On the Lingual Articulation in Vowel Production: Case Study from Ningbo Chinese[λ]

Fang Hu

Phonetics Lab, Institute of Linguistics, Chinese Academy of Social Sciences
5 Jian Guo Men Nei Street, Beijing 100732, China

`hufang@cass.org.cn`

Abstract. *Lingual articulation is of particular importance to understand both physiological and acoustic aspects of speech production. This paper presents a PARAFAC modeling of lingual articulation in vowel production in Ningbo Chinese. The lingual articulatory data were acquired from seven Ningbo speakers using Carstens's Eelectromagnetic Articulograph (EMA). Results show that a two-factor model best captures the tongue movement of vowel production, and the model explains about 90% of variance. The PARAFAC modeling of vowels is a speaker-independent generalization concerning the sampled tongue positions and the inferred lingual gestures. The results from Ningbo Chinese are consistent with those from English and other European languages. The fact that the extracted lingual movement mechanisms are comparable to the functional representation of tongue muscle forces from the EMG study (Maeda & Honda, 1994; Honda, 2000) suggests that the PARAFAC model of lingual articulation has physiological implications and reflexes speech motor organization for vowel articulation.*

1. Introduction

Speech articulation is controlled by the speech motor system through muscular activities. Among the articulators, the tongue plays a crucial role in configuring the vocal tract shape, which further determines speech acoustics. Lingual articulation is of particular importance to understand both physiological and acoustic aspects of speech production. This paper examines lingual articulation in vowel production in Ningbo Chinese.

Vowels are traditionally described in terms of vowel height and vowel backness. However, this classical model of vowel description has never been validated by empirical articulatory data. In fact, it has been contradicted by a large number of lingual articulatory data based on various kinds of techniques in the literature (Russel, 1928; Ladefoged, 1967; Ladefoged et al., 1972). Now scholars tend to interpret the traditional vowel chart, e.g. the IPA vowel chart, as acoustically or psycho-acoustically, rather than articulatorily based. For the Ningbo vowels, it has also been shown that there is a good correlation between major vowel features, vowel height and vowel backness, and vowel acoustics (Hu, 2005). Figure 1 shows the acoustic vowel space for the 10 Ningbo vowels from 10 male (left) and 10 female (right) speakers. As can be seen from the figure, both the male and female data exhibit a good correlation between the two vowel features, height and backness, and the vowel formants, for instance, a smaller F1 value is correlated with a greater degree of vowel height and a smaller F2 value with a greater degree of vowel backness

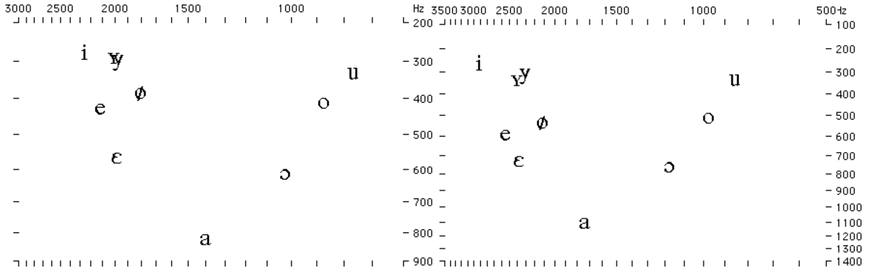

Figure 1. The acoustic vowel space for the 10 Ningbo vowels from 10 male (left) and 10 female (right) speakers.

Yet it still remains unclear how these auditorily-acoustic distinctions of vowel height and vowel backness are achieved articulatorily. Past studies have shown that the PARAFAC analysis captures the underlying mechanisms for the observed lingual gestural variations during vowel production and the results are physiologically interpretable (Harshman et al., 1977; Hoole, 1999). However, previous studies are mostly based on the data from English or other

European languages. It is then worth having a look at a wider range of languages to check the validity of the methodology and the universality of the inferred articulatory mechanisms.

2. Methodology

The lingual articulatory data of all the ten phonemic vowels [□ □ □ □ ˙∙. □ ↦ □ ◦ _] that occur in (C)V syllables were acquired from seven speakers, four male and three female, using Carstens Electromagnetic Articulograph (EMA) (see Perkell et al., 1992 for a full background of the technique, and also Hoole, 1996 for detailed methodological concerns). The receiver coils were attached at the midline on the upper lip (UL), the lower lip (LL), the gum ridge at the lower teeth (Jaw), and the three points of the tongue, namely the tongue tip (TT), tongue mid (TM), and tongue dorsum (TD). The TT coil is placed less than 1 cm to the tongue tip, the TM coil is about 3 cm to the TT coil, and the TD coil is about 6 cm to the tongue tip. In addition, two receiver coils were attached to the bridge of the nose (Ref. 1) and the gum ridge at the upper teeth (Ref. 2) respectively, serving as the two reference points.

After the speech data were collected, the orientation of the occusal plane was recorded for each speaker by having the speaker bite lightly on a plate with two transducers mounted on. The raw articulatory data were then rotated in a way such that the new horizontal (x) axis was parallel to the subject's occlusal plane. The rotation provides a base for data comparison between speakers, since articulatory data from different speakers have a same orientation after rotation. In this study, the three sampled tongue points, tongue tip (TT), tongue mid (TM) and tongue dorsum (TD) were extracted for each vowel at the vowel's target position using the velocity minima criterion (Löfqvist et al., 1993; Löfqvist, 1999).

The measured positional tongue-point data were then processed in such a way that the final data subject to the PARAFAC algorithm consisted of displacements from the mean tongue configuration of each speaker. PARAFAC (for "parallel factors") is a three-mode factor analysis model developed by Harshman (1970). Similar to the standard two-mode factor analysis, PARAFAC represents data as the sum of a set of linear components. Yet PARAFAC has the advantage in that it provides a unique solution for the set of factor loadings by extracting potentially overlapping components simultaneously and equating them in parallel, such that it resolves the problem of rotational indeterminacy of the orientation of the factor axes in the standard two-mode factor analysis. Mathematically, the PARAFAC model as applied to tongue movement measurements can be formulated as follows. Given measurements of j vowels from i articulators for k speakers, and assuming n factors are extracted, the results of the PARAFAC are provided in three loading matrices $\mathbf{V}$, $\mathbf{A}$, and $\mathbf{S}$ (for vowels, articulators and speakers respectively) with dimensions $j\times n$, $i\times n$ and $k\times n$ respectively. So, for speaker k, the complete data set matrix $\mathbf{Y}_k$ with the dimension $i\times j$ predicted by the PARAFAC algorism can be written as

$$\mathbf{Y}_k=\mathbf{A}\mathbf{S}_k\mathbf{V}^T \qquad (\mathbf{1})$$

where $\mathbf{S}_k$ is a matrix with the speaker's scaling constant for the factors n on the main diagonal and zero elsewhere, and $\mathbf{V}^T$ is the transpose of $\mathbf{V}$. Thus, in the present study, 6 pieces of articulatory information (3 tongue points × 2 coordinates) for the 10 Ningbo vowels from 7 speakers constitute the data array with the dimension of 6×10×7.

3. Results

Figure 2 shows two examples of the acquired lingual configurational data from Male Speaker 1 (left) and Female Speaker 1 (right). In the figures, the three line-connected IPA symbols represent the mean positions of TT, TM and TD, respectively, from left to right for each vowel. Each speaker's hard palate contour is also plotted in the figures for reference. And the speakers were facing left.

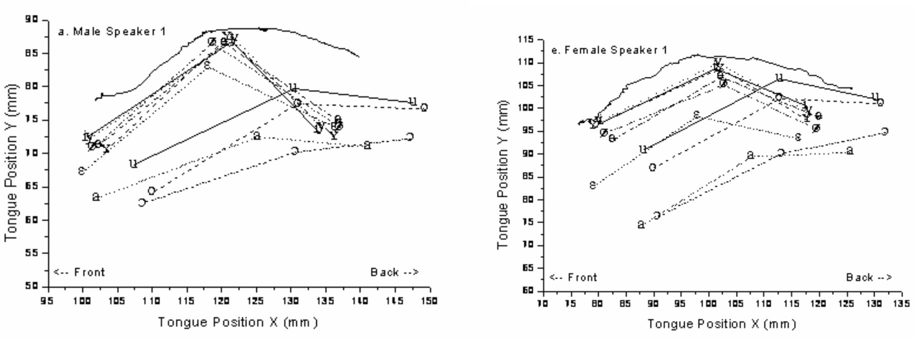

Figure 2. Examples of lingual configurations for the ten Ningbo vowels.

As mentioned above, the acquired lingual data from seven speakers were then analyzed using PARAFAC modeling. The results show that a two-factor model best captures the tongue movement of vowel production, and the model explained about 90% of variance. And modeling solutions in higher dimensionality generated degenerate models. The loading effect of tongue-point articulators was represented vividly by plotting each factor as a pattern of tongue-point displacement around the average tongue position using averaged speaker. The result is shown in Figure 3. And a mean palatal contour averaged across the seven speakers is also added in the plots for reference.

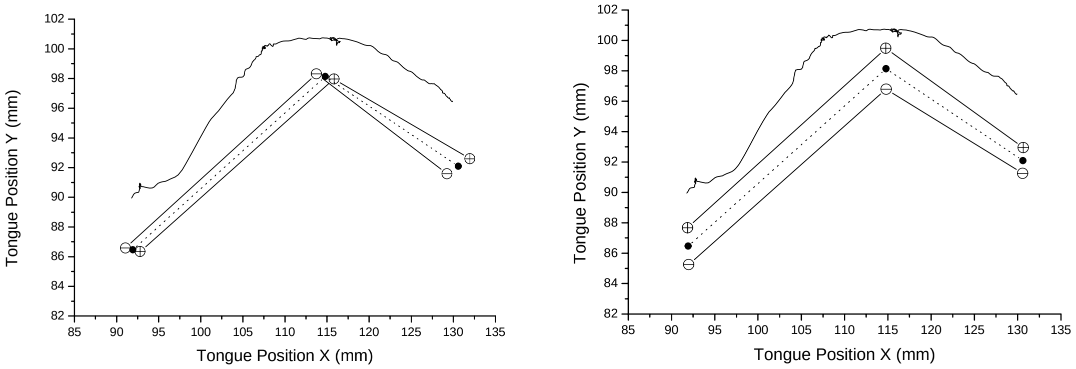

Figure 3. A graphic representation of Factor 1 (left) and Factor 2 (right) for the tongue-point articulators. Solid circles connected by the dotted line represent the mean tongue-point positions using averaged speaker. Plus and Minus centered circles connected by solid lines indicate positive and negative displacement loadings of ±2 standard deviations, respectively.

The first factor shape, as shown in the left panel of Figure 3, is similar to the second factor derived in Harshman et al. (1977) and Hoole (1999). From the figure, we can see that it is mainly characterized by a movement of tongue retraction and some amount of the raising of the posterior part of the tongue. The second factor is nearly identical to the first factor obtained in Harshman et al. and Hoole’s research, which captures a substantial raising of the tongue body towards the hard palatal region, as indicated by the tongue mid (TM) point in the right panel of Figure 3. The two factors extracted were referred to as “retraction and back raising” and “front raising” respectively hereafter.

The production of each Ningbo vowel can then be understood as a combination of the above two factors. That is, vowels can be reconstructed using the extracted two-factor model. Figure 4 shows the results of factor loadings of vowels. The values indicate the weights required for reconstructing each vowel.

It is not surprising that the distribution of the vowels in the Factor 1/Factor 2 space resembles, though is not exactly the same as, that in an acoustic F1/F2 plane or in a traditional vowel chart. And it should be noted that Figure 4 is a purely tongue-based representation of vowel space. Factor 1 values indicate the weights of “retraction and back raising”, and Factor 2 values indicate the weights of “front raising”. A positive value denotes a positive displacement, and a negative value denotes a negative displacement from the mean tongue position. For instance, the vowel [i] requires maximal front raising and a maximum amount of advancement of the tongue, the vowel [o] requires a maximal retraction and back raising, and the vowel [a] requires a maximum amount of front lowering of the tongue.

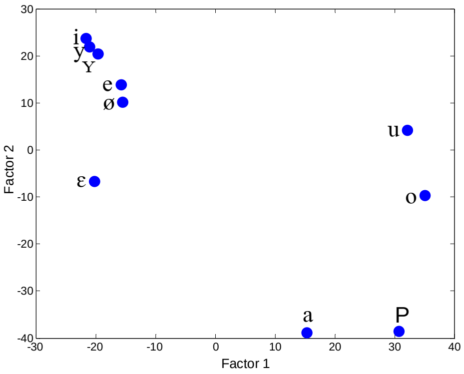

Figure 4. Factor loadings–vowels.

Linguistic vowel features such as vowel backness and vowel height can then be reinterpreted in terms of the extracted Factor 1 and Factor 2 respectively. It can be seen from the figure that all the front vowels need large negative values for Factor 1, which indicates that the tongue body is advanced and the posterior part of the tongue is not raised in producing front vowels. Similarly, all the back vowels need a retracted tongue in production. The low vowel [a] is located somewhere in-between in the figure, indicating a comparatively central tongue position, although it is produced with a relatively retracted tongue, as indicated by a positive Factor 1 value. The front vowels [i y ʏ e ø] all have positive Factor 2 values, which indicates that the front raising of tongue is needed in producing these vowels, although with different degrees. However, the front vowel [ɛ] has a negative Factor 2 value, which indicates that the tongue is lowered in producing it. The fact that the back vowels [o ɔ] and the low vowel [a] all have negative Factor 2 values means that the front part of tongue is lowered to some extent when producing these vowels. As for the high back vowel [u], the detected small positive value of Factor 2 suggests that the vowel, in addition to a great amount of retraction and back raising, needs a slight amount of front raising in production. Concerning the rounded and unrounded distinction, the rounded front vowels and their unrounded counterparts, especially the three high front vowels [i y ʏ], share a similar lingual representation in general. Meanwhile, it should be noted that there are slight differences between the rounded front vowels and their unrounded counterparts, namely the former have a slightly smaller Factor 2 value than the latter, suggesting a slightly lower tongue position for the rounded front vowels.

The speaker weights are shown in Figure 5. The factor coefficients indicate the factor weights needed for each speaker in reconstructing the vowels in the extracted PARAFAC model, and the lingual articulation strategy that speakers employ in vowel production can then be understood as different uses of Factor 1 and Factor 2 weights by different speakers. The inter-speaker difference of lingual articulation of vowels is thus simply interpreted in terms of different lingual articulation strategies employed by different speakers. Comparisons can easily be made between speakers. For instance, Male Speaker 2 (MS2) and Female Speaker 2 (FS2) make a similar light use of Factor 1 (about 0.2) but a very different use of Factor 2 (about 0.55 vs. 0.25) in vowel production; Male Speaker 1 (MS1) and 3 (MS3) and Female Speaker 1 (FS1) and 3 (FS3) are alike in that they are in the middle (about 0.35 to 0.4) among the seven speakers in terms of the use of Factor 1, but differ regarding the use of Factor 2. In summary, Male Speaker 1 and 3 make a light use of Factor 2 and an intermediate use of Factor 1, Male Speaker 2 makes the heaviest use of Factor 2 and a light use of Factor 1, Male Speaker 4 (MS4) makes the heaviest use of Factor 1 (about 0.6) and a relatively lighter use of Factor 2 (about 0.35), Female Speaker 1 makes a relatively heavier use of Factor 2 and an intermediate use of Factor 1, Female Speaker 2 makes the lightest use of both Factor 1 and Factor 2, and Female Speaker 3 makes an intermediate use of Factor 1 and a relatively lighter use of Factor 2.

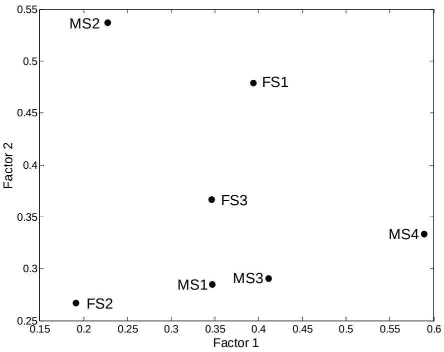

Figure 5. Factor loadings – speakers.

It should be noted that the seven speakers have different coefficients for the two factors is supportive of the stability of the extracted model. As stated in Harshman et al. (1977: 699), the PARAFAC algorithm requires that speakers (at least two among all) use different relative importance of the factors to solve the problem of rotational indeterminacy of extracted factors. What should also be noted is that all the weights have a positive sign in the seven speakers. The extracted factors in a PARAFAC modeling are assumed to capture the speaker-independent features. If different signs of a given factor were detected in the subject weights, it would constitute a violation of the modeling assumption since this would indicate that the factor itself is being used to capture speaker-specific features. The results shown here confirm that the extracted model is a satisfactory one. Thus the model captures the speaker-specific effects in terms of a simple scaling of speaker-independent factors.

4. Conclusion and discussion

The PARAFAC modeling of the tongue-point positions successfully decomposes the complicated tongue shapes into two underlying movement mechanisms, namely the “retraction and back raising” (retraction of the tongue body and the raising of the tongue dorsum towards an area around the velum or a more posterior region) and the “front raising” (the raising of the front part of the tongue towards the hard palate). Vowel backness and vowel height can be re-defined in terms of these two tongue movement mechanisms. Vowel backness can be viewed as a result of “retraction and back raising” of the tongue. The front vowels are produced with the tongue body advanced and with the back part of the tongue lowered, the back vowels are produced with tongue body retracted and with the back part of the tongue raised to some extent, and the low vowel is produced with a comparatively central degree of the tongue retraction and back raising. Vowel height can be interpreted as a result of the “front raising” of the tongue, although somewhat different from the traditionally defined vowel/tongue height, namely front vowels and back vowels should be interpreted within their own series. For the front vowels, the high and mid-high front vowels are produced with a raised front part of the tongue, though with different degrees, and generally, the rounded front vowels and their unrounded counterparts have a similar lingual representation, though the former has a slightly lower lingual configuration than the latter; the mid-low front vowel is produced with the front part of the tongue lowered. Regarding the back vowels, they are produced with different degrees of “front lowering” (i.e., negative “front raising”), as they all have minus values of “front raising”. As for the low vowel, it is produced with a maximum degree of the lowering of the front part of the tongue.

The PARAFAC modeling study is a speaker-independent generalization concerning the sampled tongue positions and the inferred lingual gestures. Meanwhile, it is observed that different speakers may use different strategies in vowel production to achieve a similar acoustic output. As indicated by the speaker-loading factors of PARAFAC modeling, different speakers use different weights of the two extracted factors, “retraction and back raising” and “front raising”.

The results from Ningbo Chinese are consistent with those from other languages. In their pilot study, Harshman et al. (1977) employed the PARAFAC model in examining the tongue shapes of the ten American English vowels produced by five male speakers, and found that the tongue postures could be decomposed into two basic tongue shapes, “a forward movement of the root of the tongue together with a raising of the front of the tongue” and “an upward and backward movement of the tongue” (p. 702). The former was referred to as “front raising” and the latter as “back raising”. Thus a two-factor PARAFAC model was developed and it accounted for 92.7% variance of data. Hoole (1999) also indicated that a two-factor PARAFAC model was sufficient to capture a very substantial proportion of the variance (r^2=92.3% for vowels in the [p _ p] context). He found that the first factor was similar to that in Harshman et al. (1977), i.e., the “front raising”. As for the second factor, mainly some degree of “back raising" was detected and it was thus characterized as the advancement of the tongue body. The consistency of PARAFAC results across languages suggests that the derived two factors are probably universal lingual movement mechanisms that underlie vowel production. Moreover, Maeda & Honda (1994) and Honda (2000), based on the EMG and MRI data, have proposed a two-dimensional physiological model of tongue articulation, in which one axis is formed by the genioglossus posterior (GGP) and Hyoglossus (HG) agonist-antagonist pair, while the other axis is formed by another pair of genioglossus anterior (GGA) and styloglossus (SG). These two physiological axes well match up the extracted factors from the PARAFAC studies-Factor 2 and Factor 1 respectively in the present study. This suggests that the PARAFAC model of lingual articulation has physiological implications and reflexes speech motor organization for vowel articulation.

References

Harshman, R. A. Foundations of the PARAFAC procedure: Models and procedures for an 'explanatory' multi-modal factor analysis. *UCLA Working Papers in Phonetics,* Vol. 16:1-84, 1970.

Harshman, R. A., Ladefoged, P., and Goldstein, L. Factor analysis of tongue shapes. *The Journal of the Acoustical Society of America,* 62:693-707, 1977.

Honda, K. Organization of tongue articulation for vowels. *Journal of Phonetics,* 24:39-52, 2000.

Hoole, P. Issues in the acquisition, processing, reduction and parameterization of articulographic data. *Forschungsberichte des Instituts für Phonetik und Sprachliche Kommunikation München (FIPKM),* 34:158-173, 1996.

Hoole, P. On the lingual organization of the German vowel system. *The Journal of the Acoustical Society of America,* 106:1020-1032, 1999.

Hu, F. An acoustic study of the vowels in Ningbo Chinese. In *Studies on Wu Chinese,* Vol. 3:341-348, Shanghai Educational Press, 2005.

Ladefoged, P. *Three Areas of Experimental Phonetics.* Oxford University Press, 1967.

Ladefoged, P., DeClerk, J., Lindau, M., and Papçun, G. An auditory-motor theory of speech production. *UCLA Working Papers in Phonetics,* Vol. 22:48-75, 1972.

Löfqvist, A. Interarticulator phasing, locus equations, and degree of coarticulation. *The Journal of the Acoustical Society of America,* 106:2022-2030, 1999.

Löfqvist, A., Gracco, V. L., and Nye, P. W. Recording speech movements using magnetometry: One laboratory's experience. *Forschungsberichte des Instituts für Phonetik und Sprachliche Kommunikation München (FIPKM),* 31:143-162, 1993.

Maeda, S., and Honda, K. From EMG to formant patterns of vowels: the implication of vowel systems and spaces. *Phonetica,* 51:17-29, 1994.

Perkell, J., Cohen, M., Svirsky, M., Matthies, M., Garabieta, I., and Jackson, M. Electro-magnetic midsagittal articulometer (EMMA) systems for transducing speech articulatory movements. *The Journal of the Acoustical Society of America,* 92:3078-3096, 1992.

Russel, G. O. *The Vowel.* Ohio State University Press, 1928.

[λ] In *Proceedings of the 7th International Seminar on Speech Production (ISSP 2006)*, pp. 303-310, Ubatuba, Brazil, 2006.