

The variations of Mandarin unaspirated stops in spontaneous speech

Chen Xiaoxia

Chinese Academy of Social Sciences, P.R.C

ABSTRACT

The variations of unaspirated stops /p/, /t/, /k/ were investigated in spontaneous speech with the aim of (1) describing the variations and (2) interpreting the variation for acoustic data in terms of articulation and environment.

In this report, 56 sentences were selected through the read speech corpus. The variations of unaspirated stops were analyzed and perception experiment was done. The primary analysis was done for voiced unaspirated stops. With the analysis, the classes of acoustic variations were explored.

INTRODUCTION

It is well known there is no phonological contrast between voiceless and voiced stops in Mandarin. The unaspirated voiceless stops are thought to be the weak stops in Mandarin [1]. But the voiceless stop often becomes voiced stop or other phoneme in spontaneous speech. This phenomenon may be usually treated as glottal spreading [2]. The variations have different types with different environment. What causes the variation and which context environment affects its types? Based on a spontaneous speech corpus CASS [3], this study investigates the relationship between the variation and environment and finds out the variation types. Also this study will explore the relationship between this variation and prosodic word. The primary experiment shows prosody strongly leads to the variation. Eventually the variation appears in prosodic word final or prosodic phrase final. Context environment is the main factor of leading to this variation. It is not easy to hear this variation in an utterance. Because of the effect of sentence meaning, listeners often

understand the utterance meaning but not consider how to produce every phoneme.

METHOD

1 Materials: 56 sentences were selected from CASS read speech corpus read by 2 male speakers. Recording materials were read with text in recording room. These materials include three stops /p/, /t/, and /k/.

2 Analysis: Praat software was used to do acoustic analysis.

3. Perception: Each word was cut from the utterance and repeated three times. 10 persons do the perception experiment. The experiment asked listeners to judge what syllables in one word and permitted them not to write unclear syllables [4].

MEASUREMENTS

Measurements include the occlusion duration from the end of the pre-syllable to the burst, the duration between burst and the following vowel, and the following vowel's duration. All the measurements were done with Praat software.

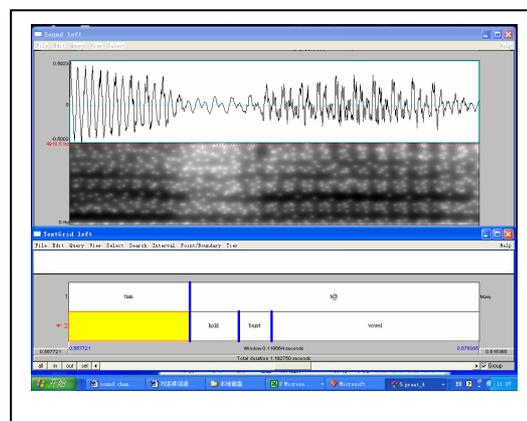


Fig.1 the different duration parts in voiced stop environment.

RESULTS

3.1 Analysis

The first measurement is the duration before burst (occlusion), burst duration and the duration of the following vowel. In this measurement, there is mostly no interval between burst and the following vowel. For three stops /p/, /t/, /k/, the durations of three parts were got as Fig.1 showed. Here is /t/'s measurement result. Fig.2 gives three parts durations. The lowest part is burst duration. It is below 20ms. The middle part is occlusion duration. Except for 8.5% bigger than 40ms, the others are 20-40ms. Here are the average durations for each part.

Three parts	Duration (ms)
Hold	0.027 (sd=0.01)
Burst	0.013 (sd=0.003)
Vowel	0.156 (sd=0.03)

Table 1 The average duration of three parts

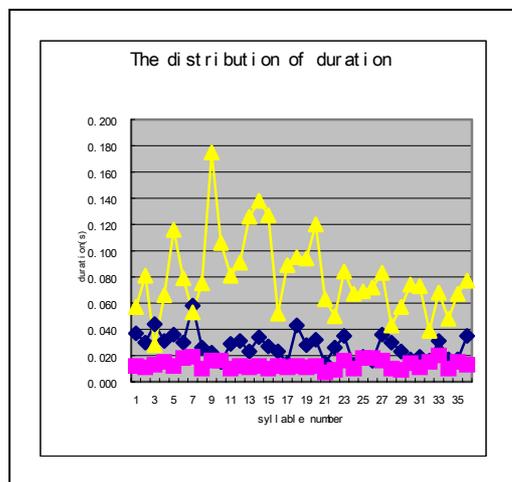


Fig.2 The distribution of duration, the x axis is syllable number. The y axis is duration. For each syllable, there are three parts.

3.2 H1, H2 and A1

It is thought H1 and H2 are very important for voiced/voiceless. The first formant amplitude is also important. The

current method is to compare the difference between H1 and H2. This research gives the measurement data. But the differences between H1 and H2, H1 and A1 are not given. Fig.3 gives the distribution of H1, H2 and A1. It shows the relative data for each sound. X axis means vowel segment. Y axis means the amplitude. The measurement is made from the vowel beginning with a 25 ms window.

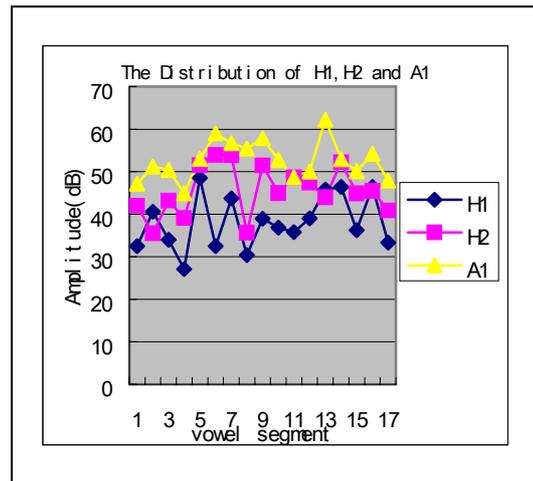


Fig.3 The distribution of H1, H2 and A1

The first harmonic amplitude H1, the second harmonic amplitude H2 and the first formant amplitude were measured. Fig. 3 gives the result for alveolar stop /t/ following schwa vowel when it is changed into voiced for one male speaker. It seems that H1 is the lowest, A1 is the highest and H2 in the middle.

3.3 The position of stop in a prosodic word

According to the position of the selected syllable in a sentence, there are two conditions for each syllable. The first is at the beginning of the word, the second is at the end of the word. Mostly, the boundaries of a word or a prosodic word are similar except for some cases. This part gives the result for the two cases. The result showed the voiceless consonant of the latter syllable in a word often changes into voiced consonant [6]. But at the

beginning position, it is seldom changed in the materials.

The position of stop in a word	Percentage
The beginning of word	5%
The end of word	95%

Table 2 the position of stop in a word (m02, a male speaker)

Table 2 gives the position of stop in a prosodic word. It is well known that Chinese has syllable structure such as CV, CVN. Generally, it is an isolated meaning unit. It can be a word or a morpheme. Two syllables or more usually combine a real word or a prosodic word. The structures are CVCV, CVNCV or CVNCVN etc.. The position of a stop in a prosodic word means it is the first consonant or the second consonant.

3.4 Perception

The experiment gives three parts as the test materials. The first is in a sentence tier. The second is in a word tier. The third is in a syllable tier. The correct rate is the highest in a sentence tier and then is in a word tier. It is the most difficult to listen and identify in a syllable tier. Mostly the stop can be justified correctly in a sentence or a word tiers. But it is not easy to listen in syllable tier. The experiment showed that stops in some cases were listened as the same place nasal or lateral. Some can be listened as zero initial syllables. Table 3 gives the listening variation results as the following:

/p/(per)%		/t/(per)%		/k/(per)%	
/p/	60	/t/	76	/k/	65
/m/	20	/n/	7.5		
/w/	6	/l/	12	/w/	20
*	14	*	4.5	*	15

Table3 the perception result of three stops (* means unclear sound)

CONCLUSION AND DISCUSSION

There are many research literatures for voiceless and voiced [5] [6] [7] [8] [9] [10] [11] [12] [13]. It is an old question. There is not the voiceless and voiced contrast in Mandarin. The aim of this research is to explore the variation of stop in spontaneous speech. The results showed the variation of stop in spontaneous speech is evident.

4.1 Acoustic analyses

Acoustic analysis shows there is continuous wave during stop occlusion. The burst keeps constant duration. In many cases, there is not evident formant at high frequency. The duration results for three parts show the occlusion of stop is constant, also the burst duration. The syllable structure in Mandarin is complex. Because of the variable in final, the vowel durations are different in the measurement data. It is well studied that there is complementation duration between consonant and vowel in a syllable. Tone can affect vowel's duration in Mandarin. But for the duration of occlusion and burst, the durations are contrary constant.

The first harmonic amplitude H1 is the lowest. The second harmonic amplitude H2 is greater than H1. The first formant amplitude A1 is the greatest. The vowels include high and low vowels. This result is similar to the Michael Jessen's male speaker result [5].

Because of the tonal effect, F0 in Mandarin changes with different tone types. This research measures F0. The results show it is not easy to compare the voiceless and voiced. This will be considered later.

4.2 Perception

The acoustic analysis shows there are different variations for stops. But the perception results show that persons can judge the variations less than the acoustic analysis. It is due to the lack of unvoiced distinction in

Mandarin. Also the listeners cannot compare the variation. This problem will be further done next.

4.3 The types of stops

According to the acoustic analysis and perception results, there are different variation types for the stops. The first is the simple voiced. This type has the continuous wave from pre-vowel to the following consonant stop. There are no other formants at high frequency domain. It usually is justified as voiceless stop. The second is the sound change. Except for the vocal vibration at low frequency, there are evident formants at high frequency. The third is unclear semi-vowel. This is due to the weak of the burst amplitude or the lack of the burst. Listeners cannot judge correctly.

4.4 The explanation of the variation

There are many literatures about the problem. There is two reasons to induce the variation. The first is the prosodic structure. The variation occurs in a word or a prosodic word. The relationship of phonemes in one word is closer than that between two words. The second is the context environment effect. It is usually because of the different context, the stops usually become the same articulation place nasal or lateral. If the pre-stop sound is nasal, the stop usually becomes labial nasal for /p/ and alveolar nasal or lateral for /t/. Because of the anticipation coarticulation in Mandarin, the place of stop can affect the preceding nasal, /p/ can affect /n/ as /m/, and /k/ can affect /n/ as glottal nasal. Why the stop can change in spontaneous speech? There is not contrary between voiceless and voiced stops in Mandarin. So it is free for the stop pronunciation. Coarticulation plays function in context environment. According to the previous studied for Mandarin, anticipatory coarticulation is the main manner in segmental

context. It leads to stop becoming different sounds.

ACKNOWLEDGEMENT

This research is supported by the national young social science foundation of China.

REFERENCES

- [1] Luo, Changpei & Wang, Jun (1956), the of general phonetics, Sciences Press, Beijing
- [2] Kenneth N.Stevens (2000), diverse acoustic cues at consonantal landmarks, vol.57, 139-151, phonetica
- [3] Li, Aijun et al, (2000) CASS: A phonetically transcribed corpus of Mandarin spontaneous speech, Report of Phonetic Research, Institute of Linguistics, Chinese Academy of Social Sciences, Beijing
- [4]David R.Williams (1995), Perception of oral release rate for initial voiced stops, ICPhS 1995
- [5] Michael Hessen & Justus C. Roux (2002), Voice quality differences associated with stops and clicks in Xhosa, Vol. 30, 1-52, Journal of Phonetics
- [6]Patricia.Keating (1995), Effects of prosodic position on /t., d/ tongue/palate contact, ICPhS, 1995
- [7]F. H.Buder, K. Williams etc. (1995), Characterizing the adult target: acoustic studies of Swedish and American English /t/ and /d/, ICPhS, 1995
- [8]David House (1995), the influence of silence on perceiving the preceding tonal contour, ICPhS, 1995
- [9]Wightman, C. W., & Shattuck-Hufnagel, S. etc. (1992), Segmental durations in the vicinity of prosodic phrase boundaries, vol. 91, J.A.S.A
- [10]Blumstein, S. E. & Stevens, K. N. (1979) Acoustic invariance in speech production: Evidence from measurements of the spectral characteristics of stop consonants. J. A. S. A. vol. 66
- [11] Dejong, K. (1995), The supraglottal articulation prominence in English: linguistics stress as localized hyperarticulation, J. A. S. A., 97(1)
- [12] Van Bergem, D. (1994), a model of coarticulatory effects on the schwa, Speech Communication, vol. 14
- [13]K. J. Kohler, etc. (1995), from scenario to segment-the controlled elicitation, transcription, segmentation and labeling of spontaneous speech, University Kiel