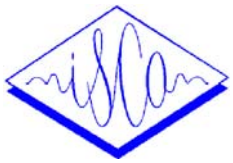

Speech Prosody 2004
Nara, Japan
March 23-26, 2004

A Study on Duration Compensation in Mandarin Chinese

Sun Lu, Hu Yu, Wang Ren-Hua, Wu Yijian

iFlyTek Speech Laboratory
University of Science and Technology of China
{sunlu,jadefox}@ustc.edu; rhw@ustc.edu.cn; jasonwu@mail.ustc.edu.cn

Abstract

This paper describes the duration compensation phenomenon in Mandarin Chinese, which means that, for two segmental units, either phonemes or syllables, their duration compensates each other with regard to different unit types. As Mandarin Chinese is such a complexity, it is impossible to find invariable rules in it. However, we observed that the general trend of compensating remains. In this paper, we investigate how durations of within syllable and between syllable segments compensate. We also enter into details about how different phoneme types, tone types and location in sentences effect on the extent of compensation, and give our analysis from the points of phonetics and phonology.

1. Introduction

Characteristics of segmental duration have long been a hot subject of scientific research, for the simple reason that it is the unavoidable and most efficient way to master how speech tempo is controlled and thus one of the cruxes of speech synthesis and recognition. Colin W. Wightman [1] reported boundary-type-dependent pre-boundary lengthening in English, and many researchers also consolidated this phenomenon in other languages, like Beckman and Pierrehumbert [2] in Japanese and Li Aijun in Chinese [3]. Beckman and Edwards reasoned about the stress-timed shortening effect, which is an "isochroous spacing of prosodically strong syllables" [4]. One of the coarticulatory variations in spoken Mandarin Chinese, namely syllable contraction, has been elaborated in [5]. End truncation, a phenomenon that denotes how duration shortening effects F0 contour, has been observed in English, Swedish and Thai [6] [7] [8].

However, duration compensation has not received so much attention, in that, there is troublesome measurement difficulty, as Beckman had warned [9]. But recently Hassan demonstrated temporal compensation between vowels and consonants in Swedish [10], and Mitsuhiko claimed to have approved this characteristic in Japanese [11]. Yet none has ever provided any thorough analysis of how this rule works under different prosodic environment, whereas prosodic cues have so much influence on segmental duration. What's more, in the aspect of Mandarin Chinese this research still remains a virgin domain. We therefore endeavor to fetch up this vacancy through this research.

Just as its name implies, duration compensation attempts to balance the overall duration by lengthening the vicinity of a phonetically short unit and shortening that of a long unit, and this here-mentioned unit could be either disyllables or initial-final pairs labeled using SAMPA-C Pinyin set [12]. Both type of compensation have been investigated in this paper. Particularly, for each type we expatiated on the compensating effect of both the anterior and posterior unit. For sake of definitude, the phrase "compensating effect of A on B", sometimes appearing as "A compensating B" is defined this way: for different types of A, their inherent duration varies, thus cause the variation of segment length of B.

Mandarin Chinese is, as is well known, a tone language. The five lexical tone types in Mandarin Chinese include high level (the first tone marked by musical points 55), rising (the second one, 35), contour (the third one, 214), falling (the fourth one, 51) and the unmarked neutral tone which usually lies on functional words or unstressed words [13], In [14], we have elucidated that tone type is an essential factor that determines segment duration, and another significant factor is location in sentences. Therefore we studied the compensating effect under different prosodic environments engendered by different tones and locations. We use 0 through 4 to denote the five tone types and as for location; we follow the definition in [14], using 1, 2, and 3 to denote head, middle and end of L3 level sentence. Thus 15 kinds of prosodic environment are considered.

This paper is organized as follows: the second part elaborates on how our experiments were carried out both for the disyllables and for the initial-final pairs, the results of our pilot experiment and analysis from the point of both phonetics and phonology. Further discussions and visualizations are arranged in the last part.

2. Methodology and Results

2.1. Materials

We experimentized on a large speech corpus of 10450 sentences, including almost 200,000 syllables. A large collection of news, essays and novels were read by an announcer from China National Radio Station with a detached tone, labeled lexical and syntactical information that includes Pinyin, tone type, location, boundary etc., and packed into a corpus that was formerly used for speech synthesis by iFlyTek. This comprehensive coverage guarantees the validity of our results.

2.2. Within Syllable Compensation

Within syllable compensation rests in two aspects, finals compensating initials, which means that finals of different length have compensating effect on their correspondent initials, and vice versa. To avoid Beckman's warning, we here didn't apply the approach of calculating correlations of two arrays of original duration data but adopted a linear regression procedure to quantify the compensating effect. Detailed descriptions and explanations of this approach are in the following major part.

We first study the compensating effect of finals on initials. In Mandarin Chinese Pinyin, there are 21 initials and 38 finals. Table1 displays the whole set of Pinyin symbols.

10.21437/SpeechProsody.2004-55

Table 1: *Initials and its classification*

Stops	b, d, g
Aspirated stops	p, t, k
Affricatives	z, zh, j
Aspirated affricatives	c, ch, q
Nasals	m, n
Fricatives	f, s, sh, r, x, h
Laterals	l

Table 2: *Finals and its classification*

Simple finals	a, o , e, i, u , v, ii, iii, er
Compound finals	ai, ei, ao, ou, ia, ie, ua, uo, ve, iao, iou, uai, uei
Nasal finals	an, en, in, vn, ang, eng, ing, ong, ian, uan, van, uen, iang, uang, ueng, iong

2.2.1. Compensating effect of finals

Figure 1 and 2 displays the duration distribution of finals and fixed initials /d/. Obviously, the length of /ang/ is comparatively longer, and /i/ shorter, and conversely, the initials before /i/ are generally longer than that before /ang/. However, we couldn't see equally clear compensating effect in other finals.

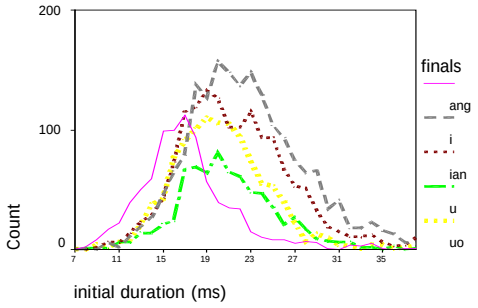

Figure 1: *Duration of finals with /d/ as initials*

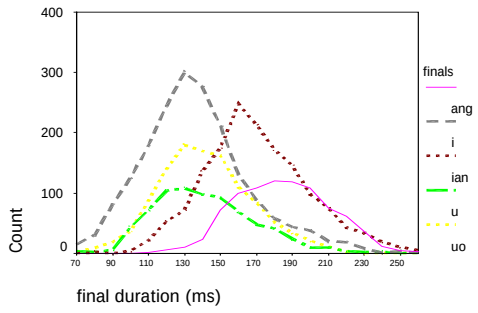

Figure 2: D*uration of initias /d/ with respect to different finals*

This might be the original observation of duration compensation, however, many researchers just stopped here and hastened to their conclusion. To get a closer and clearer look at this phenomenon as to what initials are more compensatable under what prosodic environments, we continued our research.

Before going deep into this study, one point should be firmly kept in our mind that our research objective is how different finals compensating initials, but not the correlation of stochastic duration pairs. One reason is that, the observed final duration of each sample varies greatly, and to clarify the phonetic characteristics of each final is out of the question. On the other hand, for different finals, their intrinsic duration differs a lot. According to [14], this is the most important factor determining the length of finals. So we sort the finals by average duration ascending order and make out the relation between the length of them and their correspondent initials.

We drew box plots of all the 21 initials duration by the order of corresponding finals duration, each in 15 prosodic environments, and found no prominent cues of duration compensation in stops, aspirated stops, aspirated affricatives and nasals. For stops, laterals and nasals, whose intrinsic duration is short (about 20-30ms for stops, 45ms for laterals and about 60ms for nasals) and whose variation therefore small, the ensuing finals can't exert much influence on their duration. Figure 3 is such a box plot example of 1 tone /d/ in the middle of sentences. The box shows data distribution of 0.25-quantile to 0.75-quantile, with the dark line as the median. It's apparent that its duration doesn't shorten with the lengthening of finals. For aspirate stops and aspirate affricatives, whose enunciation includes the unavoidable procedure of exhaling and thus more subject to coarticulation, they are comparatively stable in length.

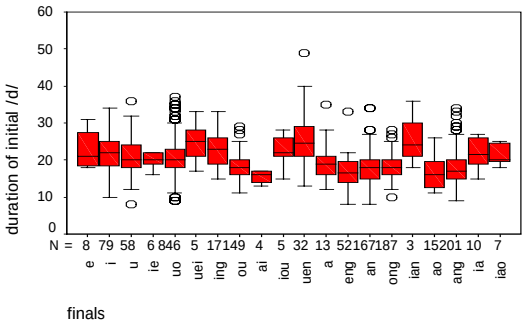

Figure 3: *Box plot of initial /d/ duration by finals duration which is in ascending order*

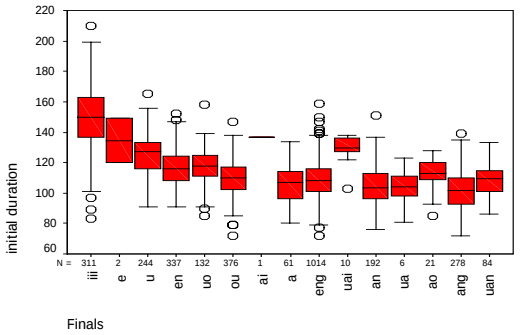

Figure 4: *Box plot of initial /sh/ duration by finals duration which is in ascending order*

But for affricatives and fricatives, the compensating effect is rather strong. Figurer 4 is an example of initial /sh/ of 1

tone and sentence middle position. It's a patency that the initial duration distribution descends with finals duration order except some special case. To assess the extent of compensating, we quantified the properties connoted in the box plots by linear regression with initials duration as the dependent variable and the finals duration order as the independent variable. We did not introduce the original finals duration as the independent variable because they are statistically too longer than that of initials, so it will be hard to fit the linear model and get appropriate coefficients for further analysis. The statistic tool SPSS [15] was adopted to perform this regression. The regression coefficient B is the slope of the fitted line, denoting how drastically the initials duration change with the finals duration, and R^2, the determinant, shows how good the model fits. Here we can't expect too good a fit due to the protean articulation and the diversity of duration distribution.

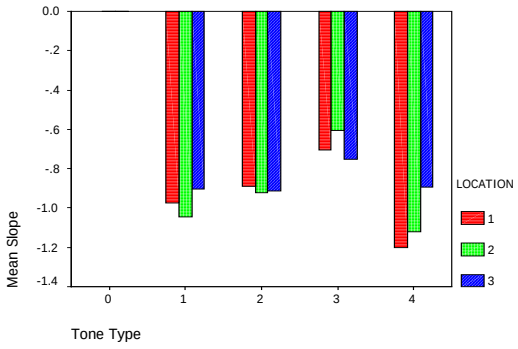

Figure 5: *Bars of slope of affricatives*

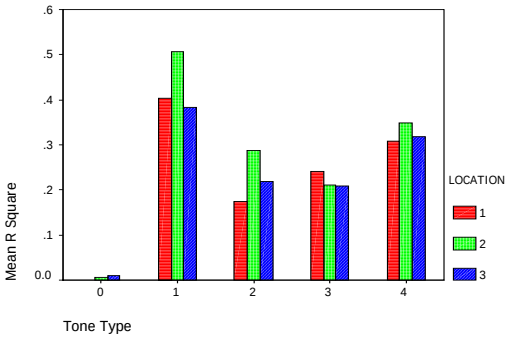

Figure 6: *Bars of R^2 of affricatives*

The above figures (Figure 5 through 8) summarizes the regression coefficients of affricative and fricative initials, showing five remarkable points: (1) the compensating effect of 0 tone seems very delicate from the mean slope value, and for some initials, the slope is even positive. Actually, this is caused by the scarcity and randomicity of 0 tone samples, and we found the trend remains for cases with enough samples; (2) samples of the first tone and the fourth tone conform better to this rule, as their determinant are generally rather high; (3) affricatives of the fourth tone and fricatives of the second tone are more sensible to its finals duration; (4) location in sentences attributes slightly for affricatives; (5) for fricatives, compensation is stronger in sentence end than in the head.

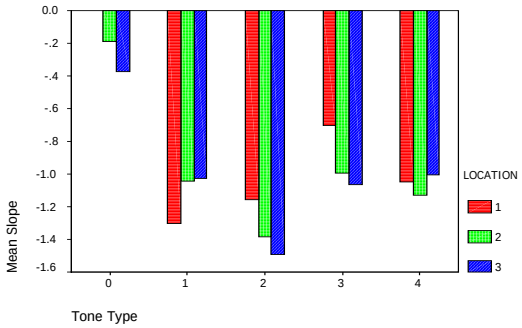

Figure 7: *Bars of slope of fricatives*

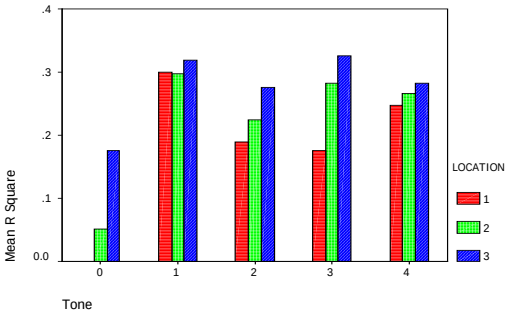

Figure 8: *Bars of R^2 of fricatives*

2.2.2. Compensating of the initials

Initials don't compensate as much to finals. We have observed the compensating effect in only several finals like /i/, /v/, /ai/ and /in/ etc. For most other finals, their length seems to have no straight relationship with correspondent initials.

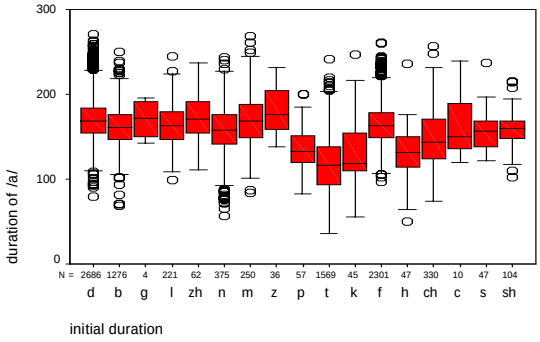

Figure 9: *Box plot of final /a/ duration by initial ascending order*

Figure 9 shows how the length of /a/ changes with the ascending of its initials duration. This box plot is congruous with Mitsuhiko's observation, who investigated only cases of /ka/ and /ma/. The duration of /k/ is generally longer than /m/, and the duration of /a/ is otherwise. But for other initials, the duration compensation theory just can't hold water.

One reason might be that finals are intrinsically long in duration and theirs length is not so subject to initials as to its

tone, thus their duration don't alter markedly with initials. Another reason could be that the number of finals is much more than initials, the samples of each final is hence inadequate and are more stochastic. Moreover, we notice that the extent of conformity of each final is closely relevant to their eta square, which we found is a demonstrable measure of one factor's influencing effect on another [14]. For example, the eta square of initials on finals duration for samples of /i/, /v/, /in/ and /ai/ are higher than 0.45, in consequence these finals are more likely to be compensated.

2.3. Disyllable Compensation

For disyllable cases, we studied the compensating effect of both the anterior and the posterior syllables. That is, how the length of the anterior syllables compensates the posterior one, and vice versa. Besides, to examine whether phonetic ingredients jumble the result, we go further to use filtered experimental data where the anterior syllable are open or closed (closed syllables are those with nasal finals, and open syllables otherwise), or where the second syllable contains no initials.

As we will always get a negative correlation between the duration of the first and second syllable if using the observed duration data, we normalized the duration of the studied unit by zscore and the other by a certain method as following:

$$Ratio = \frac{dur - \mu}{\sigma} \quad (1)$$

where *dur* denotes the observed duration of each second syllable, and μ, σ is the average and mean variance of all second syllables. The zscore represent the comparative length of each observation among its counterpart, while ratio the comparative value among all anterior or posterior syllables. If the second syllable is phonetically longer in the second-syllable set, its ratio will be positive and vise versa. Thus zscore and ratio wipe off intrinsic syllable factors, therefore we can use their correlation in our analysis.

Table 3 illustrates the compensating effect of the posterior on the anterior one. We may conclude that the compensating effect is not strong for this case.

Table 3: C*ompensating effect of posterior syllable on anterior syllable*

	Correlation
No restrictions	-0.033
Anterior syllable open	-0.016
Anterior syllable closed	-0.061
Second syllable null initials	-0.011

Table 4: *Compensating effect of anterior syllable on posterior syllable*

	Correlation
No restrictions	-0.474
Anterior syllable open	-0.458
Anterior syllable closed	-0.443
Second syllable null initials	-0.484

Table 4 displays the compensating effect of the anterior syllables on the posterior syllables.

The correlation is rather strong in all the three cases, which necessitates the compensating effect of the anterior syllable on the posterior one. The effect is even stronger if the second syllable is composed of finals only, because without initials, the finals is more subject to its prior syllable. And from the results, we can see that open anterior syllable has a little stronger compensating effect on the posterior syllable, might because that open finals are more prone to be coarticulated with the following initials.

3. Conclusions

This paper systematically and thoroughly describes the compensation phenomenon in Mandarin Chinese, both on the phoneme level and on the syllable level, and found in it a tanglesome but general rule. Whether duration compensation is an underlying factor of prosodic influences acting on segment duration and whether it is the representation of co-articulation are for further discussion, and this might be a new point of view in segmental duration research.

4. References

[1] Colin W. Wightman, Stefanie Shattuck-Hufnagel, et al, 1992. Segmental duration in the vicinity of prosodic phrase boundaries, *JASA, 91: 1707-1717.*

[2] Beckman, M. E., & Pierrehumbert, J. 1986. Intonational Structure in Japanese and English, *Phonology Yearbook, 3*, 255-309.

[3] Zu, Yiqing, 1999. Segmental Duration and Lengthened Syllables. *Proceedings of ICPhS'99.*

[4] Beckman, M. E., & Edwards, J. 1990. "Lengthening and Shortening and the nature of prosodic constituency", *Laboratory Phonology I*, by J. Kingston and M. E. Beckman(eds.), Cambridge University Press, Cambridge, 152-178.

[5] Hartmut Traunmuller, 1999, Coarticulatory effects of Consonants on vowels and their reflection in perception, *Proceedings from the XIIth Swedish Phonetics Conference*, 141-144.

[6] Grabe, Esther, Brechtje Post, Francis Nolan, and Kimberley Farrar. 2000. Pitch Accent Realization in Four Varieties of British English. *Journal of Phonetics*. 28: 161-185.

[7] Alstermark, M. & Y. Erikson, 1971 Speech Analysis A. Swedish word accent as a function of word length. *Speech Transmission Laboratory Quarterly Progress and Status Report(STL-QPSR).*

[8] Rungpat Roenpitya, 2003, End Truncation of Thai Contour Tones on Different Durations of TBU's. *Proceedings of ICPhS'03.*

[9] M. Beckman, 1982, Segmental Duration and the "Mora" in Japanese, *Phonetica* 40, 113-135.

[10] Zeki Majeed Hassan, 2003, Temporal Compensation between vowel and consonant in Swedish & Arabic in sequences of CV:C & CVC: and the word overall duration, *PHONUM* 9 (2003), 45-48.

[11] Mitsuhiko Ota, et. al, Effect of foot structure on mora duration in Japanese. *Proceedings of ICPhS'03.*

[12] http://www.cass.net.cn/chinese/s18_yys/yuyin/sampac/sampac.htm.

[13] Yuen Ren Chao, A Grammar of Spoken Chinese, *University of California Press*, 1968.

[14] Sun Lu, Hu Yu, 2003, the Statistics and Analysis of Initials and Finals Duration based on Eta Square, *NCMMSC2003*.

[15] http://www.spss.com/spssmr/regression/