Article

Language and Speech

Language *and* Speech
2016, Vol. 59(1) 59–82
© The Author(s) 2015
Reprints and permissions:
sagepub.co.uk/journalsPermissions.nav
DOI: 10.1177/0023830915578000
las.sagepub.com
SAGE

Constraints of Tones, Vowels and Consonants on Lexical Selection in Mandarin Chinese

Seth Wiener
Department of East Asian Languages and Literatures, The Ohio State University, USA; Department of Modern Languages, Carnegie Mellon University, USA

Rory Turnbull
Department of Linguistics, The Ohio State University, USA

Abstract

Previous studies have shown that when speakers of European languages are asked to turn nonwords into words by altering either a vowel or consonant, they tend to treat vowels as more mutable than consonants. These results inspired the universal vowel mutability hypothesis: listeners learn to cope with vowel variability because vowel information constrains lexical selection less tightly and allows for more potential candidates than does consonant information. The present study extends the word reconstruction paradigm to Mandarin Chinese–a Sino-Tibetan language, which makes use of lexically contrastive tone. Native speakers listened to word-like nonwords (e.g., *su3*) and were asked to change them into words by manipulating a single consonant (e.g., *tu3*), vowel (e.g., *si3*), or tone (e.g., *su4*). Additionally, items were presented in a fourth condition in which participants could change any part. The participants' reaction times and responses were recorded. Results revealed that participants responded faster and more accurately in both the free response and the tonal change conditions. Unlike previous reconstruction studies on European languages, where vowels were changed faster and more often than consonants, these results demonstrate that, in Mandarin, changes to vowels and consonants were both overshadowed by changes to tone, which was the preferred modification to the stimulus nonwords, while changes to vowels were the slowest and least accurate. Our findings show that the universal vowel mutability hypothesis is not consistent with a tonal language, that Mandarin tonal information is lower-priority than consonants and vowels and that vowel information most tightly constrains Mandarin lexical access.

Keywords

Lexical tone, Mandarin Chinese, segments and suprasegmentals, lexical selection

Corresponding author:
Seth Wiener, Department of Modern Languages, Carnegie Mellon University, 160 Baker Hall, 5000 Forbes Avenue, Pittsburgh, PA, 15213, USA.
Email: sethw1@cmu.edu

1 Introduction

During spoken-word recognition, listeners activate multiple lexical candidates, all of which compete with one another. Competition may be due to candidates sharing an initial sound (Connine, Blasko, & Titone, 1993; Marslen-Wilson, 1990), partial overlap (McQueen, Norris, & Cutler, 1994; Tabossi, Burani, & Scott, 1995), or even sharing segments that are embedded (Gow & Gordon, 1995; Shillcock, 1990). Because activation occurs continuously as the speech signal unfolds, listeners tolerate a considerable amount of mismatch and activate similar-sounding phonemes (Connine, Blasko, & Wang, 1994; McQueen, Norris, & Cutler, 1999). From this perspective, minimally different lexical neighbors play an important role in word activation. Yet, not all information in the signal contributes equally to word activation or inhibition. It may be the case that within the speech signal some phonemes contribute more to word recognition, while other phonemes constrain word recognition more effectively.

This theory was directly tested in several European languages by comparing how consonants and vowels differ in their contribution to spoken word recognition. In van Ooijen's (1996) word reconstruction task, listeners heard spoken nonwords that could be changed to real words by altering either a consonant or a vowel. For example, listeners heard the nonword *kebra* and could either change a consonant to make *zebra* or a vowel to make *cobra.* The listener's task was to respond as quickly as possible with the first real word that came to mind. Van Ooijen tested participants in three conditions: a forced condition in which participants were told they must change only the vowel, a forced condition in which listeners must only change consonants, and a free choice condition in which listeners could change whichever phoneme they wanted; this setup allowed for multiple comparisons between phoneme type and across conditions. Van Ooijen's results showed a clear asymmetry: when given a free choice, participants tended to change vowels more rapidly than consonants. Additionally, participants tended to make more erroneous vowel changes in the forced consonant condition than erroneous consonant changes in the forced vowel condition. Van Ooijen concluded that English listeners have learned to treat vowels as less exact and more mutable than consonants.

In a follow-up study, Cutler, Sebastián-Gallés, Soler-Vilageliu, and van Ooijen (2000) explored whether this vowel asymmetry was an artifact of English's skewed vowel to consonant ratio. Since there were fewer candidates for vowel substitution, the authors argued that accuracy and reaction time might have been the result of a simpler lexical search. This explanation was contrasted with an acoustic closeness hypothesis in which the observed effect may have stemmed from English vowels having more acoustic overlap and potential for perceptual confusion than English consonants. These two hypotheses were tested by replicating van Ooijen's reconstruction study with Dutch and Spanish listeners. Dutch allowed for a relatively balanced vowel–consonant ratio along with several acoustically similar vowels, while Spanish provided a much higher ratio of consonants to vowels and a high degree of distinctiveness among the vowels. Cutler et al.'s findings corroborated van Ooijen's English results: Dutch and Spanish vowels were altered significantly more quickly and more accurately than consonants, and participants more often chose to change the vowel than the consonant in the free choice condition. These results led Cutler et al. to reject both the vowel–consonant ratio hypothesis and the acoustic closeness hypothesis. Cutler et al., together with van Ooijen's original study and a third reconstruction study by Marks, Moates, Bond and Stockmal (2002),[1] offer evidence from three European languages which together suggests that vowel information constrains lexical selection less tightly than consonant information and thus allows for more potential lexical candidates.

While Cutler et al. and van Ooijen's studies offer a convincing argument for a potentially universal intrinsic vowel mutability hypothesis, the studies were all limited to European languages.

Additional data from unrelated, non-European languages are needed to strengthen this claim. Furthermore, it remains unclear how suprasegmental information interacts with vocalic information and how readily listeners access candidates that differ only in suprasegmentals. Although both languages Cutler et al. examined use stress to mark lexical distinctions, the facilitatory (or inhibitory) effect of suprasegmental information on their results was not discussed. This oversight leaves open the possibility that vowels are more mutable only if lexically contrastive suprasegmental cues are absent.

Previous work has shown that suprasegmental information can constrain lexical access in lexical tone languages (Fox & Unkefer, 1985), lexical pitch-accented languages (Cutler & Otake, 1999), and lexical stress languages (Cutler & van Donselaar, 2001). Crucially, it appears that the use and role of suprasegmental cues in lexical access is language specific. English listeners do not rely heavily on stress information, presumably due to the limited number of stress minimal pairs (Cooper, Cutler, & Wales, 2002; Creel, Tanenhaus, & Aslin, 2006; Cutler, 1986). Dutch and Italian listeners tend to rely slightly more on stress, since it can temporarily provide more disambiguating information than segmental cues (Cutler & van Donselaar, 2001; Sulpizio & McQueen, 2012; Tagliapietra & Tabossi, 2005). Soto-Faraco, Sebastián-Gallés, & Cutler (2001) argued that, for Spanish listeners, suprasegmental information can constrain lexical activation as effectively as segmental cues can. Listeners of a lexical tone language, in which syllables can be differentiated by contrastive pitch cues, may rely on suprasegmental information to an even greater degree. Therefore, a tone language such as Mandarin Chinese can provide insight into the degree to which suprasegmental cues constrain lexical access vis-à-vis vowels and consonants, while also providing the first test of the universal intrinsic vowel mutability hypothesis with a non-Indo-European language.

1.1 Mandarin Chinese

Standard Beijing Mandarin Chinese (*pu3tong1hua4*) features four lexical tones: tone 1–high level; tone 2–low rising; tone 3–low dipping; tone 4–high falling. Owing to their acoustic features, Mandarin and Cantonese tones were assumed to be processed much later than segmental cues and contribute less to word recognition (Cutler & Chen, 1997; Taft & Chen, 1992). A large body of experimental evidence from conscious judgment tasks (i.e., tone monitoring, same–different judgments and lexical decision) concluded that tone recognition is slower and less accurate than segmental recognition, and thus plays a reduced, secondary role in word recognition (Cutler & Chen, 1997; Repp & Lin, 1990; Sereno & Lee, 2014; Taft & Chen, 1992; Ye & Connine, 1999). However, more recent studies utilizing online measures such as event-related potentials (ERP) and eye-tracking present a slightly different account of tone processing, since these measures allow for a continuous examination into the natural time course of tone processing. This body of work argues for parallel processing of segmental and tonal information as the speech signal unfolds in time (Brown-Schmidt & Canseco-Gonzalez, 2004; Malins & Joanisse, 2010, 2012; Schirmer, Tang, Penny, Gunter, & Chen, 2005; Zhao, Guo, Zhou, & Shu, 2011). For example, Zhao et al. (2011) had participants compare whether two subsequent pictures belonged to the same semantic category. Two images were presented with a spoken word played in between. The word matched or mismatched the first picture in four ways: onset mismatch, rime mismatch, tone mismatch or full syllable mismatch. All three partial mismatches led to similar N400 modulations–an ERP component related to semantic integration–suggesting that tone information may play an equally important role as segmental cues in lexical semantic integration. Likewise, Malins and Joanisse (2012) had participants judge whether auditory words matched visually presented images. For example, participants saw a visual image of a flower (*hua1*) and then were presented with an auditory match (*hua1*) or one of five auditory mismatches: tonal mismatch (*hua4*, “to draw”), rime mismatch (*hui1*, “ash”), onset mismatch (*gua1*, “melon”), segmental mismatch (*jing1*, “energy”) or unrelated mismatch (*lang2*, “wolf”). The authors found similar phonological mapping negativity and N400 modulations for tone and phonemic anomalies, suggesting that listeners use tonal information as soon as it becomes available to constrain word recognition.

Additionally, Wiener and Ito (2014) reported eye-tracking results that suggest the role of tone in lexical access is highly gradient and dependent on the uniqueness of each syllable–tone combination in the lexicon. The authors point out that the approximately 400 syllables and four distinct tone types should result in 1600 syllable–tone combinations.[2] Yet of the theoretical 1600 combinations, only approximately 1300 unique syllable–tone combinations are attested morphemes (Duanmu, 2007). Because of the unpredictable assignment of lexical tone to syllables and the vast frequency and probability differences, Wiener and Ito concluded that speakers learn that certain syllables may have a greater informative need to lexically represent tone. Their finding also adds to the growing consensus of the syllable as the essential, or proximate, unit of Mandarin perception and production (Chen & Chen, 2013; Chen, Chen & Dell, 2002; Mok, 2009; O’Seaghda, Chen, & Chen, 2010; Tong, Francis, & Gandour, 2008; Sereno & Lee, 2014; You, Zhang, & Verdonschot, 2012).

In sum, our present understanding of Mandarin and Cantonese tone processing suggests that listeners can make use of suprasegmental information as quickly as segmental information, though its relative contribution may be task, context and frequency dependent (Lee, 2007; Liu & Samuel, 2007; Mattys, White, & Melhorn, 2005; Soto-Faraco et al., 2001; Wiener & Ito, 2014; Xu, 1997, 1998, 1999; Ye & Connine, 1999;). Therefore we seek to clarify to what degree tonal information is used to constrain lexical selection. Furthermore, if tonal information contributes less to lexical selection than segments (e.g., Cutler & Chen, 1997; Repp & Lin, 1990; Sereno & Lee, 2014; Taft & Chen, 1992; Ye & Connine, 1999), we wish to determine whether vowels or consonants play a more important role in Mandarin word recognition. Given that the experimental paradigm can strongly affect how tonal information is used, we believe the word reconstruction task offers the ideal means to tease apart the processing of tone and, most importantly, to test tone’s relative contribution as compared to vowels and consonants. As van Ooijen (1996) and Cutler et al. (2000) have outlined, the word reconstruction task is unique in that it (1) insures that lexical access takes place; (2) allows for a direct comparison of the effects of tone, consonant and vowel with the same experimental item; and (3) requires listeners to directly alter their acoustic percept of the stimulus without orthographic stimulation–a crucial, yet often overlooked component to studying Mandarin and Cantonese word recognition given the nature of the native orthography (see Zhou and Marslen-Wilson, 1999a, 1999b and Perfetti and Zhang, 1991 for further discussion). In other words, the nonword stimulus participants hear serves as a perceptual template via which multiple real words can be recognized (Cutler et al., 2000; van Ooijen, 1996).

The present study uses the word reconstruction task to clarify (1) whether the intrinsic vowel mutability hypothesis can be extended to a Sino-Tibetan language, (2) to what degree tonal mutability constrains lexical selection in Mandarin Chinese, and (3) if tone information contributes less to lexical selection than segmental information, whether vowels or consonants play a more important role in the process. To answer these questions, two experiments were conducted. Experiment 1 first assessed nonword stimuli by lexical decision, while experiment 2 tested the nonwords using the word reconstruction task.

2 Experiment 1–lexical decision

Unlike the Dutch, Spanish and English nonword stimuli used in previous word reconstruction studies (e.g., *kebra*), which were all highly word-like templates assembled in the same phonotactic manner, Mandarin stimuli can come from multiple different categories of nonwords. Of particular interest to the present study are tonotactic accidental gap nonwords (see Wang, 1998). These nonwords consist of phonotactic combinations that do not violate Mandarin phonotactics and form morphemes with one, two or three of the four lexical tones, but not all four tones, thus forming a "gap" in the possible syllable–tone combinations. These gaps can partly be accounted for due to the historic evolution of tone within Mandarin (Ho, 1976; Norman, 1988), though some are simply accidental gaps that lack systematic explanations (Wang, 1998). Since these gaps exist only on the tonal level, it is an open question as to whether they are processed similar to previous reconstruction nonwords such as the English nonword *kebra*.

Therefore, given that tonotactic accidental gaps are nonwords only when the suprasegmental level is considered, we first report results from a Mandarin lexical decision task. These results allowed us to evaluate whether our findings from the word reconstruction task were an artifact of the stimuli. Additionally, lexical decision allowed us to explore how the syllable–tone gaps were processed relative to systematic nonwords, allowed for planned comparisons between analogous real words, and allowed for a comparison between our results and those of The English Lexicon Project (Balota et al., 2007). Furthermore, to the best of our knowledge, only Wang (1998), Myers (2002) and Myers and Tsay (2005) have explored the processing of tonotactic accidental gap nonwords in Mandarin (see Kirby and Yu (2007) for Cantonese), albeit using word acceptability judgments rather than lexical decision. The results reported here therefore represent one of the first systematic investigations of the lexical processing of tonotactic accidental gaps in Mandarin Chinese.

2.1 Methods

2.1.1 Participants. Forty subjects (29 females, 11 males) participated in experiment 1, and were paid for their time. The mean age of all subjects was 26 years and all subjects were native speakers of Mandarin. All subjects came from the Northern Mandarin region (Norman, 1988), ensuring that they did not speak non-standard dialects. All the subjects were students at The Ohio State University and therefore spoke English as a second language at a highly proficient level. None of the subjects reported any speech or hearing disorders.

2.1.2 Materials. The experimental stimuli included 80 targets, consisting of 40 monosyllabic words (abbreviation: WORD) (e.g., *gong1* "work") and 40 nonwords. The nonword stimuli consisted of two different types: systematic violation of phonotactics (abbreviation: VIOL) such as consonant clusters or illegal codas (e.g., *yom3*) and tonotactic accidental gaps (abbreviation: TAG) in which the occurrence of a particular tone with a legal syllable results in a syllable–tone nonword (e.g., *su3*). Appendix 1 includes the complete stimuli set.

The 40 words were balanced for tone (10 targets per tone), syllable frequency and syllable–tone frequency using SUBTLEX-CH (Cai & Brysbaert, 2010), with 20 high-frequency targets and 20 low-frequency targets. Nonword targets that consisted of a legal syllable (i.e., the TAG condition) were similarly balanced for syllable frequency.

2.1.3 Procedure. The stimuli were recorded by a phonetically trained female native speaker of Mandarin Chinese using a TASCAM DR-40 digital recorder at a sampling rate of 44.1 kHz. The experiment was conducted using OpenSesame 0.25 (Mathôt, Schreij, & Theeuwes, 2012). Participants were tested in a quiet room on the campus of The Ohio State University. The entire experiment was conducted in Chinese, including the informed consent and debriefing procedures. Stimuli were presented over headphones, and participants were instructed to respond with a word (是) or nonword (不是) response as quickly and as accurately as possible. The response keys were "Z" and "/" on a standard US keyboard, with the meaning counterbalanced so that for half of the participants the "word" response was on the left, and for the other half of the participants the "word" response was on the right.

The stimuli were counterbalanced across two lists and pseudo-randomized such that no more than two consecutive trials were from the same condition. Participants began the experiment with 10 practice trials to provide familiarization with the task. The interval between trials was one second. After completion of all trials, participants completed a brief language background questionnaire. The experiment took approximately 15 minutes.

2.1.4 Analysis. Reaction time from the offset of the stimulus was recorded, as was the keyboard response. Reaction times shorter than 100ms were excluded from the analysis (resulting in the removal of 86 measurements), as were reaction times longer than 3500ms (48 measurements). Thus, 3066 tokens were available for analysis. Mixed effects logistic regression modeling was used to predict response ("word" or "nonword"). A fixed effect of syllable type (WORD, VIOL, or TAG) was entered, along with random intercepts of participant and stimulus token, with a random slope of syllable type for participant. Two models were constructed: one where the baseline for the syllable type factor was the WORD level, and another where the baseline was the TAG level. These two models thus allowed all three possible comparisons between the factor levels (each model only permits *n*-1 comparisons, where *n* is the number of factor levels; see Clopper, 2013).

Mixed effects linear regression modeling was used to predict log reaction time. For this analysis, only the subset of correct responses was analyzed, resulting in 2626 tokens for analysis. A fixed effect of syllable type (WORD, VIOL, or TAG) was entered, along with random intercepts of participant and stimulus token, with a random slope of syllable type for participant. As in the response type analysis, one model with WORD as the baseline was constructed along with a model with TAG as the baseline, to allow for comparisons between all three conditions. P-values were calculated from the t-values, using the number of observations minus the number of fixed effect parameters as an estimate of the upper bound on the degrees of freedom (Baayen, Davidson & Bates, 2008). All mixed effects modeling used the lme4 package (Bates, Maechler, & Bolker, 2014) in R (version 3.0.2).

2.2 Results

Figure 1 shows mean "word" endorsement rates for each syllable type. The logistic regression model revealed significant differences between WORD stimuli with every other syllable type, such that WORD stimuli were far more likely to be endorsed as words than TAG or VIOL stimuli. Additionally, TAG stimuli were significantly more likely to be endorsed as words than VIOL stimuli, suggesting that the tonotactic gaps were more acceptably word-like than the other nonwords. The output of these models is summarized in Table 1.

Similar effects were observed for the analysis of reaction time. Figure 2 shows mean reaction times of correct responses to the three stimuli types. Reactions to the TAG stimuli were significantly slower than reactions to the WORD or VIOL stimuli. These models are summarized in Table 2.

2.3 Discussion

As expected, participants correctly identified the WORD stimuli as real words, and did so relatively quickly. Similarly, the VIOL stimuli, which consisted of phonotactic violations such as *yom3*, were also responded to relatively quickly. The low number of "word" responses to the VIOL stimuli demonstrates that participants correctly identified these stimuli as nonwords. The TAG stimuli, on the other hand, were nonwords that conceivably could be words, and this difference between the TAG and the VIOL stimuli was reflected in the results. Overall, participants were much slower and less accurate with TAG stimuli compared to either of the other types. Taken together, these results

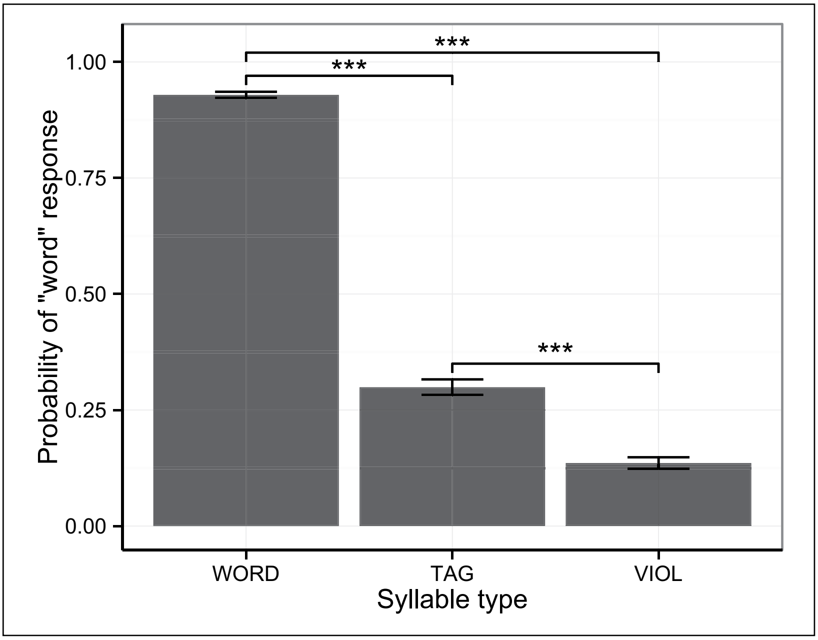

Figure 1. Mean “word” endorsement rates for each syllable type.

Table 1. Model output from the logistic mixed effects models predicting response.

	B	*SE*	*z*	*p*
Baseline: WORD				
(Intercept)	3.7823	0.3672	10.30	< 0.001
TAG	–5.0240	0.5503	–9.13	< 0.001
VIOL	–7.1564	0.5903	–12.12	< 0.001
Baseline: TAG				
(Intercept)	–1.2416	0.4381	–2.834	0.005
WORD	5.0239	0.5503	9.130	< 0.001
VIOL	–2.1325	0.6329	–3.369	< 0.001

suggest that the VIOL and WORD stimuli were very quickly and easily identified by the participants as being nonwords or words, respectively, while the TAG stimuli appeared word-like and initiated lexical access, but required further processing (such as lexical search) to rule them out. This finding underpins the notion that the TAG stimuli are extremely word-like, much like the English nonword *kebra* (cf. Balota et al., 2007). The results of this experiment therefore motivate and justify the use of the TAG stimuli as stimuli in a word reconstruction task.

3 Experiment 2–word reconstruction

3.1 Methods

3.1.1 Participants. Thirty-two subjects (20 females, 12 males) participated in Experiment 2. The mean age of all subjects was 26. The same criteria as in Experiment 1 were used in selecting these

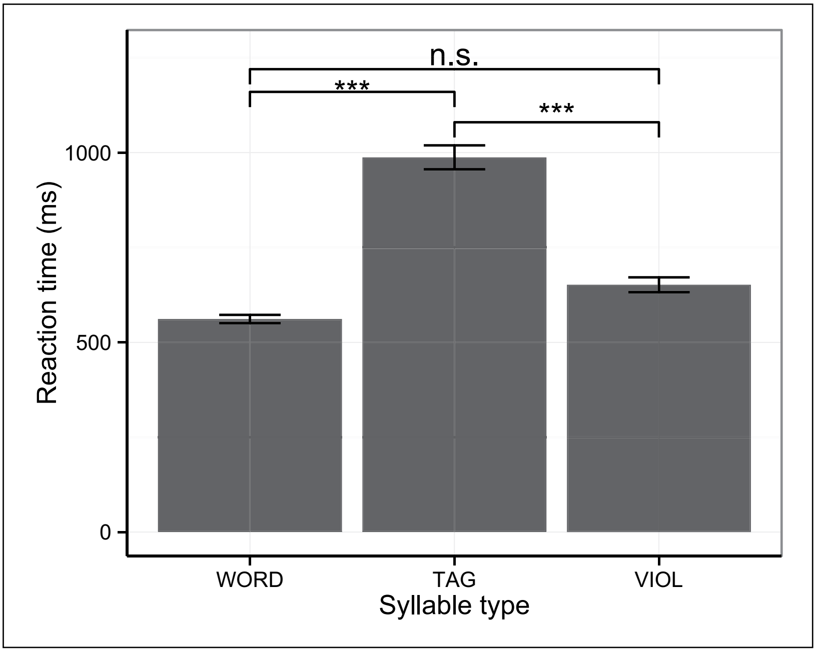

Figure 2. Mean lexical decision reaction times (for correct responses only). n.s.: not significant.

Table 2. Model output from the linear mixed effects models predicting reaction time.

	B	*SE*	*t*	*p*
Baseline: WORD				
(Intercept)	6.12919	0.05498	111.49	< 0.001
TAG	0.52825	0.09008	5.86	< 0.001
VIOL	0.14305	0.08024	1.78	0.075
Baseline: TAG				
(Intercept)	6.65744	0.08618	77.25	< 0.001
WORD	–0.52825	0.09008	–5.86	< 0.001
VIOL	–0.38521	0.09368	–4.11	< 0.001

participants. None of the subjects had participated in Experiment 1. All participants were paid for their time.

3.1.2 Materials. Ninety-six monosyllabic TAG nonwords were chosen. These stimuli consist of syllable–tone gaps in standard Mandarin. No syllable was repeated, allowing for an even distribution of the overall Mandarin syllable frequency. Each nonword could be changed into a real Mandarin word by substitution of vowel, consonant or tone. For example, the nonword *su3* could become *su4* (“fast”) via tone change, *si3* (“to die”) via vowel change or *tu3* (“soil”) via consonant change, among other potential changes. Sixty-four of the syllables were (C)(G)V syllables without a coda, while 32 of the syllables contained a nasal coda. This approximates the natural ratio of open to nasal coda syllables in the Mandarin lexicon (Duanmu, 2007). For every Mandarin stimulus word, there were multiple substitutions possible–for example, the nonword stimulus *su3* could be changed in the tone condition to either of the words *su4* ("fast") or *su2* ("custom"). This fact was true of some but not all of the stimuli in the English, Dutch and Spanish sets of van Ooijen (1996) and Cutler et al. (2000). This is an artifact of Mandarin's relatively simple, unmarked syllable structure (with no complex onsets or codas) and the lack of ambisyllabicity (Duanmu, 2007, 2008).

It is clear that the number of possible substitutions for each stimulus nonword varies as a function of the neighborhood density of the nonword (Vitevitch, Luce, Pisoni, & Auer, 1999). Therefore, neighborhood density was calculated for each nonword as the number of lexical neighbors differing from the nonword by one phoneme deletion, substitution, or insertion; this measure was cross-referenced with MiniCorpJS (Myers, 2008). Measures of "sub"-neighborhood density were also calculated for changes to the consonant alone, the vowel alone, and the tone alone. For example, the neighborhood density of the nonword *su3* is 34; 21 of these neighbors involve a consonant change, ten involve a vowel change, and three involve a tone change. Finally, from these neighbors, an estimated "nearest" neighbor was calculated by searching SUBTLEX-CH (Cai & Brysbaert, 2010) to control for the most frequent word in each sub-neighborhood set. For example, the nonword *su3*'s "nearest" tone change was *su4*, the most common form of the syllable *su*. See Appendix 2 for the full stimuli set.

The nonwords were divided into four sets (A, B, C, D) of 24 items each. Each set also included four practice trials. The four sets were rotated through four different substitution conditions: vowel substitution, consonant substitution, tone substitution and free substitution (vowel, consonant or tone) resulting in a within-subject design. For example, participants in list one heard the set A nonwords in the vowel substitution condition, while participants in list two heard the set A nonwords in the consonant substitution condition. Therefore, while all the participants heard all 96 items, the particular items heard in each substitution condition were counterbalanced across participants. In other words, responses for the nonword *su3* were observed in all four substitution conditions, although each participant only responded to *su3* once.

Additionally, to further ensure our materials were interpretable and that all participants were familiar with the terminology in the instructions, five native Mandarin speakers, none of whom participated in the study, listened to the full stimuli set and were asked to report the initial, final (see "procedure" for explanation of final/rhyme terminology) and tone of each nonword. There was 100% agreement across all five listeners.

3.1.3 Procedure. The experiment was conducted using OpenSesame 0.25 (Mathôt et al., 2012). Participants were tested in a quiet room on the campus of The Ohio State University. As with experiment 1, the entire experiment was conducted in Chinese. Stimuli were presented over headphones, and participants were instructed to listen carefully to the nonword presented, press the space bar as soon as they had thought of a real word in standard Mandarin (*pu3tong1hua4*) according to the condition requirements (e.g., change tones only), and then say the word aloud. Participants were not made aware beforehand that they would be asked to change different phonemic categories after each condition. Each new condition was always prefaced by new instructions and four practice trials. Because traditional Chinese phonology (and modern Chinese education) divides the syllable into two non-decomposable parts–an "initial" *sheng1mu3* and a "final" *yun4mu3* (Baxter, 1992; Chao, 1968)–this wording rather than the Mandarin words for "consonant" and "vowel" were used. For each condition, all experimental items were presented in the same pseudo-randomized order, but the order of conditions presented was counterbalanced across subjects. For example, one fourth of the participants started with the consonant substitution condition, while another fourth started with the vowel substitution condition. Following van Ooijen (1996) and Cutler et al. (2000), each trial had a 10 second response window before hearing a timeout error sound. The interval between trials was two seconds. Reaction time measured by the participants' keypress responses and oral responses (recorded using a TASCAM DR-40 digital recorder) were collected. After completion of all trials, participants completed a brief language background questionnaire. The experiment took approximately 25 minutes.

3.1.4 Analysis. 52 responses (less than 2%) were removed due to inaudibility and recording errors. Remaining verbal responses were coded as correct or erroneous.

Accuracy data were analyzed using mixed effects logistic regression modeling. Reaction time data were analyzed using mixed effects linear regression modeling. Fixed effects entered were condition (free choice, vowel change, tone change, or consonant change, with free choice as baseline) and number of lexical neighbors of the stimulus. Random intercepts of subject and stimulus were included, as were random slopes of condition and number of neighbors for subject, and a random slope of condition for stimulus. As in experiment 1, p-values for the linear model were calculated from the t-values and an estimate of the degrees of freedom.

The free change condition, where participants were free to alter any of the consonant, vowel, or tone, can be regarded as a control or baseline condition. It is in this condition where we expect the highest accuracy rates and fastest response times, following Cutler et al. (2000). In this condition, the participant is able to respond with the first most highly activated neighbor of the nonword. In all other conditions, the participant must inhibit inappropriate neighbors–such as a tonal change in the consonant change condition–and thus responses are expected to be somewhat slower and less accurate. The level of inhibition differs from condition to condition: for example, owing to Mandarin's small tonal inventory, relative to the consonant or vowel inventories, the tonal changes possible are necessarily few and thus require high levels of neighborhood inhibition. Although words and nonwords are processed through different mechanisms (Vitevitch & Luce, 1998, 1999; Vitevitch, 2003), since this task involves both the perception of a nonword and the activation and production of a word, the predicted effects of lexical versus sublexical processing on this task are unclear. Nevertheless, if number of possible neighbors and the required level of inhibition influences speed and accuracy of response, then we expect the tonal change condition to be the most different from the free change condition.

3.1.5 Results. Table 3 presents the mean correct response latencies and the mean overall error rates for the four conditions. Figure 3 depicts mean accuracy rates in each of the four response conditions. The logistic mixed effects regression model, summarized in Table 4, revealed significant effects for the consonant and vowel condition but not for the tone condition: the free choice condition was significantly more accurate than the consonant and vowel change conditions. Unexpectedly, the vowel change condition was significantly *less* accurate than the consonant change condition, and the tone change condition was significantly *more* accurate than the consonant change condition. A significant effect of neighborhood density was also observed, such that stimuli with more lexical neighbors had more accurate responses than stimuli with fewer lexical neighbors.

Figure 4 depicts mean reaction times per condition, and Table 5 provides a summary of the linear model output. Unlike in the accuracy data, significant effects of condition were observed for all three conditions relative to the baseline: both the vowel and consonant change conditions were significantly slower than the free change condition (with the vowel change condition slower than the consonant change condition), while the tone change condition was significantly faster than the free change condition. Additionally, an effect of neighborhood density was observed, such that words with more lexical neighbors had faster responses than words with few lexical neighbors. This effect is consistent with a framing of the word completion task as a search through the lexicon for a word that is appropriately similar to the stimulus.

Table 3. Error rates and mean reaction times per condition (RTs for correct responses only).

Condition	free	tone	consonant	vowel
% error	19.8	22.2	43.1	57.3
mean RT (ms)	2404	2073	3517	4103

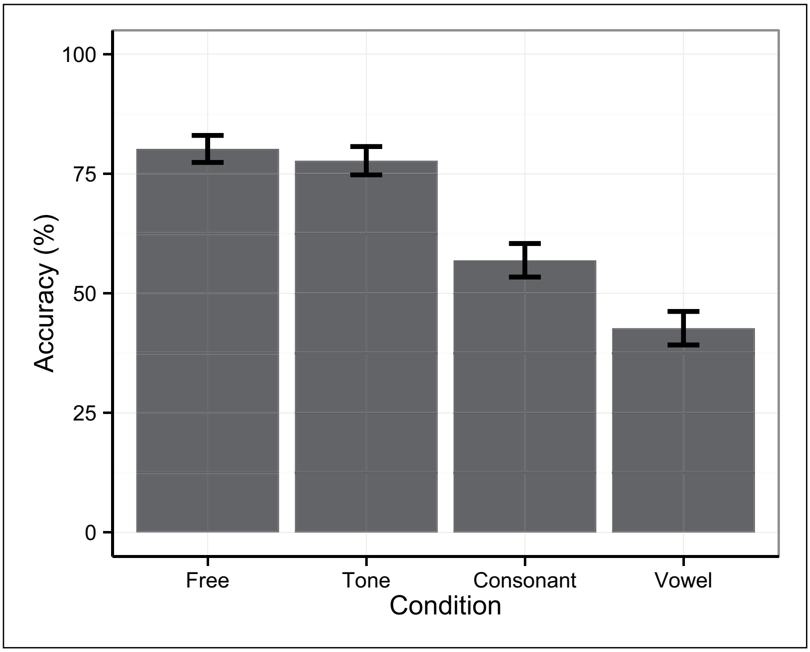

Figure 3. Mean accuracy rates in each of the four response conditions.

Table 4. Output from logistic mixed effects model predicting response accuracy.

	β	*SE*	*z*	*p*
(Intercept)	1.01817	0.31548	3.227	<0.01
Neighborhood density	0.02018	0.01017	1.984	<0.05
Consonant change	−1.18902	0.23290	–5.105	<0.001
Tone change	0.27719	0.28324	0.979	0.327
Vowel change	−1.88541	0.25165	–7.492	<0.001

Table 6 shows the counts of types of changes (tone, consonant, or vowel) of correct responses in the free choice condition. The overwhelming preference for tonal changes is significant, $\chi^2(2) = 207.3$, $p < 0.001$, again challenging the notion that vowel change is the default option. Indeed, changing the vowel was the least common type of change.

3.1.6 Discussion. In contrast with previous studies on European languages, these data do not support the hypothesis that vowel substitution in word reconstruction is the default or fastest operation. Indeed, these data demonstrate that the vowel change condition had among the longest

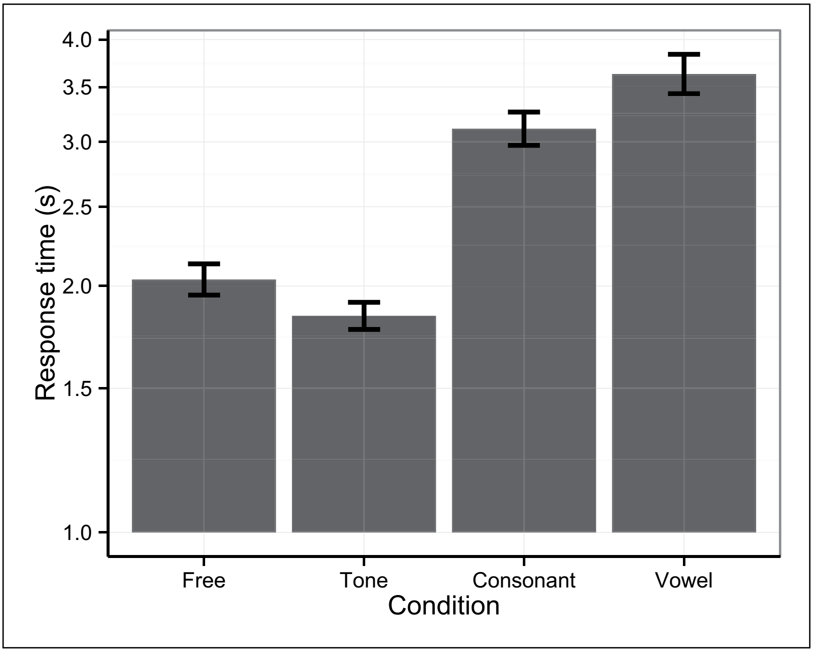

Figure 4. Mean log reaction times per condition (for correct responses only).

Table 5. Output from linear mixed effects model predicting log reaction time.

	β	*SE*	*t*	*p*
(Intercept)	0.033	0.088	0.380	0.704
Neighborhood density	–0.006	0.001	–3.161	<0.01
Consonant change	0.384	0.072	5.287	<0.001
Tone change	–0.133	0.062	–2.137	<0.05
Vowel change	0.481	0.057	8.386	<0.001

Table 6. Count of different types of change in the free choice condition.

	Tone change	Consonant change	Vowel change
Count (percentage)	366 (59.5%)	167 (27.2%)	82 (13.3%)

reaction times and the highest error rate of any condition. In contrast, the tone change condition was practically equivalent with the free choice condition in terms of accuracy and significantly faster in terms of reaction time. Finally, consonant changes were made faster and more accurately than vowel changes, though slower and less accurately than tone changes. In the free choice condition, participants overwhelmingly preferred to change the tone rather than the consonant or vowel. Neighborhood density significantly affected both accuracy and response times corroborating previous research that has shown that nonwords that are phonologically similar to a large number of real words are responded to more quickly than nonwords occurring in sparse neighborhoods (Vitevitch, Luce, Charles-Luce, & Kemmerer, 1997; Vitevitch et al., 1999).

4 General discussion

The present study set out to use the word reconstruction task to clarify (1) whether the intrinsic vowel mutability hypothesis can be extended to a Sino-Tibetan language, (2) to what degree tonal mutability constrains lexical selection in Mandarin Chinese, and (3) if tone information constrains lexical selection less than segmental information, whether vowels or consonants play a more important role in the process. With regards to the first research question, it is clear that the intrinsic vowel mutability hypothesis cannot be extended to Mandarin Chinese. The results from our word reconstruction task suggest that speakers of a tonal language like Mandarin rely on vowel information more than tone or consonant information. Mandarin speakers, therefore, do not treat vowel information as mutably as speakers of European languages do, most likely because of the lexical role of tone. Therefore we feel confident in concluding that the vowel effect is not universal. The intrinsic vowel mutability hypothesis should be amended to note that vowels are intrinsically more mutability only if lexically contrastive suprasegmental cues are absent.

In fact, a close examination of the stimuli used in van Ooijen's (1996) original study suggests that the observed inherent mutability of vowels may be due to suprasegmental factors. In the stimuli set, the vowel to be changed was always in the stressed syllable of the word (e.g., *poidon* could be changed to *pardon*, by altering the stressed vowel). The consonant to be changed, however, was often in an unstressed syllable (e.g., *poidon* could be changed to *poison*, by modifying a consonant in the unstressed syllable). An alternative explanation for the results could simply be that material in stressed syllables is more accessible and are therefore responded to faster. The notion that stressed syllables are more accessible than unstressed is well-supported by decades of psycholinguistic research (see e.g., Cutler, 2008, 2012; Grosjean & Gee, 1987, for review). This possibility could be tested via a post-hoc item analysis of van Ooijen's data, to determine if changes to the consonant inside the stressed syllable were faster than changes to the consonant outside of the stressed syllable.

Our second research question addressed the role of tone mutability in Mandarin lexical selection. The present study adds to the growing consensus that regardless of its time course, tone information constrains lexical selection less tightly than consonant, vowel (or syllable) information and thus allows for more potential lexical candidates (e.g., Cutler & Chen, 1997; Repp & Lin, 1990; Sereno & Lee, 2014; Taft & Chen, 1992; Ye & Connine, 1999). Listeners made fewer tone errors and changed tones faster than they did in the consonant and vowel conditions. Additionally, in the free choice condition participants overwhelmingly preferred to change the tone than a segment. We evaluate the following four potential explanations as to why tone was changed rather than the vowel.

Firstly, it may have been the case that Mandarin listeners simply did not know what the vowel/final was or that there was less agreement among participants as opposed to the tone or consonant largely due to the Chinese writing system. Given Mandarin's opaque, deep orthography, especially in juxtaposition with the transparent, shallow orthographies of Spanish and Dutch, it may be that native listeners lack the metalinguistic knowledge seemingly readily accessible to native speakers of alphabetic languages. However, all literate Mandarin speakers are fluent in the *pin1yin1* Romanization system which follows a relatively straightforward set of sound to letter correspondences. With the proliferation of keyboard based *pin1yin1* input on cell phones and computers, native speakers are highly proficient in *pin1yin1* (Liu, 2005). Because the word reconstruction task does not use orthographic stimulation, we see no reason to believe orthography can account for the results. Additionally, based on our norming of the materials in which there was 100% agreement of the *pin1yin1* initial and final, we believe listeners performed the task with the correct knowledge of what a vowel or final was.

Secondly, there may have been confusion stemming from the nasal coda or off-glide. The instructions did not specify whether participants needed to preserve the coda or off-glide since traditional Chinese phonology divides the syllable into either the initial or the final (Baxter, 1992; Chao, 1968), with no distinction between the vowel itself and the off-glide or coda. In the case of a nasal coda, both the preservation and deletion were accepted so long as the vowel was changed (e.g., *men3* could be changed to *min3* or *ma3*). Accuracy in the vowel change condition of open syllables was 43% and 41% for closed syllables. Reaction time was similarly congruent at 4126ms for open syllables and 4053ms for closed. We therefore feel certain the results were not due to interference from the nasal codas.

Third, because the tone change condition had the fewest potential changes, the findings may have been a result of the tone change condition being more constrained than the segmental change conditions, both of which had far more possible changes. The tone change condition therefore required inhibiting a large number of inappropriate neighbors, and then the selection of a single neighbor from a relatively small set. This process can be contrasted with the free choice condition, which was the least constrained condition and required the selection of a single neighbor from a large set. Given those different processes, this explanation predicts that the tone change condition should be the most different from the free change condition. This prediction was not borne out as the tone change and free choice conditions patterned almost equivalently, both in terms of accuracy and reaction time.

An alternative interpretation based on similar reasoning appeals not to the *number* of changes, but the *magnitude* of the changes in terms of perceptual similarity. This reasoning follows the assumption that tonal changes involve less phonetic substance being altered than vowel changes; that is, the perceptual-acoustic difference between [ká] and [kà] is necessarily smaller than the difference between [ká] and [kú]. However, findings in both acoustic and auditory phonetics do not support this assumption. The acoustic cues involved in vowels and tones are quite different (see e.g., Fu et al., 1998) such that comparing them is not straightforward; additionally, perceptual similarity is specific to the native language of the listener (Bradlow et al., 2010). More specifically, the limited evidence on Mandarin demonstrates that vowels and tones are confused at roughly equal rates (McLoughlin, 2010), suggesting that their perceptual similarity is roughly equivalent. Perceptual similarity, therefore, cannot account for the present results.

A fourth account as to why tone was changed more often than vowel is that tone carries the lowest information load in Mandarin. Tong, Francis and Gandour (2008) used a speeded classification paradigm (cf. Repp & Lin, 1990) to examine the processing interactions between each pair of segmental and suprasegmental dimensions (i.e., tone–vowel, tone–consonant, vowel–consonant). Asymmetric interference effects were observed between segmental and suprasegmental dimensions with segmental interfering more with tone than vice-versa. According to Tong et al.'s calculations, rimes are significantly more informative than consonants and consonants significantly more informative than tones. In other words, vowels provide the most information value to Mandarin listeners while tones provide the least information. The findings from the present reconstruction study strongly support this information load hypothesis.

This information load hypothesis also efficiently helps answer our third research question - if tone information constrains lexical selection less than segmental information does, do vowels or consonants play a more important role in the process? Our findings strongly suggest that vowels play a more important role in Mandarin lexical access and selection, largely owing to their high information content. Taken together, the findings from this study help clarify our understanding of Mandarin spoken word recognition. Under the theoretical framework of Marslen-Wilson (1984, 1987; Marslen-Wilson & Tyler, 1980), word recognition involves access, selection and integration. As the Mandarin speech signal unfolds, vowel and tone information is processed in tandem (Brown-Schmidt & Canseco-Gonzalez, 2004; Lee, 2007; Malins & Joanisse, 2010, 2012;

Schirmer et al., 2005; Zhao et al., 2011). Since vowels carry the most information, they are largely used as the access code to the syllable or proximate unit (Chen & Chen, 2013; Chen et al., 2002; Mok, 2009; O'Seaghda et al., 2010; Sereno & Lee, 2014; Tong et al., 2008; You et al., 2012). As our lexical decision results showed, the syllable initiates lexical access and search. At the syllable-level, tone helps constrain lexical selection less tightly, but plays a dynamic role given the syllable's frequency and tonal probability (Wiener & Ito, 2014). In the case of tone still not able to sufficiently reduce the number of candidates–spoken Mandarin has a relatively high rate of homophony, with some syllable–tone combinations resulting in as many as 90 different morphemes (Yin, 1984)–sentential-semantic contexts have their effects during the selection stage (Zwitserlood, 1989), further reducing the number of candidates. As Wiener and Ito (2014) have shown, access and selection take place in a largely probabilistic manner, with native speakers tracking and storing syllable–specific tonal probabilities. The present study helps underscore this last point: in the tone condition, the most probable tone was selected over 60% of the time. The nonword *che2*, for example, was routinely changed to the most probable syllable–tone combination *che1* in both the tone and free conditions. This suggests that tone is primarily used to constrain lexical selection when it is unexpected. In most predictable communication, tone carries little informative value. For highly frequent syllables tested in the present study, tone changes were often overshadowed by consonant (or even vowel) changes, supporting the claim that tone is dynamically represented and its role fluctuates during word recognition (Wiener & Ito, 2014). To a Mandarin listener, the nonword *che2* is much more like *che1* than *she2* or *chi2*, but crucially, *che2* is more like *she2* or even *de2* than *chi2* or *cha2*.

Open questions still remain as to whether all speakers who use lexically contrastive suprasegmental cues rely on vowels more than onsets for lexical selection. For example, McQueen, Cutler and Otake (2001) tested speakers of Japanese (a pitch-accent language) and found that preserved Japanese consonant information was more useful to listeners than preserved vowel information. However, this study did not explicitly require participants to manipulate phonemes, and the results are therefore not directly comparable to those of the present study. Other tonal languages such as Cantonese, Thai or Taiwanese (*min3nan2hua4)* would serve as excellent follow-up languages since they have larger tonal inventories and allow for different tone and vowel information value accounts. It may also be worth exploring whether L1 speakers of a tonal language learn to treat vowel information as less reliable and more mutable in their non-tonal L2.

In sum, our findings have shown a stark difference in the way Mandarin listeners perform the word reconstruction task as compared to English, Dutch and Spanish listeners. Unlike native listeners of European languages, native Mandarin listeners found it easier to alter the tone and then the onset rather than the vowel of a nonword. These results allow us to reject the universal intrinsic vowel mutability hypothesis. For speakers of a lexical tone language, it appears that vowel information is more reliable than it is to speakers of non-tonal languages.

Acknowledgements

This article reflects ideas and analysis that have benefited from critique, advice and support from the following individuals: Shari Speer, Marjorie Chan, Kiwako Ito, and Cynthia Clopper. Shuyan Sheng assisted in the recruiting of participants. The authors also wish to express thanks to Michael Vitevitch and an anonymous reviewer for comments that considerably improved an earlier version of the manuscript. An earlier version of this work was presented at the 88th meeting of the Linguistic Society of America in Minneapolis, Minnesota.

Funding

This research received no specific grant from any funding agency in the public, commercial, or not-for-profit sectors.

Notes

[1] Marks, Moates, Bond and Stockmal (2002) explored whether phonological features were responsible for Van Ooijen and Cutler et al.'s results. The authors tested the contribution of the root node feature [sonorant] in both English and Spanish; some types of changes that involved crossing a feature boundary were more difficult than others. Marks et al. concluded that rather than searching for individual segments, participants seemed to search for holistic patterns.

[2] In the following discussion, and the rest of the paper, we use the term "syllable" to refer to purely segmental content, and "syllable–tone combinations" to refer to syllables realized with a particular lexical tone.

References

Baayen, R. H., Davidson, D. J., & Bates, D. M. (2008). Mixed-effects modeling with crossed random effects for participants and items. *Journal of Memory and Language, 59*, 413–425.

Balota, D.A., Yap, M.J., Cortese, M.J., Hutchison, K.A., Kessler, B., Loftis, B., … Treiman, R. (2007). The English Lexicon Project. *Behavior Research Methods, 39*, 445–459.

Bates, D., Maechler, M., & Bolker, B. (2014). lme4: Linear mixed-effects models using Eigen and S4 [R package version 1.1–7]. Retrieved from https://github.com/lme4/lme4/

Baxter, W. H. (1992). *A Handbook of old Chinese phonology*. Berlin: Mouton de Gruyter.

Bradlow, A., Clopper, C. G., Smiljanic, R., & Walter, M. A. (2010). A perceptual phonetic similarity space for languages: Evidence from five native language listener groups. *Speech Communication, 52*, 930–934.

Brown-Schmidt, S., & Canseco-Gonzalez, E. (2004). Who do you love, your mother or your horse? An event-related brain potential analysis of tone processing in Mandarin Chinese. *Journal of Psycholinguistic Research, 33*(2), 103–135, doi: 10.1023/B:JOPR.0000017223.98667.10.

Cai, Q., & Brysbaert, M. (2010). SUBTLEX-CH: Chinese word and character frequencies based on FILM SUBTITLES. *Plos ONE, 5*(6), e10729, doi: 10.1371/journal.pone.0010729.

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

Chen, T.-M., & Chen, J.-Y. (2013). The syllable as the proximate unit in Mandarin Chinese word production: An intrinsic or accidental property of the production system? *Psychonomic Bulletin & Review, 20*, 154–162.

Chen, J.-Y., Chen, T.-M., & 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.

Clopper, C. G. (2013). Modeling multi-level factors using linear mixed effects. *Proceedings of Meetings on Acoustics*, 19, 060028.

Connine, C. M., Blasko, D. G., & Titone, D. (1993). Do the beginnings of spoken words have a special status in auditory word recognition? *Journal of Memory and Language, 32*, 193–210.

Connine, C. M., Blasko, D. G., & Wang, J. (1994). Vertical similarity in spoken word recognition: Multiple lexical activation, individual differences, and the role of sentence context. *Perception & Psychophysics, 56*, 624–636.

Cooper, N., Cutler, A., & Wales, R. (2002). Constraints of lexical stress on lexical access in English: Evidence from native and non-native listeners. *Language and Speech, 45*, 207–228, doi: 10.1177/00238309020450030101.

Creel, S. C., Tanenhaus, M. K., & Aslin, R. N. (2006). Consequences of lexical stress on learning an artificial lexicon. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 32*, 15–32, doi: 10.1037/0278–7393.32.1.15.

Cutler, A. (1986). Forbear is a homophone: Lexical prosody does not constrain lexical access. *Language and Speech, 29* (3), 201–220, doi: 10.1177/002383090104400 20301.

Cutler, A. (2008). Lexical stress. In D. Pisoni & R. Remez (Eds.), *The handbook of speech perception* (pp. 264–289). Malden, MA: Blackwell Publishing.

Cutler, A. (2012). *Native listening.* Cambridge, MA: MIT Press.

Cutler, A., & Chen, H. (1997). Lexical tone in Cantonese spoken-word processing. *Perception & Psychophysics, 59*, 165–179, doi: 10.3758/BF03211886.

Cutler, A., & Otake, T. (1999). Pitch accent in spoken-word recognition in Japanese. *Journal of the Acoustic Society of America, 105*(3), 1877–1888.

Cutler, A., Sebastián-Gallés, N., Soler-Vilageliu, O., & van Ooijen B. (2000). Constraints of vowels and consonants on lexical selection: Cross-linguistic comparisons. *Memory & Cognition, 28*(5), 746–755.

Cutler, A., & van Donselaar, W. (2001). Voornaam is not (really) a homophone: Lexical prosody and lexical access in Dutch. *Language and Speech, 44*(2), 171–195, doi: 10.1177/00238309010440020301.

Duanmu, S. (2007). *The phonology of standard Chinese* (2nd edn). New York: Oxford University Press.

Duanmu, S. (2008). *Syllable structure: The limits of variation*. New York: Oxford University Press.

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

Fu, Q. J., Zeng, F. G., Shannon, R. V., & Soli, S. D. (1998). Importance of tonal envelope cues in Chinese speech recognition. *Journal of the Acoustical Society of America, 104*, 505–510.

Gow, D.W, & Gordon, P.C. (1995). Lexical and prelexical influences on word segmentation: Evidence from priming. *Journal of Experimental Psychology: Human Perception & Performance, 21*, 344–359.

Grosjean, F., & Gee, J. (1987). Prosodic structure and spoken word recognition. *Cognition, 25*, 135–155, doi:10.1016/0010–0277(87)90007–2.

Ho, A. T. (1976). The acoustic variation of Mandarin tones. *Phonetica, 33*, 353–367.

Kirby, J. P., & Yu, A. C. L. (2007). Lexical and phonotactic effects on wordlikeness judgments in Cantonese. In the *Proceedings of the international congress of the phonetic sciences XVI*, 1161–1164.

Lee, C.-Y. (2007). Does horse activate mother? Processing lexical tone in form priming. *Language and Speech, 50*(1), 101–123, doi: 10.1177/00238309070500010501.

Li, C. N., & Thompson, S. A. (1981). *Mandarin Chinese: a functional reference grammar*. Berkeley: University of California Press.

Liu, Y. (2005). A pedagogy for digraphia: An analysis of the impact of pinyin on literacy teaching in China and its Implications for curricular and pedagogical innovations in a wider community. *Language and Education, 19*(5), 400–414, doi: 10.1080/09500780508668693.

Liu, S., & Samuel, A. G. (2007). The role of Mandarin lexical tones in lexical access under different contextual conditions. *Language and Cognitive Processes, 22*, 566–594.

Malins, J. G., & Joanisse, M. F. (2010). The roles of tonal and segmental information in Mandarin spoken word recognition: An eyetracking study. *Journal of Memory and Language, 64*(4), 407–420, doi: 10.1016/j.jml.2010.02.004.

Malins, J. G., & Joanisse, M. F. (2012). Setting the tone: An ERP investigation of the influences of phonological similarity on spoken word recognition in Mandarin Chinese. *Neuropsychologia, 50*, 2032–2043, doi: 10.1016/j.neuropsychologia.2012.05.002.

Marks, E. A., Moates, D. R., Bond, Z. S., & Stockmal, V. (2002). Word reconstruction and consonant features in English and Spanish. *Linguistics, 40*(2), 421–438.

Marslen-Wilson, W. D. (1984). Function and process in spoken word-recognition. In H. Bouma & D. G. Bouwhuis (Eds.), *Attention and performance X: Control of language processes* (pp. 125–150). Hillsdale, NJ: Erlbaum.

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

Marslen-Wilson, W. D. (1990). Activation, competition, and frequency in lexical access. In G. T. M. Altmann (Ed.), *Cognitive models of speech processing: Psycholinguistic and computational perspectives* (pp. 148-172). Cambridge, MA: MIT Press.

Marslen-Wilson, W. D., & Tyler, L. K. (1980). The temporal structure of spoken language understanding. *Cognition, 8*, 1–71.

Mathôt, S., Schreij, D., & Theeuwes, J. (2012). OpenSesame: An open-source, graphical experiment builder for the social sciences. *Behavior Research Methods, 44*(2), 314–324. doi:10.3758/s13428–011–0168–7

Mattys, S. L., White, L., & Melhorn, J. F. (2005). Integration of multiple speech recognition cues: A hierarchical framework. *Journal of Experimental Psychology: General, 134*, 477–500.

McLoughlin, I. (2010). Vowel intelligibility in Chinese. *IEEE Transactions on Audio, Speech, and Language Processing, 18*, 117–125.

McQueen, J. M., Norris, D. G., & Cutler, A. (1994). Competition in spoken word recognition: Spotting words in other words. *Journal of Experimental Psychology: Learning, Memory & Cognition, 20*, 621–638.

McQueen, J. M., Norris, D. G., & Cutler, A. (1999). Lexical influence in phonetic decision-making: Evidence from subcategorical mismatches. *Journal of Experimental Psychology: Human Perception & Performance, 25*, 1363–1389.

McQueen, J. M., Otake, T., & Cutler, A. (2001). Rhythmic cues and possible-word constraints in Japanese speech segmentation. *Journal of Memory and Language, 45*, 103–132.

Mok, P. (2009). On the syllable-timing of Cantonese and Beijing Mandarin. *Chinese Journal of Phonetics, 2*, 148–154.

Myers, J. (2002). An analogical approach to the Mandarin syllabary. *Journal of Chinese Phonology, 11*, 163–190.

Myers, J. (2008). MiniCorpJS (Version 0.5) [Computer software]. Retrieved from http://www.ccunix.ccu.edu.tw/~lngproc/MiniCorpJS.htm

Myers, J., & Tsay, J. (2005). The processing of phonological acceptability judgments. *Proceedings of symposium on 90–92 NSC projects* (pp. 26–45). Taipei, Taiwan, May.

Norman, J. (1988). *Chinese*. Cambridge: Cambridge University Press.

O'Seaghdha, P. G., Chen, J.-Y., & Chen, T.-M. (2010). Proximate units in word production: Phonological encoding begins with syllables in Mandarin Chinese but segments in English. *Cognition, 115*, 282–302.

Perfetti, C. A., & Zhang, S. (1991). Phonological processes in reading Chinese characters. *Journal of Experimental Psychology: Learning, Memory and Cognition, 17*(4), 633–643.

Repp, B. H., & Lin, H.-B. (1990). Integration of segmental and tonal information in speech perception: a cross-linguistic study. *Journal of Phonetics, 18*, 481–495.

Schirmer, A., Tang, S. L., Penney, T. B., Gunter, T. C., & Chen, H. C. (2005). Brain responses to segmentally and tonally induced semantic violations in Cantonese. *Journal of Cognitive Neuroscience, 17*, 1–12, doi: 10.1162/0898929052880057.

Sereno, J. A., & Lee, H. (2014). The contribution of segmental and tonal information in Mandarin spoken word processing. *Language and Speech*, 1–21.

Shillcock, R. C. (1990). Lexical hypotheses in continuous speech. In G. T. M. Altmann (Ed.), *Cognitive models of speech processing: Psycholinguistic and computational perspectives* (pp. 24–49). Cambridge, MA: MIT Press.

Soto-Faraco, S., Sebastián-Gallés, N., & Cutler, A. (2001). Segmental and suprasegmental mismatch in lexical access. *Journal of Memory and Language, 45*, 412–432, doi: 10.1006/jmla.2000.2783.

Sulpizio, S., & McQueen, J.M. (2012). Italians use abstract knowledge about lexical stress during spoken-word recognition. *Journal of Memory and Language, 66*, 177–193.

Tabossi, P., Burani, C., & Scott, D. (1995). Word identification in fluent speech. *Journal of Memory & Language, 34*, 440–467.

Taft, M., & Chen, H.-C. (1992). Judging homophony in Chinese: The influence of tones. In H. –C. Chen, & O. J. Tzeng (Eds.), *Language processing in Chinese* (pp. 151-172). Amsterdam: North-Holland Elsevier Science Publishers.

Tagliapietra, L., & Tabossi, P. (2005). Lexical stress effects in Italian spoken word recognition. In *Proceeding in XXVII conference of the cognitive science society* (pp. 2140–2144).

Tong, Y., Francis, A. L., & Gandour, J. T. (2008). Processing dependencies between segmental and suprasegmental features in Mandarin Chinese. *Language and Cognitive Processes, 23*, 698–708.

van Ooijen, B. (1996). Vowel mutability and lexical selection in English: Evidence from a word reconstruction task. *Memory & Cognition, 24*, 573–583.

Vitevitch, M. S. (2003). The influence of sublexical and lexical representations on the processing of spoken words in English. *Clinical Linguistics & Phonetics, 17*, 487–499.

Vitevitch, M.S., & Luce, P.A. (1998). When words compete: levels of processing in perception of spoken words. *Psychological Science, 9*, 325–329.

Vitevitch, M. S., & Luce, P. A. (1999). Probabilistic phonotactics and neighborhood activation in spoken word recognition. *Journal of Memory and Language, 40*, 374–408.

Vitevitch, M. S., Luce, P. A., Charles-Luce, J., & Kemmerer, D. (1997). Phonotactics and syllable stress: Implications for the processing of spoken nonsense words. *Language and Speech, 40*, 47–62.

Vitevitch, M. S., Luce, P. A., Pisoni, D. B., & Auer, E. T. (1999). Phonotactics, neighborhood activation, and lexical access for spoken words. *Brain and Language, 68*, 306-311.

Wang, H. S. (1998). An experimental study on the phonetic constraints of Mandarin Chinese. In B. K. Tsou (Ed.), *Studia Linguistica Serica* (pp. 259–268). Hong Kong: City University of Hong Kong Language Information Sciences Research Center.

Wiener, S., & Ito, K. (2014). Do syllable-specific tonal probabilities guide lexical access? Evidence from Mandarin, Shanghai and Cantonese speakers. *Language, Cognition & Neuroscience*, doi: 10.1080/23273798.2014.946934.

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.

Xu, Y. (1999). Effects of tone and focus on the formation and alignment of F0 contours. *Journal of Phonetics, 27*, 55–105.

Ye, Y., & Connine, C. M. (1999). Processing spoken Chinese: The role of tone information. *Language and Cognitive Processes, 14*(5–6), 609-630, doi: 10.1080/016909699386202.

Yin, B. Y. (1984). Hanyu yusu de dingliang yanjiu [A quantitative research of Chinese morphemes]. *Zhongguo Yuwen* [Chinese Literature and Language], *182*, 338–347.

You, W., Zhang, Q., & Verdonschot, R. G. (2012). Masked syllable priming effects in words and picture naming in Chinese. *PLoS ONE 7*(10): e46595, doi:10.1371/journal.pone.0046595.

Zhao, J., Guo, J., Zhou, F., & Shu, H. (2011). Time course of Chinese monosyllabic spoken word recognition: Evidence from ERP analyes. *Neuropsychologia, 49*, 1761–1770, doi: 10.1016/j.neuropsycholo gia.2011.02.054.

Zhou, X., & Marslen-Wilson, W. D. (1999a). The nature of sublexical processing in reading Chinese characters. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 25*(4), 819–837.

Zhou, X., & Marslen-Wilson, W. D. (1999b). Phonology, orthography, and semantic activation in reading Chinese. *Journal of Memory and Language, 41*, 579–606.

Zwitserlood, P. (1989). The locus of the effects of sentential-semantic context in spoken-word processing. *Cognition, 32*, 25–64.

Appendix 1.

Table 7. Experiment 1 stimuli.

Syllable	Tone	Condition
mei	1	TAG
ming	1	TAG
niao	1	TAG
nong	1	TAG
ren	1	TAG
dai	2	TAG
juan	2	TAG
ri	2	TAG
zhen	2	TAG
zuan	2	TAG
bin	3	TAG
he	3	TAG
hun	3	TAG
le	3	TAG
xia	3	TAG
cong	4	TAG
gei	4	TAG
mang	4	TAG
neng	4	TAG
tian	4	TAG
awk	1	VIOL
dua	1	VIOL
kie	1	VIOL
niz	1	VIOL
sti	1	VIOL
bak	2	VIOL
fiz	2	VIOL
lap	2	VIOL
noth	2	VIOL
nua	2	VIOL
chuf	3	VIOL
gian	3	VIOL
puai	3	VIOL
tiang	3	VIOL
yom	3	VIOL
dri	4	VIOL
kib	4	VIOL
rit	4	VIOL
thra	4	VIOL
tra	4	VIOL
bao	1	WORD
beng	1	WORD
chu	1	WORD
cu	1	WORD
da	1	WORD

Table 7. (Continued))

Syllable	Tone	Condition
fu	1	WORD
gong	1	WORD
hei	1	WORD
nian	1	WORD
piao	1	WORD
yu	1	WORD
ai	2	WORD
cha	2	WORD
de	2	WORD
fa	2	WORD
ke	2	WORD
lei	2	WORD
men	2	WORD
nu	2	WORD
sui	2	WORD
tou	2	WORD
ze	2	WORD
bang	3	WORD
bing	3	WORD
di	3	WORD
ji	3	WORD
kao	3	WORD
mai	3	WORD
rao	3	WORD
shang	3	WORD
wei	3	WORD
wu	3	WORD
xiang	3	WORD
ang	4	WORD
cai	4	WORD
dao	4	WORD
duo	4	WORD
han	4	WORD
jiao	4	WORD
pa	4	WORD
ri	4	WORD
shen	4	WORD
xiao	4	WORD
yong	4	WORD

Appendix 2

Table 8. Experiment 2 stimuli.

Group	Syllable	Tone
A	ca	2
A	cu	3
A	diu	4
A	dou	2
A	gan	2
A	gou	2
A	gui	2
A	huai	I
A	kai	2
A	le	3
A	liang	I
A	mang	4
A	mu	I
A	nen	3
A	ning	I
A	nong	I
A	ri	3
A	sen	4
A	teng	3
A	xia	3
A	yue	3
A	zai	2
A	zhuo	4
A	zui	I
B	bei	2
B	bin	3
B	ce	3
B	chui	4
B	chun	4
B	chuo	3
B	he	3
B	kou	2
B	kua	2
B	kuai	I
B	lan	I
B	lue	I
B	men	3
B	ne	I
B	nin	4
B	se	3
B	si	2
B	sun	4
B	ta	2
B	tian	4

Table 8. (Continued)

Group	Syllable	Tone
B	tie	2
B	wai	2
B	xiu	2
B	xun	3
C	cui	2
C	cuo	3
C	dai	2
C	fou	4
C	gai	2
C	gei	4
C	hun	3
C	jiong	4
C	ken	1
C	kong	2
C	kuan	4
C	mei	1
C	pen	3
C	que	3
C	re	2
C	rong	4
C	ruo	3
C	shua	2
C	su	3
C	tai	3
C	te	1
C	wo	2
C	zhun	4
C	zi	2
D	bang	2
D	bing	2
D	cang	4
D	che	2
D	cou	1
D	de	4
D	diao	2
D	dui	3
D	fo	4
D	hen	1
D	hua	3
D	ka	4
D	kao	2
D	ku	2
D	mai	1
D	mie	3

(Continued)

Table 8. (Continued)

Group	Syllable	Tone
D	nai	1
D	neng	4
D	nei	1
D	niao	1
D	nu	1
D	pian	3
D	shang	2
D	xiong	3