Constructing a three-dimension physiological vowel space of the Mandarin language using electromagnetic articulography

Jian Sun [*†], Nan Yan [*†] and Lan Wang [*†]
[*] Shenzhen Institutes of Advanced Technology, Chinese Academy of Sciences, Shenzhen, China
E-mail: {sun.jian, nan.yan and lan.wang}@siat.ac.cn Tel: +86-0755-86392178
[†] The Chinese University of Hong Kong, Hong Kong, China

Abstract— **The spatial relations of vowels are traditionally depicted by using an acoustic quadrilateral. However, the accuracy of vowel chart has been controversial. In the present study, the lingual movements of different mandarin Chinese vowels were investigated using electromagnetic articulography (EMA, AG501) and the physiological equivalent of Chinese vowels was then developed and compared to the traditional acoustical vowel quadrilateral. Six native mandarin speakers repeated each one of the six vowels (/a/, /o/, /ɤ/, /i/, /u/ and /y/) by four times. The key region of each vowel, which was characterized by x, y, z tongue position at the intermediate temporal point of vowel pronunciation, was extracted. The tongue movement distance was calculated between the key region of each vowel and static tongue position. Clustering method was used to find out the centroid of the distances for each vowel. It was found that there are considerable differences between the actual lingual positions and acoustic quadrilateral's relative position depictions in the pronunciation of vowels like /u/. Results indicated that acoustic quadrilateral was insufficient to describe the lingual movement information of vowels pronunciation.**

I. INTRODUCTION

The vowel chart, or vowel quadrilateral, was developed to represent the relative tongue positions of different vowels by comparing the highest points of tongue position during each vowel's production. Acoustic vowel quadrilateral, which uses formants value to represent relative tongue position, is widely used by linguists. For example, Joos [1] proposed an acoustic vowel quadrilateral plotted by using F1 as the first dimension (high-low) and F2 as the second dimension (front-back). However, acoustics-based vowel quadrilateral does not faithfully represent the relative spatial relationship among vowels because it lacks of a one-to-one correspondence between acoustics (formants) and real tongue positions. Therefore, a research on the relationship between the physiological articulation mechanism and contradiction of acoustics and auditory is necessary.

In mandarin Chinese, many studies have been developed to construct Chinese vowel quadrilateral, which most of them focused on the acoustic quadrilateral [2, 3]. Fig.1, which was plotted and named as formant chart of basic vowels in Mandarin by Shi et al. [3], shows a portrait of acoustic quadrilateral of mandarin vowels (/a/, /ɤ/, /i/, /u/, /y/, and /ɨ/ (with two varieties)) which used F1 and F2 to imply the relative position of tongue during vowel production. Meanwhile, some studies were designed to investigate the relationship between the physiological vowel charts and perceptually based acoustic charts. For example, Tang et al. [4] estimated comparisons between physiological vowel quadrilateral and acoustic quadrilateral for Cantonese. They found that there were considerable differences between acoustically and physiologically based vowel charts in Cantonese language. Hu also investigated the correlation between physiological vowel quadrilateral and acoustic quadrilateral in Ningbo dialect, which is another southern dialect in China [5]. He also implied that there was no one-to-one correspondence between acoustics and articulation in Ningbo dialect. However, there is no comprehensive research focused on the correspondence correlation between physiological and traditional vowel chart in mandarin vowels.

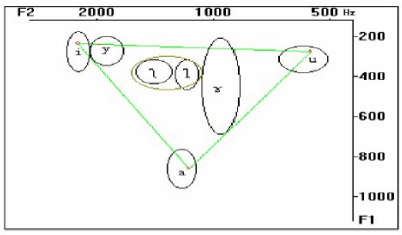


Fig.1 An acoustic quadrilateral of Mandarin vowels using F1 to show high-low dimension and F2 to show front-back dimension.

The aim of the present study is to provide an objective study on the relationship between the physiological articulation and acoustics characters in Mandarin Chinese. Using electromagnetic articulography (EMA) device might give the researchers an alternative measurement to overcome the methodological difficulties on visualizing the articulatory movements inside the oral cavity. The EMA device is a safe and non-invasive, yet accurate instrument used for physiological measurement (Kuruvilla, Murdoch, & Goozee, 2007 [6]). Many researchers have been developed to investigate lingual movements utilizing electromagnetic articulography. For example, Yunusova *et al.* [7] compared the tongue positions of different consonants and identified which consonants occupied distinctively unique locations. And Wong *et al.* [8] made comparisons of lingual kinematics

in dysarthric and nondysarthric speakers with Parkinson's disease, and normal participants using EMA.

In present study, 3D Mandarin vowels articulatory data was collected using EMA AG501 device. The lingual movements were then calculated by determining the highest tongue position during vowel production. Clustering method was used to find out the centroids of the features for each vowel. The objective physiological vowel chart was developed using the centroids of cardinal vowels. The similarities and differences between physiological vowel quadrilateral and acoustic quadrilateral were also discussed. Finally, an inter-speaker comparison was performed.

II. METHODOLOGY

A. Participants and data collection

Six adult participants (Male = 5, Female = 1) were recruited in this study. All participants are native speakers of Mandarin Chinese. They were all adult speakers with no known history of speech, language or hearing impairments. The average age of them is 25.6 and the age ranges from 23 to 29. They were able to produce the required speech materials comfortably with the EMA sensors attached to their tongue. The speech tasks included the six major mandarin vowels (/a/, /o/, / ɤ /, /i/, /u/ and /y/) produced four times in isolation at high level tone. The use of isolated vowels produced at high level tone instead of CV syllables with a non-specific tone was to avoid the possibility of coarticulation during CV syllable production (Katz & Bharadwaj, 2001 [9]) and effect of tone to vowel production (Hoole & Hu, 2004 [10]) which might confound the results.

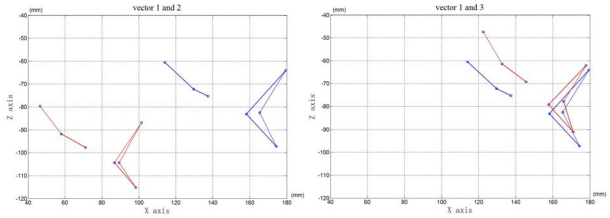
A three-dimensional electromagnetic articulography (EMA) (AG501, Carstens Medizinelektronik GmbH) was used to record the articulatory movement during vowel production. During the experiment, the participant was seated properly with his/her head positioned inside the EMA cube. All 16 available channels were used. Three of them were attached to the tongue at 1cm, 2cm, and 3cm from tongue tip respectively along the mid-sagittal plane, while 4 other sensors were put on the lips (upper lip, lower lip, left corner and right corner). Four sensors which adhere to left, right processus mastoideusand, nose and upper lip, were used as reference. Kinematic and acoustic data were recorded simultaneously. Head movements during recording were eliminated by EMA's normalization procedure automatically.

B. Data processing

To calculate the lingual movements during vowel production, the notions of 'key frame' and 'static frame' are used to mark the maximum tongue movement during vowel production and the relax condition of tongue without voice production [11, 12]. The static frame is selected from the data in relax state of tongue to define the starting point of each articulatory movement. The key frame is defined as the peak position of lingual movement during the pronunciation. For instance, the peak position of the vowel /a/ should be selected with the maximally opened mouth, while the tongue is also at its lowest point.

C. Speaker normalization

Procrustes transformation was used to overcome the speaker variability [13, 14]. It has been proved that Procrustes method could be effectively used to eliminate the speaker variability in the articulatory data [15]. The target coordinate system based on a specific speaker was selected while any other speaker's articulatory data should be transformed to it. The transform matrix was then calculated by using a 3D vector consists of 4 standard sensors' (one is on the nose, two adhered to processus mastoideus and another attached on the upper lip) XYZ information. Fig.2 shows the transformation between static frames of two speakers.



(a) Before transformation (b) After transformation

Fig.2 Transformation result on static frames of two speakers. The target coordinate was based on source vector 1 (blue dotted line), and the vector 2 (red solid line) in Fig.2 (a) was transformed to vector3 (red)in fig.2 (b).

D. Clustering

The lingual movement distance, which was monitored using 3 sensors on the tongue, was determined by calculating the difference of these sensors' positions between static frame and key frame. X, Y, and Z coordinate of tongue tip's movement distance for each vowel could be used to form a points cloud, which was referred to as a vowel target region (Yunusova *et al* [7]). Gauss kernel function clustering method was used to find the distance centroid of each vowel. 3D ellipsoid fit which covers 90% of total points was used to find out the clustered boundary for each vowel.

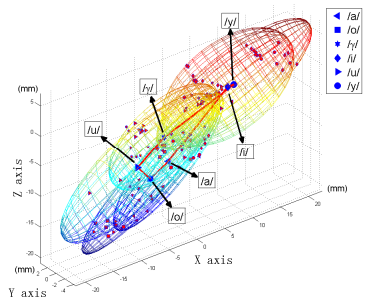


Fig.3 Ellipsoids were fit around points clouds which represent movement distances of tongue tip. The center points of these ellipsoids were determined as the centroids for tongue tip movement of mandarin vowels (/a/, /o/, /ɤ/, /i/, /u/, and /y/).

Fig.3 shows the ellipsoids fit for each of the 6 mandarin vowels, and the centroids of clustering were marked. Values on the X axis represent the front-back dimension of tongue movement and values on Z axis represent the upper-lower dimension of tongue movement. Values on the Y axis represent the left-right dimension of tongue movement.

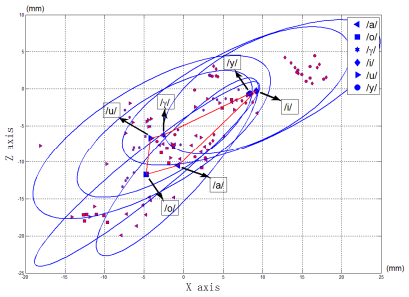


Fig.4 Ellipses were fit around points clouds representing movement distances of tongue tip on X, Z plane. The center points of these ellipses were also determined as the centroids for tongue tip movement.

Similarly, a 2D illustration of clustering could be implemented by simply projecting the 3D points cloud onto the X, Z plane. Here, SD ellipse fit was implemented to cluster the points cloud. And the value of the standard deviation was set to cover 90% of all points. Fig.4 showed the results of ellipses fit for each of the 6 mandarin vowels on X, Z plane, and centroids of ellipses were also marked as it shown the movement information for each vowel.

Based on these results, physiological vowel quadrilateral were developed using the centroids of cardinal vowels. In this vowel quadrilateral, the abscissa indicates tongue advancement and the ordinate is tongue height. The developed physiological vowel charts were then visually compared with the acoustic vowel chart of Mandarin vowels.

III. RESULTS

A. 3D view

Fig.5 showed the 3D vowel space which was developed using the clustered centroids of points clouds of tongue tip movements. Observing the distribution of points cloud (Fig.3) and their clustered centroids (Fig.5), some characters of physiological vowel chart can be found:

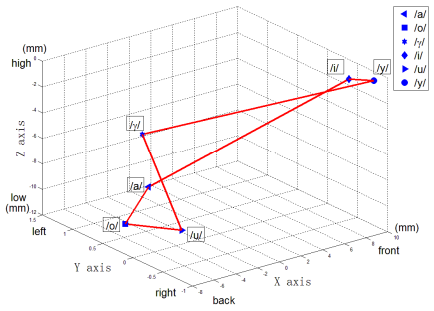


Fig.5 3D vowel space of mandarin Chinese. In this chart, X, Y, and Z axes reflect the relative front-back, left-right, high-low position respectively.

- Both /i/ and /y/ are on the far right and upper position in the 3D vowel space, which means tongue moves forward significantly while staying in the similar vertical position during sound production. Overlap between these two vowels was obviously observed from Fig.3, and the two vowels' centroids are close to each other.
- /ɤ/, /a/, /u/, and /o/ are on the left side in the 3D vowel space, and their vertical positions are lower compared with /i/ and /y/. In addition, the tongue position of /ɤ/ seems have higher attitude compare to the tongue position of vowel /a/, /u/, and /o/.
- Tongue movements in left-right dimension were minor and they have little impact on vowel distinguishment.

Based on the above results, /i/ and /y/ were categorized as group A, /ɤ/ were categorized as group B, and /a/, /u/, and /o/ were as group C. In order to illustrate the difference between relative tongue positions inside groups, 2D vowel space with X, Z plane was developed.

B. 2D view

Fig.6 shows a 2D physiological vowel quadrilateral with X, Z plane to illustrate the tongue location of each vowel's in the mid-sagittal plane. For better comparison between this physiological vowel quadrilateral and traditional acoustic physiological vowel quadrilateral (Fig.1), the X axis of the chart was reversed so that its left-right dimension shows tongue moves from front to back.

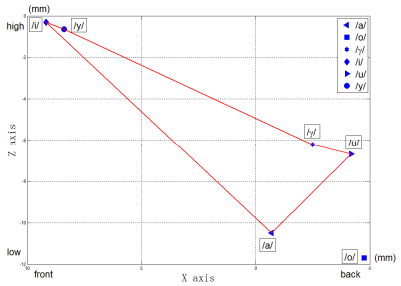


Fig.6 2D physiological vowel quadrilateral which X axis means the front-back dimension and Y axis means the low-high dimension.

According to Fig.6, differences among relative tongue positions inside groups were compared. We found that although /ɤ/, /a/, /u/, and /o/ are geographically close to each other on the 2D chart, they are still distinguishable. Tongue positions of /ɤ/ and /u/ are much higher than /a/and /o/. And /a/ does not fall back as much as /o/ does.

IV. DISCUSSION

A. Comparison between physiological vowel quadrilateral and acoustic vowel quadrilateral

The comparison between physiological vowel quadrilateral and acoustic vowel quadrilateral was based on the acoustic vowel quadrilateral we cited in Fig.1 and physiological vowel quadrilateral we plotted in Fig.6.

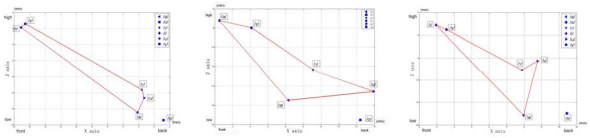
The consistent parts of two quadrilaterals are matched relative locations of /i/, /y/ and /a/. The location of /i/ and /y/ are on the front and higher position in both figures and /a/ is on the relative lower and backward position. The position of /ɤ/ on acoustic quadrilaterals is defined inside an ellipse but it still matches the relative position on physiological vowel quadrilateral.

The most significant difference between physiological vowel quadrilateral and acoustic vowel quadrilateral is that the position of /u/ is much lower in the 2D physiological

quadrilateral. It could also be observed that the forward movements of tongue for pronouncing /i/ and /y/ are actually much greater than the backward movements of /u/ and /ɤ/ while the acoustic quadrilateral shows an almost equal effect. This makes the proportions of distances between each two vowels on these two quadrilaterals unequal. It proves that there is no one-to-one correspondence between the acoustic quadrilateral and physiological vowel chart. Therefore, acoustic vowel quadrilateral is not sufficient to show the tongue's relative position information in mandarin vowels' production.

B. Inter-speaker Comparison

Because of the existence of differences in the anatomical structure and pronunciation habits of different speakers, an inter-speaker comparison was performed to illustrate how much the vowel space we constructed could be influenced by this inequity.



(a) Speaker 1 (b) Speaker 2 (c) Speaker 3

Fig.7 2D physiological vowel quadrilaterals of three speakers.

Fig.7 shows 2D physiological vowel quadrilaterals of three speakers. Speaker 1 is a male participant with a southern Chinese accent while speaker 2&3 are a male participant and a female participant with northern Chinese accents. The relative positions of vowels in Group A and vowels in Group B defined above are matched in all four figures while the relative positions of vowels in each group may vary in a limited extent, e.g. /i/ and /y/ in vertical dimension and /ɤ/ and /u/ in vertical dimension.

The comparison shows that the positions of different vowels for a single speaker may deviate from the center points as plotted in the clustered 2D view (Fig.6), and the shape of the quadrilateral may change among different speakers. Despite that, the relative position information of tongue movement during vowel production was protected to a large extent. Therefore, in the aspect of presenting relative position of vowels based on tongue movements, the physiological vowel quadrilateral is reliable.

V. CONCLUSIONS AND FUTURE WORK

In present study, a 3D physiological vowel space of mandarin Chinese was build. Based on this vowel space, a direct comparison between lingual movements and acoustic information of mandarin vowels was illustrated. The results implied that there is no one-to-one correspondence between the acoustic vowel quadrilateral and physiological vowel chart in mandarin Chinese. The acoustic vowel quadrilateral is insufficient to show the tongue's relative position information in mandarin vowels' production. The result of inter-speaker comparison supported that the relative physiological vowel quadrilateral was reliable. It should be noted that the methodology we proposed above may become a standard methodology in investigating the tongue movement during pronunciation, and it also can be utilized on the research of consonant productions.

Because the present study is a pilot study, there are some limitations in this research. First of all, the number of participants should be enlarged to enhance the effectiveness of our finding. Secondly, more detailed comparisons among speakers with different backgrounds and different genders should also be investigated. Statistical analysis could also be utilized to show the overlap among neighboring vowels.

ACKNOWLEDGMENT

This work is supported by National Natural Science Foundation of China (NSFC 61135003, NSFC 90920002), and Guangdong Innovative Research Team Program No. 201001D0104648280.

REFERENCES

[1] Joos, M. (1948). Acoustic phonetics. Language, 24(2), 5-136.
[2] Shi, F. (2002). The vowel pattern of Beijing Mandarin. Nankai Linguistics.
[3] Shi, F., Ran, Q., Wang, P., Wen, B., & Liang, L. (2011). Exploration on Acoustic Sound Pattern. International Congress of Phonetic Sciences, 1810-1813.
[4] Tang, C., Ng, M., Chen, Y., & Yan, N. (2012). Constructing a physiological equivalent of the Cantonese vowel quadrilateral using electromagnetic articulography (EMA). Asia Pacific Journal of Speech, Language and Hearing, 15(3), 163-173.
[5] Hu, F. (2003). An acoustic and articulatory analysis of vowels in Ningbo Chinese. In Proceedings of the 15th International Congress on Phonetic Sciences, Barcelona, 3017-3020.
[6] Kuruvilla, M., Murdoch, B., & Goozee, J. (2007). Electromagnetic articulography assessment of articulatory function in adults with dysarthria following traumatic brain injury. Brain Injury, 21(6), 601-613.
[7] Yunusova, Y., Rosenthal, J. S., Rudy, K., Baljko, M., & Daskalogiannakis, J. (2012). Positional targets for lingual consonants defined using electromagnetic articulography. The Journal of the Acoustical Society of America, 132, 1027.
[8] Wong, M. N., Murdoch, B. E., & Whelan, B. M. (2011). Lingual kinematics in dysarthric and nondysarthric speakers with Parkinson's disease. Parkinson's Disease, 2011, 1-8.
[9] Katz, W. F., & Bharadwaj, S. (2001). Coarticulation in fricative-vowel syllables produced by children and adults: a preliminary report. Clinical linguistics & phonetics, 15(1-2), 139-143.
[10] Hoole, Philip / Hu, Fang (2004): "Tone-vowel interaction in standard Chinese", In TAL-2004, 89-92.
[11] Wang, L., Chen, H., Li, S., & Meng, H. M. (2012). Phoneme-level articulatory animation in pronunciation training. Speech Communication, 54(7), 845-856.
[12] Chen, H., Wang, L., Liu, W., & Heng, P. A. (2010). Combined X-ray and facial videos for phoneme-level articulator dynamics. The Visual Computer, 26(6), 477-486.
[13] Sheng, L. & Lan, W. (2012). Cross Linguistic Comparison of Mandarin and English EMA Articulatory Data. INTERSPEECH-2012, 903-906.
[14] Gower, J. C. (1975). Generalized procrustes analysis. Psychometrika, 40(1), 33-51.
[15] Geng, C., & Mooshammer, C. (2009). How to stretch and shrink vowel systems: Results from a vowel normalization procedure. The Journal of the Acoustical Society of America, 125, 3278.