SEGMENTAL DURATIONS AND LENGTHENED SYLLABLES

Zu, Yiqing & Chen, Xiaoxia
Institute of Linguistics
Chinese Academy of Social Sciences, China

ABSTRACT

This study examined three speakers' 1560 utterances in aim of getting good understanding of segmental durations in continuous speech of standard Chinese. By using normalized speaking rate which can reduces inherent/intrinsic duration, the levels of segmental lengthening are calculated. The results show that different tones, final nasals, tongue position of vowels and numbers of phonemes of final are factors to influence compensation between initial, transition and final (38% for initials, 40% c-v transitions and 22% finals). Aarticulatory manners of consonants are found to be a factor affecting initial lengthening. We also found that nearly half lengthening is implemented in middle part of syllable and voice duration which includes transition and final is easily lengthened. At last, we try to find the role of syllable lengthening as a prosodic feature by perception test and find that the lengthening syllables mainly play the stress role, but not prosodic boundaries.

1. INTRODUCTION

The issue of segmental duration is important not only for speech synthesis and speech recognition but also for phonetic research. Based on labeled speech database, this study is restricted to syllable lengthening that is one of important cues of prosody.

There is one more prosodic structure in an utterance that may have some relationship with syntactic constraint [1]. The boundaries of the prosodic structure are breaks whose realization can be a pause, pre-lengthening / final lengthening, or pitch movement / F0 reset [2,3]. It can be inferred that perceived breaks are not necessarily signaled by silence pause, but probably by other acoustic cues, such as final lengthening. By using normalized speaking rate which can reduces intrinsic duration, the levels of pause and segmental lengthening are found very consistent with prosodic boundaries within an utterance [4]. Previous study has reported that prosodic boundaries within an utterance are divided into major phrase (MAP) and minor phrase (MIP) and that nearly 100% MAP is distinguished by pause, whereas both pause and syllable lengthening will mark MIP. Syllable lengthening will occur at both pre-boundary and post-boundary. We have also found that in many cases there is no pause or lengthening at MIP at all and a lot of syllable lengthening do not occur at boundary but within prosodic phrase. This study used three speakers' 1560 utterances and examined the syllable lengthening, aiming at indicating the roles of syllable lengthening in prosody.

2. SPEAKING RATE AND SEGMENTAL LENGTHENING

2.1 The Material

A national continuous speech database suitable for training and testing Chinese speech recognizers is used in this study, which contains 200 speakers and each has about 500 utterances. In the procedure of designing reading text of speech database, phonetic knowledge has contributed to select speech units, which can describe variability in continuous speech such as junctures between syllables. We use syllables, inter-syllabic diphones, inter-syllabic triphones and final-initial structure as speech units to control the coverage [5]. None of those four sets of speech units involved prosodic information including tones, stress and other prosodic structures. Among 200 speakers' data, initials, finals and c-v transitions of 3 speakers' 1560 utterances had been hand-labeled [6]. The labeled utterances in the continuous speech database provides us the mean and standard deviation of duration of following segments [7]:

(1) 21 consonants:

b, c, ch, d, f, g, h, j, k, l, m, n, p, q, r, s, sh, t, x, z, zh;

(2) 17 voiced consonants:

bv, cv, chv, dv, fv, gv, hv, jv, kv, pv, qv, sv, shv, tv, xv, zv, zhv;

(3) 38 finals with five tones (38 finals $\times$ 5 tone (1,2,3,4,0) = 190):

a, e, ai, an ,ang, ao, ei, en, eng, er, o, ou, ong, i, ia, iao, iu, ian, ie, in, ing, iong, iang, i1, i2, i3, ua, uan, uai, ui, uang, ueng, uo, un, v, ve, van, vn };

(4) 374 c-v transitions

The data of speaker M00, M01 and M45 are used in this study.

2.1 Normalizing speaking rate

There are numerous studies on this area [8,9,10]. The duration of each segment has to be expressed in term of how longer or shorter it is than expected, i.e. we should de-emphasis the variation caused by the inherent duration of the segment. So we use normalized segment duration, instead of absolute durations [11,12]:

$$\tau = (d-\mu)/\sigma \qquad (1)$$

Where μ is mean and σ is standard deviation. The speaking rate r of an utterance is defined as:

$$r = (\sum\tau_i) / N \qquad (2)$$

Where N is the number of segments in that utterance.

The speaking rate of 1560 sentences uttered by speaker M00, M01 and M45 show normal-like distribution (See Figure 1, Figure 2 and Figure 3). The centers are 0. The normalizing

speaking rate provides a set of inherent duration to determine which segment is lengthening in m01's every sentence.

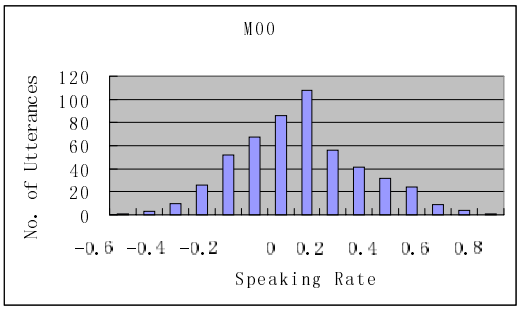

Figure 1. Distribution of speaking rates of speaker M00

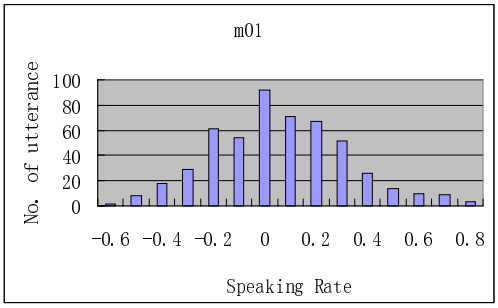

Figure 2. Distribution of speaking rates of speaker M01

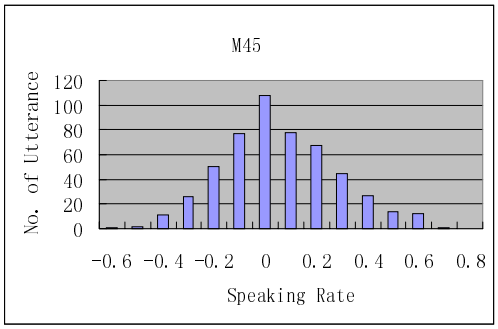

Figure 3. Distribution of speaking rates of speaker M45

2.3 The relative duration of segments

As mentioned above, to determine the segmental lengthening the inherent duration must be used. The normalized duration can not only deduce the difference of inherent duration, but also the different deviation range of durations. The speech rate should also be concerned. So we measure the relative duration by:

$$dur = \tau - r \qquad (3)$$

There is much evidence to demonstrate that segmental lengthening is one of important acoustic cues of prosodic boundaries. But how the role of lengthening plays in Standard Chinese is still a problem. Based on the data we can easily count the occurrence of segmental lengthening and classify the lengthening into several levels. To simplify problem, at first we concern the syllable lengthening, which is the sum of lengthening of initial, transition and final. Figure 4 is the procedure of this study.

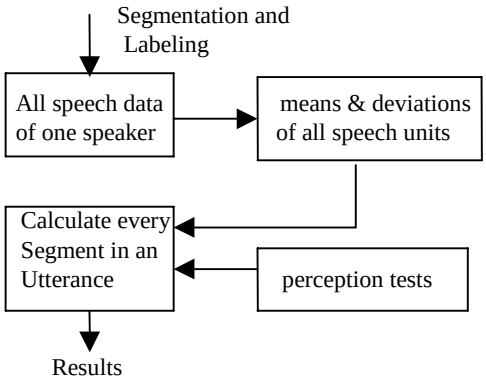

Figure 4 the procedure of data processing

3. THE SEGMENTS IN LENGTHENED SYLLABLES

3.1 The segmental compensation between initial, transition and final

Three speakers' (speakers M00, M01, M45) 1560 labeled sentences in the national database have been examined by normalized speaking rate. If the duration calculated by formula (3) is greater than zero it is defined as lengthened segment and sum of durations within a syllable can determine whether longer or shorter of a syllable is. Table 1 shows the counts of longest relative duration of initial (c), transition (c-v) and final (v) in lengthened syllables.

	c	c-v	v	uv/v
M00	1199	1033	299	1/1.11
M01	823	974	627	1/1.91
M45	930	947	856	1/1.94

Table 1. Distribution of the count for longest relative segments and ratio of unvoiced to voiced segments

The voiced segment are sum of c-v and v. Table 1 shows that the voiced segments have greater contribution to lengthened syllable. It also demonstrates that in continuous speech, the durations of consonant are relatively stable. For voiced part in a syllable, c-v transition plays more important role for syllable lengthening. In other word, middle part of a syllable or c-v transition is very important to lengthen a syllable.

To get further result, we have examined the data of speaker M01 carefully. Table 2 gives the contribution of segments c, c-v and v with different finals that are in different conditions such as tones carried by syllable, tongue position, finals with different number of phones and nasals. In Table 5:

tone0-tone4 is 5 Chinese tones;
vow1: a-set (a, e, ai, an, ang, ao, ei, en, eng, er, o, ou, ong);vow2: I-set (i, ia, iao, iu, ian, ie, in, ing, iong, iang, i1, i2);vow3: u-set (u, ua, uan, uai, ui, uang, ueng, uo, un);
vow4: v-set (v, ve, van, vn};
vo1-vo3: finals with different number (1-3) of phones
n: nasal final n (an, en, in, van, vn, uan, ian, un, van);
ng: nasal final ng (ang, eng, ong, ing, uang, ueng, iong, iang);

In M01's lengthened syllables, we found that the articulatory manner is a factor to influence the consonant lengthening (See Figure 5). Stops and fricative stops tend limited to be lengthened.

	c	c-v	v
tone0	24.7	37.0	38.4
tone1	35.9	36.7	27.4
tone2	39.8	34.2	26.1
tone3	26.1	43.5	30.4
tone4	34.7	45.84	19.4
vow1	30.0	37.9	32.1
vow2	36.9	40.2	22.9
vow3	43.9	43.9	12.2
vow4	34.2	52.1	13.7
vo1	43.8	41.6	14.6
vo2	28.0	39.4	32.6
vo3	31.85	37.3	30.8
n	31.4	34.2	34.5
ng	24.3	37.8	37.9

Table 2. Distribution of longest relative segments with different final conditions

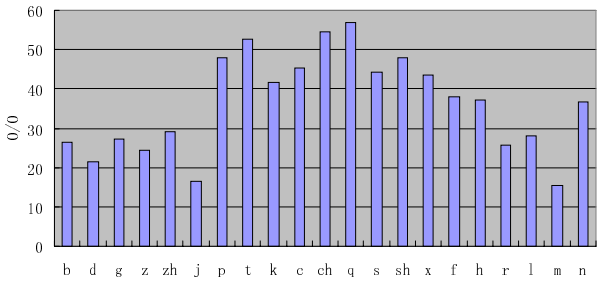

Figure 5. Distribution of proportion of lengthened initial consonants

3.2 Role of syllable lengthening

There is one more prosodic structure in an utterance, which may have some relationship with syntactic and semantic constraints. The hierarchical prosodic structure [13] in continuous speech is assumed as follows (from large to small): intonational phrase, phonological phrase, prosodic word and foot. The boundaries of the prosodic structure are breaks whose realization can be a pause, pre-lengthening/final lengthening, or pitch movement / F0 reset. To indicate the phenomena of segmental lengthening in Chinese, we have contrasted the results to a perception experiment done by Li Aijun [14], which concentrates on the m01's 157 sentences. The result shows that pause and segmental lengthening are aligned with prosodic boundaries within an utterance and that segmental lengthening not only occurs at per-boundary but also at post-boundary. Table 3 gives the count of syllable lengthening in M01's 520 sentence. The lengthening levels are 4 equal intervals between 0 to maximum of relative durations of syllable by all data of M01. Table 4 shows the relationship between lengthening levels and prosodic boundaries.

	Sum of lengthening	lengthening levels			
		1	2	3	4
no.of syll	1983	636	193	38	10

Table 3. Count of syllable lengthening on M01's data

No.	Minor boundary 58		Major boundary 213	
level	Left	right	left	right
1	21	22	69	6
2	10	7	28	3
3	1	2	2	0
4	0	0	0	0

Table 4. The relationship between lengthening levels and prosodic boundaries

We know from table 3 that 157 sentences include 1983 syllables and that among them, there are 877 syllable, less than 50%, are lengthened. In further, table 4 tells us that on 58 minor boundaries and 213 major boundaries, only 171 syllables are lengthened, which are very lower than 877. The fact shows that the lengthened syllables seem to be not closely aligned to the prosodic boundaries.

What is the role of lengthened syllable? To find the answer we use 50 sentences from 157 to test where are stress syllables and what is the relationship between stress and syllable lengthening. In this test, three native listeners, who don't know the aim of the expriment, were asked to answer where are the stress syllables in those 50 sentences.

	First syll.	Stressed syll.
0	13	4
1	19	21
2	10	15
3	3	9
4	5	8

Table 5. Count of lengthened syllable on first syllable of utterance and stressed syllable

Table 5 shows the results that on the first syllables in all sentences, 74% syllable will be lengthened and in stressed syllables, 93% will be lengthened. Therefore the results indicate that the main role of lengthened syllable plays on stress syllables. The first syllable in an utterance is always stressed. Of course, there are other factors affecting on stress, such as F0 and energy.

4. CONCLUSIONS

In this paper we describe the study on syllable lengthening based on labeled continuous speech database. With the results shown above, we get the following conclusions:

(1) We used normalized speaking rate to determine the lengthened syllable and examined the compensation between initial, transition and final and find that 38% initials, 40% c-v transitions and 22% finals play main roles in syllable lengthening. It reveals the fact that transition is a very important segment in a syllable.

(2) The analysis on lengthened syllable has been contrasted to perception experiment on prosodic boundaries, the result shows that pause is very consistent with major boundary and syllable lengthening may occur at left or right of minor boundary. But a large amount of lengthened syllables not occur at boundary.

(3) With perception test on stress of 50 sentences, lengthened syllables are found to be very closely related with stress or prominence in utterance. In many cases syllable group is stressed and all syllables in the group will be lengthened.

(4) Syllable lengthening is one of prosodic features. This initial study has not concern with F0 movement, energy and other features that will play the roles together with lengthening in prosody. For example, F0 movement may be an important cues in prosodic boundary (See Prof. Lin Maocan's report in this Conference).

ACKNOWKEDGMENT

We are grateful for Ms. Li Aijun who provided us the result of perception experiment on prosodic boundary.

REFERENCES

[1] Rossi, Mario 1997. Is Syntactic Structure Prosodically Retrievable. *Proceedings of 5th European Conference of Speech Communication and Technology*, 1, KN 1-8.

[2] Fant, Gunnar & Anita Kruckenberg 1989. Preliminaries to the Study of Swedish Prose Reading and Reading Style. *STL-QPSR 2*.

[3] Blaauw, Eleonora 1994. The Contribution of Prosodic Boundary Marks to the Perceptual Difference between Read and Spontaneous Speech. *Speech Communication*, 14, 359-375.

[4] Zu, Yiqing & Chen Xiaoxia 1998. Segmental Durations of a Labeled Speech Database and its Relation to Prosodic Boundaries. *Proceedings of the Conference on ISCSLP'98*, 83-88.

[5] Zu Yiqing 1999. The Text Design for Continuous Speech Database of Standard Chinese. *Chinese Journal of ACOUSTICS*, vol.18, No.1, 56-69

[6] Chen, Xiaoxia 1997, Segmenting and labeling on continuous speech database. in printing.

[7] Chen, Xiaoxia and Zu, Yiqing 1998. The factors affecting consonant durations in continuous speech. *The Proceedings of the Conference on Phonetics of the Languages in China*, 29-33.

[8] Crystal, Thomas H. And Arthur S. House 1982. Segmental durations in connected speech signals: Preliminary results. *J. Acoust. Soc. Am.*, 72(3), 705-716.

[9] Crystal, Thomas H. And Arthur S. House 1988. Segmental durations in connected speech signals: Current results. *J. Acoust. Soc. Am.*, 83(4), 1553-1573

[10] Santen, J P. H. Van 1992. Contextual effects on vowel duration. *Speech communication* 11, 513-546.

[11] Wightman, Colin W.& Stefanie Shattuck_Hufnagel 1992. Segmental durations in the vicinity of prosodic phrase boundaries. *J. Acoust. Soc. Am.*, vol.91, No.3, 1707-1717.

[12] Wang Xue 1997. Incorporating Knowledge on Segmental Duration in HMM-based Continuous Speech Recognition. Foris Publications, The Nethlands.

[13] Selkirk, E 1990. Phonolog and syntaxt: the relation between sound and structure, Cambridge. MA: MIT Press.

[14] Li, Aijun 1998. Dirational Characteristics of the Prosodic Phrase in Standard Chinese. *The Proceedings of the Conference on Phonetics of the Languages in China*, 65-68