EFFECT OF CONSONANT AND VOWEL CONTEXT ON MANDARIN CHINESE VOT: PRODUCTION AND PERCEPTION

Bernard L. Rochet and Yanmei Fei
Department of Romance Languages, University of Alberta

1.0 Introduction:
The following is a preliminary report on a study of voice onset time (VOT) in Mandarin Chinese. The issues to be focused on here are the effects of consonantal place of articulation and vowel quality on VOT duration (at the production level) and cross-over boundary values (at the perceptual level).

2.0 Subjects:
Ten native speakers of Mandarin Chinese were recruited from various departments at the University of Alberta. All of them completed the recordings which yielded the production data, and eight of them completed the perceptual tests.

3.0 Materials:
The production data were derived from the subjects' readings of a randomized word-list and were recorded on a good quality cassette recorder (Sony TC-D5 PROII), via a Beyer 111 dynamic microphone in a soundproof room. The words recorded consisted of an initial occlusive (/p, t, k, b, d, g/) and a high vowel (/i/ or /u/) or a low vowel (/a/).

The stimuli used for the perceptual tests were synthesized in cascade at a 10 kHz sampling rate, using Klatt's cascade/parallel speech synthesizer (Klatt 1980), implemented on an IBM AT microcomputer, with Canadian Speech Research Environment software. Several continua were used, 3 each for labial and dental occlusives (for the vowels /i/, /a/, and /u/); and 2 for the velar occlusives (for the vowels /u/ and /a/ only, because velar occlusives do not occur before the vowel /i/ in Mandarin Chinese). In each continuum, VOT values were incremented in 100 ms steps from -60 ms (lead VOT) to +130 ms (lag VOT).

4.0 Measurements:
The signals recorded for each subject were digitized at a sampling rate of 22 kHz, and temporal measurements of VOT were obtained from these digitized units by means of waveform analysis sofware developed at the University of Alberta. The results were analyzed by means of repeated measures ANOVAs.

5.0 Results:
5.1.0 <u>Production</u>
5.1.1 The voiceless aspirated occlusives
Mean VOT durations for the voiceless aspirated occlusives /p/, /t/ and /k/ are shown in Figure 1 (for the vowels /a/ and /u/). Significant main effects were found for both PLACE ($F[2, 198] = 18.44$, $p<.01$) and VOWEL ($F[1, 94] = 19.341$, $p<.01$), with no significant two-way interaction. Tukey HSD tests confirmed that VOT for /k/ (at 110.3 ms) was significantly greater than for /t/ (98.7 ms, $p<.01$) and for /p/ (99.6 ms, $p<.01$). On the other hand, the VOT values for /p/ and /t/ were not significantly different. When the following vowel was /u/, the mean VOT duration was 106.7 ms, compared with 99.1 ms when the following vowel was /a/.

Figure 2 shows the effects of the three vowels /i/, /a/, and /u/ on the preceding initial occlusives /p/ and /t/ (in this case, the consonant /k/ is not represented because it does not occur before the vowel /i/). The nature of the VOWEL had a significant effect on the VOT values of the preceding consonants ($F[2, 196] = 6.248$, $p<.01$), while the PLACE of articulation of the consonant was not significant. The two-way interaction was significant ($F[2, 196] = 3.343$, $p<.05$). Tukey HSD tests for main vowel effects indicated significantly longer VOT values (105.6 ms) when the following vowel was /i/ than when it was /a/ (96.1 ms; $p<.01$), while no other comparisons reached significance (with /u/ at 102.6 ms).

5.1.2 The voiceless unaspirated occlusives
Very similar patterns were observed for the occlusives /b, d, g/, for which the mean VOT values are summarized in Figures 3 and 4. For both series (/b, d, g/ and /p, t, k/), the velar is characterized by a longer VOT than the dental and the labial. As far as vowel effect is concerned, occlusives followed by the low vowel /a/ are always accompanied by a shorter VOT than those followed by the high vowels /u/ and /i/. Both series also show an interaction between place of articulation and identity of the vowel: specifically, the longest voicing lags occur before /u/ for the labials, and before /i/ for the dentals.

5.2.0 <u>Perception</u>
Mean cross-over boundary values for the labial, dental, and velar occlusive series (/p-b/, /t-d/ and /k-g/) are shown in Figure 5 (for the vowels /a/ and /u/). Significant main effects were found for both PLACE ($F[2, 14] = 25.865$, $p<.001$) and VOWEL ($F[1, 7 = 116.227$, $p<.001$), with a significant two-way interaction ($F[2, 14] = 13.598$, $p<.001$). Tukey tests confirmed that the cross-over boundary for the velars (at 37.2 ms) was significantly higher than that for the dentals (25.3 ms, $p<.01$). The boundary for the labials (33.9 ms) was also significantly higher than that of the dentals ($p<.01$), but there was no significant difference between the boundaries of the velars and labials. For the vowels, a following /u/ was associated with a significantly higher cross-over boundary (39.1 ms) than a following /a/ (25.2 ms).

Mean cross-over boundary values for the labial and dental series are shown in Figure 6 (for the vowels /i/, /a/, and /u/). The nature of the VOWEL had a significant effect on the cross-over boundary values of the preceding consonant ($F[2, 14] = 50.234$, $p<.001$), while the PLACE of articulation of the consonant was not significant. The two-way interaction was significant ($F[2, 14] = 19.94$, $p<.001$). Tukey tests for main vowel effects indicated that /i/ was associated with a higher cross-over boundary (41.3 ms) than /u/ (35.2 ms), which in turn was characterized by a higher boundary than /a/ (24.0 ms; $p<.01$).

A comparison of Figures 1, 3 and 5, and 2, 4 and 6 respectively suggests that the interaction between consonant place of articulation and vowel quality observed for the production data is also at work in perception.

6.0 Discussion:
The present results are in keeping with trends reported for VOT durations in other languages. In particular, the occurrence of longer voicing lags with velars observed here is a common phenomenon. The tendency for high vowels to be associated with longer VOT values and the interaction between consonant place and vowel quality have also been documented by Rochet et

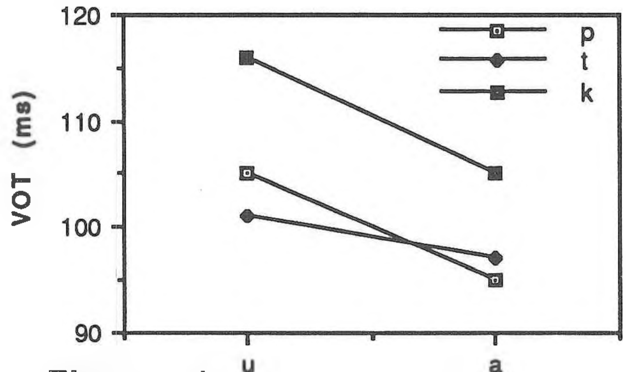

Figure 1.

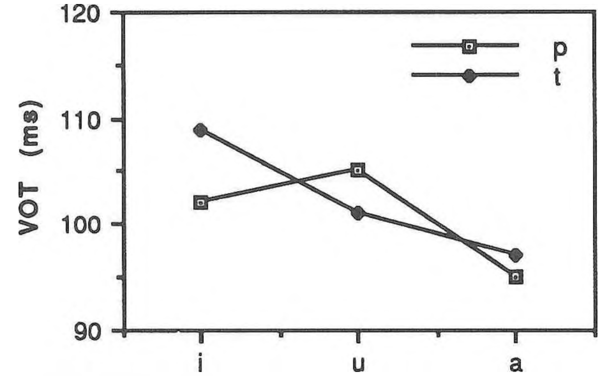

Figure 2.

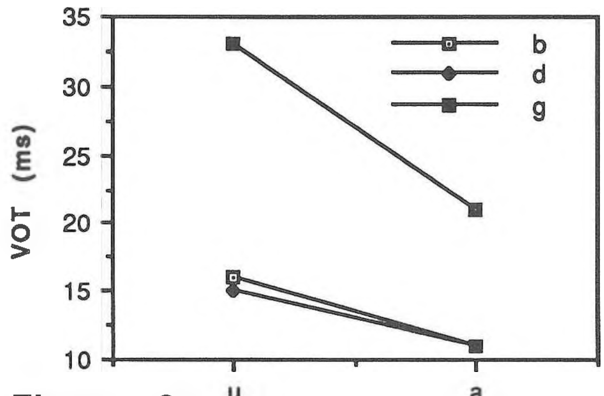

Figure 3.

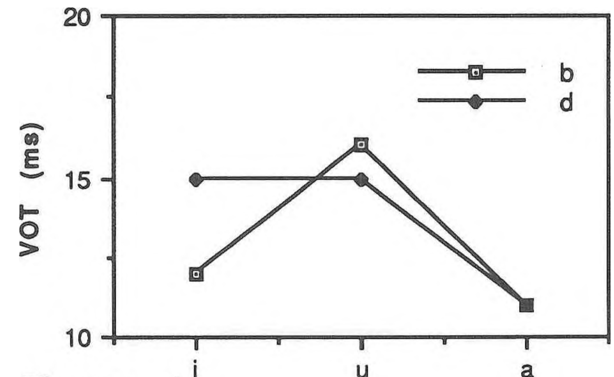

Figure 4.

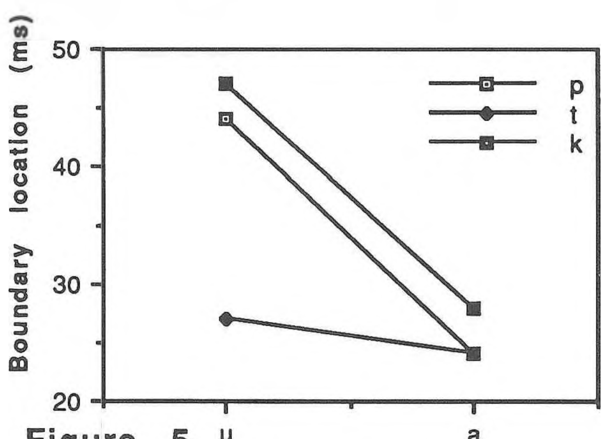

Figure 5.

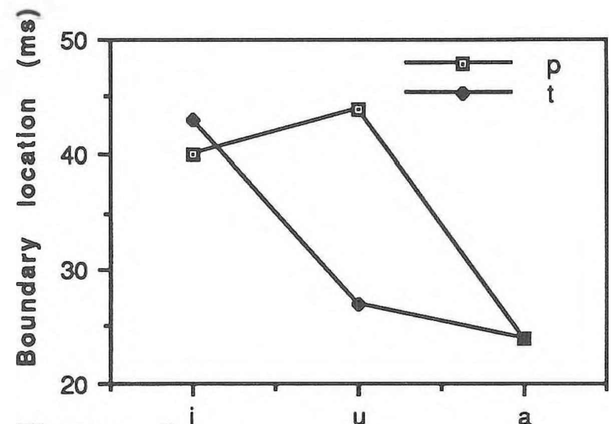

Figure 6.

al. (1987) and Fischer-Jørgensen (1972), respectively.

The parallelism noted in the present study between the patterns of variation of VOT duration (in production) and cross-over boundary location (perception) has rarely been observed (but see Summerfield 1975). Further research is needed to establish in what languages and under what conditions VOT characteristics are mirrored or ignored in perception.

7.0 References:

Fischer-Jørgensen, E. (1972). "'ptk' et 'bdg' français en position intervocalique accentuée." In A. Valdman, ed., Papers in Linguistics and Phonetics to the Memory of Pierre Delattre. The Hague: Mouton. Pp. 143-200.

Klatt, D. (1980). "Software for a cascade/parallel formant synthesizer." JASA 67: 971-995.

Rochet, B., T. Nearey, & M. Munro (1987). "Effects of voicing, place and vowel context on VOT for French and English stops." JASA 81: S65.

Summerfield, Q. (1975). "How a full account of segmental perception depends on prosody." In A. Cohen et al., eds., Structure and Process in Speech Perception. New York: Springer-Verlag. Pp. 51-68.