John Benjamins Publishing Company

Chinese disyllabic words in conversation

Shu-Chuan Tseng
Institute of Linguistics, Academia Sinica

This paper presents a study of segment duration in Chinese disyllabic words. The study accounts for boundary-related factors at levels of syllable, word, prosodic unit, and discourse unit. Face-to-face conversational speech data annotated with signal-aligned, multi-layer linguistic information was used for the analysis. A series of quantitative results show that Chinese disyllabic words have a long first syllable onset and a long second syllable rhyme, suggesting an edge effect of disyllabic words. This is in line with disyllabic merger in Chinese that preserves the onset of the first syllable and the rhyme of the second syllable. A shortening effect at prosodic and discourse unit initiation locations is due to a duration reduction of the second syllable onset, whereas the common phenomenon of pre-boundary lengthening is mainly a result of the second syllable rhyme prolongation including the glide, nucleus, and coda. Morphologically inseparable disyllabic words in principle follow the "long first onset and long second rhyme" duration pattern. But diverse duration patterns were found in words with a head-complement and a stem-suffix construction, suggesting that word morphology may also play a role in determining the duration pattern of Chinese disyllabic words in conversational speech.

Keywords: Segment duration, within-word and boundary-related factors, discourse segmentation units, conversational speech

關鍵詞: 音段音長, 詞組邊界, 言談單位, 對話口語

1. Introduction

Phonetic variability occurring in natural, continuous speech is shaped by both of the segmental and suprasegmental properties of spoken words. Among the well-studied features, e.g. formants, loci, intensity, pitch, and duration etc., duration is clearly one of the most stable features representing timing patterns in various languages (Beckman & Edwards 1990, Crystal & House 1990, Crompton

Chinese Language and Discourse 5:2 (2014), **231–251**. DOI 10.1075/cld.5.2.05tse
ISSN 1877-7031 / E-ISSN 1877-8798

1980, Klatt 1975, Lehiste 1972, Lindblom 1963, Shen 1993). According to the isochrony hypothesis, segments are in principle shorter in long suprasegmental units than in short ones (Lehiste 1977). Van Santen and Shih (2000) empirically identified this effect in American English and Mandarin Chinese. Van Santen and Shih (2000: 1017) also mentioned contextual factors, though not necessarily the most important one: "The duration of any segment in a syllable can be predicted from the predicted duration of the syllable, the identity of the segment, and only minimal information about contextual factors." Correlated with these so-called contextual factors, higher-level discourse structures hosting *boundaries* and *units* contribute to the local and global temporal organization (Dankovičová 1997), as natural speech production needs to be structured in a prosodically coherent and conceptually comprehensive way to provide a clear discourse segmentation and to enhance efficient understanding (Tao 1996, Liu & Tseng 2009). Thus, two boundary-related factors standing for the contextual effect were found significantly correlated with the segment/syllable duration in Mandarin Chinese by Van Santen and Shih (2000): Phrase (final vs. non-final) and word (initial vs. non-initial). Similarly, but using controlled data of Mandarin Chinese, Xu and Wang (2009) found that syllable duration varies according to its phrasal position, i.e. phrase-final syllables are the longest, followed by phrase-medial and phrase-initial syllables. Also related to boundary effect in Chinese, Lai et al. (2010) found that the duration of the first syllable is slightly longer than the second syllable in disyllabic words, whereas the final syllable is consistently the longest one in tri- and quadrasyllabic words.

Concerning units and boundaries in conversation, phonetic details in the produced speech may be relevant to the perception of linguistic boundaries (Smith & Hawkins 2012). As a spoken discourse is structured by boundaries in the sense of segmentation, the resulting units should be able to be perceived and used by the recipients to construct discourse context and build up meaningful comprehension in conversational interaction. An empirical, comparative study of French and Mandarin Chinese discourse and prosodic units has shown that approximately the half of discourse unit completion locations are aligned with prosodic unit boundaries at both initial and final ends. Furthermore, around 80% of discourse units are aligned with prosodic units at either the initial or the final end in both languages (Prévot et al. 2013). That is to say that in addition to syntactic account that expectedly leads to syntax-prosody matches (Clark & Wasow 1998, Selting 1996, Shattuck-Hufnagel & Turk 1996, Watson 2010), factors such as pragmatic function (including disfluencies) may also influence the way speech is produced. Beyond Selkirk's (1984) theoretical proposal of a syntax-prosody mechanism, pre-boundary lengthening and pausing in general blueprint the so-called syntactic timing pattern in speech communication, as pre-boundary lengthening is a clear phonetic indicator marking constituent units (Lehiste 1972, Klatt 1975). In particular, a number of duration studies have been conducted for boundary effects. Lengthening at the level of words was found in the final position of intonation units in English, including accented, stressed, or unstressed words (Bouzon & Hirst 2002). The first phonological word within an intonation phrase tends to be short in Czech (Dankovičová 1997). Similarly, initial syllables in phonological phrases are subject to a significant shortening effect in French (Crompton 1980). But differently, word-initial lengthening instead of shortening was found by Turk & Shattuck-Hufnagel (2000) for English. Among these previous studies conducted for pre- and post-boundary effects at different levels of units, pre-boundary lengthening is a rather common phenomenon in the aforementioned studies, whereas post-boundary effect seems to vary across languages.

This paper aims to provide a detailed, quantitative corpus-based study on segment duration in conversational speech by accounting for boundaries resulting from various supragmental units (discourse unit, prosodic unit, word, and syllable). Segment duration that appears to be a local phonetic aspect, as a matter of fact, contributes to the global, temporal organization of speech. Although duration pattern has been widely examined, most of the studies focus on effects from single levels only. The study presented in this paper was conducted on Chinese conversational data that have been annotated with linguistic information about various production unit boundaries. Conversational data, unlike well-controlled laboratory data, requires proper data normalization before analyses can be conducted, because differences across speakers and phoneme classes have to be eliminated first. Thus, scaling data normalization was applied to our dataset of Chinese disyllabic words to achieve a reasonable control on the data. Our aim is to study the boundary effect as well as to identify within-word duration patterns of Chinese disyllabic words, e.g. syllable components and onset-rhyme relationship.

2. Data and method

2.1 Multi-layer linguistic annotation

We used a corpus subset extracted from the Taiwan Mandarin Conversational Corpus (the TMC Corpus) for our study (Tseng 2013).[1] The dataset contains three and a half hours of speech data produced by sixteen Mandarin Chinese native speakers, aged between 14 and 59. It has been annotated with linguistic boundaries at levels of discourse unit (DU), prosodic unit (PU), word, syllable, and segment — completed in several research projects (Liu & Tseng 2009, Chen 2011, Prévot et al. 2013, Liu et al. 2014). An annotated example with these various units is shown in Figure 1. A DU is defined to include all complements and modifiers (tense, aspect, modality) adhering to an identified main predicate. Additional criteria were necessary to include discourse connectives and interactive dialogue phenomena, which normally appear in isolation, but with a specific pragmatic function in conversational speech (Givón 1993, Chen 2011, Prévot et al. 2013). PU was manually annotated based on pauses, pitch reset, final lengthening, and tempo change according to impressionistic judgments (Liu & Tseng 2009). Although the definition of a PU is oriented at that of an Intonation Unit used for conversational analysis (Du Bois et al. 1993, Tao 1996), we give no account of the content of speech in our PU annotation in the current study, but rather regard PU as a type of prosodic segmentation. Comparing boundary alignment between DU and PU, we found that final boundaries of DUs are aligned with approximately 80% of those of the PUs and initial boundaries with 70% of the PUs. Despite independent annotation stages and procedures, the match between DU and PU suggests that the organization of discourse meaning and structure is manifested by means of prosody, too (Prévot et al. 2013).

Syllable boundaries were manually annotated according to perceptual judgments and visualization of acoustic features (spectrogram, pitch, and intensity) that were agreed by at least two annotators. With signal-aligned syllable boundaries, the output of the CKIP automatic word segmentation and part of speech (POS) tagging system was directly implemented to obtain the respective word boundaries in the speech data (Chen et al. 1996, Tseng 2013). Word segmentation is a well-known, difficult task in processing Chinese data, as the definition of words directly affects the segmentation result. In our study, we do not follow any particular theory of Chinese word morphology, but instead we based on the CKIP system in which a list of common words constructed by referring to Chinese dictionaries provides the main basis for word segmentation with an additional set of morphological rules to deal with derived words and compounds. Details about the CKIP word segmentation standard please refer to Chen et al. (1996). In this paper, we are concerned with disyllabic words, for which the segmentation is often uncontroversial.

The annotation of segment boundaries was accomplished by adopting a combined verification procedure. Preliminarily, a phone aligner trained with monophone acoustic models using the HTK toolkit was applied to the dataset to obtain automatic phone boundaries. These boundaries then underwent a verification procedure (instead of "actually labeling") by six labelers divided into three pairs. The instruction we gave to the labelers was that they should judge (only by listening) whether the preliminary boundaries suitably separate the adjacent sounds. If it is not the case, they should adjust the boundary to where they think it best separates the two segments. All six labelers had many years' of experiences working with phonetic labeling before the verification experiment. After a successful training session, 90% of the verified phone boundaries deviate within 20ms among the labelers. The final version of our dataset was then completed by time marking the middle position of the two boundaries verified by each of the two labelers without further human intervention (Liu et al. 2014).

As a result, the dataset contains 122,945 segments produced in 49,236 syllables, 33,330 words, 8,043 prosodic units, and 4,332 discourse units. In Figure 1, the speech excerpt presented in PRAAT (Boersma & Weenink 2013) contains three discourse units and five prosodic units, presented at six linguistic levels: DU, PU, word, POS, syllable, and segment.

DU: *Nangang guoqu de hua* (**PU**)
(If you) go from Nangang
DU: *wo jiu bijiao bu qingchu* (**PU**)
I don't know it exactly
DU: *xiang* (**PU**) *wo meitian shangban shi cong* (**PU**) *jieyun Yongchun Zhan* (**PU**)
For instance (**PU**) I go for work every day from (**PU**) Yongchun MRT station (**PU**)

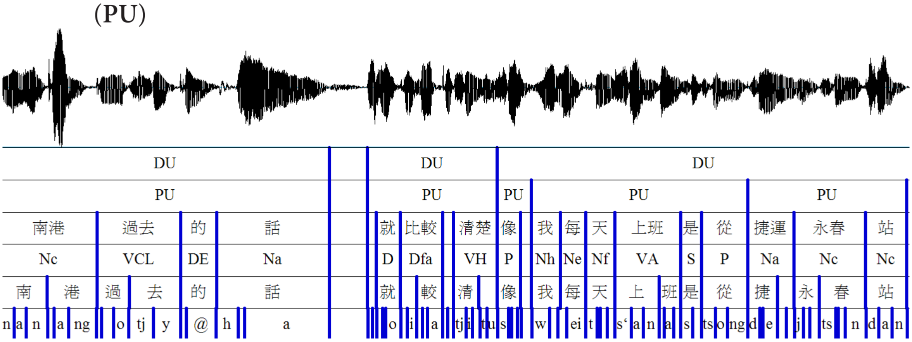

Figure 1. Data with signal-aligned multi-layer linguistic annotation

Discourse particles, markers, and fillers are often used in Chinese conversation (Tseng 2013). They were excluded from our analysis for the reason that their prosodic representation (due to strongly discourse-oriented pragmatic functions in conversation) is normally very different from that of ordinary words that we focus on in the current study. Concerning the location of words, whenever we need to identify the location of a disyllabic word relative to a PU or DU boundary, these discourse-relevant items are taken into consideration. For example, a disyllabic word followed by a sentence-final particle is regarded as located at a prosodically penultimate position, i.e. a PU non-final position. The same criterion also applies to DU positions.

2.2 Inventory and data normalization

Mandarin Chinese has a syllable structure of **consonant-glide-vowel-consonant** (CGVC) with no consonant clusters. The inventory of consonants and vowels in Mandarin Chinese can slightly differ according to the respective sound system (Duanmu 2000, Ho 1996). In our phoneme inventory, we include two glides /j/ and /w/ only and classify instances of monophthongs preceded by /ɥ/ into the diphthong classes (e.g. *yue* the moon). The occurrences of /ɥ/ are significantly less frequent than /w/ and /j/ and the variants of /ɥ/ are not as stable as /j/ and /w/ in Taiwan Mandarin. Thus, it would have caused problems when training our phone aligner, if /ɥ/ were trained as an independent acoustic model. For the phoneme inventory, four syllable components (**onset**, **glide**, **nucleus**, and **coda**) are distinguished. **Onset** can be occupied by a set of 21 consonants /p t k pʰ tʰ kʰ m n l f s ts tsʰ ʂ tʂ tʂʰ ʐ ɕ tɕ tɕʰ x/. **Glides** are /j/ and /w/. **Coda** can only be /n/ or /ɲ/ in Mandarin Chinese. The **nucleus** inventory consists of /a ə e o ɨ ʉ i u y ai au əɻ ei ou ye/.

Two data normalization approaches that are often adopted nowadays were used to normalize formant values of Russian vowels in Lobanov (1971). The scaling approach normalizes a formant value F by calculating (F-Fmin)/(Fmax-Fmin), whereas Fmax and Fmin are the maximum and minimum formant values produced by a particular speaker. The z-score approach normalizes a formant value by calculating $(F-\mu)/\sigma$, whereas μ is the mean of all formant values produced by a particular speaker and σ the standard error. In this study, we applied the scaling normalization approach to our data. That is, the normalized (scaled) duration was defined as (*x-min*)/(*max-min*), where *x* is the original observation value (segment duration) and *max* and *min* are the maximum and the minimum values calculated for each segment produced by each speaker. The original duration values were normalized relative to each of the 39 segment classes separately. After the normalization procedure, differences across phonemes and speakers are diminished by scaling the duration of all segments in our dataset to values between 0 and 1.

2.3 The dataset

55% of word tokens in the present dataset are monosyllabic. 40% and 4% are disyllabic and trisyllabic, respectively. Equivalently, 45,347 segments were found in monosyllabic (37%), 65,817 in disyllabic (54%), and 9,628 in trisyllabic words (8%). Irrespective of word length, CV, CVC, CGV, CGVC are the four most frequently produced syllable structures in Chinese conversation, followed by GV, V, VC, and GVC. The statistical summary is given in Table 1. The aim of this study is to examine the effects from multiple levels including word-internal ones. As monosyllabic words give no insight into within-word effects, we focus on disyllabic words that form the largest set of words having more than one syllable. Please also note that the numbers of **tokens per type** of disyllabic words in Table 1 are consistently high in all syllable structure categories, i.e., we have an appropriate data distribution in terms of syllable types among our data of disyllabic words. Moreover, segment duration that was normalized by the scaling method is normally distributed, thus eligible for further statistical tests.

Table 1. Data overview

Syllable structure	# of word tokens			# of word types			Tokens per type		
	1-syll.	2-syll.	3-syll.	1-syll.	2-syll.	3-syll.	1-syll.	2-syll.	3-syll.
CV	9,555	12,123	1,910	309	590	242	30.92	20.55	7.89
CVC	1,973	5,325	649	198	333	145	9.96	15.99	4.48
CGV	2,767	3,293	654	119	197	74	23.25	16.72	8.84
CGVC	997	1,927	175	89	152	56	11.2	12.68	3.13
GV	2,288	1,448	201	16	42	19	143	34.48	10.58
V	686	1,400	181	34	67	30	20.18	20.9	6.03
VC	33	601	55	10	33	12	3.3	18.21	4.58
GVC	130	569	150	17	38	14	7.65	14.97	10.71
Total syllables	18,429	26,686	3,975	792	1,452	592	23.27	18.38	6.71
Total words	18,429	13,343	1,325						

3 Segment duration

3.1 Overall data

Figure 2a shows normalized segment duration means of the overall data in terms of the eight syllable structure types. A clear tendency is observed for Chinese syllables produced in conversation that the onset is longer than the glide, the nucleus, and the coda. In a closed syllable, the coda is usually shorter than the onset, but longer than the nucleus. Our result shows that the glide is produced with a longer duration than the nucleus in GV and GVC syllables, suggesting that glides, at least in zero-onset syllables may not be subject to a peripheral production, but phonetically very prominent with regard to duration in Chinese. In Figure 2b, where the accumulated syllable component means are shown, although we observed a kind of isochronous effect of vowels (the nucleus) that are longer in syllables with fewer segments than in those with more segments, syllables with a small number of segments are in general accordingly shorter than those with a larger number of segments. As a result, it is shown that in conversational speech data, Chinese syllables tend to have a long onset and in cases of zero-onset syllables a long glide.

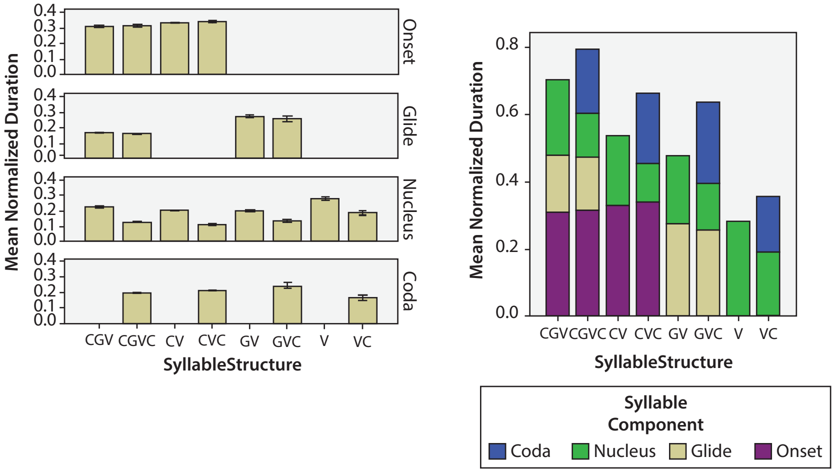

Figure 2. Syllable component means: (a) Separately (b) Accumulated

3.2 Within-word factors in disyllabic words

We then concretely considered three word-internal factors for studying the duration patterns in Chinese disyllabic words: (1) segment class (consonants by manner of articulation, glides, and vowels), (2) syllable position in disyllabic words (S1, S2), and (3) syllable structure (CV, CVC, CGV, CGVC, GV, V, VC, GVC). Figure 3 summarizes duration means of syllable components in terms of these three factors. Figure 3a shows that **onset** is longer than **glide**, **nucleus**, and **coda**, similar to the aforementioned general tendency observed in the overall data in Figure 2. Concerning segment classes, affricates, fricatives, retroflex, and plosives are longer than nasals, lateral, and approximants. The latter ones normally vary according to the immediately neighboring sounds. It is likely that the resulting assimilation leads to a short duration. Moreover, the only one consonant that can be both of the onset and the coda is /n/, so onset and coda /n/ were given different labels in our data. As shown in Figure 3a, /n/ is on average longer in the onset position than in the coda position. Figure 3b shows that **onset** is longer in S1 than in S2, whereas the **glide**, **nucleus**, and **coda** in S2 are slightly longer than in S1. Concerning syllable structure types, the data of disyllabic words (Figure 3c) show exactly the same tendencies as the overall data (Figure 2a). As mentioned above, the duration of **nucleus** tends to be short in a more complex syllable such as in CVC, GVC, and CGVC syllables with an onset or a coda, but clearly to be long when produced in V syllables.

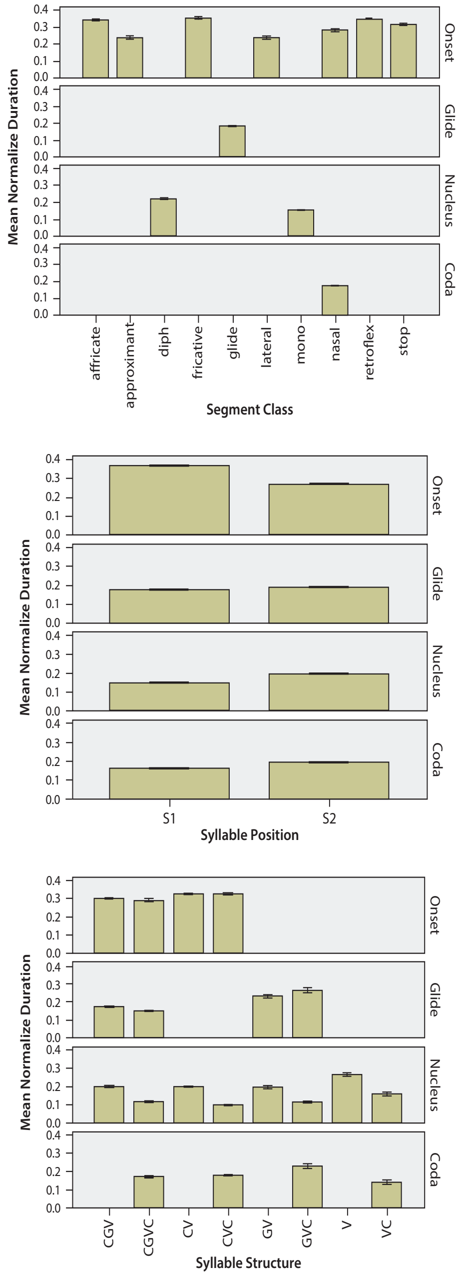

Figure 3. Duration of syllable components with regard to (a) segment class, (b) syllable position, and (c) syllable structure

3.3 Higher level factors

It is likely that discourse structuration in conversational speech to a certain degree determines the phonetic representation of spoken words, as discourse segmentation leads to different levels of boundary effects. Prosodically initial and final positions are often regarded as the positions that correlate with boundaries in a spoken discourse. Therefore, we labeled our disyllabic words according to their position relative to PU and DU boundaries by distinguishing three positions: **initial**, **medial**, and **final**. For DU and PU that are composed of only one word, they are dealt with separately and were classified as **1word DU** and **1wordPU**. To be more specifically, **1word DU** and **1wordPU** consist of only one disyllabic word. The duration pattern in Figures 4a and 4b shows that all four syllable components were produced with a longer duration in DU- and PU-final positions than in initial and medial positions, suggesting an effect of final lengthening. In addition to final lengthening, an initial shortening effect is also observed in Figures 4a and 4b. For in-depth observations, when an initial disyllabic word is shortened, it is the onset in S2 that is reduced. When a final word is lengthened, it is the rhyme in S2 that is prolonged. Lai et al. (2010) reported a slightly longer S1 in Chinese disyllabic words. As shown in Figure 4a and 4b, the durations of S1 and S2 in disyllabic words vary according to their positions within DU and PU in our conversational data. It is clear that within-word duration patterns are dependent on boundary effects as well, particularly in continuous speech. The patterns found in Lai et al.

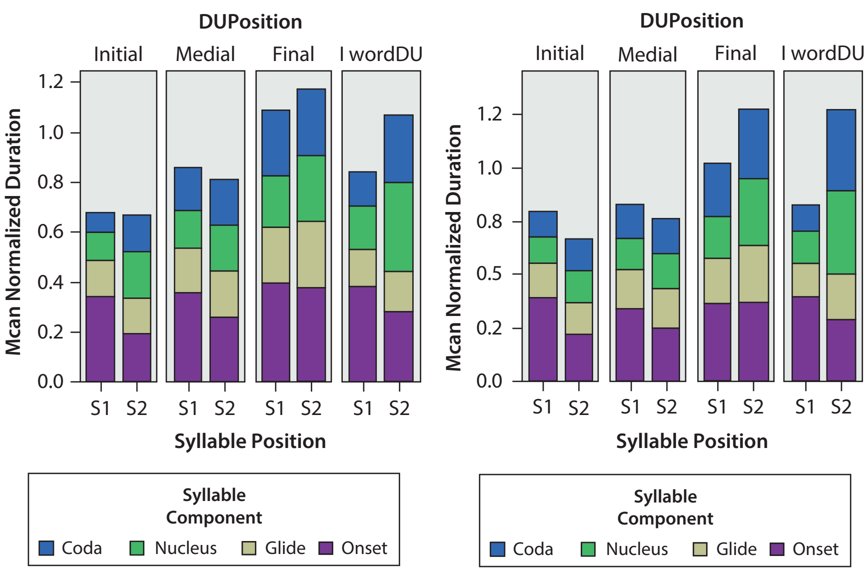

Figure 4. Segment and syllable durations wrt. (a) DU position and (b) PU position

(2010) were identified in words located at initial and medial positions in DU as well as in PU, but not in words at final positions and in **1DUword** and **1PUword**, either. As we can see from Figure 4 and 5, in these cases, S2 is longer than S1.

3.4 Statistical verification

To statistically verify the aforementioned duration patterns, we ran a generalized linear modal with the normalized duration as the dependent variable, PositionComponent, PUPosition, DUPosition, SyllableStructure, SegmentClass, and POS as predictors.

PositionComponent:	S1_Onset, S2_Onset, S1_Glide, S2_Glide, S1_Nucleus, S2_Nucleus, S1_Coda, S2_COda
PUPosition:	Initial, Medial, Final, 1wordPU
DUPosition:	Initial, Medial, Final, 1wordDU
SyllableStructure:	V, VC, GV, GVC, CV, CVC, CGV, CGVC
SegmentClass:	Fricative, plosive, affricate, nasal, approximant, lateral, retroflex, glide, monophthong, diphthong
POS:	A, Caa, Cab, Cbb, D, Da, Dfa, Dfb, Di, Dk, I, Na, Nb, Nc, Ncd, Nd, Nep, Neqa, Neu, Nf, Ng, Nh, P, SHI, T, VA, VAC, VB, VC, VCL, VD, VE, VF, VG, VH, VHC, VI, VJ, VK, VL (Appendix A)

The position relative to a PU as well as that to a DU is significant in determining the duration of segments in disyllabic words, Wald $\chi2(3) = 104.1$; $p < 0.01$ for PUPosition and Wald $\chi2(3) = 35.5$; $p < 0.01$ for DUPosition. So are SegmentClass and SyllableStructure, Wald $\chi2(7) = 64.3$; $p < 0.01$ and Wald $\chi2(7) = 107.9$; $p < 0.01$, respectively. Concerning individual syllable components separated in terms of S1 and S2, PositionComponent also shows a significant effect in segment duration, Wald $\chi2(5) = 80.8$; $p < 0.01$. So is the POS of words to which the segments belong, Wald $\chi2(39) = 165.2$; $p < 0.01$. The statistical result is summarized in Appendix B. Despite the significant effect of POS, we will not further discuss this factor, as disyllabic words with varying POS lead to a large number of combinations of syllable structures. In this case, controlled laboratory data with restricted variety types are required.

We ran univariate variance analyses with Scheffe's post hoc comparisons on each of the predictors separately. The statistical results confirmed that **onset** is significantly longer than **glide**, **nucleus**, and **coda** ($p < 0.01$). **Glide** is significantly longer than **nucleus** and **coda** ($p < 0.01$), whereas there is no significant difference between **nucleus** and **coda**. For DUPosition and PUPosition, the segments produced in **initial** position are statistically the shortest, followed by those in **medial** position, those in **1wordDU/PU**, and those in **final** position ($p < 0.01$ for all the differences). That is, an initial shortening and a final lengthening of segments (similarly also in terms of syllables and disyllabic words) were clearly identified at prosodic as well as discourse unit boundaries, as illustrated in Figure 4. For the onset consonants, **affricates**, **retroflex**, and **fricatives** were produced with a longer duration, compared to the other segment classes ($p < 0.01$), followed by **stops** ($p < 0.01$). Surprisingly, **monophthongs** were produced with a shortest duration, followed by **glides** and **nasals**. These three segment classes were significantly shorter than the others ($p < 0.01$). It may be probably due to the facts that monophthongs and glides are often reduced or omitted in conversation and nasals are often reduced or assimilated with the adjacent segment to form nasalized variants. Concerning syllable components in different within-word positions, **S1_Onset** was produced with a longest duration and **S1_Nucleus** with a shortest duration, compared with the other syllable components ($p < 0.01$). For S2-syllables, the onset (**S2_Onset)** is the longest component ($p < 0.01$). **S2_Nucleus** is longer than **S2_Glide** and **S2_Coda**

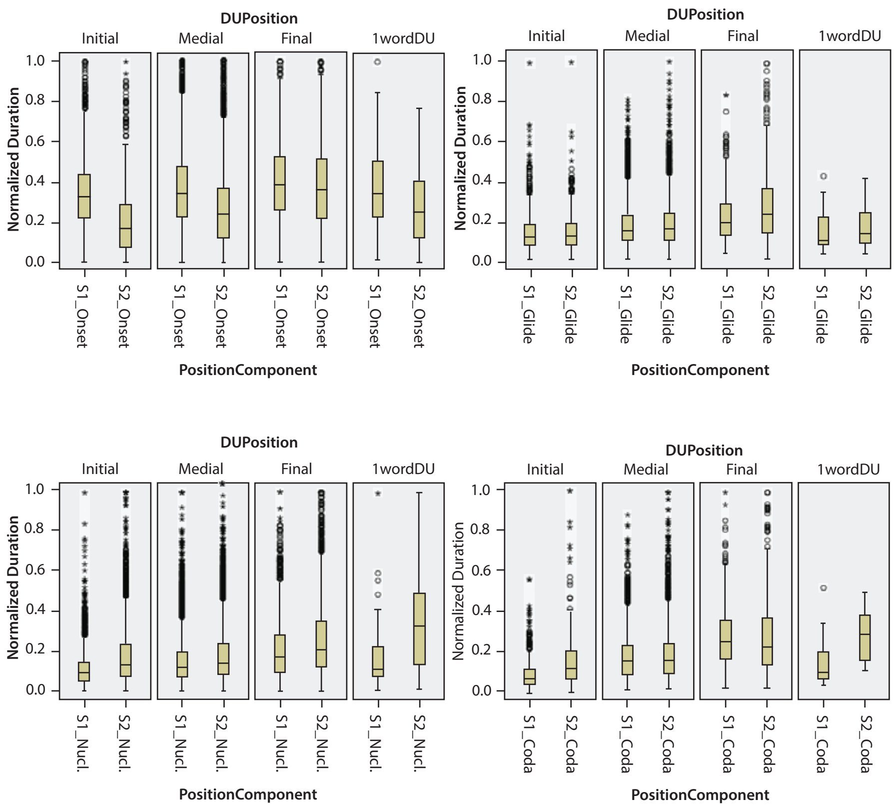

Figure 5. Boundary effects concerning DU position

($p < 0.01$). The within-word duration pattern in S1 and S2 is in principle similar, in which the onset is longer than the other components in each of the two syllables. However, the nucleus is the shortest component in S1, but it is not the case for S2.

4. An integrated account for disyllabic word duration pattern

4.1 Boundary effects

PU and DU annotation has been conducted independently at two stages, each based on strictly defined prosodic and structural cues. We mentioned earlier that these two higher levels of units are aligned at both ends in the majority of the data. Figures 5 and 6 detail the original duration distribution (in boxplots) of the four syllable components in terms of syllable position and DU/PU position. Generally speaking, duration patterns of segments do not deviate too much in terms of these

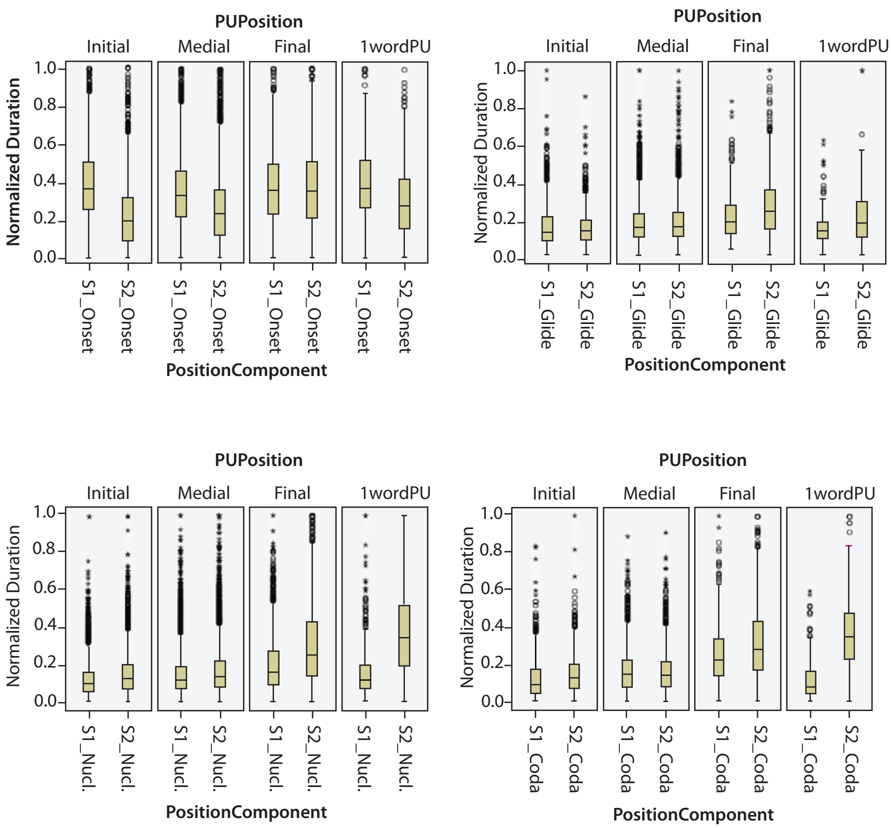

Figure 6. Boundary effects concerning PU position

two unit boundaries. For studying changes in disyllabic words, we regard the duration pattern of **1wordPU** and **1wordDU** as the basic pattern, as they are produced in isolation with no (obvious) influences from boundary effects. The basic pattern is described as follows. The S1 onset is longer than the S2 onset, whereas the glide, the nucleus, and the coda are all longer in S2 than in S1. Figures 5 and 6 showing the original data distribution detail our results. The difference in the onset duration is enlarged, when a disyllabic word is produced at a unit initiating position, but shrunk at an ending position. Contrastively, the difference in all four components is kept at a medium scale for those at a medial position. In disyllabic words located at a unit ending position, almost all four components in S1 and S2 are produced with a longer duration than their counterparts in the **1word** cases. Based on these results, we draw a clear and concrete conclusion for Chinese disyllabic words: the shortening effect at a unit initiating position is mainly caused by the heavily reduced duration of the onset in S2, whereas the final lengthening effect is caused by a prolongation of all components in both S1 and S2.

4.2 Onset-rhyme relation

As duration pattern is the end output mixing influences from both of the segmental and suprasegmental levels, we extended our scope of analysis from segment to onset/rhyme. CV, CVC, CGV, and CGVC syllables (those with an onset), were studied in terms of syllable position (S1, S2). An ANOVA test was conducted and reported with a significant result, $F(3,49392) = 963.4$; $p < 0.01$. Scheffe's post hoc tests show that the onset in S1 is the longest, followed by the rhyme in S2 and the onset in S2 ($p < 0.01$). The rhyme in S1 is the shortest among all ($p < 0.01$). This result shows that Chinese disyllabic words have a long first syllable onset and a long second syllable rhyme. Figure 7 illustrates the results in terms of syllable structure and syllable position. The onset is longer when it appears in S1 than in S2; and the rhyme is longer in S2 than in S1. It seems that this result is the consequence of a kind of edge effect of disyllabic words in Mandarin Chinese. Syllable merger, regarded as a kind of phonological phenomenon, can be predicted by the Edge-in theory in which the onset of S1 and the rhyme of S2 are preserved, accompanied with certain changes in the merging of the two nuclei (Chung 1997, Hsu 2003). In respect of phonetic representations of Chinese spoken words, disyllabic word merger and those not completely merged disyllabic words are as a matter of fact reduced spoken words. Applying the edge effect in disyllabic words, we should be able to predict phonetic forms of reduced disyllabic words, i.e. to preserve or delete certain syllable component according to the degree of pronunciation clarity and their position within different suprasegmental units in conversation.

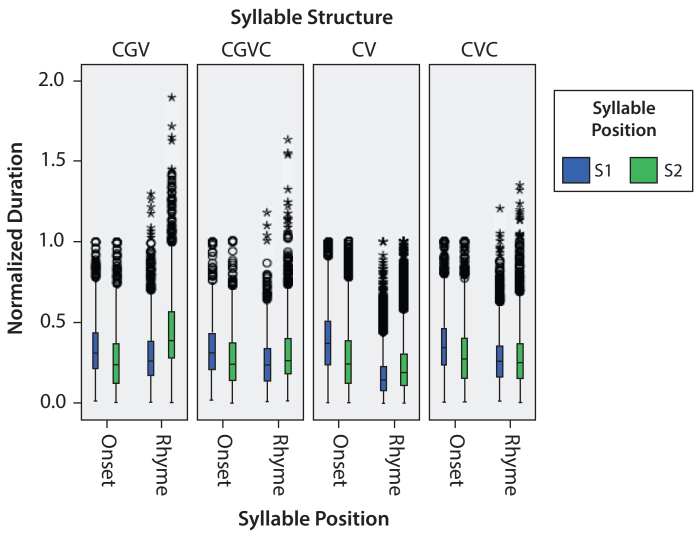

Figure 7. Long S1-onset and long S2-rhyme pattern in disyllabic words

However, despite this significant statistical result, different tendencies of the onset-rhyme pattern may still exist due to the structure of syllables and the production frequency of spoken words. For example, the long S1-onset and long S2-rhyme pattern is the most prominent for CGV syllables. But it does not appear to be very obvious for CV syllables. The reason may lie in the high word frequency of CV-CV disyllabic cue phrases in conversation that often lead to extreme merger such as keyi /kə i/ (can) to /kei/ in which the S2-rhyme is largely shortened.

4.3 Word morphology

The above analysis identified a common duration pattern of Chinese disyllabic words in conversational data. However, disyllabic words have varying morphological structures that may lead to different degrees of prominence in word meaning (Chao 1968). We would like to explore the issue whether word morphology also plays a role in duration pattern. A pilot study was conducted to examine three small groups of disyllabic words from our dataset with (possibly) the same syllable structure, representing morphologically inseparable words and words with a head-complement and a head-suffix construction. These include four content words, four verbs with directional complements, and two plural pronouns. These words are all high frequency words. The data size and distribution of these selected words are comparable to each other. The four content words that cannot be decomposed into smaller morphological units are *zhidao* (know), *houlai* (afterwards), *keshi* (but), and *dagai* (probably). The directional verbs that have a head-complement construction are *huilai* (return-hither), *huiqu* (return-thither), *chulai* (exit-hither), and *chuqu* (exit-thither). The plural pronouns that consist of a pronoun and a plural suffix -*men* are *women* (we) and *tamen* (they). Concerning syllable structure, except for *huilai*, *huiqu*, *tamen*, and *women*, the selected disyllabic words contain only CV syllables.

As shown in Figure 8, the duration pattern of disyllabic words systematically varies according to their morphological pattern. The morphologically inseparable words have a consistent *long S1-onset and long S2-rhyme pattern*, as shown in Figure 7. For words with a head-complement construction, this pattern is not observed, but more like two independent monosyllabic morphemes, i.e. *the long-onset pattern* for both of S1 and S2, as identified in Figure 2. Different from these two duration patterns, disyllabic words with a suffix have *a strong S1 onset and S1 rhyme, but a relatively weak S2 rhyme*. The S2 rhyme tends to be heavily reduced and the bilabial nasal onset in S2 merged with S1, becoming a bilabial nasal coda in the merger. Because bilabial nasals cannot be an eligible coda in Mandarin Chinese, the merger syllable *wom* with a bilabial nasal coda sufficiently and successfully marks the existence of a plural suffix. It should be easily recognized by the listeners in interactional conversation. These results suggest a likely relationship between word morphology, phonology, and the surface form of spoken words. Not only boundary-related factors such as discourse and prosodic unit boundaries may affect the duration pattern of segments in conversation, within-word factors

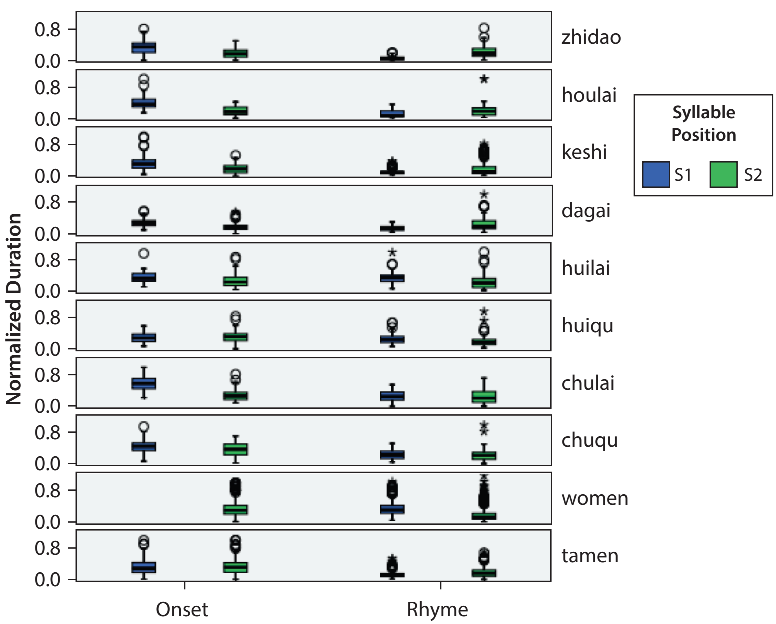

Figure 8. Duration patterns in morphologically different disyllabic words

such as word morphology and within-syllable factors such as segment class, number, and component type may all play a role.

5. Conclusion

The study presented in this paper utilized conversational data for analyzing segment duration in an integrated account. Corpus-based, quantitative analyses bring up in-depth, insightful issues for phonetic studies by benefiting from a large dataset with reliable annotation works. Uncontrolled data like conversational speech may cause difficulties in drawing deterministic conclusions. For instance, we are not able to investigate the effects of tones on duration pattern in disyllabic words due to the lack of balanced data with an appropriate size of tone combinations. Nevertheless, this study demonstrates our efforts of using well-annotated conversational data to account for multiple boundary-related effects on segment duration in disyllabic words. As prosodic and discourse units are results of spoken discourse segmentation, phonetically reduced spoken words are not merely phonetic reduction. They are results of boundary effects as well, as heavily reduced words in conversation are either located at phrase- or sentence-initial positions or they are high-frequency function words or repeated word sequences, acting as a kind of prosodic marking cue (Clark & Wasow 1998, Liu & Tseng 2009, Pluymaeker et al. 2005, Tseng 2006). This study identified a shortening effect of the initial position in discourse and prosodic units. Cue phrases in Chinese (very often disyllabic words) such as connectives that discourse units often start with are normally produced with a short duration in conversation (Liu & Tseng 2009). However, this type of prosodic marking in initial positions can also be performed by lengthening in other languages (Turk & Shattuck-Hufnagel 2000). Thus, post-boundary effects seem to be a language dependent feature, which is different from final lengthening that has been commonly found in many languages (Prévot et al. 2013, Shattuck-Hufnagel & Turk 1996, Tao 1996). In our duration analysis, it became clear that the onset-rhyme relation marking an edge effect in disyllabic words enhances a systematic disyllabic contraction procedure that makes the phonetic forms of reduced disyllabic words in Mandarin Chinese predictable. If this is true, then pronunciation variants of spoken disyllabic words may be theoretically derivable for Chinese, and these rules may probably be represented in the mental lexicon of Chinese speakers as well.

Acknowledgements

The author would like to sincerely thank two anonymous reviewers of *Chinese Language and Discourse* for their precious comments on the manuscript, and the team members who have been working on the Chinese conversational corpus data along the years. The annotation work presented in this paper was financially supported by the France-Taiwan ORCHID Program (under grant 100-2911-I-011–504), Academia Sinica, and the National Science Council (under grant 100-2411-H-001–093).

References

Beckman, M., and J. Edwards. 1990. "Lengthenings and Shortenings and the Nature of Prosodic Constituency." In *Papers in Laboratory Phonology I*, ed. by J. Kingston, and M. E. Beckman, 152–178. Cambridge: Cambridge University Press.

Boersma, P., and D. Weenink. 2013. "PRAAT: Doing Phonetics by Computer." http://www.fon.hum.uva/praat5.3.60.

Bouzon, C., and D. Hirst. 2002. "The Influence of Prosodic Factors on the duration of Words in British English". In *The Proceedings of Speech Prosody 2002*, 191–194. Aix-en-Provence.

Chao, Y.-R. 1968. *A Grammar of Spoken Chinese*. Berkeley and Los Angeles: University of California Press.

Chen, K.-J., C.-R. Huang, L.-P. Chang, and H.-L. Hsu. 1996. "SINICA CORPUS: Design Methodology for Balanced Corpora". In *The Proceedings of the Eleventh Pacific Asia Conference on Language, Information, and Computation*, 167–176.

Chen, A. C.-H. 2011. *Prosodic Phrasing in Mandarin Conversational Discourse: A Computational-acoustic Perspective*. Doctoral dissertation. Graduate Institute of Linguistics, National Taiwan University.

Chung, R.-F. 1997. "Syllable Contraction in Chinese". In *Chinese Languages and Linguistics III. Morphology and Lexicon*, ed. by F. Tsao, and S. Wang, 199–235. Taipei: Academia Sinica.

Clark, Herbert, and T. Wasow. 1998. "Repeating Words in Spontaneous Speech." *Cognitive Psychology* 37: 201–242. DOI: 10.1006/cogp.1998.0693

Crompton, A. 1980. "Timing Patterns in French." *Phonetica* 37: 205–234. DOI: 10.1159/000259993

Crystal, T., and A. House. 1990. "Articulation Rate and the duration of Syllables and Stress Groups in Connected Speech." *Journal of the Acoustical Society of America* 88 (1): 101–111. DOI: 10.1121/1.399955

Dankovičová, J. 1997. "The Domain of Articulation Rate Variation in Czech." *Journal of Phonetics* 25: 287–312. DOI: 10.1006/jpho.1997.0045

Du Bois, J., S. Schuetze-Coburn, S. Cumming, and D. Paolino. 1993. "Outline of Discourse Transcription". In *Talking Data Transcription and Coding in Discourse Research*, ed. by J. A. Edwards, and M. D. Lampert, 45–90. Hilsdale: Lawrence Erlbaum Associates.

Duanmu, S. 2000. *The Phonology of Standard Chinese*. New York: Oxford University Press.

Givón, T. 1993. *English Grammar: A Function-based Introduction*, volume 2. Amsterdam: John Benjamins.

Ho, D.-a. 1996. *Some Concepts and Methodology of Phonology*. Taipei: Da-An Press. (in Chinese)

Hsu, H-c. 2003. "A Sonority Model of Syllable Contraction in Taiwanese Southern Min." *Journal of East Asian Linguistics* 12: 349–377. DOI: 10.1023/A:1026108613211
Klatt, D. 1975. "Vowel Lengthening is Syntactically Determined in a Connected Discourse." *Journal of Phonetics* 3: 129–140.
Lai, C., Y. Sui., and J. Yuan. 2010. "A corpus Study of the Prosody of Polysyllabic Words in Mandarin Chinese". In *The Proceedings of Speech Prosody 2010*. Chicago.
Lehiste, I. 1977. "Isochrony Reconsidered." *Journal of Phonetcs* 5: 253–263.
Lehiste, I. 1972. "The Timing of Utterances and Linguistic Boundaries." *Journal of the Acoustical Society of America* 51 (6): 2018–2024. DOI: 10.1121/1.1913062
Lindblom, B. 1963. "Spectrographic Study of Vowel Reduction." *Journal of the Acoustical Society of America* 35: 1773–1781. DOI: 10.1121/1.1918816
Liu, Y.-F., and S.-C. Tseng. 2009. "Linguistic Patterns Detected Through a Prosodic Segmentation in Spontaneous Taiwan Mandarin Speech". In *Linguistic Patterns in Spontaneous Speech*, ed. by S.-C. Tseng, 147–166. Institute of Linguistics, Academia Sinica.
Liu, Y.-F., S.-C. Tseng, and J.-S. R. Jang. 2014. "Phone Boundary Annotation in Conversational Speech". In *The Proceedings of LREC 2014*, 848–853, Reykjavik.
Lobanov, B. M. 1971. "Classification of Russian Vowels Spoken by Different Speakers." *Journal of the Acoustical Society of America* 49: 606–608. DOI: 10.1121/1.1912396
Pluymaeker, M., M. Ernestus, and R. H. Baayen. 2005. "Lexical Frequency and Acoustic Reduction in Spoken Dutch." *Journal of the Acoustical Society of America* 118: 2561–2569. DOI: 10.1121/1.2011150
Prévot, L., S.-C. Tseng, A. C.-H. Chen, and K. Peshkov. 2013. "A Quantitative Comparative Study of Prosodic and Discourse Units, the Case of French and Taiwan Mandarin". In *The Proceedings of the 27th Pacific Asia Conference on Language, Information, and Computation* (PACLIC 27), 92–101. NCCU, Taipei.
Selkirk, E. 1984. *Phonology and Syntax. The Relation between Sound and Structure*. Cambridge, Mass. & London: MIT Press.
Selting, M. 1996. "On the Interplay of Syntax and Prosody in the Constitution of Turn-constructional Units and Turns in Conversation." *Pragmatics* 6 (3): 371–388.
Shattuck-Hufnagel, S., and A. E. Turk. 1996. "A Prosody Tutorial for Investigators of Auditory Sentence Processing." *Journal of Psycholinguistic Research* 25 (2): 193–247. DOI: 10.1007/BF01708572
Shen, X. S. 1993. "Relative Duration as a Perceptual cue to Stress in Mandarin." *Language and Speech* 36 (4): 415–433.
Smith, R., and S. Hawkins. 2012. "Production and Perception of Speaker-specific Phonetic detail at Word Boundaries." *Journal of Phonetics* 40 (2): 213–233. DOI: 10.1016/j.wocn.2011.11.003
Tao, H. 1996. *Units in Mandarin Conversation: Prosody, Discourse, and Grammar*. Amsterdam: John Benjamins. DOI: 10.1075/sidag.5
Tseng, S.-C. 2006. "Repairs in Mandarin Conversation." *Journal of Chinese Linguistics* 34 (1): 80–120.
Tseng, S.-C. 2013. "Lexical Coverage in Taiwan Mandarin Conversation." *International Journal of Computational Linguistics and Chinese Language Processing* 18 (1): 1–18.
Turk, A. E., and S. Shattuck-Hufnagel. 2000. "Word-Boundary-Related Duration Patterns in English." *Journal of Phonetics* 28: 397–440. DOI: 10.1006/jpho.2000.0123
Van Santen, J. P. H., and C. Shih. 2000. "Suprasegmental and Segmental Timing Models in Mandarin Chinese and American English." *Journal of the Acoustical Society of America* 107 (2): 1012–1026. DOI: 10.1121/1.428281

Watson, D. G. 2010. "The Many Roads to Prominence: Understanding Emphasis in Conversation." *Psychology of Learning and Motivation* 52: 163–183. DOI: 10.1016/S0079-7421(10)52004-8
Xu, Y., and M. Wang. 2009. "Organizing Syllables into Groups - Evidence from F0 and Duration Patterns in Mandarin." *Journal of Phonetics* 37 (4): 502–520. DOI: 10.1016/j.wocn.2009.08.003

Appendix A.

CKIP POS Tagging system (Tseng 2013, p. 7)

Adjectives:	Non-predicative adjective (A)
Adverbs:	Adverb (D), quantitative adverb (Da), pre-verbal adverb of degree (Dfa), post-verbal adverb of degree (Dfb), sentential adverb (Dk), aspectual adverb (Di)
Conjunctions:	Coordinate conjunction (Caa), correlative conjunction (Cbb), conjunction: *deng3deng3* (Cab), conjunction: *de5hua4* (Cba)
Determinatives:	Demonstrative determinatives (Nep), quantitative determinatives (Neqa), specific determinatives (Nes), numeral determinatives (Neu), post-quantitative determinatives (Neqb)
Foreign words:	Foreign words (FW)
Interjections:	Interjection (I)
Nouns:	Measure (Nf), common noun (Na), proper noun (Nb), place noun (Nc), localizer (Ncd), time noun (Nd), postposition (Ng), nominalization (Nv)
Particles:	Particle (T)
Prepositions:	Preposition (P)
Pronouns:	Pronoun (Nh)
Sentence:	Nominal expression, idioms (S)
Verbs:	Active intransitive verb (VA), active pseudo-transitive verb (VB), stative intransitive verb (VH), stative pseudo-transitive verb (VI), active causative verb (VAC), active transitive verb (VC), active verb with a locative object (VCL), ditransitive verb (VD), active verb with a sentential object (VE), active verb with a verbal object (VF), classificatory verb (VG), stative causative verb (VHC), stative transitive verb (VJ), stative verb with a sentential object (VK), stative verb with a verbal object (VL), you3 (V_2)
DE:	Structural particles: *de5, zhi1, de2, di4*
SHI:	Copula: *shi4*

Appendix B.

GLM results

Dependent variable: Normalized segment duration	Wald χ2	df	Sig.
(Intercept)	1741.251	1	.000
PUPosition	104.125	3	.000
DUPosition	35.499	3	.000
SegmentClass	64.307	7	.000
SyllableStructure	107.864	7	.000
PositionComponent	80.835	5	.000
POS	165.162	39	.000

Author's address

Shu-Chuan Tseng
Institute of Linguistics, Academia Sinica
Nankang 115, Taipei, Taiwan

tsengsc@gate.sinica.edu.tw