The roles of tonal and segmental information in Mandarin spoken word recognition: An eyetracking study

Jeffrey G. Malins, Marc F. Joanisse *

Department of Psychology & Program in Neuroscience, The University of Western Ontario, London, Ontario, N6A 5C2 Canada

ARTICLE INFO

ABSTRACT

We used eyetracking to examine how tonal versus segmental information influence spoken word recognition in Mandarin Chinese. Participants heard an auditory word and were required to identify its corresponding picture from an array that included the target item (*chuang2* 'bed'), a phonological competitor (segmental: *chuang1* 'window'; cohort: *chuan2* 'ship'; rhyme: *huang2* 'yellow'; tonal: *niu2* 'cow'), and two phonologically unrelated distractors. Growth curve analysis was used to characterize the trajectory of looks to target and competitor items during word processing. We found similar model fits for the segmental and cohort conditions characterized by slower eye movements to correct targets compared to baseline, suggesting that tonal and segmental information are accessed concurrently and play comparable roles in constraining activation. These findings are discussed with respect to current models of spoken word recognition that have not previously accounted for the role of tone.

Introduction

Listeners comprehend spoken utterances quickly and effortlessly, which belies the complex cognitive processing that contributes to their understanding. A central component of this task is simply recognizing the words that have been said. This involves taking incoming sound information and mapping it onto phoneme-level (sublexical) and word-level (lexical) representations in the brain so that the meanings of the individual words in the utterance can be accessed. This conceptualization is the inspiration for a number of current models of spoken word recognition (such as TRACE: McClelland & Elman, 1986 and Shortlist/ Merge: Norris, 1994; Norris & McQueen, 2008; Norris, McQueen, & Cutler, 2000).

Such models of speech recognition fall short in representing the totality of this process for all speakers however. The reason for this is that they have been developed using Indo-European languages such as English and Dutch, which do not make extensive use of suprasegmental phonological features, such as stress and tone, in the word recognition process. However, many of the world's languages do make extensive use of these features. In Mandarin Chinese, a word can have different meanings depending on tonal contrasts signaled by modulations in pitch during articulation; when pronounced in a high and level pitch (Tone 1), the word *ma* means 'mother', but when produced with a pitch that rises from mid-range to high (Tone 2), *ma* means 'hemp' (hereafter, pitch distinctions like these are indicated using numerical notation; e.g., *ma1*, *ma2*). Importantly, these contrasts are thought to be realized only through variation of suprasegmental features, while segmental (phonemic) features are left intact. These suprasegmental features are fundamentally different from segmental features, as they can span multiple individual segments in speech, and are not intrinsic properties of phonetic segments themselves (Cutler & Chen, 1997).

Despite widespread use across the world's languages, there have been relatively few studies of how suprasegmental features such as tone influence on-line auditory word recognition, both from the behavioral and modeling

* Corresponding author. Fax: +1 519 661 3961.
 E-mail address: marcj@uwo.ca (M.F. Joanisse).

standpoints. A greater understanding of this is thus an important step in expanding current theories – and thereby informing current models – of speech recognition to account for these types of languages.

Studies of lexical stress

There is some evidence for how suprasegmental features influence auditory word recognition from studies of lexical stress in Indo-European languages such as English and Dutch. In these languages, stressed syllables are distinguished from unstressed ones using suprasegmental features such as duration and amplitude, and patterns of stress can differentiate otherwise similar words. Cutler (1986) found that in a cross-modal priming paradigm, English readers showed similar semantic priming effects for correctly and incorrectly stressed auditory words (e.g., RElay vs. reLAY both showed equivalent priming of the target *race*). Therefore, suprasegmental features did not appear to be used to constrain activation prior to lexical access in English, since otherwise the stress difference should have been used to distinguish the words before the meanings of both were accessed, and thus only activated words related to the appropriately stressed word.

Cutler (1986) suggests that one of the reasons that suprasegmental features (in this case stress) might not constrain word recognition prior to lexical access in English is that they are almost always irrelevant; only a few pairs of words in English are distinguished solely on the basis of stress. Indeed, similar experiments using Dutch (in which a greater number of word pairs differ solely on the basis of stress) have shown that stress cues are exploited during spoken word recognition, even though mismatches in stress may not constrain activation in the same way as mismatches in segmental information (Cutler & van Donselaar, 2001; van Donselaar, Koster, & Cutler, 2005). The authors propose that Dutch is an intermediate case, in that more pairs of words differ solely on the basis of stress than in English, yet the prevalence of these distinctions is still much less in Dutch than in a language like Spanish. In Spanish, there are many words that differ from each other solely in stress, and indeed stress distinctions often serve grammatically important roles such as indicating verb tense. Indeed, when the influence of stress was examined in Spanish using cross-modal priming, it was found that suprasegmental cues were not only exploited when recognizing spoken words, but that these cues were as effective as segmental cues in constraining activation (Soto-Faraco, Sebastián-Gallés, & Cutler, 2001). Taken together, these results seem to suggest that listeners do exploit suprasegmental cues in word recognition, but the role of these cues in constraining activation is proportional to the extent to which these cues are used to distinguish words in their language.

Tone processing

One of the reasons it is relatively uncommon for languages to differentiate words using only stress is that stress contrasts are frequently accompanied by changes in segmental structure as well (e.g., the vowel change in CONtent vs. conTENT). Thus, lexical distinctions realized by variation in suprasegmental features alone are less pervasive in stress languages than they are in tonal languages. For this reason, studying tone processing might offer a unique perspective into the relative weighting of segmental and suprasegmental cues in constraining word recognition. In addition, studies of tone processing can also address another component of the word recognition process: timing of access to these cues. Unlike patterns of stress, which are realized over the duration of polysyllabic words, pitch variations in tonal languages are realized within monosyllables themselves, and thus listeners must use suprasegmental cues to establish word identity at a more local level. This is not to say that listeners of non-tonal languages are not capable of this, and indeed there is some evidence that speakers of non-tonal languages can identify the stress or pitch accent of syllables in isolation (Cutler & Otake, 1999; Cutler, Wales, Cooper, & Janssen, 2007); rather, speech perception in tonal languages *requires* the listener to distinguish words on the basis of suprasegmental cues. As it is thought that tonal information is usually realized on the vowel portion of a syllable and not on onset consonants (Cutler & Chen, 1997), studies often compare access to tonal information to access to vowel information in syllables.

Early studies of lexical tone established a consistent pattern of results. Taft and Chen (1992) had participants make homophone judgments for written characters in Mandarin, and found that response times were slower when the pronunciations of the two characters differed in tone than when they differed in vowel, suggesting later access to tone. This result was also replicated in Cantonese (Taft & Chen, 1992), a Chinese language in which there are six tones as opposed to Mandarin's four. Repp and Lin (1990) also found evidence for later access to tone, as listeners categorized Mandarin nonwords on the basis of tone less rapidly than they performed consonant or vowel (segmental) categorizations for these stimuli. Ye and Connine (1999) corroborated this finding by showing that in the absence of linguistic context, Mandarin listeners were able to identify vowel mismatches more quickly than tonal mismatches in a syllable monitoring task. Cutler and Chen (1997) offered further support for this, showing that in a lexical decision task Cantonese listeners made more errors on nonwords that differed tonally from real words than they did on those that differed segmentally from real words. Furthermore, in a same-different judgment task, listeners were slower and less accurate when the stimuli differed in tonal information than when they differed in segmental information. Because tone was more likely to be misprocessed and result in errors, the authors proposed that in addition to being accessed later than segmental information, perhaps tone might be a weaker cue for word recognition as well.

While the above studies appeared to reach a consensus on the relative timing of access to tonal versus segmental information, and thus extrapolated from this to suggest a weaker role for tonal information in constraining word recognition, Liu and Samuel (2007) point out that these tasks are mostly sublexical in nature, meaning that they could be performed without having to access word knowledge. In

addition, the tasks lacked contextual constraints. In a task that demanded greater access to word knowledge in varying levels of contextual constraint (monitoring for nonwords in spoken Mandarin words, sentences, and idioms), they found that segmental information was more influential than tonal information in constraining word recognition only in a sentential context, as indicated by greater lexical decision accuracy. In the single word and idiom contexts, there was an equal weighting of the two types of information. In a different task emphasizing sublexical information (identifying vowels of words masked by white noise, produced either in isolation or embedded in idioms and sentences), they actually found a shift in the relative weighting of the two types of information: segmental information dominated when there was little context available (words in isolation), whereas tonal information dominated in conditions of greater contextual constraint (sentences and idioms). The authors argued that their findings support a much stronger role for tonal information in lexical access than had been previously believed.[1]

The proposal that tone plays a relatively strong role in spoken word recognition – perhaps even equal to that of segmental information – has also garnered support from some more recent studies. In an auditory lexical decision task in Mandarin, Lee (2007) found that primes that were segmentally identical to targets yet different in tone did not give rise to faster lexical decision times than phonologically unrelated control primes. This suggested that tonal information strongly constrains word activation in Mandarin by serving to inhibit segmentally identical words that do not share tone. Likewise, Schirmer, Tang, Penney, Gunter, and Chen (2005) used event related potentials (ERPs) to examine processing of spoken Cantonese sentences, finding that tonal and segmental information are accessed at a similar point in time and play comparable roles during word recognition in Cantonese. Changing the rime phonemes of an anticipated word in a sentence modulated the N400 component in the same manner as changing its tone.

While the results of Schirmer et al. (2005) are in line with Liu and Samuel (2007) and Lee (2007), the conclusions are still controversial, and indeed some recent studies still suggest a weaker role for tone (Tong, Francis, & Gandour, 2008). Also, the Schirmer et al. (2005) study offered unique insights into the timing of access to tonal and segmental information that other studies were unable to offer, as the conclusions from this study were based on an on-line language processing measure, and so it stands alone in some of its claims.

Motivation for the present study

The present study examined some of the outstanding issues concerning the role of tone in Mandarin auditory

word recognition using the "visual world paradigm", which allows the use of eyetracking to measure on-line auditory comprehension. In this task, participants are presented with an array of pictures on a computer screen and subsequently hear an auditory stimulus corresponding to one of these items. The participant is typically asked to overtly identify the word that was perceived, for instance via a button press or a mouse click. The experimental manipulation in this type of study comes in the form of a competitor picture that is also present on the screen, the name of which has a phonological relationship with the name of the target picture. For example, Tanenhaus, Spivey-Knowlton, Eberhard, and Sedivy (1995) presented participants with displays of objects, and found that participants were slower to make eye movements to correct targets (e.g., *candy*) when a competitor picture was also present in the display (e.g., *candle*, which shares an onset with the target word). In this example, hearing the spoken instruction "Pick up the candy" initially cues both *candy* and *candle* due to shared onsets, and it is not until later on in the word that information is received that disambiguates the target and competitor items. Allopenna, Magnuson, and Tanenhaus (1998) used a similar methodology and presented participants with arrays that contained a target picture (e.g., *beaker*), a competitor picture (e.g., *beetle*, which shares an onset with the target word), and pictures of distractor items that were phonologically unrelated to the target or competitor. When hearing the instruction "Pick up the beaker", participants initially produced eye movements to both the pictures of the beaker and beetle with an equal probability, but later in the trial produced a greater proportion of looks to the target item, as unique identifying information was perceived. Interestingly, Allopenna et al. (1998) also demonstrated that these effects were also seen for words that rhymed with the target (e.g., *speaker* for the target *beaker*), suggesting that competition between words in this paradigm is not restricted solely to words sharing onset. Implicit in this type of work is the "linking hypothesis": the probability of fixating a visual object is a direct reflection of a word's neural activation (Tanenhaus, Magnuson, Dahan, & Chambers, 2000). By this logic, the extent of competition between targets and competitors can be assessed by quantifying the probability of fixation to each of these items.

We used the visual world paradigm to test Mandarin speakers' sensitivity to segmental versus tonal cues in recognizing familiar auditory words. Of interest are competition effects stemming from words sharing segmental structure and differing only in tone (segmental competitors; e.g., *chuang1* 'window' for the target *chuang2* 'bed' – note that competitor labels are defined by what is shared between targets and competitors), words sharing word-initial phonemes and tone (cohort competitors; e.g., *qian2* 'money' for the target *qiu2* 'ball'), words sharing word-final phonemes and tone (rhyme competitors; e.g., *gua1* 'melon' for the target *hua1* 'flower'), and words sharing only tone and thus having different segmental structure (tonal competitors; e.g., *mi3* 'rice' for the target *shu3* 'mouse'). Note that we have defined a target's cohort competitor as sharing not only onset consonant but also the initial vowel of the rime (as per Marslen-Wilson & Tyler, 1980) as well as

---

[1] This view has been presented as a more recent trend; however, one exception to this is Experiment 2 in Ye and Connine (1999), in which participants performed tone and vowel monitoring tasks in both neutral and highly constraining (idiomatic) contexts. It was found that tone showed an advantage in an idiomatic context, and this finding is in fact offered by Liu and Samuel (2007) as evidence supporting their claim that tones play a stronger role in more constraining linguistic contexts.

tone, such that cohort competitors diverge from targets not at the beginning of the rime of a syllable, but at some point within it. This was done to closely mimic the point of divergence of segmental competitors from targets, which should also happen within the rime of a syllable, as the pitch contour profiles of tones are thought to be realized over rimes of syllables.

Participants were presented with a monosyllabic Mandarin word spoken in isolation and were asked to match it to one of four pictures in an array. We monitored looks to items in these arrays during the unfolding of auditory stimuli. We hypothesized that we would observe a similar pattern of results to those of Schirmer et al. (2005): tonal and segmental information should be accessed at a similar point in time, and both types of information should play a comparable role in constraining word recognition. Therefore we predicted that because segmental competitors diverge from targets (in tone) at a point in time similar to when cohort competitors diverge from targets (in phonemes), listeners should be able to detect these differences at a similar point in time, reflected in similar trajectories of change in the proportion of looks to targets and competitors for these two trial types.

In addition, following the findings of Allopenna et al. (1998) and Desroches, Joanisse, and Robertson (2006) for rhyme competitors in English, we might expect that Mandarin rhyme competitors would be looked at more than unrelated items later on in word processing due to phonemic and tonal overlap with target items. Lastly, following the findings of Lee (2007), as well as converging evidence from implicit priming tasks (Chen, Chen, & Dell, 2002), we expect that competitors which share only tones with target items should not show interference effects, and so we should not observe slower looks to target or increased looks to competitors in this condition.

Methods

Participants

Twenty-four native speakers of Mandarin were recruited to participate in this study; seven individuals performed the task using peripheral vision, resulting in few eye movements, and were excluded from analyses for this purpose. The remaining seventeen participants (13 female) had a mean age of 28. All were from Mainland China and had been living in North America for a mean of 3.9 years. Participants were screened for native speaking ability with the help of a native Mandarin speaker who made an eligibility judgment prior to participation in the experiment.

Stimuli

The experimental stimuli consisted of 27 monosyllabic Mandarin words, which were all easily imageable nouns (refer to Appendix A for a complete list of the stimuli, as well as morphemic frequency for each item). Seven of these were critical stimuli around which sets of competitors were constructed. Four competitor items were selected per critical stimulus: a segmental competitor, which shared all pho-

nemes but differed in tone; a cohort competitor, which shared initial CV and tone but differed in word-final phonemes; a rhyme competitor, which shared rime and tone but differed in onset consonant; a tonal competitor, which shared only tone and differed in all phonemes. For example, for the critical stimulus *chuang2* 'bed', the competitors were: *chuang1* 'window' (segmental), *chuan2* 'ship' (cohort), *huang2* 'yellow' (rhyme), *niu2* 'cow' (tonal). When possible, competitor items were selected from the set of competitors for the other critical stimuli, resulting in some items participating in more than one competitor condition. Frequency was balanced across conditions such that targets did not significantly differ in frequency from competitors [$t(27) = -0.262$, $p = .80$], nor did the competitor conditions differ in frequency from each other [$F(3, 18) = 0.149$, $p = .93$].

Five tokens of each word were digitally recorded as produced by an adult male native speaker of Mandarin (16 bits; 44,100 Hz). Pictures matching these items were selected with the assistance of this speaker to ensure they depicted words in a way that was culturally appropriate. To ensure that listeners were familiar with the words of interest, participants also performed a naming task prior to the eyetracking study. In this task they were presented with each picture and asked to say aloud the Mandarin word that they thought was most appropriate for it. In cases where the name given was different from the one used in the experiment, participants were given the intended name.

Procedure

Participants performed an auditory word-visual picture matching task in which eye movements were recorded at 60 Hz using an SMI RED-II remote eyetracker (SensoriMotoric Instruments, Inc.; Cambridge, USA). On each trial they were presented with an array of four pictures on an LCD monitor located 50–60 cm away. Pictures were oriented directly above, below, to the left, or to the right of a fixation point consisting of a red circle (see Fig. 1 for a sample display). On the experimental trials, each array consisted of a target item, a phonological competitor (in one of four conditions: segmental, cohort, rhyme, or tonal), and two phonologically unrelated distractor items drawn from a different competitor set and thus sharing neither phonemes nor tone with the target or competitor. These distractor items were balanced for frequency across competitor conditions [$F(3, 52) = 0.622$, $p = .60$]. Baseline trials consisted of target items presented instead with three phonologically unrelated distractor items and thus no competitors. The positions of the target and competitor in the display were counterbalanced across trials, as was the relationship between the two pictures in the array (adjacent or opposite), to offset the effect of any influence of target/competitor orientation on eye movement data.

Each trial proceeded as follows: first, the picture array was presented on screen for 1500 ms; next, the central fixation cue appeared; then, following a 500 ms delay, an auditory word was presented over headphones. Participants were instructed to look at the fixation cue when it appeared, and to push a button on a keypad that corre-
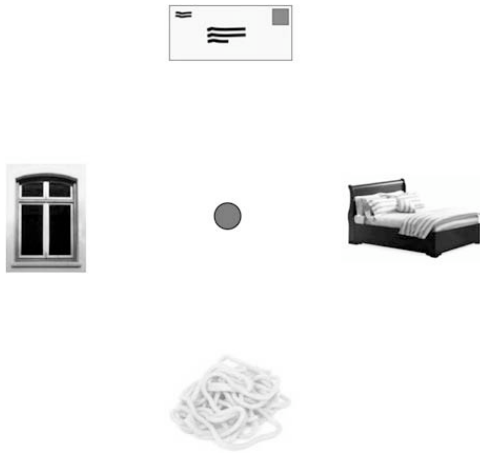
Fig. 1. A sample display containing pictures of a target item (*chuang2* 'bed'), a phonological competitor item (*chuang1* 'window') and two phonologically unrelated distractor items (*xin4* 'envelope' and *mian4* 'noodles') oriented around a central fixation point.

sponded with the position of the picture matching the word that they heard. Note that participants were not explicitly instructed to look at the picture of the target item. However, we initially observed that some participants (*N* = 7, not included in our analyses) misunderstood the task and maintained fixation on the center of the screen for the entire trial, using only peripheral vision to complete the task. To correct this, an instruction was added to inform participants that they were permitted to look freely to any of the objects after fixating the central point.

There were five blocks of 66 trials, with block order counterbalanced across subjects, for a total of 330 trials. Of these, 280 were experimental trials (70 in each of the four competitor conditions), 35 were baseline control trials (target item presented with only phonologically unrelated distractors), and 15 were filler trials to appropriately balance the experiment. The 70 experimental trials consisted of 35 trials with the appropriate target and competitor relationship outlined in Appendix A, and 35 reciprocal trials in which the roles of target and competitor were inverted (so the competitor in Appendix A was in fact the auditory stimulus). This was done so that when a picture of a critical stimulus was on the screen, there was a .50 probability of hearing the name of that picture, thus ensuring that participants were not biased to look to the target picture before hearing the auditory stimulus. Prior to beginning the task, participants also completed a practice block of ten trials that did not use any of the 27 items from the actual experiment.

After completing the eyetracking task, a language history and experience questionnaire was administered to participants. This questionnaire further verified native speaking ability by requesting information regarding daily language use and exposure as well as self-ratings of proficiency. All testing was conducted by a native Mandarin speaker, and all instructions were given in Mandarin. Efforts were made to keep the experimental setting as monolingual as possible to reduce interference from English.

Analysis of eye movement data

Eye gaze position was coded relative to visual stimuli by dividing the display space into discrete regions corresponding to the four items and the central fixation point. Only looks within these regions were included in analyses. We next calculated the proportion of looks to targets and competitors at each time point, as a proportion of total looks to all objects and the central fixation. Analyses were restricted to the trials in which the target was the appropriate critical stimulus as outlined in Appendix A. As a result, the target items were identical for all conditions, such that any differences between experimental conditions in gaze to target were only the result of the competitor pictures that appeared along with the target on screen, and not the result of differences in the availability of acoustic information between conditions.

Eye movement data were analyzed from 200 to 1100 ms post stimulus onset. The lower limit takes into account the 200 ms delay for planning and executing an eye movement, while the upper limit was chosen to reflect the point at which all conditions had reached maximum looks to target. Data were subjected to growth curve analysis using the R software package (Mirman, Dixon, & Magnuson, 2008; see http://magnuson.psy.uconn.edu/GCA/ for further details). The goal of this analysis was to describe the functional form of the probability distribution of gaze location over time by quantifying the major aspects that result from underlying processes. Thus, the analysis identified model fit components for a curve that captures this probability distribution. Using this approach, experimental conditions could be compared to each other by assessing these model fit components, each of which indexes a different aspect of gaze trajectories over time. The intercept term indexes the average height of the curve, while the linear component (slope) indicates the overall angle of the curve. The higher-order components primarily concern the shape of the curve about inflection points, specific points at which the graph changes from convex to concave, or vice versa. The quadratic component tends to index the symmetric rise and fall of the curve about inflection points, while cubic and quartic components primarily reflect the steepness of the curve about inflection points.

To conduct the analysis of looks to target, a base model was first constructed using the baseline control condition. This model contained only the effects of time and subjects. Condition was then added onto this model as a factor with four levels corresponding to the four competitor conditions. Subsequently, the model was built up by adding interaction terms for the condition variable with time in a stepwise fashion to determine the effect of condition on the linear, quadratic, cubic, and quartic components of the model respectively. Model fits were tested at each step by assessing change in deviance, which tests whether including a parameter increases the fit of the model. Once the effect of condition on the quartic component of the model had been assessed, parameter estimates were generated for the additional contribution made by each experimental condition on the model components beyond what the base model was able to capture. These parameter estimates were tested for significance using *t*-tests.

Looks to competitor were analyzed in a similar fashion, except a separate analysis was performed for each condition using looks to unrelated distractors within the same condition to construct a base model. This analysis was restricted to the point in time at which looks to competitor appeared greatest (400–800 ms post stimulus onset), and only a second order function was used to generate model fits, reflecting prior observations that looks to competitor follow an inverse U-shaped function (e.g., Allopenna et al., 1998). Generation of parameter estimates was unnecessary, as change in deviance directly indexed differences between looks to competitors and looks to unrelated distractors in each condition.

Results

Behavioral data

Mean reaction time and percent accuracy for button press responses are given in Table 1. Trials with reaction times 2.5 standard deviations above condition-wise subject means, or shorter than 100 ms, were rejected. A one-way repeated measures ANOVA showed no main effect of condition on reaction time or accuracy [$F(4, 64) = 0.99$, $p = .42$; $F(4, 64) = 0.62$, $p = .65$].

Eye movement data

Looks to target

Model fits for each experimental condition are plotted against baseline in Fig. 2, along with the observed proportion of looks to target at each time point. Estimates for the parameters that characterize the model fits for each condition are presented in Table 2. As illustrated in Fig. 2a, the model fit for the segmental condition (*chuang2–chuang1*, where *chuang2* represents a sample target and *chuang1* a sample competitor) differed from baseline in the latency of peak looks to target; the baseline condition reached its maximum at approximately 925 ms, while the segmental condition reached its maximum at approximately 1025 ms. The inflection points for the curves occurred near where the two diverged, at around 600 ms. Following the curves upwards from these points, we see that the baseline curve rose much more steeply to its maximum than did the segmental curve. As shown in Table 2, there were significantly different quadratic, cubic, and quartic components for the segmental condition from baseline, which is indicative of a difference in how steeply the curves rose from the central inflection point.[2] These results suggest that there was greater competition between targets and competitors in the segmental condition than in the baseline condition, as after the curves diverged at 600 ms, the proportion of looks to target

Table 1
Mean reaction time and mean percent accuracy for the button press response.

Condition	Reaction time (ms)	Percent accuracy
Segmental	482.1 (25.8)	99.7 (0.002)
Cohort	486.1 (28.1)	99.5 (0.004)
Rhyme	489.8 (33.5)	99.3 (0.005)
Tonal	464.1 (26.2)	99.5 (0.003)
Baseline	469.8 (29.4)	100.0 (0)

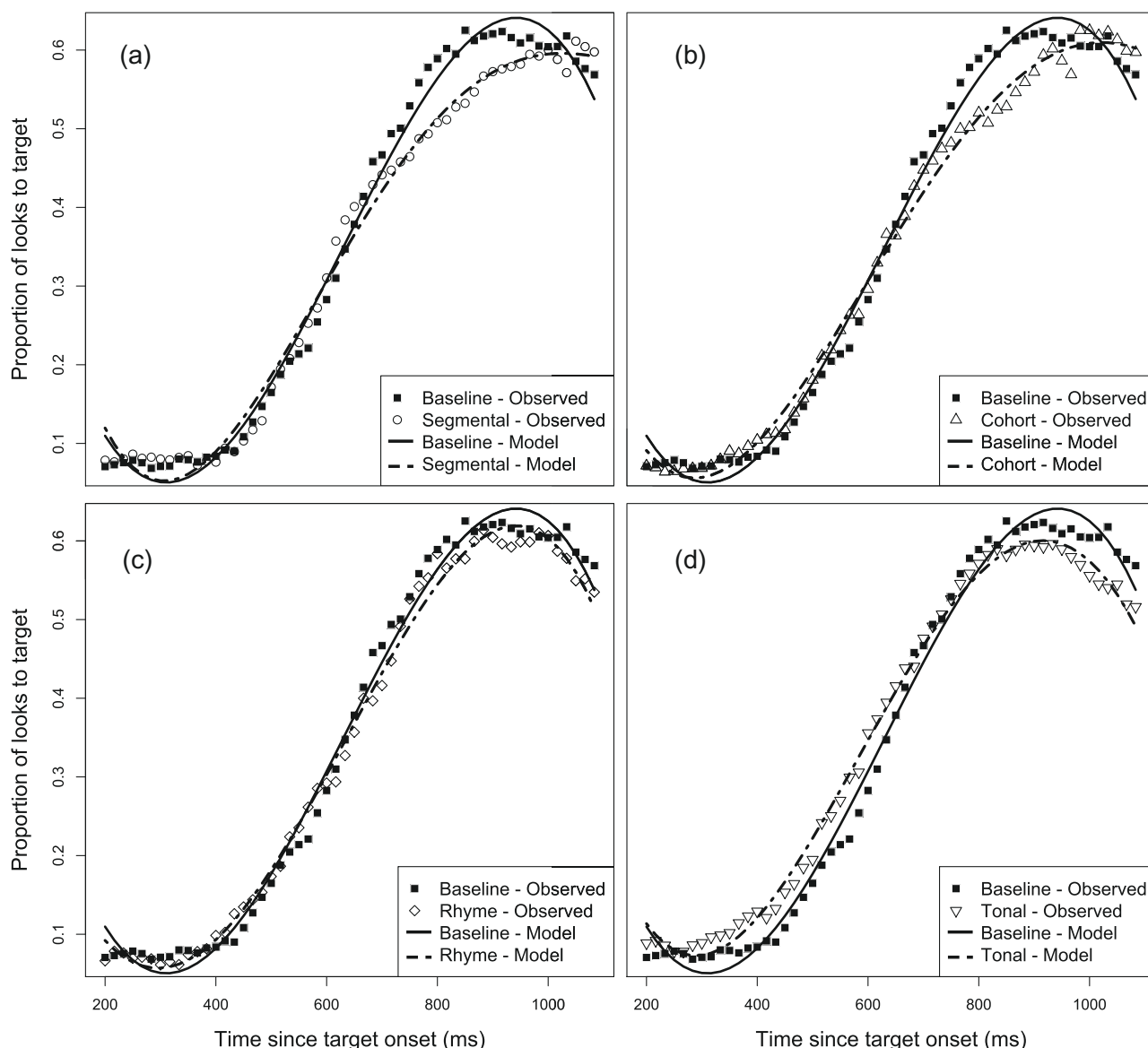*Note.* Values in parentheses represent standard errors.

in the baseline condition rose more steeply and reached its maximum earlier than it did in the segmental condition.

We interpret this delay in looks to target pictures as the result of competition among partially activated target and competitor representations. One conceptualization of this draws on the TRACE model of word recognition (McClelland & Elman, 1986) in which word-level representations are mutually inhibitory, such that bottom-up phonological information can lead to the simultaneous activation of two concepts in a way that inhibits the selection of the correct form in memory. The result of this mutual inhibition is a delay in activation of target words, reflected here as a delay in making eye movements to target pictures (Tanenhaus et al., 2000).

Fig. 2b shows the data from the cohort condition (*chuang2–chuan2*) plotted against baseline. The curves differed from one another in much the same way as the curves in the segmental condition (a); that is, they diverged at around 600 ms, and from this point the baseline curve rose to its maximum much more steeply. This was confirmed by significant quadratic, cubic and quartic components (Table 2) between the two model fits. This suggests that there was also a delay in looks to target in the cohort condition due to inhibitory effects of competitors.

For the rhyme condition (*chuang2–huang2*), there was no significant difference in model fit from baseline (Table 2), confirming the observation in Fig. 2c that the presence of rhyme competitors did not delay activation of target words. Lastly, analyses of the tonal condition (*chuang2–niu2*; Fig. 2d) showed that the tonal condition curve had a shallower slope than baseline until the curves intersected at around 800 ms, reflected by a significantly different linear component between the two conditions (Table 2), suggesting that there were greater looks to target in the tonal condition in the initial portion of the trial compared to baseline. However, this pattern reversed towards the end of the trial, as looks to target rose less steeply to their maximum in the tonal condition compared to baseline. This is indexed by significantly different quadratic and cubic components in the tonal condition from baseline, which suggest that there were inhibitory effects of tonal competitors towards the end of the trial.

The above analysis indicates that both the segmental and cohort conditions showed significant slowing in looks to target, suggesting interference effects for both types of phonological relationships. Of interest was whether these two types of interference differed in terms of time course. This was addressed using a follow-up analysis that directly compared the two conditions to each other by

---

[2] Note that our analyses focus on findings of significant linear as well as higher-order terms, reflecting the assumption that differences in lexical dynamics can be reflected in both the simple rate with which fixations increase over time as well as differences in the shape of these curve profiles. That said, we also note some caution in interpreting higher-order terms given that they may also reflect other sources of variation such as early-going differences in time course; see Mirman et al. (2008), for further discussion.

Fig. 2. Observed data (symbols) and model fits (lines) for the data from each experimental condition ((a) segmental; (b) cohort; (c) rhyme; (d) tonal) plotted separately against baseline. Note that each observed data point corresponds to mean proportion of looks to target at each particular point in time, averaged across participants and items.

applying the same growth curve analysis procedure as above to a subset of the data containing only these two conditions. The model fits for the two conditions are presented along with observed data in Fig. 3. As the analysis was restricted to only two conditions, change in deviance at each step directly indexed the difference between the two conditions in model fits (Table 3), and so presentation of separate parameter estimates for each condition is unnecessary. As we see in Table 3, the analysis failed to reveal a significant difference between the model fits for the two conditions in any of the components of interest, with the probability values associated with change in deviance relatively high in each case.

Looks to competitors

Observed proportions of looks to competitors and unrelated distractors at each time point are plotted in Fig. 4,

along with model fits for the 400–800 ms time window. The results of significance tests for the model fit components characterizing the curves are presented in Table 4. Fig. 4a shows a greater rate of looks to competitors than to unrelated distractors in the segmental condition, confirmed by significant linear and quadratic components in Table 4. Similar effects are observed for looks to the cohort competitors in Fig. 4b, although only the quadratic component was significantly different in this case. These effects of segmental and cohort competitors complement the patterns seen for the target data, as does the lack of competitor effects in the rhyme condition observed in Fig. 4c and confirmed in Table 4. Lastly, Fig. 4d suggests that there were greater looks to competitors than unrelated distractors in the tonal condition in the 400–800 ms time window, indexed by significant linear and quadratic components (Table 4).

Table 2
Growth curve analysis of looks to target.

| Model | Model fit | | | Parameter estimates | | | | | |
| | | | | Segmental | | | Cohort | | |
	−2LL[a]	ΔD[b]	p<	Est.	t	p<	Est.	t	p<
Base	10582.1	–	–	–	–	–	–	–	–
Intercept	10585.9	3.8	n.s.[c]	−0.0141	−0.97	n.s.	−0.0099	−0.68	n.s.
Linear	10591.1	5.2	n.s.	−0.1036	−1.20	n.s.	−0.0746	−0.86	n.s.
Quadratic	10674.7	83.6	0.01	0.0597	2.41	.05	0.0560	2.27	.05
Cubic	10748.4	73.7	0.01	0.1434	5.80	.01	0.1828	7.40	.01
Quartic	10771.3	22.9	0.01	0.0855	3.46	.01	0.0492	1.99	.05
				Rhyme			Tonal		
Base	10582.1	–	–	–	–	–	–	–	–
Intercept	10585.9	3.8	n.s.	−0.0084	−0.58	n.s.	0.0066	0.45	n.s.
Linear	10591.1	5.2	n.s.	−0.0750	−0.87	n.s.	−0.1953	−2.26	.05
Quadratic	10674.7	83.6	0.01	−0.0187	−0.76	n.s.	−0.1379	−5.58	.01
Cubic	10748.4	73.7	0.01	0.0450	1.82	n.s.	0.0625	2.53	.05
Quartic	10771.3	22.9	0.01	−0.0195	−0.79	n.s.	0.0352	1.43	n.s.

Note. The method of growth curve analysis used to generate the values in this table did not remove lower-order terms once they were found to be non-significant. It is thus possible that these terms are unduly accounting for portions of the total variance. For this reason, a second analysis was conducted in which non-significant terms were removed before adding higher-order components. While the parameter estimates were marginally different as a result, the significance level associated with each parameter estimate did not change from what is presented here.
  [a] The deviance statistic, equal to minus two times the log-likelihood.
  [b] Change in deviance.
  [c] DF = 4 in the chi-square distribution used to assess model fit; DF = 64 for the parameter estimates of intercepts (*t* distribution), and 4421 for the parameter estimates of all other components (*t* distribution).
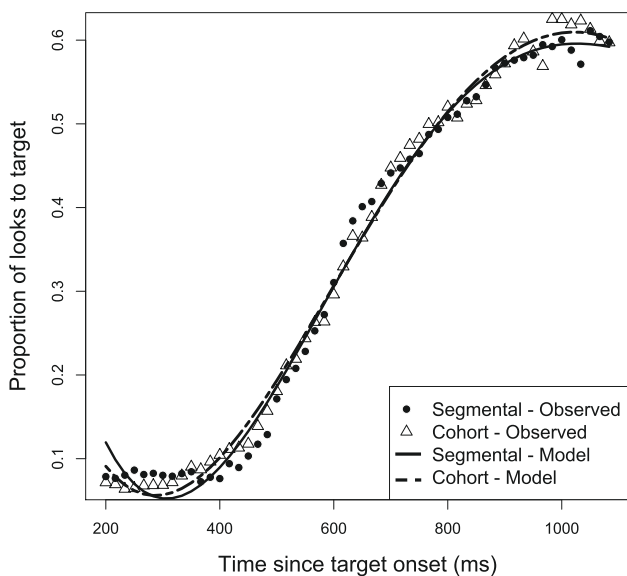
Fig. 3. Observed data (symbols) and model fits (lines) for the segmental and cohort conditions. Note that the each observed data point corresponds to the mean proportion of looks to target at each particular point in time, averaged across participants and items.

Table 3
Direct comparison of model fits for looks to target in the segmental and cohort conditions.

Model	−2LL	ΔD	p
Base	4751.1	–	
Intercept	4751.1	0.04	.83
Linear	4751.3	0.2	.68
Quadratic	4751.3	0.03	.87
Cubic	4754.6	3.3	.07
Quartic	4757.4	2.8	.10

Discussion

The present study used eyetracking to examine the time course of access to tonal and segmental information in Mandarin spoken word recognition as well as the role of both types of information in constraining word recognition. In the experimental task, conditions differed with respect to whether competitors overlapped targets in either tonal or segmental information (or both), as well as the point at which competitors diverged from targets. The trajectory of change in the proportion of looks to targets and competitors for each competitor condition was characterized by growth curve analysis to assess the time course of resolution of targets from competitors in each condition. These trajectories can now in turn be compared to the phonological relationship between targets and competitors in each condition to see how listeners used disambiguating information (either tonal or segmental in nature) to establish target identity.

Of particular interest were the cohort and segmental conditions, as these were designed such that competitors diverged from targets at similar points in time, yet the nature of the information signaling this divergence was different. In the cohort condition, competitors diverged from targets in segmental information (word-final phonemes); in the segmental condition, competitors diverged from targets in tonal information (yet shared all phonemes). However, the point of divergence between targets and competitors for both conditions occurred in similar locations – the rimes of syllables (a detailed acoustic analysis of stimuli in these two conditions is presented in Appendix A, and confirms that this was true for the stimulus sets used in this experiment). A similar time course of resolution of targets from competitors between the two conditions should thus
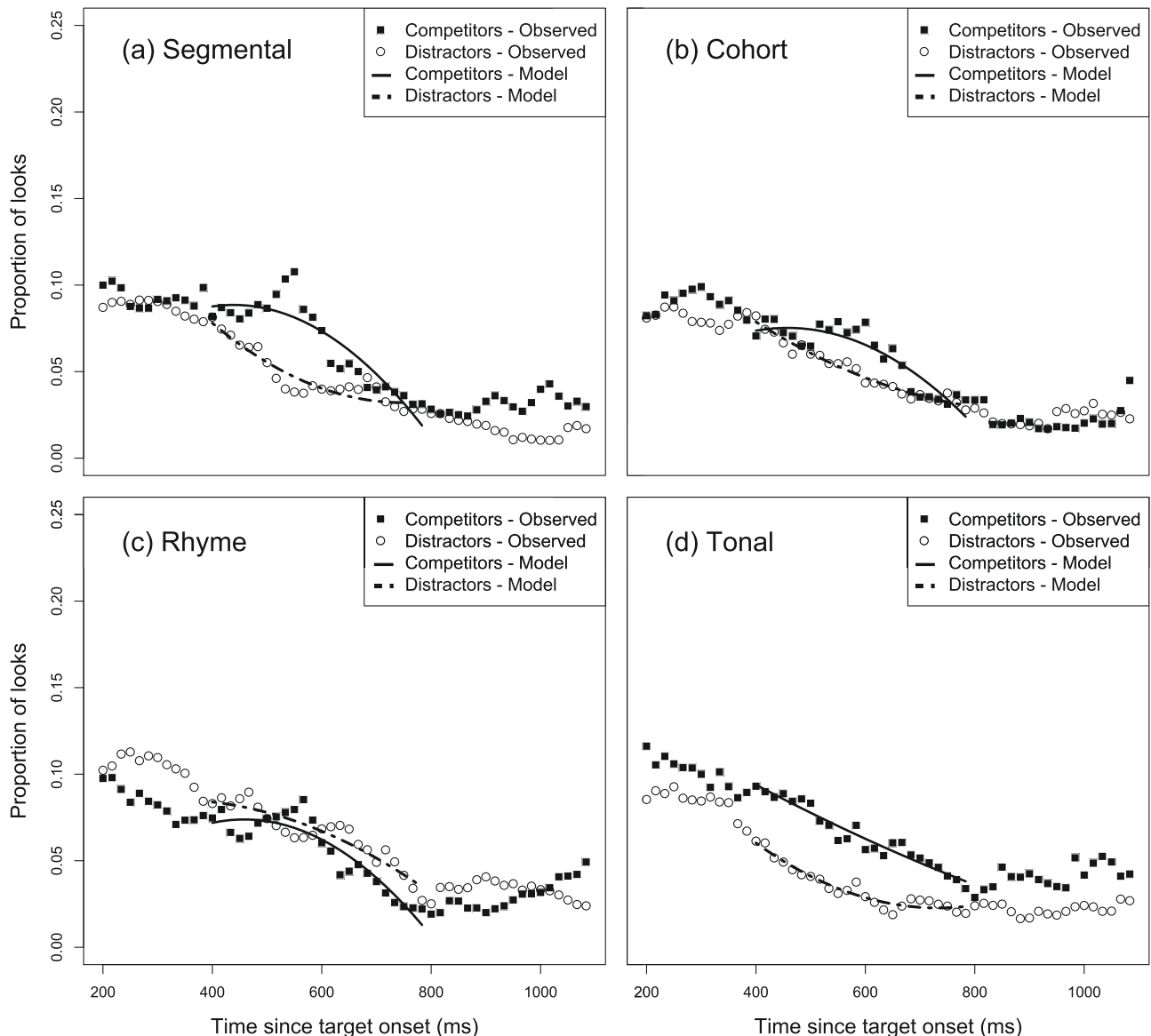
Fig. 4. Observed data (symbols) and model fits (lines) for looks to competitor in each experimental condition ((a) segmental; (b) cohort; (c) rhyme; (d) tonal) plotted against looks to unrelated distractors within the same condition. Note that each observed data point corresponds to mean proportion of looks to competitors or distractors at each particular point in time, averaged across participants and items.

indicate that segmental and tonal information are accessed concurrently and play a comparable role in constraining recognition. Results indicate that this is the case. As predicted, both conditions showed slower looks to target compared to baseline (Fig. 2a and b). Further, when the two conditions were compared directly, the trajectory of change in looks to target was almost identical for the two conditions (Fig. 3) yielding no differences in any of the model fit components (Table 3). This pattern of results was also complemented by an observation of competitor effects in both conditions (Fig. 4a and b).

While the results from the cohort and segmental conditions matched our predictions, the results from the rhyme condition did not. This condition showed no effect on the time course of looks to target stimuli, suggesting that the presence of rhyme competitors did not alter the time course of recognizing the auditory target stimulus any more than the presence of phonologically unrelated items.

This is corroborated by our analysis of looks to competitors, which were statistically indistinguishable from looks to phonologically unrelated distractors. These results for Mandarin differ from eyetracking studies of English that have found evidence for rhyme competition in adults (Allopenna et al., 1998) and typically-developing children (Desroches et al., 2006). This difference might reflect a key difference in how the two languages weight rhyming information during word recognition – perhaps the initial phonemes of words play a more important role in constraining word recognition in Mandarin than they do in English. Alternatively, this could be due to a difference in stimuli between the experiments. In the present study, only monosyllables were used, and in most cases the overlap between targets and rhyme competitors was restricted to only one or two phonemes. In contrast, Allopenna et al. (1998) and Desroches et al. (2006) used words that overlapped with targets in as many as five phonemes, and so

Table 4
Growth curve analysis of looks to competitors.

Model	−2LL	ΔD	p<
Segmental			
Base	3332.6	–	–
Intercept	3342.3	9.8	.01[a]
Linear	3345.2	2.9	n.s.
Quadratic	3378.2	33.0	.01
Cohort			
Base	3267.9	–	–
Intercept	3269.5	1.5	n.s.
Linear	3269.5	0.005	n.s.
Quadratic	3284.9	15.4	.01
Rhyme			
Base	3220.5	–	–
Intercept	3222.2	1.7	n.s.
Linear	3222.4	0.2	n.s.
Quadratic	3225.9	3.5	n.s.
Tonal			
Base	3647.9	–	–
Intercept	3657.9	10.0	.01
Linear	3659.1	1.2	n.s.
Quadratic	3663.5	4.4	.05

Note. The method of growth curve analysis used to generate the values in this table did not remove lower-order terms once they were found to be non-significant. As with looks to target, a second analysis was conducted in which non-significant terms were removed before adding higher-order components. Patterns of significance were not different for any component from what is presented here.

[a] DF = 1 in the chi-square distribution for all model fit components.

this more extensive phonemic overlap might explain the greater amount of rhyme competition that they observed. In addition, a number of the words they used were disyllabic. It thus remains an open question whether comparable effects could be observed with disyllabic Mandarin words, in which targets and competitors overlap entirely in the second syllable and only differ in the onset of the first (e.g., *sandal–candle* in the Allopenna et al. study). This might yield a broader inference on the extent of rhyme competition given that competitors and targets would overlap in terms of phonemes, but also would share an entire syllable with targets (this is especially relevant given the proposed special status of the syllable in Mandarin; Zhou & Marslen-Wilson, 1994).

The results from the tonal condition also differed from what we expected. Based on previous findings from priming studies (Chen et al., 2002; Lee, 2007), we did not expect that the presence of a tonal competitor (i.e., words overlapping only with respect to tone; e.g., *chuang2–niu2*) to have caused any difference in looks to target compared to baseline. Notably however, we saw a rather inconsistent pattern: early on in word recognition, the proportion of looks to target was actually greater in the tonal condition compared to baseline, suggestive of faster rather than slower processing in the presence of a competitor, whereas this pattern reversed at later time points. Further, we saw evidence of competitor effects in the 400–800 ms window, suggestive of interference effects at this time stage. The looks to competitor may reflect subtle interference effects due to overlapping tone in stimuli that nevertheless had no segmental overlap. However, there is an alternative expla-

nation for the pattern observed in looks to target, having to do with the nature of the distractor items on these trials. Distractor items were selected from among competitor items in the other stimulus sets in order to have a closed set of stimuli across all trials. For this reason, even though the most phonologically dissimilar stimuli to targets were selected as distractors, in some cases one phoneme did overlap between target and distractor items. This is in contrast to the tonal competitors, which were selected to ensure that they shared no phonemes with targets in any position. Thus the mean phonemic overlap between baseline targets and distractor items was 14.7% (calculated by dividing the number of shared phonemes in the same position in a pair of words by the total number of phonemes in the longer word of the pair; see Pastizzo & Feldman, 2002), whereas the mean phonemic overlap between targets and all competitors/distractors in the tonal condition was somewhat smaller at 11.5%. This lower amount of phonemic overlap might explain why gaze to target was initially higher in the tonal condition compared to baseline, as the other items on the screen impeded word recognition to a lesser extent. A future study that more closely controls for phonemic overlap might reveal in more detail the nature of these subtle interference effects due to shared tonal contours.

Overall, our results corroborate recent findings concerning the role of tonal information in spoken word recognition. The eye movement data were completely in line with the findings of Schirmer et al. (2005), whose critical stimuli mismatched expected words in a very similar way as cohort and segmental competitors differed from targets in the present study, and who also found evidence for concurrent access to tonal and segmental information. This lends the support of a second on-line language measure (eyetracking) to their conclusion of similarity in timing of access, and also offers complementary data from Mandarin language processing that supports their results from Cantonese. Our behavioral results also lend some support for their findings, and offer a reason why conclusions based on ERP and eyetracking evidence contradict those of earlier behavioral studies. The eyetracking and ERP data reveal dynamic processing of auditory words occurring prior to the end-state, yielding differences between conditions that were not evident from end-state processing (as measured by the button press). This suggests that the end-state alone may often give a different pattern of results than the processing leading up to it (Spivey, 2007).

We also saw support for the hypothesis set forth by Liu and Samuel (2007) that an experimental task that demands lexical access in an environment of contextual constraint promotes the use of tonal information. In the present study, the button press demanded lexical access in order to identify the picture that matched the word that was heard. In addition, the presence of the four pictures on the screen provided contextual constraint, as listeners were well familiar with the names of the pictures after completing the naming task at the beginning of the experiment, and so the presence of the pictures on the screen should have generated some expectations of incoming auditory stimuli. Together, these conditions explain why

tone was likely assigned a high priority by listeners due to its relatively informative nature in terms of constraining this limited set of competitors, and also might explain why the results of the present study differ from those of other tasks that have suggested a weaker role for tone (Tong et al., 2008).

Implications for models of spoken word recognition

In addition to strengthening our understanding of tonal processing, the present findings offer some insight into which models of spoken word recognition are best able to accommodate the data and how these models might be updated to account for tonal languages. One of the most influential models of speech recognition is the TRACE model (McClelland & Elman, 1986). This model incorporates featural, phonemic, and word-level processing units, which interact via excitatory and inhibitory connections. A spoken input of a word consists of a pattern of phonetic features for the particular phonemes that comprise it, leading to activation of feature detectors representing the different dimensions of acoustic information. These feature units then spread activation to phoneme units, with activation of a particular phoneme unit a result of activation of its constituent features. Activation of a word, in turn, corresponds to spreading activation from its constituent phonemes. Thus after being introduced to the model, a spoken input is left as a pattern of activation, or trace, across these three levels of processing. One of the key features of this model is the interaction across these three levels of processing, which is bidirectional in nature, and continues in a dynamic fashion throughout processing of speech. Models such as TRACE (as well as Shortlist/Merge: Norris, 1994; Norris & McQueen, 2008; Norris et al., 2000) are termed continuous mapping models, for as speech unfolds over time, input is continuously mapped onto units in the model in a dynamic fashion.

The present data fit well with this view, given evidence of significant competition effects as listeners chose among potential visual targets. In the experimental trials, the target word provided bottom-up phonological information that evolved over time, and activated a corresponding set of candidate word forms. Competition occurred when multiple candidate word forms received partially compatible phonological inputs; lateral inhibition among word-level representations led to delayed activation of the correct target form; with respect to eye movements this was reflected in a shallower upward slope in the rate of fixations to target pictures over time. The increased activation of competitors was also reflected in subtly higher rates of looks to competitor pictures at earlier points in recognition, again reflecting the activation of both a target word and a phonologically related competitor, especially early on within a trial.

As discussed above, we did not observe rhyme effects in the present study, which might appear to be consistent with the Cohort model of speech recognition (Marslen-Wilson, 1987). In Cohort, acoustic–phonetic information in an utterance is extracted from the speech stream, and a number of potential candidates consistent with this information are activated early in the process of word recognition. This pool of candidates is narrowed down as the speech stream unfolds and unique information is received that disambiguates a word from its competitors. Importantly, the set of lexical competitors is restricted to words that share the same onset, since speech perception occurs in a linear fashion over time.

We interpret this result with caution however; as discussed above, the present experiment may have failed to detect rhyme effects due to the lesser amount of phonemic overlap between targets and rhyme competitors compared to previous studies that found these effects in English. Similarly, it is also difficult to address models such as the Neighborhood Activation Model (Luce & Pisoni, 1998), in which neighbors are defined solely on the basis of segmental criteria. In this case, it is unclear exactly what a neighbor would be for a Mandarin word, given the extra feature of tone; for example, would two segmentally identical words that differ only in tone be considered neighbors on this account? The present data would suggest this is the case, and that a definition of neighbors that includes suprasegmental criteria might be more appropriate in this type of model. Likewise, if we take at face value the finding of increased looks to tonal competitors (i.e., forms that differed completely in terms of segments, and overlapped only with respect to tone), this again supports a non-linear model of word recognition that considers similarity beyond simple cohort effects.

Regardless of which model is most appropriate, several modifications would be necessary to accommodate the Mandarin data: most notably, suprasegmental feature detectors would need to be incorporated in such a way that they are activated over a similar time course as phonemic feature detectors. This would also need to be instantiated not only at the lower acoustic feature level of processing, but also in higher-up phoneme-level representations. Ye and Connine (1999) have gone so far as to posit that "toneme" nodes be added to continuous mapping models such as TRACE to reflect the notion that tones merit a similar status to phonemes in the models. This is complemented by recent evidence suggesting that tones and phonemes might be psychologically equivalent in terms of representation. Certainly, tones do act like phonemes in Mandarin in that they are used in a lexically contrastive sense; there are minimal pairs that differ in tones (i.e., the segmental competitors used in the present study) just as there are minimal pairs that differ in phonemes. In addition, there is some evidence that contour tones are perceived categorically (in Mandarin: Xu, Gandour, & Francis, 2006; in Cantonese: Francis, Ciocca, & Ng, 2003) in the same way that consonants are in English and other languages. Furthermore, to help children learn these categories, parents often exaggerate differences in pitch contour between tones when directing Mandarin speech towards infants (Liu, Tsao, & Kuhl, 2007), thus helping to form discrete representations of contour profiles.

Incorporating tone into higher levels of speech recognition models could be done either by having separate units for each of the four tones in Mandarin, which could then spread activation to words expressed in a respective tone, or it could be done by having a layer of units for individual phonemes articulated in each of the four tones. For example, perception of the phoneme /a/ produced with a rising

tone (Tone 2) could concurrently activate a phoneme unit corresponding with the /a/ phoneme, as well as a tone unit corresponding with the rising tone, or it could activate a distinct /a2/ toneme unit. The latter seems unlikely, however, because pitch contour is expressed over the entire rime of syllables, so pitch varies in a dynamic fashion throughout articulation (except in the case of the high level Tone 1). Thus, the /a/ in *han2* is quite different from the one in *hua2*. In the former, it is at the lower part of the rising contour, while in the latter, it is at the higher part. One way to redress this is to have units for each possible rime articulated in each possible tone, but as of yet, there is little evidence suggesting that rimes in Mandarin merit this special status in the word recognition process.

Regardless of how tone is instantiated in the models, it should be done in such a way as to embody similar roles for tonal and segmental information in the word recognition process. For instance, Liu and Samuel (2007) suggest there should be additional types of interaction in models to account for the role of tone, such as inhibitory connections to morphemes that are inconsistent with tonal cues. A change such as this would reflect on-line use of tonal information to restrict lexical access to a more constrained subset of words, thereby making the word recognition process more efficient.

Conclusions

Results of the present study suggest both concurrent access to tonal and segmental information during Mandarin spoken word recognition, as well as a comparable role for both types of information in constraining the activation of phonologically similar words. This implies that Mandarin listeners might integrate tonal and segmental information in parallel, a finding that has important implications for theories and models of spoken word recognition. As well, the current use of eyetracking offers further evidence that on-line language measures can give us insights into language processing that are not always apparent in end-state results.

Acknowledgments

Appendix A

Acoustic analysis of segmental and cohort competitors

In order to test if segmental and cohort competitors diverged from targets acoustically at similar or different points in time, we calculated fundamental frequency (F0) and first and second formant frequency (F1 and F2) for segmental and cohort competitor stimuli, respectively. F0, F1
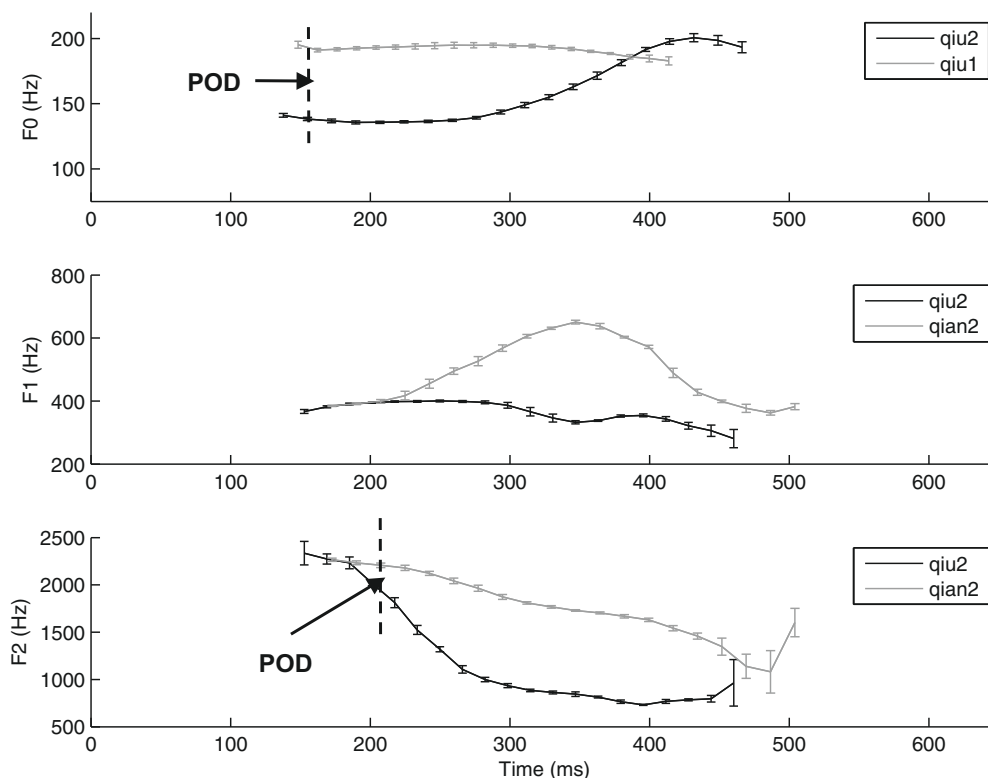
Fig. A1. Mean F0, F1 and F2 frequencies for the target *qiu2* compared to its segmental and cohort competitors (stimulus set 4 in Table A1). Point of divergence (POD) for the target vs. segmental competitor was determined using F0 measurements (top plot), while POD for the target vs. cohort competitor was determined using F1 and F2 measurements (whichever differed first, in this case F2; bottom plot). All curves have been generated using mean frequency values in each time window across the five tokens of each auditory word. Error bars represent standard error of the mean.
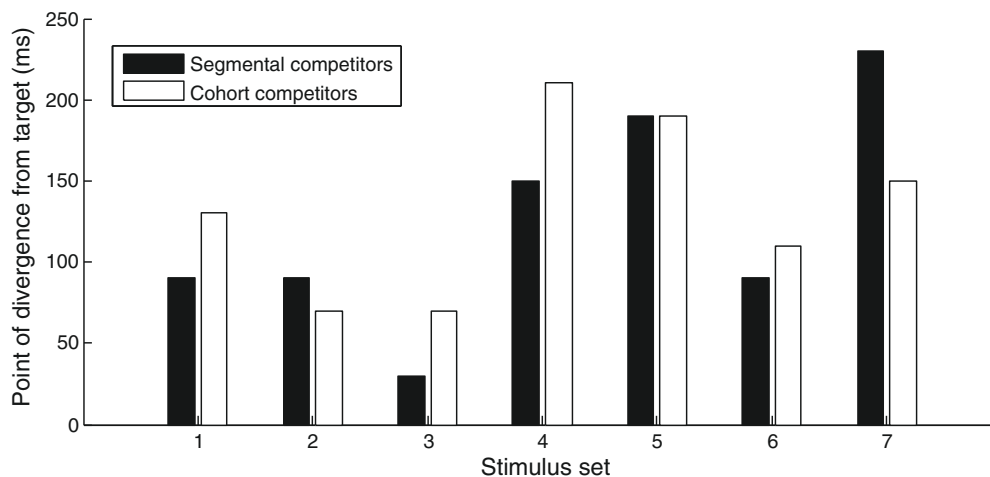
Fig. A2. Points of divergence (POD) of target vs. segmental or cohort competitors, for all seven stimulus sets (see Table A1).

Table A1
Experimental stimuli.

Stimulus set	Target	Segmental competitor	Cohort competitor	Rhyme competitor	Tonal competitor
1	*chuang2* (bed) [454]	*chuang1* (window) [401]	*chuan2* (ship) [1016]	*huang2* (yellow) [508]	*niu2* (cow) [373]
2	*hua1* (flower) [1574]	*hua4* (painting) [602]	*hui1* (gray) [284]	*gua1* (melon) [174]	*jin1* (gold) [695]
3	*mi4* (honey) [125]	*mi3* (rice) [852]	*mian4* (noodles) [4754]	*di4* (land) [4809]	*tu4* (rabbit) [103]
4	*qiu2* (ball) [550]	*qiu1* (autumn) [226]	*qian2* (money) [953]	*niu2* (cow) [373]	*huang2* (yellow) [508]
5	*shu3* (mouse) [40]	*shu1* (book) [1583]	*shui3* (water) [3318]	*tu3* (dirt) [1014]	*mi3* (rice) [852]
6	*tu3* (dirt) [1014]	*tu4* (rabbit) [103]	*tui3* (leg) [321]	*shu3* (mouse) [40]	*mi3* (rice) [852]
7	*xin1* (heart) [3963]	*xin4* (envelope) [1202]	*xia1* (shrimp) [54]	*jin1* (gold) [695]	*gua1* (melon) [174]

Note. Values in square brackets denote morphemic frequency of each item, as indicated in Modern Chinese frequency dictionary (1986).

and F2 values were calculated for each of the five tokens using Praat version 5.1.20 (Boersma & Weenink, 2009). To offset differences in duration across the five tokens, frequency values were time normalized by first dividing tokens into twenty time intervals of equal size, and then recalibrating the intervals according to the mean duration of the five tokens. These re-scaled curves were then binned into 20 ms windows starting from word onset, and two-sample *t*-tests (independent samples) were performed to compare frequencies at each window. Note that samples that did not yield frequency values due to the absence of phonation (such as voiceless initial consonants) were excluded from comparisons.

The point of divergence (POD) between a segmental competitor pair was defined as the first of three consecutive time windows in which the target and competitor F0 differed at $p < .01$. This ensured that the POD reflected the time at which the two forms differed reliably; this also reflects Robinson & Patterson's (1995) finding that participants require six to eight cycles of F0 to reliably identify pitch of vowels, which for the F0 range of this speaker (around 75–200 Hz) would be about three windows (60 ms). In the case of cohort competitor pairs the POD was the first of three consecutive time windows at which either F1 or F2 differed at $p < .01$ (whichever came first). Fig. A1 illustrates POD values for one of the target-competitor sets.

Mean POD values were pooled across all stimulus pairs in each condition. As indicated in Fig. A2, a paired-sample *t*-test [$t(6) = -0.478$; $p = .65$] showed that mean POD timing did not differ systematically across the segmental and cohort competitor pairs, confirming our assertion that the divergence in the two types of mismatch occurred at comparable points in time.

References

Allopenna, P. D., Magnuson, J. S., & Tanenhaus, M. K. (1998). Tracking the time course of spoken word recognition using eye movements: Evidence for continuous mapping models. *Journal of Memory and Language, 38*, 419–439.

Boersma, P., & Weenink, D. (2009). Praat: Doing phonetics by computer (Version 5.1.20) [Computer program]. <http://www.praat.org/> Retrieved 28.11.09.

Chen, J., Chen, T., & 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. (1986). Forbear is a homophone: Lexical prosody does not constrain lexical access. *Language and Speech, 29*(3), 201–220.

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

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

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.

Cutler, A., Wales, R., Cooper, N., & Janssen, J. (2007). Dutch listeners' use of suprasegmental cues to English stress. In J. Trouvain & W. J. Barry

(Eds.), *Proceedings of the 16th international congress of phonetic sciences (ICPhS 2007)* (pp. 1913–1916). Dudweiler: Pirrot.

Desroches, A. S., Joanisse, M. F., & Robertson, E. K. (2006). Specific phonological impairments in dyslexia revealed by eyetracking. *Cognition, 100*, B32–B42.

Francis, A. L., Ciocca, V., & Ng, B. K. C. (2003). On the (non)categorical perception of lexical tones. *Perception & Psychophysics, 65*, 1029–1044.

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

Liu, S., & Samuel, A. G. (2007). The role of Mandarin lexical tones in lexical access under different contextual conditions. *Language and Cognitive Processes, 22*(4), 566–594.

Liu, H., Tsao, F., & Kuhl, P. K. (2007). Acoustic analysis of lexical tone in Mandarin infant-directed speech. *Developmental Psychology, 43*, 912–917.

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

Marslen-Wilson, W. D. (1987). Functional parallelism in spoken word-recognition. *Cognition Special Issue: Spoken Word Recognition, 25*(1–2), 71–102.

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

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

Mirman, D., Dixon, J. A., & Magnuson, J. S. (2008). Statistical and computational models of the visual world paradigm: Growth curves and individual differences. *Journal of Memory and Language, 59*, 475–494.

Modern Chinese frequency dictionary (1986). Beijing: Beijing Language Institute Publisher [in Chinese].

Norris, D. (1994). Shortlist: A connectionist model of continuous speech recognition. *Cognition, 52*(3), 189–234.

Norris, D., & McQueen, J. M. (2008). Shortlist B: A Bayesian model of continuous speech recognition. *Psychological Review, 115*, 357–395.

Norris, D., McQueen, J. M., & Cutler, A. (2000). Merging information in speech recognition: Feedback is never necessary. *Behavioral and Brain Sciences, 23*(3), 299–370.

Pastizzo, M. J., & Feldman, L. B. (2002). Discrepancies between orthographic and unrelated baselines in masked priming undermine a decompositional account of morphological facilitation. *Journal of Experimental Psychology: Learning, Memory and Cognition, 28*, 244–249.

Repp, B. H., & Lin, H. (1990). Integration of segmental and tonal information in speech perception: A cross-linguistic study. *Journal of Phonetics, 18*(4), 481–495.

Robinson, K., & Patterson, R. D. (1995). The stimulus duration required to identify vowels, their octave, and their pitch chroma. *Journal of the Acoustical Society of America, 98*(4), 1858–1865.

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

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

Spivey, M. (2007). *The continuity of mind*. New York: Oxford University Press.

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

Tanenhaus, M. K., Magnuson, J. S., Dahan, D., & Chambers, C. (2000). Eye movements and lexical access in spoken-language comprehension: Evaluating a linking hypothesis between fixations and linguistic processing. *Journal of Psycholinguistic Research, 29*, 557–580.

Tanenhaus, M. K., Spivey-Knowlton, M. J., Eberhard, K. M., & Sedivy, J. C. (1995). Integration of visual and linguistic information in spoken language comprehension. *Science, 268*, 1632–1634.

Tong, Y., Francis, A. L., & Gandour, J. T. (2008). Processing dependencies between segmental and suprasegmental features in Mandarin Chinese. *Language and Cognitive Processes, 23*(5), 689–708.

van Donselaar, W., Koster, M., & Cutler, A. (2005). Exploring the role of lexical stress in lexical recognition. *The Quarterly Journal of Experimental Psychology A: Human Experimental Psychology, 58A*, 251–273.

Xu, Y., Gandour, J. T., & Francis, A. L. (2006). Effects of language experience and stimulus complexity on the categorical perception of pitch direction. *The Journal of the Acoustical Society of America, 120*(2), 1063–1074.

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

Zhou, X., & Marslen-Wilson, W. (1994). Words, morphemes, and syllables in the Chinese mental lexicon. *Language and Cognitive Processes, 9*(3), 393–422.