International Journal of

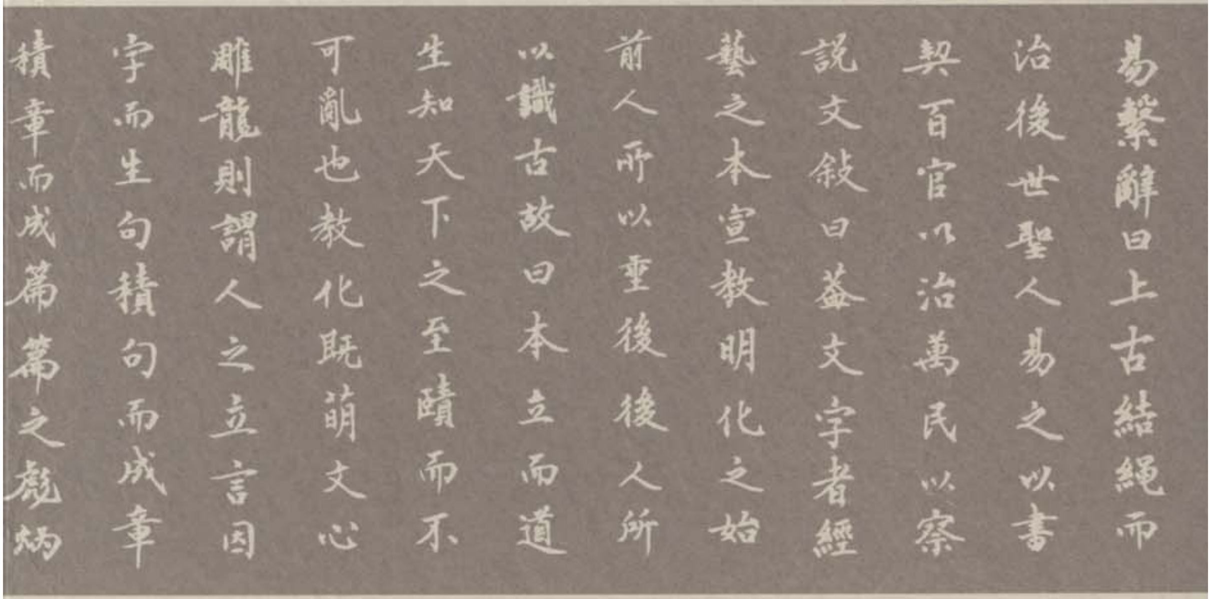
Computational Linguistics & Chinese Language Processing

中文計算語言學期刊

A Publication of the Association for Computational Linguistics and Chinese Language Processing

This journal is included in THCI, Linguistics Abstracts, and ACL Anthology.

Special Issue on "Processing Lexical Tones in Natural Speech"
Guest Editor: Shu-Chuan Tseng



Vol.18 No.3 September 2013 ISSN: 1027-376X

International Journal of Computational Linguistics & Chinese Language Processing

International Journal of

Computational Linguistics & Chinese Language Processing

Aims and Scope

International Journal of Computational Linguistics and Chinese Language Processing (IJCLCLP) is an international journal published by the Association for Computational Linguistics and Chinese Language Processing (ACLCLP). This journal was founded in August 1996 and is published four issues per year since 2005. This journal covers all aspects related to computational linguistics and speech/text processing of all natural languages. Possible topics for manuscript submitted to the journal include, but are not limited to:

- Computational Linguistics
- Natural Language Processing
- Machine Translation
- Language Generation
- Language Learning
- Speech Analysis/Synthesis
- Speech Recognition/Understanding
- Spoken Dialog Systems
- Information Retrieval and Extraction
- Web Information Extraction/Mining
- Corpus Linguistics
- Multilingual/Cross-lingual Language Processing

Membership & Subscriptions

If you are interested in joining ACLCLP, please see appendix for further information.

Copyright

Cover

Calligraphy by Professor Ching-Chun Hsieh, founding president of ACLCLP
Text excerpted and compiled from ancient Chinese classics, dating back to 700 B.C.
This calligraphy honors the interaction and influence between text and language

International Journal of
Computational Linguistics and Chinese Language Processing
vol. 18, no. 3, September 2013

Contents

Special Issue Articles:

Processing Lexical Tones in Natural Speech

The Association for Computational Linguistics and Chinese Language Processing

Foreword

Phonologically, lexical tones in Chinese are abstract categories. In everyday communication, however, the acoustic-sensory properties in the speech input accessible to us may not always correspond to the canonical forms of tones. This special issue is aimed at attracting attention from researchers in interdisciplinary fields concerning the issue of how lexical tones are produced, perceived, and processed in continuous speech. I am happy that the papers submitted present a wide range of research results, adopting phonological, corpus-based, and experimental approaches. Tone is an assigned categorical label. It would be meaningless to produce or perceive a tone without semantic context. For Chinese, the characters invoke meaning with tone information. The first paper, "*Implicit Priming Effects in Chinese Word Recall: The Role of Orthography and Tones in the Mental Lexicon,*" specifically addresses the issue of orthography and the representation of tone category in the mental lexicon by adopting psycholinguistic experiments. The second paper, "*Understanding Mandarin Prosody: Tonal and Contextual Variations in Spontaneous Conversation,*" studies pitch variability of tones in read and spontaneous speech. Pitch contours of tones in continuous speech, in spite of having great diversity, show systematic shapes dependent on the tonal context. "*Acoustic Correlates of Contrastive Stress in Compound Words versus Verbal Phrase in Mandarin Chinese*" explores the stress patterns in compound words and verbal phrases by examining acoustic-prosodic features of tonal syllables. Furthermore, "*Non-segmental Cues for Syllable Perception: the Role of Local Tonal f0 and Global Speech Rate in Syllabification*" shows not only the local F0 cue but also the global speech rate affect how tonal syllables in Chinese sentences are perceived. The final paper, "*Tones of Reduced T1-T4 Mandarin Disyllables,*" presents results of two tone category identification experiments on fully pronounced and reduced disyllables whose stimuli were designed based on a corpus-based speech production study.

With this special issue, I hope to demonstrate the importance of "across-team works" for extending the research scope of language and speech processing to a realistic, communicative context. Finally, I would like to thank all of the authors who submitted their works to this special issue as well as the reviewers who kindly gave valuable comments and suggestions to each of the submissions.

Shu-Chuan Tseng

Guest Editor

Institute of Linguistics, Academia Sinica

Computational Linguistics and Chinese Language Processing
Vol. 18, No. 3, September 2013, pp. 1-20

Implicit Priming Effects in Chinese Word Recall: The Role of Orthography and Tones in the Mental Lexicon

Nian Liu*

Abstract

This paper explores the relative contributions made by orthography, syllabic segment, and lexical tone in the word recognition and retrieval process. It also challenges recent assumptions regarding the role of orthography and tones in mental lexicon architecture. Using an implicit priming paradigm, a word recognition experiment was conducted with native speakers of two tonal languages, Chinese and Vietnamese, that use a logographic orthography and a phonetically-based orthography, respectively. Contradicting prior findings, response time differences indicate that orthography plays a crucial role in the word recognition process, a finding that has implications for Chinese language teaching.

Keywords: Orthography, Tones, Word Recall, Chinese Teaching

1. Introduction

Chinese is one of the few existing languages that use a logographic writing system. In addition, it uses four lexical tones to distinguish otherwise identical words. Neither of these features is shared by Indo-European languages, which makes teaching Chinese to native speakers of these languages a difficult task for teachers and a painstaking process for learners.

There has been abundant literature discussing efficient ways of teaching all aspects of Chinese as a second language, including tonal pronunciation (Miracle, 1989; McGinnis, 1997; Wang *et al*., 1999; Li, 2007) and character writing (Everson, 1998), as well as grammar points. Many of these pedagogical methodologies certainly pave the way to learning Chinese, yet, even with intensive training and practicing of the target language, Chinese second language (L2) learners still have a difficult time achieving native-like language proficiency. Therefore, the fundamental issue in improving L2 Chinese speakers' proficiency lies not only in teaching methods, but also in the language knowledge within each of the learners.

* Department of Modern Languages, Literatures and Linguistics, University of Oklahoma, U.S.A.
E-mail: nian.liu@ou.edu

The major difficulty in learning Chinese as a second language usually concerns the acquisition of tones and characters. How can tones be recognized and produced by L2 Chinese speakers if these phonological features are not used in distinguishing meaning in their native languages? Also, how are characters to be read by L2 Chinese learners if the writing symbols are, for the most part, not logically associated with the phonology, as the writing symbols are in an alphabetic language? How is each word represented in the mental lexicon of an L2 speaker? Do they use the same strategies for processing and producing words as native speakers?

In addressing this query, this paper explores the role of orthography and tones in the word recall process in terms of the way word entries are stored in the mental lexicon and how they are retrieved in word recognition and production.

Considerable research on lexicon architecture has been done on Indo-European languages, with English and Dutch commanding most of the attention. Nevertheless, Chen, Chen, and Dell (2002) examined the relative contribution of three potential structural components in Mandarin Chinese to investigate lexicon architecture. They used an implicit priming paradigm, in which exposure to a series of syllables (embedded in real words) that share a set of features can serve to prime other words with similar syllables and facilitate their recall, as measured by response time (RT). To take an English example, after being exposed acoustically to a series of words, like *season*, *ceiling*, and *secret*, people identify and name the word *Caesar* more quickly, because all of these words share the same onset, which creates a priming effect.

In Mandarin's orthography, individual syllables are represented by symbolic characters in logographic form. In their experiment, Chen *et al.* (2002) examined the implicit priming effect of different components of Mandarin Chinese. The first characters of the response words were controlled-response words in a set under the homogeneous condition shared some features (*i.e.*, syllable + tone + orthography in the 3-way syllable Match condition and syllable + tone in the 2-way syllable Match condition), whereas, in the heterogeneous condition, which served as the baseline, none of the three components were shared. They found that the word recall response time was shortened by 53 milliseconds over baseline times (659 ms-606 ms) for Mandarin words that shared phonological segments (音节), lexical tones (音调), and written characters (汉字) (3-way syllable match). Chen *et al.* then demonstrated that priming from syllables that shared segments and tones, but not characters (2-way syllable match) resulted in 46 milliseconds of response time improvement over the baseline (664 ms-618 ms). They concluded that the combination of segment and tone is a critical unit in lexical representation; therefore, it is a vital component in the lexicon architecture for tonal languages. Based on a series of experiments that using the same paradigm but different materials in combining segment, tone and orthography, they also argued that, due to the relatively small difference between RTs for the 3-way match versus the 2-way match, orthographic characters should not be considered a significant architectural component.

The current study is motivated by two objectives. The first is to see if, because of its logographic orthography, Chinese is unique among tonal languages in how priming effects function in it. The second is to re-examine the minor role to which Chen *et al.* relegated orthography in lexicon architecture. From these objectives, two researchable questions emerge. (1) Is the apparently significant combination of segment and supra-segmental tone a common unit, equally important in the word recognition process for all tonal languages? Or does the process differ in Chinese due to its logographic writing system? (2) To what extent does orthography facilitate the word recognition task in a tonal language whose written characters do not actually add any information that could be used during the retrieval process? To answer these questions, it is useful to introduce another language to serve as a baseline. Vietnamese serves the inquiry purposes well in that, like Mandarin, it is a tonal language, but, unlike Mandarin, it uses a phonetically-based orthography. Therefore, it is used as a control language in this experiment, which is designed to partially replicate Chen *et al.*'s design and method using both Mandarin and Vietnamese.

2. Experiment

The study consists of two phases designed as separate experiments: a Mandarin implicit priming experiment and a Vietnamese implicit priming experiment. The Mandarin participants each saw four sets of words in four different permutations of a 2 x 2 factor design, one set for each of the four lexical tones in Mandarin. The Vietnamese participants each saw six sets of words in two different conditions, one set for each of the six lexical tones in Vietnamese. The following sections on procedures and materials give a detailed description of these conditions.

2.1 Participants

Ten native Mandarin speakers and 10 native Vietnamese speakers were recruited for the test. Each participant was given a small financial compensation for their time (approximately 15 minutes) and effort.

2.2 Procedure

The experiments were programmed in E-Prime. A Dell computer with a 17-inch monitor was used. The experiments were conducted in a sound booth, and the stimuli were displayed on the computer screen. Both the Mandarin and Vietnamese experiments followed the same sequence of steps. Each word set was presented in two phases: a learning phase and a testing phase.

2.2.1 The Learning Phase

In this phase, the participant was presented with a complete set of four word-pairs. The participant was instructed to take as long as required to memorize the words in the set. When the participant was finished with the memorization task, he or she proceeded to the testing phase.

2.2.2 The Testing Phase

In this phase, a fixation cross first appeared on the screen for one second to draw the participant's attention to the center of the screen, where a prompt word would appear. Then, one of the prompt words from the just-learned set was selected at random and displayed in the center of the screen. The participant spoke the corresponding response word, which was semantically related to the prompt word, into a microphone as quickly as possible. The microphone was connected to a serial-response box (SR-box) that sensed the moment the participant began a verbal response. The response time (RT) in milliseconds from the moment the prompt word was displayed to the onset of verbal response was recorded, and a new fixation cross appeared on the screen to announce the start of a new trial and, subsequently, the random selection of another prompt word from the word set. A digital recorder recorded each participant's response. Each of the four word-pairs in a set appeared four times for a total of 16 trials in each set for each condition. Thus, there were a total of 256 trials per participant in the Mandarin experiment (16 x 4 tone sets x 4 conditions) and 192 trials per participant in the Vietnamese experiment (16 x 6 tone sets x 2 conditions). A practice session containing eight trials was given before the experiment began.

2.3 Materials

2.3.1 Mandarin Experiment

Following the methodology employed by Chen *et al.* (2002), a 2 x 2 factor experimental design was used. For each condition, four sets of word-pairs were assembled, with four word-pairs in each set. Each pair consisted of a two-syllable prompt word and a two-syllable response word. The two words in a given pair were semantically related to facilitate ease of memorization; for example, a prompt word 生意 'business' is related to the response word 客户 'client'.

The first factor, syllable match type, consists of two conditions, a 3-way Match condition and a 2-way Match condition. In the 3-way Match condition, the first syllables of the response words all share the same segment, tone, and character. For instance, the response words for the word-pairs in Word Set 4 were 客户 *ke4-hu4* 'client,' 客人 *ke4-ren2* 'guests,' 客厅 *ke4-ting1* 'living room,' and 客栈 *ke4-zhan4* 'hotel'. The initial syllable for each has the same segment, *ke*; the same tone, tone 4 (high-falling); and the same written character, 客. In the 2-way Match condition, the first syllables of the response words in a single word set have different written characters, but still match in segment and tone, such as *xi2* - that is, *xi* with tone 2 (rising) - written as: 席, 媳, 习, and 袭.

Each of these two conditions, 3-way and 2-way, was also crossed with the other factor, set type, which has two conditions, Homogeneous and Heterogeneous. In the Homogeneous condition, each response word in a set is configured in the manner described above, that is, by matching the other response words in a set in all three features or in only two features. In the Heterogeneous condition, the first syllables of the response words in a set have different segments, different tones, and different written characters. In this way, the Heterogeneous condition serves as the baseline by providing a test of recall time under control conditions where words are not matched by component features. Table 1 shows a sample word set for each of the 2 x 2 conditions. A full list of stimuli for the Mandarin experiment can be found in Appendix 1.

Table 1. Sample word sets for the Mandarin experiment.

	Homogeneous Condition		**Heterogeneous Condition**	
	Prompt	Response	Prompt	Response
3-way syllable match	*sheng1-yi4* 生意 business	*ke4-hu4* 客户 client	*zhang1-qian1* 张骞 ancient character	*xi1-yu4* 西域 Tibet
	kuan3-dai4 款待 host	*ke4-ren2* 客人 guest	*zhuang1-jia1* 庄稼 crop	*fei2-liao4* 肥料 fertilizer
	sha1-fa1 沙发 sofa	*ke4-ting1* 客厅 living room	*fu4-jing1* 负荆 carry wipes	*qing3-zui4* 请罪 ask for punishment
	long2-men2 龙门 name of a hotel	*ke4-zhan4* 客栈 hotel	*sheng1-yi4* 生意 business	*ke4-hu4* 客户 client
2-way syllable match	*xia4-tian1* 夏天 summer	*xi2-zi3* 席子 cooling mats	*bing1-dai4* 冰袋 ice bag	*qing1-liang2* 清凉 cool
	feng1-su2 风俗 custom	*xi2-guan4* 习惯 custom	*feng1-su2* 风俗 custom	*xi2-guan4* 习惯 custom
	di2-ren2 敌人 enemy	*xi2-ji1* 袭击 attack	*wu1-mie4* 诬蔑 defame	*fei3-bang4* 诽谤 slander
	gong1-po2 公婆 parents in law	*xi2-fu4* 媳妇 daughter-in-law	*shu1-ji2* 书籍 book	*ke4-ben3* 课本 textbook

Four sets of word-pairs were assembled for the 3-way condition, and four sets were assembled for the 2-way condition. Each of the four sets focused on one of the four lexical tones in Mandarin. Thus, each tone was represented equally throughout the stimuli. The same word-pairs were used in the Heterogeneous condition as in the Homogeneous condition, except that the pairs in the heterogeneous set were created by drawing one word from each of the homogeneous sets, ensuring that no two heterogeneous response words shared any features.

2.3.2 Vietnamese Experiment

The same general configuration of materials was used for Vietnamese as for Mandarin. In other words, sets of four prompt and response word-pairs were assembled, with the first syllables of response words carrying matching features within each set. Nevertheless, there were two major differences that resulted from a linguistic difference between Mandarin and Vietnamese. First, there was no conceivable way to construct a 2-way condition. Since Vietnamese uses a phonetic alphabet, it is never the case that a given segment-tone can be expressed by more than one written symbol. Thus, only a 3-way condition is supportable in the Vietnamese experiment. Of course, the set type factor, consisting of Homogeneous (matching syllable features) and Heterogeneous (no matching features) conditions, still was used with the Vietnamese word sets, as with Mandarin. Second, Vietnamese has six lexical tones (see Appendix 3). So, in order to ensure equal distribution across the tonal spectrum, six Vietnamese word sets were assembled for each condition. Table 2 shows sample word sets for the two conditions in the Vietnamese experiment. A full list of the stimuli used in the Vietnamese experiment can be found in Appendix 2.

Table 2. Sample word sets for the Vietnamese experiment.

	Homogeneous Condition		**Heterogeneous Condition**	
	Prompt	Response	Prompt	Response
3-way syllable match	*hợp lý* logical	*phải chăng* reasonable	*con chim* bird	*cánh chuồn* wing
	đanh trống to beat down	*phải đòn* get a spanking	*lờ đi* ignore	*mặc thây* leave alone
	tình yêu romantic love	*phải lòng* to be just	*tình yêu* romantic love	*phải lòng* to be just
	lạnh lẽo wintery	*phải gió* caught in a draft	*ấn nút* to press a button	*lăn tay* to take fingerprints

3. Results

First, the incorrect answers were eliminated from the data. Also, any answers that were more than 2.5 SD from that participant's mean RT were excluded, resulting in approximately 12.5% of the data being eliminated. No participants or items were excluded because of low accuracy rates. The dependent measure is the RTs for the response words. A within-subjects analysis of variance was computed to determine the significance of the difference in RTs by condition.

The data from the Chinese experiment showed the following pattern (Figure 1). In the word recall process, participants had longer reaction times for the 3-way (syllable, tone, and orthography) Match than for the 2-way (syllable and tone only) Match condition, *F*(1, 9) = 2.728, *p* = 0.023. More surprisingly, participants had longer RTs in the Homogeneous (3-way and 2-way together) condition than in the Heterogeneous (Non-match) condition, although the difference was not statistically significant, *F*(1, 9) = 1.509, *p* = 0.166, contrary to Chen *et al.*'s (2002) results. A closer one-by-one check confirmed that all of the participants were faster in the Heterogeneous (Non-match) condition. In other words, no priming effect was found.

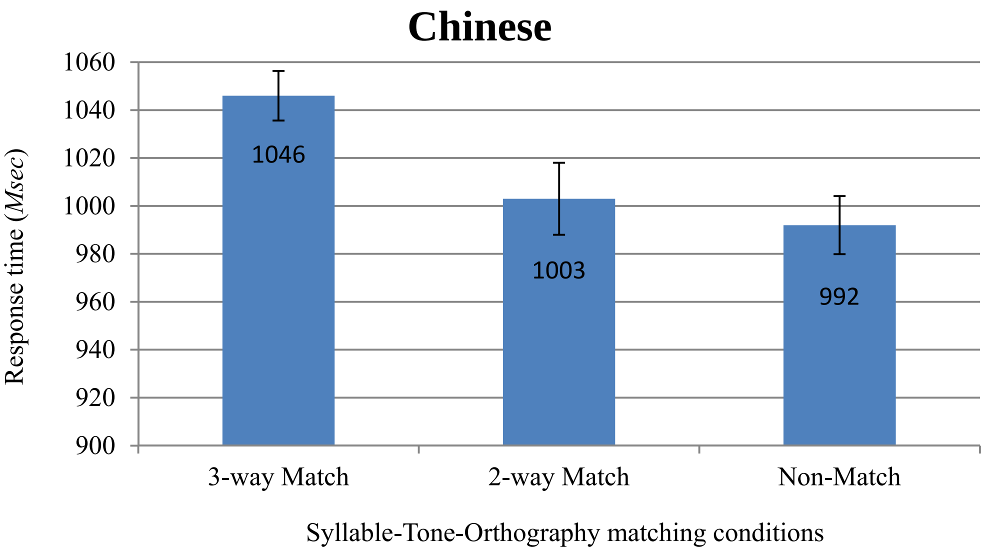


Figure 1. Mean reaction times (in milliseconds) of Chinese word recall.
Note: Error bars indicate standard error.

In contrast, the Vietnamese group showed the opposite pattern, as can be seen in Figure 2, in that the Homogeneous condition had a stronger priming effect compared to the Heterogeneous condition, $F_1(1, 9) = 15.99$, $p = 0.03$.

A comparison between the two target language groups, Chinese and Vietnamese, is shown in Figure 3. A 2-way repeated measures ANOVA, with Condition (Homogeneous and Heterogeneous) as a within-subjects factor and Language as a between-subjects factor, was performed to detect differences between the recall strategies of the Chinese and Vietnamese groups. The result shows a large main effect of Language *F*(1, 18) = 33.78, *p* < 0.01, which indicates that the Chinese speakers generally performed much faster in the experiment. More importantly, there is an interaction effect between Language and Condition, *F*(1, 18) = 18.64, *p* < 0.01, suggesting that the Chinese and Vietnamese groups were indeed using different word recall strategies.

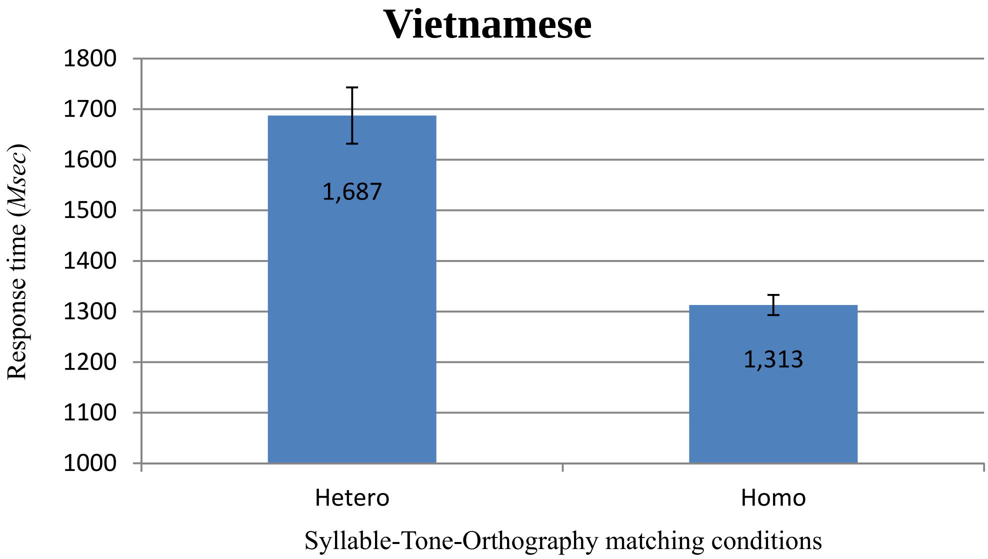


Figure 2. Mean reaction times (in milliseconds) of Vietnamese word recall. Note: Error bars indicate standard error.

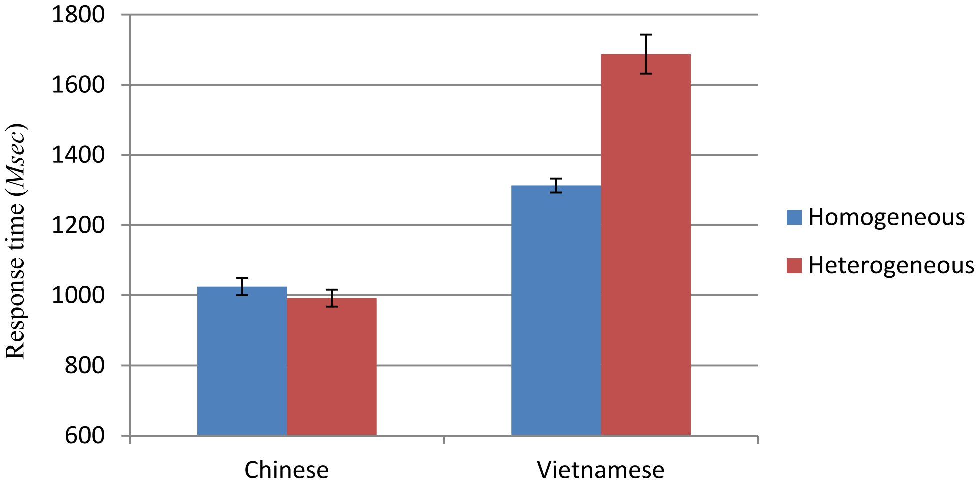


Figure 3. Comparison of reaction times (in milliseconds) between Chinese and Vietnamese word recall.

3.1 Discussion

The experiment testing Chinese speakers did not replicate the results of Chen *et al.*'s (2002) study, and there is a large gap between the average response times revealed in these two experiments (about 1000 ms in this study and 600+ ms in Chen *et al.*). The author attributes this apparent difference to two possible reasons. First, the response-time-measuring methodology employed in this study was not identical with that used by Chen *et al.* More specifically, participants in Chen *et al.*'s study were facing an arbitrary cutoff time of 1000 milliseconds. In their experiment, if "no response was initiated within 1000 ms of the presentation of the cue word" (p. 757), a feedback tone would sound and the trial would be terminated automatically. In other words, this setting eliminated all potential responses that were longer than 1000 milliseconds, regardless of individual differences in performing the word recall task. Although adding time pressure in psycholinguistic experiments is a common manipulation and might urge participants to focus and respond more quickly, it is not clear why an arbitrary cutoff time of 1000 ms was chosen. The current experiment, instead, set the cutoff time in the data analysis stage instead of the testing phase, and it was set according to each participant's average response times and standard deviations. This methodological difference is probably the main reason for the large difference in response times. Second, the participants in the two experiments were drawn from different language backgrounds. Although Mandarin Chinese was the mother tongue of the participants in both experiments, Chen *et al.*'s participants were mainly from Taiwan while the ones in the current study were from mainland China. There should be minimal influence resulting from this demographic difference, yet this factor may be worth our attention.

Considering the above differences, there are two main preliminary findings from this experiment, each with implications for alternative interpretations of Chen *et al.*'s results that ultimately may provide evidence for or against the relative importance of the three components as structural units in lexicon architecture. First, neither the Chinese 2-way Match condition nor the 3-way Match condition produced priming. Moreover, the 2-way Match condition resulted in faster RTs than the 3-way Match condition. This suggests that orthography may have played a major role in the implicit priming test, and that the logographic nature of Mandarin characters adds an extra layer of processing load in word production tasks. Second, although the Chinese participants had longer reaction times in the Heterogeneous condition than in the Homogeneous condition, the Chinese group showed a significant advantage in word recall times compared to their Vietnamese-speaking counterparts in all conditions. It is possible that unexpected phonological differences might be involved in these cross-linguistic findings, but, given the fact that Mandarin Chinese and Vietnamese are both tonal and monosyllabic, the difference in RTs is highly likely to be due to the different orthographies the two languages employ. In general, the graphic dimension of the

Chinese writing system may add an identifying element to the configuration of a mental lexical representation, which is absent in Vietnamese orthography. Mandarin orthography may provide an extra symbolic or iconic dimension that goes beyond the representation of morphemic or prosodic aspects to become a component in the word recognition mechanism for Mandarin speakers. This finding adds evidence in support of a suggestion made in some previous studies (Biederman & Tsao, 1979; Treiman *et al*., 1981) that Chinese characters invoke meaning much quicker than words in an alphabetic language. The current study supports the idea that the Chinese and Vietnamese speakers were using different word retrieval processes. More specifically, by using a logographic writing system that has no relationship to the pronunciation of words, the Chinese speakers may have developed integrated lexical entries that facilitate their word recognition process. They can give a brief glance at the character and recall all of the factors related to it, such as sounds and tones. Using an alphabetical writing system, the Vietnamese speakers have to read a sequence of letters, which relate to the words' sound and tone combinations, in their word processing. Without an additional indexing logographic symbol, they may have separate entries for syllable and tone, since the two elements are presented separately in the writing (prosodic components of Vietnamese words [tonality] are represented by diacritical marks above vowel symbols in Vietnamese writing. See Appendix 2). This may cause the delay in their word recognition process shown in this experiment.

4. General Discussion

4.1 Lexicon Architectures and Implications in Language Teaching Methodology

As mentioned in the introduction, I believe that the difficulty that learners, especially the native speakers of Indo-European languages, face in learning Chinese lies in the inherent difference between first language acquisition and second language acquisition. More specifically, it seems intuitively true that using an alphabetic native language would form the habit of focusing solely on the syllable part of the word and would prevent the speakers from taking syllable, tone, and orthography as an integrated word entry.

Research regarding the processes involved in word recognition has developed along diverse pathways, producing various theoretical models that address the nature and architecture of the human mental lexicon. They run the gamut from early approaches based on pure frequency of occurrence, such as the LOGOGEN model (Morton, 1969), to more recent and more elaborate mechanisms of activation, such as those proposed by the Distributed COHORT model of Gaskell & Marslen-Wilson (1997) and the NEIGHBORHOOD activation model of Luce & Pisoni (1998). They also have reflected the modular versus interactive debate with such theories as the Autonomous View of Norris *et al*. (2000) in his MERGE model and the connectionist approach taken by the TRACE model (McClelland & Elman, 1986). Even though some researchers have argued for no substantial representation of word forms in the mental lexical at all, as in Goldinger's conceptualization of episodic memory (1998), still the so-called traditional approach in describing the word recognition process has persisted. In their article "Spoken Word Recognition," Dahan and Magnuson (2006, p. 253) have defined a traditional approach that conceives of the lexicon as the "mapping of the speech input onto abstract lexical representations, with abstract units standing for the word's subcomponents, the phonemes, mediating this mapping." A crucial factor in determining the nature and function of the lexicon architecture is the identification of the number of components that comprise the structure of a single lexical representation. An even more fundamental question is whether the collection of components is universal across languages or whether a subset of such components may be language-specific.

Based on the results of this experiment, I propose that there are differences in the lexicon architectures between lexical tone languages with characters (*i.e.*, Chinese), lexical tone languages without characters (*i.e.*, Vietnamese), and non-lexical tone languages (for example, Dutch). It is possible that, in a native Chinese speaker's mind, every word (character) is an integrated entry consisting of three indispensable elements: the phonological segment, tones, and orthography, which are fused together and form a single entry. In contrast, in a Vietnamese speaker's mental lexicon, each word entry is a combination, in which the phonological segment is the major element of the construction and tones are likely attached to the phonological segment as supra-segmental elements. Orthography, as a direct combination of segment and tones, does not act as an independent element in the mental lexicon. As for non-lexical tone languages, such as Dutch, only segment is involved in the word retrieval process, as it represents the alphabetic orthography, and lexical tone is simply not an element in these languages.

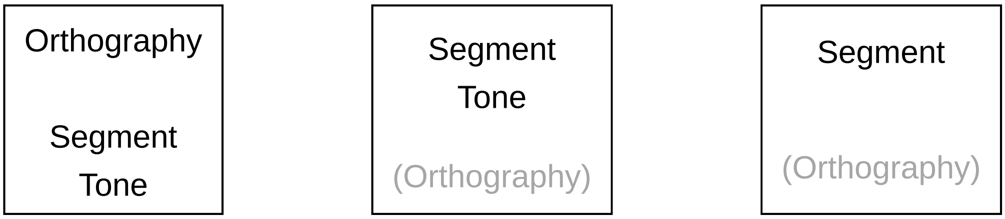


Figure 4. Possible lexical representations in speakers' mental lexicon of lexical tone language with characters (left), lexical tone language without characters (middle), and non-lexical tone language (right).

If this is the case, then we can speculate that speakers of languages that do not offer an ideographic character writing system may, when learning Chinese, build a different lexicon architecture than that of native Chinese speakers, due to transfer from their first language.

More specifically, they may have a poly-variant schema (presented graphically in Figure 5), in which segment, tone, and orthography form a loose combination, and in which segment is the major element of the construction. The tones are attached to the phonological segment, while orthography is a peripheral element and plays a very limited role in the storing and retrieving process. This hypothesis is supported by previous studies that have shown that L2 Chinese speakers may process tones and characters differently than L1 speakers. Wang *et al.* (2001) discovered that Mandarin tones are predominantly processed in the left hemisphere by native Mandarin speakers, whereas they are bilaterally processed by American English speakers with no prior tone experience. In addition, Hayes (1988) claimed that native and non-native Chinese speakers may use different strategies in reading Chinese characters. More specifically, the strategy used by native Chinese readers for holding words in short-term memory was acoustically oriented, in that they immediately associated the symbol with a particular sound, while the non-native Chinese readers rely more on graphic processing.

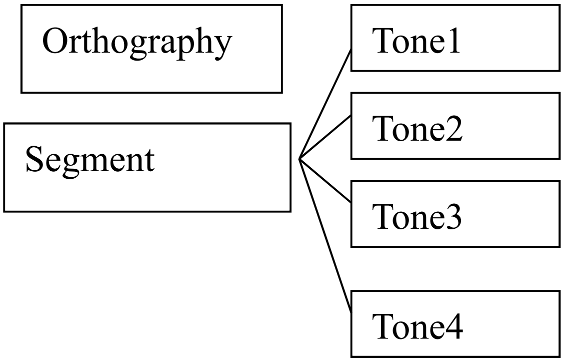


Figure 5. Possible lexical representations in L2 Chinese speakers' mental lexicon.

Therefore, teachers of Chinese need to find ways, through our teaching practice, to prevent our students from forming habits of neglecting orthography and marginalizing tone information. Intuitively, it may seem easier for the learners to learn through acquiring the familiar syllable forms, which are like those that exist in their own language, and later learn the new and unfamiliar tonal features. Nevertheless, it is not feasible to first introduce only syllables (*e.g.*, *ma*), and later teach the four lexical tones and ask the students to try to pair the syllable with each of the four tones (ma1, ma2, ma3, ma4). Learning these elements in such an order would be likely to cause students even more difficulty in the later stages of learning, since the students may lose the initial opportunity to form an integrated combination of syllable and tone, as the Chinese native speakers do. All teachers of Chinese as a foreign language can probably recall from time to time encountering students who can pronounce the syllable correctly but constantly make mistakes on the tones, which is probably the consequence of the students' learning syllables and tones separately. Hence, I propose that it may be better for teachers to initially ask the students to only imitate word pronunciation in a repeated manner, without even introducing the notion of tones. This extreme method might be hard to practice, but it would avoid emphasizing the tone features separately from the phonological segment. The goal would be to duplicate the natural learning process in first language acquisition and to help the students form native-like word entries in their lexicon as much as possible.

Regarding the role of writing systems, the experiment proved, through comparison with another tonal language, Vietnamese, which uses an alphabetic writing system, that the special logographic characters employed in Chinese have a facilitation effect in word recall. This result is instructive for teachers in that it points to the benefits of introducing orthography simultaneously with the teaching of the segment and tone when students are already very familiar with the pronunciation of a certain word. In addition, it emphasizes the fact that, for each lesson, introducing too many characters will actually lower the rate of learning, because students will not have enough time to integrate the characters they just learned with pronunciation, and instead will probably rely more heavily on sound noting systems (such as *pinyin*), which may further discourage them from remembering characters. Moreover, as soon as they master the characters, the students should be encouraged to read texts *sans* pinyin in order to fortify their memory of integrated mental lexical entries.

To conclude, the preliminary results of this study suggest that Chinese teachers should pay careful attention to the teaching order of the three main elements in Chinese. The best strategy might be to introduce syllable and tone for each word simultaneously, and only when the recognition of the two are stable to teach the writing of characters.

4.2 Limitations and Possible Future Studies

Revisiting Chen *et al.*'s (2002) experiment, this study limited its examination to the relationship of orthography to a conceptual amalgam of segment and tone in word recall. This experiment partially replicates the design of Chen *et al.*'s (2002) study and adds Vietnamese as a testing group in order to explore the mental lexicon representations of Chinese speakers; however, there are interesting discrepancies between the findings described here on implicit priming effect and those of Chen and colleagues. The current study suggests that orthography affects word retrieval processes, while it was considered to have a minor role in the mental lexicon in Chen *et al.*'s study.

Clearly, however, the experimental methodology employed in both studies has room for improvement. For example, semantic relationships between cue words and response words were inconsistent. In some pairs, the cue words may be more closely related to their response words than in other pairs. As a result, reliance on compound word form can give rise to inadequate control over testing materials. In addition, the experimental design actually required both perception and production, and using a task combining these processes may have complicated the results, which can hardly be considered to only measure word recognition independent of the production process.

Therefore, the author hopes that more data can be collected using fine-tuned methods, such as an eye-tracking paradigm (Cutler *et al.*, 2006) in the near future. To test the proposed model of mental lexicon representation showed in Figure 5, I intend to conduct a word recognition-based experiment employing an eye-tracking paradigm that relies on mono-morphemic picture stimuli, thus avoiding any confusion due to the use of compound forms. I will further isolate the effects of word production by collecting the percentages of eye-gazes toward target and control pictures as dependent variables. Once the effects of orthography are satisfactorily isolated or eliminated, the author intends to probe more deeply into the specific nature of the segment-tone combination. For instance, in a more recent study, Xu and Speer (2007) report the results from a cross-modal priming experiment that led them to conclude that "prosodic tonal information is not processed at a separate 'toneme level' from the 'phoneme level' during lexical access. Instead, lexical tone is an integrated component of the auditory signal used in Mandarin word recognition." I find this conclusion to be consistent with my own theoretical position on the role of tonality in the process of word recognition (although it seems somewhat speculative, given the limited and problematic nature of their stimulus materials, which rely on sandhi-driven tone changes). In a word, the employment of new methodologies will hopefully provide more definitive answers in the ongoing discussion, yielding stronger evidence and more meaningful implications in the field of teaching Chinese as a second language.

References

Biederman, I., & Tsao, Y.-C. (1979). On processing Chinese ideographs and English words: Some implications from Stroop-test results. *Cognitive Psychology*, *11*, 125-132.

Chen, J. Y., Chen, T. N., & Dell, G. S. (2002). Word-form encoding in Mandarin Chinese as assessed by the implicit priming task. *Journal of Memory and Language*, *46*, 751-781.

Cutler, A., Weber, A., & Otake, T. (2006). Asymmetric mapping from phonetic to lexical representations in second-language listening. *Journal of Phonetics*, *34*(2), 269-284.

Dahan, D., & Magnuson, J. S. (2006). Spoken word recognition. In M. J. Traxler & M. A. Gernsbacher (Eds.), *Handbook of Psycholinguistics, 2nd Edition* (pp. 249-283). London: Academic Press.

Everson, M. E. (1998). Word recognition among learners of Chinese as a foreign language: Investigating the relationship between naming and knowing. *The Modern Language Journal*, *82*, 194-204.

Gaskell, M. G., & Marslen-Wilson, W. D. (1997). Integrating form and meaning: A distributed model of speech perception. *Language and Cognitive Processes. Special*

Cognitive Models of Speech Processing: Psycholinguistic and Computational Perspectives on the Lexicon, *12*, 613-656.

Goldinger, S. D. (1998). Echoes of echoes? An episodic theory of lexical access. *Psychological Review*, *105*, 251-279.

Hayes, E. B. (1988). Encoding strategies used by native and non-native readers of Chinese Mandarin. *Modern Language Journal, 72*, 188-195.

Li, C. W-C. (2007). What Chinese language instructors should know about Mandarin tone: Phonological representation, acoustic targets, and cognitive processes. Paper presented at the *Chinese Language Teachers Association of California Spring 2007 Workshop*, Stanford University, Palo Alto, CA.

Luce, P. A., & Pisoni, D. B. (1998). Recognizing spoken words: The neighborhood activation model. *Ear and Hearing*, *19*, 1-36.

McClelland, J. L., & Elman, J. L. (1986). The TRACE model of speech perception. *Cognitive Psychology*, *18*, 1-86.

McGinnis, S. (1997). Tonal spelling versus diacritics for teaching pronunciation of Mandarin Chinese. *The Modern Language Journal*, *81*, 229-236.

Miracle, W. C. (1989). Tone production of American students of Chinese: A preliminary study. *Journal of the Chinese Language Teachers Association*, *24*(3), 49-65.

Morton, J. (1969). Interaction of information in word recognition. *Psychological Review*, *76*, 165-178.

Norris, D., McQueen, J. M., & Cutler, A. (2000). Merging information in speech recognition: Feedback is never necessary. *Behavioral & Brain Sciences*, *23*, 299-370.

Treiman, R., Baron, J., & Luk, K. (1981). Speech recording in silent reading: A comparison of Chinese and English. *Journal of Chinese Linguistics*, *9*, 116-125.

Xu, L., & Speer, S. R. (2007). Integrated representation and access of tone and segments. Paper presented at the *20th Annual CUNY Conference on Human Sentence Processing*, La Jolla, CA.

Wang, Y., Spence, M., Jongman, A., & Sereno, J. A. (1999). Training American listeners to perceive Mandarin tones. *Journal of the Acoustical Society of America*, *166*(6), 3649-3658.

Wang, Y., Spence, M., Jongman, A., & Sereno, J. A. (2001). Dichotic perception of Mandarin tones by Chinese and American listeners. *Brain and Language, 78*, 332-348.

Appendix 1: Mandarin Word Sets

	Word Set		Prompt	Meaning	Response	Meaning
3-way Syllable Match	**Set 1**	1	shui3-guo3 水果	*fruit*	xi1-gua1 西瓜	*watermelon*
		2	zhang1-qian1 张骞	*ancient character*	xi1-yu4 西域	*west land*
		3	mian4-shi4 面试	*interview*	xi1-zhuang1 西装	*suit*
		4	da2-lai4 达赖	*Dalai Lama*	xi1-zang4 西藏	*Tibet*
	Set 2	5	jie2-shi2 节食	*diet*	fei2-pang4 肥胖	*fat*
		6	guo3-shi2 果实	*fruit*	fei2-shuo4 肥硕	*big*
		7	zhuang1-jia1 庄稼	*crop*	fei2-liao4 肥料	*fertilizer*
		8	xi3-zao3 洗澡	*take a bath*	fei2-zao4 肥皂	*soap*
	Set 3	9	sheng1-bing4 生病	*become ill*	qing3-jia4 请假	*leave of absence*
		10	wei4-min2 为民	*for the people*	qing3-ming4 请命	*request*
		11	gong1-zhu3 公主	*princess*	qing3-an1 请安	*inquire after*
		12	fu4-jing1 负荆	*carry wipes*	qing3-zui4 请罪	*ask for punishment*
	Set 4	13	sheng1-yi4 生意	*business*	ke4-hu4 客户	*client*
		14	kuan3-dai4 款待	*host*	ke4-ren2 客人	*guests*
		15	sha1-fa1 沙发	*sofa*	ke4-ting1 客厅	*living room*
		16	long2-men2 龙门	*name of a hotel*	ke4-zhan4 客栈	*hotel*

	Word Set		**Prompt**	**Meaning**	**Response**	**Meaning**
2-way Syllable Match	**Set 5**	17	bing1-dai4 冰袋	*ice bag*	qing1-liang2 清凉	*cool*
		18	zi1-sha1 自杀	*suicide*	qing1-sheng1 轻生	*suicide*
		19	kun1-chong2 昆虫	*insect*	qing1-ting2 蜻蜓	*dragonfly*
		20	shu1-cai4 蔬菜	*vegetable*	qing1-jiao1 青椒	*green pepper*
	Set 6	21	xia4-tian1 夏天	*summer*	xi2-zi3 席子	*cooling mat*
		22	feng1-su2 风俗	*custom*	xi2-guan4 习惯	*habit*
		23	di2-ren2 敌人	*enemy*	xi2-ji1 袭击	*attack*
		24	gong1-po2 公婆	*parents-in-law*	xi2-fu4 媳妇	*daughter-in-law*
	Set 7	25	zhu1-bao3 珠宝	*jewelry*	fei3-cui4 翡翠	*emerald*
		26	dao3-guo2 岛国	*island country*	fei3-ji4 斐济	*Fiji*
		27	wu1-mie4 诬蔑	*defame*	fei3-bang4 诽谤	*slander*
		28	chan2-mian2 缠绵	*lingering*	fei3-ce4 悱恻	*sorrowful*
	Set 8	29	sha1-fa1 沙发	*sofa*	ke4-ting1 客厅	*living room*
		30	ren4-zhen1 认真	*serious*	ke4-ku3 刻苦	*painstaking*
		31	kun4-nan2 困难	*difficulty*	ke4-fu2 克服	*conquer*
		32	shu1-ji2 书籍	*book*	ke4-ben3 课本	*textbook*

Appendix 2: Vietnamese Word Sets

Word Set	**Prompt**	**Meaning**	**Response**	**Meaning**
set 1				
1	con chim	*bird*	cánh chuồn	*wing*
2	thủy thủ	*sailor*	cánh buồm	*sail*
3	màu đỏ	*red*	cánh hồng	*rose petal*
4	bàn chân	*foot*	cánh tay	*arm*
set 2				
5	món cá	*fish*	mặc ngư	*squid*
6	lờ đi	*ignore*	mặc thây	*leave alone*
7	dẫ cho	*although*	mặc dũ	*although*
8	đồng ý	*to agree*	mặc ước	*tacit agreement*
set 3				
9	hợp lý	*logical*	phải chăng	*reasonable*
10	đanh trống	*to beat down*	phải đòn	*get a spanking*
11	tình yêu	*romantic love*	phải lòng	*to be just*
12	lạnh lẽo	*wintery*	phải gió	*caught in a draft*
set 4				
13	cuôn giấy	*a scroll of paper*	lăn cù	*to roll*
14	chết trận	*to die in battle*	lăn đùng	*to drop dead*
15	gian khổ	*tribulation*	lăn lộn	*to experience hardship*
16	ấn nút	*to press a button*	lăn tay	*to take fingerprints*
set 5				
17	tiền lãi	*dividend*	đồ lợi	*to seek profit*
18	trái cây	*fruit*	đồ ăn	*food*
19	bức tranh	*painting*	đồ hoạch	*drawing*
20	đại học	*college*	đồ đệ	*student*
set 6				
21	đường giao	*intersection*	ngã tư	*crossroads*
22	rơi xuống	*to collapse*	ngã gục	*to be set up*

23	kết cục	*conclusion*	ngã ngũ	*concluded*
24	bất đồ	*unexpectedly*	ngã ngửa	*to be shocked*

Appendix 3: Explanation of Chinese and Vietnamese Orthographic Tone Marks

Chinese

Tone #	Description	Diacritical Mark	Example	Meaning
1	High level tone	Horizontal bar	mā	*mother*
2	Rising tone	Left to right rising slant	má	*hemp*
3	Dipping tone	falling slant connected with rising slant	mă	*horse*
4	High falling tone	Left to right falling slant	mà	*scold*

Vietnamese

Tone #	Description	Diacritical Mark	Example	Meaning
1	High level tone	No mark	ma	*ghost*
2	High rising tone	Left to right rising slant	má	*mother*
3	Low falling tone	Left to right falling slant	mà	*that*
4	Dipping tone – low to mid	Question mark (no dot)	mả	*tomb*
5	High broken tone – low to high	Tilde	mã	*horse*
6	Low broken tone – low to lower	A dot below the vowel	mạ	*burgeon*

Computational Linguistics and Chinese Language Processing
Vol. 18, No. 3, September 2013, pp. 21-44

Understanding Mandarin Prosody: Tonal and Contextual Variations in Spontaneous Conversation

Li-chiung Yang*, and Richard Esposito+

Abstract

Tonal identity and tonal variation in Mandarin have been the focus of intensive research that has long sought to bring out the underlying causes of variations in realized pitch values. Included among the variables studied in tonal variation are syntactic, contextual, emotional, and interactional influences. In the current study, we present results of our comparative research into tonal pitch variation in read speech and spontaneous Mandarin conversations. We acoustically and quantitatively characterize differences in the degree of pitch variability of these two modes of speech, and we present our results on tonal variability, as well as the influence of tone sequencing, syllable amplitude, and contextual factors on realized tonal shape. We show that, although tones are manifested in great diversity of pitch in spontaneous speech, there is a consistency of pitch shape that is dependent on tonal lexical identity.

Keywords: Tone, Prosody, Mandarin, Tonal Variability, Spontaneous Speech

1. Introduction

Previous research on read speech and spontaneous speech in Mandarin has demonstrated the wider variability of pitch movement in the latter, and researchers have attributed the greater variability to a number of different factors. Research on read speech (Xu, 1997; Shih, 1992; Shen, 1990) has predominantly focused on production and perception of tones in read or experimentally elicited speech. In these studies, prominent results have been the elucidation of rules for tone sandhi and the modification of tonal shapes under differing aspects, such as questioning, focus, or emphasis.

* English Language Center, College of Arts, Tunghai University, Taichung Taiwan
TEL: 011-886-04-2359-0121x31923 FAX:011-886-04-2359-0232
E-mail: yang_lc@thu.edu.tw
The author for correspondence is Li-chiung Yang.
+ Spoken Language Research, USA
E-mail: esposito_r@sprynet.com

Recent research has demonstrated the wide divergence between defined tonal pitch values that occur in spontaneous speech (Tseng, 2005) and values in read speech. The interactive and cognitively intense environment of spontaneous speech provides an abundance of differing factors that often lead to tonal sequences with tones that seemingly rarely reach their defined values.

Researchers have suggested several avenues to explain systematic divergences from intrinsic tone shape, including tone sequence patterns, targeting of adjacent tones, and stress and metrical patterns of speech. Prior research on speech prosody in both tone and non-tonal languages (Hirst *et al.*, 1998; Shriberg *et al.*, 2000; Tseng, 2005; 2009; Tseng, 2010) has shown that there are a number of important influences on the prosody of natural speech, including interactive monitoring activity, emotional state, and the level of uncertainty, as well as topic organization and phrasal position.

Our view is that experimental and spontaneous speech corpora are complementary and each has its important role to play in the discovery process. Experimental and read speech data are ideally suited to testing hypotheses on relationships among known variables while controlling for confounding effects of other variables. Study of spontaneous speech, on the other hand, is ideally suited to the discovery of new contributing variables and to the forming of new hypotheses. In addition, the study of spontaneous speech, as in other natural science fields, has the potential to yield valid information on underlying processes that are uncovered through the identification of systematic parallels observed across the available data. This discovery of relevant variables is especially important when a particular phenomenon could arise from more than one underlying cause or from an alternate cause. By combining the results of read or experimental studies with those of spontaneous speech, the relative robustness of experimental results in a spontaneous setting can be used as a marker or clue to the existence of other potentially important variables.

With this view in focus, in the current paper, we study how tonal sequences and amplitude affect the realization of lexical tone shape and compare results obtained from a read speech corpus to results found in spontaneous Mandarin. We further investigate differences between the speech modes, and study the roles of tone sequence, speaker variability, and lexical identity in shaping realized tone values.

Section 2 describes our methodology, including the nature and extent of the data corpora, and the automatic and post-processing steps taken to extract acoustic parameters from the speech signal.

In Section 3, we introduce an innovative technique to facilitate consistent comparison of measures of Chinese tonal f_0 contour in large speech corpora. We describe the importance of spontaneous speech to realized tonal variability and also present a benchmark comparison of spontaneous to read f_0 shape. We explore several important factors in the determination of f_0 shape in spontaneous speech, including tonal identity, word syllable number and position, and tonal sequence patterns, including tone sandhi. We present results of our comparisons grouped by lexical tone, by tonal sequence patterns, and by mono- or di-syllabic words, and we indicate the degree to which our results match prior theories on anticipatory and carryover effects of tone sequences. Section 3.3 introduces the data reduction techniques used to extract comparable measures of f_0 shape and provides graphical representations of the distributions of these measures as important guides to the behavior of syllable f_0 in natural spontaneous conversations. In Section 3.4, we show individual instances of tonal variation in spontaneous speech and suggest mechanisms for the specific variational tendencies revealed by the data.

Section 4 provides a summarization of the key findings of the paper, the importance of tone variability in spontaneous Mandarin, and how future work on tonal variability can be enhanced through the techniques introduced.

2. Data and Methodology

2.1 Data, Participants, and Approach

The data utilized in this study are part of a larger project on Mandarin conversational speech, totaling over 20 hours of speech. For this study, a subset of continuous speech from two conversations, one between two female speakers (Speaker P and Speaker S) and the other between one female (Speaker T) and one male speaker (Speaker B), totaling approximately 40 minutes in duration for spontaneous speech were selected and analyzed. In addition, to provide a baseline comparison between the read and spontaneous speech modes of the same speaker, 10 minutes of read speech, consisting of four read stories, by the same male speaker in the spontaneous conversation, were also analyzed. As previous research on Mandarin has concentrated almost exclusively on read or experimentally controlled speech (with the exception of Tseng, 2004, 2005, 2009), in this study, our goal is to concentrate on exploring tonal and prosodic variations in spontaneous Mandarin Chinese.

The spontaneous conversation data were collected in informal settings, and the read speech corpus was collected in a laboratory. Speech data were recorded using a SONY PCM-M1 DAT recorder with a SONY ECM lavalier microphone at a sampling rate of 22,050 kHz. Data were manually segmented to the phrase, word, and syllable levels using Wavesurfer and ESPS/xwaves for their ease with extended speech data processing capability, and acoustic-prosodic features such as time, amplitude, and pitch (f_0) values were automatically extracted from the speech files using ESPS/xwaves function *get_f0*. F_0 values were further corrected for formant outliers (*e.g.* doubling and halving), and slight errors in segmentation among the different label files were adjusted automatically.

A total of 7,710 lexical syllables from the 40 continuous minutes of segmented spontaneous speech were obtained, after eliminating overlapping syllables for which f_0 values were ambiguous between speakers. For read speech, there was one speaker, and a total of 1,804 syllables of speech. Tables 1-3 show the breakdown by speech mode, corpus, speaker, and tone. Altogether, the corpora investigated contained a total of 9,514 syllables over 50 minutes of speech.

In this study, we investigate the f_0 contours of tones as they are realized in spontaneous speech by utilizing several measures of f_0 contour to facilitate data reduction and comparisons across a large corpus of syllables. Our approach is to examine the data first, without any preconceived assumptions about specific tonal variations, show the patterns that emerge from this large corpus, and then relate our results to previous claims and findings.

The results between spontaneous and read speech are based on spontaneous and read speech corpora from one male speaker. While the sample size of the read speech is small, use of the same speaker highlights differences in syllable f_0 contour while controlling for speaker variability. Focusing on finding metrics to evaluate the variations in f_0 contour in spontaneous speech with respect to defined tonal shape, we present overall results on the consistency of tonal change across speakers in the wider spontaneous corpora.

Table 1. Number of syllables by tone, spontaneous conversation MC1, 2 female speakers, Speaker P and Speaker S.

	Tone 1	Tone 2	Tone 3	Tone 4	Tone 0	Total
P	439	338	455	902	357	2491
S	270	222	297	579	239	1607
Total	709	560	752	1481	596	4098

Table 2. Number of syllables by tone, spontaneous conversation MC2, 1 male speaker, Speaker B and 1 female speaker, Speaker T.

	Tone 1	Tone 2	Tone 3	Tone 4	Tone 0	Total
B	404	353	485	807	340	2389
T	151	182	261	449	180	1223
Total	555	535	746	1256	520	3612

Table 3. Number of syllables by tone, read speech, RS1, male speaker, Speaker B.

	Tone 1	Tone 2	Tone 3	Tone 4	Tone 0	Total
B	356	300	318	502	328	1804

3. Results: Variability of Tonal Shapes

The defined lexical tones in Mandarin are commonly recognized as Tone 1, which has a high pitch level and is flat (55); Tone 2, which starts at a mid to low pitch level, and rises (35); Tone 3, which starts at a mid to high level, falls, then rises (214); Tone 4, which is high and falling (51); and a neutral tone, Tone 0 (Chao, 1968). Tone 3 has two recognized variants in spontaneous speech: a high to low fall with no final rise and a short duration, low pitch value (21). Tone sequence pitch values are also governed by generally accepted tone sandhi rules (Chao, 1968): a Tone 3 that is followed by a Tone 3 takes on a rising pitch (33-->23); Tone 3 when followed by a non-3rd tone takes on a half-Tone 3, without an ending rise; and a Tone 4 that is followed by Tone 4 takes on a less steep fall.

As a simple first measure of relative concordance of syllable shapes with the defined pitch movement for tones, we used the linear slope of f_0 as a simple shape indicator to approximate f_0 values. Syllable f_0 values were extracted automatically and corrected as described in Section 2. The resulting f_0 values of each syllable were fitted using S-plus, and linear slopes and intercepts of each syllable were calculated. A linear slope would be most applicable for Tones 1 and 4, and, to a large degree, rising Tone 2. A quadratic fit would better capture the curvature of all tones, especially Tones 2 and 3, and was also produced for each syllable and used in measuring the effects of amplitude.

3.1 Tonal Values in Read and Spontaneous speech

Read, controlled, or experimental speech data are frequently considered as benchmarks that preserve a number of relatively stable phonological relationships in their realization. Table 4 shows the overall measure of tonal shape for our read speech data, using averages of linear approximations to each syllable's f_0 data, restricted to monosyllables occurring in the pre-pause or at phrase end position, to avoid the influence of adjacent tonal values. This is likely to have some downward bias due to pitch declination at phrase end, and this may cause the slight negative slope of level Tone 1. As expected, Tone 4 has a large falling slope, while Tones 2 and 3 have an overall positive rise. The f_0 minimum values shown in Table 4 indicate the percentage point within the syllable that attains the syllable minimum, and they provide additional information on syllable shape, as they mark the location of syllable f_0 slope direction change. The f_0 minimum point is very informative on the shape as well and is more robust with respect to averaging over different speakers and different speech situations, as this point is defined in percentage terms. In read speech, Tones 1 and 4 reach their minimum pitch point relatively close to the end, while Tones 2 and 3 reach their minimum pitch point very close to the midpoint of the syllable, matching the defined fall-rise shape of Tone 3, and rising Tone 2 with an initial fall.

Table 4. Mean slope in Hz change per second and percentage point of syllable f_0 minimum of each token, pre-pausal monosyllables, read speech, Speaker B.

	Tone 1	Tone 2	Tone 3	Tone 4	Tone 0
Slope	-39.73	36.42	58.98	-212.94	-75.49
f_0 min	0.64	0.48	0.48	0.72	0.55

Table 5 shows the parallel results for this same speaker as he engaged in spontaneous conversation. It is immediately clear that tone values in spontaneous speech diverge widely from their lexically defined values and from their realized values in read speech. In particular, Tone 1 exhibits a greater average fall than expected and Tone 4 displays an average fall that is just barely greater than defined level Tone 1. Notably, lexically rising Tone 2, which on average rises in read speech, also has an overall negative slope in spontaneous speech. Of the four tones, only Tone 3 has an overall rise that is similar to its read speech counterpart. Both Tables 4 and 5 are restricted to monosyllables in pre-pausal position, so the average pitch slopes for Tone 3 are independent of tone sandhi rules.

The f_0 minimum points for Tones 2, 3, and 4 occur earlier than in read speech and occur at nearly the same percentage position for Tone 1. For this speaker, Tones 1 and 2 in spontaneous speech became more negatively sloped, while Tone 3 remained similar to read speech. The most striking change occurs with Tone 4, which falls much less in spontaneous speech than in read speech and has about the same average slope as spontaneous Tone 1. Neutral tone (Tone 0) for this speaker becomes significantly more neutral, that is, *flatter*, in spontaneous speech, with a slope near zero, indicating very little pitch change.

Table 5. Mean slope in Hz change per second and percentage point of syllable f0 minimum of each tone, pre-pausal monosyllables, spontaneous speech, Speaker B.

	Tone 1	Tone 2	Tone 3	Tone 4	Tone 0
Slope	-62.31	-11.86	54.54	-64.97	3.91
f_0 min	0.64	0.41	0.34	0.54	0.38

The overall tone slope results, including mono-, di-, and tri-syllables, for the two participants in the current spontaneous corpus are presented in Table 6.

Table 6 indicates that the slope directions agree for each of the four tones across the two speakers, although the *strength* of directional changes varies, and the tones vary substantially from their defined lexical values. For both speakers, there is a striking similarity in average pitch slope for Tones 1 and 4, but there is also a consistent difference in degree between the speakers. For Speaker B, Tones 1 and 4 fall by about the same amount. For Speaker T, Tone 1 actually falls more than defined falling Tone 4, on average. Relative to each speaker’s pattern for all tones, Table 6 shows that Speaker T’s Tone 2 syllables fall relatively more, highlighting the importance of speaker variation in the realized tone shapes of spontaneous speech.

Table 7 further breaks out the results for the two speakers in spontaneous conversation MC2 by whether the syllable is a monosyllabic word (mono), the 1st syllable in a disyllabic word (D1), or the 2nd syllable in a disyllabic word.

Table 6. Mean slope in Hz change per second over all syllables, spontaneous speech, Speaker B and Speaker T.

	Tone 1	Tone 2	Tone 3	Tone 4	Tone 0
B	-79.84	-15.72	21.47	-77.64	-104.38
T	-188.39	-104.14	7.87	-138.69	1.16

Table 7. Mean slope in Hz change per second over monosyllables, 1st and 2nd syllables of disyllabic words, spontaneous speech, Speaker B and Speaker T.

Speaker	Tone 1	Tone 2	Tone 3	Tone 4	Tone 0
Mono-B	-62.31	-11.86	54.54	-64.97	3.91
Mono-T	-192.31	-124.76	14.43	-150.49	-127.10
D1-B	-90.49	-30.22	-23.42	-105.59	-
D1-T	-189.30	-79.72	-10.39	-227.56	-
D2-B	-52.57	7.24	-16.64	-69.99	-17.70
D2-T	-118.26	-128.95	1.86	-45.33	-72.46

The most striking pattern seen in Table 7 is the evident pervasive strength of a falling pitch, with almost all values showing an average negative slope. Table 7 clearly indicates that this pattern is similar for both speakers. As the two participants differ in overall pitch level, in Table 8 we show the same data normalized to Z-score values with respect to each speaker's overall average pitch mean and standard deviation, calculated across all f_0 values, by speaker.

Table 8. Mean slope in Hz change per second over monosyllables, 1st and 2nd syllables of disyllabic words, normalized to each speaker's average syllable pitch range, spontaneous speech, Speaker B and Speaker T.

Speaker	Tone 1	Tone 2	Tone 3	Tone 4	Neutral
Mono-B	-2.13	-0.41	1.87	-2.22	0.13
Mono-T	-4.36	-2.83	0.33	-3.41	-2.88
D1-B	-3.09	-1.03	-0.80	-3.61	-
D1-T	-4.29	-1.81	-0.24	-5.16	-
D2-B	-1.80	0.25	-0.57	-2.39	-0.61
D2-T	-2.68	-2.92	0.04	-1.03	-2.68

From Table 8, calculated from the spontaneous speech of corpus MC2, we can see that Tones 4 and 1 have the largest negative slope, but the mean values for Tone 4 do not have a consistently greater negative slope than Tone 1 for both speakers. For Speaker B, Tone 4 is marginally more falling than Tone 1, but for Speaker T, Tone 1 falls marginally more than Tone 4 for monosyllables and for the 2nd syllable of a disyllabic word.

The normalized data of Table 8 indicate that Speaker T has a greater propensity for a falling pitch over all tones except Tone 3, and for Tone 4 when it is the 2nd syllable of a disyllabic word; such differences in tonal modification may form a component of a speaker's characteristic or general speech style.

The data from Tables 7 and 8 show that, except for Tone 4, the slope of syllables in spontaneous speech exhibit a divergence from the lexically defined shape, on average. As these results are consistent across syllable types, Table 8 further suggests that the results of divergence from lexically defined tonal values may not be due solely to a pattern that affects only monosyllables or that is effected through syllable position in the word, and may arise from other causes as well.

3.2 Tonal Sequencing and the Influence of Preceding and Following Lexical Tone

Researchers on Mandarin tonal pitch values have proposed that local tone sequences can theoretically affect the realized target in two primary ways: through anticipatory effects of the upcoming syllable or through carryover effects of the preceding syllable. A succeeding tone value with a high f_0 onset should lead to a higher offset for the previous syllable, while a succeeding syllable with a low onset should induce a low offset in the previous syllable, according to the anticipatory theory. Analogously, carryover theory predicts that a high f_0 offset in a preceding syllable should lead to a higher onset for the succeeding syllable, while a preceding syllable with a low offset (half-Tone 3 and Tone 4) should induce a low onset in the succeeding syllable.

For example, Xu (1997) found greater evidence for the strength of carryover effects than for anticipatory effects using balanced sequences in experimental speech for Mandarin, while Chang & Hsieh (2012) found a more balanced effect of carryover and anticipatory effects for the more complex tonal system of Eng Choon Hokkien in their experimental data. In the current study, we were interested in investigating if similar effects on the realized target tones hold in our spontaneous speech data.

In Tables 9 through 12, we compare the average slope of each lexical syllable of two speakers from spontaneous conversation MC1, grouped by immediately preceding and succeeding syllable tone, using the linear regression slopes for the two speakers. The resulting slope coefficients were grouped by the subsequent lexical tone in Tables 9 and 10, and by the preceding tone for Tables 11 and 12.

Table 9. Mean slope in Hz change per second of each tone, by tone of the following syllable, Speaker P.

Slope of	X-Tone 1	X-Tone 2	X-Tone3	X-Tone 4	X-Tone 0
Tone 1	-78.4	-76.2	-125.3	-35.6	-119.5
Tone 2	-66.1	24.7	-115.8	-16.0	-53.2
Tone 3	-185.4	-76.7	-0.4	-149.5	-97.8
Tone 4	-133.6	-254.7	-248.8	-187.1	-202.9

Table 10. Mean slope in Hz change per second of each tone, by tone of the following syllable, Speaker S.

Slope of	X-Tone 1	X-Tone 2	X-Tone3	X-Tone 4	X-Tone 0
Tone 1	71.1	-114.3	-51.3	-53.4	-81.3
Tone 2	-23.7	-48.1	-58.7	52.6	-83.9
Tone 3	-211.1	-145.4	-87.8	-98.0	-205.5
Tone 4	-35.6	-233.8	-177.5	-139.9	-119.9

Table 11. Mean slope in Hz change per second of each tone, by tone of the preceding syllable, Speaker P.

Slope of	Tone 1-X	Tone 2-X	Tone 3-X	Tone 4-X	X-Tone 0
Tone 1	-56.4	-35.2	-67.4	-75.2	-102.6
Tone 2	-160.0	-130.0	30.3	-18.5	17.0
Tone 3	-211.3	-163.6	-118.5	-106.2	-26.7
Tone 4	-225.5	-249.7	-122.9	-201.4	-180.1

Table 12. Mean slope in Hz change per second of each tone, by tone of the preceding syllable, Speaker S.

Slope of	Tone 1-X	Tone 2-X	Tone 3-X	Tone 4-X	X-Tone 0
Tone 1	-116.5	-35.3	57.4	-104.3	-32.1
Tone 2	-124.7	-61.5	34.6	-12.3	-1.7
Tone 3	-332.8	-87.3	-149.9	-163.8	-114.4
Tone 4	-212.2	-127.7	-69.4	-151.9	-27.5

According to the anticipatory theory, high succeeding onset Tones 1 and 4 should show a relatively positive slope on the previous syllable, and conversely for low and low onset tones 2 and 3. In forward assimilation not based on the succeeding onset, a succeeding rising tone 2 can cause a rising effect. We can see from Tables 9-12 that there is considerable variability in the slopes of each tone depending on the tonal sequence; we can also see that this is in part speaker dependent. The data from Tables 9-12 suggest that different tone combinations merge contextually in different ways.

For example, for both speakers, Tables 9-10 show that the anticipatory condition is true for Tone 4 followed by all tones, that is, the following tone influences the pitch slope of the previous syllable in the hypothesized direction, on average. For both speakers, succeeding Tones 3 and 4 induce a deeper fall for Tone 4 than succeeding Tones 1 and 4. A succeeding neutral tone induces a fall for Tone 4 roughly between these two cases. Similarly, high onset succeeding Tone 4 induces a relatively flatter (but still falling) pitch for all Tones except initial Tone 2 for Speaker S in Table 10. Initial Tone 2 achieves an overall average rising pitch only when it is followed by Tone 2 for Speaker P, contrary to the anticipation account, and by Tone 4 for Speaker S, in agreement with the anticipation effect. Initial Tone 2 falls the most for both speakers when it is followed by Tones 2 and 3, again consistent with the anticipation effect. A succeeding Tone 1 appears as the most problematic: for both speakers, an initial Tone 3 achieves its greatest falling pitch when followed by Tone 1. Based on our data, we suggest that the defined level nature of Tone 1 may introduce a discrete jump rather than a transformation to a gradient-sloped succeeding tone.

The tone sequence values seen in these tables also show some evidence of tone sandhi effects on pitch shapes. Tables 9-12 show some evidence for the 3-3 tone sandhi rule, since, for both speakers, the first of two 3rd Tones falls the least of all 3rd Tones, especially for Speaker P. When followed by another tone, Tone 3 exhibits a strong falling contour, as predicted by tone sandhi. Similarly, comparing the 4-4 tonal sequences for both speakers in Tables 9-12 shows that the 1st of two 4th Tones will have a slightly flatter pitch slope than the 2nd. Tone sandhi rules can override assimilation, as seen when Tone 3 is followed by Tone 1, where Tone 3 remains low, rather than assimilating to high Tone 1. However, in our data, succeeding Tone 4 has a relative positive effect on Tone 3 for both speakers.

When looking at the influence of the preceding tone in Tables 11 and 12, we see that a preceding Tone 1 has a negative effect on the subsequent tonal slope. When preceded by Tones 2 or 3, the tonal slope becomes relatively higher for Tone 1 and Tone 2 for both speakers, and these results are in accordance with the carryover predictions. Also, preceding Tone 1 has a negative effect on all tones for Speaker S, and for all tones except Tone 1 for Speaker P. For both speakers, a succeeding Tone 4 has the greatest fall in pitch when it is preceded by high ending Tones 1 and 2, consistent with carryover.

Both anticipatory and lag effects are found in the above tables. Because of the greater consistency of anticipatory results, there may be a marginally greater influence of anticipatory effects than carryover effects, and the overall pattern provides support for Chang and Hsieh’s findings of more balanced effects from both anticipatory and carryover. The differences that are evident between the speakers suggest that the influence of sequencing in speech on tonal targets may be conditioned by a number of speaker factors, such as speech rate and speaking style. The results obtained confirm the results of prior researchers, who have found substantial effects of preceding and subsequent tones on the realized tonal pitch values in read and experimental speech (Xu, 1997; Chang & Hsieh, 2012).

3.3 Variability of Tonal Shape and Amplitude

As seen in the above results, average results to a certain degree can be accounted for by tonal sequencing and by consistency of speaker style in producing the defined lexical tones. Nevertheless, when we look at individual tokens, our data indicate that it is much more difficult to account for the wide range of pitch shapes found in spontaneous speech for a given tone *only through* the factors presented in the tables above. A wide range of phonological and linguistic phenomena has been cited as also affecting the pitch values of syllables. Among the factors affecting pitch are the position in the phrase, phonemic identity, emotional and interactive effects, communicative function, and prosodic environment.

3.3.1 Comparative Measures of Tonal Shape

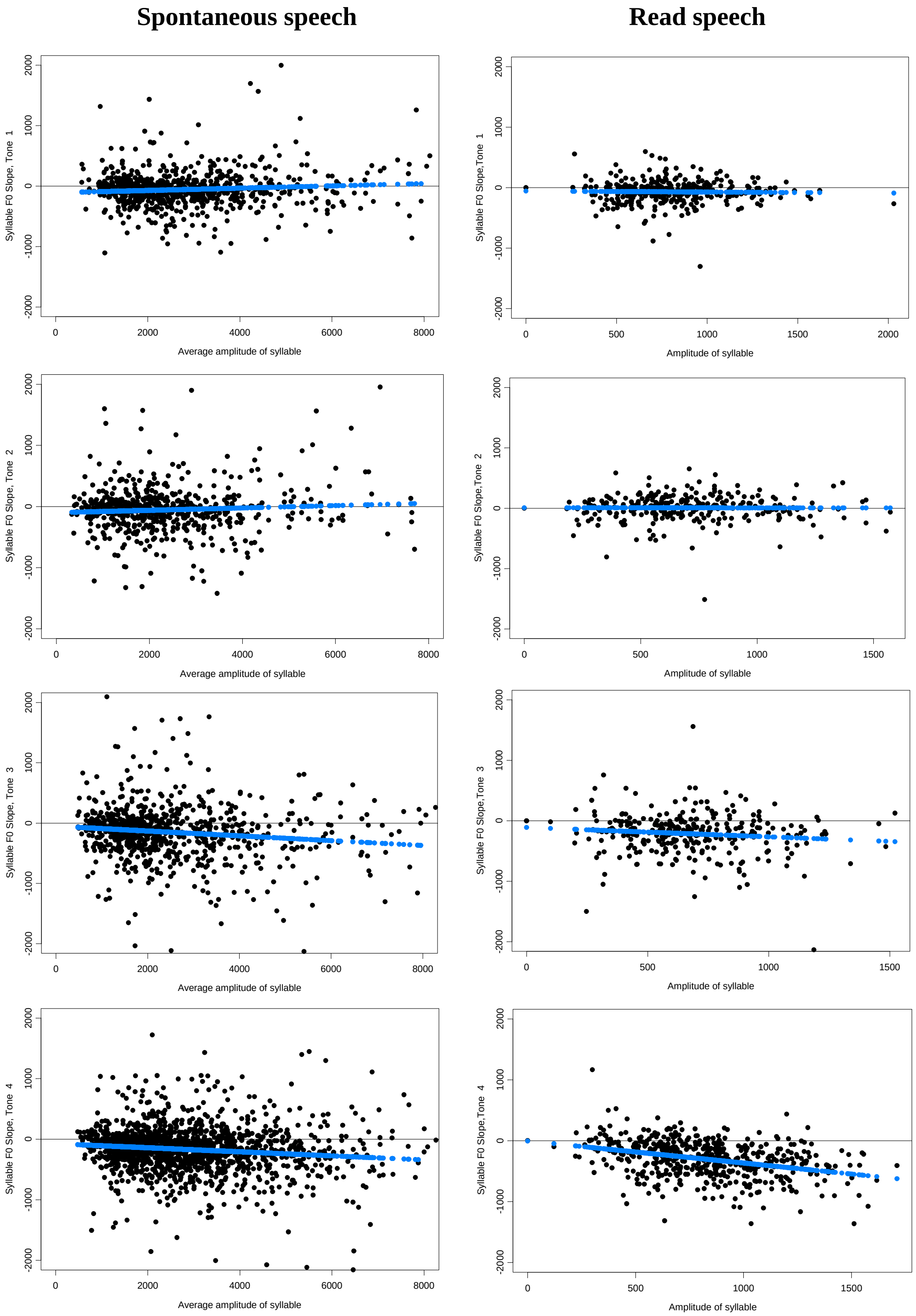
Figures 1 through 8 present a comparison of read speech to spontaneous speech, based on the simple linear slope measure of each syllable in the read speech corpus and the corresponding spontaneous speech counterpart, in which the fitted linear syllable slope is plotted against syllable average amplitude. A further regression line of slope vs. amplitude is superimposed on each figure.

Figures 1-4 show the slopes for spontaneous speech, and Figures 5-8 show the slopes for read speech. Comparison of spontaneous to read speech syllable slopes by tone shows that spontaneous speech is dispersed much more widely around the linear regression line. Furthermore, for each tone, for spontaneous speech the variation of pitch slope extends widely into both negative and positive slope regions, showing that the basic slope direction for spontaneous speech occurs with great frequency as either rising or falling.

By contrast, the slope value points for read speech in Figures 5-8 are clustered more tightly around the regression line, and their slope values are more in accordance with their defined lexical pitch values. In particular, for read speech, the slope values for Tone 1 cluster around zero, consistent with Tone 1's level pitch definition, and virtually all Tone 4 slopes are falling for read speech. Tone 3 has a preponderance of falling slopes, while Tone 2 has a somewhat greater preponderance of rising pitch values for read speech. The much greater dispersion of spontaneous speech tone pitch values into both negative and positive slope regions indicates the presence and greater influence of contextual variables in determining pitch shapes in spontaneous speech.

These figures indicate that *tonal identity* is still an important factor in the realized pitch shapes of syllables because of the consistencies with respect to tones that exist in the figures for both read and spontaneous speech. For example, comparing across tones for spontaneous speech shows that, like read speech, Tone 1 syllables tend to cluster relatively symmetrically around a zero level pitch slope value, in accordance with the defined flat slope of Tone 1, although, for both read and spontaneous, the overall slope average was slightly below zero. Spontaneous Tone 2 syllables generally do not have the steeper falling slopes associated with Tones 3 and Tones 4. Spontaneous Tone 4 syllables seem to be the most consistent in their adherence to falling pitch, with the great majority of syllables still having a negative slope, although not as consistently as with read speech. The overall tendency for a greater falling pitch for syllables than is predicted by the lexical identity may be the reason that Tone 4 exhibits the greatest resilience and adherence to its defined falling pitch value, as it combines both of these effects in its pitch realization. Pair-wise non-parametric Wilcoxon rank sum tests comparing read and spontaneous speech by tone indicate the existence of systematic contour differences between spontaneous and read speech at high significance levels (See Appendix A).

Results shown in Figures 1-8 indicate that amplitude also correlates with changes in tonal shape. A steeper falling slope for high amplitude syllables is seen in the figures for spontaneous and read speech Tones 3 and 4, and, for Tones 1 and 2, greater amplitude is moderately associated with flatter or rising slope values. Prior research using experimental speech has found that the distribution of energy in syllables is correlated with tone identity in Mandarin tones (Whalen *et al*., 1992). The current finding measures across syllables and finds that in both spontaneous and read speech, *higher average amplitude* of a syllable is associated with a greater degree of adherence to the defined lexical tone shape for the syllable as a whole.



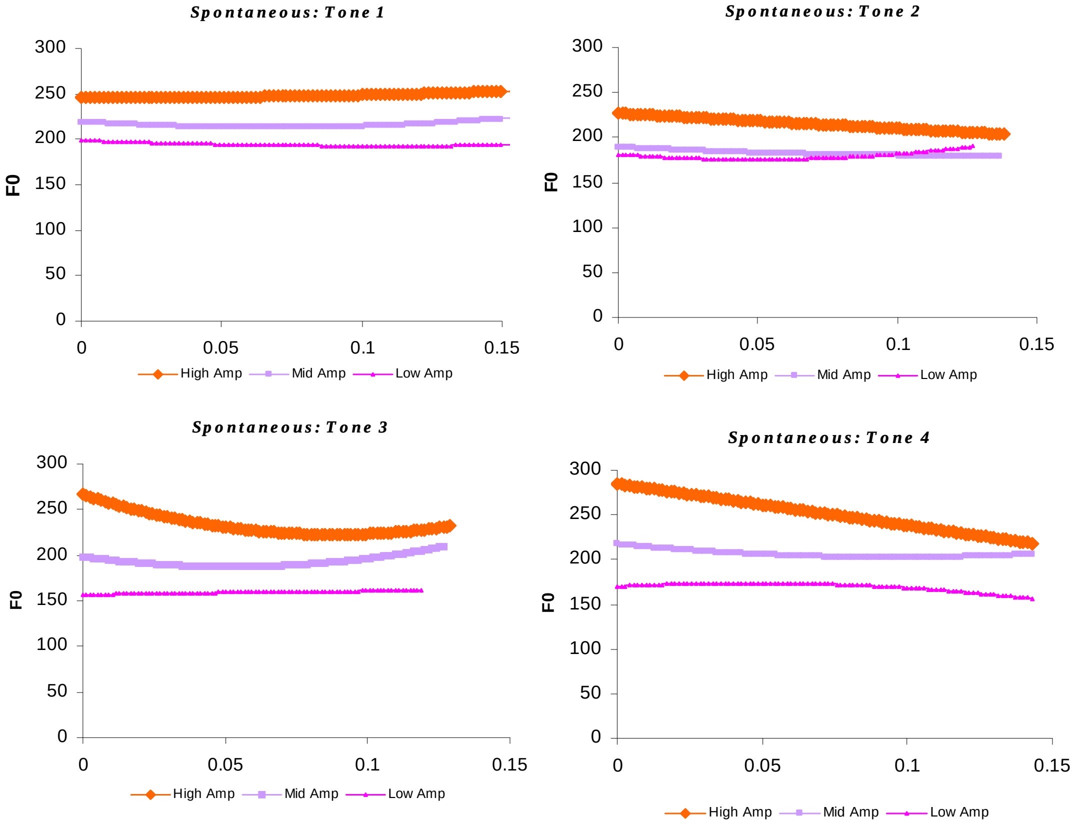
Spontaneous speech

Read speech

Figures 1-8. Slope of f_0 by amplitude for Tones 1-4, spontaneous speech shown in the left column and read speech in the right column. x-axis= syllable amplitude level; y-axis=syllable f_0 slope in Hz change per second

To achieve a more precise representation of the different tonal shapes, we fitted each syllable's normalized f_0 values to a quadratic polynomial. For all syllables in the spontaneous corpora used in this study, all f_0 values output were normalized by calculating Z-score values with respect to each speaker's mean syllable f_0 and standard deviation over all f_0 values for that speaker. After manual segmentation to the syllable level for all speech data, an average amplitude was calculated as the mean of all non-zero amplitude values. For each syllable, a quadratic fit of the f_0 normalized values also was calculated. Within each of the 4 tones' sub-groups, the average syllable amplitudes were categorized as low, medium, or high, such that each amplitude category had equal numbers of syllables within a tone sub-group, that is, each amplitude grouping contained 1/3 of the syllables within that tone. An average duration was also calculated over all syllables within each subgroup, resulting in 12 average syllable duration values. For each quadratic contour, the time dimension then was compressed or extended linearly to the average duration within its tone and amplitude subgroup, in order to avoid averaging across inconsistent segments of each syllable. Within each of the 12 tone and amplitude sub-groups, a model quadratic contour then was calculated as the average of the f_0 values over all quadratic contours within that subgroup, which, because of the normalization to average duration, is accomplished by averaging over the quadratic coefficients. Quadratic contours provide better overall model representations of average syllable shape than simple linear slopes, especially for Tones 2 and 3, because of the varying curvatures of the lexically defined f_0 values of the different tones.

The resulting quadratics are plotted in Figures 9-12. The model pitch slope shapes for spontaneous speech depicted in these figures corroborate the above conclusion that higher amplitude is associated with greater conformity to the tonal target: Tone 1 becomes higher and flatter and Tones 3 and 4 fall more with higher amplitude. Only Tone 2 breaks this pattern: higher amplitude Tone 2 becomes higher in average pitch, but exhibits a relatively greater fall in slope. The average durations are similar across amplitude groups for each tone, except for a slightly shorter average duration for low and mid amplitude syllables for Tone 3 tokens, so that the amplitude effect is not due to greater duration tokens having a fuller manifestation of the lexical shape. The model shapes seen in Figures 9-12 may provide a partial explanation of the similarity of slope between different tones, as seen previously with Tone 1 and Tone 4 in Table 5. The low and mid amplitude model shapes are more similar than dissimilar across tones, and this will cause greater similarity in the overall averages.



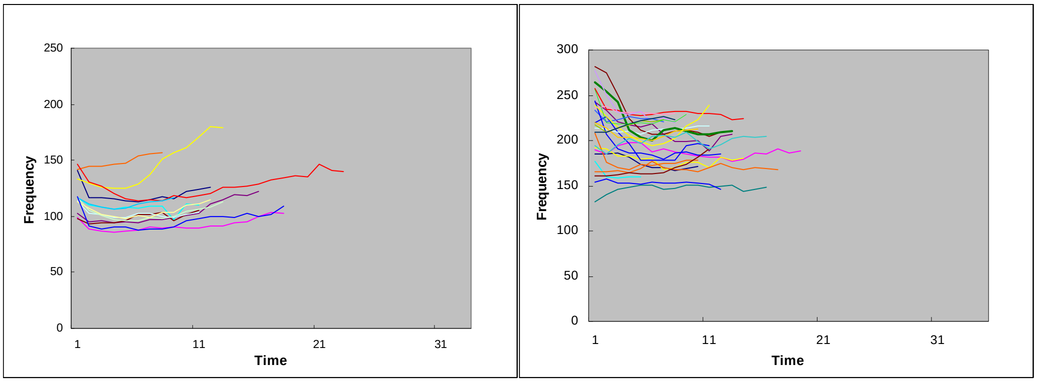
Figures 9-12. Slope of f_0 for Tones 1-4 by amplitude, spontaneous speech, in upper left, right, lower left, right order. x-axis= normalized syllable duration; y-axis= f_0 in Hz

The greater overall conformity to defined shape for read shape may arise because of the relative lack of other contextual influences. However, the greater conformity for greater amplitudes f_0 for both read and spontaneous speech must arise from a different cause, as it also occurs in the highly contextualized environment of interactive spontaneous conversation. Amplitude frequently is associated with *emphasis*, and emphasis arises naturally in both read and spontaneous speech. The above results suggest that one way to give emphasis to a lexical item in a tonal language, such as Mandarin, is to provide a more prominent and distinct lexical tone shape that marks the lexical meaning as more salient.

3.4 Visualizing Tonal Variation in Read and Spontaneous Speech

When viewed without respect to amplitude level, Figures 1-8 also give us a clearer picture of the differences between read and spontaneous speech. A comparison of the read to spontaneous speech in these figures shows that, although the pitch level varies more in spontaneous speech, the level nature of Tone 1 is similar in both read and spontaneous speech.

To analyze the differences between read and spontaneous in greater detail, we compared spontaneous vs. read speech for a number of identical tokens that were among the most frequently used syllables in both the read and the spontaneous corpora. This comparison suggested that a difference in speech mode does not affect all tones equally. The contours from read speech for identical Tone 1 tokens substantiated the result seen in Figures 1 and 5 above that Tone 1 remains essentially flat in both read and spontaneous speech. For Tones 2-4, however, there were more noticeable differences between the speech modes, as seen in the following Tone 2 *hai* 'still' example.



Figures 13-14. f_0 contours of 2^{nd}Tone 'hai' in read (left) and spontaneous speech (right)

3.4.1 Data Example 1: Tone 2 *hai*

The f_0 contours for Tone 2 *hai* 'still' of Figure 13 for read speech show clearly the defined rising nature of Tone 2, while, in the spontaneous speech shown in Figure 14, the pitch contours, for the most part, are much flatter, after an initial drop. Similar results were found for Tone 2 with other tokens. This *flattening* effect for spontaneous Tone 2 can also be seen from Figures 2 and 6 above, with a greater proportion of read speech slope measures in the rising range above zero than spontaneous Tone 2. A similar result occurs for Tones 3 and 4: the read speech slopes depicted in Figures 7 and 8 reflect an overall falling slope for almost all Tone 3 and Tone 4 read syllables, while Figures 3 and 4 show that spontaneous Tones 3 and 4 have greater numbers of syllables that *fall less* and frequently are *rising*.

These results (also see the results in Appendix A) suggest that, while spontaneous speech has more variability in pitch height and pitch contour, it also has a greater *flattening* effect on f_0 adherence to the defined tonal contours. Tone 1 remains overall flatter in spontaneous speech than in read speech. Spontaneous Tone 3 is similar to read speech, while Tone 2 is distinctly flatter in spontaneous speech than in read speech. Spontaneous Tone 3 is less likely to have an ending rise, and Tone 4 falls less than in read speech.

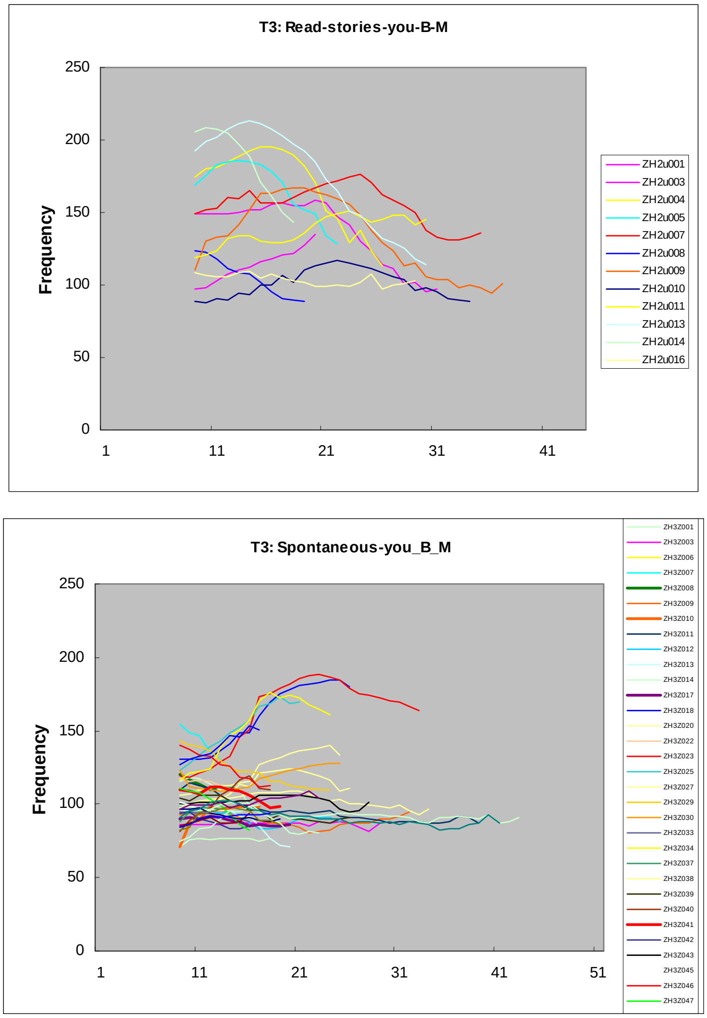
Our analysis suggests that it is the greater multi-functional usage of f_0 variation in spontaneous speech that leads to these results. In conversation, there is a greater use of pitch to indicate topic, provide signals of emotional and cognitive state, and to signal interactive intentions, and that the 'flattening' or reduced adherence to the lexically defined shape may be a more efficient use of lexical pitch so that a greater proportion of f_0 variability for the discourse functions is adopted.

3.4.2 Data Example 2: Tone 3 *you*

In Figures 15 and 16, we show individual instances of *you3* 'have/has' or 'is' for the male speaker as read speech in Figure 15 and in spontaneous conversation in Figure 16. *You* in Mandarin not only has a very high frequency of use, but frequently occurs in environments where cognitive or emotional intensity is high, such as in both questioning and in answers to questions, when uncertainty of information is present. On the other hand, *you* also occurs as a unremarkable syntactic object within a stream of more relevant lexical items or simple statements of facts. From Figures 15 and 16, Tone 3 *you* achieves its lexical full fall-rise shape or half-tone falling shape very rarely for this speaker, even in read speech. Since the read speech is a story in this case, the rise-fall pattern that is most predominant in Figure 15 reflects the communicative functions of emphasis added to the underlying lexical token. In spontaneous speech, *you* is reduced in pitch range in most cases to a flattened pitch contour, with a small number of exceptions that rise and fall as in read speech. In neither the read nor the spontaneous case do the pitch shapes approximate the lexical shape. In spontaneous speech, *you* has a very high frequency of use in interactive exchange, comparable to 'has/have' and 'is' in English, and its frequency of use in matter-of-fact statements de-emphasizes the need for a prominent signal for lexical comprehension; thus flatter pitch shape would not hinder comprehension and de-emphasis may aid in placing focus on the informative lexical target. Therefore, high frequency tokens, such as *you*, may rarely hit their defined shape: on one hand, de-emphasis flattens the contour, while, when communicating a cognitively or emotionally high intensity content, the commonplace nature of this token may be what allows it to take on primarily communicative prosody.

Spontaneous Tone 3 is similar to read speech Tone 3, while Tone 2 is distinctly flatter in spontaneous speech than in read speech. Spontaneous Tone 3 is less likely to have an ending rise, and Tone 4 falls less than in read speech. From research on experimental data, it has been proposed that flattening effects in speech are correlated with speech rate, and that shorter duration syllables should have greater flattening. The longer duration tokens for *hai* in Figure 13 that approximate continuous ending rise suggest that this may be a partial factor in flattening. However, in the corresponding spontaneous contours of Figure 14, there does not appear to be a correlation between short duration and flatness of slope. The flatter average slopes found for mid and low amplitude tones in the quadratic approximations in Figures 9-12 also do not appear to reflect the influence of duration, as, for all tones except Tone 3, all amplitude classes had approximately the same average duration.

Our preliminary analysis suggests that it may be the greater multi-functional usage of f_0 variation in spontaneous speech instead that leads to these results. In spontaneous speech, there is a greater use of pitch to indicate topic, provide signals of emotional and cognitive state, and to signal interactive intentions than in read speech, as exemplified in the emotional emphasis signaled in the high arching contours in Figure 16. The general 'flattening' or reduced adherence to the lexically defined shape may be a more efficient use of lexical pitch so that a greater proportion of f_0 variability for the discourse functions is adopted.

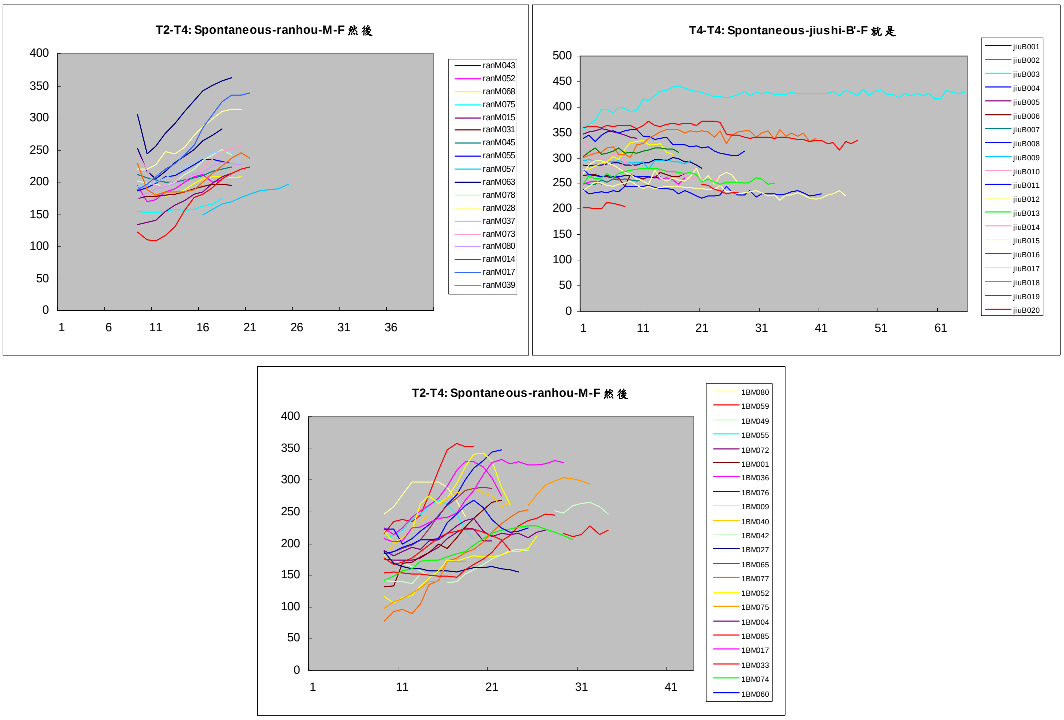


Figures 15-16. f_0 contours of 3rd Tone 'you' of the same speaker in read (top) and spontaneous speech (bottom)

3.5 Lexical Identity and Prosodic Shape

3.5.1 Data Example 3: *ranhou (2-4) & jiushi (4-4)*

It is illuminating to compare individual instances of tone shapes for specific lexical tokens. In Figures 17, 18, and 19, we compare tokens of two frequently used lexical items from our spontaneous corpus: 2-4 tone sequence *ranhou* '*then*', and 4-4 tone sequence *jiushi* '*it's just*'. The tokens illustrate how individual characters or words in Mandarin exhibit differing tendencies or abilities to vary their tonal shape, suggesting that a token's lexical identity and its propensity to vary its tonal pitch contour are closely related.



Figures 17-19. f_0 contours of 2-4 tone sequence ranhou & 4-4 tone sequence of jiushi in spontaneous speech, in top left, top right and bottom order

In conversation, *ranhou* has a number of frequently used functions: it can act as a simple, direct logical or temporal connector, linking events that are sequential in time or in logical progression, similar to '*then*' in English. As it links lexical and pragmatic meaning, it often takes on a rising form, where the falling Tone 4 '*hou*' has been transformed prosodically. This is especially evident in the very short rising slopes of Figure 17. However, when it functions to emphasize the temporal transition from one previous state to the following one, *ranhou* is often realized as a pitch sequence that mirrors its lexical rise-fall tonal shape, as in Figure 19.

Expressivity through pitch can also depend on speaker or speaker state, as well as the pragmatic communicative functions that a particular lexical token assumes. *Jiushi* is often used as a hesitation marker as a speaker retrieves information for the next step in the communicative sequence, and the long and flattened pitch slopes of Figure 18 frequently occur. When *jiushi* is used to signal varying degrees of affirmation and agreement, the certainty and emphasis is reflected in a greater adherence to falling Tone 4 *shi* as the final syllable.

These examples illustrate a number of important points that provide insight into both the opportunities and the constraints governing tonal shape in spontaneous speech. The very high frequency and familiarity of both *ranhou* and *jiushi* imply that there is low cognitive effort in auditory interpretation, and this feature allows a high reduction in pitch shape under rapid speech or coarticulation. This is most likely to occur when these words are used as simple links between statement sequences that are the semantic focal targets of the communication. Moreover, as links that have specific lexical function, the familiarity and high frequency enable both *ranhou* and *jiushi* to more fully take on prosody that expresses emotional and cognitive qualifications on the nature of the linkage, especially including the marking of hesitation through duration and pitch shape. Conversely, when the meaning in the lexical link itself is the focus, emphasis on that lexical meaning is reinforced through a high, even exaggerated, use of the defined lexical shapes. Thus, the propensity of a given token to conform to defined tonal shapes may be conditioned by its frequency of use and its lexical identity, as well as its intrinsic ability to take on a number of different pragmatic communicative functions.

4. Conclusion

In this study, we have presented a framework for characterizing tonal variability over spontaneous conversations and have compared the degree of adherence to defined tonal values in read and spontaneous Mandarin. The results show that realized tonal pitch shapes in spontaneous speech have greater internal variance as well as greater overall divergence from defined pitch shapes. We have investigated several factors that contribute to this divergence, and we have shown evidence that corroborates, for spontaneous conversations, prior research on the existence and importance of anticipatory and carryover effects. The numerous exceptions to sequencing effects indicate that these effects were not *sufficient* to account for the wide variability of tonal shapes in spontaneous speech. A key finding of the study is that adherence to a defined shape is greater when syllable *amplitude* is relatively high. The study also identified two tendencies for pitch shape in spontaneous speech. First, pitch slope tends to have more negative slopes, compared to the defined shape for Tones 1, 2, and 3. A second related tendency is that slopes in spontaneous speech operate under a flattening tendency, with 4th Tones less steep than in read speech and with level and rising tones more negative.

The current study provides an initial platform from which to extend research to the role of conversational contextual factors, including cognition, emotion, and communicative function, on realized tonal shapes, and we have shown initial evidence on the importance of lexical meaning and speaker state to tonal and prosodic variation. Finally, results of the study indicate that, although there is great variability in individual tone pitch shape, there are also systematic relationships among the tones and in tone sequences that are dependent on tonal identity. Thus, our results suggest that the diverse variations in realized tonal shape are evidence of the great ability of Mandarin to simultaneously express both lexical meaning and speaker state in a unified system of lexical and prosodic form and demonstrate the high potential for expressive prosody in Mandarin.

Acknowledgements

This work was supported in part by the National Science Council of Taiwan under grant number 100-2410-H-029-039 to the first author. We would like to thank the two anonymous reviewers whose comments have substantially improved the paper. We would also like to thank Dr. Shu-Chuan Tseng and Dr. Chiu-yu Tseng of Academia Sinica for very helpful discussions while the first author was a Visiting Scholar at the Institute of Linguistics, Academia Sinica, Taiwan, during the year 2012.

References

Chang, Y.-C, & Hsieh, F.-F. (2012). Tonal coarticulation in Malaysian Hokkien: A typological anomaly? *The Linguistic Review, 29*, 37-73.

Chao, Y. R. (1968). *A grammar of spoken Chinese*, Berkeley, University of California Press.

Gussenhoven, C. (2004). *The phonology of tone and intonation*, Cambridge University Press.

Hirst, D. & Di Cristo, A. eds. (1998). *Intonation systems: a survey of twenty languages*, Cambridge, Cambridge University Press.

Hsieh, F.-F. (2008). Preservation of the marked as slope correspondence in Hangzhou Chinese disyllabic tone sandhi. *Interfaces in Chinese Phonology*, 223-242.

Rose, P. (2012). Two sides of the same coin: between - speaker F0 differences in linguistic-phonetic description and forensic voice comparison. TAL 2012. *Interanational Conference on Tonal Aspects across Tone and Non-tone Languages*, keynote speech, Nanjin, China.

Tseng, C.-Y. (2010). Beyond sentence prosody. Keynote speech, In *Proceedings of Interspeech 2010*, Makuhari, Japan, 20-29.

Tseng, C.-Y. (2010). An f0 analysis of discourse construction and global information in realized narrative prosody. *Language & Linguistics, 11*(2), 183-218.

Tseng, S.-C. (2004). Processing spoken Mandarin corpora. *Traitement automatique des langues. Special Issue: Spoken Corpus Processing*, *45*(2), 89-108.

Tseng, S.-C. (2005a). Syllable contractions in a Mandarin conversational dialogue corpus. *International Journal of Corpus Linguistics*, *10*(1), 63-83.

Tseng, S.-C. (2005b). Mandarin topic-oriented conversations. *International Journal of Computational Linguistics and Chinese Language Processing. Special Issue: Annotated Speech Corpora*, *10*(2), 201-218.

Tseng, S.-C. (2008). Spoken corpora and analysis of natural speech. *Taiwan Journal of Linguistics*, *6*(2), 1-26.

Tseng, S.-C. (2013). Lexical Coverage in Taiwan Mandarin. *International Journal of Computational Linguistics and Chinese Language Processing*, *18*(1), 1-18.

Shen, X. S. (1990). *The Prosody of Mandarin Chinese*, University of California Press.

Shih, C. & Sproat, R. (1992). Variations of the Mandarin rising tone. *IRCS workshop on prosody in natural speech*, 193-200.

Shriberg, E., Stolcke, A., Hakkani-Tur, D. & Tur, G. (2000). Prosody-Based automatic segmentation of speech into sentences and topics. *Speech Communiccation*, *32*(1-2), 127-154 (Special Issue on Accessing Information in Spoken Audio).

Whalen, D. H. & Xu, Y. (1992). Information for Mandarin tones in the amplitude contour and in brief segments. *Phonetica*, *49*, 25-47.

Xu, Y. (1997). Contextual tonal variations in Mandarin. *Journal of Phonetics*, *25*, 61-83.

Xu, Y. (2011). Functions and mechanisms in linguistic research -- Lessons from speech prosody. In *Proceedings of Workshop on Experimental Linguistics*. Paris: 1-10.

Xu, Y., & Wang, Q. (2001). Pitch targets and their realization: evidence from Mandarin Chinese. *Speech Communication*, *33*(4), 319-337.

Yang, L.-C. (1995). *Intonational Structures of Mandarin Discourse*, Ph.D. dissertation, Georgetown University.

Appendix A: Supplementary Statistical Results on Tonal Variation

Table A1 compares the distributional spread of spontaneous speech to read speech of Speaker B. For Tone 1, the standard deviation for spontaneous speech is about twice as large as for read speech, with the kurtosis values for spontaneous tones being greater than for read speech, especially for Tones 1 and 4, indicating a greater occurrence of extreme variations in slope for spontaneous speech.

Table A1. Comparison of syllable of f_0 slope, read speech to spontaneous speech, Speaker B

	Read speech				Spontaneous speech			
	Tone 1	Tone 2	Tone 3	Tone 4	Tone 1	Tone 2	Tone 3	Tone 4
Mean	-29.82	36.95	-22.61	-271.06	-84.33	-14.45	27.20	-95.64
Median	-22.92	51.12	-76.22	-260.38	-62.76	-5.99	16.33	-75.03
SD	155.20	232.12	457.43	319.33	328.65	247.37	334.43	297.83
Kurtosis	9.51	7.34	6.65	1.92	21.11	8.82	8.90	14.49
Skewness	-1.20	-1.47	0.54	0.35	-1.66	-0.62	0.16	0.84
R^2	0.0005	2e-07	0.019	0.1419	0.0419	0.0078	0.0075	0.0358

The skewness and large kurtosis values for both read and spontaneous speech show significant departures from normality, so non-parametric Wilcoxon rank sum tests were computed to test the hypothesis of no shift in average slope between read and spontaneous speech. Table A2 presents the results of the pair-wise Wilcoxon tests between read and spontaneous syllable slope by tone for Speaker B. The very low p-values shown indicate that the alternative hypothesis of a change in syllable slope holds at high significance levels and indicate the existence of systematic contour differences between spontaneous and read speech.

Table A2. Wilcoxon rank sum test comparing read to spontaneous speech by tone, Speaker B

Read vs. Spontaneous	W-value	p-value
Tone 1	73760	6.035e-07
Tone 2	52604	4.882e-06
Tone 3	45319	0.0001075
Tone 4	97593	< 2.2e-16

The Wilcoxon results corroborate the shifts in syllable slope between read and spontaneous speech shown in Table A1. The decrease in median slope of spontaneous speech for Tones 1 and 2 and the increase in median slope for Tones 3 and 4 suggest a tendency towards flatter slopes, on average, in spontaneous speech. The exception to this is Tone 1, which tends to fall more in spontaneous speech than in read speech.

Computational Linguistics and Chinese Language Processing
Vol. 18, No. 3, September 2013, pp. 45-58

Acoustic Correlates of Contrastive Stress in Compound Words versus Verbal Phrase in Mandarin Chinese

Weilin Shen[*], Jacqueline Vaissière[+,#], and Frédéric Isel[*,+,#]

Abstract

Duanmu (2000) proposed that tonal languages, such as Chinese, follow the same Compound and Nuclear Stress Rules (Chomsky & Halle, 1968) for phrasal stress as English. This study investigates the acoustic correlates of contrastive stress between compound words and verbal phrases in Mandarin Chinese. We focused on the durational, fundamental frequency, and intensity correlates of stress within minimal pair MN modifier-head compounds and VO verb-object phrases. Our results demonstrated that (1) the final syllable was more lengthened in [VO] than in [MN] and that (2) the F_0 range was larger in [VO] than in [MN]. Moreover, the duration of the pause between the two syllables seems to play a role in distinguishing between [MN] and [VO]. In contrast, we showed that intensity contributed less to this distinction. Our results confirmed the right stress pattern in [VO]; however, we failed to find the lexical stress on the Left syllable we had expected, at least with the speakers we examined. Taken together, the present acoustic study lends support to the hypothesis that principles of stress upward of word level are universal through different languages.

Keywords: Morpholexical Ambiguity, Compounding, Compound versus Nuclear Stress, Acoustic Features.

1. Introduction

In stress languages, such as in English, most words have stable lexical stress patterns and it is often easy to tell which syllables have stress. For a typical tonal language, *e.g.* Mandarin, the word stress is often less obvious. Although, lexical stress has been shown to be highly

[*] Institute of Psychology, Paris Descartes University, France
[+] Laboratory of Phonetics and Phonology, University Sorbonne Nouvelle Paris 3, France
[#] Laboratoire d'Excellence EFL, PRES Paris Sorbonne, France
Phone: ++33 1 55205924 Fax: ++33 1 55205745
E-mail: frederic.isel@parisdescartes.fr
The author for correspondence is Frédéric Isel.

language-dependent, principles of stress upward of word level (*i.e.* compound stress and phrasal stress) are more universal in different languages. Chomsky and Halle (1968) proposed two rules for English compound and phrasal stress.

- **Compound stress rule**: stress is assigned to the leftmost stressable vowel in nouns, verbs, or adjectives, *e.g.* bláckbird.
- **Nuclear stress rule (NSR)**: stress is assigned to the rightmost stressable vowel in a major constituent, *e.g.* [the [black bírd]].

It has been proposed that the Compound stress rule and NSR are true for Mandarin Chinese, and they permit one to distinguish between compounds and phrases (Duanmu, 2000). Nevertheless, there is no empirical evidence supporting this hypothesis to date. The first goal of the present study was to understand to what extent morphology affects abstract stress using acoustic-phonetic evidence. Moreover, we were interested in discerning the acoustic phonetic cues, which reflect abstract stress. In Chinese, a V-N construction is sometimes ambiguous, possibly representing both a modifier-head compound [MN] and a verb-object phrase [VO]. For example, a V-N construction ‘chao-fan’ (fry-rice) may be a compound, in which the verbal constituent ‘chao’ (fry) modifies the nominal head ‘fan’ (rice); it may also represent a verb-object relation (to fry rice). The ambiguous pairs have the same segmental characteristics and are assumed to differ from each other only in the stress pattern, showing left stress for compounds and right stress for phrases. There is, however, no phonetic evidence that compound stress and phrasal stress are implemented in [MN] compounds and in [VO] phrases in Mandarin Chinese.

Fraisse (1956) proposed two basic rhythmic tendencies 1) “*rythmitisation intensive*,” sensitive to strengthening of the initial element, and 2) “*rythmitisationtemporelle*,” building on the lengthening of the final element. The supposed basic rhythmic tendencies predict initial extra loudness and final lengthening. From phonetic studies on the acoustic correlates of stress since the 1950s, researchers have agreed that linguistic stress correlates with a complex configuration of events of increased duration, larger F_0 range, and raised intensity (Lehiste, 1970) and that several cues may be functionally equivalent cross-linguistically (Vaissière, 2004).

Duration

From a series of experiments, Fry (1955, 1958) showed that duration is a consistent correlate of stress at the word level in English and that it is a more effective cue than intensity. Since then, researchers have started to give up the classical view that stress is equated to a higher degree of intensity. Studies on the neutral tone (*i.e.* destressed syllable) in Chinese have confirmed the crucial role of duration on the perception of a destressed syllable for Chinese. Lin (1980, 1990) and Cao (1992) showed that duration of the destressed neutral tone syllable is systematically shorter (reduced by approximately 50%) than a syllable with full tone.

Fundamental frequency

The F_0 has been showed to be a major acoustic manifestation of suprasegmental structures. It is claimed by some researchers to be the strongest cue of stress for stress languages (Cooper *et al.*, 1985; Lieberman, 1960; Gussenhoven *et al.*, 1997). Nevertheless, others have shown that F_0 is not a necessary cue because stress can be identified on the basis of duration and intensity alone (Cutler & Darwin, 1981). The situation is the same for tonal languages, such as Mandarin. The pitch range has been shown wider when syllables are stressed (Shen, 1985; Liu & Xu, 2005). More specifically, when a 3rd Tone is stressed, it is dipped lower and, when a 4th Tone is stressed, it starts higher and falls lower (Chao, 1968). Moreover, computational corpus studies (Kochanski *et al.*, 2003) have established quantitative F_0 predictions in terms of the lexical tones and the prosodic strength of each word. Shen (1993), however, found that stress in Mandarin could be identified without F_0 information.

Intensity

In literature, the role of intensity for stress is not agreed upon. Fry (1955, 1958) showed that intensity was a less effective cue than duration on the perception of linguistic stress patterns. Nevertheless, some authors have argued that the strongest cue to prominence is intensity for English (*e.g.*, Beckman, 1986; Turk & Sawusch, 1996). For Mandarin Chinese, the effect of the intensity is only secondary. Studies on the neutral tone in Chinese showed that the intensity of the destressed neutral tone is not necessary lower than the one with full tone (Cao, 1986). Moreover, the destressed neutral tone raises its intensity after Tone 3 (Lin, 2006). Phonetic data (Cao, 1992) has illustrated that the destressing of the neutral tone syllable is not related simply to its intensity. The intensity of a neutral tone syllable is lower than that of one with full tone in general, but the situation is reversed when it is preceded by a Tone 3 syllable.

The present study investigates the acoustic correlates of stress between compound and phrase in Mandarin Chinese. We focused on the durational, fundamental frequency, and intensity correlates of stress within minimal pair [MN] modifier-head compound and [VO] verb-object phrase. Our hypotheses were that 1) [MN] modifier-head compound and [VO] phrases differ phonetically with left stress in [MN] modifier-head compounds and right stress in [VO] phrases and that 2) a different prosodic pattern is reflected in acoustic features in F_0, duration and intensity.

2. Methodology

2.1 Materials

One hundred thirty-five minimal pairs presenting a morpholexical ambiguity (*i.e.* [MN] modifier-head compound vs. [VO] phrases) were selected from the Contemporary Chinese Dictionary 5th edition (Lu & Ding, 2008). Each pair had the same segmental characteristics and was assumed to differ from each other only in the stress pattern. The target words were not recorded in isolation and were embedded in an utterance fragment:

1)我说的不是名词编号而是动词编号.
[I did not say noun "bian-hao" but said verb "bian-hao".]

The critical words in each pair change their position in the utterance fragment, giving

2)我说的不是动词编号而是名词编号.
[I did not say verb "bian-hao" but said noun "bian-hao".]

In all, 270 sentences were created. The order of the sentences was randomized.

2.2 Recording Procedure

Before the recording session, the participants were instructed in the goal of the recording and how the recording would proceed. The material was carried out in the laboratory of Phonetics and Phonology of University Sorbonne Nouvelle Paris 3. Speakers were recorded individually in an acoustic chamber, using an attached microphone, placed at a distance of about 5 centimeters from the speaker's mouth. Speech samples were recorded digitally at 44,100 Hz, 16-bit mono.

2.3 Subjects

Three Mandarin speakers (two females) in Paris participated in the experiment. One female speaker is an international student aged 25 years that was born in Xi'an, China. Her mother tongue and language of schooling is Mandarin. The others speakers are Beijing Mandarin speakers (one female 26 years; one male 32 years).

2.4 Acoustic Measurements

The first syllable, the second syllable, and the pause between them for each critical word were manually marked in Praat, yielding four marks, one at the beginning of the first syllable, a second mark at the offset of the first syllable, a third mark between the offset of the first syllable and the onset of the second syllable, and a fourth one at the offset of the second syllable. A Praat script extracted the duration and intensity value of each segment in msec. F_0 onset and offset were measured at the beginning and at the end of the vowel. In the study, we divided the vowel into ten segments normalized in time, with the mean F_0 of the first segment as F_0 onset and the mean F_0 of the last segment as F_0 offset.

3. Results

Three-way repeated analysis of variance ANOVA tests were performed separately for each acoustic feature (duration, F_0, and intensity). Word type ([MN] modifier-head compound vs. [VO] verb-object phrase) and syllable position (left syllable: S1 vs. right syllable: S2) were the within groups factors, and word position in the utterance fragment (*i.e.* final vs. non-final) was the between groups factor.

3.1 Duration

3.1.1 Left Syllable vs. Right Syllable

Results of the three-way ANOVA for the duration revealed a significant main effect for word type [$F(1, 134) = 440.8$, $p<0.001$; $\eta_p^2= 0.77$], a significant main effect for word position [$F(1, 134) = 25.6$, $p< 0.001$; $\eta_p^2 = 0.16$], a significant main effect for syllable position [$F(1, 134) = 105.3$, $p< 0.001$; $\eta_p^2 = 0.44$], a significant interaction word type x syllable position [$F(1, 134) = 87.4$, $p< 0.001$; $\eta_p^2 = 0.40$], and a significant interaction word position x syllable position [$F(1, 134) = 440.8$, $p< 0.001$; $\eta_p^2 = 0.77$]. Word position showed no interaction with word type. In order to increase the statistical power, we token the word position out, and ran a two-way ANOVA (word type x syllable position). The two-way ANOVA showed significant main effect for word type [$F(1, 269) = 846.7$, $p< 0.001$; $\eta_p^2 = 0.64$] and for syllable position [$F(1, 269) = 166.3$, $p< 0.001$; $\eta_p^2 = 0.38$] and a significant interaction word position x syllable position [$F(1, 269) = 217.1$, $p< 0.001$; $\eta_p^2 = 0.45$]. *Post hoc* analyses showed a larger effect of syllable position for [VO] [$F(1, 269) = 217.0$, $p< 0.001$; $\eta_p^2 = 0.45$] than for [MN] [$F(1, 269) = 28.6$, $p< 0.001$; $\eta_p^2 = 0.10$].

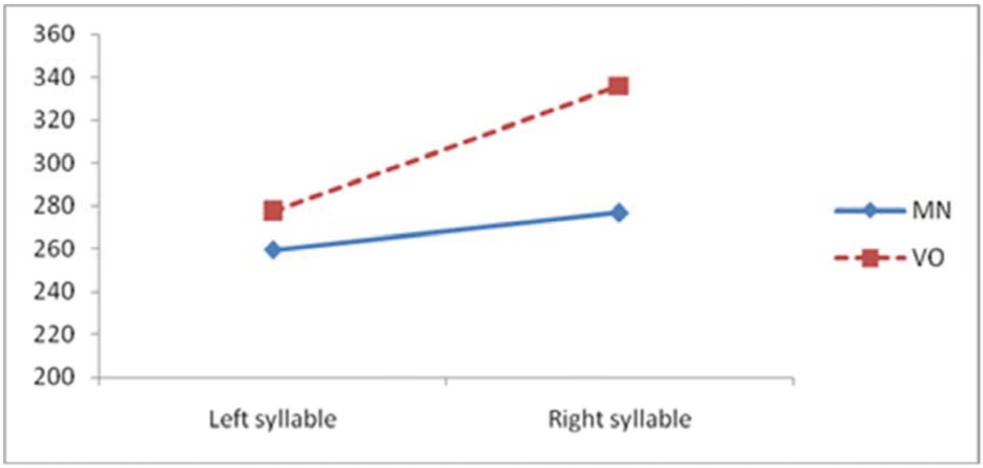


Figure 1. Mean syllable durations in msec for each syllable in [MN] and [VO].

3.1.2 Duration of the Pause

As one can notice a pause between the two syllables in VO, we decided to perform measures of pause duration. The ANOVA on the average pause duration between the syllables showed a significant main effect for word type [$F(1, 269) = 33.5$, $p< 0.001$; $\eta_p^2 = 0.11$].

3.2 F_0

The F_0 was analyzed separately for each of the four tones. A three-way ANOVA with word type (MN vs. VO), syllable position (Left syllable vs. Right syllable), and measure point (onset vs. set) was applied to Tone 1 and Tone 4, and a two-way ANOVA with word type (MN vs. VO) and syllable position (Left syllable vs. Right syllable) was calculated on the difference between F_{0max} and F_{0min} for Tone 2 and Tone 3.

3.2.1 Tone 1

Results showed a significant main effect for measure point [$F(1, 45) = 10.7$, $p<0.01$; $\eta_p^2 = 0.19$], a significant interaction between word type and syllable position [$F(1, 45) = 4.7$, $p<0.05$; $\eta_p^2 = 0.10$], and a significant interaction between measure point and syllable position [$F(1, 45) = 6.6$, $p<0.05$; $\eta_p^2 = 0.13$]. Nevertheless, neither significant interaction between word type and measure point [$F< 1$], nor significant interaction between word type, syllable position, and measure point [$F< 1$] was found.

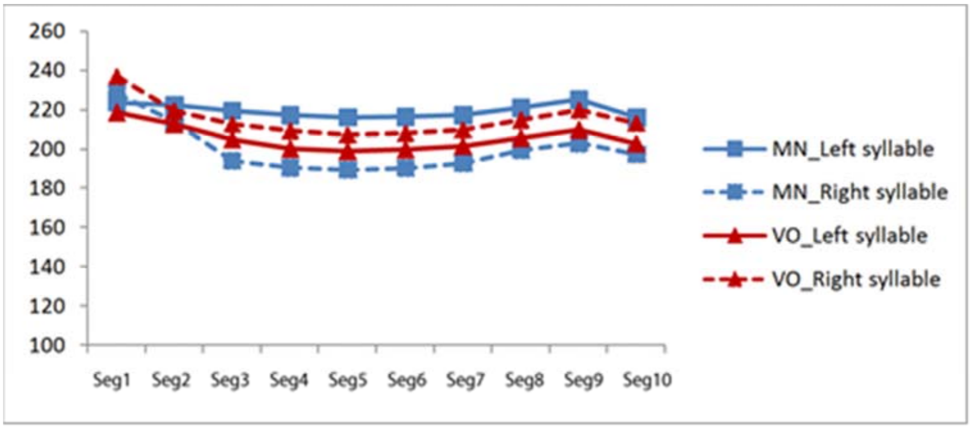


Figure 2. F_0 values on ten segments for Tone 1 for Left and Right syllable in [MN] and [VO].

3.2.2 Tone 2

Neither significant main effect for word type and syllable position [*F*< 1] nor significant interaction [*F*< 1] was found on the difference between F_{0max} and F_{0min}.

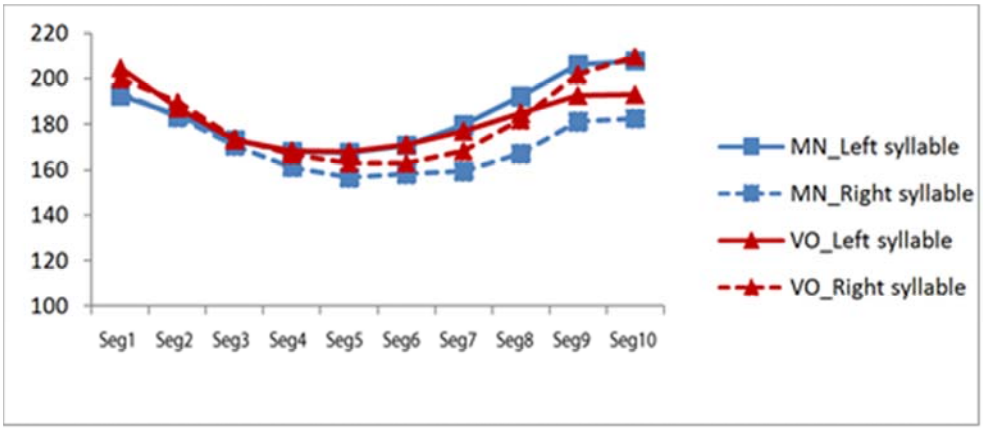


Figure 3. F_0 values on ten segments for Tone 2 for Left and Right syllable in [MN] and [VO].

3.2.3 Tone 3

In order to not confound tone sandhi influence for these analyses, we took out two items in our experimental material with a Tone 3-Tone 3 combination. The two-way ANOVA on the difference between F_{0max} and F_{0min} revealed a significant interaction word type x syllable position [$F(1, 53) = 217.1$, $p< 0.001$; $\eta_p^2 = 0.31$]. *Post hoc* analyses showed a larger effect of syllable position for [VO] [$F(1, 53) = 37.9$, $p< 0.001$; $\eta_p^2 = 0.42$] than for [MN] [$F(1, 53) = 37.9$, $p< 0.05$; $\eta_p^2 = 0.80$].

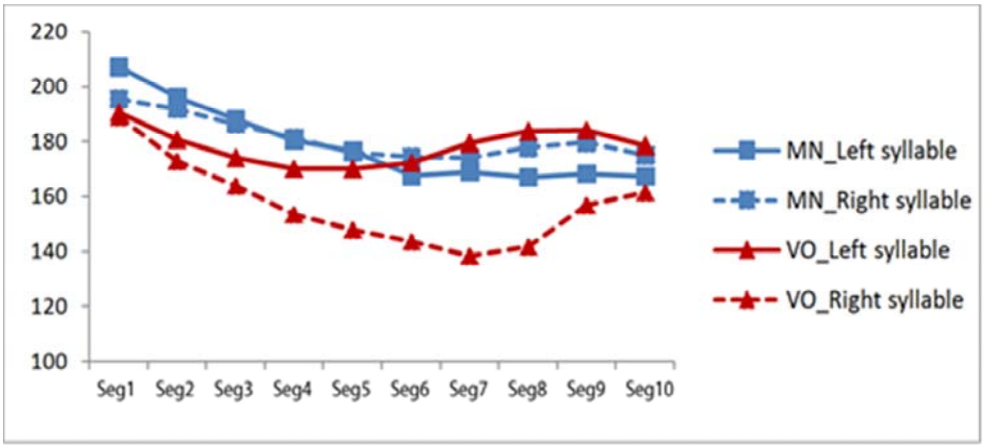


Figure 4. F_0 values on ten segments for Tone 3 for Left and Right syllable in [MN] and [VO].

3.2.4 Tone 4

Results showed a significant main effect for measure point [$F(1, 91) = 339.1$, $p<0.001$; $\eta_p^2 = 0.79$] and for syllable position [$F(1, 91) = 14.0$, $p<0.001$; $\eta_p^2 = 0.13$], a significant interaction between word type and syllable position [$F(1, 91) = 23.5$, $p<0.001$; $\eta_p^2 = 0.21$], a significant interaction between measure point and syllable position [$F(1, 91) = 14.4$, $p<0.001$; $\eta_p^2 = 0.14$], and a significant interaction of word type x measure point x syllable position [$F(1, 91) = 6.8$, $p<0.05$; $\eta_p^2 = 0.07$]. *Post hoc* analyses showed a main effect of syllable position for the Left syllable of [VO] [$F(1, 91) = 36.8$, $p< 0.001$; $\eta_p^2 = 0.29$], however, there was no main effect of syllable position for the Left syllable of [VO] [$F<1$].

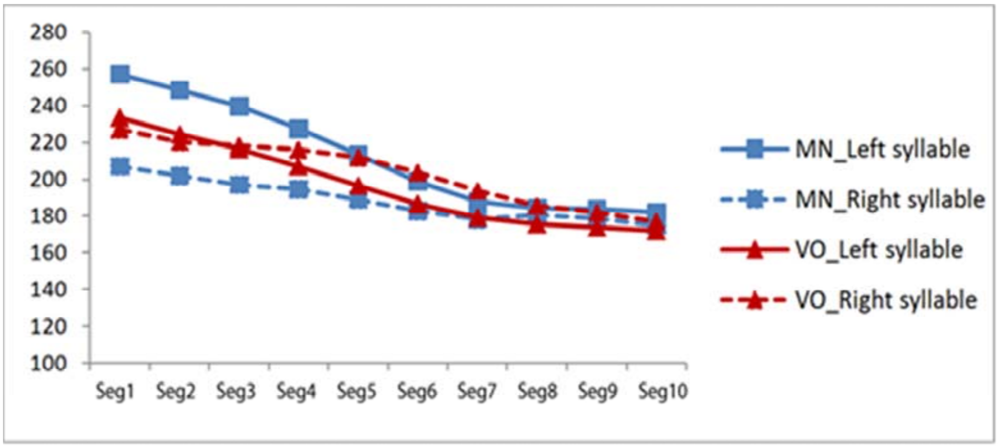


Figure 5. F_0 values on ten segments for Tone 4 for Left and Right syllable in [MN] and [VO].

3.3 Intensity

A three-way ANOVA with repeated measure was performed on the average intensity. Results showed a significant main effect for word type [$F(1, 134) = 139.0$, $p< 0.001$; $\eta_p^2 = 0.64$], a significant main effect for word position [$F(1, 134) = 234.8$, $p< 0.001$; $\eta_p^2 = 0.16$], a significant main effect for syllable position [$F(1, 134) = 58.9$, $p< 0.001$; $\eta_p^2 = 0.31$], a significant interaction word type x syllable position [$F(1, 134) = 87.4$, $p< 0.001$; $\eta_p^2 = 0.40$], and a significant interaction word position x syllable position [$F(1, 134) = 440.8$, $p< 0.001$; $\eta_p^2 = 0.77$]. No interaction was found.

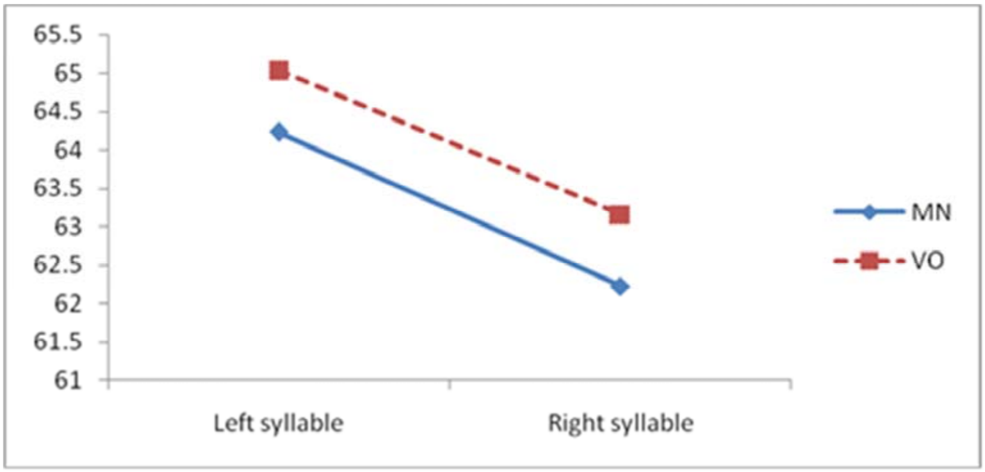


Figure 6. Mean intensity (dB) for each syllable in [MN] and [VO].

4. Discussion

This article investigated the acoustic correlates of linguistic stress on the ambiguous structure Verb-Noun (*i.e*. [MN] vs. [VO]) in Mandarin Chinese. Moreover, the acoustic feature associated with this stress pattern was analyzed. As explained in the introduction, duration, F_0, and intensity are the main correlates of stress. Results showed the implication of duration, F_0, and intensity in the production of compound and phrasal stress in Mandarin.

Our preliminary data showed that the duration was longer for the right syllable in [VO], which was consistent with previous studies on the acoustic correlates of linguistic stress for stress languages and for tone languages, such as Mandarin. Nevertheless, the ‘assumed stressed’ left syllable in [MN] was not longer than the Right syllable. We also performed measures of pause duration, and the results on the average pause duration between the Left and Right syllables showed that average pause duration is longer in [VO] than in [MN]. Nevertheless, we considered that this larger pause duration was not an acoustic manifestation of stress but a mark of the syntactic boundary in the verb-object phrase.

Despite the fact that, in tone languages, F_0 information should be attributed to its lexical usage, our results showed that F_0 would be a reliable cue for the stress pattern in [MN] and [VO]. The F_0 range was shown to link to the stress for Tone 3 and Tone 4, which was in line with the predictions (Chao, 1968) that pitch range is wider for stressed syllables, specifically, when a 3rd Tone is stressed, it dips lower, and, when a 4th Tone is stressed, it starts higher. Our results showed that, for Tone, 3 the right syllable in [VO] had a larger F_0 range than the left syllable. For Tone 4 the left syllable in [MN] showed higher onset F_0 than the right one.

The analyses on the intensity were in line with previous studies, which showed a less important role of the intensity for stress. In our preliminary data, the intensity was shown to have larger amplitude in [VO] than in [MN] for the two syllables. Nevertheless, we failed to find the strengthening of the Left syllable in [MN], as proposed by Fraisse, that left-headed feet should show extra loudness on the initial syllable than the second initial. Our results showed the same pattern of intensity between [VO] and [MN]. Therefore, we considered that, unless the [VO] and [MN] were presented together, the intensity was not an effective cue for distinguishing between [VO] and [MN].

In sum, our preliminary data suggested an implementation of the final lengthening for the stressed syllable in [VO], but no initial extra loudness in [MN]. The F_0 information suggested that, for Tone 3, the Right syllable was stressed in [VO] and, for Tone 4, the Left syllable was stressed in [MN]. The results confirmed the right stress pattern in [VO]; however, with the only support in Tone 4, we did not consider a lexical stress on the Left syllable in [MN].

The prosodic information, such as stress, duration, and pause was shown to be critical for the processing of the compound words (Isel *et al*., 2003). Once we have shown that compound word and verbal phrase present different acoustic patterns with respect to the position of stress, the next step would be to verify whether this stress pattern is used by the listeners to differentiate the two forms in cases of segmental ambiguities. For this purpose, we plan to conduct different perception and categorization experiments. At the same time, more speakers would be added to the production study.

5. Conclusion

Our results showed a right stress pattern in [VO] with longer duration in the Right syllable, larger range F_0, and longer pause duration between the syllables; in contrast, no initial strengthening in [MN] was found. Only the F_0 range information in Tone 4 supported a lexical stress on the Left syllable in [MN].

Acknowledgments

We are grateful to the anonymous reviewers whose insightful suggestions have appreciably improved the manuscript. We also thank Mr. Joe Harwood for the correction of English.

References

Boersma, P. (2001). Praat, a system for doing phonetics by computer. *Glot International, 5*(9/10), 341-345.

Cao, J. f. (1992). On Neutral-Tone Syllables in Mandarin Chinese. *Canadian Acoustics*, *20*(3).

Chao, Y. R. (1968). *A Grammar of Spoken Chinese*. Berkeley, CA: University of California Press.

Chomsky, N., & Halle, M. (1968). *The sound pattern of English*. New York: Harper and Row.

Cooper, W. E., Eady, S. J., & Mueller, P. R.. (1985). Acoustical aspects of contrastive stress in question-answer contexts. *Journal of the Acoustical Society of America*, *77*, 2142-2156.

Duanmu, S. (2000). *The phonology of Standard Chinese*. Oxford: Oxford University Press.

Fraisse, P. (1956). *Les structures rhythmiques*. Louvain: Publication Universitaires de Louvain.

Fry, D. B. (1955). Duration and intensity as physical correlates of linguistic stress. *Journal of the Acoustical Society of America*, *27*(4), 765-768.

Fry, D. B. (1958). Experiments in the perception of stress. *Language and Speech*, *1*, 126-152.

Gussenhoven, C., Repp, B. H., Rietveld, A. C. M., Rump, H. H., & Terken, J. (1997). The perceptual prominence of fundamental frequency peaks. *Journal of the Acoustical Society of America*, *102*, 3009-3021.

Isel, F., Gunter, T. C., & Friederici, A. D. (2003). Prosody-assisted head-driven access to spoken German compounds. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *29*, 277-288.

Kochanski, G., Shih, C., & Jing, H. Y. (2003) Hierarchical structure and word strength prediction of Mandarin prosody. *International Journal of Speech Technology*, *6*(1), 33-43.

Lehiste, I. (1970). *Suprasegmentals*. Cambridge, Mass.: M.I.T. Press.

Lieberman, P. (1960). Some acoustic correlates of word stress in American English. *Journal of the Acoustical Society of America*, *33*, 451-454.

Lin, M. C., & Yan, J. Z. (1990). Putonghua qingsheng yu qingzhong yin [The neutral tone and stress in Mandarin.]. *Yuyan Jiaoxue yu Yanjiu [Language Teaching and Linguistic Studies]*, *1990*(2), 88-104.

Lin, H. (2006). Mandarin Neutral Tone Is a Phonologically Low Tone. *Journal of Chinese Language and Computing*, *16*(2), 121-134.

Lin, M. C., & Yan, J. Z. (1980). The acoustic characteristics of neutral-tone syllables in Standard Chinese. *Fangyan*, *3*, 166-178.

Liu, F., & Xu, Y. (2005). Parallel Encoding of Focus and Interrogative Meaning in Mandarin Intonation. *Phonetica*, *62*, 70-87.

Lu, S. X., & Ding, S. S. (2008). *Modern Chinese Dictionary* (5th ed.). Beijing: Commercial Press.

Shen, J. (1985). *Beijinghua shengdiaode yinyu he yudiao*. [Pitch range of tone and intonation in Beijing dialect]. In: Beijing Yuyin Shiyan Lu [*Working Papers in Experimental Phonetics*]. Beijing: Beijing Daxue Chubanshe. 73-130.

Shen, X. S. (1993). Relative duration as a perceptual cue to stress in Mandarin. *Language and Speech*, *36*, 415-433.

Vaissière, J. (2005). Perception of intonation. In D. B. Pisoni & R. E. Remez (Eds.), *Handbook of Speech Perception*. Oxford: Blackwell.

Appendix

Table 1. One hundred thirty-five minimal pairs [MN] [VO] selected in the Contemporary Chinese Dictionary 5th edition (Lu & Ding, 2008).

		Verb	Noun			Verb	Noun
1	帮工	help with farm work	helper	71	讲话	speak	speech
2	报价	to quote	a quote	72	剪影	to sketch	a sketch
3	报料	reveal a news	news	73	兼职	hold two or more posts concurrently	concurrent post
4	保险	to assure	insurance	74	剪纸	cut paper	paper-cut
5	备份	to backup	backups	75	结晶	crystallize	crystal
6	备料	to stock	stock preparation	76	借款	to loan	a loan
7	编号	to number	a number	77	结尾	to end	ending
8	编剧	write a play	playwright	78	进口	to import	importation
9	编码	encoded	code	79	进账	register an income	income
10	标价	mark the price	price	80	纪实	to record actual events	record of actual events
11	表情	express one's feelings	expression	81	纪事	to record	a record
12	比价	compare the price	price parity	82	捐款	contribute money	donation
13	拨款	allocate funds	Appropriation	83	决策	make a strategic decision	a strategic decision
14	补液	supply water	fluid supplementation	84	开局	to open	opening
15	补益	to benefit	benefits	85	开头	make a start	beginning period
16	藏书	collect books	a collection of books	86	考绩	check achievement	merit of professional performance
17	插话	interpose	episode	87	理事	to manage	member of a council
18	超人	exceed	superman	88	留言	leave a message	a message
19	成人	grow up	adult	89	留影	take a photo as a memento	photograph
20	成文	become official dispatch	existing writings	90	满堂	fill the hall	whole hall
21	吃水	absorb water	drinking water	91	满月	complet the first month of life	full moon
22	创意	create	originality	92	逆流	against the flow	backset
23	创作	create	creation	93	配方	make up a prescription	a prescription
24	传人	teach	successor	94	平价	stabilize the price	fair price
25	传闻	to rumour	a rumour	95	评价	evaluate	evaluation
26	传言	pass on a message	hearsay	96	品味	to taste	taste
27	出口	to export	exports	97	欠款	owe	debt
28	存货	to stock	a stock	98	欠债	run into debt	amount due
29	存粮	store up grain	grain in stock	99	欠账	to debit	debit
30	出身	come from	class origin	100	签证	to visa	a visa
31	出账	enter an item of expenditure in the accounts	payment	101	签字	to sign	signature
32	存款	to deposit	a deposit	102	起价	make a price	starting price
33	代表	to delegate	a delegate	103	起头	to start	beginning
34	答卷	answer questions in an examination paper	answer sheet	104	融资	to finance	a finance
35	倒账	repudiate a debt	bad debts	105	说理	argue	argument
36	得分	to score	a score	106	谈话	to talk	a talk
37	定案	decide on a verdict	verdict	107	提包	carry a bag	handbag
38	订单	to order	an order	108	题词	write an inscription	an inscription
39	顶风	against the wind	head wind	109	贴息	pay interest	interest so deducted
40	定稿	finalize a manuscript	final version	110	替工	work as a substitute	temporary worker
41	定价	make a price	fixed price	111	投资	invest	investment
42	定量	to quantify	norm	112	限价	limit the price	limited price
43	定时	definea time	definite time	113	限量	limit the quantity of	limited quantity
44	定义	to define	definition	114	限令	order sb. to do sth. within a certain time	an order
45	定员	designate members	fixed number of staff members	115	限期	set a time limit	deadline
46	定址	select a venue	permanent venue	116	显效	take effect	effect
47	断层	fault	faultage	117	选项	delect an option	option
48	对话	have a dialogue	dialogue	118	续约	renew a contract	renewal term
49	罚金	to fine	a fine	119	押款	borrow money on security	a loan on security
50	罚款	impose a fine	a fine	120	演义	explain some reason or fact	historical novel
51	发面	leaven dough	leavened dough	121	议价	negotiate a price	negotiated price
52	返程	to return	return	122	引例	to cite	citation
53	返利	to rebate	a rebate	123	用语	choose words	choice of words
54	发言	speak	speech	124	约期	fix a date	date of appointment
55	封口	to seal	seal	125	渔利	reap unfair gains	easy gains
56	分界	have as the boundary	boundary	126	造型	to model	modelling
57	耕地	to plough	cultivated land	127	摘要	make a summary	summary
58	管家	manage	manager	128	掌舵	steer a boat	the man who steers a boat
59	鼓包	to lump	a lump	129	征文	solicit articles	essay writing
60	雇工	hire labour	hired labourer	130	转机	transfer	a favourable turn
61	耗材	consume	consumptive material	131	转年	pass to the coming year	the coming year
62	合力	join forces	resultant	132	铸币	to coin	coined money
63	护法	defend	custodian	133	驻军	to garrison	garrison
64	回话	to reply	a reply	134	作文	write a composition	composition
65	汇款	make a remittance	remittance	135	作业	to operate	a job
66	回礼	present a gift in return	a gift in return				
67	回味	recollect the pleasant flavour of ...	aftertaste				
68	寄语	send word	message				
69	剪报	cut newspaper	cuttings				
70	兼差	work part-time	part-time jobs				

Computational Linguistics and Chinese Language Processing
Vol. 18, No. 3, September 2013, pp. 59-80

Non-segmental Cues for Syllable Perception: the Role of Local Tonal f0 and Global Speech Rate in Syllabification[1]

Iris Chuoying Ouyang*

Abstract

This study is aimed at a better understanding of the perception of syllables. As the traditional view seems to associate syllable perception with segmental cues that result from local (*i.e.* present only within or adjacent to the syllable) supralaryngeal events, we are particularly interested in whether non-segmental and non-local laryngeal information contribute to syllable perception as well. Existing works on Indo-European languages show that local stress patterns and global (*i.e.* non-local) speech rates provide perceptual cues to words and phonemes. While we believe that the effects of the global speech rate hold across languages, based on the long-developed notion of language-specific perception, we expect that lexical tones, rather than stress patterns, serve as an important local non-segmental cue in tonal languages. We conducted a perception study on Mandarin to investigate whether tonal f0 patterns and speech rates interfere with spectral information in determining the number of syllables in an utterance. F0 contours were generated using the qTA model (Prom-on, Xu & Thipakorn, 2009). Our results show that the perceptual number of syllables depends on the perception of tonal f0 patterns and speech rates to a substantial extent. Combining our findings with prior claims (Olsberg, Xu & Green, 2007), it appears that a variety of cues – segments, lexical tones, and speech rate – compete in perceiving Mandarin syllables. In relating this study to the existing works on word segmentation, lexical access, and phoneme identification,

[1] Earlier versions of this work were presented at LabPhon (Conference on Lab Phonology) 13 and InterSpeech (Conference of the International Speech Communication Association) 2012; I would like to thank the audience members for the valuable comments and suggestions. Thanks also go to the USC Phonetics Lab group for feedback during the development of this project. I would also like to thank the editors of 'International Journal of Computational Linguistics and Chinese Language Processing' and two anonymous reviewers for their helpful comments.

* Department of Linguistics, University of Southern California, U.S.A.
E-mail: chuoyino@usc.edu

we find that the language comprehension system integrates local with global, supralaryngeal with laryngeal information, in perceiving linguistic units – not only words and phonemes, but also syllables.

Keywords: Cue Integration, Syllable Perception, Tone Perception, Speech Rate, Mandarin Chinese

1. Introduction

The phonetic cues that determine the perception of phonological entities constitute one of the central issues in the research on speech processing. The syllable, a fundamental phonological unit across languages, is generally thought to be perceived via segmental cues and phonotactics. In other words, syllable identification and syllabification in a given language usually are believed to be determined by supralaryngeal *and* local (*i.e.* temporally only present within or adjacent to the syllable in question) information. Local yet laryngeal information such as rhythmic patterns, or global (*i.e.* non-local, temporally present throughout a larger utterance that includes the syllable in question) cues such as speech rate, have not been considered important factors in syllable perception to our knowledge. Nevertheless, existing studies on word perception have shown that listeners use both local and global prosodic cues in word segmentation and/or lexical access (*e.g.* local rhythmic pattern: Nakatani & Schaffer, 1978; Mattys, Jusczyk, Luce & Morgan, 1999; local prosodic boundary: Christophe, Peperkamp, Pallier, Block & Mehler, 2004; Gout, Christophe & Morgan, 2004; global rhythmic grouping: Dilley & Mcauley, 2008; Brown, Salverda, Dilley, Laura & Tanenhaus, 2011). Moreover, phoneme identification also has been found to be sensitive to both local and distant rhythmic patterns (*e.g.* Shields, McHugh & Martin, 1974; Pitt & Samuel, 1990), as well as global speech rate (*e.g.* Summerfield, 1975; Miller & Grosjean, 1981). Based on these findings, it is reasonable to speculate that local and global laryngeal information are involved in the perception of syllables, just as they are in the perception of phonemes and words.

Mandarin Chinese may be one of the languages where the syllable plays a particularly important role. In addition to the argument that it may be a syllable-timed language (*e.g.* Lin & Wang, 2007), Mandarin is known for its close relationship between syllables, morphemes and lexical tones, as most morphemes consist of one syllable and *vice-versa* (*e.g.* DeFrancis, 1984) and every syllable is temporally aligned with a lexical tone and *vice-versa* (*e.g.* Xu, 1998; Gao, 2008). As the syllable appears to have a special status in Mandarin, we conducted this study on Mandarin to investigate non-segmental cues for syllables and hoped to improve our understanding of syllable perception in general.

The next question that naturally arises is what kinds of non-segmental information might affect syllable perception. Prior work on word perception extensively has examined rhythmic representations, such as stress and stress-based prosodic grouping in Indo-European languages (see the beginning of Section 1 for relevant citations). Nevertheless, the mappings between phonetic representations and phonological categories are language-dependent (*e.g.* Caramazza & Yeni-Komshian, 1974); what counts as perceptual cues for syllables may vary depending on the characteristics of a given language. For tonal languages, where pitch patterns provide contrastive information that differentiates one word from another, we consider pitch as a potential factor in segmenting and identifying linguistic units. Given the fixed timing relationship between syllables and tones in Mandarin (*e.g.* Xu, 1998; Gao, 2008), we expect lexical pitch patterns to serve as a local cue and interact with another local cue – segments – in perceiving Mandarin syllables.

Degradation of segmental information has been shown to influence the perception of tones and syllables in Mandarin. Olsberg, Xu, and Green (2007) found that, when high-frequency spectral information was removed from a sequence of [ma] syllables, listeners could not accurately identify either the number of syllables or the categories of tones in the sequence, despite the presence of intact f0 information. What remains unclear is whether the effect goes both ways: does tonal information impact the perception of segments and syllables as well? Particularly, can lexical f0 cues override formant patterns in determining the number of syllables?

Thus far, we have discussed tonal f0 as providing local laryngeal cues for syllable perception. Let us now consider global laryngeal cues. Prior work on phoneme identification has examined speech rate and found that it influences the expectation of VOT length (*e.g.* Summerfield, 1975; Miller & Grosjean, 1981). Since speech rate is not a language-specific phenomenon, we expect to see a similar effect in Mandarin. Global speech rate should affect expected length of linguistic units – in our case, the expected duration of syllables – which should in turn affect the perceptual syllable count within a given time window.

1.1 Aims of this Study

In this paper, we report a perception study that investigates whether local and global non-segmental information influence syllable perception. As discussed in the preceding paragraphs, existing studies that have looked into the role of non-segmental information in segmenting and identifying speech units mostly have focused on phonemes and words. The syllable, another fundamental speech unit, has not been brought into this conversation. To shed light on this issue, we examined the effects of lexical pitch contours[2] and overall speech

[2] In this paper, the terms ‘lexical pitch/f0 patterns’, ‘lexical pitch/f0 contours’, ‘tonal pitch/f0 patterns’, and ‘tonal pitch/f0 contours’ are used interchangeably. We avoid using simply ‘tonal (cues)’ to refer to lexical f0 cues, because tones also involve other acoustic dimensions, such as amplitude (*e.g.*

rate on the perception of syllable numbers. We asked if changes in either lexical pitch contours or overall speech rate, without changes in formant patterns, alter the count of syllables in an utterance. If so, to what extent can they impact syllabification?

1.2 Background: Mandarin Tones

There are four lexical tones in Mandarin: high (Tone 1), rising (Tone 2), low (Tone 3), and falling (Tone 4). They distinguish lexical items from one another, as illustrated in (1).

(1)	Tone 1	ma [High]	'mother'
	Tone 2	ma [Rising]	'hemp'
	Tone 3	ma [Low]	'horse'
	Tone 4	ma [Falling]	'scold'

The simplicity of the tone system in Mandarin allows us to examine how lexical pitch patterns affect a listener's judgment of syllable numbers in a natural way. Crucially, when two level tones (*i.e.* High and Low) are adjacent to each other in continuous speech, they form a pitch pattern that, shape-wise, looks similar to a contour tone (*i.e.* Rising or Falling), although typically different from the contour tone in f0 onsets, offsets, ranges, the turning points of f0 movement, *etc*. (*e.g.* Shih, 1986; Shen, 1990). This opens up possibilities where two words consist of different numbers of syllables yet minimally differ from each other in segments and tones to an extent that they may 'sound similar'. For example, a bisyllabic word with two different level tones (*i.e.* High-Low or Low-High) out of context can potentially be perceived as a monosyllabic word with a contour tone, if there is no consonant at the syllable boundary, *e.g.* [CV.VC] → [CVVC]. Thus, we were able to use real words in the experiment, where the stimuli only differed in tonal f0 contours while retaining a chance of being identified as either bisyllabic or monosyllabic.

2. Perception Study: Method

Participants performed an identification task, where they heard a target word in isolation or embedded in a carrier sentence, saw pictures on a computer screen, and determined which picture matched the word they heard. Using pictures allowed us to avoid presenting participants with written words, which might carry additional phonetic information that is not in the aural stimuli.

Whalen & Xu, 1992, "Information for Mandarin tones in the amplitude contour and in brief segments", *Phonetica,* 49, 25-47) and duration, in addition to f0.

We focused on two factors that were expected to influence the perceptual syllable count of a word: the local tonal pitch contour and the global speech rate. In this study, the term 'local' refers to cues that only occur within the target word region of a sentence, such as the tonal pitch contour carried by the target word; whereas 'global' refers to cues that occur throughout a sentence that includes a target word, such as the overall speech rate of a sentence. In the following subsections, we will first go over the experiment design and procedures before discussing the hypothesis and predictions.

2.1 Design and Stimuli

Participants heard target words or sentences containing a target word one at a time. To investigate whether syllabification depends on local tonal pitch contours and global speech rate, we manipulated the pitch contour in a target word and the speech rate of its carrier sentence while controlling the segments. Specifically, a repeated-measures within-subjects design with two independent variables was used: (i) **'monotone-ness' of the f0 contour in a word** (with six steps, on a continuum between a bitonal sequence to a monotonal sequence) and (ii) **absence or the average syllable length of a carrier sentence** (with four levels: fast, medium, slow, and no carrier).

2.1.1 Target Words

All target words originally consisted of two syllables, as words are predominately bisyllabic in Mandarin. The options of words were limited to the sequences whose tonal and vocalic properties allowed ambiguity about the number of syllables in a word, as discussed in Section 1.1. We used bisyllabic words of which the two syllables had different level tones, namely, Low-High or High-Low, given that the f0 contour of a Low-High tonal sequence (LH) appears similar to a Rising tone (R), and the contour of a High-Low tonal sequence (HL) appears similar to a Falling tone (F). To examine syllable perception across different types of syllable structure, we used bisyllabic words that contained one of the three vocalic sequences, with the same vowel [u] at both sides of the syllable boundary: [u.u], [Vu.u], and [u.uV] (in which V refers to a vowel other than [u]). Thus, for a target word to be identified as monosyllabic, besides perception as a single tone (*i.e.* R or F), the two [u] segments had to be 'misperceived' as one segment. In addition, for the words with a diphthong [Vu] or [uV], the perceptually merged [u] had to be 'misperceived' as part of the diphthong. There were six target words, each in one of the two tonal sequences and one of the three vocalic sequences. (See Appendix A for the list of bisyllabic target words and Appendix B for the list of their monosyllabic alternatives.) Due to the limitations regarding tonal and vocalic properties on the selection of target words, the identity of segments in the target words was not controlled across tonal and vocalic sequences. These limitations also made it difficult to control lexical properties of target words, such as word frequency and word class, which might raise concerns about whether such kinds of differences between bisyllabic target words and their monosyllabic alternatives could have impacted our data. Nevertheless, as the goal of this study was not to compare utterances of different numbers of syllables, but rather to compare segmental material occuring with different tonal f0 at different speech rates, an imbalance in lexical properties between the two members of a bisyllabic-monosyllabic word pair (*e.g.* [tsu.u] - [tsu]) should not distort our results. For example, even if the differences in word frequency or word class between [tsu.u] and [tsu] had biased participants towards [tsu], such bias had probably existed across different levels of tonal f0 and speech rate for this word pair. In other words, response tendencies associated with particular word pairs should not affect data patterns with respect to the effects of tonal f0 and speech rate. Thus, although it would have been ideal for lexical properties to be controlled, we do not regard this as a problem for our findings.

2.1.2 Carrier Sentences and Speech Rate

In one-fourth of the trials, target words were played in isolation; in the other three-fourths of the trials, target words were embedded in a sentence frame, as illustrated in (2). Sentence frames were at one of three different speech rates: fast, medium, or slow. The use of carrier sentences served two major purposes. First, it enabled us to examine the effect of tonal f0 across different speech rates while keeping the duration of target words constant. In other words, we manipulated global speech rate by presenting the same target item with sentence frames that differed in speech rate. Second, it also allowed us to vary the f0 excursion of target words over a wider range and test more levels of f0 contours. As the content of sentence frames were identical throughout the experiment, target words were the only new information in a sentence. Existing works on information structure and discourse-level intonation show that new information is produced with larger f0 ranges than given information in Mandarin (Chen & Braun, 2006; Ouyang & Kaiser, 2013). By placing target words in sentence frames, we were able to use more extreme f0 values in target words, *i.e.* higher f0 for high tone targets and lower f0 for low tone targets, without compromising the naturalness of sentence intonations.

(2)	wo	ba	TARGET	shuo-le	san-ci
	PRO.1st.sg	BA	TARGET	say-PERP	three-time

‘*I said TARGET three times*’

A male speaker of Beijing Mandarin recorded target words in isolation at a medium speech rate (254 ms/syllable) and sentence frames at three different speech rates: fast (131 ms/syllable), medium (259.5 ms/syllable), and slow (441.5 ms/syllable). To obtain natural-sounding tonal coarticulation between a target word and its adjacent words for the sentence stimuli, sentence frames for LH and HL tonal sequences were recorded separately, with a target word embedded as a placeholder. Among the target words in each tonal sequence, the one with the simplest vocalic structure, *i.e.* [tsu.u] in LH and [tʂʰu.u] in HL, was used for recording sentence frames. The speaker was first told to produce all materials at the rate he naturally spoke at, and then to produce the sentences at a rate faster and another rate slower than the first one. Sentence stimuli were later made by replacing the placeholder word in each recording with target words whose f0 had been altered synthetically. There was no apparent pause between a placeholder word and its adjacent words in the original recordings, and no pause was added during the sentence synthesis.

To make sure that embedding f0-altered words in naturally-produced sentence frames did not create artifact effects, we included the conditions of isolated words to be compared against the conditions of medium speech rate. Since target words and medium-rate sentence frames were both recorded at the speaker's natural speaking rate, target words presented in isolation should not yield different results from those flanked by the medium-rate sentence frames.

2.1.3 Tonal f0 Synthesis

To systematically vary tonal f0 patterns in target words, we generated f0 contours via the qTA model (Prom-on, Xu & Thipakorn, 2009), and synthesized them with natural words using the PSOLA (Pitch Synchronous Overlap Add) method implemented in the Praat software. Synthesized pitch patterns were on a continuum between an f0 contour that indicated a two-tone sequence and one that suggested a single tone (*i.e.* LH-R and HL-F). The qTA model was configured as a third-order linear system, where the level, velocity, and acceleration of f0 at a given time point depend on those at the previous time point. There were three parameters in the model: f0 target slope m, f0 target height b, and the rate at which f0 changes λ. In addition, there was an initial f0 state consisting of an initial f0 level $f_0(0)$, initial f0 velocity $f'_0(0)$, and initial f0 acceleration $f''_0(0)$. F0 contours were generated using formulae for one of the two-tone sequences: LH and HL. We kept all parameters in the qTA model fixed and only manipulated λ in the second tone of a sequence, namely the High tone in LH and the Low tone in HL, as shown in Table 1. Since λ corresponds to the speed at which f0 rises or falls, it determines whether and when an f0 contour arrives at its target: the f0 target may be reached later or eventually undershot when λ is low. As λ in the first tone of a sequence was set at a value such that the first tone always reached its (fixed) target, we effectively only varied the f0 contour during the second tone. This choice was motivated by two reasons. *First*, since f0 offsets have been suggested to play an important role in the perception of tones (see Xu, 1997 for relevant discussion), altering the latter part of an f0 contour might be a way of evoking different tone perceptions with minimal manipulation. Pre-empting our findings somewhat, the f0 contour in the second tone of a two-tone sequence did influence the perceptual number of tones that listeners heard. *Second*, for tonal f0 steps that favored a two-tone perception, this manipulation simulated the unequal variability of f0 contours between two consecutive tones in real data. Prior work on Mandarin tones indicates that the f0 contour of a two-tone sequence varies substantially more in the first tone than in the second tone, due to an asymmetry between carryover and anticipatory tonal coarticulation (Xu, 1997).

Table 1. Parameter settings of f0 synthesis using the qTA model

	Low-High		High-Low	
	First tone	Second tone	First tone	Second tone
m	0	0	0	0
b	105	205	230	100
λ	80	65, 80, 95, 110, 125, 140	200	120, 150, 180, 210, 240, 270
f_0	135	-	190	-
f'_0	0	-	0	-
f''_0	0	-	0	-

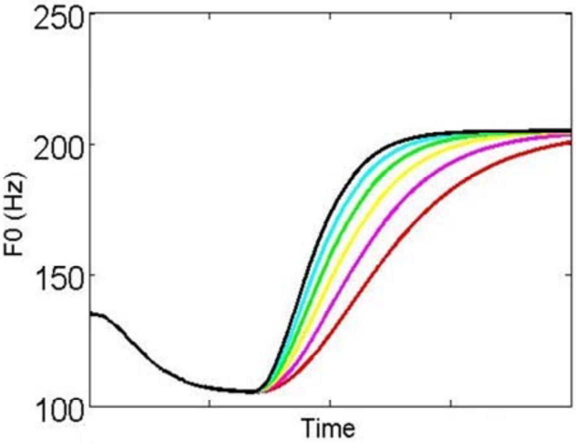


Figure 1. f0 contours for the Low-High target words

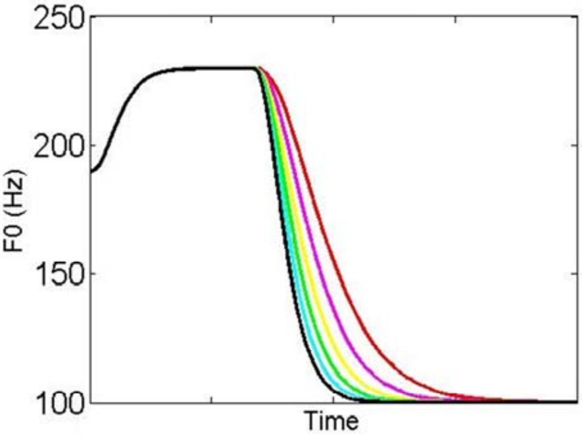


Figure 2. f0 contours for the High-Low target words

It is important to note that, since the tonal sequences used in this study contained two level tones rather than including any contour tones, the tone boundary (*i.e.* where the second tone began) specified in a formula matched the turning point of the f0 contour (*i.e.* the point where f0 started rising or falling for the second tone) it generated. This, to a large extent, resembled real f0 data in Mandarin: the f0 turning point in a two-tone sequence was near the tone boundary when the two tones were in an LH or HL combination (Xu, 1997, Figure 3, Tone 3-1 and Tone 1-3). Essentially, by changing the λ value of the second tone in a formula, we manipulated the part of an f0 contour after the turning point. With a high λ value after the turning point, *i.e.* the higher lines in Figure 1 and the lower ones in Figure 2, an f0 contour sounded more like a sequence of two tones; whereas, when the λ value was low after the turning point, *i.e.* the lower lines in Figure 1 and the higher ones in Figure 2, the entire f0 contour sounded more like one single tone. We aligned the turning point of a synthesized f0 contour with the turning point of the original f0 contour that was produced by the speaker in each bisyllabic target word. An original turning point was defined as f0 minimum for an LH contour and f0 maximum for an HL contour in natural words. Thus, the timing relationship between segments and pitch patterns was preserved to some extent.

A number of six λ's (hence, six tonal f0 steps) was used for each tonal sequence due to reasons of statistical power and experiment length. F0 target heights *b*'s were determined such that f0 values fell between 105-205 Hz in LH target words and 100-230 Hz in HL target words. These tonal f0 targets were reasonable given the f0 ranges of sentence frames produced by the speaker: the average f0 minimum and maximum of sentence frames were 119 and 204 Hz for the LH conditions and 120 and 207 Hz for the HL conditions.

2.1.4 Identification Task

The main experiment was a two-alternative forced-choice identification test (2AFC). Participants saw two pictures side by side on a computer screen while listening to a target word in isolation or flanked by a carrier sentence through headphones. As illustrated in Figure 3, one of the pictures represented the bisyllabic target word that was being played (*e.g.* [tsu.u] in the LH tone), and the other represented a monosyllabic word that participants could potentially hear in the case of 'misperception' induced by our manipulation of tonal f0 and speech rate (*e.g.* [tsu] in the R tone). Participants were asked to choose the picture that matched what they heard. (See Appendix A and B for the pictures used for each word.) They were told that the study was interested in how people recognize words.

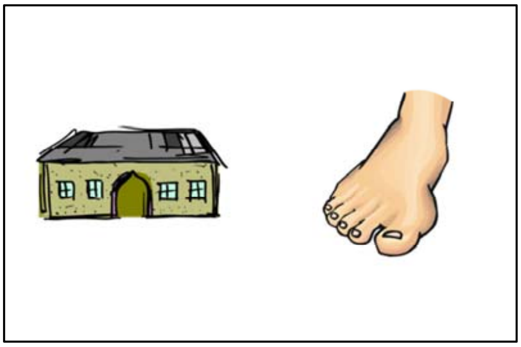


Figure 3. Sample display

Before the main experiment, participants completed a familiarization phase where they learned the names of the pictures (*i.e.* the words represented by the pictures) used in the main experiment. The familiarization phase consisted of two parts. In the first part, each picture was shown on a computer screen one by one with the word it represented underneath. Participants read aloud and memorized the names of the pictures. In the second part, each picture was shown with two words on top of it, one being the word it represented and the other being its monosyllabic or bisyllabic counterpart, *e.g.* the picture of a foot with the words for 'foot' ([tsu]) and 'ancestral house' ([tsu.u]). Participants were asked to pick the name of the picture from the two words; none of them made any mistakes in this task.

Every participant responded to four repetitions of an item. There were 144 items (6 target words * 6 levels of f0 contours * 4 levels of speech rate) and 576 trials in total. The pictures were counter-balanced for the left and right positions across trials. The dependent variable was the percentage of monosyllabic-word responses, namely, the percentage of trials where participants chose the picture representing the monosyllabic alternative rather than the one representing the bisyllabic word (*e.g.* [tsu] instead of [tsu.u]).

2.2 Participants

Twenty-seven native speakers of Mandarin who were born in China participated in the study. The experiment was conducted in Los Angeles, California, USA. All of the participants were between 24 to 44 years old and had left China no earlier than the age of 22. Each participant received 10 U.S. dollars for their participation.

2.3 Hypothesis and Predictions

As discussed in Section 1, the perception of linguistic entities, such as words and phonemes, has been shown to depend on local and on global laryngeal information. Along this line, we believe that syllables should not be an exception; syllabification should also be sensitive to both local and global laryngeal cues. This hypothesis was tested with Mandarin Chinese, a tonal language where lexical pitch patterns and syllables have a fixed timing relationship. We investigated whether Mandarin syllabification is affected by lexical pitch patterns – a local laryngeal factor that indicates the number of syllables in Mandarin – and speech rate – a global laryngeal factor that may crosslinguistically suggest average syllable duration. Our general prediction was that the change of lexical pitch patterns and speech rates both would impact the perception of the number of syllables in Mandarin substantially, regardless of segmental information. Specifically, we expected that the percentages of monosyllabic-word responses would depend on the tonal f0 contour that a target word bears and the rate of the carrier sentences where a target word occurs. Monosyllabic-word responses would increase as the tonal f0 contours became more like a single tone event and the carrier sentences became slower, despite the intact formant patterns that originally produced bisyllabic words. Interaction between speech rates and tonal f0 contours, however, was not expected, as there is no existing evidence that would lead us to believe so. In other words, we predicted that the effect of speech rate would hold across tonal f0 contours, and the effect of tonal f0 contours would hold across different speech rates. Finally, we expected the conditions of isolated words not to differ from the conditions of the medium speech rate, since the target words were originally produced at the medium rate, as mentioned in Section 2.1.2.

3. Results

Overall, the predictions outlined in Section 2.3 were borne out in our results: Tonal f0 contours and speech rate each considerably influenced participants' judgments about the number of syllables in a target word. In this section, the results of LH and HL target words will be analyzed separately, as the segmental properties were not fully controlled across the tonal combinations.

As can be seen in Figures 4-5, target words are perceived as monosyllabic more often when their f0 contours are more monotone-like and when their carrier sentences are slower. In the presence of original, bisyllabic segmental material, participants still chose the monosyllabic alternatives for some percentage of the time, depending on the conditions of tonal f0 and speech rate. In terms of tonal f0, the most monotonal f0 contours on average had 30% more of the monosyllabic choices than the most bitonal f0 contours for the LH words and 21% for the HL words. In the conditions of fast carriers, medium-rate carriers, slow carriers, and isolated words, the mean percentage of monosyllabic-word responses in Step 1 of tonal f0 was higher than that in Step 6 by 27, 35, 33, and 27, respectively, for the LH words and 19, 27, 19, and 19, respectively, for the HL words. With regard to speech rate, the slow carrier sentences on average had 28% more of the monosyllabic choices than the fast carrier sentences for the LH words and 22% for the HL words. From Steps 1 to 6 of tonal f0 contours respectively, the mean percentage of monosyllabic-word responses at the slow rate was higher than that at the fast rate by 27, 33, 27, 32, 31, and 21 for the LH words and 23, 21, 29, 19, 21, and 23 for the HL words. Furthermore, by comparing the two 'extreme' conditions with each other, namely the condition of tonal f0 Step 1 at the slow rate with the condition of tonal f0 Step 6 at the fast rate, we see that tonal f0 contours and speech rate together increased the chance for target words to be identified as monosyllabic by 54% in the LH words and 42% in the HL words.

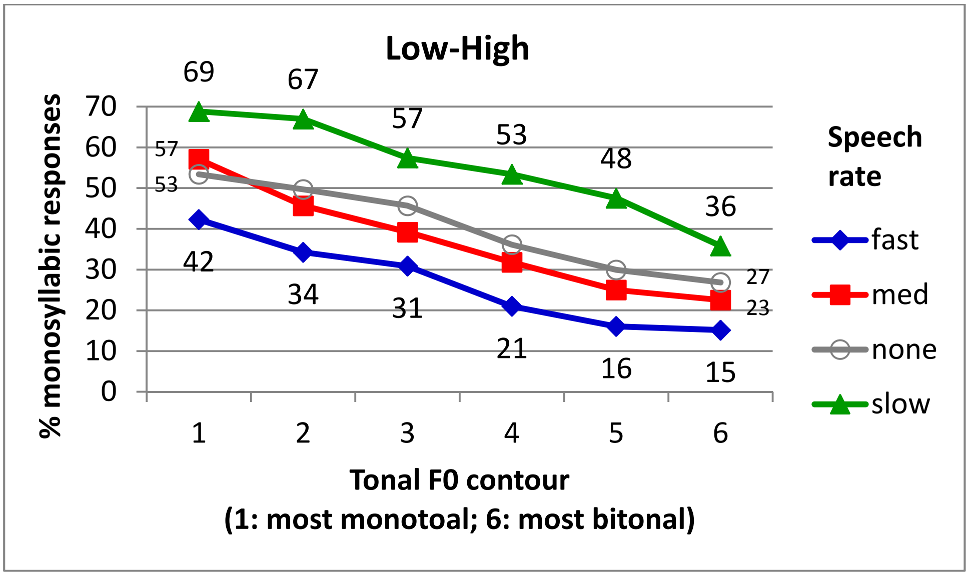


Figure 4. Percentage of monosyllabic-word responses in each condition of the Low-High tonal sequences

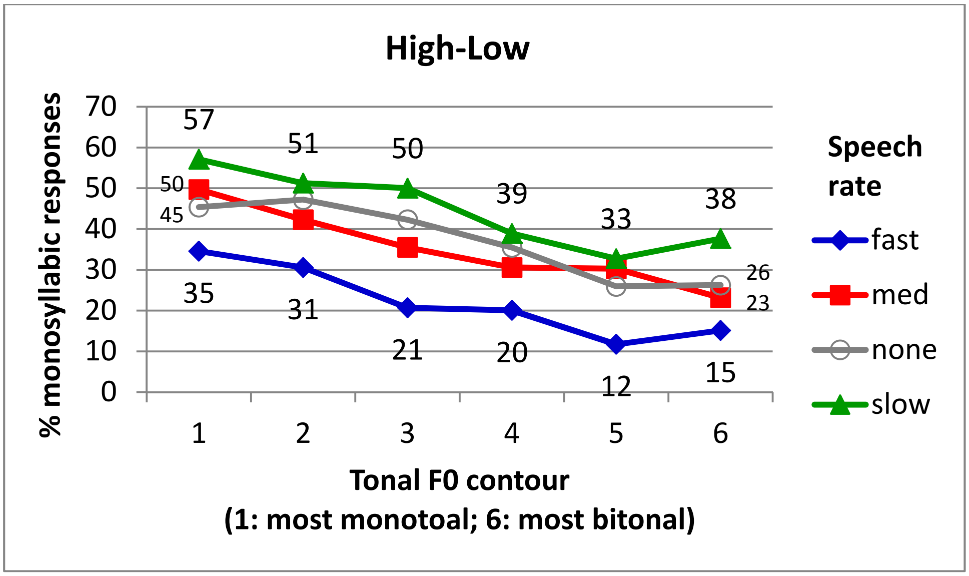


Figure 5. Percentage of monosyllabic-word responses in each condition of the High-Low tonal sequences

Our observations based on visual inspection are confirmed by statistical analysis. The percentages of monosyllabic-word choices were transformed into arcsine data and analyzed using R: A Language and Environment for Statistical Computing (R Core Team, 2013) and the R packages *lmerTest* (Kuznetsova, Brockhoff & Christensen, 2012) and *multcomp* (Hothorn, Bretz & Westfall, 2008). In all of the analyses presented in this paper, the hypotheses were tested at a significant level of α = 0.05 and a marginally significant level of α = 0.1. Mixed effect models were conducted using the *lmer* function in the *lme4* package. We used tonal f0 contours and speech rate as fixed effects and used subjects and items (*i.e.* target word pairs) as random effects. Models with different structures of fixed effects and random effects were compared using the *anova* function in the standard R distribution. We rejected models including a given effect or interaction when they did not differ significantly from the models excluding it. The results show the main effects of tonal f0 contours and speech rate in both LH and HL sequences (tonal f0 contours: F(5,1891.5)'s >= 38.498, p's < 0.001; speech rate: F(3,1891.5)'s >=72.141, p's < 0.001). No interaction between tonal f0 contours and speech rates was found in either LH or HL sequences (F(15,1891.5)'s <= 1.193, p's >= 0.269), as predicted. Planned comparisons were conducted using the *glht* function in the *multcomp* package. For both LH and HL sequences, tonal f0 contours and speech rate each robustly differed between the two ends on their scale. In terms of tonal f0, bisyllabic words with the most monotonal f0 contours were misidentified significantly more often as their monosyllabic alternatives than those with the most bitonal f0 contours at all four levels of speech rate (z's >= 4.576, p's < 0.001). With regard to speech rate, bisyllabic words with the slow carrier sentences were misidentified significantly more often as monosyllabic than those with the fast carrier sentences in all six steps of tonal f0 contours (z's >= 4.692, p's < 0.001). Words presented in isolation did not differ from those flanked by the medium-rate carrier sentences (|z|'s <= 2.201, p's > 0.664), as we expected (see Section 2.1.2 for relevant discussion).

Let us now move on to the intermediate levels of tonal f0 contours and speech rate, namely words with an f0 contour between Steps 2-5 and those in the medium-rate carrier sentences. Although some of the conditions did not significantly differ from each other, the patterns discussed in the preceding paragraph emerge in the intermediate levels as well. In fact, when we look at the pairs of conditions that have a significant difference, all of these response tendencies fit our predictions (*i.e.* monosyllabic-word responses increased when f0 was more monotonal and speech rate was slower), as shown in Tables 2-3. Thus, the hypothesis discussed in Section 2.3 is largely supported by our results. In what follows, we will first look at the middle steps of tonal f0 contours, *i.e.* comparisons between those less than five steps apart, and move on to comparing the medium rate with the slow and fast rate.

Adjacent steps of tonal f0 contours: There was no significant difference between any adjacent steps of f0 contours (z’s <= 2.937, p’s > 0.153, except Step 5 vs. Step 6 in HL was marginal: z = 3.128, p = 0.090). *Tonal f0 contours two steps apart (Step 1 vs. Step 3, Step 2 vs. Step 4, Step 3 vs. Step 5, Step 4 vs. Step 6):* For LH words, monosyllabic responses occurred significantly more in Step 1 than Step 3 at the medium rate, in Step 3 than Step 5 at the fast and the medium rate, and in Step 4 than Step 6 at the slow rate. For HL words, such a difference was significant between Steps 1 and 3 at the medium rate, and Steps 3 and 5 at the slow rate (Step 1 vs. Steps 3 at the medium rate in LH and HL: z’s >= 3.533, p’s < 0.05; Step 3 vs. Steps 5 at the fast and medium rate in LH: z’s >= 3.372, p’s < 0.05; Step 3 vs. Steps 5 at the slow rate in HL: z = 4.634, p < 0.001; Step 4 vs. Step 6 at the slow rate in LH: z = 4.406, p < 0.001; the remainder had z’s <= 3.046, p’s >= 0.114, except these pairs were marginal: Step 1 vs. Step 3 in HL at the fast rate, z = 3.244, p = 0.062, and Step 2 vs. Step 4 in LH at all three rates, z’s >= 3.155, p’s <= 0.084). *Tonal f0 contours three steps apart (Step 1 vs. Step 4, Step 2 vs. Step 5, Step 3 vs. Step 6):* In LH words with f0 contours that were three steps away from each other, the more monotonal contours yielded significantly more monosyllabic responses than the more bitonal contours at all three speech rates. In HL words, only the slow rate conditions showed significant differences between all of the pairs of f0 contours that were three steps apart. Among the conditions of the other two speech rates, significant differences only appeared between Steps 1 and 4 at the medium and the fast rate and Steps 2 and 5 at the fast rate (all three pairs at all three rates in LH: z’s >= 3.699, p’s < 0.05; all three pairs at the slow rate in HL: z’s >= 3.359, p’s < 0.05; Step 1 vs. Step 4 at the medium rate in HL: z = 4.344, p < 0.001; Step 2 vs. Step 5 at the fast rate in HL: z = 5.213, p < 0.001; the remainder had z’s <= 2.896, p’s >= 0.170, except Step 1 vs. Step 3 at fast the rate in HL was marginal, z = 3.128, p = 0.090). *Tonal f0 contours four steps apart (Step 1 vs. Step 5, Step 2 vs. Step 6):* Words with more monotonal contours were significantly more often perceived as monosyllabic than words with more bitonal contours at all three speech rates (z’s >= 4.170, p’s < 0.01), except in HL words at the slow rate where Steps 2 and 6 only marginally differ (z = 3.128, p = 0.089). *Medium vs. fast rate:* Monosyllabic responses occurred significantly more in the medium rate than in the fast rate in Step 1 of the f0 contours for LH words, and Steps 1, 3, and 5 of the f0 contours for HL words (Step 1 in LH: z = 3.427, p < 0.05; Steps 1, 3, and 5 in HL: z’s >= 3.707, p’s < 0.05; the remainder: z’s <= 2.780, p’s >= 0.231). *Slow vs. medium rate:* Words in slow carrier sentences were significantly more often perceived as single syllables than words in medium-rate carrier sentences for LH words with all of the steps of f0 contours, except only marginally so in Step 1. For HL words, such a difference was only significant in Steps 3 and 6 of the f0 contours (Steps 2-5 in LH: z’s >= 3.427, p’s < 0.05; Steps 3 and 6 in HL: z’s >= 3.765, p’s < 0.05; the remainder: z’s <= 2.259, p’s >= 0.617, except Step 1 in LH had z = 3.100, p = 0.098).

Table 2. Effects of tonal f0 at each speech rate

Distance in tonal f0 steps	**Contrast**	**LH**	**HL**
1 step apart	Step 1 vs. Step 2	F M S	F M S
	Step 2 vs. Step 3	F M S	F M S
	Step 3 vs. Step 4	F M S	F M S
	Step 4 vs. Step 5	F M S	F M S
	Step 5 vs. Step 6	F M S	F M S•
2 steps apart	Step 1 vs. Step 3	F M* S	F• M* S
	Step 2 vs. Step 4	F• M• S•	F M S
	Step 3 vs. Step 5	F* M* S	F M S*
	Step 4 vs. Step 6	F M S*	F M S
3 steps apart	Step 1 vs. Step 4	F* M* S*	F• M* S*
	Step 2 vs. Step 5	F* M* S*	F* M S*
	Step 3 vs. Step 6	F* M* S*	F M S*
4 steps apart	Step 1 vs. Step 5	F* M* S*	F* M* S*
	Step 2 vs. Step 6	F* M* S*	F* M* S•
5 steps apart	Step 1 vs. Step 6	F* M* S*	F* M* S*

(F: fast rate; M: medium rate; S: slow rate; *: significantly more monosyllabic-word responses in the more monotonal step than in the more bitonal step; •: marginally more monosyllabic-word responses in the more monotonal step than in the more bitonal step)

Table 3. Effects of speech rate in each tonal f0 step

	Step 1	**Step 2**	**Step 3**	**Step 4**	**Step 5**	**Step 6**
Medium vs. Fast	LH* HL*	LH HL	LH HL*	LH HL	LH HL*	LH HL
Slow vs. Medium	LH• HL	LH* HL	LH* HL*	LH* HL	LH* HL	LH* HL*
Fast vs. Slow	LH* HL*	LH* HL*	LH* HL*	LH* HL*	LH* HL*	LH* HL*

(*: for the two speech rates compared in a given cell, there are significantly more monosyllabic-word responses at the slower rate than at the faster rate; •: for the two speech rates compared in a given cell, there are marginally more monosyllabic-word responses at the slower rate than at the faster rate)

4. Discussion

The study presented in this paper investigates two kinds of perceptual cues for syllabification in Mandarin: lexical pitch patterns that are locally associated with words and the overall speech rate of a sentence that is determined by global measures (*e.g.* average syllable length). As the traditional view seems to connect syllable perception with segmental cues that result from local supralaryngeal events, we are particularly interested in whether non-segmental and non-local laryngeal information contribute to syllable perception as well. A better understanding of these factors is important because they figure fundamentally in how the language perception system integrates different sources of information. In the following subsections, we will discuss how our results correspond to our research questions, sketched out in Section 1.1, as well as their broader implications.

Our key aims were to investigate whether and to what extent the syllable count of a Mandarin word depends on the tonal f0 contour that the word carries and the speech rate of the sentence that includes the word. Our results show that local tonal f0 and global speech rate highly impact the perception of the number of syllables. Across the two types of tonal sequences examined in this study (LH and HL), a change in tonal f0 contours increases the chance of bisyllabic words being perceived as monosyllabic by as much as 35%, and a change in speech rate does so by as much as 33%; changes in tonal f0 and speech rate together impact the likelihood by as much as 54%. These findings indicate that non-segmental and non-local information can contribute substantially to the perception of the number of syllables.

Note that the segmental content of an item in this study was taken from a naturally-produced, bisyllabic word token; its formant patterns strongly encouraged participants to choose the bisyllabic word. If segmental information had dominated syllable perception, we would have obtained equally low percentages of monosyllabic-word responses, no matter what kinds of tonal f0 contours or speech rate had been presented to the participants. The fact that the changes in tonal f0 contours and speech rate changed the likelihood of whether an item was interpreted as monosyllabic or bisyllabic tells the importance of non-segmental information in syllabification. Moreover, when segmental material and non-segmental material provide conflicting cues, for the language perception system to comply with the cues provided by non-segmental material, it must somewhat ignore the cues provided by segmental material – in our study, the two [u] vowels across the syllable boundary needed to be perceived as one. Results of our study suggest that tonal f0 contours and speech rate interfere with segments in perceiving the number of syllables.

Can the cues provided by tonal f0 and speech rate override segmental information in determining syllable counts? This strong claim is not supported by our results. When the tonal f0 content is ambiguous with regard to the number of syllables in a word, *i.e.* in the conditions with the middle steps of tonal f0 contours (Steps 3 and 4), changes in speech rate increase the chance that listeners are 'mistaken' about syllable numbers by only 19-32%. Similarly, changes in tonal f0 raise the likelihood by only 27-35% when the speech rate is neutral, *i.e.* under the condition of medium-rate carrier sentences. In other words, when one of the non-segmental cues (*i.e.* tonal f0 or speech rate) conflicts with segmental information and the other stays neutral, perception of syllable numbers is at best altered to an extent where the likelihood of misperception increases by 35%. It does not seem that either of these two factors decisively influences the syllable count of a word. One might ask, however, can a combination of non-segmental cues 'beat' segmental information? If this were the case, we would have observed a big boost in the percentage of monosyllabic-word responses when both tonal f0 and speech rate strongly favored a monosyllabic interpretation. In fact, our manipulation of tonal f0 and speech rate only produced maximally 54% greater likelihood for bisyllabic words to be identified as monosyllabic. It seems that, even when these two sources of non-segmental information both conflict with the segmental cues, they do not dominate listeners' perception but rather only create confusion about the number of syllables in a word. On the other hand, prior work has shown that degradation of spectral information impacts the perception of the number of syllables and the perception of tone categories (Olsberg *et al.*, 2007). Combining our findings with theirs, it appears that syllable perception in a tonal language involves integration of spectral patterns, tonal f0 contours, as well as the durational information provided by the speech rate.

As we saw in Section 1, existing studies have shown that listeners are sensitive to both local and global non-segmental cues in word segmentation, lexical access, and phoneme identification. Our study adds to the body of literature on the perception of linguistic units by providing evidence for non-segmental and non-local cues in syllable perception. Listeners are capable of using a vast range of information – local or global, supralaryngeal or laryngeal – to segment and identify linguistic units of different levels. Although segmental material alone might be sufficient for syllabification, the presence of cues from other sources impacts the perceptual number of syllables. This highlights cue integration (see Repp, 1982 for an early review on this topic) as the nature of the language comprehension system to utilize multiple domains of cues when available.

5. Conclusion

Based on the perception study reported in this paper, we conclude that the perception of syllable numbers in Mandarin depends on the global speech rate of the entire utterance and local tonal f0 contours that are associated with lexical tones, as well as local segmental material (see also Olsberg *et al.*, 2007). Against the traditional view that syllable perception by and large involves the segmental content of an utterance, we found that the lexical pitch patterns also play an important role in syllabification. Furthermore, not only local information, such as segments and tonal f0, but also global information, such as speech rate, serve as cues for syllable counts. Our findings provide further evidence for cue integration as the nature of the language comprehension system. Even though syllabification and syllable identification supposedly can be accomplished based on local segmental information alone, the presence of non-segmental and non-local information impacts the perception of syllables.

References

Brown, M., Salverda, A. P., Dilley, L. C. & Tanenhaus, M. K. (2011). Expectations from preceding prosody influence segmentation in online sentence processing. *Psychonomic Bulletin & Review*, *18* (6), 1189-1196.

Caramazza, A. & Yeni-Komshian, G. H. (1974). Voice onset time in two French dialects. *Journal of Phonetics*, *2*, 239-245.

Chen, Y. & Braun, B. (2006). Prosodic realization in information structure categories in standard Chinese. In R. Hoffmann & H. Mixdorff (Eds.), *Speech Prosody 2006*. Dresden, Germany: TUD Press.

Christophe, A., Peperkamp, S., Pallier, C., Block, E. & Mehler, J. (2004). Phonological phrase boundaries constrain lexical access: I. Adult data. *Journal of Memory and Language*, *51*, 523-547.

DeFrancis, John. (1984). *The Chinese language: Fact and fantasy*. Honolulu, HI: University of Hawaii Press.

Dilley, L. C. & McAuley, J. D. (2008). Distal prosodic context affects word segmentation and lexical processing. *Journal of Memory and Language*, *59*, 294-311.

Gao, M. (2008). *Tonal Alignment in Mandarin Chinese: An Articulatory Phonology Account*. (Doctoral dissertation). Yale University, New Haven, CT.

Gout, A., Christophe, A. & Morgan, J. (2004). Phonological phrase boundaries constrain lexical access: II. Infant data. *Journal of Memory and Language*, *51*, 547-567.

Hothorn, T., Bretz, F. & Westfall, F. (2008). Simultaneous Inference in General Parametric Models. *Biometrical Journal*, *50* (3), 346-363.

Lin, H. & Wang, Q. (2007). Mandarin rhythm: An acoustic study. *Journal of Chinese Linguistics and Computing*, *17* (3), 127-140.

Kuznetsova, A., Brockho, P. B. & Christensen, R. H. B. (2012). lmerTest: Tests for random and fixed effects for linear mixed effect models (lmer objects of lme4 package). URL http://www.cran.r-project.org/package=lmerTest/.

Mattys, S. L., Jusczyk, P. W., Luce, P. A. & Morgan, J. L. (1999). Phonotactic and prosodic effects on word segmentation in infants. *Cognitive Psychology*, *38*, 465-494.

Miller, J. & Grosjean, F. (1981). How the components of speaking rate influence the perception of phonetic segments. *Journal of Experimental Psychology: Human Perception and Performance*, *7*, 208-215.

Nakatani, L. H. & Schaffer, J. A. (1978). Hearing 'words' without words: Prosodic cues for word perception. *Journal of the Acoustical Society of America*, *63*, 234-245.

Olsberg, M., Xu, Y. & Green, G. (2007). Dependence of tone perception on syllable perception. *Proceedings of Interspeech 2007*, Antwerp, Belgium, 2649-2652.

Ouyang, I. C. & Kaiser, E. (2013). Prosody and information structure in a tone language: an investigation of Mandarin Chinese. *Language and Cognitive Processes*. DOI:10.1080/01690965.2013.805795.

Prom-on, S., Xu, Y. & Thipakorn, B. (2009). Modeling tone and intonation in Mandarin and English as a process of target approximation. *Journal of the Acoustical Society of America*, *125*, 405-424.

R Core Team (2013). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL http://www.R-project.org/.

Repp, B. (1982). Phonetic trading relations and context effects: New experimental evidence for a speech mode of perception. *Psychological Bulletin*, *92* (1), 81-110.

Shen, X. S. (1990). Tonal coarticulation in Mandarin. *Journal of Phonetics*, *18*, 281-295.

Shields, J. L., McHugh, A. & Martin, J. G. (1974). Reaction time to phoneme targets as a function of rhythmic cues in continuous speech. *Journal of Experimental Psychology*, *102*, 250-255.

Shih, C. (1986). The phonetics of the Chinese tonal system. Technical memorandum, AT & T Bell Laboratories.

Summerfield, A. Q. (1975). Aerodynamics versus mechanics in the control of voicing onset in consonant-vowel syllables. *Speech Perception*, *2* (4), 61-72. Belfast, Northern Ireland: Queen's University.

Xu, Y. (1997). Contextual tonal variations in Mandarin. *Journal of Phonetics*, *25*, 61-83.

Xu, Y. (1998). Consistency of tone-syllable alignment across different syllable structures and speaking rates. *Phonetica*, *55*, 179-203.

Appendix A: Target words

Tone	Vocalic structure	Word	Image
LH	u.u	(a) tsu.u ‘ancestral house’	
	Vu.u	(b) pau.u ‘treasure valley’	
	u.uV	(c) hu.uo ‘tiger’s lair’	
HL	u.u	(d) tʂ^{h}u.u ‘first dance’	
	Vu.u	(e) tʂou.u ‘Friday’	
	u.uV	(f) ku.ua ‘single roof tile’	

Appendix B: Monosyllabic alternative of each target word

Tone	Vocalic structure	Word	Image
R	u	(a) tsu 'foot'	
	Vu	(b) pau 'thin'	
	uV	(c) huo 'alive'	
F	u	(d) tʂʰu 'touch'	
	Vu	(e) tʂou 'wrinkle'	
	uV	(f) kua 'Chinese hexagram'	

Computational Linguistics and Chinese Language Processing
Vol. 18, No. 3, September 2013, pp. 81-106

Tones of Reduced T1-T4 Mandarin Disyllables

Shu-Chuan Tseng[*], Alexander Soemer[*], and Tzu-Lun Lee[*]

Abstract

The lexical meaning of Chinese words is determined by syllables and lexical tones. Phonologically, there are four full tones. Empirically, however, it remains a puzzle how tones are recognized when they are reduced in natural speech. This article presents three studies on tones of reduced disyllables: (1) a corpus study on disyllabic reduction, (2) two tone categorical identification experiments on fully pronounced and reduced disyllables, and (3) an analysis of word identification responses of two disyllables. Utilizing a segment-aligned corpus, disyllables were classified by ear into four degrees of contraction (from none to full), *i.e.*, where a disyllable is gradiently reduced towards one syllable. The results suggested that the onset of the second syllable was most likely to be shortened or deleted. For studying the lexical effect of tones, a Ganong-style word bias experiment was conducted on T1-T4 continua of three T1-T4 disyllables. Results of the fully pronounced stimuli confirmed that the lexical status of the disyllables affected the tone classification of F0 contours along a continuum from T1 to T4, showing distinct differences of tone identification in real words and nonwords. Then, this effect disappeared when the onset of the second syllable was removed to simulate a partly reduced disyllable. Insufficient segmental information seemed to deactivate the word-nonword contrast, *i.e.* lexical status seemed to override any acoustic information available. Tones tended to be recognized as those from a real word throughout the continua. Finally, responses to two T1-T4 disyllables from the identification experiment done by Tseng & Lee (2010) were re-analyzed. The results suggested that reduction degree, F0 shapes, word unit type, and exposure frequency seemed to play a role in the recognition of words and tones.

Keywords: Taiwan Mandarin, Disyllabic Words, Tone Perception, Reduced Speech.

[*] Institute of Linguistics, Academia Sinica, Taipei, Taiwan
E-mail: {tsengsc, soemer }@gate.sinica.edu.tw; tzulun@gmail.com

1. Introduction

How tones are recognized in natural conversation remains a puzzle, partly because a number of other factors are acting simultaneously. Some of these factors, for instance pragmatic contextual information and the degree of speech clarity, can be directly observed and manipulated when researchers analyze natural speech. The speaker- and recipient-related individuality, such as their prior experiences about the world, prior experiences about their acquired language, and their speech competence, also are important but difficult to control. This means that the substantial construction of the mental lexicon of a speaker or a group of speakers of a language cannot be investigated in-depth unless we can cope with the variability derived from discrepancies and similarities among individuals, language communities, and their concurrent interaction in speech communication. Working towards this research goal, we made an initial but significant move in this article by beginning to examine the issue of tone recognition in reduced disyllables in Mandarin Chinese. The production and the perception of lexical tones closely and dynamically correlate with the context of lexical constituency. Thus, we first focused on disyllabic words by examining their phonetic forms. Selected disyllabic content words extracted from a corpus of conversational speech were analyzed to obtain typical reduction patterns in terms of reduction degree. The reduction pattern then was applied in our later tone categorical identification experiments to simulate the contrast of fully pronounced and reduced speech. As disyllables with an internal word boundary may not have the same reduction pattern as disyllabic words, we further re-analyzed the responses of a previously conducted word identification experiment, focusing on one disyllabic word and one monosyllabic disyllable to study the role of tone as modulated by the degree of reduction.

1.1 Processing of Spoken Words

The most commonly used speech form is daily face-to-face conversation. The consensus in building natural speech corpora is the belief that we should look at realistic data for studying linguistic patterns and human behavior. The fine (and complex) details in natural speech cannot be thoroughly investigated until annotated speech corpora have been made available for different languages, for instance the Kiel Corpus of Spontaneous Speech for German (IPDS, 1995), the Chinese Annotated Corpus of Spontaneous Speech (Li *et al*., 2000), the Spoken Dutch Corpus (Goddijn & Binnenpoorte, 2003), the Buckeye Corpus of Conversational Speech (Pitt *et al*., 2006), the Corpus of Interactional Data for French (Bertrand *et al*., 2008), the Corpus of Spontaneous Japanese (Maekawa, 2009), and the Taiwan Mandarin Conversational Corpus (Tseng, 2013). Variation of spoken words has been studied since then on different levels, including segmental, word category, word frequency, and prosodic position. Corpus-based studies are required to provide quantitative empirical description of reduction in natural speech. For instance, using the Buckeye Corpus, Jurafsky *et* *al.* (2001) studied the effect of lexical frequency and local contextual information on the probability of reduction for function and content words. They suggested that this shortening process was not necessarily correlated with vowel reduction, but probably with the lexical frequency. Similarly, Meunier & Espesser (2011) analyzed the Corpus of Interactional Data for French. Although not directly affected by the lexical frequency effect, they found that vowel reduction, including duration shortening and vowel centralization, occurred more often in monosyllabic function words than in monosyllabic content words. Both studies addressed the importance of lexical frequency on the surface forms of reduced word in natural speech. This leaves the question of how reduced spoken words are recognized by humans.

A number of psycholinguistic models have been proposed to explain how humans perform the task of spoken word recognition and how words are stored in the mental lexicon, including the phonetic-acoustic form and the associated higher level lexical information. The Cohort Theory, the Interactive-Activation Model (TRACE model), and the Neighborhood Activation Model were the classical models accounting for abstract lexical representation and extraction (Marslen-Wilson, 1987; McClelland & Elman, 1986; Luce & Pisoni, 1998). The Cohort Theory proposes that spoken word recognition involves both early bottom-up processing mainly utilizing word-initial onsets and late top-down contextual information. The Interactive-Activation Model states that (once the sensory acoustic-phonetic input comes in) different levels of a lexical item are activated, including feature, phoneme, and word. The Neighborhood Activation Model stresses the importance of similarity neighborhood density between words and the relative effects of word frequency on the spoken word recognition. Different from these abstract form extraction models, Lacerda (1995) suggested that prototypes of word forms are stored in memory and discrimination extent is used to filter the best exemplar, which is most similar to the prototype. To some extent, the concept of best exemplar is similar to the concept of episodic memory traces, which may preserve the (most likely) surface phonetic details (Goldinger, 1996). In speech communication, we constantly encounter different types of phonetic forms of words, which we may classify into phonological variants. Thus, it is not surprising that production frequency of words has some influence on the phonological representation of pronunciation variants, as found by Ranbom & Connine (2007) in their studies on the nasal flap in English.

While pronunciation variants are normally concerned with substitution of segments, the phonetic forms of words can deviate substantially from the canonical forms in the way that segments are omitted. Nevertheless, spoken words with a number of deleted segments still can be recognized easily, given proper contextual information. Frauenfelder & Tyler (1987) emphasized the importance of context in empirical practice of spoken language, in addition to the five typical, sequential lexical processing phases they proposed: initial lexical contact, activation, selection, word recognition, and lexical access. A number of previous studies have made similar proposals. Words in semantically coherent sentences were more accurately recognized than those in isolation, and words heard in context often are recognized long before their full acoustic signal has been delivered (Grant & Seitz, 2000; Grosjean, 1980; Marslen-Wilson, 1987). Nevertheless, words with omitted segments have been studied only recently. Highly reduced words were well recognizable only when presented in their original context with semantic and syntactic information. Without context or only with limited contextual information of adjacent syllables, they cannot be properly recognized (Ernestus *et al*., 2002; Tseng & Lee, 2010). Thus, questions are raised. Do the same sequential lexical processing phases apply to the recognition process of reduced words without context (Frauenfelder & Tyler, 1987)? Will we observe a lexical effect of tones in reduced words without contextual information? In this study, we aim to study the reduction patterns of Mandarin words in natural speech and how tones in reduced words are recognized.

1.2 Tones

Lexical tones in Mandarin Chinese constitute an abstract, contrasting phonological system of four full tones and a neutral tone (Ho, 1996; Duanmu, 2000). The four full tones include the high level tone (T1), the mid-rising tone (T2), the dipping tone (T3), and the high falling tone (T4). The neutral tone is normally noted as T5. The lexical meaning of Chinese words is determined by the syllable and tone information simultaneously. For decades, the issue of how lexical tones are produced and recognized has attracted tremendous attention and interest from linguists and psycholinguists. Tones are often illustrated by means of fundamental frequency contour (F0). F0 contour is a kind of acoustic information transmitted through the air to the ears of the listeners. Canonical forms of lexical tones in monosyllables show similar F0 contour shapes (Xu, 1997: 67; Lai & Zhang, 2008: 184). A clear high level F0 contour is observed for T1. T4 starts higher than T1 with a falling contour. T2 and T3 start lower in pitch height, with a rising and dipping contour, respectively. In addition to studies on monosyllables produced in isolation, contextual effects of tones have also been investigated. Very detailed phonetic cues have been studied or used for manipulating the stimulus tokens, such as the onset and the offset F0 values of syllables and those in the adjacent syllables, as well as the degrees of rise and fall of the F0 contours (Lin & Wang, 1984; Xu, 1997; Xu, 2004; Cutler & Chen, 1997). A number of different paradigms have been applied to test how tones are perceived or recognized (Lai & Zhang, 2008; Lee, 2007; Lee *et al*., 2008; Hallé *et al*., 2004; Ye & Connine, 1999; Fox & Unkefer, 1985). Testing monosyllables with four lexical tones in the gating experiment of Lai & Zhang (2008), T1 and T4 are confusing to the listeners in the first 40 to 80 ms after onset. After a period of 80 ms, the recognition turns to be correct in cases of both T1 and T4. The vowel and tone monitoring tasks conducted by Ye & Connine (1999) provided support for the notion that tones are activated differently (from vowels) when the listeners hear the syllable-tone stimulus tokens in isolation or in context. To more concretely involve the semantic association of experiment stimuli, Lee (2007) and Lee *et al.* (2008) took into account semantic and acoustic cues in their form priming and identification experiments. Their results suggested that tonal information can be used to reduce activated candidates and can be compensated for when segmental information is insufficient. To examine the lexical status effect of tones in an implicit way, the Ganong paradigm (Ganong, 1980) was adopted by Fox & Unkefer (1985). When using the Ganong paradigm, listeners are given a standard categorical perception identification task, *i.e.*, they are played a series of phonetic stimuli that vary from one phonemic category to another in a gradient way and must identify each stimulus as belonging to one of these two phonemic categories. One end of the continuum is a real word and the other is not, usually finding a response preference for the phonemic category that forms a real word (*e.g.*, in a tash-dash continuum, English listeners will identify the /d/ in more stimuli than /t/, since "dash" is a real word but "tash" is not). Fox & Unkefer (1985) tested the lexical status effect of tones by having native Chinese and non-native English subjects identify T1 or T2 in the stimulus tokens of T1-T2 continua of Mandarin Chinese monosyllabic words and nonwords. Differences were observed between the two groups of subjects. Nevertheless, no differences were found between the nonword/nonword continuum and the other continua involving real words. Hallé *et al.* (2004) conducted a tone identification task on tone continua of T1-T2, T2-T4, and T3-T4 monosyllables to native Taiwan Mandarin and non-native French listeners. Despite observable differences between the native and non-native groups, the French listeners were also sensitive to the changes of F0 shapes in perceiving the tonal contrast to a certain degree. As we speculate on the lack of an obvious semantic association of the stimulus tokens in the above experiments, we used disyllabic words instead of monosyllabic words in our tone perception experiment to concretely enhance the link to the lexical constituency.

2. Disyllabic Words in Mandarin Chinese

2.1 Why T1-T4 Disyllabic Words?

The majority of words in modern Mandarin Chinese are either monosyllabic or disyllabic, according to the 5-million words of a textual corpus, the Sinica Balanced Corpus[1] (Chen & Huang, 1996). In terms of word tokens, mono- and disyllabic words together make up approximately 90% of the corpus. Tri- and quadra-syllabic words normally are composed of mono- or disyllabic words. **Figure 1** illustrates the frequency percentages of monosyllabic words, disyllabic words, and words of more than two syllables in the Sinica Balanced Corpus of 5 million words, in the Taiwan Mandarin Conversational Corpus of 500K transcribed

[1] Website: http://db1x.sinica.edu.tw/kiwi/mkiwi/.

words[2] (Tseng, 2013), and in the historical work *Zuozhuan* of 160K words from approximately 400 B.C. (the Academia Sinica Tagged Corpus of Old Chinese[3], Wei *et al*., 1997).

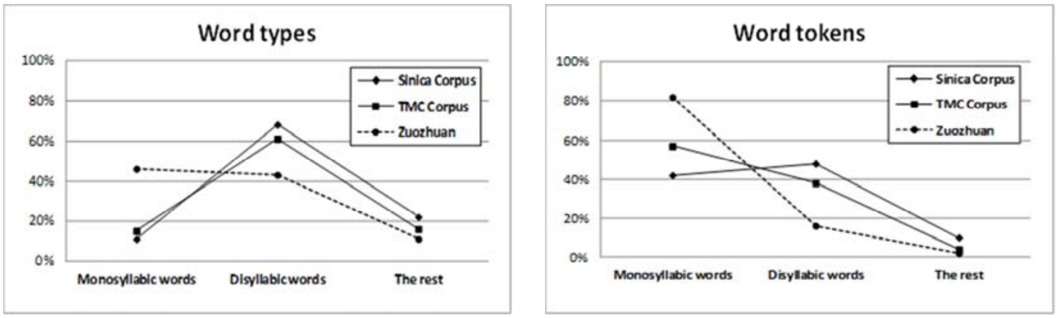


Figure 1. Word distribution in Mandarin Chinese corpora of modern texts, modern conversational speech, and ancient text.

The use of monosyllabic words descends and that of disyllabic words ascends in modern Mandarin in both the written and spoken uses. The Sinica Balanced Corpus and the Taiwan Mandarin Conversational Corpus share a similar distribution of disyllabic words, both making up 60% of the word types and 40% of the word tokens in the corpus. According to the publicly-distributed Taiwan Mandarin Spoken Wordlist derived from the Taiwan Mandarin Conversational Corpus (Tseng, 2013), the four most frequently produced disyllabic tone pairs are T4-T4, T1-T4, T2-T4, and T3-T4 with a proportion of 13.4%, 9.7%, 8.9%, and 8%, respectively. One of the reasons we used T1-T4 tone pair in our later tone perception experiment was both T1 and T4 are produced with a high pitch onset. This shared property makes it easier to manipulate the tone continua than the other tone pairs.

2.2 Using Syllable Contraction to Define the Reduction Degree of Disyllables

There should be a wide variety of reduced word forms, given different syllable structure, phonological neighborhood, lexical constituency, sentence structure, *etc*. Thus, how to represent the typical reduced word form is the first task we encounter. As we are concerned with disyllabic words, we chose syllable contraction to annotate the degree of reduction. Syllable contraction is a phenomenon of producing words of more than one syllable in a shorter way, resulting in a decrease of the number of segments and possibly also syllables. Some of the syllable mergers are predictable in Chinese phonology and dialectology (Lung, 1976; Cheng, 1985; Chung, 1997; Lien, 1997). The onset of the first syllable and the rhyme of

[2] Website: http://mmc.sinica.edu.tw.

[3] Website: http://old_chinese.ling.sinica.edu.tw/.

the second syllable make up the final form of the merger, with the rhyme of the first syllable and the onset of the second syllable omitted, which is also known as the Edge-in Theory (Chung, 1997; Hsu, 2003). For instance, the merger of two sentence-final particles *zhi55* and *hu11* in Old Chinese is represented by a new character with the pronunciation *zhu55*. This kind of lexicalized syllable merger may be due partly to speech reduction in natural speech use, as different degrees of syllable contraction including the merger are observed in production data of natural speech (Tseng, 2005). In our corpus analysis of disyllabic words, we adopted the concept of syllable contraction to annotate four different degrees of reduction.

3. Reduction of Disyllabic Words in Natural Speech

In our study of disyllabic word reduction, we examined a dataset of 3.5 hours of conversational data produced by 16 speakers to account for speaker differences and to create a more or less near-"naturalness" of the speech data. This dataset was extracted from the Taiwan Mandarin Conversational Corpus (Tseng, 2013). The transcripts of the conversations were automatically segmented and POS tagged using the CKIP (1998) system. Segmentation errors, including errors of proper nouns, idioms, constructions with numbers, directional complements, and disfluencies, were manually corrected (Tseng, 2013). To avoid highly frequent function words, as they may have specific phonetic forms due to their semantic predictability in speech communication, we analyzed disyllabic verbs and nouns. Those that contained reduplicated syllables, such as *ba4ba5* (colloquial form for "father"), and homophonous words, such as *ge4ren2*, which can mean oneself or each one, were excluded. As a result, 1,496 tokens of 50 disyllabic proper nouns and verbs that were produced at least three times in the corpus were studied (**Appendix A**).

3.1 Human Labeling of Reduction Degree in Terms of Syllable Contraction

Although previous speech production studies on syllable contraction of Taiwan Mandarin observed word forms contracted to different extents (Tseng, 2005), they lack an adequate definition for operationally annotating the instances. In our current study, we propose an annotation scheme consisting of four reduction degrees. If there is an audible syllable boundary within a disyllabic word, it is marked as **uncontracted** one. In the case of a vague syllable boundary, it is marked as **moderately contracted**. If the two nuclei of a disyllabic word are merged and it is practically impossible to separate the two syllables anymore, the boundary is regarded as **considerably contracted**. In the case of syllable merger, it is a **merged** case. Two annotators independently labeled the reduction degree of each disyllabic word following the above criteria. The inter-rater's agreement is high, Kappa = .803, $p < .01$. Forty inconsistent cases were discussed until a consensus was achieved. As a result, half of the instances were marked as **merged** (724 occurrences). There were 171 and 300 **Moderately** and **considerably contracted** instances, followed by 301 **uncontracted** occurrences. The four annotated categories are well-balanced, which provides a good basis for the later analysis.

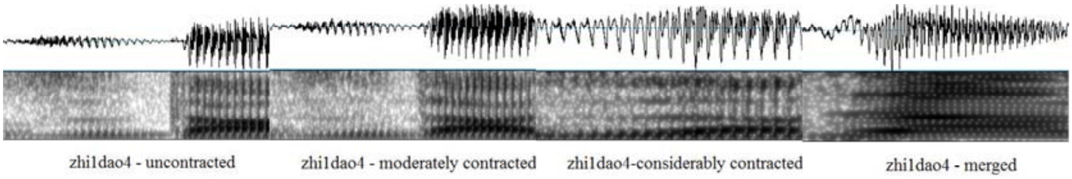


Figure 2. Four reduction degrees.

Figure 2 illustrates the spectrograms of the T1-T4 disyllabic verb *zhi1dao4* /tʂʅ tau/ (to know) annotated with the four reduction degrees. They were all located in a prosodically medial position. The first three examples of *zhi1dao4* were spoken by a male speaker, and the merged example was spoken by a female speaker. In the uncontracted instance, the spike of /t/ is clearly observed, whereas it is not the case in the moderately contracted example. The syllabic boundary in the moderately contracted example is clear, but it is hardly visible in the considerably contracted example. In the merged case, no syllabic boundary can be observed in the spectrogram.

3.2 Segmental Reduction in Disyllabic Words

To study segment reduction in disyllabic words, we need segment-labeled data. A forced segment aligner was adopted to process the segment labeling automatically. A number of supervised and unsupervised automatic aligners have been developed to obtain segment-labeled data of realistic speech (Pitt *et al*., 2006; Scharenborg *et al*., 2010; Kuo *et al*., 2007). For our study, we trained the aligner with a human-checked, segment-labeled subset extracted from the Taiwan Mandarin Conversational Corpus with mono-phone acoustic models using the HTK toolkit (Liu *et al*., 2013). Applying this to our dataset of 1,496 tokens of disyllabic verbs and nouns, we obtained time-aligned boundaries for 7,166 segments. Please note that the forced alignment of segments was conducted independently from the labeling of reduction degree above. Given a text and a sound file, a segment aligner forcedly assigned segment boundaries to all segments in the text according to the similarity in terms of the previously trained acoustic models for each segment. That is, for the merged and contracted cases, segments that were not distinguishable by the human ear were assigned with boundaries as well. In such cases, the aligner could not find a suitable acoustic model to match the probably deleted segments, so a minimum duration of 15 ms was assigned, suggesting that it may be a case of segment substitution or deletion. Thus, we regard segments with duration of 15 ms as our candidates for deletion. In adopting this approach, instead of human judgment of phonological variants of spoken words, we referred to the prosodic-acoustic features of the disyllabic words for our analysis of reduction degree.

The duration of all 7,166 segments was extracted via Praat (Boersma & Weenink, 2013). Please note that the coda position can only be occupied by nasals [n, ŋ] and there are only two glides [j, w]. The originally extracted segment duration was normalized by taking the z-score for each speaker, noted as **Z_dur** (Lobanov, 1971). **Table 1** summarizes the descriptive statistics of the syllabic positions of segments, including the distribution of the deletion candidates and the **Z_dur** mean duration. S1 denotes the first syllable; S2 the second.

Table 1. Segments categorized in syllabic positions.

Segment Position	**Total**	**Deletion candidates**	**%**	**Z_dur mean**
Onset_S1	1468	73	4.97%	-0.009
Glide_S1	374	34	9.09%	-0.433
Nucleus_S1	1496	273	18.25%	0.003
Coda_S1	524	229	43.70%	-0.329
Onset_S2	1340	781	58.28%	-0.484
Glide_S2	258	119	46.12%	-0.379
Nucleus_S2	1496	84	5.61%	0.720
Coda_S2	210	6	2.86%	0.062

Generalized linear models were conducted to statistically verify the contribution of the position of segment within the disyllabic words and the annotated reduction degree to the duration and the deletion likelihood of syllabic positions. Dependent variables were **Z_dur** mean and the percentage of deletion candidates. Predictors were the syllabic position of the segment and the reduction degree. The results confirmed significant effects of segment position and reduction degree on the duration and deletion percentage. Detailed results are summarized in **Table 2**.

Table 2. Summary of the Generalized Linear Models.

Dependent variable: **Z_dur** mean				Dependent variable: Deletion percentage			
	Wald χ^2	*df*	*P*		Wald χ^2	*df*	*p*
Intercept	.01	1	.92	Intercept	55.81	1	< .01
Segment position	91.04	3	< .01	Segment position	19.76	3	< .01
Reduction degree	486.01	7	< .01	Reduction degree	35.57	7	< .01

Using the uncontracted tokens as the comparison baseline, their duration was significantly longer than the other three categories (p < .01) and the percentage of deletion was significantly lower than the merged cases (p < .01), considerably contracted cases p = .011, and moderately contracted cases p= .45. To take Onset_S2 as the comparison baseline, its duration was significantly shorter than Onset_S1, Nucleus_S1, and Nucleus_S2, Coda_S2 (p < .01) and was shorter, but not significantly, than Coda_S1 (p = .68) and Glide_S2 (p = .37). Onset_S2 was longer than Glide_S1, but this was not statistically significant (p = .28). With regard to deletion percentage, that of Onset_S2 was significantly larger (p < .01) than Onset_S1, Glide_S1 (p < .05), Nucleus_S1, Nucleus_S2, and Coda_S2. The deletion percentage of Onset_S2 was larger (but not significantly) than Coda_S1 (p = .6) and Glide_S2 (p = .58).

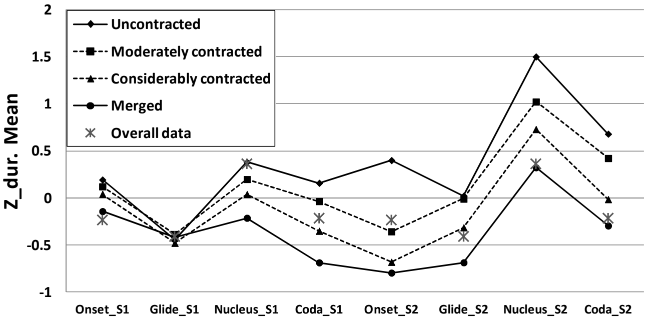


Figure 3. Temporal patterns of different reduction degrees.

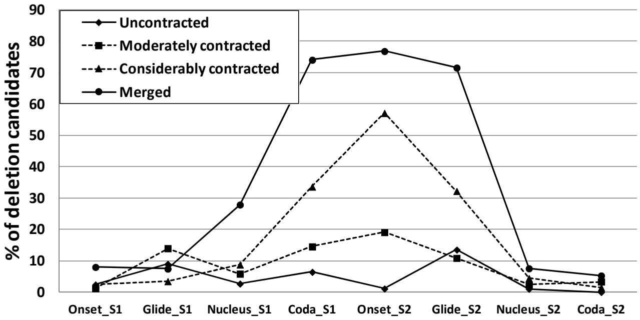


Figure 4. Distribution of deletion candidates of segments.

Figure 3 and **Figure 4** illustrate these tendencies. The asterisks mark the **Z_dur** means for each syllabic position in the overall data. Generally speaking, when a word is reduced more, the duration of segments also declines more. The onset of S2 is the shortest segment position, with an exception of the glide of S1 (Glide_S1). **Figure 3** shows a longer duration of the nucleus of S2 (Nucleus_S2) than all of the other segment positions. It became even longer in the contracted and merged cases. A similar tendency also was observed for Onset_S1 and Coda_S2. They are both longer than the average mean durations. These are supportive evidence for the preservation of Onset_S1, Nucleus_S2, and Coda_S2, corresponding to the final form predicted by the Edge-in Theory for disyllables. As proven in the statistical results, Onset_S2 may be the most likely position to be deleted when producing the reduced disyllabic words, shown in **Figure 4**. Coda_S1, Onset_S2, and Glide_S2 contain more deletion candidates; Nucleus_S2, Coda_S2, and Onset_S1 considerably fewer. Among them, Onset_S2 contains the most deletion candidates, with significantly more than Coda_S2, Onset_S1, Nucleus_S2, and Glide_S1. The results of the deletion percentage and the duration of syllabic positions suggest that Onset_S2 may be the most likely segment to be deleted in naturally spoken, reduced disyllabic words. Please also note that, as we regarded the likelihood of segment deletion as the percentage of deleted segments in each given position separately, the syllable structure should not cause obvious artifacts on this conclusion.

Despite the widely-accepted and approved conclusions that reduced segments are more likely to be shortened, we found that changes in the duration of segments vary depending on their position in our data of disyllabic words. The results of the corpus-based segment analysis suggested that the device of reduction in disyllabic words in natural speech is systematic. Taking into account the pattern presented above, the possibly reduced forms of the CV-CV disyllabic word *zhi1dao4* (to know) may be [tʂʅ-au], [tʂau], and [au] by omitting the Onset_S2 first, then Nucleus_S1, finally also Onset_S1. In the subsequent tone categorical identification experiment, we will remove Onset_S2 to simulate the reduced version of our CV-CV disyllabic stimuli.

4. Tone Identification in T1-T4 Disyllabic Words

4.1 Data and Method

Using the Ganong paradigm for examining lexical effect of tones, the response preference should be found for the tone category that forms a real word. Nevertheless, Fox & Unkefer (1985) experimented on T1-T2 continua of monosyllabic Chinese words and nonwords without finding clear differences between nonword-nonword pairs and other pairs involving real words. This may be due to the lack of a sufficient link to semantic association of monosyllables, especially when presented in isolation. To enforce the semantic involvement in the test items, we used disyllabic instead of monosyllabic words in our study. Two T1-T4 disyllabic words *zhi1dao4* (to know) and *zhi4du4* (the system) were used as our stimulus tokens, whose T4-T4 counterparts *zhi4dao4* and *zhi1du4* are nonwords. *Zhi1dao4* was chosen, because it is the most frequently used disyllabic verb that is composed of T1 and T4 and because *zhi1dao4* and *zhi4du4* have no homophones in the Taiwan Mandarin Spoken Wordlist

(Tseng, 2013). Fox & Unkefer (1985) used DEI in their T1-T2 continua study as the nonword-nonword item. In our study, we also used *dei4* to be the nonword-nonword token. Empirically confirmed, *dei4* was not found in the Taiwan Mandarin Spoken Wordlist. As *dei4* is a non-syllable, we used it to contrast the non-word tokens with the real syllable *du4*. **Table 3** illustrates the design of the stimulus tokens.

Table 3. Stimuli design.

T1⇨T4 *zhi* continua	Contrast: Real word vs. nonword
DAO4	Real word⇨Nonword (with a real tonal syllable) *Zhi1dao4* ⇨ *zhi4dao4*
DU4	Nonword (with a real tonal syllable)⇨Real word *Zhi1du4* ⇨ *zhi4du4*
DEI4	Nonword (with a non-tonal syllable)⇨Nonword (with a non-tonal syllable) *Zhi1dei4* ⇨ *zhi4dei4*

The stimulus tokens were recorded and manipulated using Praat. *Zhi1dao4* was recorded and *zhi1* was cut out to manipulate the continuum from T1 to T4. As *zhi1* is produced with the high level T1, followed by the plosive /t/ as Onset_S2, the F0 contour of *zhi1* is slightly rising for the main portion of the entire syllable and slightly falling towards the end due to the closure of the following plosive /t/ and the high falling T4 in the second syllable. A T1-T4 continuum was synthesized by shifting the highest F0 value of *zhi1* 5 ms each time backwards while maintaining the original rising and falling slopes of the F0 contours. In this way, we increased the falling part of *zhi1*. Finally, we obtained the high-falling *zhi4*. When we concatenated the T1-T4 continua with *dao4*, *du4*, and *dei4*, we accordingly modified the pitch registers fitting those of the subsequent syllables to make each transition phase within the disyllables sound as smooth as possible. As proposed by Lin & Wang (1984), the tone recognition of a Mandarin Chinese syllable can be affected by the pitch height of the following syllable, which was taken into account when we were manipulating the tokens. At first, the tone continuum contained thirteen steps. Ten subjects were asked to judge the similarity of the thirteen steps that were presented to them in a row. As a result, we excluded five steps that most of the subjects reported as indistinguishable. Please note that the recording of the three disyllables maintained the same duration over all of the steps. While manipulating the stimulus tokens, we did not change the temporal quality of the words at all. For the first experiment, the disyllabic stimuli were fully pronounced. In the second experiment, the Onset_S2 /t/ and its transition phase to the nucleus of the second syllable were removed.

4.2 Tone Identification in Fully Pronounced Disyllables

36 subjects, 18 female and 18 male, aged between 20 and 30 were recruited. They were paid for the experiment. In the training session, the subjects heard the examples of *dian4shi4* (television) with different F0 contours contrasting with a nonword *dian1shi4* to familiarize the subjects with the procedure of the experiment. In the main session, the eight steps in each of the three T1-T4 continua *zhi1dao4 -> zhi4dao4*, *zhi1du4 -> zhi4du4*, and *zhi1dei4 -> zhi4dei4* were used as our stimuli. The acoustic properties of the first tone in each of the word-nonword, nonword-word, and nonword-nonword disyllables were the same in each step. In total, 48 trials (3 word types x 8 steps x 2 repetitions) were presented to the subjects in a randomized order and their responses were recorded using the E-Prime Software (Schneider *et al.*, 2002). The task of the subjects was to decide whether the first syllable was T1 or T4 by pressing one of the two buttons marked as T1 and T4.

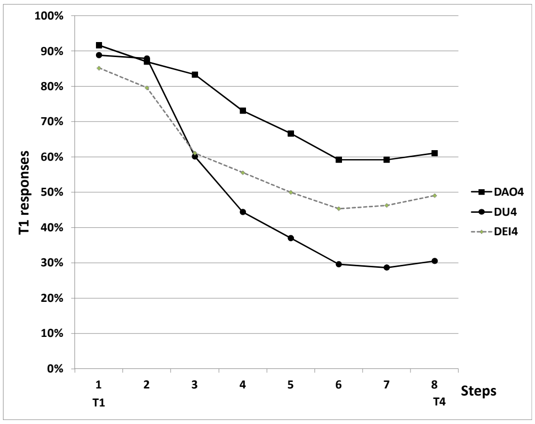


Figure 5. Tone identification responses (T1) for T1 to T4 continua in fully pronounced disyllables zhi1dao4, zhi1du4, and zhi1dei4.

A look at **Figure 5** gives the impression of a clear lexical status effect on the recognition of the first syllable. In the first two steps, the difference seems to be marginal. Starting from Step 3, DU4 and DEI4 drop drastically, compared with DAO4. The listeners still tend to hear T1, because T1 would form the real word, *zhi1dao4*. In the following five steps, the listeners continued showing this preference for the DAO4 trials. This also shows the neutral role the nonword-nonword continuum of DEI4 trials plays in the experiment. The T1 recognition rate of DEI4 is in the middle of the word-nonword DAO4 continuum and the nonword-word DU4 continuum.

We sought to confirm this subjective impression by fitting mixed-effects logistic regression models on our data (see Baayen *et al.*, 2008 for a related tutorial). The basic method consists of establishing a reasonable random effects structure (which is equivalent to testing the significance of these factors with χ^2-Tests (Baayen *et al.*, 2008)) before testing for the significance of the fixed factors using an alpha level of 5%. Our base model consisted of *word type* as the only fixed factor and *step*, *subjects*, and *repetitions* as random factors. Exclusion of the random factors *step* and *subjects* in this base model led to a significant decrease in model fit (both p < .01), while there was no significant difference between the basic model and a model without the random factor *repetition* (p=.48). Thus, the factors *subjects* and *steps* were kept in the model and the factor *repetitions* was discarded. The random effects structure of this model was further extended step by step to include interactions between the fixed and random factors. The final model included a *word type* by *step* interaction (p < .01) and a *word type* by *subject* interaction (p < .01). After establishing the random effects structure, we tested the fixed effect of word type on tone recognition. The χ^2-Test showed that exclusion of the factor *word type* resulted in a significant decrease of model fit (p < .05). Planned pairwise comparisons between the three word types showed that DAO4 differed significantly from DU4 (p < .01), while DEI4 did not differ significantly from DAO4 (p = .13) and from DU4 (p = .24).

Consistent with a visual inspection of **Figure 5** we found an interaction between *word type* and *step*. Fitting models and carrying out pairwise comparisons for each step separately showed that the DAO4 and DU4 differed significantly for all steps (p < .01) except for Steps 1 and 2 (p = .44 and p =.83, respectively). DAO4 differed from DEI4 significantly for Steps 3 to 7 (p < .05) and marginally for Steps 1 and 8 (p = .08 and p = .06, respectively). Nevertheless, for Steps 6, 7, and 8, DU4 also differed significantly from DEI4. In addition, we found marginal differences between these two word types for Steps 4 (p = .05) and 5 (p = .06).

In sum, our statistical analysis fully confirmed the visual impression of **Figure 5**. Subjects' responses differed between the three word types. The word type DAO4 led to significantly more T1 responses for the first syllable than DU4, while DEI4 resided between these two. We conclude that, in spite of the same acoustic cues being provided to the listeners in the three word type conditions, the perception of tones in disyllabic stimulus tokens is affected by whether the disyllabic stimuli are real words and real syllables or not.

4.3 Tone Identification in Reduced Disyllables: Onset_S2 Removed

A different group of subjects from the previous experiment was recruited for the second experiment on the reduced stimuli. 36 subjects, 18 female and 18 male, aged between 20 and 30, were paid for the experiment. We removed /t/ in all three continua of *dao4*, *du4*, and *dei4*, including the transition phase to the Nucleus_S2. In total, 72 trials (3 word types x 8 steps x 3 repetitions) were randomized and presented to the subjects using E-Prime. The task of the subjects was to answer whether the first syllable was T1 or T4. The result of the reduced stimuli was completely different from the previous one.

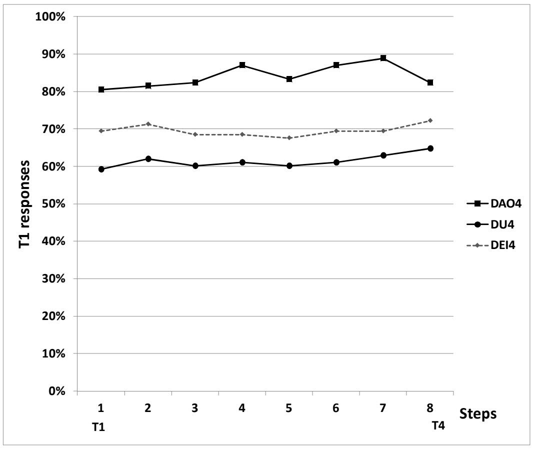


Figure 6. T Tone identification responses (T1) for T1 to T4 continua in reduced disyllables zhi1dao4, zhi1du4, and zhi1dei4.

Visual inspection of **Figure 6** suggests that reduced stimuli had a different effect on tone recognition from the fully pronounced stimuli of the previous experiment. Most obvious, there seems to be no difference in the judgment for the different steps. The listeners constantly tended to hear T1, but with different degrees dependent on word type. When the first syllable was combined with DAO4, which formed a real word, they tended to hear T1 the most. When it was combined with DU4, which formed a nonword, they tended to hear T1 the least. The concatenation with the non-lexical syllable DEI4 lies right in the middle. The question is if the visible differences between the three word types are also statistically significant.

We used the same statistical procedure. Logistic regression models were fit on our data, with *word type* as the fixed factor, and *subjects*, *steps*, *repetitions* as random factors. This time, exclusion of both *steps* and *repetitions* did not lead to a significant decrease of model fit (p = .99). This confirms that *step* did not have a significant influence on tone judgment. Nevertheless, adding a *word type* versus *subjects* interaction yielded a significantly better fit (p < .01), suggesting that the subjects responded quite differently to the word types. After establishing the random effects structure, we tested for significance of the fixed factor *word type*. The main effect of word type failed to reach significance (p = 0.13 in case of DAO4 versus DU4).

As pointed out before, our fitted model contained a significant *word type* versus *subjects* interaction. This suggested that, for some subjects, the difference visible in **Figure 6** between the word types reached significance. We carried out a per-subject analysis. Out of all 36 subjects, we obtained a *word type* effect for 5 subjects. For Subjects 15 and 22, there was a significant difference between DU4 and DEI4 (p < .01). For Subject 19, both the differences between DU4 and DEI4 (p < .05) and between DU4 and DAO4 (p < .01) were significant.

Subject 25 showed significant differences in all comparisons (all p < .05).

Overall, the results suggest that when the segmental information is not sufficient, listeners are more likely to interpret the F0 contour as a tone in a disyllabic word in the way that is consistent with an existing word in the mental lexicon. Interestingly, the differences in F0 contour between the steps seem to have been completely overridden by lexical information in this experiment. One caveat to this interpretation is that differences in word type were not statistically significant for all subjects. We suggest that this was due to a ceiling effect (many subjects reached 100% decision scores for T1), and we are currently carrying out further experiments to tackle this issue. In sum, we interpret this result with caution as showing the tendency of an influence of the mental lexicon on tone recognition. In this context, we further suggest that the overall tendency to hear T1 could be due to the higher lexical frequency of *zhi1dao4* versus *zhi4du4*.

4.4 Interim Discussion

We constituted a more concrete association of tones with word meaning through use of disyllabic words, instead of monosyllables, as recognizing tones of monosyllables would not necessarily imply recognizing spoken words. It could also only mean a perceptual processing of pitch differences (Cutler & Chen, 1997; Ye & Connine, 1999). As shown in previous studies, lexical tones can be determined very quickly in monosyllabic words by native listeners (Lai & Zhang, 2008), while non-native listeners also performed quite well in recognizing tones in monosyllabic items (Fox & Unkefer, 1985; Hallé *et al*., 2004). Different from the results proposed by Fox & Unkefer (1985), we did find difference among the tone continua of word-nonword, nonword-word, and nonword-nonword pairs via the fully pronounced disyllabic stimuli. At what stage tones start to function in recognizing spoken words has not yet been studied well. One of the reasons may be that most of the studies have focused on tones of monosyllables produced in isolation. So far, psycholinguistic research is still far from being able to propose a theory of the role of tone in the classical models of spoken word processing, as it is difficult to find a place to include tonal information on the basis of a hierarchy of feature, phoneme, and word (Marslen-Wilson, 1987; McClelland & Elman, 1986). In our results, the non-lexical counterpart *dei4* provides a clear threshold for the distinction between word and nonword from Step 3. This means the lexical effect is observed by shifting the highest F0 value by 20 ms leftwards for *dao4* and *du4* contrasts, which is a very subtle difference. This threshold, however, does not necessarily tell us how tones are processed. As previous studies have pointed out, tonal information in disyllabic words varies to a large extent depending on the tonal context (Xu, 1997) and the perception of tones in multiple syllables are dependent on the pitch register and duration of the following syllable (Lin & Wang, 1984). Thus, tone processing is complex. Could it be possible that tonal information is processed in a way similar to the late top-down contextual information, as mentioned in the Cohort Theory (Marslen-Wilson, 1987), but maybe at a different level? The recognition rates of tone continua of monosyllabic stimulus tokens in Fox & Unkefer (1985) and Hallé *et al*. (2004) range from nearly 0% to 100%. In our experiment with the fully pronounced disyllables, however, the rates range from 30% to 90%. This difference suggests that the mapping between tone categories and lexical decision deviates depending on how tones are associated with the reduced word forms stored in the lexicon (or in the memory). Furthermore, based on the results of our first experiment, lexical processing of fully pronounced disyllabic words may involve a direct activation of mapping the phonetic form into the canonical form. When segment information is missing, however, acoustic cues of tones did not play a role at all, as no step difference was found. Instead, the response preference was related to the more frequently heard reduced word forms. This suggests that listeners tend to hear the words (thus, the tone) as those they encounter more often in speech communication. Production frequency of words may play a role.

5. Observations in a Case Study of Reduced T1-T4 Disyllable Recognition

Reduction does not result in any obvious problems hindering the understanding of the listeners if enough contextual information is available to them (Ernestus *et al*., 2002). Nevertheless, the question remains of what kinds of cues are there for language users to use to associate a reduced, ambiguous sound (including segmental and tonal information) with a meaning when there is no contextual information. In this section, we conducted a re-analysis of a previous word identification experiment (Tseng & Lee, 2010). Tokens of 8 disyllabic words and 8 disyllables composed of two monosyllabic words that were spoken in two reduction degrees (contracted and merged) and were extracted from the Taiwan Mandarin Conversational Corpus were used as stimuli. Uncontracted stimuli were recorded separately by the same speaker. They were presented to 48 subjects in three conditions: with the full context, in isolation, and in a carrier sentence. The subjects had to identify what they heard in terms of Chinese characters or the phonetic symbols of Zhuyin Fuhao, which is a phonetic transcription method used widely in Taiwan. As all stimulus tokens were spoken by the same speaker, discrepancies of pitch height and pitch range should not be an obvious problem. The previous results in Tseng & Lee (2010) showed that reduced words in their original sentential context were easier to recognize than those in an irrelevant context, such as a less meaningful carrier sentence or in isolation. It was more difficult to recognize more reduced than less reduced words. In consideration of word unit type, it also showed that semantically coherent disyllabic words were more correctly identified than combinations of monosyllabic words that were semantically ambiguous without context. This effect was especially apparent under the isolation condition. In Tseng & Lee (2010), correct responses were counted and analyzed without examining all of the responses given by the subjects. In the current analysis, we closely transcribed and classified the responses to two T1-T4 disyllable stimuli. As word unit type played a role in the accuracy rate of word identification, we analyzed both *yin1wei4* /in wei/ (because) and *ta1-shi4* /ta ʂʅ/ (it/he/she - to be). *Yin1wei4* is a disyllabic word, and *ta1-shi4* a disyllable with two monosyllabic words.

5.1 Word Identification Responses

For *yin1wei4*, if provided with full context, all 48 subjects had no problems recognizing it in both of the contracted and merged cases, as shown in **Table 4**. All subjects used the correct Chinese characters to transcribe it. Nevertheless, when the context was missing, only 24 subjects recognized *yin1wei4* in the contracted case. When the reduction degree increased, the percentage of the correct answers also declined. In the merged cases, only 17 answers were correct. Among the not-recognized responses, there was a consensus on the onset and nucleus of the first syllable. The recognition of segments did not differ much among the subjects, but that of tones did. Different from *yin1wei4*, even provided with full context, not all subjects correctly recognized *ta1-shi4*. 34 subjects recognized it in the contracted case and only 27 in the merged case. To a certain extent, the degree of reduction and the word boundary within the disyllable led to recognition problems. Nevertheless, regardless of whether full context was provided or not, if the responses were not completely correct, at least the first syllable was correctly recognized by the subjects. Even in the merged case without context, 16 answered with *ta1* and 28 subjects with the correct onset of the first syllable. Here, it was shown that the reduction patterns of disyllables differ in terms of the pattern of lexical constituency. Disyllabic words and two monosyllabic words, although both are disyllables, use different reduction devices. The first onset of *ta1-shi4* is mostly preserved, and the second syllable is completely reduced and weakened. Thus, the subjects consistently recognized the first tone as T1. For *yin1wei4*, however, in spite of some consensus on the onset consonant and nucleus, the recognition of tones differed to a large extent. As this may have to do with the F0 shapes of the stimulus tokens, we further analyzed the F0 patterns.

5.2 F0 Shape and Exposure Frequency

We stylized the F0 shapes of the disyllables with two linear lines for the voiced regions in each syllable and calculated the slopes. Baseline slopes were obtained by using the first syllable of the uncontracted tokens that were recorded by the same female speaker. If the slopes of the stylized F0 contour in the contracted and merged tokens rose more than the rising T2, we regarded the F0 shape as a “rise”. If the slopes fell more than the falling T4, we regarded it as a “fall”. Otherwise, the F0 contour was classified as “level”. The F0 shape was categorized only if the duration ratio of the stretch of the F0 was larger than one-third of the entire syllable, as it was not representative enough otherwise. Applying this procedure, the F0 shape of the contracted *yin1wei4* is rise-fall. The merged *yin1wei4* has a level F0 shape. The contracted stimulus tokens of *ta1-shi4* both have a level F0 shape.

Table 4. Word identification responses to reduced disyllables.

Yin1wei4	Contracted: [inei], rise-fall		Merged: [ỹɛ̃], level	
With context	**Yin1wei4**	48	**Yin1wei4**	48
In isolation	**Yin1wei4**	24	**Yin1wei4**	17
	Onset_S1 recognized as the nasal /n/ or /m/ (S1 answered with T1: 3, T2: 6, T3: 5, T4: 2, T5: 1)	17	Onset_S1 recognized as the bilabial stop /p/ (S1 answered with T1: 2, T4: 6)	8
	Onset_S1 recognized as zero onset with /i/ or /y/ (S1 answered with T1: 6, T2: 1)	7	Onset_S1 recognized as zero onset with /i/ or /y/ (S1 answered with T1: 3, T4: 10, T5: 5)	18
			Others	5
Ta1-shi4	Contracted: [t^{h}az̩], level		Merged: [t^{h}aʔ], level	
With context	**Ta1-shi4**	34	**Ta1-shi4**	27
	Ta1-de5	13	**Ta1/ta1-zai4/ta1-jiu4**	16
	Not answered	1	Others	5
In isolation	**Ta1-shi4**	9	**Ta1**	16
	Ta1-de5	30	Onset_S1 recognized as /t^h/ (S1 answered with T1: 19)	28
	Ta1-others	6		
	Others	3	Others	4

As shown in **Table 4**, more answers were correct in the recognition of the citation tone of *ta1-shi4* (45 in contracted cases and 35 in merged cases) than of *yin1wei4* (33 in contracted cases and 22 in merged cases). This may be due to the reason that the F0 shapes of *ta1-shi4* are level. Still, the reduced disyllables were easier to recognize if the disyllables formed a word unit. When the contracted *yin1wei4* was spoken in a rise-fall F0 shape, it clearly indicated that there were two syllables. This was correctly recognized by 24 subjects. The merged *yin1wei4* token was produced with a level F0 shape, but 17 subjects were still able to recognize it. *Yin1wei4* is a frequently produced item, so its reduced phonetic form should be familiar to the subjects. In contrast, the internal word boundary in *ta1-shi4* and the greatly reduced second syllable seemed to hinder the recognition; only nine answers were correct in the contracted case and none in the merged case.

We also examined the exposure frequency of *yin1wei4* and *ta1-shi4* by calculating the bi-gram frequency from the Taiwan Mandarin Conversational Corpus. *Ta1-shi4* appeared 376 times, and *yin1wei4* appeared 1,573 times. The disyllable *yin1wei4* empirically is encountered more often in conversational speech than *ta1-shi4*. The other responses *ta1-de5* and *ta1-jiu4* occur 406 and 259 times, respectively. When segmental information is largely missing, tone information does not seem to act the same way as when the segmental information is complete. The mapping to a lexical representation via a phonetic form encountered often could be more efficient than via a sequential or hierarchical mapping route through the abstract units. As shown in **Table 4**, with insufficient segmental information, the identification of tone categories seems to be closely bound to the "already" recognized words, instead of being processed separately. Exposure frequency may affect the method of mapping a reduced, ambiguous phonetic form to a meaningful word, as it normally determines the degree of familiarity of the "ambiguous phonetic forms".

6. Conclusion

The recognition process of tone and meaning of spoken words may differ when listeners are conducting tasks with different extents of linguistic information, as proposed by Ye & Connine (1999). When presented in isolation, surface phonetic forms of spoken words that are often heard in everyday speech communication influence the way in which we recognize reduced words. This was shown in our tone category identification experiment with reduced stimulus tokens. The simulated disyllabic tokens became ambiguous when the onset consonant of the second syllable was removed, as a direct association of the phonetic form with a canonical word form was nearly impossible. At the same time, due to a clearly audible boundary between the two syllables, we can be sure that the subjects should map the tokens they heard with disyllables. The preference for tones towards real words disappeared with the ambiguous reduced stimuli. Distinction in the acoustic information had no effect on the decision of tone selection, as was evident from the absence of a step effect. We would like to interpret this result as a kind of support for the influence from the production frequency of spoken words (Jurafsky *et al.*, 2001; Myers & Li, 2009). With insufficient segment information, a word with a high production frequency would attract responses towards the word. In our case, as *zhi1dao4* is a frequently spoken word in conversation, its reduced form is more familiar than *zhi4du4*. As a consequence, the preference for hearing a T1 could simply be a frequency effect. Listeners tended to base their judgment on real words, thus more T1 for *zhi1dao4*, less T1 for *zhi4du4*, and the non-lexical counterpart with *dei4* sat in between. Moreover, the word identification experiment suggested that the segmental and tonal information of reduced, ambiguous word forms were recognized as one unit, not separately.

The recognition of spoken words would be based mainly on how frequent the surface phonetic forms, *i.e.*, reduced word forms, language users encounter and store/memorize in their mental lexicon. In the consideration of word unit type, production frequency should not merely be lexical frequency, but we should also take into account co-occurrences of words (n-grams). Is production frequency, however, the only factor that determines the recognition of reduced words and the way words are stored in the mental lexicon? To answer this question, we plan to use words that are used rarely in the spoken language as stimuli to study the effect of production frequency. Finally, with regard to our experimental design, it is possible that listeners may shift to different strategies when encountering clear versus reduced speech (see the word type versus subject interaction in the experiment with reduced stimuli). At this point, we cannot determine whether this shift is automatic and rapid (as it should be in a natural conversation that includes both clear and reduced speech) or is induced by the blocked design of our (and other researchers') experiments. We plan to include both clear and reduced stimuli in our next experiments and to further explore the role of individual strategies in listening to natural speech.

Acknowledgements

The authors are grateful to the useful comments provided by two anonymous reviewers of the *International Journal of Computational Linguistics and Chinese Language Processing*. The authors also sincerely thank the team members who have been working on the corpus data along the years. The studies presented in this article were funded by Academia Sinica and the National Science Council under Grant NSC-101-2410-H-001-085.

References

Baayen, R.H., Davidson, D.J., & Bates, D.M. (2008). Mixed-effects modeling with crossed random effects for subjects and items. *Journal of Memory and Language, 59*, 390-412.

Bertrand, R., Blache, P., Espesser, R., Ferre', G., Meunier, C., Priego-Valverde, B., & Rauzy, S. (2008). Le CID: Corpus of Interactional Data. Annotation et Exploitation Multimodale de Parole Conversationnelle. *Traitement Automatique des Langues*, *49*(3), 1-30.

Boersma, P. & Weenink, D. (2013). Praat: doing phonetics by computer [Computer program]. Version 5.3.48, retrieved 1 May 2013 from http://www.praat.org/.

Chen, K.J. & Huang, C.-R. (1996). Sinica corpus: Design methodology for balanced corpora. *PACLIC 11*, 167-176.

Cheng, R. L. (1985). Sub-Syllabic Morphemes in Taiwanese. *Journal of Chinese Linguistics*, *13*(1), 12-43.

Chung, R. (1997). Syllable Contraction in Chinese. In *Chinese Languages and Linguistics III. Morphology and Lexicon*. Tsao and Wang (Eds.). Symposium Series of the Institute of History and Philology. Academia Sinica. Taipei. 199-235.

CKIP. (1998). *The Sinica Corpus 3.0*. The Chinese Knowledge Information Processing Group - technical report 98-04. Academia Sinica. (In Chinese)

Cutler, A. & Chen, H. C. (1997). Lexical tone in Cantonese spoken-word processing. *Attention, Perception, & Psychophysics, 59*(2), 165-179.

Duanmu, S. (2007). *The Phonology of Standard Chinese*. 2nd Edition. Oxford University Press.

Ernestus, M., Baayan, H., & Schreuder, R. (2002). The Recognition of Reduced Word Forms. *Brain and Language*, *81*(1-3), 162-173.

Fox, R. A. & Unkefer, J. (1985). The effect of lexical status on the perception of tone. *Journal of Chinese Linguistics*, *13*(2), 69-89.

Frauenfelder, U.H. & Tyler, L.K. (1987). The process of spoken word recognition: An introduction. *Cognition*, *25*(1-2), 1-20.

Ganong, W. F. (1980). Phonetic categorization in auditory word perception. *Journal of Experimental Psychology: Human Perception and Performance*, *6*(1), 110-125.

Goldinger, S. D. (1996). Words and voices: Episodic traces in spoken word identification and recognition memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *22*, 1166-83.

Goddijn, S.M.A. & Binnenpoorte, D. (2003). Assessing manually corrected broad phonetic transcriptions in the Spoken Dutch Corpus. In *Proceedings of the International Congress of Phonetic Sciences (ICPhS)*, Barcelona, Spain, 1361-1364.

Grant, K.W. & Seitz, P.F. (2000). The recognition of isolated words and words in sentences: Individual variability in the use of sentence context. *The Journal of the Acoustical Society of America*, *107*, 1000-1011.

Grosjean, F. (1980). Spoken word recognition processes and the gating paradigm. *Perception & Psychophysics*, *28*(4), 267-283.

Hallé, P. A., Chang, Y. C., & Best, C. T. (2004). Identification and discrimination of Mandarin Chinese tones by Mandarin Chinese vs. French listeners. *Journal of Phonetics, 32*(3), 395-421.

Ho, D. (1996). *Concept and Method of Phonology*. Daan Publishing (in Chinese).

Hsu, H. (2003). A Sonority Model of Syllable Contraction in Taiwanese Southern Min. *Journal of East Asian Linguistics*, *12*, 349-377.

IPDS. (1995). The Kiel Corpus of Spontaneous Speech. IPDS vol. 1. University of Kiel.

Jeffreys, H. (1961). *The Theory of Probability* (3 ed.). Oxford: Oxford University Press.

Jurafsky, D., Bell, A., Gregory, M., & Raymond, D. (2001). Probabilistic relations between words: evidence from reduction in lexical production. In Bybee, J. and Hopper, P. (Eds.)

Frequency and the emergence of linguistic structure, 229-54. Amsterdam: John Benjamins.

Kuo, J.-W., Lo, H.-Y., & Wang, H.-M. (2007). Improved HMM/SVM methods for automatic phoneme segmentation. *Interspeech 2007*, Antwerp, Belgium. 2057-2060.

Lacerda, F. (1995). The perceptual-magnet effect: An emergent consequence of exemplar-based phonetic memory. In Elenius, K. & Branderyd, P. (Eds.), *The XIIIth International Congress of Phonetic Sciences*. Stockholm, Sweden. 140-147.

Lai, Y. & Zhang, J. (2008). Mandarin lexical tone recognition: The gating paradigm. *Kansas Working Papers in Linguistics*, *30*, 183-194.

Lee, C. Y. (2007). Does horse activate mother? Processing lexical tone in form priming. *Language and Speech, 50*(1), 101-123.

Lee, C. Y., Tao, L., & Bond, Z. S. (2008). Identification of acoustically modified Mandarin tones by native listeners. *Journal of Phonetics, 36*(4), 537-563.

Li, A., Zheng, F., & Byrne, W. (2000). CASS: A phonetically transcribed corpus of Mandarin spontaneous speech. In *Proceedings of the International Conference on Spoken Language Processing (ICSLP)*, 485-488.

Lien, C. (1997). Studies on Directional Complement in Taiwan Southern Min – Dialect Typology and Historical Study. In Chiu-yu Tseng (Ed.), *Chinese Languages and Linguistics IV, Typological Studies of Languages in China*, vol. 2. Academia Sinica, Taipei. 379-404.

Lin, T. & Wang, W. S.-Y. (1984). On the issues of tone perception. *Zhongguo Yuwen Xuebao*, *2*, 59-69. (in Chinese)

Liu, Y.-F., Tseng, S.-C., & Chang, R. J.-S. (2013). A phone-aligned conversational corpus of Taiwan Mandarin. Manuscript.

Lobanov, B. M. (1971). Classification of Russian vowels spoken by different speakers. *Journal of the Acoustical Society of America*, *49*, 606-608.

Luce, P. A. & D. B. Pisoni. (1998). Recognizing spoken words: The neighborhood activation model. *Ear and Hearing*, *19*, 1-36.

Lung, Y. (1979). A Discussion of the Theory that Yin-sheng Words End with Final Consonants. *Bulletin of the Institute of History and Philology*. *50*(4), 679-716. (in Chinese)

Maekawa, K. (2009). Analysis of language variation using a large-scale corpus of spontaneous speech. In S.-C. Tseng (Ed.) *Linguistic Patterns in Spontaneous Speech*, 27-50. Academia Sinica: Taipei.

Marslen-Wilson, W.D. (1987). Functional parallelism in spoken word-recognition. *Cognition*, *25*(1-2), 71-102.

McClelland, J. L. & Elman, J. L. (1986). The TRACE model of speech perception. *Cognitive Psychology*, *18*, 1-86.

Meunier, C. & Espesser, R. (2011). Vowel reduction in conversational speech in French: The role of lexical factors. *Journal of Phonetics*, *39*(3), 271-278.

Myers, J. & Li, Y. (2009). Lexical frequency effect in Taiwan Southern Min syllable contraction. *Journal of Phonetics*, *37*, 212-230.

Pitt, M.A., Dilley, L., Johnson, K., Kiesling, S., Raymond, W., Hume, E., & Fosler-Lussier, E. (2006). Buckeye Corpus of Conversational Speech (1st release). Columbus, OH: Department of Psychology, Ohio State University, USA.

Ranbom, L. J. & Connine, C. M. (2007). Lexical representation of phonological variation in spoken word recognition. *Journal of Memory and Language*, *57*, 273-298.

Scharenborg, O., Wan, V., & Ernestus, M. (2010). Unsupervised speech segmentation: An analysis of the hypothesized phone boundaries. *Journal of the Acoustical Society of America*, *127*(2), 1084-1095.

Schneider, W., Eschman, A., & Zuccolotto, A. (2002). *E-Prime Reference Guide*. Pittsburgh: Psychology Software Tools, Inc.

Tseng, S.-C. (2005). Syllable Contractions in a Mandarin Conversational Dialogue Corpus. *International Journal of Corpus Linguistics*, *10*(1), 63-83.

Tseng, S.-C. (2013). Lexical coverage in Taiwan Mandarin conversation. *International Journal for Computational Linguistics and Chinese Language Processing*, *18*(1), 1-18.

Tseng, S.-C. & Lee, T.-L. (2010). Contextual effects in recognizing reduced words in spontaneous speech. *Proceedings of DiSS-LPSS Joint Workshop*, Tokyo. 39-42.

Wei, P.-c., Thompson, P. M., Liu, C.-h., Huang, C.-R., & Sun, C. (1997). Historical corpora for synchronic and diachronic linguistics studies. *International Journal for Computational Linguistics and Chinese Language Processing*, *2*(1), 131-145.

Xu, Y. (1997). Contextual tonal variations in Mandarin. *Journal of Phonetics, 25,* 61-84.

Xu, Y. (2004). Understanding tone from the perspective of production and perception. *Language and Linguistics, 5*(4), 757-797.

Ye, Y. & Connine, C. M. (1999). Processing spoken Chinese: The role of tone information. *Language and Cognitive Processes*, *14*(5-6), 609-630.

Appendix A

List of Selected Disyllabic Nouns and Verbs

Word	Tokens	Word	Tokens	Word	Tokens
shi2hou4(time)	226	kai1shi3(start)	21	xiao3de2(know)	7
jue2de2(think)	216	jing1ji4(economy)	21	xiang3yao4(want)	7
xian4zai4(now)	165	bie2ren2(someone else)	20	dian4hua4(telephone)	6
zi4ji3(oneself)	82	tong2xue2(classmate)	18	dian4ying3(movie)	4
zhi1dao4(know)	80	xiao3hai2(children)	15	zi4ran2(nature)	4
dong1xi1(things)	60	ping2chang2(usually)	17	gang1cai2(just)	4
da4jia1(all)	47	ji4de2(remember)	17	fa1sheng1(happen)	4
hou4lai2(later)	45	xu1yao4(need)	16	sheng1yi4(business)	3
jin1tian1(today)	44	zui4hou4(at last)	16	sheng1huo2(life)	3
jie2guo3(result)	42	guo2wang2(king)	16	guo2yu3	2
ban4fa3(solution)	29	xi1wang4(hope)	15	(national language)	
yuan4yi4(will)	27	deng3yu2(equal)	13	shu3yu2(belong)	2
xi3huan1(like)	26	she4hui4(society)	12	xia4wu3(afternoon)	2
bian4cheng2(turn into)	26	zheng4fu3(government)	12	yao4shi4(what if)	1
jie2yun4(MRT)	25	dian4shi4(TV)	11	nyu3hai2(girls)	1
peng2you3(friend)	24	shi4shi2(fact)	10	di4li3(geography)	1
gan3jue2(feel)	22	xian1sheng1(Mr.)	8	jia1ren2(family)	1

The Association for Computational Linguistics and Chinese Language Processing

(new members are welcomed)

Aims：

1. To conduct research in computational linguistics.
2. To promote the utilization and development of computational linguistics.
3. To encourage research in and development of the field of Chinese computational linguistics both domestically and internationally.
4. To maintain contact with international groups who have similar goals and to cultivate academic exchange.

Activities：

1. Holding the Republic of China Computational Linguistics Conference (ROCLING) annually.
2. Facilitating and promoting academic research, seminars, training, discussions, comparative evaluations and other activities related to computational linguistics.
3. Collecting information and materials on recent developments in the field of computational linguistics, domestically and internationally.
4. Publishing pertinent journals, proceedings and newsletters.
5. Setting of the Chinese-language technical terminology and symbols related to computational linguistics.
6. Maintaining contact with international computational linguistics academic organizations.
7. Dealing with various other matters related to the development of computational linguistics.

To Register：

Please send application to:

The Association for Computational Linguistics and Chinese Language Processing
Institute of Information Science, Academia Sinica
128, Sec. 2, Academy Rd., Nankang, Taipei 11529, Taiwan, R.O.C.

payment： Credit cards(please fill in the order form), cheque, or money orders.

Annual Fees：

regular/overseas member：	NT$ 1,000 (US$50.-)
group membership：	NT$20,000 (US$1,000.-)

life member：ten times the annual fee for regular/ group/ overseas members

Contact：

Address： The Association for Computational Linguistics and Chinese Language Processing
Institute of Information Science, Academia Sinica
128, Sec. 2, Academy Rd., Nankang, Taipei 11529, Taiwan, R.O.C.

Tel.：886-2-2788-3799 ext. 1502 Fax：886-2-2788-1638

E-mail: aclclp@hp.iis.sinica.edu.tw Web Site: http://www.aclclp.org.tw

Please address all correspondence to Miss Qi Huang, or Miss Abby Ho

The Association for Computational Linguistics and Chinese Language Processing

Membership Application Form

Member ID#： ________________

Name： __________________________________ Date of Birth： _______________

Country of Residence： ____________________ Province/State： ____________________

Passport No.： ___________________________ Sex: ______________________

Education(highest degree obtained)： ____

Work Experience： ____

____

Present Occupation： ____

Address： ____

____

Email Add： ____

Tel. No： ____________________________ Fax No： __________________________

Membership Category：□ Regular Member □ Life Member

Date： _____/____/____（Y-M-D）

Applicant's Signature：

Remarks： Please indicated clearly in which membership category you wish to register, according to the following scale of annual membership dues：

Regular Member ： US$ 50.- （NT$ 1,000）
Life Member ： US$500.-（NT$10,000）

Please feel free to make copies of this application for others to use.

Committee Assessment：

中華民國計算語言學學會

宗旨：

（一） 從事計算語言學之研究
（二） 推行計算語言學之應用與發展
（三） 促進國內外中文計算語言學之研究與發展
（四） 聯繫國際有關組織並推動學術交流

活動項目：

（一）定期舉辦中華民國計算語言學學術會議（Rocling）

（二）舉行有關計算語言學之學術研究講習、訓練、討論、觀摩等活動項目

（三）收集國內外有關計算語言學知識之圖書及最新發展之資料

（四）發行有關之學術刊物，論文集及通訊

（五）研定有關計算語言學專用名稱術語及符號

（六）與國際計算語言學學術機構聯繫交流

（七）其他有關計算語言發展事項

報名方式：

1. 入會申請書：請至本會網頁下載入會申請表，填妥後郵寄或E-mail至本會

2. 繳交會費：劃撥：帳號：19166251，戶名：中華民國計算語言學學會
信用卡：請至本會網頁下載信用卡付款單

年費：

終身會員：	10,000.-	（US$ 500.-）
個人會員：	1,000.-	（US$ 50.-）
學生會員：	500.-	（限國內學生）
團體會員：	20,000.-	（US$ 1,000.-）

連絡處：

地址：台北市115南港區研究院路二段128號 中研院資訊所(轉)
電話：(02) 2788-3799 ext.1502 傳真：(02) 2788-1638
E-mail：aclclp@hp.iis.sinica.edu.tw 網址: http://www.aclclp.org.tw
連絡人：黃琪 小姐、何婉如 小姐

中華民國計算語言學學會
個人會員入會申請書

會員類別	□終身 □個人 □學生	會員編號	（由本會填寫）		
姓名		性別		出生日期	年 月 日
				身分證號碼	
現職		學歷			
通訊地址	□□□				
戶籍地址	□□□				
電話		E-Mail			
				申請人:	（簽章）
				中華民國	年 月 日

審查結果:

1. 年費：

   終身會員： 10,000.-
   個人會員： 1,000.-
   學生會員： 500.-（限國內學生）
   團體會員： 20,000.-

2. 連絡處：

   地址：台北市南港區研究院路二段128號 中研院資訊所(轉)
   電話：(02) 2788-3799 ext.1502 傳真：(02) 2788-1638
   E-mail：aclclp@hp.iis.sinica.edu.tw 網址: http://www.aclclp.org.tw
   連絡人：黃琪 小姐、何婉如 小姐

3. 本表可自行影印

The Association for Computational Linguistics and Chinese Language Processing (ACLCLP)
PAYMENT FORM

Name: ____________________________________(Please print)   Date: _______________

Please debit my credit card as follows: US$ _____________

❑ VISA CARD ❑ MASTER CARD ❑ JCB CARD   Issue Bank:___________________

Card No.: ________ -_________-________ -_________   Exp. Date:____________(M/Y)

3-digit code: _________ (on the back card, inside the signature area, the last three digits)

CARD HOLDER SIGNATURE: ____________________________________

Phone No.: __________________________E-mail: _____________________________________

Address: ____

PAYMENT FOR

US$ _____ ❑ Computational Linguistics & Chinese Languages Processing (IJCLCLP)

Quantity Wanted: ______

US$ _____ ❑ Journal of Information Science and Engineering (JISE)

Quantity Wanted: ______

US$ _____ ❑ Publications:_________________________________

US$ _____ ❑ Text Corpora: _______________________________

US$ _____ ❑ Speech Corpora:______________________________

US$ _____ ❑ Others: ___________________________________

US$ _________ ❑ Membership Fees ❑ Life Membership ❑ New Membership ❑Renew

US$ _________ = Total

Fax 886-2-2788-1638 or Mail this form to:

ACLCLP
% IIS, Academia Sinica
Rm502, No.128, Sec.2, Academia Rd., Nankang, Taipei 115, Taiwan

E-mail: aclclp@hp.iis.sinica.edu.tw
Website: http://www.aclclp.org.tw

中華民國計算語言學學會
信用卡付款單

姓名: ______________________________(請以正楷書寫)　日期:：______________________

卡別：❑ VISA CARD　❑ MASTER CARD ❑ JCB CARD　發卡銀行：_______________

信用卡號：__________-__________-__________-__________　有效日期：__________(m/y)

卡片後三碼：_______________（卡片背面簽名欄上數字後三碼）

持卡人簽名：____(簽名方式請與信用卡背面相同)

通訊地址：____

聯絡電話：______________________________E-mail：___________________________________

備註：為順利取得信用卡授權，請提供與發卡銀行相同之聯絡資料。

付款內容及金額：

NT$___________ ❑ 中文計算語言學期刊(IJCLCLP) ________________

NT$___________ ❑ Journal of Information Science and Engineering (JISE)

NT$___________ ❑ 中研院詞庫小組技術報告______________________

NT$___________ ❑ 文字語料庫 ______________________________________

NT$___________ ❑ 語音資料庫 ______________________________________

NT$___________ ❑ 光華雜誌語料庫1976~2010

NT$___________ ❑ 中文資訊檢索標竿測試集/文件集

NT$___________ ❑ 會員年費：❑續會　❑新會員　❑終身會員

NT$___________ ❑ 其他: ______________________________________

NT$________________ = 合計

填妥後請傳真至 02-27881638 或郵寄至:

11529台北市南港區研究院路2段128號中研院資訊所(轉)中華民國計算語言學學會 收

E-mail: aclclp@hp.iis.sinica.edu.tw

Website: http://www.aclclp.org.tw

Publications of the Association for Computational Linguistics and Chinese Language Processing

			AIR	AIR		
		Surface	(US&EURP)	(ASIA)	VOLUME	AMOUNT
1.	no.92-01, no. 92-04(合訂本) ICG 中的論旨角色與 A Conceptual Structure for Parsing Mandarin -- Its Frame and General Applications--	US$ 9	US$ 19	US$15	_____	_____
2.	no.92-02 V-N 複合名詞討論篇 & 92-03 V-R 複合動詞討論篇	12	21	17	_____	_____
3.	no.93-01 新聞語料庫字頻統計表	8	13	11	_____	_____
4.	no.93-02 新聞語料庫詞頻統計表	18	30	24	_____	_____
5.	no.93-03 新聞常用動詞詞頻與分類	10	15	13	_____	_____
6.	no.93-05 中文詞類分析	10	15	13	_____	_____
7.	no.93-06 現代漢語中的法相詞	5	10	8	_____	_____
8.	no.94-01 中文書面語頻率詞典（新聞語料詞頻統計）	18	30	24	_____	_____
9.	no.94-02 古漢語字頻表	11	16	14	_____	_____
10.	no.95-01 注音檢索現代漢語字頻表	8	13	10	_____	_____
11.	no.95-02/98-04 中央研究院平衡語料庫的內容與説明	3	8	6	_____	_____
12.	no.95-03 訊息爲本的格位語法與其剖析方法	3	8	6	_____	_____
13.	no.96-01 「搜」文解字－中文詞界研究與資訊用分詞標準	8	13	11	_____	_____
14.	no.97-01 古漢語詞頻表（甲）	19	31	25	_____	_____
15.	no.97-02 論語詞頻表	9	14	12	_____	_____
16.	no.98-01 詞頻詞典	18	30	26	_____	_____
17.	no.98-02 Accumulated Word Frequency in CKIP Corpus	15	25	21	_____	_____
18.	no.98-03 自然語言處理及計算語言學相關術語中英對譯表	4	9	7	_____	_____
19.	no.02-01 現代漢語口語對話語料庫標註系統說明	8	13	11	_____	_____
20.	Computational Linguistics & Chinese Languages Processing (One year) (Back issues of *IJCLCLP*: US$ 20 per copy)	---	100	100	_____	_____
21.	Readings in Chinese Language Processing	25	25	21	_____	_____
				TOTAL	_____	_____

10% member discount: ___________Total Due:__________

- **OVERSEAS USE ONLY**
- PAYMENT： □ Credit Card (Preferred)
  □ Money Order or Check payable to "The Association for Computation Linguistics and Chinese Language Processing " or “中華民國計算語言學學會”
- E-mail：aclclp@hp.iis.sinica.edu.tw

Name (please print): ____________________________ Signature: ____________________________

Fax: ____ E-mail: ______________________________

Address：____

中華民國計算語言學學會

相關出版品價格表及訂購單

編號	書目	會　員	非會員	冊數	金額
1.	no.92-01, no. 92-04 (合訂本)　ICG 中的論旨角色 與 A conceptual Structure for Parsing Mandarin--its Frame and General Applications--	NT$ 80	NT$ 100	_____	_____
2.	no.92-02, no. 92-03 (合訂本) V-N 複合名詞討論篇 與V-R 複合動詞討論篇	120	150	_____	_____
3.	no.93-01　新聞語料庫字頻統計表	120	130	_____	_____
4.	no.93-02　新聞語料庫詞頻統計表	360	400	_____	_____
5.	no.93-03　新聞常用動詞詞頻與分類	180	200	_____	_____
6.	no.93-05　中文詞類分析	185	205	_____	_____
7.	no.93-06　現代漢語中的法相詞	40	50	_____	_____
8.	no.94-01　中文書面語頻率詞典（新聞語料詞頻統計）	380	450	_____	_____
9.	no.94-02　古漢語字頻表	180	200	_____	_____
10.	no.95-01　注音檢索現代漢語字頻表	75	85	_____	_____
11.	no.95-02/98-04　中央研究院平衡語料庫的內容與說明	75	85	_____	_____
12.	no.95-03　訊息為本的格位語法與其剖析方法	75	80	_____	_____
13.	no.96-01　「搜」文解字－中文詞界研究與資訊用分詞標準	110	120	_____	_____
14.	no.97-01　古漢語詞頻表（甲）	400	450	_____	_____
15.	no.97-02　論語詞頻表	90	100	_____	_____
16	no.98-01　詞頻詞典	395	440	_____	_____
17.	no.98-02　Accumulated Word Frequency in CKIP Corpus	340	380	_____	_____
18.	no.98-03 自然語言處理及計算語言學相關術語中英對譯表	90	100	_____	_____
19.	no.02-01 現代漢語口語對話語料庫標註系統說明	75	85	_____	_____
20	論文集 COLING 2002 紙本	100	200	_____	_____
21.	論文集 COLING 2002 光碟片	300	400	_____	_____
22.	論文集 COLING 2002 Workshop 光碟片	300	400	_____	_____
23.	論文集 ISCSLP 2002 光碟片	300	400	_____	_____
24.	交談系統暨語境分析研討會講義 （中華民國計算語言學學會1997第四季學術活動）	130	150	_____	_____
25.	中文計算語言學期刊（一年四期） 年份：______ （過期期刊每本售價500元）	---	2,500	_____	_____
26.	Readings of Chinese Language Processing	675	675	_____	_____
27.	剖析策略與機器翻譯 1990	150	165	_____	_____
			合　計	_____	_____

※　此價格表僅限國內 (台灣地區) 使用

劃撥帳戶：中華民國計算語言學學會　　劃撥帳號：19166251

聯絡電話：(02) 2788-3799 轉1502

聯絡人： 黃琪 小姐、何婉如 小姐　　E-mail:aclclp@hp.iis.sinica.edu.tw

訂購者：________________________ 收據抬頭：______________________________

地　　址：____

電　　話：__________________________ E-mail: _____________________________

Information for Authors

International Journal of Computational Linguistics and Chinese Language Processing (IJCLCLP) invites submission of original research papers in the area of computational linguistics and speech/text processing of natural language. All papers must be written in English or Chinese. Manuscripts submitted must be previously unpublished and cannot be under consideration elsewhere. Submissions should report significant new research results in computational linguistics, speech and language processing or new system implementation involving significant theoretical and/or technological innovation. The submitted papers are divided into the categories of regular papers, short paper, and survey papers. Regular papers are expected to explore a research topic in full details. Short papers can focus on a smaller research issue. And survey papers should cover emerging research trends and have a tutorial or review nature of sufficiently large interest to the Journal audience. There is no strict length limitation on the regular and survey papers. But it is suggested that the manuscript should not exceed 40 double-spaced A4 pages. In contrast, short papers are restricted to no more than 20 double-spaced A4 pages. All contributions will be anonymously reviewed by at least two reviewers.

Copyright：It is the author's responsibility to obtain written permission from both author and publisher to reproduce material which has appeared in another publication. Copies of this permission must also be enclosed with the manuscript. It is the policy of the CLCLP society to own the copyright to all its publications in order to facilitate the appropriate reuse and sharing of their academic content. A signed copy of the IJCLCLP copyright form, which transfers copyright from the authors (or their employers, if they hold the copyright) to the CLCLP society, will be required before the manuscript can be accepted for publication. The papers published by IJCLCLP will be also accessed online via the IJCLCLP official website and the contracted electronic database services.

Style for Manuscripts: The paper should conform to the following instructions.
1. Typescript: Manuscript should be typed double-spaced on standard A4 (or letter-size) white paper using size of 11 points or larger.
2. Title and Author: The first page of the manuscript should consist of the title, the authors' names and institutional affiliations, the abstract, and the corresponding author's address, telephone and fax numbers, and e-mail address. The title of the paper should use normal capitalization. Capitalize only the first words and such other words as the orthography of the language requires beginning with a capital letter. The author's name should appear below the title.
3. Abstracts and keywords: An informative abstract of not more than 250 words, together with 4 to 6 keywords is required. The abstract should not only indicate the scope of the paper but should also summarize the author's conclusions.
4. Headings: Headings for sections should be numbered in Arabic numerals (i.e. 1.,2....) and start form the left-hand margin. Headings for subsections should also be numbered in Arabic numerals (i.e. 1.1. 1.2...).
5. Footnotes: The footnote reference number should be kept to a minimum and indicated in the text with superscript numbers. Footnotes may appear at the end of manuscript
6. Equations and Mathematical Formulas: All equations and mathematical formulas should be typewritten or written clearly in ink. Equations should be numbered serially on the right-hand side by Arabic numerals in parentheses.
7. References: All the citations and references should follow the APA format. The basic form for a reference looks like

Authora, A. A., Authorb, B. B., & Authorc, C. C. (Year). Title of article. *Title of Periodical, volume number*(issue number), pages.

Here shows an example.

Scruton, R. (1996). The eclipse of listening. *The New Criterion, 15*(30), 5-13.

The basic form for a citation looks like (Authora, Authorb, and Authorc, Year). Here shows an example. (Scruton, 1996).
Please visit the following websites for details.
(1) APA Formatting and Style Guide (http://owl.english.purdue.edu/owl/resource/560/01/)
(2) APA Stytle (http://www.apastyle.org/)

No page charges are levied on authors or their institutions.

Final Manuscripts Submission: If a manuscript is accepted for publication, the author will be asked to supply final manuscript in MS Word or PDF files to clp@hp.iis.sinica.edu.tw

Online Submission: http://www.aclclp.org.tw/journal/submit.php

Please visit the IJCLCLP Web page at http://www.aclclp.org.tw/journal/index.php

International Journal of
Computational Linguistics & Chinese Language Processing
vol.18 no.3 September 2013

Contents

Special Issue Articles:
Processing Lexical Tones in Natural Speech

ISSN: 1027-376X
The Association for Computational Linguistics and Chinese Language Processing