ISCA Archive
http://www.isca-speech.org/archive

International Symposium on Chinese Spoken Language Processing (ISCSLP 2000)
Fragrant Hill Hotel, Beijing
October 13-15, 2000

# COARTICULATION AND APPLICATION OF LATERAL IN STANDARD CHINESE IN SPEAKER IDENTIFICATION

*ZHANG Cuiling LIU Xiaoli TAN Tiejun CUI Jingxu*
Department of Criminal Technology, Chinese Criminal Police College , Shenyang
zclt@webpc.edu.cn, http://www.webpc.edu.cn

## ABSTRACT

Lateral is one of the four voiced consonants in Standard Chinese and it often displays many variants in pronunciation because of different following vowels. It's distribution of formant frequency changes greatly with different vowels and assumes strong coarticulation. It is suggested that the coarticulation is different from person to person. Whereas lateral has relative stability and value of formant frequency of the same speaker assumes relative stable state. Therefore, the individual features of the coarticulation may be anticipated in speaker's sound. The aim of the article is to study coarticulation of lateral with different vowels, it's behavior in different speakers and application in speaker identification.

## 1. INTRODUCTION

There are several definitions of coarticulation, some point a language phenomena that two and above articulation characteristics appear at the same time, some indicate influence of one speech segment to the other, that is, the effect of phonetic context on given speech segment. Many researches were conducted on coarticulation in speaker identification. What we are interested in is that coarticulation is the result of speech activity of speaker, some individual difference should be exist and may be useful in speaker identification.

Lateral is one of the only four voiced consonants in Standard Chinese and it often displays many variants in pronunciation for different following vowels. It's formant frequency changes greatly with following vowels and displays strong coarticulation. It is founded that lateral has relative stability and the value of formant frequency of the same speaker assumes relative stable state. Whereas the extent of coarticulation of lateral with following vowels varies between speakers even with the same vowel. Therefore, the individual feature of the coarticulation should be useful in speaker identification.. The aim of the article is to study formant frequency distribution and coarticulation of lateral with different vowels, it's behavior in different speakers and application in speaker identification.

## 2. MATERIAL AND METHOD

### 2.1 Experimental Material

5 monophthongs（/i/、/a/、/u/、/e/ and /y/）of Standard Chinese were selected to combine with /l/ and obtained 20 monosyllables of four tones. Then entire permutation and combination of them was made and 25 two-character words were formed (7 meaningful words and 18 meaningless words were included). Finally 7 meaningful words were picked out and put in sentences. All 10 speakers are male teachers selected from Chinese Criminal Police College, 28—37 years old, born in three provinces of the Northeast of China. They can speak good Standard Chinese and have not dialect features or special speech habits.

Let ten speakers know the text well and read these monosyllables, words and sentences as naturally as possible. The recordings were made with AIWA high quality digital micro-recorder and matching microphone supported at collapsible of the chest, approximately 10 centimeters from the lips of the speaker. The recording tapes were SONY standard cassette tapes.

### 2.2 Experimental Method

All recorded sound of ten speakers was put into computer. Spectrograph analysis was made on Computer Speech Laboratory (CSL –4300). The sampling rate is 16kHz, display range 0~8kHz, effective bandwidth 234 Hz. Firstly, a wideband spectrogram was made of the item; secondly formant frequency of lateral and vowel was measured by visual observation and digital analysis. Only the first four formants of /l/ and vowels were adopted to analyze because the energy of other higher formants was too poor to measure accurately.

## 3. RESULTS AND ANALYSIS

### 3.1 Coarticulation Between /l/ and Monophthongs

Above all to deal with the coarticulation of /l/ and different monophthongs of all speakers as a whole. Table 1 presents the frequency means of the first four formants of /l/ over all speakers in each vowel environment, together with that of the vowels.

Table 1 Frequency means ( across all speaker) of formants of /l/ and different following vowels, below, the means of the means. ( Hz )

| | $LF_1$ | $VF_1$ | $LF_2$ | $VF_2$ | $LF_3$ | $VF_3$ | $LF_4$ | $VF_4$ |
|---|---|---|---|---|---|---|---|---|
| li | 323 | 348 | 1702 | 2308 | 2681 | 3033 | 3774 | 3771 |
| la | 436 | 807 | 1267 | 1315 | 2830 | 2679 | 3895 | 3681 |
| lu | 345 | 378 | 1251 | 908 | 2744 | 2756 | 3774 | 3586 |
| le | 366 | 535 | 1325 | 1370 | 2736 | 2778 | 3762 | 3740 |
| ly | 321 | 350 | 1675 | 1976 | 2670 | 2628 | 3779 | 3558 |
| mean | 358 | 484 | 1444 | 1575 | 2732 | 2775 | 3797 | 3667 |

Table 1 shows that the formant frequencies of /l/ differs with different vowels. The means of the first four formants of /l/ are $LF_1$: 358Hz, $LF_2$:1444Hz, $LF_3$: 2732Hz， $LF_4$: 3797Hz. It seems that the changes on $F_1$ and $F_2$ are much more than $F_3$ and $F_4$ with vowel environment. So $F_1$ and $F_2$ are more sensitive to different following vowels while $F_2$ is the most sensitive in the first four formants. This means the effect of coarticulation of $F_1$ and $F_2$ of /l/ with vowels is stronger than that of $F_3$ and $F_4$ while the strongest coarticulation is on $F_2$. This result can be also confirmed in Table 2.

## 3.2 Correlation Analysis on Formant Frequency Between /l/ and Following Monophthongs

Table 2 presents the correlation coefficients r for all speakers between the first four formants of /l/ and the equivalent formants of the following monophthong as a measure of coarticulation. The higher is the correlation coefficient the stronger of coarticulation is. From Table 2 it can be seen that the effect of coarticulation for /l + front i/ is weaker than /l + back u/ but stronger than /l + y/. This is probably because pronouncing position of /l/ is near with /i/ while round lip of /y/ makes all formants frequency value low.

Table 2 The ABS of Correlation coefficients on equivalent formants of /l/ and 5 monophthongs for all speakers.

| syllables | $rF_1$ | $rF_2$ | $rF_3$ | $rF_4$ |
|---|---|---|---|---|
| li | 0.5015 | 0.0408 | 0.3391 | 0.3749 |
| la | 0.0196 | 0.4177 | 0.4718 | 0.2331 |
| lu | 0.5084 | 0.3684 | 0.5022 | 0.1030 |
| le | 0.1452 | 0.4421 | 0.4460 | 0.0623 |
| ly | 0.4764 | 0.2610 | 0.2546 | 0.0343 |

It is clear form Table 2 that the extent of coarticulation on each formant varies even with the same vowel environment. For /l + i/, the coarticulation of $F_1$ (r=0.5015) is stronger than other three formants obviously, but it's $F_2$ is poor. For /l + a/, the coarticulation of $F_3$ (r=0.4718) is the best while $F_1$(r=0.0196) is the worst. This is because /i/ is a front vowel, the distance between $F_1$ and $F_2$ is the longest in all monophthongs. $F_2$ of /i/ is centred on the 2300Hz or so, the frequency mean of $F_2$ of /l/ is 1444Hz, whose great difference results in the smallest coarticulation; but in /a/, the position of tongue is very low, it's $F_1$ is the highest in all monophthongs. Though the tongue position of /l/ have drawn back and lowed a lot when producing /la/ for pronouncing /a/ there is rather long distance between $F_1$ of /l/ and $F_1$ of /a/, which explained the reason of the smallest correlation coefficient (r =0.0196) appeared in Table 2.

Table 3 shows formant frequency of lateral and it's coarticulation in different phonetic context. From Table 3 it can be seen that the formant frequency of /l/ and the extent of coarticulation varies from front and back environment, but the difference is not too significant. Variance analysis of mean was made on formant frequency of /l/ in different phonetic context (isolation, words, sentence) to verify if there exists substantial variation among different environments (in Table 3 the same letters indicate no significant difference while the different letters show significant). The result shows that the variation on $F_1$ between words and sentences is significant at 0.05 level, on $F_2$ it is significant between sentences and other two items, there is not obvious difference on $F_3$ and $F_4$. This indicates $F_1$ and $F_2$ of /l/ is affected by neighbor phonetic segment greatly, which also confirmed the coarticulation of /l/ on $F_1$ and $F_2$ is stronger than that on $F_3$ and $F_4$.

Table 3 The formant frequency means of /l/ in different environments and the result of variance analysis (Hz)

| Phonetic context | $F_1$ | | $F_2$ | | $F_3$ | $F_4$ |
|---|---|---|---|---|---|---|
| isolation | 352 | AB | 1445 | B | 2720 | 3773 |
| word | 359 | A | 1444 | B | 2747 | 3801 |
| sentence | 335 | B | 1587 | A | 2672 | 3793 |

## 3.3 Individual Characteristics of Coarticulation Between Lateral and Monophthongs

Table 4 presents the formant frequency means of /l/ for ten speakers and the coarticulation coefficients with following vowel. Statistical result shows the formant frequency distribution of /l/ differs from person to person, so does the extent of coarticulation with following vowel, which exhibits variation between speakers. It can be seen from Table 4 that frequency means of /l/ on $F_1$ is centred 330~400 Hz region, not too much variation; but $F_2$, $F_3$ and $F_4$ of them exist greater difference. In speakers $F_2$ of the N0.2 is the lowest (1352Hz), N0.7 the highest (1612Hz); the range of variation on $F_3$ is the greatest, from 2432Hz (N0.4) to 2900Hz (N0.7); the variation of $F_4$ is a little less, from 3662Hz (N0.3) to 3980Hz (N0.7). We can see individual variation from these values, especially for N0.7, whose $F_2$, $F_3$ and $F_4$ are the highest in all speakers and higher than the others evidently. The greater individuality makes it easier to identify him. This tendency also can be seen from Fig 1.

Table 4 The formant frequency means of /l/ for each speaker and the correlation coefficients on equivalent formants with following vowels.

| speakers | $lF_1$ | $lF_2$ | $lF_3$ | $lF_4$ | $rF_1$ | $rF_2$ | $rF_3$ | $rF_4$ |
|---|---|---|---|---|---|---|---|---|
| NO.1 | 353 | 1513 | 2849 | 3672 | 0.7142 | 0.8192 | 0.4147 | 0.4877 |
| NO.2 | 401 | 1352 | 2865 | 3710 | 0.7057 | 0.7983 | 0.0840 | 0.3424 |
| NO.3 | 371 | 1383 | 2761 | 3662 | 0.8014 | 0.6350 | 0.0758 | 0.1225 |
| NO.4 | 336 | 1462 | 2432 | 3763 | 0.6854 | 0.8042 | 0.3692 | 0.1155 |
| NO.5 | 352 | 1367 | 2745 | 3802 | 0.7969 | 0.8172 | 0.2420 | 0.3748 |
| NO.6 | 341 | 1495 | 2611 | 3829 | 0.7731 | 0.8699 | 0.2536 | 0.0180 |
| NO.7 | 331 | 1612 | 2900 | 3980 | 0.7766 | 0.7461 | 0.0265 | 0.2123 |
| NO.8 | 354 | 1583 | 2858 | 3705 | 0.6898 | 0.7283 | 0.0872 | 0.1956 |
| NO.9 | 337 | 1470 | 2653 | 3880 | 0.7796 | 0.7963 | 0.3109 | 0.0808 |
| NO.10 | 356 | 1444 | 2607 | 3922 | 0.4200 | 0.8536 | 0.2746 | 0.0881 |

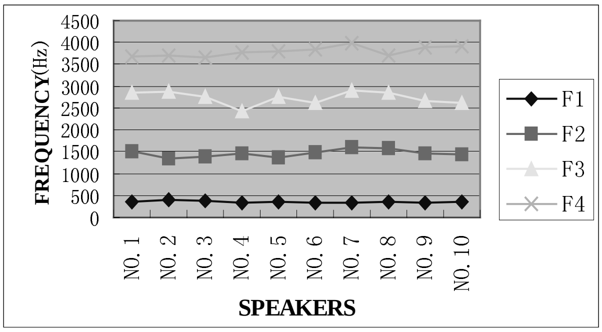


Fig 1. Frequency means of /l/ on the first four formants for ten speakers.

Correlation analysis of coarticulation shows that the extent of coarticulation of /l/ with following vowel differs for ten speakers, variation between speakers is also obvious. In them the correlation coefficient of $F_2$ for N0.6 is the highest (0.8699), which shows strong effect of coarticulation; but it's $F_4$ is the lowest (0.018) in all speakers. From this we can see significant individuality. On the other hand the information from Table 4 indicates $F_1$ and $F_2$ show stronger coarticulation than $F_3$ and $F_4$, coarticulation on $F_2$ is the strongest in the first four formants. This result also verifies a fact that common features of speakers mainly exist in low frequency part such as $F_1$ and $F_2$, while individual characteristics of a sound are usually reflected in high frequency region such as $F_3$ and above. Speaker identification should be more reliable by high frequency region of sound. Inter-speaker variation is displayed more significant in Fig2 than Fig1.

### 3.4 Standard Deviation analysis

Standard deviation (SD) on $F_1$ of /l/ in monosyllable /l.i/ for each speaker is analyzed as a parameter of intra-speaker variation and SD between different speakers as inter-speaker variation. The result of research shows SD of intra-speaker is obviously smaller than that of inter-speaker. The same SD analysis was done on $F_2$, $F_3$ and $F_4$ of /l/ in other syllables, the result gained is identical. So this result confirms quantitatively the intra-speaker variation is smaller than inter-speaker variation, which supports the basis hypothesis of speaker identification and offers foundation for forensic phonetics applying in law.

Table 5 SD of $LF_1$ in /li/ for ten speakers (Hz)

| SD | NO.1 | NO.2 | NO.3 | NO.4 | NO.5 | NO.6 | NO.7 | NO.8 | NO.9 | NO.10 |
|---|---|---|---|---|---|---|---|---|---|---|
| NO.1 | 42 | 62 | 53 | 59 | 53 | 42 | 56 | 58 | 62 | 69 |
| NO.2 | | 42 | 48 | 48 | 56 | 40 | 58 | 56 | 70 | 52 |
| NO.3 | | | 27 | 40 | 48 | 32 | 42 | 40 | 48 | 49 |
| NO.4 | | | | 24 | 37 | 29 | 34 | 39 | 42 | 47 |
| NO.5 | | | | | 29 | 34 | 49 | 47 | 59 | 52 |
| NO.6 | | | | | | 18 | 29 | 32 | 32 | 35 |
| NO.7 | | | | | | | 31 | 43 | 67 | 52 |
| NO.8 | | | | | | | | 30 | 51 | 51 |
| NO.9 | | | | | | | | | 39 | 60 |
| NO.10 | | | | | | | | | | 35 |

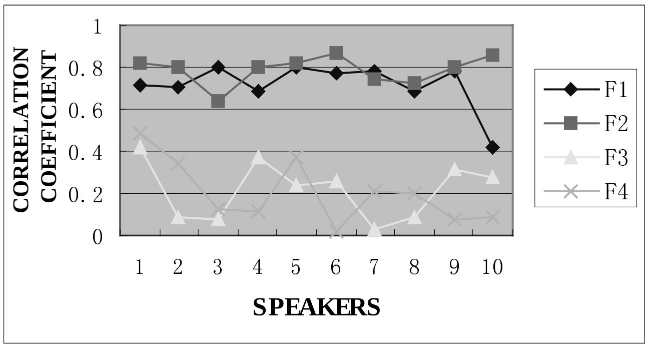


Fig 2. Correlation coefficients on equivalent formants of /l/ and following vowels for ten speakers as measure inter-speaker variation.

## 4. CONCLUSION

(1) The formant frequency distribution and degree of coarticulation of /l/ differs with different vowels. The degree of coarticulation of different formants varies even with same vowel. Generally the degree of coarticulation of /l/+ front vowel /i/ is weaker than /l/+ back vowel /u/, but stronger than /l/ + /y/. In addition, the formant frequency distribution and degree of coarticulation varies with front and back environments (isolation, monosyllabic word and sentence).

(2) The degree of coarticulation is depended on the formants, $F_1$ and $F_2$ shows greater coarticulation than $F_3$ and $F_4$, $F_2$ exhibits the strongest coarticulation effect. This result also confirmed the fact that individual

characteristics of speech exist in high frequency section such as $F_3$ and above and speaker identification should has greater credibility by high frequency section.

(3) The degree of coarticulation of /l/ with vowels is different in inter-speaker, and assumed difference between speakers. Therefore, the coarticulation of /l/ may be useful in speaker identification, /l/ is worth utilizing in a speaker identification scheme making use of segmental information. Furthermore coarticulation of lateral would be a better clue to identity than lateral alone.

(4) Formant frequency distribution of lateral and other monophthongs indicated that each speaker shows relative stability within himself though of certain variance. It seems that the variance intra-speaker is smaller than inter-speaker variance in our statistical data. This is a good result for speaker identification.

## 5. REFERENCES


[1] Wu, Zongji and Lin, Maochan, *An Outline of Experimental Phonetics*, Higher Education Press, Beijing, 1989.

[2] Cui, Jingxu, *Numerical Value Analysis of the Effect of consonant on Vowel Formants*, The 11th Science Conference Paper Volume on Calculation Physics in China, 1999.

[3] Chen, Xiaoxia, *An Acoustic Analysis and Temporal Perception Test of Lateral in Standard Chinese*, Report of Phonetic Research Institute of Linguistics (CASS) 1991, Phonetic Laboratory Institute of Linguistics, Chinese Academy of Social Sciences

[4] Yan, Jingzhu, *A Study of the Vowel Formant Pattern and the Coarticulation in the Voiceless Stop Initial Monosyllable*, Report of Phonetic Research Institute of Linguistics (CASS) 1992-1993, Phonetic Laboratory Institute of Linguistics, Chinese Academy of Social Sciences

[5] Francis Nolan, *The Phonetic Bases of Speaker Recognition*, Cambridge University Press, 1983

[6] Tosi O.et al, *Experiment on Voice Identification*, The Journal of the Acoustical Society of America, Vol.51, NO.6: 2030-2043, JAN 1972