Sandhi-Tone Words Prolong Fixation Duration during Silent Sentence Reading in Chinese

Jinger Pan [(1)], Caicai Zhang [(2), (3)], Xunan Huang [(2), (3)] and Ming Yan [(4)]

(1) Department of Psychology, The Education University of Hong Kong, New Territories, Hong Kong SAR, China

(2) Department of Chinese and Bilingual Studies, The Hong Kong Polytechnic University, Kowloon, Hong Kong SAR, China

(3) Research Centre for Language, Cognition, and Neuroscience, The Hong Kong Polytechnic University, Kowloon, Hong Kong SAR, China

(4) Department of Psychology, University of Macau, Taipa, Macau SAR, China

Abstract

The current study examined whether or not lexical access is influenced by detailed phonological features during the silent reading of Chinese sentences. We used two types of two-character target words (Mandarin sandhi-tone and base-tone). The first characters of the words were in the sandhi-tone condition subject to a tonal alternation, but no tonal alternation was involved in the base-tone condition. Recordings of eye movements revealed that native Mandarin Chinese readers viewed the base-tone target words more briefly than the sandhi-tone target words when they were infrequent. Such articulation-specific effects on visual word processing, however, diminished for frequent words. We suggest that a conflict in tonal representation at a character/morpheme level and at a word level induces prolongation in fixation

duration on infrequent sandhi-tone words, and conclude that these tonal effects appear to reflect articulation simulation of words during the silent reading of Chinese sentences.

Keywords: eye movement, tone sandhi, sentence reading, Chinese

Sandhi-Tone Words Prolong Fixation Duration during Silent Sentence Reading in Chinese

During the reading of continuous text, sound-defining properties of orthographic patterns are activated automatically for readers' lexical processing. One topic that has been under considerable debate is whether or not detailed phonological features are used during visual word identification. One one hand, according to the *minimality hypothesis* (Frost, 1998), the early use of phonological codes during visual word recognition implies that the effective phonological form must be simple and abstract, devoid of speech-specific features including articulation specific sub- and supra-phonemic variations. The hypothesis suggests that the use of phonological knowledge for visual word recognition has little in common with the use of acoustic features for auditory word recognition. On the other hand, a number of recent studies have showed that readers of alphabetic languages use relatively detailed phonological features, such as vowel duration, spoken syllable boundaries, and lexical stress, during visual word identification (e.g., Huestegge, 2010; Inhoff, Connine, & Radach, 2002). For instance, there have been reports of shorter reaction times to target words pronounced with **short** vowels (e.g., "deaf") than to orthographically matched words with long vowels (e.g., "deal"; c.f., Abramson & Goldinger, 1997; Lukatela, Eaton, Sabadini, & Turvey, 2004).

Event-related potential (ERP) studies have also revealed an early use of sub- and supra-phonemic features in isolated word recognition. Ashby, Sanders, and Kingston (2009) reported that, when reading words with a final unvoiced consonant (e.g., *fat*) or voiced consonant (e.g., *fad*), preceded by masked nonword primes that were either incongruent or congruent with the target word in terms of the voicing of the final consonant, participants showed an early congruency effect such that the incongruent

condition elicited more negative ERP amplitudes, as early as 80 ms after stimulus onset. In a later study of phonological syllable priming effects, Ashby (2010) observed a smaller amplitude of the N1 component in the phonologically congruent (i.e., same first syllable e.g., *po## – PONY*) than in the incongruent condition (i.e., first syllable with one letter more or fewer, e.g., *pon# – PONY*). In combination, these experiments demonstrate that skilled English readers process detailed phonological information early in visual word recognition.

All the behavioral and ERP studies reviewed above employed an isolated word recognition task using a priming paradigm. It is worth pointing out that, in daily life, words seldom occur in isolation, but rather appear in context. As such, it is less known whether detailed phonological features are activated in more ecologically valid reading conditions, such as sentence reading. Both isolated word recognition and sentence reading involve immediate lexical access. However, in sentence reading, readers can fixate and refixate freely on words, as well as skip them. They can also process words in the parafovea before fixating on them and regress back to words that have been processed previously, for reanalysis. Indeed, eye-tracking allows psychological and linguistic researchers to understand processes underlying natural reading through nonintrusive measurement of reading behaviors in an online sentence-reading task. Fixation durations and other oculomotor parameters are then used to make inferences about the underlying cognitive processes. It has been documented widely that eye-movement measures such as fixation duration are highly sensitive to linguistic variables. For example, shorter time is spent fixating words with frequent lexical forms, words with regular phonology and words that are more likely to be guessed from prior context, than those having infrequent forms, irregular phonology and low predictability (see Rayner, 2009 and Kliegl, Nuthmann, &

Engbert, 2006, for reviews). Therefore, fixation duration is accepted widely as a reliable index for processing difficulty.

Eye-movement studies have shown that, during the silent reading of sentences, relatively detailed phonological features can be obtained. For instance, Huestegge (2010) reported longer fixations on words with longer vowels. Using stress-alternating homographs (e.g., *PREsent* vs. *preSENT*) embedded in limericks, Breen and Clifton (2011) reported reading costs when readers encountered mismatches between the predicted and actual stress patterns. A series of studies by Ashby and her colleagues provided consistent evidence for an early use of articulation-specific sub- and supra-phonemic features during visual word recognition. Ashby and Clifton (2005) found that lexical stress could influence fixation durations: words with two stressed syllables took longer to read and received more fixations than those with one stressed syllable. Ashby, Treisman, Kessler, and Rayner (2006) showed that English readers spent shorter time processing words preceded by nonword parafoveal previews containing vowels with matching than with mismatching vowels (e.g., *cherg – chirp* vs. *chorg – chirp*). Similarly, it was demonstrated that readers spent shorter time on words following parafoveal previews with congruent than with incongruent syllables (e.g., *de_πxw – device* vs. *dev_πx – device*; Ashby & Rayner, 2004). Together, these findings in English indicate that articulation-specific features influence visual word identification during the online reading of sentences.

Although transparency of print to sound mapping (i.e., orthographic depth) varies across different writing systems, all alphabetic scripts, overall, explicitly express the identity and sequence of phonemes in oral language. Therefore, when these orthographies are processed, the extraction of visual information can be associated directly with the activation of phonemic units. In contrast, logographic characters

differ fundamentally from alphabetic letters and are not designed to represent phoneme-specific properties. In the example of Chinese, the most widely used logographic writing system, the characters are designed to represent morphemes, and visually similar characters can have different pronunciations (Hoosain, 1991). Previous studies have suggested that orthographic depth has a profound effect on the use of phonological information (see Frost, 2012; Hoosain, 1991; Reilly & Radach, 2012, for reviews). Given the low orthographic consistency, some language-specific features of Chinese may prevent its readers from early and fast access to phonology. First, sublexical information from character radicals provides unreliable phonological information: by character type, only 26% of characters have the same pronunciations as their phonetic radicals when tones are taken into account (Zhou, 1980). Additionally, the reliability of phonetic radicals in predicting character pronunciation decreases when frequency is considered (i.e., character token): more frequent characters tend to be less regular (Shu, Chen, Anderson, Wu, & Xuan, 2003). Second, there is a high degree of homophony in Chinese, with over 5000 commonly used characters mapped to about 400 unique syllables in Mandarin Chinese, leading to the situation that phonological access to a character does not lead to precise access to the intended lexical representation of a specific character form. For instance, the syllable /yi4/ is shared by over 200 characters.

Given the language-specific properties of the Chinese writing system, it is of great theoretical interest to explore how phonological information is represented and activated in this deep orthography. Along this line, it is worth noting that the role of phonology during Chinese lexical access has been under considerable debate. Overall, the use of the sound code exists also for Chinese visual word recognition (see Perfetti, Liu, & Tan, 2005, for a review). Eye-movement studies of Chinese sentence reading

further demonstrated that words in the parafovea can be processed (Yan & Sommer, 2019; Yang, Rayner, Li, & Wang, 2012) and phonological information of a parafoveal word is available for lexical processing even before it is fixated (e.g., Liu, Inhoff, Ye, & Wu, 2002; Tsai, Lee, Tzeng, Hung, & Yen, 2004; Yan, Richter, Shu, & Kliegl, 2009). Nevertheless, considerable experimental evidence suggests that semantic knowledge can be accessed directly in Chinese without phonological mediation (e.g., Chen & Shu, 2001; Feng, Miller, Shu, & Zhang, 2001;Yang, Wang, Tong, & Rayner, 2012; Zhou & Marslen-Wilson, 2000). Furthermore, a direct comparison of phonological and semantic parafoveal preview effects revealed that the latter is more robust (Tsai, Kliegl, & Yan, 2012; Yan et al., 2009), while other studies have suggested that access to parafoveal phonology is enhanced in oral reading where overt articulation is required (Pan, Laubrock, & Yan, 2016; Pan, Yan, Laubrock, & Shu, 2019). Given the somewhat subordinate role of phonology in Chinese orthography, it is reasonable to expect a word's activated phonological form to be impoverished and relatively abstract, as maintained by the minimality hypothesis. Likely, word recognition with the Chinese script may bypass these detailed properties. The involvement of sub- and supra-phonemic[1] details during early stages of visual word recognition in alphabetic scripts thus cannot be assumed for Chinese. As an alternative theoretical hypothesis, if the use of detailed phonological features for visual word recognition is obligatory and universal, it should be considered as strong evidence for the early use of such properties independent of the script type, when such effects can be obtained in scripts less optimized for phonology such as Chinese.

To the best of our knowledge, so far, two studies have examined the nature of the phonological code during the silent reading of Chinese sentences (Luo, Yan, Yan, Zhou, & Inhoff, 2016; Yan, Luo, & Inhoff, 2014), but both studies focused on the

neutral-tone phenomenon. Mandarin Chinese is a tonal language with four full tones, each having a different pitch pattern to express lexical distinctions. It also includes a licensed tonal variant, generally referred to as neutral tone, involving reduced syllable articulation duration and intensity, as compared to its full-tone form (Cao, 1986; Lin & Yan, 1980). Comparing two types of two-character compound words with respect to their articulation variation in full-tone and neutral-tone conditions (e.g., *苍鹰*, /ts‘ɑŋ55 iŋ55/, Goshawk vs. *苍蝇*, /ts‘ɑŋ55 iŋ0/, Fly), these studies reported that native speakers of Mandarin Chinese spent less viewing time on neutral-tone words and made fewer refixations on them, than full-tone words. Electrophysiologically, neutral-tone words elicited smaller N100 (i.e., negative-going evoked potential that peaks around 100 milliseconds after the onset of stimulus) and anterior N250 amplitudes and larger N400 amplitude than full-tone words. These results suggest that Chinese readers can use articulation-specific sub- or supra-phonemic features rather than abstract phonological forms, as do readers of alphabetic scripts. The effects of lexically conditioned syllabic tone articulation were not clear-cut, however, because early (N100 and N250) and late (N400) ERP components appeared in diverging directions. It was hypothesized that the ease in early processing, due to reduced syllable articulation duration and intensity, was followed by late processing costs, due to a conflict between tonal features at the morpheme (lexical-tone) and whole word levels (neutral-tone) (Luo et al., 2016). In order to test this hypothesis, verification using other tonal variance phenomena without changes in articulation duration and intensity is needed.

In the present study, we take advantage of another tonal variance phenomenon in Mandarin Chinese, namely *tone sandhi*. Tone sandhi primarily occurs with Tone 3 (T3; a low dipping tone /213/), where a syllable carrying T3 alternates to Tone 2 (T2;

a mid rising tone /35/) when followed by another syllable carrying T3 (e.g., 宝马 /pau**213** ma213/ 'BMW' becomes like /pau**35** ma213/; Wang, 1967). It has been proposed that the base tone form and sandhi tone form may be listed under the same morpheme (Hsieh, 1976). In support of this account, Li and Chen (2015) found reduced mismatch negativity when a syllable carrying T3 was presented as the standard than when the same syllable carrying T2 was presented as the deviant in an oddball paradigm. The authors attributed this effect to the co-activation of the sandhi form of the T3 syllable when serving as the standard, thereby reducing the acoustic distance from the incoming deviant T2 stimuli. As for the production of tone sandhi, Zhang, Xia, and Peng (2015) presented two isolated spoken syllables with a pause in between to native Mandarin speakers and asked them to produce the two isolated syllables as a compound word (e.g., 赌-瘾 /tu213/-/in213/ 'gambling addiction' vs. 毒-瘾 /tu35/-/in213/ 'drug addiction'). The first T3 in T3+T3 words alternates to T2 due to Mandarin T3 sandhi in speech planning prior to articulation (cf. Politzer-Ahles & Zhang, in press). The authors found that the production of T3+T3 words elicited greater P2 amplitude than that of T2+T3 words. This is presumably due to the cognitive effort to phonologically re-encode the first T3 syllable into its sandhi form or the retrieval of its stored sandhi form in online production.

According to the minimality hypothesis, the articulation-defining properties of two types of target words, sandhi-tone and base-tone (i.e., non-sandhi) words should not influence viewing during the silent reading of Chinese sentences. However, if detailed phonological features are activated and utilized, we predict that fixation duration should be longer for sandhi-tone words than for base-tone words, because only the former type involves a conflict in tonal representations between morpheme and word levels. Additionally, another important factor to consider is word frequency.

The word frequency effect is a well-established phenomenon that lexical access is faster for more commonly used words than for those seen less often (see Brysbaert, Mandera, & Keuleers, 2018, for a recent review). According to the dual-process theory (Coltheart, 1978), when words are familiar to subjects, they tend to bypass phonology in favor of direct access, which leads to a general prediction that experimental effects due to phonological mediation should diminish or disappear (Van Orden, 1987). In support of this bypass hypothesis, for instance in a verification task in which participants were requested to judge the correctness of sentences, older children who were more familiar with the target words were less likely to accept meaningless sentences that are meaningful after phonological recoding (e.g., "*He ran **threw** the street*" for "*He ran **through** the street*"; Doctor & Coltheart, 1980). Instead of designing a categorized factor of word frequency, we allowed word frequency to vary naturally across target pairs and thus to cover a wide frequency range. During data analysis, target word frequency was treated as a continuous predictor (log-transformed and centered; Kliegl et al., 2006) to test whether activation of articulation-defining properties depended on word frequency. It is possible that reading cost, as indexed by prolonged fixation for sandhi-tone words compared to base-tone words, due to the conflict of tonal features at morpheme and whole-word levels, would be even greater for less frequent words.

In the present study, we aimed at exploring lexical processing of sandhi-tone Chinese words in a natural sentence-reading task using eye-tracking methodology. Specifically, we hypothesized that readers should spend more time processing sandhi-tone words than base-tone ones, because of a conflict of tonal features at morpheme and word levels for the former type, leading to longer viewing time. Additionally, however, this effect should be larger for infrequent words, because

frequent ones are more likely to be accessed directly, bypassing articulation-specific properties.

Method

Participants

Thirty-two readers (27 females), aged between 18 and 30 years (M = 22.8, SD = 2.6), participated in the eye-tracking experiment. Three additional groups of 30, 30 and 19 readers, who were from the same subject-pool, were recruited to obtain age-of-acquisition, imageability and contextual predictability estimates for the two groups of target words. All participants were native speakers of Chinese and were graduate and undergraduate students with normal or corrected-to-normal vision. Due to the purpose of the present study, we carefully chose native Mandarin Chinese speaking participants for the eye-tracking experiment. These participants were all born and grew up in Heilongjiang, Jilin, Liaoning, Shandong, or Hebei provinces, and had not lived in southern China for more than one year, before moving to Beijing for undergraduate or graduate studies. Therefore, their use of Mandarin Chinese was well-established. Experimental procedures were approved by the Human Research Ethics Committee of [name of institution masked for anonymity]. Participants gave their written informed consent prior to the experiment, which conformed to the tenets of the Declaration of Helsinki.

Material

A total of 146 pairs of two-character sandhi- and base-tone target words were selected for the experimental manipulation. The first characters of each word pair were always identical and had T3. The second characters in the sandhi-tone condition also had dipping tone, so that the two consecutive characters/syllables induced tone sandhi, with T3 alternated to T2 in production. In contrast, the second characters in

the base-tone condition varied in their tones, including high-level, rising or falling tones (tone 1, 2 and 4), so that the combination of the two characters/syllables did not lead to tone sandhi. Additionally, we carefully avoided using neutral tone words as target words. To rule out potential confounding of articulation-specific properties with other word properties, we matched the second characters of the target word pairs closely with regard to a number of visual and linguistic features. As shown in Table 1, there were no statistically significant differences in word frequency, second character frequency, age-of-acquisition (AoA), imageability (i.e., the extent to which a word can elicit a mental image; Paivio, Yuille, & Madigan, 1968), or visual complexity as indexed by numbers of strokes (all F-values < 1). Word AoA and imageability were controlled because they play critical roles in lexical access (e.g., Bird, Franklin & Howard, 2001). To make the target words fully comparable, we ensured that both members of each target word pair had very similar word frequencies as reflected by within-pair frequency difference (in occurrences per million; M = 2.7, SD = 3.4, max = 15.8). Although target word frequency was closely matched in an item-by-item manner, nevertheless, we allowed the frequency to cover a wide range (between 0.03 and 211.11). Therefore the influence of word frequency could be tested by including it as a predictor.

--TABLE 1--

In order to minimize possible influences of context and parafoveal processing, each target word in a pair was embedded in a sentence. The two sentences for each word pair had identical words preceding the targets and at least two words immediately following them. The last few words in the two sentences were allowed to vary between conditions, for fluent continuation. Critically, we carefully ensured that the two characters immediately preceding and following the target words did not have

T3, to avoid complicating the tone sandhi. Sentence contexts preceding the targets were written to be non-predictive for the two types of target words, to minimize any difference in top-down processing between conditions. A cloze test indicated that the targets were equally constrained (F-value < 1). All sentences contained 17 to 24 characters (M = 20.6, SD = 1.9), and the target words were never among the first three or the last three words. Each sentence was only presented once, with the two members of a sentence pair counterbalanced across participants. An example of sentences is shown in Figure 1. To summarize, using a between-item and within-subject manipulation, we designed a factor of word type (sandhi-tone versus base tone) and included word frequency as a continuous predictor.

--FIGURE 1--

Apparatus

The participants' eye-movements were recorded with an Eyelink 1000 system running at 1000 Hz. Each sentence was presented in a single line on a 21-inch ViewSonic G220f monitor (resolution: 1024-768 pixels; frame rate: 120 Hz) using Song font. The participants were seated 65 cm from the monitor, with their heads positioned by a forehead-and-chin rest. Each character had 36x36 pixels and subtended 1.2 degrees of visual angle. All recordings and calibrations were done monocularly, based on the right eye, and viewing was binocular.

Procedure

Gaze-positions were calibrated and validated with a 9-point grid (error<.5°). During each trial, a fixation-point appeared on the left side of the screen for drift check. A participant's gaze on the initial fixation-point initiated presentation of the next sentence, with its first character occupying the fixation-point. If the eye-tracker failed to detect the participant's eyes around the initial fixation-point, an extra

calibration was performed. The participants were instructed to read the sentences silently for comprehension, then fixate on a dot in the lower-right corner of the monitor, and finally press a joystick button to signal trial completion. Thirty-five sentences were followed randomly by easy yes-no comprehension questions, which were designed primarily to ensure participants' engagement with the reading task. The participants correctly answered 92% of these questions (*SD* = 4%).

Data Analysis

Fixations were determined with an algorithm for saccade detection (Engbert & Kliegl, 2003). Trials were removed in cases of participants' blinking or coughing, or tracker errors (n = 91, 2%). Data analyses were carried out based on the two-character target word region. Following standard procedures (e.g., Yan et al., 2014), target words with FFDs (first-fixation duration; durations of the first-fixations on a word irrespective of number of fixations) shorter than 60 ms or longer than 600 ms and GDs (gaze duration; accumulative durations of fixations during first-pass reading) longer than 800 ms were removed (n = 118, 3%). Arguably, processing effects that emerge within the first fixation on a word (i.e., FFD) are generally assumed to occur earlier than those that require a target to be refixated (i.e., GD), which is in turn earlier than those that appear in second-pass rereading such as TRT (total reading time; sum of all fixations on a word including regressive fixations). Therefore, analyses of these three duration measures are often used to obtain indexes for the time course of information processing (c.f., Inhoff, 1984).

Estimates were based on linear mixed models (LMMs) using the lme4 package (Version 1.1-21; Bates, Maechler, Bolker, & Walker, 2015) in the R environment (Version 3.5.3; R Development Core Team, 2019). We included in LMMs the word type (sandhi-tone versus base-tone words), a continuous predictor of word frequency

(log-transformed and centered) and their interaction as fixed-effects, as well as intercepts for subjects and items, and slopes (word type and frequency) for subjects as random-effects, given successful model convergence, which was achieved by dropping the smallest variance components (Bates, Kliegl, Vasishth, & Baayen, 2015). Fixation-durations were log-transformed for LMM and analyses for untransformed and log-transformed durations yielded the same pattern of significance. We applied two-tailed criterion ($|t| > 1.96$), corresponding to a 5 % error criterion for significance. *P*-values were reported using the lmerTest package (Version 3.1-0; Kuznetsova, Brockhoff, & Christensen, 2017). Complete model outputs are reported in the Appendix.

Results

Overall, when frequent target words were compared to infrequent target words, for the former the readers were more likely to skip words (*b*=.272, *SE*=.090, *z*=3.02, *p*=.003), made fewer refixations (*b*=-.263, *SE*=.077, *z*=-3.41, *p*=.001), and had shorter first-pass and second-pass reading time (FFD: *b*=-.018, *SE*=.009, *t*=-1.99, *p*=.048; GD: *b*=-.041, *SE*=.015, *t*=-2.79, *p*=.006; TRT: *b*=-.056, *SE*=.017, *t*=-3.35, *p*=.001). These word frequency effects on skipping, refixation, and fixation duration are canonical effects in sentence-reading experiments, which were reported consistently across a large number of previous studies (see Kliegl et al., 2006, for a review). These effects are not directly relevant to the research interest of the present study and thus are not discussed; however, their existence does suggest the reliability of the present data set.

--TABLE 2--

More importantly, we found a significant interaction between word type and word frequency in all three fixation duration measures (FFD: *b*=-.044, *SE*=.017, *t*=-2.64, *p*=.009; GD: *b*=-.060, *SE*=.024, *t*=-2.49, *p*=.014; TRT: *b*=-.061, *SE*=.025,

t=-2.39, *p*=.018; Figure 2). These effects were decomposed further, by dichotomizing word frequency into two approximately equal-sized groups (because the two words in each target pair had very similar word frequencies, as mentioned in the Method section, this indicated that different word types were comparable in terms of frequency). This indicated that readers fixated on sandhi-tone words for significantly longer than on base-tone words only when they were infrequent in terms of GD and TRT (Table 2; GD: *b*=.047, *SE*=.022, *t*=2.19, *p*=.029; TRT: *b*=.050, *SE*=.022, *t*=2.25, *p*=.025). The effect was marginally significant in FFD (*b*=.029, *SE*=.016, *t*=1.78, *p*=.076), and there was no significant difference in fixation durations when they were frequent (FFD: *b*=-.013, *SE*=.014, *t*=-0.91, *p*=.365; GD: *b*=-.013, *SE*=.019, *t*=-0.68, *p*=.497; TRT: *b*=-.023, *SE*=.020, *t*=-1.15, *p*=.249). None of the remaining main effects was significant (*p*-values>.1).

--FIGURE 2--

Discussion

In the present study we investigated how articulation-specific properties of visual words influenced oculomotor behaviors during the silent reading of Chinese sentences, as a test to determine the use of spoken language properties. The key manipulation was the use of target word pairs with specific articulation properties, one of which bore sandhi-tone alternation while the other did not. We additionally controlled a number of potentially confounding factors of visual and linguistic properties in the two experimental conditions to ensure a clean pattern of results, but allowed word frequency to vary; this was included as a predictor in the analysis. Recordings of eye movements revealed that native Mandarin Chinese speakers spent more time reading sandhi- than base-tone infrequent target words, and that the two types of words did not differ when they were frequent. Importantly, the effects of word type emerged in a

relatively early lexical processing stage during visual word recognition.

In previous research on alphabetic scripts, there is a vast amount of evidence for the effects of word-specific articulation properties on visual word recognition. For instance, vowel articulation duration can influence not only reaction time in isolated visual word recognition (Abramson & Goldinger, 1997; Lukatela et al., 2004), but also word viewing duration in silent sentence reading (Ashby & Clifton, 2005; Huestegge, 2010). Additionally, neurophysiological evidence from ERP recordings showed that detailed phonological features can influence early processing, as reflected by less negative early components when sub- and supra-phonemic properties of prime and target words are consistent than when they are inconsistent (Ashby, Sanders, & Kingston, 2009; Wheat, Cornelissen, Frost, & Hansen, 2010). In many alphabetic languages, vowel duration is linguistically informative: changes in vowel duration can be used to distinguish similar lexical alternatives (e.g., *pitch* vs. *peach*). It is thus not surprising to see that alphabetic readers are sensitive to vowel duration information during silent reading.

The theoretical significance of articulation-specific effects in Chinese in the present study is that it puts the minimality hypothesis to test through the use of a logographic script that was not designed to represent spoken language properties at the phoneme level. In contrast to alphabetic languages, standard Chinese words are not distinguishable by syllable duration, that is, a neutral-tone alternation denotes the same word as the original full-tone form. Yan et al. (2014, Exp. 1) reported the first evidence for the influence of articulation-specific tonal information on the silent reading of Chinese sentences. They found that native Mandarin Chinese (i.e., northern Chinese dialect) speakers processed neutral-tone target words more briefly than full-tone ones. Arguably, it is likely that visual word identification involves the

generation of articulatory gestures, and simulated articulation of a short syllable took less time. Equivalent processing time for the two types of words among non-native Mandarin Chinese (i.e., southern Chinese dialects, including Min, Wu, Yue/Cantonese, etc.) speakers implies that the differences between neutral- and full-tone words should not be attributed to visual or linguistic properties. The joint experimental evidence of reduced parafoveal processing of upcoming words (Yan et al., 2014, Exp. 2) and a larger N400 ERP component (Luo et al., 2016, Exp. 1) for neutral than full-tone Chinese words among native Mandarin readers suggests that the reduction in viewing duration reported by Yan et al. (2014, Exp. 1) does not imply an ease in lexical processing of neutral-tone words. In addition to a more negative N400 peak amplitude, Luo et al. (2016) also reported smaller N100 and N250, for neutral-tone words than for full-tone ones. It can be argued that early benefits for neutral-tone words (N100 and N250), due to a faster simulated articulation during the convergence of lexical forms, were followed by a subsequent processing cost (N400), due to a conflict between lexical (full) and neutral tone features that occurred during a late stage of word processing. Taken together, differences between these two types of words with Mandarin Chinese readers can thus be attributed to dialect-specific articulation properties, providing novel evidence for the use of detailed phonological features during visual word identification, even when they did not distinguish between lexical alternatives.

The present study takes the issue of whether detailed phonological features can be accessed early during visual word identification in Chinese silent reading one step further, by examining words involving T3 sandhi versus those that did not. A factor that was not tested previously in this issue was word frequency. The word frequency effect in itself was not of particular interest to us. Instead, we were theoretically

motivated to test an interaction between word type and word frequency. This is because frequent and infrequent words are known to have different lexical access routes (Coltheart, 1978): readers are more likely to access lexical information directly for more frequent words, diminishing the role of phonological mediation. Indeed, phonological properties often have little impact on naming high frequency words and their effects tend to be more robust on low frequency words (e.g., Seidenberg, Waters, Barnes, & Tanenhaus, 1984; Seidenberg 1985). This is of particular importance in the context of the minimality hypothesis. On the one hand, we found that native Mandarin Chinese readers spent more time, in both early and late eye-movement measures, on sandhi- than on base-tone targets when they were infrequent words, which is difficult to account for by the minimality hypothesis. We argue that such a prolongation in viewing time is a consequence of a conflict for the sandhi-tone words between their base tones represented at a morpheme/character level and surface realizations represented at a word level (i.e., sandhi tones). In contrast, phonological representations of the base-tone words are consistent at different levels. Note that FFD typically reflects early processing effects (Inhoff, 1984). As such, the early involvement of sandhi variance in the present study mirrors previous findings of sub- and supra-phonemic features during early stages of visual word recognition in both alphabetic (e.g., Ashby & Clifton, 2005; Ashby et al., 2006) and Chinese scripts (e.g., Yan et al., 2014), suggesting that the early use of articulation-specific information during visual word recognition is likely a language-universal phenomenon for accessing lexical knowledge among readers of various writing systems.

On the other hand, processing was equivalent for the two types of words when they were frequent, indicating that readers may have followed a more direct lexical access route bypassing detailed phonological features. Alternatively, it is also possible

that these detailed phonological features were not yet activated by the time the lexical access of the frequent target words was completed. In either case, the results do not support an early use of phonological features in silent reading for frequent words. Together, rather than simply being a challenge to or a support for the minimality hypothesis, our results make an interesting complement and suggest that it is possible that the early use of detailed phonological features in visual word processing depends on other linguistic properties such as word frequency. These results, however, are compatible with the spirit of the well-known dual route model of reading (Coltheart, Curtis, Atkins, & Haller, 1993), which suggests that word reading can be achieved through two separte routes. Relying on a direct connection between the graphemic form and lexical representation of words, the lexical route is used for frequent words, potentially bypassing detailed phonological features. On the other hand, when encountering infrequent words, readers are known to achieve lexical access by identifying the word's phonological constituents following grapheme-to-phoneme rules. Such indirect access is termed the sublexical route and allows detailed phonological features to build up. It is also important to keep in mind that the present findings are based on Chinese *adults*. It is of great theoretical interest and importance for future research to determine the usage of articulation-specific properties among Chinese *children*, who are still acquiring proficiency in the language and are known to have a different lexical processing priority (Feng et al., 2001; Zhou, Shu, Miller, & Yan, 2018).

To conclude, the present study provides solid evidence for language universality in the early use of detailed phonological features during visual word identification in silent sentence reading, a phenomenon that the minimality hypothesis may not be able to account for fully. Together with previous studies of English, it is likely that readers

make use of articulation-specific properties irrespective of the nature of the script, even for logographic scripts that are not designed to capture sound- and articulation-related word properties, such as Chinese.

Footnote

1. There is considerable debate on whether lexical tone is phonemic or not. Lexical tone is supra-segmental and many would also argue that it is phonemic, in the sense that a change in lexical tone modifies the lexical meaning in a systematic way. For better reconciliation with previous studies, we chose to use the term phonemic (or sub-phonemic) in the present study.

References

Abramson, M., & Goldinger, S.D. (1997). What the reader's eye tells the mind's ear: Silent reading activates inner speech. *Perception & Psychophysics, 59,* 1059-1068.

Ashby, J. (2010). Phonology is fundamental in skilled reading: Evidence from ERPs. *Psychonomic Bulletin & Review, 17*, 95-100. Doi:10.3758/PBR.17.1.95

Ashby, J. & Clifton, C. Jr. (2005). The prosodic property of lexical stress affects eye movements in silent reading: Evidence from eye movements. *Cognition, 96,* B89-B100.

Ashby, J., & Rayner, K. (2004). Representing syllable information during silent reading: evidence from eye movements. *Language and Cognitive Processes, 19,* 391-426.

Ashby, J., Sanders, L.D., & Kingston, J. (2009). Skilled readers begin processing subphonemic feature by 80 ms during visual word recognition. *Biological Psychology, 80,* 84-94.

Ashby, J., Treiman, R., Kessler, B., & Rayner, K. (2006). Vowel processing during silent reading: Evidence from eye movements. *Journal of Experimental Psychology: Learning, Memory, & Cognition, 32,* 416-424.

Bates, D., Kliegl, R., Vasishth, S., & Baayen, R. H. (2015). Parsimonious mixed models. Available from arXiv:1506.04967 (stat.ME).

Bates, D., Maechler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software, 67*,1-48.

Bird, H., Franklin, S., & Howard, D. (2001). Age of acquisition and imageability ratings for a large set of words, including verbs and function words. *Behavior Research Methods, Instruments, and Computers, 33*, 73–79.

Doi:10.3758/BF03195349

Breen, M., & Clifton, C. Jr. (2011). Stress matters: Effects of anticipated lexical stress on silent reading. *Journal of Memory and Language, 64*, 153–170. Doi:10.1016/j.jml.2010.11.001.

Brysbaert, M., Mandera, P., & Keuleers, E. (2018). The word frequency effect in word processing: an updated review. *Current directions in psychological science, 27*(1), 45–50.

Cao, J. (1986). Putonghua qingsheng yinjie texing fenxi. *Applied Acoustics* (in Chinese)*, 5 (4),* 1-6.

Cai, Q., & Brysbaert, M. (2010). SUBTLEX-CH: Chinese word and character frequencies based on film subtitles. *PLoS One, 5*(6), e10729. doi:10.1371/journal.pone.0010729

Chen, H.-C., & Shu, H. (2001). Lexical activation during the recognition of Chinese characters: evidence against early phonological activation. *Psychonomic Bulletin & Review, 8 (3)*, 511-518.

Coltheart, M. (1978). Lexical access in simple reading tasks. In G. Underwood (Ed.), *Strategies in information processing* (pp. 151-216). London: Academic Press.

Coltheart, M., Curtis, B., Atkins, P., & Haller, M. (1993). Models of reading aloud: Dual-route and parallel-distributed-processing approaches. *Psychological Review, 100*, 589–608.

Doctor, E.A. & Coltheart, M. (1980). Children's use of phonological encoding when reading for meaning. *Memory & Cognition, 8*, 195–209. Doi:10.3758/BF03197607

Engbert, R., & Kliegl, R. (2003). Microsaccades uncover the orientation of covert attention. *Vision Research, 43*, 1035-1045.

Feng, G., Miller, K., Shu, H., & Zhang, H. (2001). Rowed to recovery: The use of phonological and orthographic information in reading Chinese and English. *Journal of Experimental Psychology: Learning, Memory, & Cognition, 27,* 1079-1100.

Frost, R. (1998). Toward a strong phonological theory of visual word recognition: True issues and false trails. *Psychological Bulletin, 123,* 71-99.

Frost, R. (2012). Towards a universal model of reading. *Behavioral and Brain Sciences, 35,* 263-279. doi:10.1017/S0140525X11001841

Hohenstein, S., & Kliegl, R. (2015). *remef: Remove partial effects. [Computer software]* Retrieved from https://github.com/hohenstein/remef/

Hoosain, R. (1991). *Psycholinguistic implications for linguistic relativity: A case study of Chinese*. Hillsdale, NJ: LEA.

Hsieh, H. I. (1976). On the unreality of some phonological rules. *Lingua, 38*(1), 1–19. doi:10.1016/0024-3841(76)90038-3

Huestegge, L. (2010). Effects of vowel length on gaze durations in silent and oral reading. *Journal of Eye Movement Research, 3(5),* 1-18.

Inhoff, A.W. (1984). Two stages of word processing during eye fixations in the reading of prose. *Journal of Verbal Learning and Verbal Behavior, 23*, 612–624. Doi:10.1016/S0022-5371(84)90382-7

Inhoff, A.W., Connine, C., & Radach, R. (2002). A contingent speech technique in eye movement research on reading. *Behavior Research Methods, Instruments, & Computers, 34*, 471–480. Doi:10.3758/BF03195476

Kliegl, R., Masson, M.E.J., & Richter, E.M. (2010). A linear mixed model analysis of masked repetition priming. *Visual Cognition, 18*, 655-681.

Kliegl, R., Nuthmann, A., & Engbert, R. (2006). Tracking the mind during reading:

The influence of past, present, and future words on fixation durations. *Journal of Experimental Psychology: General*, *135,*13-35.

Kuznetsova, A., Brockhoff, P.B., & Christensen, R.H.B. (2017). lmerTest Package: Tests in Linear Mixed Effects Models. *Journal of Statistical Software, 82 (13),* 1-26. doi:10.18637/jss.v082.i13

Li, X., & Chen, Y. (2015). Representation and processing of lexical tone and tonal variants: Evidence from the mismatch negativity. *PLoS ONE, 10*(12), 1–24. doi:10.1371/journal.pone.0143097

Lin, M., & Yan, J. (1980). Beijinghua qingsheng de shengxue xingzhi. *Dialect* (in Chinese)*, 3,* 166-178.

Liu, W., Inhoff, A.W., Ye, Y., & Wu, C. (2002). Use of parafoveally visible characters during the reading of Chinese sentences. *Journal of Experimental Psychology: Human Perception and Performance, 28,* 1213–1227.

Lukatela, G., Eaton, T., Sabadini, L., & Turvey, M.T. (2004). Vowel duration affects visual word identification: Evidence that the mediating phonology is phonetically informed. *Journal of Experimental Psychology: Human Perception & Performance, 30,* 151-162.

Luo, Y., Yan, M., Yan, S., Zhou, X., & Inhoff, A.W. (2016). Syllabic tone articulation influences the identification and use of words during Chinese sentence reading: Evidence from ERP and eye movement recordings. *Cognitive, Affective, & Behavioral Neuroscience, 16*(1), 72-92. doi:10.3758/s13415-015-0368-1

Paivio, A., Yuille, J.C., & Madigan, S.A. (1968). Concreteness, imagery, and meaningfulness values for 925 nouns. *Journal of Experimental Psychology, 76*, 1–25. doi:10.1037/h0025327

Pan, J., Laubrock, J., & Yan, M. (2016). Parafoveal processing in silent and oral

reading: Reading mode influences the relative weighting of phonological and semantic information in Chinese. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 42,* 1257-1273. doi:10.1037/xlm0000242

Pan, J., Yan, M., Laubrock, J., & Shu, H. (2019). Lexical and sublexical phonological effects in Chinese silent and oral reading. *Scientific Studies of Reading*. doi:10.1080/10888438.2019.1583232

Perfetti, C.A., Liu, Y., & Tan, L.H. (2005). The lexical constituency model: Some implications of research on Chinese for general theories of reading. *Psychological Review, 112(1)*, 43-59.

Politzer-Ahles, S. and Zhang, J. (in press) 'Evidence for the role of tone sandhi in Mandarin speech production'. *Journal of Chinese Linguistics*.

R Development Core Team (2019). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL: http://www.R-project.org/.

Rayner, K. (2009). Eye movements and attention in reading, scene perception, and visual search. *Quarterly Journal of Experimental Psychology, 62*(8), 1457-1506. doi:10.1080/17470210902816461

Reilly, R., & Radach, R. (2012). The dynamics of reading in non-Roman writing systems. *Reading and Writing, 25,* 935-950. Doi:10.1007/s11145-012-9369-4

Seidenberg, M. S. (1985). Constraining models of word recognition. *Cognition, 20*(2), 169-190.

Seidenberg, M. S., Waters, G. S., Barnes, M. A., & Tanenhaus, M. K. (1984). When does irregular spelling or pronunciation influence word recognition? *Journal of Verbal Learning and Verbal Behavior, 23*(3), 383-404.

Shu, H., Chen, X., Anderson, R.C., Wu, N., & Xuan, Y. (2003). Properties of school

Chinese: Implicatins for learning to read. *Child Development, 74,* 27-47. doi:10.1111/1467-8624.00519

Tsai, J.-L., Kliegl, R., & Yan, M. (2012). Parafoveal semantic information extraction in Traditional Chinese reading. *Acta Psychologica, 141,* 17-23. doi:10.1016/j.actpsy.2012.06.004

Tsai, J.-L., Lee, C.-Y., Tzeng, O.J.L., Hung, D.L., & Yen, N.-S. (2004). Use of phonological codes for Chinese characters: evidence from processing of parafoveal preview when reading sentences. *Brain and Language, 91,* 235–244.

Van Orden, G.C. (1987). A rows is a rose: Spelling, sound, and reading. *Memory & Cognition, 15, 181–198*. doi:10.3758/BF03197716

Wang, S.-Y. (1967). Phonological features of tone. *International Journal of American Linguistics, 33*, 93-105. doi:10.2307/1263953

Wickham, H. (2009). *ggplot2: Elegant graphics for data analysis.* Dordrecht, Netherlands: Springer.

Yan, M., Richter, E.M., Shu, H., & Kliegl, R. (2009). Chinese readers extract semantic information from parafoveal words during reading. *Psychonomic Bulletin & Review, 16,* 561-566. doi:10.3758/PBR.16.3.561

Yan, M., & Sommer, W. (2015). Parafoveal-on-foveal effects of emotional word semantics in reading Chinese sentences: Evidence from eye movements. *Journal of Experimental Psychology: Learning, Memory and Cognition, 41(4),* 1237-1243. doi:10.1037/xlm0000095

Yan, M., Luo, Y., & Inhoff, A.W. (2014). Syllable articulation influences foveal and parafoveal processing of words during the silent reading of Chinese sentences. *Journal of Memory and Language, 75,* 93-103. doi:10.1016/j.jml.2014.05.007

Yang, J., Rayner, K., Li, N., & Wang, S. (2012). Is preview benefit from word n + 2 a

common effect in reading Chinese? Evidence from eye movements. *Reading and Writing, 25,* 1079-1091. doi: 10.1007/s11145-010-9282-7

Yang, J., Wang, S., Tong, X., & Rayner, K. (2012). Semantic and plausibility effects on preview benefit during eye fixations in Chinese reading. *Reading and Writing, 25,* 1031-1052. doi: 10.1007/s11145-010-9281-8

Zhang, C., Xia, Q., & Peng, G. (2015). Mandarin third tone sandhi requires more effortful phonological encoding in speech production: Evidence from an ERP study. *Journal of Neurolinguistics, 33*, 149-162. doi:10.1016/j.jneuroling.2014.07.002

Zhou, W., Wang, A., Shu, H., Kliegl, R., & Yan, M. (2018). Word segmentation by alternating colors facilitates eye guidance in Chinese reading. *Memory & Cognition, 46*(5), 729-740. doi: 10.3758/s13421-018-0797-5

Zhou, X., & Marslen-Wilson, W. (2000). The relative time course of semantic and phonological activation in reading Chinese. *Journal of Experimental Psychology: Learning, Memory and Cognition, 26,* 1245–1265. doi:10.1037/0278-7393.26.5.1245

Zhou, Y. (1980). *Pronunciations of phonetics within compound characters*. Changchun, China: Jilin People's Publisher. (Original work in Chinese)

Table 1

Word Properties

	Sandhi Tone	Base Tone
Example word	赶紧	赶快
Meaning	Immediately	immediately
Pronunciation	Gan3-Jing3	Gan3-Kuai4
Log-word freq.	0.43 (0.71)	0.42 (0.72)
N. of strokes	16.6 (4.7)	16.5 (4.7)
Log-2nd char. freq.	0.85 (0.76)	0.79 (0.75)
AoA	8.3 (1.2)	8.3 (1.2)
Imageability	4.0 (0.6)	3.9 (0.7)
Predictability	2.3% (8.4%)	1.6% (5.1%)

Means (and standard deviations, in parentheses) of word frequency and second character frequency (in log-transformed number of occurrence per million; Cai & Brysbaert, 2010), number of strokes, age of acquisition (AoA), imageability and predictability of the target words.

Table 2

Fixation Measures

	High Frequency (N = 1865)		Low Frequency (N = 1986)	
	Sandhi Tone	Base Tone	Sandhi Tone	Base Tone
SP	14.0 (15.9)	14.4 (13.6)	12.1 (11.2)	12.7 (12.7)
RP	16.8 (11.8)	18.3 (13.3)	22.1 (19.0)	17.4 (15.2)
FFD	249 (39)	250 (31)	259 (32)	249 (32)
GD	288 (48)	293 (43)	315 (59)	292 (51)
TRT	301 (52)	308 (54)	342 (76)	315 (64)

Note. Means (and standard deviations, in parentheses) of skipping probabilities (SP; %), refixation probabilities (RP; %), first-fixation duration (FFD; ms), gaze duration (GD; ms) and total reading time (TRT; ms) of the target words. Values are computed across participants' means.

Figure Captions

Figure 1. An example pair of target words, 赶紧 (Gan3-Jing3, sandhi tone) and 赶快 (Gan3-Kuai4, base tone) with identical first character (i.e., 赶), embedded into an identical sentence frame. The target words are highlighted with gray background, only for the purpose of illustration but not during the experiment. The sentence is translated as: *The local people are very eager to immediately end such a tumultuous day*.

Figure 2. Partial effects (i.e., LMM estimates after statistical control for between-subject and between-item differences) on first-fixation duration as a function of word type and word frequency, generated using remef (version 0.6.10; Hohenstein & Kliegl, 2015) and ggplot2 packages (version 2.1.0; Wickham, 2009).

Figure 1

Sandhi tone condition

当地百姓非常希望赶紧结束这样颠沛流离的日子。

Base tone condition

当地百姓非常希望赶快结束这样颠沛流离的日子。

Word-to-word translation:

Sandhi tone condition

当地 百姓 非常 希望 赶紧 结束 这样 颠沛流离 的 日子。

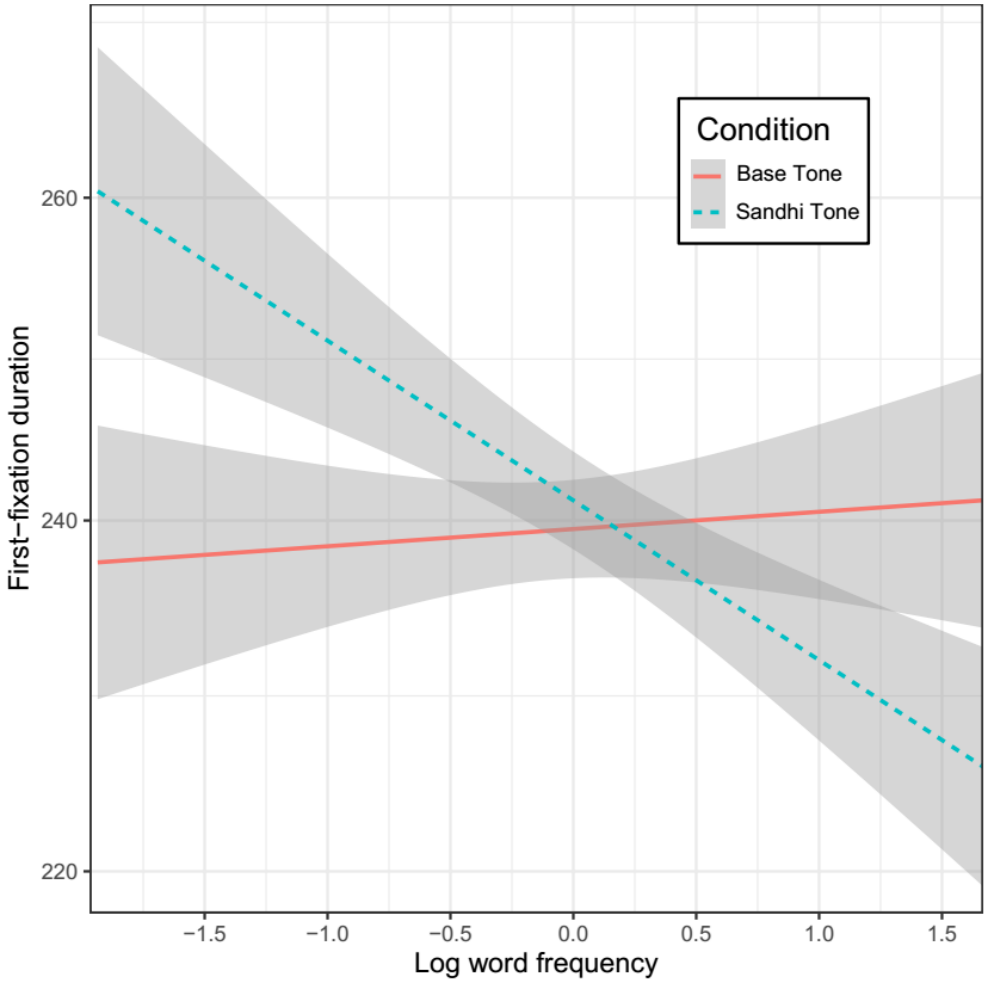
Local residents very eager speedily end this tumultuous of day.

Base tone condition

当地 百姓 非常 希望 赶快 结束 这样 颠沛流离 的 日子。

Local residents very eager quickly end this tumultuous of day.

Figure 2


Condition
Base Tone
Sandhi Tone
First−fixation duration
260
240
220
−1.5
−1.0
−0.5
0.0
0.5
1.0
1.5
Log word frequency

Appendix

Linear mixed model estimates for durations

	First fixation duration			Gaze duration			Total reading time		
Fixed effect	Est	SE	p-value	Est	SE	p-value	Est	SE	p-value
(Intercept)	5.482	0.021	<.001 *	5.606	0.025	<.001 *	5.653	0.027	<.001 *
Word Type	0.007	0.009	.447	0.004	0.014	.769	-0.003	0.014	.841
Word Frequency	-0.018	0.009	.048 *	-0.041	0.015	.006 *	-0.056	0.017	.001 *
WT x WF	-0.044	0.017	.009 *	-0.060	0.024	.014 *	-0.061	0.025	.018 *
Variance component	SD			SD			SD		
Subject – WT	0.031			0.078			0.086		
Subject – (Intercept)	0.055			0.105			0.125		
Item – (Intercept)	0.113			0.125			0.139		
Residual	0.275			0.366			0.380		