on Tone 1, then on Tone 2, and so on. The vowel /o/, which has a rather peripheral status in the Mandarin vowel system, was recorded on Tone 1 only. (Schwa, diphthongs, apical vowels and rhotacized vowels were also recorded but are not discussed here.) The complete material was repeated 12 times. To date, one female subject has been recorded, a trained phonetician, native of Beijing.

4. Results

4.1 Overview of articulator configurations for the target vowels

The first group of results simply involved extracting articulator positions at the midpoint of the vowels.

Fig. 1 shows average tongue positions for all vowels as spoken with Tone 1. Since the rearmost sensor was not located very far back, essentially only the oral region of the tongue is seen. Basically, the configurations conform to expectations. /i/ has the highest and most fronted tongue shape. /y/ is marginally lower and further back, a pattern which appears to be typical of many languages. /u/, /ɤ/ and /o/ share a similar tongue shape, distinguishing them from /a/. The back unrounded vowel is overall closer to /u/ than to /o/, but the picture is complicated by it being relatively high at the front part of the mouth

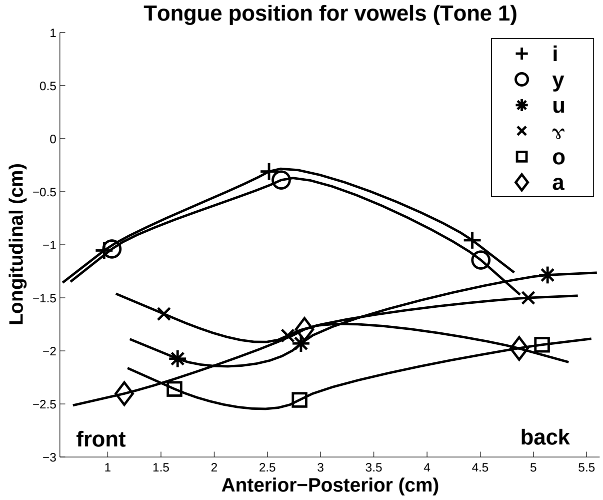

Figure 1: *Reconstructed tongue contours for each target vowel (spoken on Tone 1). Tongue contours are extrapolated 5mm beyond the front and rear sensor using sensor orientation information.*

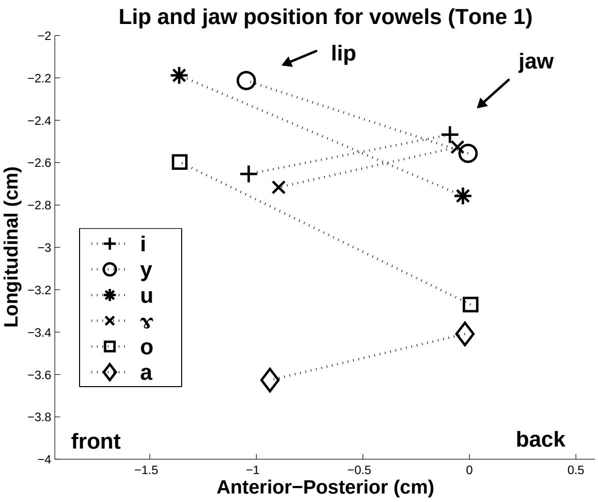

Figure 2: *Lower-lip and jaw coordinates (joined by dotted line) for each target vowel.*

The corresponding lip and jaw positions are shown in Fig. 2. /i/, /y/ and /ɤ/ share higher jaw positions, /u/ is marginally lower, and /a/ and /o/ share low jaw positions. These jaw positions will be returned to briefly below. The lip position is slightly more retracted for /ɤ/ than for /i/, and less protruded for /y/ than for /u/. The video indicated a possibly more inrounded quality for /y/.

4.2 Tone-related articulatory differences

4.2.1. Tongue

In order to test for effects of tone on articulator position the following procedure was followed: The essential articulatory data for the tongue consists of the anterior-posterior and longitudinal (up-down) coordinates of the three sensors on the tongue (the angular coordinates and possibly even the lateral coordinates of the sensors may also be of interest but will be considered in detail at a later date). In order to bypass some of the problems inherent in making multiple comparisons on this total of 6 coordinates the data was first subjected to principal components analysis. This indicated that two factors together explained about 96% of the variance: 87% for Factor 1, and 9% Factor 2. Further statistical testing was based on the factor scores for these two factors. The distribution of each vowel-tone combination in the space of the first two factors is shown in Fig. 3, each data point representing the average over the 12 repetitions. An appreciation of the families of tongue shapes associated with each factor can be obtained by referring back to Fig. 1.

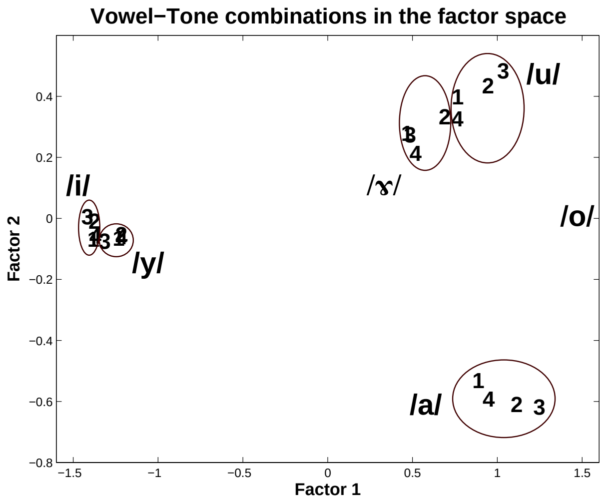

Figure 3: *All vowel-tone combinations in the space of the first two principal components of the tongue data. Tones indicated by numerals. Ellipses indicate which data points belong to which vowel, but do not indicate the actual dispersion of the raw data. /o/ was spoken on Tone 1 only.*

Factor 1 essentially captures the distinction between high and low position of the front part of the tongue (i.e /i, y/ versus all other vowels), while Factor 2 captures the difference in tongue shape between /a/ on the one hand, and high back vowels on the other. It should be emphasized that the PC analysis should be seen here as a data reduction technique and not as a definitive view of the articulatory space for Mandarin vowels. A somewhat different picture could emerge from recordings with sensors located further back on the tongue, and with the inclusion of rhotacized and apical vowels in the corpus.

In terms of effects of tone, the main feature to emerge from Fig. 3 is a tendency forTone 3 vowels to be located at more extreme locations of the factor space. This can be observed for all vowels except /ɤ/ (the magnitude of the effects being admittedly pretty small for the front vowels /i, y/). Since there was clearly no articulatory pattern common to all vowels (in line with the previous results of Zee and Torng [10,7]) we carried out one-way ANOVAs separately for each vowel, with Tone as independent variable. Scheffé post-hoc tests were then used to determine which tones differed significantly from each other. For /a/, Tone 3 was highly significantly different from Tones 1 and 4 with respect to Factor 1 ($p<0.01$). In addition, Tone 2 differed from Tone 1 at $p<0.01$. For /u/, Tone 3 was highly significantly different from Tone 4 with respect to both Factor 1 and 2. There was a weakly significant difference between Tone 2 and Tone 4 for Factor 2. For /y/ Tone 3 differed from all other tones at $p<0.01$ for Factor 1. For /i/ only a weakly significant difference between Tone 3 and Tone 1 was found for Factor 2. For /ɤ/ no significant effects were found.

Fig. 4 shows the tongue configuration for two of the main cases where strong effects were found, namely Tone 3 vs. Tone 4 for /a, u/. Both vowels show a lowered and more retracted configuration of the order of 2mm for Tone 3 (most clearly at more posterior locations on the tongue). The problem for an articulatory generalization of these effects lies in the fact that the significant effect for /y/ was essentially in the opposite direction (albeit of restricted magnitude: about 1mm at the front two sensor locations).

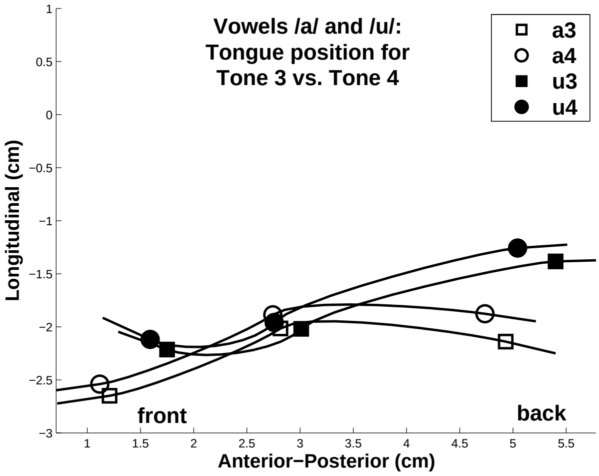

Figure 4: *Tongue contours for two of the clearest examples of tone-related effects.*

4.2.2 *Jaw and lip*

An obvious question is whether any of the tone-related differences in tongue position can be attributed to jaw position. The evidence for this turned out to be somewhat equivocal. For /a/, jaw height for Tone 3 tended to be relatively low (as might be expected from the tongue results), but failed to reach significance. However, there was a highly significantly more *posterior* jaw position for Tone 3 than Tone 4. The same tendency applied to /u/, although the result was only weakly significant. For /i/ and /y/, the jaw was *higher* for Tone 3 than Tone 4 at $p<0.01$, corresponding to a difference of about 1mm.

For the lip, few significant effects were found, the main exception involving greater retraction for Tone 3 vs. Tone 4 for /a, u/, thus paralleling the results for the jaw.

Future analysis will look at this area in more detail by re-examining the results for tongue and lip after subtracting out the jaw contribution.

4.3 Analysis of head position

The first group of analyses simply extracted head position and orientation at the midpoint of the vowel, as done for the articulatory analyses above. Additional analyses below look at peak movement velocities determined over the vowel as a whole.

The absolute values of head position are unlikely to be of great interest, since clearly speakers are quite able to speak any desired intonation contour (and presumably tone) while adopting a wide range of head postures. Accordingly all head positions were expressed relative to the mean position for each block of repetitions (this also roughly compensates for overall changes in body posture over the course of the experiment).

Fig. 5 shows the results for vertical head position, broken down by vowel and tone. It will be observed that Tone 3 shows overall the lowest head position, especially in comparison to Tone 2 (of the order of 2mm). Statistical analysis indicated that Tone 3 was indeed lower than Tone 2 at $p<0.01$ for all vowels except /a/. (It was also lower than Tones 1 and 4 for /ɤ/ and Tone 4 for /y/. The only other significant pairwise difference was between Tones 1 and 2 for /u/.) Exactly the same pattern of siginificances was found for rotation in the sagittal plane. Essentially, the head sensors rotated downwards by about 1deg. for Tone 3 relative to Tone 2. Even clearer effects were found for anterior-posterior position of the head. Tone 3 showed a highly significantly more retracted position compared to Tone 2 for all vowels except /a/. Moreover, the magnitude of the effects was larger (up to about 4mm), and for /u, ɤ, y/ Tone 3 was also significantly more retracted than Tones 1 and 4.

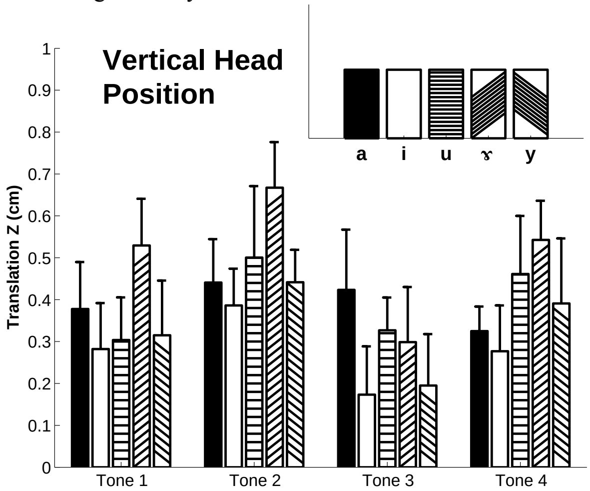

Figure 5: *Average vertical head position (error bars indicate 1 sd) for each vowel-tone combination, relative to an arbitrary reference position.*

From this analysis quite consistent evidence thus emerges for characteristic differences in head posture, especially with respect to Tone 3. As a supplement to this static analysis we also considered head movement velocity. Velocities represent attractive measures in this context, since they do not require normalization to some assumed reference head posture. They may also be suspected to be communicatively robust, i.e movement in a particular direction (or even just movement) may be more salient to perceivers than subtle differences in position (cf. Keating et al., [6]). For this preliminary analysis we simply extracted maximum velocities in the upward and downward, forward and backward directions over the course of each vowel. Given the crude nature of these measures, the results were surprisingly consistent. Briefly, upward and forward movement velocities tended to be higher (by the modest amount of about 10mm/s) for Tone 3, distinguishing this tone from various combinations of the other tones (as in the positional analyses, no significant effects were found for vowel /a/). This is an interesting result since it represents movement away from the typically lower and more retracted head position for Tone 3, and indicates that closer consideration of the complete time-course of head-movement would be worthwhile.

4.4 Intrinsic Pitch

The final set of results will look just at one rather striking effect presented by Torng et al. [8]: They found for several speakers an unexpectedly low F0 for /i/, and at the same time a low jaw position (tongue position was nonetheless high, as would be expected for this vowel). Conversely, /o/ had a high intrinsic pitch, and at the same time an unexpectedly high jaw position. Confirmation of this result would be very interesting, because there remain unresolved issues regarding the relative weight to be attached to jaw and tongue position as influences on intrinsic pitch (see e.g Hoole & Mooshammer [4], for discussion of German). Regarding jaw position, neither effect could be replicated: Referring back to Fig. 2 shows that /i/ has a high jaw position, while /o/ has a low one. The results for F0 were also unremarkable: neither /i/ nor /o/ had especially high or low F0 values (the material for /o/ being rather restricted as it was only recorded on Tone 1). The main tendencies involved /u/, which had the highest F0 for every tone, and /a/ which was sigificantly lower than at least 3 of 4 vowels on Tones 2 and 4.

The isolated vowels recorded in this study are admittedly not ideal for detailed examinination of intrinsic pitch. We would simply speculate in conclusion that at least the vowel /i/ may have a wide latitude in jaw position in Mandarin because of the absence of low front vowels (though the same may not apply to high front /y/); consequently, it could be interesting to examine tongue-jaw coordination for /i/, and relationships to F0, over a range of vocalic and consonantal contexts that could be expected to induce coarticulatory effects on jaw position in /i/.

5. Discussion

The most consistent pattern to emerge from this pilot study is that when vowel midpoint is examined, then Tone 3, the lowest tone, tends to show both articulatory and head-position differences from the other tones. Regarding articulation, the retraction (plus some lowering) of tongue and jaw for /a/ and /u/ fits in well with the results of Erickson et al. [2], and may be part of the mechansim identified by Honda et al. [3] for pitch lowering associated with larynx lowering. However, this result did not generalize to all vowels. While it is acoustically conceivable that retraction and lowering may unduly compromise high front vowels, it remains unclear why /y/ should actually show small but significant tongue *raising* (and also unclear why back unrounded /ɤ/ did not behave like /u/). For the head-movement results, too, a number of interesting issues remain open. Above all, there is the question of the closeness of the relationship between head-movement and F0. This involves looking at the complete time-course of F0 and head-movement for all vowels and tones (Tones 2 and 4 of course involve considerable F0 movement, which our midpoint analysis has ignored). Yehia et al. [9] found that while F0 could be predicted quite well from head position at the level of individual utterances, the nature of the prediction varied considerably from utterance to utterance, suggesting some kind of functional (rather than a mechanical) relationship. Our guess at the moment is that the speaker may have learnt to associate a preferred pattern of head posturing with Tone 3 as a kind of additional visual prosodic cue, but that the posture is unlikely to track the F0 pattern in any close way.

Finally, it is obviously important to test the reproducibility both of the articulatory and head-movement effects by examining tone production in a wider range of word and sentence contexts.

References

[1] Burnham, D.; Lau, S.; Tam, H.; Schoknecht, C., 2001. Visual discrimination of Cantonese tone by tonal but non-Cantonese speakers, and by non-tonal language speakers. *Proc. AVSP 2001*, 155-160.

[2] Erickson, D.; Iwata, R.; Endo, M.; Fujino, A., 2004. Effect of tone height on jaw and tongue articulation in Mandarin Chinese. *Proc. Tonal aspects of languages, Beijing.*

[3] 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.

[4] Hoole, P.; Mooshammer, C., 2002. Articulatory analysis of the German vowel system. In *Silbenschnitt und Tonakzente*, P. Auer, P. Gilles & H. Spiekermann (eds.),. Niemeyer, Tübingen.

[5] Hoole, P.; Zierdt, A.; Geng, C., 2003. Beyond 2D in articulatory data acquisition and analysis. *Proc. XVth Int. Cong. Phonetic Sci.*, 265-268.
An updated version with more recent analysis is available from:
http://www.phonetik.uni-muenchen.de/~hoole/pdf/icphs03_talk_ann.pdf

[6] Keating, P.; Baroni, M.; Mattys, S.; Scarborough, R.; Alwan, A.; Auer, E.; Bernstein, L., 2003. Optical Phonetics and Visual Perception of Lexical and Phrasal Stress in English. *Proc. XVth Int. Cong. Phonetic Sci.*, 2071-2074

[7] Torng, P.-C., 2000. *Supralaryngeal articulator movements and laryngeal control in Mandarin Chinese tonal production*. Doctoral dissertation, University of Illinois at Urbana-Champaign.

[8] Torng, P.-C.; van Lieshout, P.H.H.M.; Alfonso, P. J., 2001. Articulator position and laryngeal control in Mandarin vowels. In *Speech motor control in normal and disorderd speech*, B. Maassen, W. Hulstijn, R. Kent, H. Peters, P. van Lieshout (eds.). 4th International Speech Motor Conference, Nijmegen, 82-85

[9] Yehia, H.; Kuratate, T.; Vatikiotis-Bateson, E., 2002. Linking facial animation, head motion and speech acoustics. *J. Phonetics* 30 (3), 555-568.

[10] Zee, E., 1980. Tone and vowel quality. *J. Phonetics*, 6, 247-258.