Lip rounding and vowel formant frequencies in Nantong Chinese

Benjamin Ao

Department of Linguistics, The Ohio State University

Abstract

A study of the vowel system of Nantong Chinese, which has as many as seven high vowels, suggests that instead of the traditional two formant model, a three formant model is needed to approximate the Nantong vowel space. Of the three formants used in the model, F3 appears to be the major acoustic cue of lip rounding. It also appears that the "maximum dispersion" hypothesis is not true with regard to the Nantong vowel system.

Introduction

In the Nantong dialect of Chinese spoken on the northern Yangtze Delta, there are 13 oral vowels, as follows:

i y ʒ ʒ̹ ƶ β u

e ɜ o

ɛ ɔ

a

Of the seven high vowels, four sound like syllablic fricatives represented by the IPA symbols used: [ʒ] is a syllabic [ʒ], [ʒ̹] a syllabic [ʒ̹] (the rounded version of [ʒ]), [ƶ] a syllabic [z], and [β] a syllabic [β]. Ignoring [ƶ], the six other high vowels can be divided into three pairs (front, central and back), each contrasting in rounding.

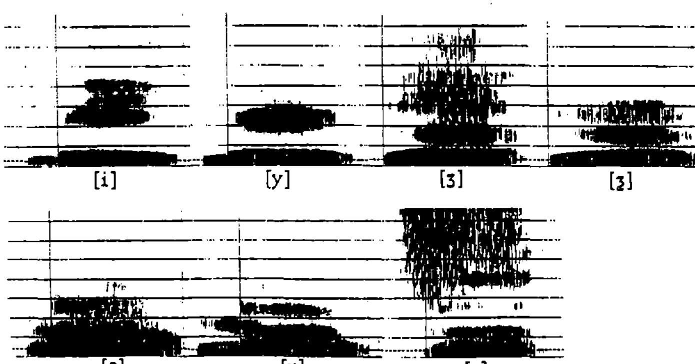

Figure 1. Spectrograms of the seven Nantong high vowels.

A number of questions can be raised about the Nantong vowel system. First of all, is it possible to approximate the crowded Nantong vowel space, especially that of the high vowels, with the traditional two formant model? Secondly, what is the acoustic effect of lip rounding, particularly on the high vowels? Finally, does the "maximum dispersion" hypothesis a la Liljencrants & Lindblom (1972) hold true to the Nantong vowel system, especially in the high vowel area? To answer these questions, I conducted an experiment, which will be explained below.

Experiment

Method

Subjects

Three native speakers of Nantong Chinese, two males (AX and ST) and one female (LP), all in their 20's or 30's, were asked to read out the syllable list.

Material

A list of all the (131) phonotactically possible CV syllables with all the 13 vowels was used as the experiment material. The sequence of syllables was randomized.

Procedure

Video and audio signals of the subjects reading the syllable list were simultaneously recorded with a camcorder placed two feet in front of the mouth of the speaker. A large mirror was placed beside the speaker's head at a 45 degree angle to allow the sideview of the lip movement to be also recorded. A cushion was placed between the speaker's head and the wall behind to prevent any head movement during the recording. The audio recording was fed into a Kay 5500 speech analyzer, and formant frequency values were taken from the average power spectrum across the entire vowel duration.

To quantify lip rounding movements, 11 measurements between geometrically defined points (CF, CE, CD, BF, BE, GJ, GI, PS, LM, LN and LO, see Fig 2) were taken from the playback of the video recording on a 19" TV screen.

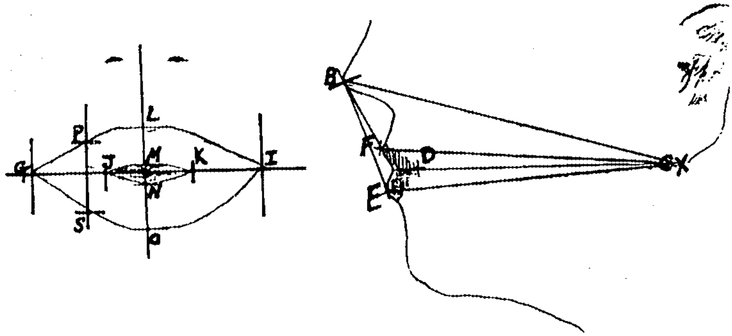

Figure 2. Reference points for lip gesture measurements

Analysis

The traditional two formant (F1 by F2) model for the vowel space proves insufficient for the Nantong vowel system. Overlap between different vowels exists for all three speakers. Figure 3 shows this.

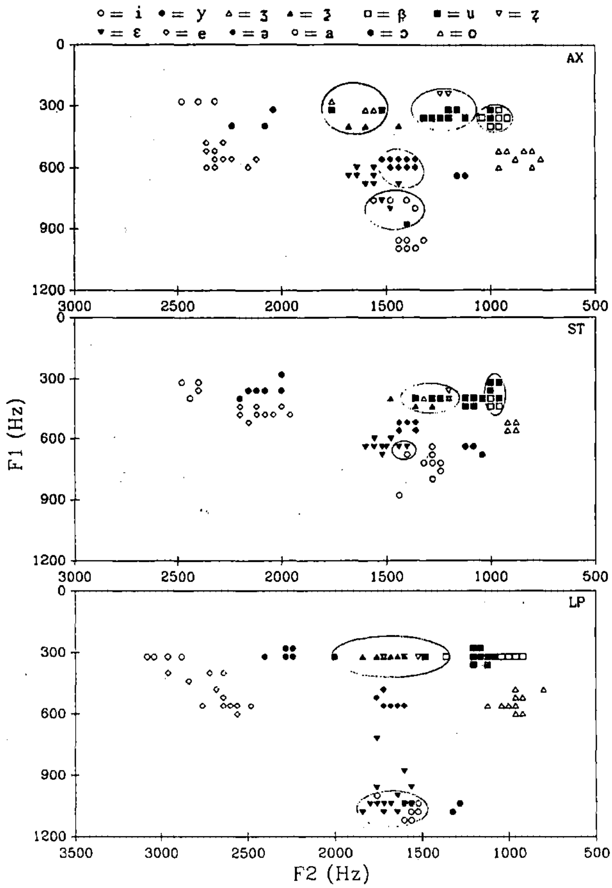

Figure 3. Nantong vowel space in F1 x F2 for three speakers; major areas of overlapping are circled.

Little improvement was achieved by substituting F2 with the weighed average F2' calculated according to the formula proposed by Bladon (1983) as in:
$F2' = F2 + C^2 \times (F3 \times F4)^{1/2} / (1+ C^2)$, where $C = [12 \times F2 \times 67 \times F2(1 - F1^2 / F2^2) \times (1 - F2^2 / F3^2) \times (1 - F2^2 / F4^2)] \times [1400 \times (F4 - F3)^2 \times (F3 \times F4 / F2^2 - 1)]^{-1}$. Figure 4 shows Nantong vowel space.

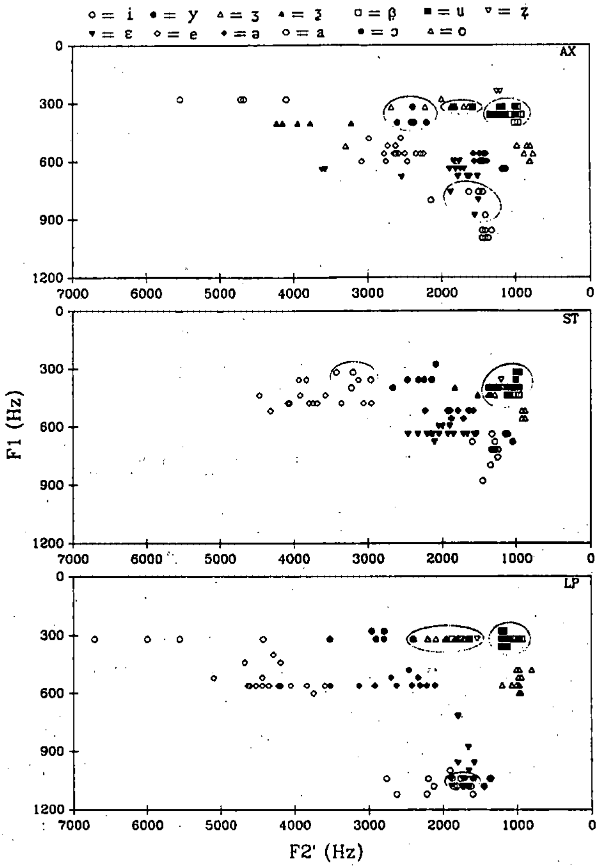

Figure 4. Nantong vowel space (F1 by F2')

Plotting F1 by F3 separates rounded and unrounded vowels, but overlaps vowels with different degrees of tongue advancement or retraction, as is seen in Figure 5.

The best separation of the seven high vowels is obtained by plotting F2 against F3, as shown in Figure 6, where like vowels are encircled. This suggests that a three dimensional model with F1, F2 and F3 as coordinates is necessary for approximating the entire Nantong vowel space, with F1 for vowel height and F2 for tongue advancement.

In order to justify adopting a three formant model instead of the traditional two formant model of vowel space, with F3 as the additional dimension, efforts were made to find out the underlying articulatory parameter responsible for the variation of F3. An ANCOVA test reveals that the single most significant ($p < .01$)

articulatory measurement that covaries with F3 is the horizontal width of horizontal lip opening (JK). Figure 7 shows the mean JK value of the seven high vowels for all three speakers.

Figure 8 shows how JK is correlated with F3. One may notice that there is a big difference between the F3 values of the two members of the rounded and unrounded pairs of front vowels and central vowels, but the difference between the F3 values of the rounded and unrounded back vowels is minimal with speakers ST and LP. In addition to the width of horizontal lip opening, speaker identity has a significant effect on F3 as well ($p < .01$), and lip protrusion (BF, CF) is also a significant covariate (BF and CF, $p < .05$).

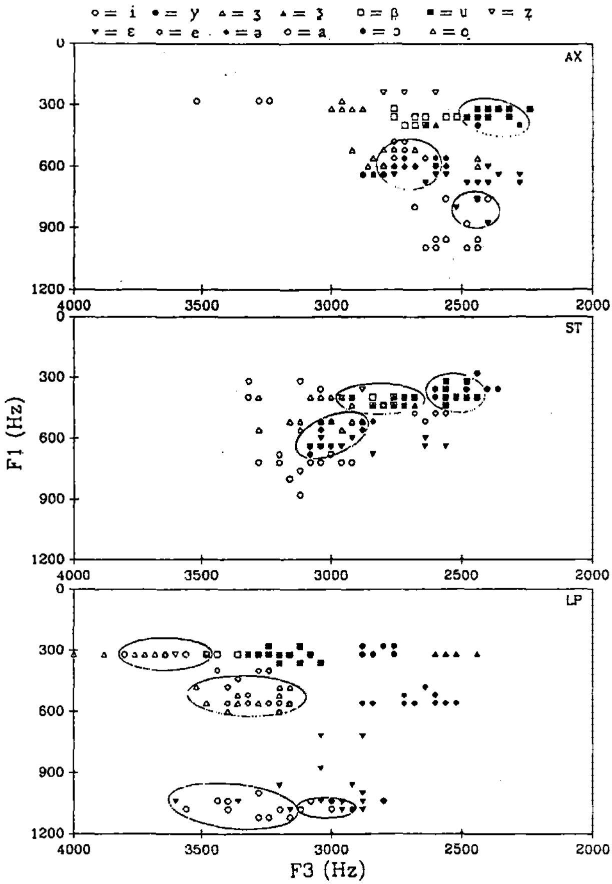

Figure 5. Nantong vowel space (F1 by F3)

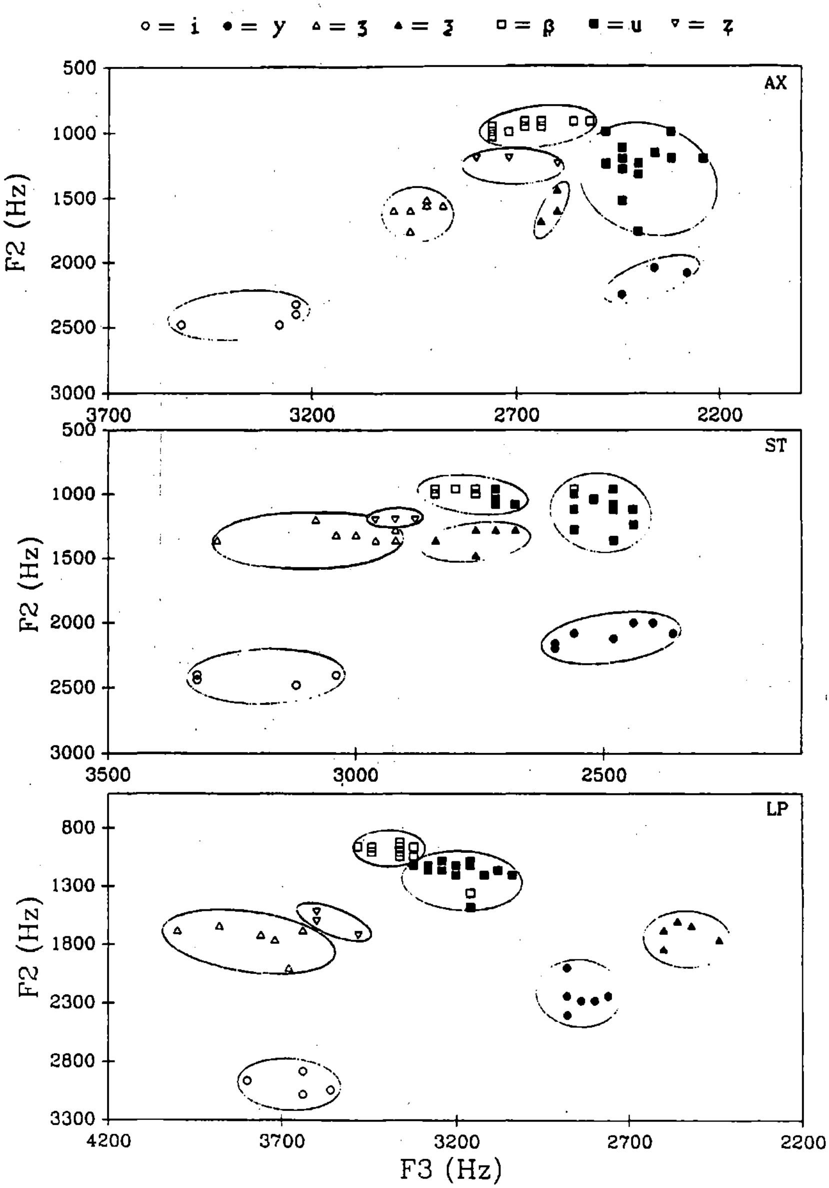

Figure 6. Nantong vowel space for high vowels (F2 by F3)

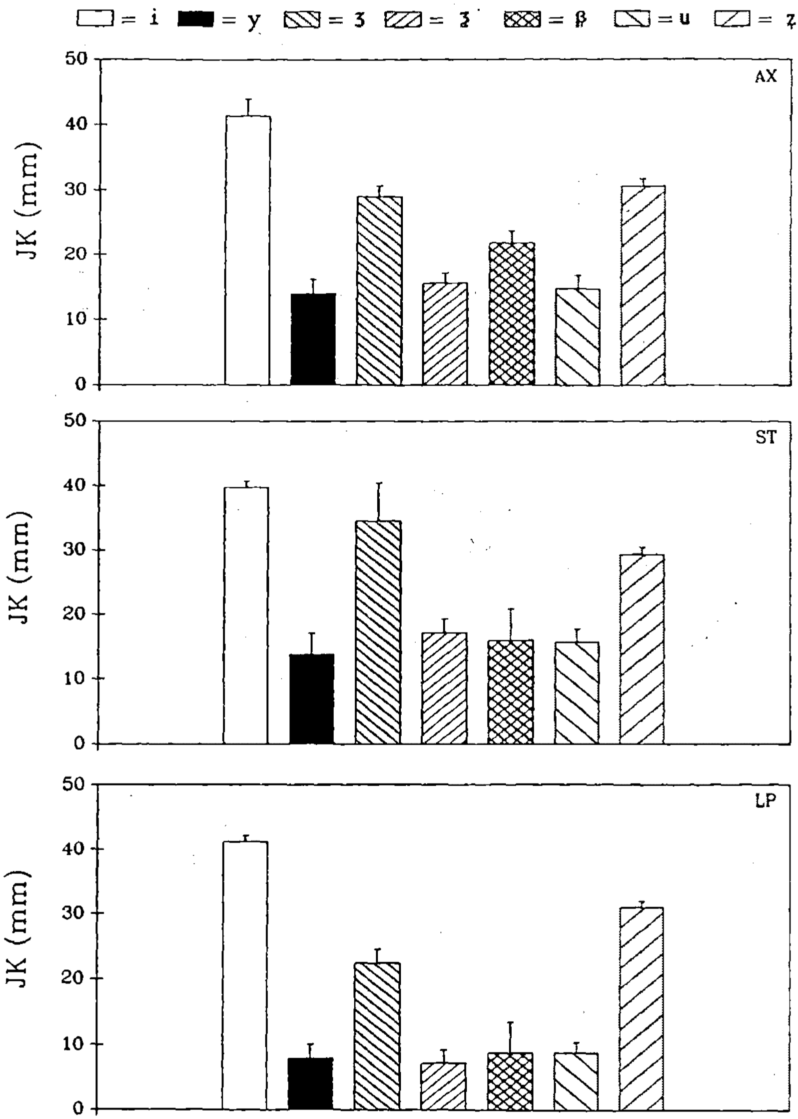

Figure 7. Width of lip opening of the Nantong high vowels

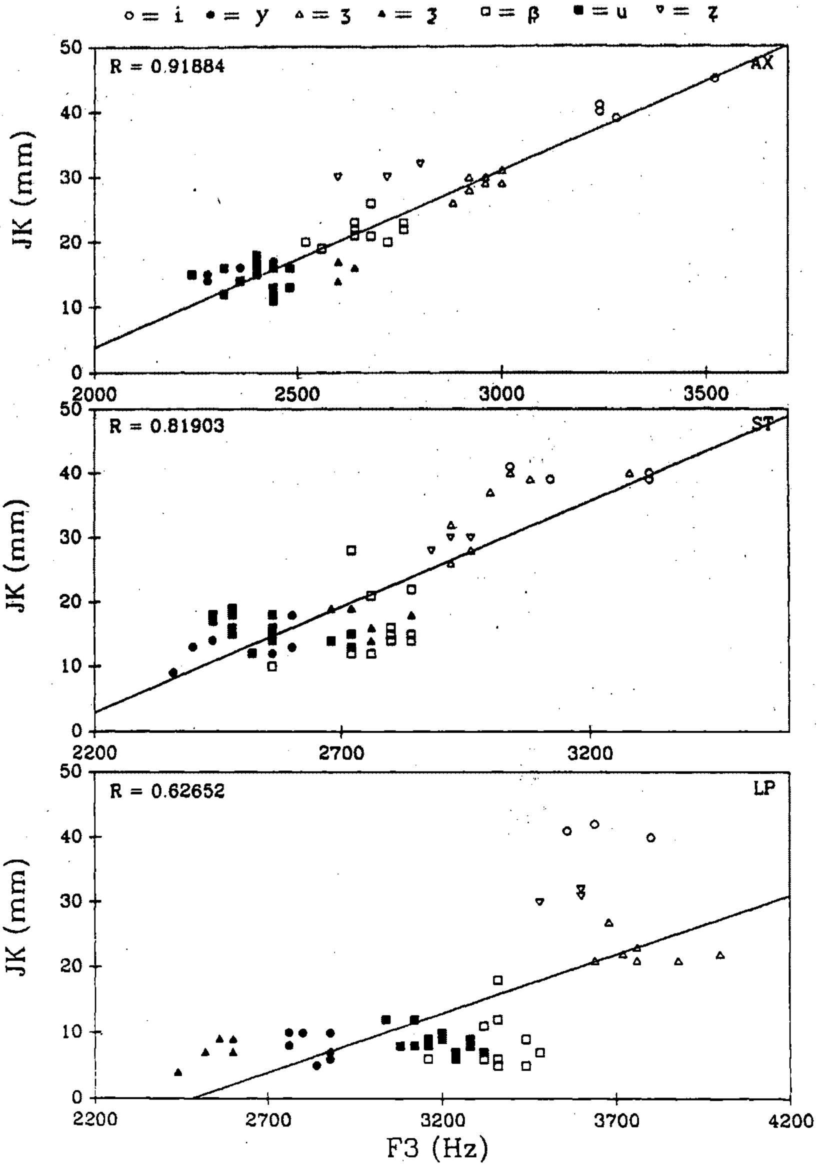

Figure 8. Correlation between F3 and width of lip opening

Discussion

Now we may return to the questions raised at the beginning of this paper. First of all, it is clear by now that the traditional two formant model of vowel space is not universally adequate, for it fails to represent the crowded Nantong vowel space. Instead, a three formant model, with F3 as the third dimension, is needed.

Secondly, it also appears that while tongue height is responsible for F1 and tongue advancement for F2, lip rounding accounts for much of the variation of F3, especially with front and central vowels.

Finally, the massive overlapping of plotted vowel formant frequencies that coexists with vacant areas in a two dimensional model of the Nantong vowel space is contradictory to the maximum dispersion hypothesis. Even in a three dimensional model where overlapping of plotted vowel formant frequencies is by and large eliminated, the distances between the centers of plotted formant frequencies of various vowel are largely unequal. This suggests that the maximum dispersion hypothesis is not universally true.

Acknowledgement

My heartfelt thanks go to Michel Jackson of the Division of Speech and Hearing Science at the Ohio State University. Without his generous help, I could not have completed this project. I am also grateful to the Division of Speech and Hearing Science for allowing me to use their facilities for part of my measurement and computation.

*The original version of this paper was presented in the 119th meeting of the Acoustical Society of America, State College, Pennsylvania, 21-25 May 1990.

References

Bladon, Anthony (1983) "Two-formant model of vowel perception: shortcomings and enhancements". *Speech Communication* 2: 305-13.

Liljencrants, Johan and Bjørn Lindblom (1972) "Numerical solution of vowel quality systems: the role of perceptual contrast". *Language* 48.4: 839-62.