

SYLLABLE DURATION AND ITS FUNCTIONS IN STANDARD CHINESE DISCOURSE

Zu Yiqing

Motorola China Research Center
Email: A16534@email.mot.com

Chen Xiaoxia, Li Aijun, Hua Wu, Sun Guohua
Institute of Linguistics
Chinese Academy of Social Sciences
5 Jian Nei Da Jie, 100732, Beijing, China

ABSTRACT

To get good understanding of prosody in continuous speech of Standard Chinese, we have collected large amount of speech in paragraph. 18 read discourse each contains 300-500 syllables are used as reading texts, which cover main discourse. We are going effort on linguistic annotation. In This paper we report works reported as follows: One male speaker's 10,000 syllables duration in discourse; the relationship between silence duration in discourse waveform and the prosodic structures.

1. INTRODUCTION

To get good understanding of prosodic features and find the basic prosodic unit in continuous speech of Standard Chinese, We are going effort on linguistic annotation such as segmental labeling, prosodic labeling and linguistics information annotation on the speech waveform and find the acoustic indicator of segmental structures and prosodic structures and their relationships to Linguistics.

Prosody in speech is very important for improving the naturalness of the synthesized speech and the accuracy of the continuous speech recognition since it communicates to hearers the intentional meaning of speakers. One of way to study prosody, such as duration and stress pattern, is extracting models from a big speech database.

Prosody involves briefly tiers: Break index (prosodic structure) tier; Stress tier; and intonation tier as described in ToBI system [1] (prosodic labeling system). Among 3 tiers, break index is easiest to access. To study the speakers how to group words together and how the duration plays in prosody the following experiments is being done.

The segmental duration plays important roles in prosody and the duration model is easy to obtain than stress model. Using duration information alone, Nick Campbell studied a way of differentiating prominence-related lengthening from boundary-related lengthening in British-English [2,3].

Duration information is significant in the encoding of two aspects of prosodic structure: making prominence and making boundaries. In the application of speech synthesis there are a lot of works in duration models [4,5,6,7].

Based on labeled speech database of isolated sentences, we have studied syllable duration of Standard Chinese, which reflects cues of prosodic structures.

2. MATERIALS

2.1 Reading Text

There are about three styles in discourse, which are narration, argumentation and conversation (see Table 1). To get good understanding of prosody in continuous speech of Standard Chinese, 18 discourses (10 narration discourse and 8 argumentation discourses) each contains 300-500 syllables are used as reading texts, which cover main discourse structures such as coherence relations as well as the Chinese phrase structures. The total materials contain about 10,000 syllables.

2.2 Recording

The recording is conducted in a sound proof room and ratio of signal to noise is greater than 50 dB. 10 higher educated native speakers, 5 males and 5 females, were asked to read 18 discourses at normal speaking rate. The speech is recorded in two channels by DAT with the sampling rate 44.4 KHz. The one channel is speech waveform; the other is impedance waveform of vocal fold detected by laryngograph that reflects opening and closure of vocal folder. Finally the speech data are converted by pass Creative live Sound Blaster with 16KHz. For speech signal is transformed from vocal fold to mouth there is a delay to impedance waveform. The time alignment of two channel signals is dealt with.

Tab. 1 Three Styles of Discourse

Narration	Sets of short stories that are fully formed narration were selected. Each story includes abstract, orientation, complicating action, evaluation, result and coda.
Argumentation	Sets of paragraphs are used to cover the logical words, which include the lasses of additive, adversative, causal, temporal.
Conversation	Main conversation structures will be concerned.

information annotation on the speech waveform.

Fig.1 - Fig. 3 are waveform, spectrum, segmental labeling, syntactic tree and prosodic tree of an example sentence in our discourse database, which is "前几年，先生因工作需要，配备了一个大哥大 (qian2 ji3 nian2, xian1 sheng0 yin1 gong1 zuo4 xu1 yao4, pei4 bei4 le0 yi2 ge4 da4 ge1 da4)", means that my husbands equip a mobile several years ago. With the annotation on speech data, we can account the linguistic and phonetic phenomena what we want. In this study we are concentrate our attention in syllable duration and their related problems.

2.3 Annotation on Speech Data

To find the acoustic indicator of segmental structures and prosodic structures and their relationships to Linguistics we are going effort on giving three levels annotation such as segmental labeling, prosodic labeling and linguistics

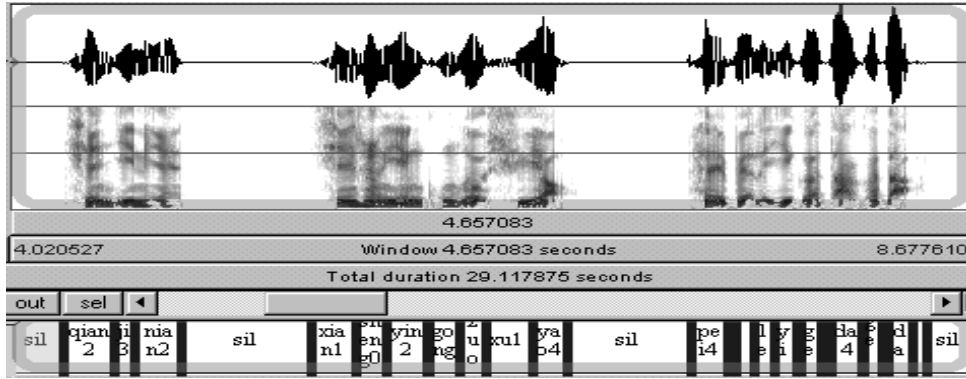


Fig. 1 The waveform, spectrum and segmental labeling on example sentence



Fig.2 The Syntactic Tree of the Example Sentence



Pw: Prosodic word, MAP: Major Phrase, IU: Intonational Utterance

Fig. 3 The Prosodic Tree of the Example Sentence

3. THE SYLLABLE DURATION IN DISCOURSE

The duration of each syllable has to be expressed in term of how longer or shorter it is than expected. To de-emphasize the variation caused by the inherent duration of syllables we use normalized syllable duration (z-score normalization, instead of absolute duration:

$$z_{ij} = (d_i - \mu_j) / \sigma_j \quad (1)$$

$$r = (\sum z_i) / N \quad (2)$$

Where N is the number of syllables in an utterance; d_i is i-th observed syllable duration in that utterance; μ_j and σ_j is j-th syllable mean and standard variances in the syllable table (totally about 400 syllable without tone) respectively. r reflects the speaking rate. z can be positive or negative, and for normally distributed data falls with a probability by 0.998 within the range of plus or minus 3. The duration of shortened syllable is shorter than the mean and will have a negative value of z; and the duration of lengthened syllable is longer than the mean and will have positive value of z.

As mentioned above, to determine the segmental lengthening the inherent duration must be used. The normalized duration cannot only deduce the difference of inherent duration, but also the different deviation range of duration. The speech rate should also be concerned.

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.

4. PERCEPTION ON SYLLABLE DURATION

There is one more prosodic structure in an utterance, which may have some relationship with syntactic and semantic constraints. The hierarchical prosodic structure [8] in continuous speech is assumed as follows (from large to small): intonation 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 studied 570 isolated sentences [9,10]. This time we labeled read discourse and still focused our attention on syllable lengthening. The read discourse speech contains 18 read discourse each with 300-500 syllables. In this test, three native listeners, who don't know the aim of the experiment, were asked to answer where are the breaks and where is the stress syllables in one of discourse (556 syllables). With the consistent result on prosodic boundary and stress syllable we gave the prosodic labels on this 556 syllables discourse.

4.1 Prosodic Boundary and the related Acoustic cues

By the perception test we defined the prosodic boundaries in 4 levels:

- I. Minor break;
- II. Major break;
- III. Phrase boundary corresponding to comma “,”;
- IV. Phrase boundary corresponding to “:”, “!”, “;”, “?”.

Table 1 shows the results. The pre-boundary lengthening is divided into 6 levels, which is defined by standard deviation, which is -3, -2, -1, 1, 2, 3. The silence duration is also listed in the table. Means and standard deviation of all syllables are obtained from all 18 read discourse of the same speaker, which contains about 10,000 syllables.

Table 2 Pre-boundary lengthening

% Levels (no. of occurrence)	Pre-boundary Lengthening					
	-3~-2	-2~-1	-1~0	0~1	1~2	2~3
I (47)	0	43	32	38	22	4
II (22)	0	5	35	45	10	5
III (47)	0	6	31	45	13	4
IV (20)	0	5	35	45	10	5

Table 3 Silence Duration

% Level (No.)	Silence Duration						
	0.00-0.25	0.25-0.50	0.50-0.7	0.75-1.00	1.00-1.50	1.50-2.00	0.00-0.25
I (47)	89	7	4	0	0	0	0
II (22)	64	64	36	0	0	0	0
III (47)	0	0	15	47	28	10	0
IV (20)	0	0	10	25	20	35	10

Table 4 Lengthening on the Stress Words

% First Syllable (62) Middle Syllable (15) Last Syllable (55)	-3~-2	-2~-1	-1~0	0~1	1~2	2~3
	First Syllable (62)	0	5	10	52	27
Middle Syllable (15)	0	13	53	13	20	0
Last Syllable (55)	0	11	44	35	5	5

5. RESULTS AND DISCUSSION

We can summarize the results as follows:

- (1) Silence duration is a parameter to distinguish the

different levels on prosodic boundary. Figure 4 shows the silence duration in different prosodic boundaries in the read discourse. The speaking rate is 3.5 syllable/sec. The silence in the discourse is included.

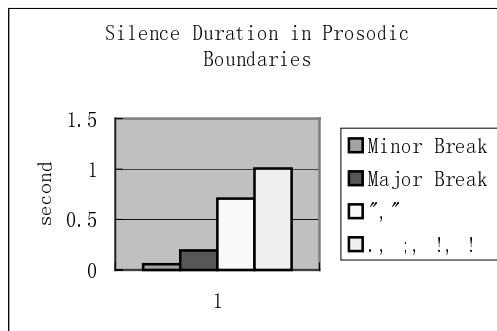


Fig. 4 Silence Duration in Prosodic Boundaries

- (2) Although syllable lengthening trends to occur in prosodic pre-boundary, the role of lengthened syllable in discourse shows little difference from the case of isolated sentences. It seems control the rhythm pattern in discourse. For example, of 47 rhythm boundaries in this discourse, there are 47 percent stress syllables just on the post of boundaries. It illustrates that before the stress word there always exist a percept minor boundary!
- (3) Among the pre-boundary syllable with no silence, for example in the case of pre-minor-boundary, the initials of post syllables are zero initials, voiced consonants and aspirated consonants. In other word, the post syllable of silence boundary is stops or stop-fricative consonants. So we need to concern on the sub-syllable duration in further.
- (4) We have examined pitch movement on the stressed syllable. In most case, there is obvious pitch reset on the stressed syllable. We will not discuss this problem in this paper. For the goals of application, in particular speech synthesis, we are going to find more acceptable results by using automatic method on bigger speech corpus. In speech synthesis a lot of studies

REFERENCE

- [1] K. Silverman, M. Beckman, J. Pitrelli, M. Ostendorf, C.W. Wightman, P.J. Price, J. Pierrehumbert and J.Hirschberg(1992), "TOBI: A standards for labeling English Prosody", Proc. ICSLP'92, pp.867-870.
- [2] Nick Campbell and S.D. Isard (1991), "Segment Duration in a Syllable Frame", Journal of Phonetics, 19, 37-47.
- [3] Nick Campbell (1993), "Automatic Detection of Prosodic Boundaries in Speech", Speech Communication 13,

pp. 343-354.

- [4] Crystal, Thomas H. and Arthur S. House (1982), "Segmental durations in connected speech signals: Preliminary results", *J. Acoust. Soc. Am.*, 72(3), 705-716.
- [5] Crystal, Thomas H. and Arthur S. House (1988), "Segmental durations in connected speech signals: Current results", *J. Acoust. Soc. Am.*, 83(4), 1553-1573
- [6] Santen, J P. H. Van (1992), "Contextual effects on vowel duration", *Speech communication* 11, 513-546.
- [7] Chilin Shih, Wentao Gu, Jan P. H. van Santen (1998), "Efficient Adaptation of TTS Dration Model to New Speakers", ICSLP98, Sidney Australia.
- [8] Selkirk, E 1990. Phonology and syntax: the relation between sound and structure, Cambridge, MA: MIT Press.
- [9] Zu, Yiqing & Chen Xiaoxia (1998), "Segmental Durations of a Labeled Speech Database and its Relation to Prosodic Boundaries", *Proceedings of the Conference on ICSLP'98*, 83-88.
- [10] Zu, Yiqing (1999), "Segmental Duration and Lengthened Syllables", Proceedings of ICPHs'99.