Processing Tone and Vowel Information in Mandarin: An Eye-tracking Study of Contextual Effects on Speech Processing

by

Xizi Deng

M.A. (Linguistics), The Chinese University of Hong Kong, 2016

B.A. (Chinese Language and Literature), Sichuan Normal University, 2015

Thesis Submitted in Partial Fulfillment of the
Requirements for the Degree of
Master of Arts

in the
Department of Linguistics
Faculty of Arts and Social Sciences

SIMON FRASER UNIVERSITY

Spring 2020

Approval

Name:	**Xizi Deng**
Degree:	**Master of Arts**
Title:	**Processing Tone and Vowel Information in Mandarin: An Eye-tracking Study of Contextual Effects on Speech Processing**
Examining Committee:	**Chair:** Suzanne Hilgendorf Associate Professor
	Henny Yeung Senior Supervisor Assistant Professor
	John Alderete Supervisor Professor
	Ashley Farris-Trimble Supervisor Assistant Professor
	Philip Monahan External Examiner Assistant Professor Department of Linguistics University of Toronto
Date Defended/Approved:	April 6th, 2020

Ethics Statement

The author, whose name appears on the title page of this work, has obtained, for the research described in this work, either:

a. human research ethics approval from the Simon Fraser University Office of Research Ethics

or

b. advance approval of the animal care protocol from the University Animal Care Committee of Simon Fraser University

or has conducted the research

c. as a co-investigator, collaborator, or research assistant in a research project approved in advance.

A copy of the approval letter has been filed with the Theses Office of the University Library at the time of submission of this thesis or project.

The original application for approval and letter of approval are filed with the relevant offices. Inquiries may be directed to those authorities.

Abstract

Prior work has suggested that rime (vowel) information is given priority over tone information in the perception of isolated words but there are flipped roles of tone and rime in a semantically constraining context. Here, I examined the eye gaze of native listeners of Mandarin Chinese, asking when and how top-down contextual effects from hearing a noun classifier constrains real-time processing of a target noun, and whether this classifier context has differential impacts on activating tone and rime information. The results show that, when hearing the classifier, average looking time to the target noun and noun competitors with the same tone or rime was significantly greater than to phonologically unrelated nouns. Moreover, fixations to the target were significantly greater to the phonological competitors only in a high-constraint classifier context. In addition, there was more distraction from a tone competitor than a rime competitor only in the high-constraint context. Results suggest that segmental and lexical tone perception follow different perceptual processes, and that tone was predicted ahead of rime when perceiving spoken words in context.

Keywords: Spoken word recognition; contextual effects; eye-tracking; Mandarin Chinese

Acknowledgements

The person who I would like to give my deepest gratitude to will definitely be my senior supervisor Dr. Henny Yeung, who has given me many insightful comments on my first, second and final drafts. Not only did he guide me all the way to explore a new field of speech perception, which I hardly set foot in before, but he also raised lots of intriguing questions about my research, which kept encouraging me to devote myself to relentless learning. He is the first professor who told me that he saw potential in me for doing research. These words moved me so much that I have decided to bear his encouragement in mind throughout my future academic life.

My gratitude also goes to Dr. Ashley Farris-Trimble and Dr. John Alderete. Dr. Farris-Trimble equipped me with foundational knowledge of designing an eye-tracking experiment and relevant techniques, and Dr. Alderete also provided me with lots of helpful suggestions on my drafts from a theoretical view.

Finally, my thanks go to my family in China and friends at SFU. They all provided me with warm emotional support whenever I needed it, and my friends who are native Mandarin speakers selflessly spent time participating in the experiment and contributed the authentic and precious data to my current study.

Table of Contents

List of Tables

List of Figures

Chapter 1.

Introduction

In most modern theories of spoken word recognition, a word is accessed by mapping continuous sensory input to abstract lexical items stored in the mental lexicon, which includes both low-level phonological representations and high-level semantic information. Marslen-Wilson & Welsh (1978) proposed “cohort model” to account for this process. When an individual word is heard, according to the cohort model, initial sensory input would activate a set of lexical candidates that share the same phonemes in listeners’ mental lexicons. During the assessment of the match between continuous acoustic signals and stored lexical candidates, unmatched candidates would be constantly abandoned until only one lexical item is consistent with all received acoustic input. However, words are usually embedded in a longer linguistic context in daily speech, which could make the mechanism of word recognition different from what the cohort model has suggested. Because, in this case, word candidates are also compared to the semantic interpretation of the preceding context. Understanding how to access a lexical entry for a target word in this richer semantic context remains controversial, and more evidence that can better reflect real-time processing is needed. The current thesis work contributes to this broader goal, helping to show how semantic constraints from prior context guide lexical access using a real-time visual world paradigm.

In this chapter, I will introduce the theoretical background and relevant literature for the current study. First, I will review the basis of most theories of lexical access, outlining differences between interactive and integrative views. Then, I will discuss prior work that has asked how different types of phonological information, i.e. tone and rime, may be differently activated during lexical access in context, compared to how they are processed when presented in isolated syllables. Subsequently, I will explain why Mandarin classifier-noun agreement is the target of research here, and review an example of prior ERP studies using this structure to investigate contextual effects. Lastly, I will outline the two major research questions that the current study is dedicated to answer.

From this review, outstanding questions will be discussed, which form the basis of an experimental manipulation in Chapter 2, where I will present the experimental methods adopted in the current study and present empirical data that addresses whether contextual cues pre-activate phonological information about an upcoming word in Mandarin lexical access, and whether tone and rime information is accessed differently in semantically constraining contexts. In Chapter 3, I will conclude with implications from the results that draw inspiration from theories of speech production.

1.1. Theories of Lexical Access

Lexical access can operate in a bottom-up fashion, where, according to the TRACE model (McClelland & Elman, 1986), as more and more acoustic information is heard, the received input activates higher semantic levels by eliciting gradually increasing excitation for certain features, phonemes, and word items. On the other hand, lexical access also simultaneously functions in a top-down manner. At a certain recognition point, even if a word has not been fully articulated yet, previously available sensory input can bias our perception of the upcoming sounds towards an existing real word (Cutler, Mehler, Norris, & Segui, 1987; Elman & McClelland, 1988; Frauenfelder, Segui, & Dijkstra, 1990; Ganong, 1980; McClelland & Elman, 1986; Segui & Frauenfelder, 1986; Segui, Frauenfelder, & Mehler, 1981), or a word with higher lexical frequency (McClelland & Elman, 1986). For example, when hearing the onset sound in the word "_ash," where the critical sound is at the phonetic category boundary of [d] and [t], participants tended to categorize this sound as [d] rather than [t] due to the different lexical status of the word "dash" and non-word "tash" in their mental lexicon (Ganong, 1980). In addition, Frauenfelder et al. (1990) found a stronger lexical effect after the uniqueness point of spoken words (time points when cohort competitors were fully excluded and successful word recognition took place) in a task of detecting phonemes [p, t] and [k], suggesting an interaction between the initial bottom-up approach and a later top-down process. There has been tremendous literature on single-word perception and more precisely on how higher-level semantic content activates lower-level phonological elements at both segmental and suprasegmental levels (Cutler & Chen, 1997; Lee, 2007; Malins & Joanisse, 2010; Sereno & Lee, 2015; Wiener & Turnbull, 2016; Ye & Connine, 1999)

Nevertheless, in natural speech, words are seldom heard in isolation, but rather in a larger linguistic context. Retrieval of the semantic aspects of a lexical item does not happen only at a local level. That is, the semantic context where the words are embedded will have a significant effect on activating the meaning of a target word in various ways (Moss & Marslen-Wilson, 1993; Rayner & Frazier, 1989; Seidenberg, Waters, Sanders, & Langer, 1984; Sereno, Brewer, & O'Donnell, 2003; Tweedy, Lapinski, & Schvaneveldt, 1977; Tyler & Wessels, 1983). For example, different semantic aspects of "tomato" are in focus within the following three sentences (Moss & Marslen-Wilson, 1993):

a. The tomato rolled across the floor (the shape)

b. The sun was a ripe tomato (the color)

c. He accidentally sat on the tomato (the squishiness)

Unlike recognizing words in isolation, which takes place in a single moment (Moss & Marslen-Wilson, 1993), the semantic context may have guided access to the meaning of a target word, and influenced perception of phonological information of the word from very early in a top-down manner. In the examples above, "rolled" implies the shape of the tomato to be round; "the sun" and "ripe" emphasizes the red color of a tomato; "sat" indicates the texture of the tomato. However, there has been a rather limited amount of previous work on word recognition in context, which resembles speech perception in natural settings more than the perception of words in isolation. The current study thus contributes to investigating how phonological details of words are perceived in semantically constraining contexts.

1.1.1. Lexical Access in Context

If contextual effects change the activation of a target word, a central question is when and how this activation operates. With regard to the timing of activation, there have been two major views. First, there is an interactive view, which proposes that listeners actively use semantic information from a prior context to make predictions about upcoming words (Dambacher, Kliegl, Hofmann, & Jacobs, 2006; DeLong, Urbach, & Kutas, 2005; Federmeier, 2007; Federmeier & Kutas, 1999; Federmeier, Wlotko, De Ochoa-Dewald, & Kutas, 2007; Liu, Shu, & Wei, 2006; Sereno et al., 2003; Van Petten & Kutas, 1990). That is, lexical activation is initiated during semantic access from a prior context. Second, there is an integrative view, which holds that all the semantic fragments in preceding context are integrated with lexical information of a target word only after recognition of the target word has been accomplished (Traxler & Foss, 2000; Tyler & Wessels, 1983). That is, lexical activation is a delayed process where semantic congruency will be scrutinized after global semantic information from the whole sentence has been accessed.

Under the theoretical framework of the two above views, how contextual activation operates could be quite different as well. Some studies (e.g., Federmeier and Kutas, 1999) propose an interactive view, and argue that a preceding context predicts a set of semantic features of the upcoming target word. This view is supported by work showing that implausible words belonging to the same semantic category as the target word received more activation than between-category implausible words. Furthermore, similar to the top-down aspect of semantic activation from the mental lexicon during single-word recognition, contextual effects extend to activate lower-level phonological details of the target word. When hearing a semantically constraining context, listeners not only anticipate possible semantic categories for what is coming next, but also predict a specific word that has the highest cloze probability from the preceding context. That is, both lexical and phonological information of the target word could be fully derived from the prior context in a pure top-down fashion in absence of the actual sensory input.

As opposed to the interactive view, studies (e.g. Norris, 1986; Seidenberg et al., 1984; Seidenberg, 1985; Traxler & Foss, 2000; Tyler & Wessels, 1983; Zwitserlood, 1989) arguing for the integrative view have claimed that lexical candidates could not be pre-selected by semantic context before integrating information from any acoustic signals of the target word. Rather, only after bottom-up processes had placed constraints on candidates could contextual effects be exerted. A constraining context merely accelerates the interaction between top-down and bottom-up approaches in word identification. That is, when limited phonological information of the target word activated a large set of lexical candidates, the prior context would rapidly scrutinize the semantic congruency between the context and the possible candidates and strengthen the confidence in selecting only one plausible candidate. Compared to single-word perception, less sensory input of the target word would thus be needed for recognition when it is embedded in constraining context, due to the rich information provided in a prior context. A detailed review and comparison of these two views follows below.

1.1.2. Interactive view

Most recent studies arguing for the interactive view have adopted the online technique of event-related potentials (ERPs). In these studies, two types of context (i.e. high-constraint vs. low-constraint contexts) were created. It was hypothesized that if the contextual effects started earlier, then a high-constraint context would pre-activate a narrower scope of possible completions, while low-constraint contexts would demonstrate a weaker prediction of subsequent words, precipitating a larger set of lexical candidates. For example, in (1a), only words like “disease” or “gift” are strongly anticipated from the preceding context, while in (1b), there is weaker activation for a specific word among a series of word candidates, because the context constraints candidates to a wider semantic range (Federmeier, 2007):

(1) a. The child was born with a rare *disease* / *gift*.
b. Mary went into her room to look at her *clothes / gift / cat / book*…

Different ERP patterns would be attested in the two context conditions above under the interactive view, but, in contrast, the integrative view holds that semantic context merely functions to screen the semantic plausibility of the upcoming words without concerning degree of constraint, and thus integrative accounts would hypothesize that similar ERP results would be found equally in the two context types. In order to factor out varying expectancy and perceptual difficulty of the target words and ensure any potential differences in the results were driven only by contextual effects, easiness of processing the target words was also controlled in many studies. Two types of target words, were designed, including high vs. low cloze-probability words (Federmeier et al., 2007), high vs. low lexical-frequency words (Van Petten and Kutas, 1990; Sereno et al., 2003; Dambacher et al., 2006) and short vs. long words (Van Petten and Kutas, 1990), all attached to the two context types, resulting in four types of experimental conditions. Smaller N400s can reflect greater semantic expectancy or congruency. In Federmeier et al. (2007), for high cloze probability words, the authors found that there was a significantly smaller N400 elicited by a high-constraint context than that by a low-constraint context, indicating a facilitatory role of the high-constraint context in word processing, whereas for low cloze probability words, the N400 elicited by the two context types was equivalent.

In addition to controlling the factor of the expectancy and perceptual difficulty of the target word, contextual activation was also instantiated by comparing the ERPs elicited by a low- or high-constraint context, compared with ERPs elicited by lexical frequency. In Van Petten and Kutas (1990), although low-frequency words were associated with larger N400 compared to high-frequency words, effect of word frequency was obscured by word position as the target word was shifted from the beginning of a sentence to the intermediate position, and then to the terminal position, suggesting that a more powerful semantic constraint from context than word frequency.

Similarly, in Sereno et al. (2003), four experimental conditions were created: a neutral context with a high-frequency word, a neutral context with a low-frequency word, a neutral context with an ambiguous word, and a biased context with an ambiguous word (where the context was biased towards the subordinate meaning of the ambiguous word, as those words had a primary meaning that was more frequent than the subordinate meaning, analogous to high-frequency relative to low-frequency conditions). Sereno et al. observed that the effect of lexical frequency was elicited in a neutral context, and reflected in significantly different N1 components (from 132 ms to 192 ms). More importantly, ambiguous words acted like high-frequency words in the neutral context, suggesting that their primary sense was activated. However, the biased context exclusively influenced activation of the subordinate meaning of an ambiguous word. The result was more supportive of an interactive model in a sense that access to lexical entry was initiated early, and showed that both lexical frequency and context effects were manifested in the N1 component (~150 milliseconds after the start of the word). As opposed to what an integrative model would claim, participants did not wait till both the primary and subordinate word senses were activated and suppressed the meaning that was incongruent with the context, which would have been reflected by different processing stages for effects of lexical frequency and context.

In addition to ERP studies, behavioural studies (Altmann & Kamide, 1999; Kamide, Altmann, & Haywood, 2003) recording participants' eye gaze, synchronized to crucial information in the speech stream, also provided more direct evidence supporting the interactive view. In Altmann and Kamide's (1999) experiment, participants would hear sentences that were either like, "The boy will eat the cake" or "The boy will move the cake," and see four visual referents among which the cake was the only edible object. They found that the probability of launching the first saccade to the target before the onset of the noun in the *eat* condition was significantly larger than that in the *move* condition. The result indicates that the semantic information extracted from the preceding context (the verb) can guide eye movements towards the object whose thematical role fits with the requirement of the verb by comparing lexical concepts of noun candidates as early as the verb-specific knowledge about thematical roles is activated, which supports the locus of contextual effect predicted by the interactive view. However, since the context in this experiment did not provide semantic information constraining a specific noun to be projected by the verb, only a thematic specification containing a range of lexical concepts was pre-activated. Kamide et al. (2003) further showed that information provided by a verb was not only capable of predicting the thematic information of an immediately subsequent object, but also an object that occurred at a later point. These studies have shown speech processing is an incremental process where listeners instantly utilize what is available in the context at the moment to make a “broad guess” about the characteristics of the upcoming word at a very early point, so as to make an interpretation as complete as possible, as early as possible in time.

The results from the studies described above argued for an independent role of semantically constraining context in accessing a lexical entry, further supporting the idea that a semantic context provides dynamic semantic information for participants to make categorical decisions about semantic plausibility at a very early stage of word recognition. A further question is what type(s) of information in the target word can this semantic contextual effect pre-activate, and previous studies have suggested both semantic and phonological information is relevant.

At the semantic level, Federmeier and Kutas (1999) found that context activated a set of semantic components of the target word. In their study, a semantically constraining context was matched with a plausible word, an implausible word from the same semantic category as the expected word (it also shares some of the semantic features with the expected word), and an implausible word from a different semantic category. For example, in the first condition, the word “palms” was the most expected completion in the following sentence:

> (2) They wanted to make the hotel look more like a tropical resort. So along the driveway, they planted rows of_____.

In the second condition, the sentence ended with a within-category, but implausible word, “pines,” which are also a type of tree, but are not tropical. In the last condition, the sentence was completed with a between-category and implausible word, “tulips,” which do not belong to the category of trees and are not tropical. When comparing these two implausible words, a significantly smaller N400 was elicited by the within-category violation (“pines”) than the between-category counterpart (“tulips”) even if they shared the same cloze probability. This result argued against the integrative view in a sense that plausibility did not appear to be the sole factor influencing context comprehension. Rather, a prior semantic context seems to pre-activate a set of semantic features of the target word (in this case, [+tropical, +tree, etc.]), resulting in categorically similar words also being primed due to featural overlap (e.g., [+tree]) in long-term semantic memory.

At phonological level, it was also found that context predicted the most expected word with its phonological details specified. In Delong et al. (2005), for example, the most expected word is “kite” in the context shown below, which begins with a consonant:

(3) The day was breezy. So the boy went outside to fly _____.

A grammatical completion of this sentence involves an article, and “a” was anticipated before the target word, “kite”, which began with a consonant. However, when an alternative completion was played, like “an airplane,” which had a lower expectancy, an N400 was observed as early as the article “an” was heard. More directly, Liu et al. (2006) found that, in a high-constraint context, the amplitude of N400 would be smaller when the target word mismatched the onset segment of the expected word with a single phonological feature, compared to when it mismatched the expected word in multiple phonological features. However, this discrepancy disappeared in low-constraint context. This latter study indicates that at least the onset of the expected word can be accessed to at both the phoneme and feature level through the preceding context.

In summary, studies supporting the interactive view make several claims: First, a semantically constraining context advances the timeline of when the lexical entry of the target word is accessed, and this can happen even before any contact with acoustic cues of the target word occurs. Second, contextual effects pre-activate information in the target word at both semantic and phonological levels. In the next subsection, studies from the integrative view on these issues will be discussed, focusing on the dynamics and types of contextual effects.

1.1.3. Integrative view

In contrast to the interactive view, the integrative view holds that a preceding context does not have any early effects on lexical access. Rather, whether words can get processed successfully depends on how easy they can be integrated in a global discourse at a later stage. Two assumptions can be generated from this account: First, both high-constraint and low-constraint contexts can prime expected words, to the same extent, as long as they are semantically congruent; and second, that both high- and low-constraint contexts can equally prime unexpected words that are semantically associated to the expected words to different degrees, as long as they are semantically congruent. Traxler & Foss (2000) designed an experiment where participants needed to listen to meaningful context including high-constraint and low-constraint sentences followed by semantically close and distant competitors of the most expected words. Experimental participants were then asked to name the final target words. They found a uniform priming effect across all the four conditions and concluded that the results fitted better with the integrative view, since any words that made the sentence meaningful received equally distributed amount of processing facilitation. This would be surprising from the interactive view, since although the constraining strength of the prior context and expectancy of the target words were manipulated, there was no apparent predictive effect from the prior context.

With regard to the specific locus of contextual effects under the integrative view, some studies (e.g. Seidenberg et al., 1984; Seidenberg, 1985; Norris, 1986) found that context effects could only be demonstrated at a post-lexical stage, where decoding lexical information did not require context, and a context could only function in structural and semantic integration. However, other studies (e.g. Tyler & Wessels, 1983; Zwitserlood, 1989) argued that contextual information facilitated rapidly narrowing down potential lexical candidates, which were semantically congruent with the prior context and thus accelerated the lexical decoding. Tyler & Wessels (1983) found that when a word was embedded in a semantically constraining context, only 50% to 65% of the word needed to be heard with 80% confidence in their choice, before participants could recognize the word. This suggests that semantically constraining context only serves to accelerate the lexical process of word recognition, instead of pre-selecting a proper lexical target. An implication of this idea is that bottom-up activation is still required. More semantic information was provided in the context to help narrow down the lexical candidates activated by limited phonetic information.

Zwitserlood (1989) designed an experiment where a spoken word in Dutch (e.g. ‘kapitein,’ which means “captain”) that had a close competitor differing in the last vowel (e.g. ‘kapitaal,’ or “capital”) was placed at the end of a sentence of different contexts (neutral and biased contexts). Then, either one of two visual target words that are semantically associated with the actual presented word (e.g. ‘schip’ - “ship”) or its close competitor (e.g. ‘geld’ - “money”) would appear at different crucial time points of the final target word. Participants were requested to make lexical decision on the visual words. The results showed that, at an early stage of processing (Position 1: 130ms on average after the onset of the target word and Position 2: 199ms on average after the target word, which were defined as the stage when the actual word was still one of the lexical candidates), the mean reaction times of the actual word probe and close competitor probe were approximately identical across both the neutral and biased contexts. Indeed, a context-effect started to be significant only after Position 3 (278ms on average after the target word, which was defined as the selection phase for the target among several candidates), reflected by a significant difference between the reaction time of the actual word probe and close competitor probe only in biased context instead of neutral context at this time point. These results indicate that both the actual presented words and their close competitors were activated at an early stage of word processing, regardless of the preceding semantic context. The effect of biased context could only be elicited considerably during later processing when the selection among series of lexical candidates was tapped into.

Overall, these results support the integrative view in the sense that a semantically constraining context did not pre-activate a target word with all its phonological information implicitly specified. That is, contextual effects cannot influence word recognition before sufficient sensory signals were provided for lexical processing. However, under the integrative view, contextual constraint does not have a predictive effect and successful lexical access still depends on a certain amount of sensory input, thus both Tyler & Wessels (1983) and Zwitserlood (1989) still argued for an important role of contextual constraint in lexical processing. Compared to neutral contexts, constraints from biased contexts increase the speed of selecting a semantically appropriate lexical item among all the lexical candidates that are already activated by current sensory input. This happens by scrutinizing the semantic consistency of the information in the context and lexical information of the candidates. This process is not necessarily post-lexical, as the integration does not happen only after all the bottom-level acoustic information is available and the target lexical entry is fully specified. Rather, the integration likely takes place during lexical processing and facilitates the selection by reducing the acoustic signals that are needed.

To summarize, these studies concluded that processing words in context was a delayed, bottom-up process, where the recognition must wait until sufficient sensory signals activate a certain amount of lexical information. The contextual constraint cannot predict the target word before any sensory signals are received. Rather, the context provides semantic information to accelerate the selection of target word among a series of lexical candidates activated by later acoustic information. That is, listeners check the semantic congruency between the context and lexical candidates to quickly decide the semantically appropriate target word. Therefore, constraining strength should have little effect before the target words are available.

1.1.4. Interim summary

The interactive model and integrative model propose different mechanisms of word recognition in context. The interactive model argues for a more involved role of context where people utilize any prior linguistic information to make predictions in a top-down manner, while the integrative model holds that the listeners delay the recognition process until the actual acoustic signals are received. The studies in favour of the integrative view adopted experimental methods such as lexical decision and naming tasks, while more recent studies supporting the interactive view preferred more real-time techniques such as eye tracking and ERPs. In the latter studies, a continuous stream of data rather than a single discrete response was obtained, which would have better captured the dynamics of word processing in a naturalistic scenario.

No matter which view has provided a more convincing account for processing both the high-level semantic content and low-level phonological representation of a target word in context, neither of these views has so far addressed another unsettled dispute in how context affects the processing of different phonological elements in a target word, i.e. segments and suprasegmentals. The current study aims at exploring the process of perceiving different phonological elements of a word in context, asking whether contextual effects would be exerted on these two types of phonological information differently. Furthermore, this study will investigate which of the two accounts can provide a more convincing explanation for the differences in processing between these two types of phonological information (if there are any), which has not been covered by the previous literature. In the next section, I will compare the results from previous studies on perception of vowels and lexical tones in single-word context, and in longer phrasal/sentential contexts, emphasizing places where there is insufficient model construction in prior work.

1.2. Perception of segments vs. lexical tones in single words and in context

There has been a vigorous debate on the roles of segmental and suprasegmental information, including lexical stress, pitch accent and lexical tones, in lexical access both when the word is presented in isolation and embedded in context. In single-word perception, some studies showed that there was a greater contribution of segmental information to lexical access than suprasegmental information, such as lexical stress (Cutler, 1986; Cutler & Clifton, 1984; Slowiaczek, 1990) and tones (Cutler & Chen,1997; Sereno & Lee, 2015; Wiener & Turnbull, 2015). Yet, others found that activation of a lexical representation was equivalently sensitive to both segmental and suprasegmental cues native to the language, including lexical stress (Cutler & Donselaar, 2001; Soto-faraco, Sebastián-gallés, & Cutler, 2001), pitch accent (Cutler & Otake, 1999) and lexical tones (Liu & Samuel, 2007; Malins & Joanisse, 2010; Schirmer, Tang, Penney, Gunter, & Chen, 2005). However, when a word is presented in a semantically constraining context, suprasegmentals such as lexical tones appeared to win more perceptual weight back, reflected by comparable time course of processing tone and vowel (Schirmer et al., 2005) or even an early access to tonal information (Ye & Connine, 1999; Liu & Samuel, 2007), which reversed one of the patterns found in single-word perception.

It is not unreasonable to suppose that the differences between processing these two types of information under these different conditions must have something to do with the mechanism of word processing in context. In single-word recognition, processing these types of phonological information involves the interaction of high-level semantic content and low-level phonological information, which takes place without a semantic context, and thus any differences in processing likely result from how phoneme- and tone-level information is constructed and how they are connected to higher word-level information in a perception model. However, in context, the high-level semantic activation is from the accumulated semantic content within a larger domain, which increases the time window for the top-down activation to become effective. Furthermore, no models have been established yet that describe how different types of phonological information is perceived in context. This study asks whether the same pattern of processing for tonal and vowel information in context would be found using a real-time processing technique. More importantly, I would like to explore what causes the differences between processing the two types of phonological information and how the results can be fitted in the bigger picture that the interactive or integrative view has created at a theoretical level.

Since, in the previous studies, the differences in processing segments and suprasegmentals in context were specifically found between vowels and lexical tones, literature on other suprasegmentals such as lexical stress and pitch accent will not be covered. In this section, I will first review previous literature on processing lexical tones and vowels in single words, followed by prior work on processing these two types of information in a longer context, in order to present a broad view on different roles of vowels and lexical tones. Mandarin Chinese is selected as the specific case for studying contextual effects on perceiving phonological information. Specifically, in the experiment design outlined in Chapter 2, classifier-noun phrases were used as experimental stimuli, where the classifiers functioned as context, while the rime and tone information in the nouns was systematically varied. In the subsequent sub-section, the characteristics of this phrasal structure in Mandarin will be described as well as the reason for choosing this phrasal structure.

1.2.1. Processing phonological information in single words

There remains a great controversy in the role of lexical tone versus segments, especially vowels, in lexical access. On the one hand, different lexical tones are cued by pitch height and contour (Chao, 1968), which is similar to the perceptual mechanism for lexical stress in using pitch height as a cue (Wang, 2008). On the other hand, lexical tones are able to differentiate word meanings and thus determine the access to different lexical entries. For example, the four tones borne by the same Mandarin syllable “ma” result in four different meanings, as showed in the following table:

Table 1. Mandarin tone inventories

	Tone 1	Tone 2	Tone 3	Tone 4
Syllable	ma1 妈	ma2 麻	ma3 马	ma4 骂
Meaning	mother	linen	horse	to scold

There have been studies that found tonal information was not as significant as segmental information in accessing the mental lexicon, which also suggest that tone processing seems to be delayed compared to segmental processing. Such studies have used tasks like lexical decision, tone priming, and word generation, which involves changing a non-word to a real word. Cutler and Chen (1997) used a lexical decision task, where foil words close to real Cantonese words were created by changing its tone or segments. Here, native speakers had a lower accuracy rate when the target word’s tone was changed to form a non-word than when the segments were changed. Similarly, in a same-different judgement task, the response times and the accuracy rates were lower when the two syllables only differed in tones. The findings suggested that segmental information was processed sooner and more accurately than tonal distinctions. Moreover, in Ye and Connine (1999), when subjects were asked to monitor for the presence of a syllable [a] with Tone 2, the response times were faster if the stimulus mismatched with the target in vowel versus tonal information, suggesting that vowel was treated as more salient in recognizing perceptual differences.

Parallel results are seen in other tasks. For example, Sereno and Lee (2014) conducted a priming task, where the primes and targets overlapped in terms of segments and tones, segments alone, or tones alone. The results showed the strongest priming effect when both segments and tones overlapped, with a reduced priming effect when only segments matched, and no priming effect when only tones overlapped. This indicated that tonal information did not appear to immediately block incorrect competitors, contributing only in a minor way to lexical access. Resonating with previous studies, in a word reconstruction task, Wiener and Turnbull (2015) discovered that when subjects were asked to change either a consonant, vowel or tone to alter a non-word to a real word, the response time was faster when subjects were asked to change a tone. Additionally, in the free-change condition, more subjects tended to change tones than segments. The results suggested that tones constrained lexical access less tightly than segments. That is, when tone is processed in isolation, it is less salient than segments in constraining lexical access.

Nevertheless, there are other researchers who argued for a more important role of tone in speech perception. They found that tone and vowel were processed simultaneously with comparable perceptual significance. Lee (2007) found that when the prime and target differed from each other only in tone, the prime failed to activate the target, indicating a preferential usage of tone to distinguish segmentally identical items. Such a result also shows that tone might be perceptually as important as segments in reducing incompatible lexical candidates during the activation. Furthermore, Malins & Joanisse (2010) designed an eye tracking experiment where competitors mismatched with the target in onset, rime and tone. They found that there was a greater delay of looking at the target in tone mismatch condition and rime mismatch condition than the baseline condition. Critically, these authors observed no greater delay in onset mismatch condition than the baseline, while the time course of the rime mismatch condition and tone mismatch condition was consistent. This suggests that the tone and rime were processed simultaneously. These studies, on the contrary, suggested an equally significant role and comparable processing time course of tone and vowel in natural speech perception.

1.2.2. Processing phonological information in longer contexts

What is more intriguing is that some studies are concerned with tonal and segmental processing in longer contexts, such as four-syllable idiomatic words in Mandarin, where the idioms themselves formed a semantically and syntactically constraining context. Ye and Connine (1999) discovered that in idiomatic context, the detection of tone of the final syllable was faster than the detection of vowel. The difference still stood, though to a lower degree, when the semantic context was manipulated by changing the tone of the third syllable. The interpretation of the results could be that, since the four-syllable idioms in Mandarin were a high-constraint context (i.e., the final syllable in each idiom was fossilized), there would be only one possibility of what the final syllable can be, after hearing the first three syllables. The segmental and tonal information of the final syllable may well be pre-activated immediately when perceiving the first three syllables. Ye and Connine (1999) hypothesized that the reason why tone monitoring was faster than vowel monitoring in this experiment was that, in semantically constraining contexts, the activation from higher lexical node was stronger for toneme node than for phoneme node.

Perhaps because Ye and Connine (1999) was the first study tackling perception of segments and tones in semantically constraining contexts, more details about the mechanism of how semantic activation from the prior context affects perception of bottom-level components were not provided. The authors proposed different activation weights for tone and vowel, but did not answer the two core questions upon which the interactive and integrative model are distinguished: First, when does the semantic activation from the prior context start to be exerted, i.e., is it pre-lexical or lexical? That is, do participants predict the information of the target word beforehand or integrate it with the context after the word was processed? During the detection task in Ye & Connine (1999), acoustic signals of the target syllable might still be needed for participants to make a decision, and whether participants made a decision before the final syllable, during the final syllable or after, once it was fully heard, remain unknown. Thus, this experimental technique was not able to monitor implicit pre-lexical effects (if there were any) resulting from the preceding context during the time course of the perception process.

Second, how does semantic activation operate? Is it a top-down or bottom-up process? That is, whether the results better fit with the interactive model or integrative model remains unknown. Without embedding this account in a theoretical framework, the reason why the activation of tone would be greater than that of vowel in a context, which exhibits a different pattern than perceiving tone and vowels in single words, remains unanswered. Nevertheless, Ye & Connine (1999) was just an initial study, and later work examined semantically constraining contexts that have a higher frequency in natural daily speech, such as phrases and sentences. For example, Liu and Samuel's (2007) Experiment 1 had participants judge whether the last word they heard was a real word or a non-word in a single-word condition, sentential condition and idiomatic condition. Critically, the target real word was sometimes changed to a non-word by only alternating its tone or vowel. However, mixed results were attested. They found that there was a vowel advantage over tone in the sentential condition, reflected by a significantly lower accuracy rate in the tone mismatch condition (where the tone of a real word was changed to create a non-word), whereas in word and idiomatic conditions, there was no difference between vowel and tone cues, which partially confirmed what Ye and Connine (1999) found. However, these results failed to replicate the tonal advantage. Since, in the first experiment, participants made decisions after the phonological information of the target syllable was fully available, tonal salience might come from a post-lexical decision. Therefore, Liu and Samuel (2007) conducted a second experiment, where they put white noise (the degree was set to 75% for the first presentation and decreased by 2.5% in every following presentation) on the target non-words or syllables that had mispronounced tones or vowels (compared to the real words). They asked subjects to push a button as soon as they heard a tone in one condition, or write down the tone or vowel they detected in other conditions. Only in the sentential and idiomatic conditions was vowel detection impaired by tone mismatch. This finding was reflected by significantly lower accuracy in tone mismatch condition when participants were asked to respond to a vowel, compared to the baseline condition where vowel detection was performed when there was no mismatch on the target word. However, tone detection was not hurt in vowel mismatch condition in any of the three context types. The authors concluded that, under these contextual conditions, tone was given more perceptual weight than vowels, and thus the tonal processing advantage could only be reflected at a sub-lexical level, where the phonological details of the target syllable were obscured by noise.

Although this study argued for a more important role of tonal information under noisy conditions (and in high-constraint contexts), one of the shortcomings of lexical decision and monitoring tasks in these studies was that one cannot keep track of the time course of when the contextual effects started to be exerted. For example, the lexical decision task in Liu and Samuel (2007) could only show that it was easier or more difficult to integrate different types of information in different context scenarios. Likewise, in the second experiment, although access to the lexical information was impaired by the noise, after which contexts seemed to exhibit more of their effects, participants still could only make a response after the phonological elements were processed under the noise. Therefore, it did not allow one to ask how contextual effects function as participants build an expectation for the upcoming words in advance. The reason for this shortcoming was that these studies failed to provide data from real-time tracking, which reflected aspects of the recognition process that could distinguish the interactive and integrative models and that could fully illustrate contextual effects on processing tone and vowel.

Researchers have recently used ERP techniques to better capture the potential activation from semantic contexts, which are indexed by the time course and amplitude of N400. Schirmer et al. (2005) found that participants showed electrophysiological detection of tonal and segmental information at the same latency in a task where participants made semantic congruity judgments for Cantonese words that sometimes had altered segments or tones. This study concluded that, at least in a semantically constraining context, tones and vowels have a comparable time course. Although the authors argued for a more important role of tone in context, compared to processing isolated words in the previous reaction-time experiments, tone did not seem to outweigh vowel, as in Ye and Connine (1999) and Liu and Samuel (2007). However, in this study, participants still needed to process the intact phonological information of the target words, which resembles the first experiment in Liu and Samuel (2007). Even though ERP is a better technique to capture real-time processing, the study still failed to reflect any potential distinctions between sub-lexical and lexical processing.

In summary, unlike the earlier studies that found dominance of vowel over tone, recent studies involving real-time processing tasks such as eye-tracking and ERPs tend to conclude that tone and vowel have equal importance and comparable time courses in accessing to lexical representation. Moreover, studies investigating contextual effects found that tone appeared to be more dominant than vowels in these constraining contexts (Ye & Connine, 1999; Liu & Samuel, 2007), but only at a sub-lexical stage, when local lexical access was impaired (Liu & Samuel, 2007). When lexical information was fully available, vowel information regains the same amount of salience as tone. The pre-dominance of tone in constraining contexts was reflected by the fact that tonal information was earlier and more quickly accessed (Ye & Connine, 1999), and by the finding that tone information outweighed vowels in accessing lexical information under some noisy conditions (Liu & Samuel, 2007). However, there have not been any studies systematically examining the detailed mechanism of processing tone and vowel information in semantically constraining contexts under a theoretical view of word processing in context (e.g. interactive view or integrative view). Furthermore, how contexts influence processing tone and vowel (rime, as studied here) under the assumptions provided by the two views remains unclear. The current study aims to fill in this gap.

1.2.3. Motivation for current study

It still remains unknown which view better accounts for the contextual effects on processing words. Although more recent studies using real-time experimental techniques, such as ERPs, provided more solid evidence for the interactive view, there have been few behavioral studies showing the specific process of semantic context activating different types of phonological information, where different information might become attainable at different stages as acoustic cues unfold continuously. In addition, contextual effects seemed to reverse the perceptual weight of tone and vowel (rime) in lexical access. While tone and vowel appeared to have comparable time course and equal perceptual salience when processed in isolated words, they were processed at different levels when presented in a high-constraint context, with earlier and stronger activation on tone information.

In the current study, two issues will be addressed and investigated: First, whether the previous result that the perceptual weights of tone and vowel were reversed can be attested in a different high-constraint context, i.e. classifier-noun agreement, using real-time processing technique; Second, whether the interactive or integrative models better explain the underlying mechanisms of contextual effects that constrain processing tone and rime (i.e. do people predict beforehand or integrate afterwards and how does the process operate differently on tone and rime?). If contextual effects guiding lexical access took place at pre-lexical stage (Sereno et al., 2003), it is expected that target words will be activated before the acoustic signals are available, which supports the interactive model. Furthermore, as greater perceptual salience of tone was discovered at the sub-lexical level in a constraining context (Liu & Samuel, 2007), it is reasonable to expect that there would be greater activation of tone than rime when the lexical information has not been locally accessed. Additionally, resonating with previous studies, it is expected that there would be significant behavioral differences as the constraining strength of context varies, if the interactive view is a more reasonable account, while no significant distinctions if the integrative view has more explanatory power.

In order to investigate the issues mentioned above, manipulating constraining strength of context and teasing tonal information from rime information is necessary. In the current study, classifier-noun phrases in Mandarin Chinese were selected as a type of constraining context for two reasons: Mandarin has lexical tones, and the classifier-noun structure in NPs creates a convenient manipulation. In the next section, the characteristics of classifier-noun phrases along with the reasons for choosing them as the experimental materials will be introduced in detail.

1.3. Word recognition in a classifier-noun structure in Mandarin Chinese

Classifier-noun phrases in Mandarin Chinese were selected as the experimental stimuli for two reasons: First, the constraining effect of classifiers is easy to manipulate, with some classifiers taking only one noun (high-constraint classifiers), some taking many nouns (low-constraint classifiers), and other classifiers in the middle; Second, Mandarin Chinese is a tonal language where both rimes and tones are used to distinguish lexical meanings, and where there are many monosyllabic words that share the same tone or the same rime, making it simpler to test independent effects of rime and tone competitors for possible noun targets.

Below, I describe the properties of classifier-noun phrases, and then review relevant studies that have tested for the contextual effect elicited by these phrases. Finally, I will conclude with a brief overview of empirical work covered in more detail in Chapters 2 and 3.

1.3.1. Classifier-noun phrases in Mandarin Chinese

Mandarin Chinese is selected as the target language of the current study in order to test processing rime and tonal information in context because Mandarin is a tonal language that has four tones distinguishing lexical meanings, and there are many monosyllabic words differing from each other only in tone or rime, as in Table 2:

Table 2. An example of tone and rime competitor in Mandarin

Syllable	Tone competitor	Rime competitor
Ke4 课 “class”	Kuang4 矿 “mine”	Ke2 壳 “shell”

Thus, it will be feasible to test how rime and tonal information are perceived separately using these stimuli in Mandarin. Classifier phrases (ClPs) in Mandarin Chinese usually follow a number / demonstrative word to form a determiner phrase, and are the default phrasing used to refer to simple nouns, as showed in the following examples:

(4) a. yi4 tiao2 yu2 (一条鱼)
One CL fish
"a fish"

b. na4 tiao2 yu2 (那条鱼)
that CL fish
"that fish"

In a ClP, the classifier is the obligatory head, while a noun is selected as the complement. Traditionally, classifiers are used with classes of nouns based on perceptual properties that are intrinsically related to the nouns (Tai, 1992), which include semantic features. Historically, most classifiers originated from verbs or nouns that were thematically, iconically, or categorically associated. For example, the classifier "ba3" originated from the literal meaning of the verb "ba3", meaning "to hold," and this classifier is thus used to refer to objects that can be held by a hand, such as "yao4shi"(key) and "mi3"(rice). The classifier "gen1" came from the original meaning of the noun "gen1", which is "root" and thus it can be used to refer to objects that share semantic features such as [+long, +thin] with roots of some plants, such as "luo2bo"(carrot), "sheng2zi"(rope) or "yan1"(cigarette).

In Mandarin Chinese, there are some classifiers that can only take a couple nouns as their complements and some other classifiers can take a much greater number of nouns. In this case, they differ in constraining strength in a way that some classifiers constrain the selection of their legal noun complements in a more strict way than some other classifiers. Interestingly, many classifiers that can only take a couple noun complements were originally nouns as well in the history which were used to denote the nominal objects sharing exactly the same meaning with the classifiers' noun complements at a later time point. For example, the classifier "du3"(堵) originating from the noun "du3", which used to mean "wall", is used to refer to an object that is [+wall]. Usually, "qiang2" (wall) is the classifier's only plausible completion but in certain contexts, some objects that share the same semantic features with a wall, such as [+be able to block], are also weakly acceptable to native speakers. For example, in the norming task that is described in the second chapter, two participants also put "men2" (gate) and "cheng2-men2" (city gate) respectively as the second completion of "du3":

(5) a. Yi4 du3[+able to block] qiang2 ?b. Yi4 du3[+able to block] men2/cheng2-men2

One CL wall One CL gate/ city gate

"a wall" "a gate/city gate"

In addition, some classifiers of this kind also denote a specific kind of container exclusively for the object they take. For example, the classifier "feng1" which means envelope, can only take "xin4" (letter) and the subcategories of letter, such "qing2-shu1" (love letter), as its complement. In contrast, some other classifiers can refer to a broader variety of nouns that share common semantic features, which is illustrated as follows:

(6) Yi4 qun2 [+animate, +plural] ya1/ yang2/niu2/zhu1

One CL duck/ sheep/ cow/ pig

"a group of ducks/ sheep/ cows/ pigs"

While classifiers like "du3" and "feng1" usually refer to a specific noun object, the classifier "qun2" is able to take a whole category of objects as long as they have semantic features of [+animate, +plural] and without further restriction, it is able to take "people" or "animals" and any of their semantic subsets in plural form as complement. Classifiers like "du3" and "feng1" are high-constraint by nature whereas classifiers like "qun2" are low-constraint, both of which have a set of semantic requirements for the following nouns.

There are nouns that satisfy the semantic requirements of several classifiers. However, when a noun is matched with different classifiers, different semantic aspects will be emphasized. Although "gen1" and "tiao2" are used to refer to objects that are [+long, +thin], "tiao2" focuses more on the [+vertical] property. For example, they can both take "sheng2" (rope) as complement, but when "tiao2" is the classifier, the situation is more likely to be a rope hanging somewhere. Nevertheless, one cannot voluntarily select the semantic properties of an object they want to express and assign it a classifier that seems to require such properties. The match between classifiers and nouns is regulated by how the semantic properties of the prototype of an entity are perceived, which varies across different Chinese languages. For example, in Mandarin, classifier "ba3", which is used to refer to objects that can be held by hand, such as "mi3"(rice) or "sao4zhou"(broom) can select "yaoshi" (key) as its complement whereas in Cantonese, the features [+long, +thin, +vertical] of a key are concentrated on and thus "so2si3" (key) goes after classifier “tiu4” (the same classifier “条” in Mandarin, with a different pronunciation “tiao2”) instead, as showed in (7) and (8):

(7) a. Mandarin

Yi4 ba3 [+able to be held by hand] yao4shi

一把钥匙

One CL key

A key.

*b. Mandarin[1]

Yi tiao2 [+long, +thin, +vertical] yao4shi

一条钥匙

One CL key

A key.

(8) a. Cantonese

Yat1 tiu4[+long, +thin, +vertical] so2si3

一条锁匙

One CL key

A key.

*b. Cantonese

Yat1 ba2[+able to be held by hand] so2si3

一把锁匙

One CL key

A key.

Since the match between classifiers and nouns has been regulated by the consensus of native speakers in the long history of language development, there is usually a determined number of nouns that can go after a certain classifier. For example, the three classifiers below form a continuum of constraining strength, depending on how many nouns they can take, with “du3” and “qun2” at the two ends and “li4” in the middle[2]:

---

[1] It is less frequent for the classifier “tiao2” to take “yao4shi” (key) as the complement and it can be grammatical only when the [+long,+thin,+vertical] features of the key are emphasized.

[2] The examples come from the norming task conducted before the main experiment. The average numbers of the nouns that participants came up with after the three classifiers “du3”, “li4” and

(9) yi4 du3: qiang2 yi2 li4: mi3/yao4/tang2 yi4 qun2: ya1/yang2/niu2/zhu1/gou3…

One CL wall One CL rice/medicine/candy One CL duck/sheep/cow/pig/dogs…

◄────────────────────────────────►

High Medium Low

Therefore, classifiers and their noun complement are semantically associated, but the match between them is also pre-determined by the grammar, creating a way of manipulating contextual effects. The classifiers approximate the two ends can form two clear-cut types of context. If the interactive view functions as a better account for processing words in context, it will not be unreasonable to expect that high-constraint classifiers are more predictive of the upcoming nouns since there are only one or two grammatical possibilities, while low-constraint classifiers may activate more plausible candidates.

1.3.2. Semantic activation in classifier-noun structures

There have only been a few studies examining the contextual effect of classifier-noun structures, but these studies confirm that the semantic information conveyed by classifiers can facilitate the prediction of the following noun. Chou, Huang, Lee, & Lee (2014) revised the experiment in Federmeier et al. (2007) by changing the strong vs. weak constraint sentences to high- vs. low-constraint classifiers, which were followed by expected and unexpected noun complements. Significantly different processing costs were found for unexpected nouns in the two types of classifier phrases, with a greater N400 found in the high-constraint context. However, the facilitative effect was later manifested for expected nouns in a frontal positivity at 600-900ms. Although slightly different results were presented in classifier-noun agreement than in sentence condition, as in Federmeier et al. (2007), the fact that both a processing boost and cost were manifested through different ERPs has confirmed that classifier-noun phrases in Mandarin are a valid medium to demonstrate contextual effects.

“qun2” are: 1.13, 2.87 and 3.87. The numbers may be different from how many grammatical nouns can follow those classifiers in a dictionary or corpus.

1.4. Research questions

Two research questions will be investigated in the current study: First, in a semantically constraining context such as classifier-noun agreement, are tonal and segmental information processed at different levels? That is, do lexical tones receive more semantic activation than rimes, which can be demonstrated by earlier availability of the tonal information in a target word? Second, does the interactive or integrative model better account for word recognition, more precisely, tone and rime processing in semantically constraining contexts? That is, do listeners anticipate tone and rime information of specific target words in context before any of their acoustic signals become available or do they integrate the new word information with the prior context after receiving some phonological information of the word?

In the next chapter, the methods of the experiment in this study and the results obtained from the experiment will be described in detail.

Chapter 2.

Methods and Results

This chapter outlines an experimental investigation driven by the two questions raised in the previous chapter: First, during lexical access, do listeners anticipate specific target words from a prior semantic/syntactic context before any acoustic signals from the target become available, or does contextual information only accelerate the selection of target word after acoustic and phonological information about the target has already activated a series of lexical candidates? Second, in a high-constraint context, is tonal and segmental information processed at different levels: Do lexical tones receive more semantic activation than rime, which can be demonstrated by earlier accessibility of the tonal information of a target word?

In the current experiment, classifier-noun agreement in Mandarin was used as the experimental context, and a visual word paradigm was used to explore these questions through eye-gaze. The methodology used in this study is based on early findings using an eye-tracking paradigm, which found that participants' eye gaze at visual items was synchronized with concurrent auditory speech input received (Allopenna, Magnuson & Tanenhaus, 1997). The sensitivity of an eye-tracking paradigm makes the method suitable for reflecting real-time processing in naturalistic scenarios. In order to investigate contextual effects on processing tones and segments, the constraining strength of context was manipulated by grouping classifiers into two categories, a high-constraint group where the classifiers can take as few as one or two nouns, versus a low-constraint group, where the classifiers can take more nouns.

Challenges in designing the stimuli were a) to balance various criteria for categorizing the classifiers as being in a high- or low-constraint group, which should be consistent with native speakers' intuition, and b) to select target nouns for each classifier as well as corresponding phonological competitors. In Section 2.1, I will discuss the selection of the 12 experimental classifiers and the criteria of determining their high vs. low classification, as well as the selection of the targets and competitors. Section 2.2 depicts the design of experimental and filler trials for the eye tracking experiment.

Section 2.3 introduces the participants of the experiment. Section 2.4 outlines the specific procedure of the eye tracking experiment. Section 2.5 summarizes the two hypotheses proposed in the study. Finally, in Section 2.6, results will be presented, which investigate contextual effects on activating target words and their two types of phonological information in both high-constraint and low-constraint contexts.

2.1. Stimuli

In this section, the selection and design of the experimental stimuli will be introduced in two sub-sections. In the first sub-section, the selection of classifiers will be described and in the second sub-section, the procedure of selecting the target nouns and their phonological competitors will be discussed.

2.1.1. Selection of classifiers

Classifiers were selected using a two-step process. First, twenty-five classifiers were initially selected from *The Dictionary of Chinese Classifiers* (Liu, 2013) because they were perceived to be in common use, and thus recognizable to experimental participants. Candidates for high-constraint classifiers were chosen because they could only take one or two nouns, while the other classifiers could take multiple nouns in the dictionary examples, and this was verified from an analysis of the Beijing Language and Culture University Corpus Centre (Gou, E., Rao, G., Xiao, X., & Zang, J., 2016). Eleven classifiers were initially categorized as high-constraint classifiers and fourteen were low-constraint ones (See Appendix A for details). Second, in order to verify that corpus results aligned with how native speakers process classifier-noun agreement in a naturalistic scenario, a norming task was conducted.

Fifteen native mainland Mandarin speakers aged 19 to 28 (mean = 23.53, *SD* = 2.03), were recruited from Simon Fraser University. All acquired mainland Mandarin since birth and had no significant exposure to other Chinese languages. The norming task was completed online, and it required participants to complete two sections, the first of which involved rating their familiarity with the twenty-five classifiers based on how often they heard or used the classifiers in their daily life. The scale of rating ranged from 1 to 5, as presented in Table 3:

Table 3. Scale of familiarity rating

Scale	Interpretation
1	I never hear or use this classifier in my daily life
2	I seldom hear or use this classifier in my daily life
3	I sometimes hear or use classifier in my daily life
4	I often hear or use classifier in my daily life
5	I hear or use this classifier all the time on a daily basis

The purpose of this first task was to ensure that all the classifiers were more commonly used in oral language instead of literary language.

The purpose of the second task was to measure the constraining strength of the classifiers by evaluating how many head nouns a classifier could activate. Here, I asked participants to come up with as many nouns as possible (up to 7) that could grammatically follow each classifier in the commonly heard template of "yi + classifier + noun", where "yi" (one) acts as an indefinite article. Participants were also instructed to produce the bare form of each noun they generated, since results could be biased if participants added adjectival morphemes to their noun responses, which may have affected the link between classifier and head noun. For example, if they put words such as "hua1" (flower), "hong2-hua1" (red flower), "bai2-hua1"(white flower) as three different noun targets of the classifier "duo3", I would have overestimated how many different entities can be activated, since the head noun would refer, in these cases, to the same object, "hua1" (flower).

The results of the familiarity rating task are shown in Appendix B. Since all experimental stimuli were intended to be familiar to most participants in their daily lives, any classifiers that had an average scored less than 2 were eliminated (i.e., one classifier, "zhong1" [盅], was excluded for this reason). One additional classifier, i.e. "zhan4" (站) was excluded, because the phrasal structure "yi2 zhan4" is ambiguous between a verb complement and the head of a ClP (classifier phrase). For example, some participants' answer for "yi2 zhan4" was "che1" (car), but this particular prompt does not necessarily function as a classifier modifying "che1". Rather, the participants may have understood the structure as if there was an omitted verb preceding "yi2 zhan4," and thus this classifier may have functioned as the complement of the verb.

The constraining score from the second task was generated for each remaining classifier by calculating the number of target nouns averaged across subjects. The lower a constraining score was, the more constraining a classifier was. Two nouns were not counted separately if they satisfied all of the following standards: If they share the same head morpheme; if one noun denoted an entity that was a semantic subcategory of another noun; and if no semantic borrowing, broadening, narrowing or metaphorical transfer could be detected between the head morpheme of the two nouns. Consider the following Example (10):

(10) a. deng1 "light" 灯

b. tai2-deng1 table-light "lamp" 台灯

c. lu4-deng1 road-light "streetlight" 路灯

The three nouns were counted as one noun "deng1" because: they share the same head morpheme "deng1" (light); "deng1" is the generic name of all the electric devices that are used for lightening while "tai2-deng1" and "lu4-deng1" are subcategories; the morpheme "deng1" in the three words all denotes the same electric device without semantic transformation. In contrast, I counted "mei2-gui1 (rose) and hua1 (flower) as different nominal categories: Although "mei2-gui1" is a subcategory of "hua1", they do not share the head morpheme.

The results of the second task are presented in Appendix C (mean rating = 3.18, range = 1.13 – 5.60). Based on a mean split[3], the classifiers that had scores higher than 3.18 were categorized as low-constraint classifiers and the ones scoring lower than that were defined as high-constraint ones. The selection of the final high-constraint stimuli started by choosing the classifier with the lowest score within all the 11 classifiers that had scores lower than 3.18, and then choosing the one with the second lowest score (et cetera) until a total of 6 high-constraint classifiers were all chosen. Likewise, for low-constraint stimuli, the selection started from choosing the one with the highest score

[3] Median split is not adopted here because the median is 3.73, which will result in 10 members in the low-constraint group. However, only five of them are qualified candidates based the standard of selection. Mean split categorizes 11 classifiers to be low-constraint and 6 of them are eligible items, which is the number needed for the experiment. Besides, mean split divides the classifiers into 12 for high-constraint group and 11for low-constraint group, which is not a big difference of number.

within all the 12 classifiers that had higher scores than 3.18 until another 6 were selected. Because some nouns take several classifiers, some exceptions to this procedure were made, which are described further in Appendix C. The final 12 classifiers selected for this study and their constraining scores are presented in Table 4:

Table 4. Constraint strength of stimulus classifiers

Classifier	Constraint type	Constraint score
(yi4) du3 (一)堵	High	1.13
(yi4) tang2 (一)堂	High	1.33
(yi4) duo3 (一)朵	High	1.67
(yi4) zhan3 (一)盏	High	1.87
(yi4) feng1 (一)封	High	2.13
(yi4) lü3 (一)缕	High	2.47
(yi4) tou2 (一)头	Low	4.13
(yi4) di1 (一)滴	Low	4.07
(yi4) qun2 (一)群	Low	3.87
(yi4) gen1 (一)根	Low	3.73
(yi4) tiao2 (一)条	Low	3.73
(yi4) shuang1 (一)双	Low	3.73

2.1.2. Selection of the target nouns and competitors

The target noun for each classifier was selected based on 4 criteria: First, it should be the best completion of the classifier; Second, it should be concrete or image-able; Third, it should have a tone competitor (sharing the same tone as the target) and rime competitor (sharing the same rime as the target) that are also concrete or image-able; Fourth, it should be monosyllabic unless the head morpheme of a disyllabic word could not have grammatically stood alone after the classifier.

The “best completion” was defined as the noun with the highest cloze probability following the classifier, which was calculated with the following procedure: the first and second completions for each classifier provided by each participant were marked down, and then the probability of a certain noun being the first or second completion for a certain classifier was calculated by averaging its occurrences across all the participants, and those probabilities were then summed for each noun. After the best completion of each classifier was decided, the remaining nouns competed with each other to be the next best completion based on their overall probability (i.e., the sum) of being the first and the second completion (See Appendix D for the details about these calculations of best and next best completions of each classifier).

For high-constraint classifiers, if the best completion of a classifier did not satisfy any of the other three criteria, that classifier was rejected altogether, and the next most-constraint classifier and its best completion (or its next best completion) would be scrutinized for selection. For low-constraint classifiers, if the best completion did not satisfy the other three criteria, the next best completion would first be selected, and if that did not work, only then would a new classifier be chosen. The reason for having this discrepancy in noun selection procedures was that high-constraint classifiers usually had a much stronger prediction probability for their best completion than for their next best completion, while low-constraint classifiers are more equally predictive of their possible continuations.

As an example, consider the high-constraint classifier “lü3”. The three completions of it, “yan1” (smoke), “yang2-guang1” (sunshine) and “tou2-fa” (hair) have an equal probability of being the best completion. In this case, “yan1” outweighs the other two because it is a monosyllabic word. Consider another example for the low-constraint classifier “gen1”, where the noun “sheng2” (rope) was chosen as the target because the better-rated completion “cong1” (green onion) did not have any tone competitors. The noun selected as the target for each of the 12 classifiers and its cloze probability as the best or the next best completion are presented in Table 5:

Table 5. Target nouns and their cloze probabilities

Classifier	Constraining type	Noun	Completion type	Cloze probability
(yi4) du3 (一)堵	High	Qiang2 墙 “wall”	Best	1
(yi4) tang2 (一)堂	High	Ke4 课 “class”	Best	0.93
(yi4) duo4 (一)朵	High	Hua1 花 “flower”	Best	0.8
(yi4) zhan3 (一)盏	High	Deng1 灯 “light”	Best	1
(yi4) feng1 (一)封	High	Xin4 信 “letter”	Best	0.93
(yi4) lü3 (一)缕	High	Yan1 烟“smoke”	Best	0.27
(yi4) tou2 (一)头	Low	Zhu1 猪 “pig”	Next best	0.3
(yi4) di1 (一)滴	Low	Lei4 泪 “tear”	Next best	0.3
(yi4) qun2 (一)群	Low	Ya1 鸭 “duck”	Third best	0.13
(yi4) gen1 (一)根	Low	Sheng2 绳“rope”	Next best	0.2
(yi4) tiao2 (一)条	Low	Yu2 鱼 “fish”	Best	0.2
(yi4) shuang1 (一)双	Low	Xie2 鞋 “shoe”	Best	0.4

The design of tone and rime competitors for the target noun was inspired by Malins & Joanisse (2010), where the main comparison was between the target and either a cohort-competitor (differing only in rime while sharing the same onset and tone) or a segment-competitor (differing only in tone while sharing the same onset and rime). In their study, it was found that these two types of competitors diverged from the target item at a similar time point and their trajectories were identical to each other, indicating an equal perceptual salience of tone and rime information. In the current study, since the focus is also the comparison between tone and rime information, the same types of competitors (cohort and segmental competitors) will be used. Note that the onsets of these competitors are kept the same as that of the target because participants may be able to differentiate the competitors from the targets before they have any contact with the rime or tone if they do not share the same onset. In fact, Malins & Joanisse (2010) found that competitors that only shared the same rime with the target word (their onsets were different) did not have significantly more influence on the eye gaze to the target than the phonologically unrelated distractors, indicating some unique characteristics of Mandarin that are different from English (e.g. onsets might play a more important role in constraining word recognition in Mandarin). Here I rename cohort and segmental competitors in Malins & Joanisse (2010) as ‘tone’ and ‘rime’ competitors, because the tone information is the same as the target in the first case (while the syllable—specifically, the rime—was different), and the rime was the same as target in the second case (while the tone was different). Three principles were considered for the selection:

Competitors also had to be concrete and image-able monosyllabic words; The average lexical frequency of targets, competitors and distractors should have been relatively balanced with each other (across all items), particularly between tone and rime competitors, so that any difference of fixation could not be driven by lexical frequency, but rather by phonological factors. Specifically, lexical frequencies per million words were taken from the *Modern Chinese Frequency Dictionary* (1986), and were converted to logarithmic values for further comparison, in order to ensure differences between two words that reflect human perception of lexical frequency.

To verify that the final stimuli set had met this constraint, log-transformed lexical frequencies between different object types (target, competitor or distractor) and context types (preceded by high-constraint or low-constraint classifiers) were shown to not significantly differ from each other in the two stimuli sets: First, where a target was paired with a tone competitor, and second, where a target was paired with a rime competitor. This was verified with a pair of two-way ANOVAs. In the ANOVA testing the rime competitor stimulus set, there were no main effects of object type, $F(2, 45) = .25$, $p = .78$, or context type, $F(1, 46) = 1.43$, $p = .24$, and no interaction, $F(2, 45) = 1.63$, $p = .21$. In the ANOVA testing the tone competitor stimulus set, again there were no effects of either object type, $F(2, 45) = .04$, $p = .96$, or context type, $F(1, 46) = .11$, $p = .75$, and no interaction, $F(2, 45) = .92$, $p = .41$. Some examples of target nouns and their tone and rime competitors along with their lexical frequency were illustrated as follows (See Appendix E for the full version):

Table 6. Examples of target nouns and their phonological competitors

Target	Tone competitor (LF)	Rime competitor (LF)
Qiang2 墙 “wall”	Qian2 钱 “money” (2.77)	Qiang1 枪 “gun” (2.31)
Ke4 课 “class”	Kuang4 矿“mine” (2.08)	Ke2 壳“shell” (1.61)
Zhu1 猪 “pig”	Zhen1 针 “needle” (1.81)	Zhu2 竹 “bamboo” (1.71)
Sheng2 绳“rope”	Shi2 石 “rock” (2.24)	Sheng3 省“province” (2.31)

2.2. Design of Trials

Twenty-four unique experimental trials were constructed for the eye tracking experiment, among which each trial consisted of an auditory stimulus and four visual items. A template for the auditory stimulus was selected to elicit the 12 classifier-noun phrases, with an example below:

(11) na4 du3 QIANG2
That CL wall
“That wall”.

Although “yi + CL” was the structure used to evaluate the classifiers, the demonstrative article “na4” replaced the numeral “yi” in the auditory recording to ensure uniformity, as the article “yi” undergoes tone sandhi with certain classifiers, which could provide information about classifiers even during the article. Auditory stimuli were recorded at 44.1 kHz, produced by the author, a trained linguist who was also a native Mandarin speaker. I produced each classifier-noun phrase five times in total using a natural pace, among which the best token acoustic token was selected (no disfluencies, using similar intonational prosody, etc.). Noise reduction and normalization was implemented on all the recordings using the corresponding functions in Audacity. Additionally, silence was added at the beginning of each sound file to ensure the sound made by mouse clicking at the beginning of a trial would not cover the onset of an audio stimulus. Since the duration of the demonstrative “na4” varies across different trials, the duration of silence was also designed to vary accordingly so that the duration of “na4” with its following classifier stayed the same (618 ms) across different trials and thus, the onsets of the classifiers all started at the same time point within each trial.

The four visual items were composed of a target item, a phonological competitor, which shared either the same tone (tone competitor) or rime (rime competitor) with the target, and two phonologically unrelated distractors, which do not overlap with the target in either tone or syllable (either onset or rime). Only the target among the four items could grammatically follow the classifier from the auditory recording. Visual stimuli were clip-art pictures that came from a database of images developed at the University of Iowa (the Audio-Visual Categorization Lab) and also available at the Phonological Processing Lab at SFU. Images were selected and edited in such a way as to choose, typical-looking images that were also culturally appropriate for Mandarin speakers. If an image could not be found in the database, candidates from www.clipart.com were edited and saved in the database for future use. All the images were re-sized to 300 *300 pixels and were edited to be consistent in color and style. Each target noun was paired with a tone competitor in one trial and a rime competitor in another trial, which means that the 12 target nouns were repeated twice in all 24 experimental trials: Half of the trials had tone competitors while the other half had vowel competitors. The distractors in any one trial consisted of the targets or competitors from other trials in order to have a closed stimulus set, as the examples in Table 7:

Table 7. Examples of experimental trials

	Target	Competitor	Distractor A	Distractor B
Trial 1	Qiang2 "wall"	Qiang1 "gun"	**Zhu1 "pig"**	**Yu3 "rain"**
Trial 2	**Zhu1 "pig"**	Zhu2 "bamboo"	Qiang2 "wall"	Xie4 "crab"
Trial 3	Yu2 "fish"	**Yu3 "rain"**	Sheng3 "province"	Deng1 "lamp"
Trial 4	Qiang2 "wall"	Qian2 "money"	**Yu3 "rain"**	**Zhu1 "pig"**
Trial 5	Deng1 "lamp"	Dao1 "knife"	**Yu2 "fish"**	Shi2 "stone"
Trial 6	**Yu2 "fish"**	You2 "oil"	Zhen1 "needle"	Kuang4 "mine"

In addition, 48 filler trials were designed to balance the number of occurrences that any particular item would be heard or seen. Half of the filler trials were composed of the same visual items as the experiment trials, except that the original competitors in the experimental trials were be used as auditory stimuli in these filler trials (and thus function as the targets for those fillers). The other half of filler trials consisted of only the tone and rime competitors from the experimental trials, such that each of the four items in one experimental trial were named across all trials (including fillers). Consequentially, each object in the experiment was heard as an auditory target twice and was seen as a visual image 8 times throughout a whole block. Five blocks of these 72 total trials were presented, resulting in a total of 360 trials during the whole experiment. The trials were randomized for each participant to disguise the objective of the experiment.

2.3. Participants

Twenty-four native mainland Mandarin speakers, aged from 19 to 28 (*mean* = 23.42, *SD* = 2.04), were recruited from Simon Fraser University, and filled out a demographic questionnaire asking their age, place of birth, age of arrival in Canada, all the languages and dialects they acquired, years of learning these languages, and their own perception of dominance of these languages, and musical experience. Speakers who had not acquired Mandarin in Mainland China since birth, or speakers of a different variety of Mandarin, such as Singaporean or Taiwanese Mandarin, were not recruited because the usage of classifier-noun agreement in an overseas context or in an alternative variety of Mandarin may be very different. Bilingual speakers who had acquired another language in addition to Mandarin before the age of five were also excluded in the current study. In addition, all the eligible participants could not have had a southern Chinese dialect such as Cantonese, Shanghainese or Hakka as their dominant language: Southern dialects are typologically more distant from standard Mandarin and there may have been different matches between classifiers and nouns. However, due to demographic factors, it is inevitable that some people in my sampled population acquired a local dialect from their family in addition to standard Mandarin, and it would be difficult to only recruit participants from a single area of China. Thus, a participant who had also acquired a southern dialect was not eliminated as long as the participant rated Mandarin to be more dominant because of more frequent use of Mandarin in daily life (and, indeed, only one participant reported that she had acquired Shanghainese at the age of two).

All participants had normal or corrected-to-normal vision. None of them had a hearing impairment or auditory-related medical issue and nor did they have any significant learning or language-related pathologies such as dyslexia. Participants received $10 or 2 course credits for the one-hour experiment.

2.4. Procedure

The participants were requested to complete the demographic questionnaire and to read through a consent form at the beginning of the study. After their consent was received, the experiment began and lasted about forty-five minutes. The whole experiment consisted of two sessions. During the first session, a naming task was conducted where participants saw all the images that would later serve as visual stimuli in the eye-tracking experiment. They were requested to say aloud the object using standard Mandarin in clear voice. Each image was accompanied by a monosyllabic word in Chinese character, which ensured that only the target noun would be named by the participants. The task was conducted twice to ensure participants could name each target object accurately. If the participants accidentally mispronounced any words, the experimenter would correct them and ask them to say the word again.

A visual world paradigm was conducted in the second session. First, each participant was calibrated in an Eyelink 1000 eye-tracker in binocular mode (SR Research) using a standard nine-point procedure on a 1280 * 1024 pixel screen. Afterwards, instructions were given both on the screen and verbally, and these instructions indicated that the participants needed to click on the center red dot when it turned blue to receive an auditory stimulus, and then click on the picture matching the auditory recording. Each trial began with four pictures at the four corners of the screen, whose outer edges lay 50 pixels from the edge of the screen, and a red dot (60 pixels in diameter) in the center of the screen (See Figure 1).

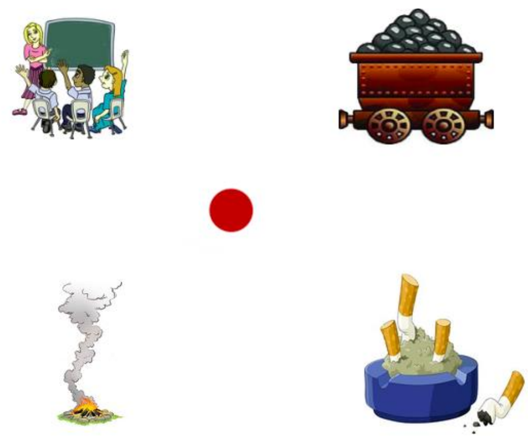

Figure 1. **A sample display screen containing a target item (ke4, "class"), a tone competitor (kuang4, "mine"), and two phonologically unrelated distractors (yan1 "smoke" and hui1 "ash"), located around a central fixation point.**

The positions of the four pictures were completely randomized so that the participants could not predict the location of their appearance. The circle turned blue after 500 milliseconds and the audio stimulus was played after the blue dot appeared. The purpose of this initial clicking was to avoid participants' fixation at any of the pictures at the beginning of the auditory stimulus. The participants were allowed to take a break every thirty trials and a drift correct was conducted at the end of each break to validate the eye track. Fixations were recorded from the onset of the display of a new trial until the participants made a response by clicking on one of the pictures.

2.5. Hypotheses

Two hypotheses were proposed in the current study: First, if the participants could anticipate the target nouns when hearing the classifiers, both semantic and phonological information from targets should be activated (which could also include information for phonological competitors). This could manifest as greater average looking time to the targets and to the phonological competitors relative to the two distractors during this early time window. Additionally, this effect could be significantly more prominent in high-constraint contexts compared to low-constraint contexts.

Second, if lexical tones receive more activation than rimes in high-constraint contexts, similar to what has been found in previous studies (Ye & Connine, 1999; Liu & Samuel, 2007), then there should be an interaction between competitor type and context. That is, there should be more distraction from the tone competitor (sharing the same tone with the target) than the rime competitor (sharing the same rime with the target) in high-constraint contexts compared to low-constraint ones, which might manifest with a significantly greater average looking time to the tone competitor than the rime competitor during the time window of the classifiers, or perhaps also in measures capturing more downstream processes, like first fixation to the target. In a low-constraint context, tone and rime information may either be equally predictive, or vowel information may be more predictive than tone information. These predictions were further pre-registered on aspredicted.org, a platform created by Penn Wharton Credibility Lab https://credlab.wharton.upenn.edu. Table 8 shows a summary of the pre-registered analyses, what hypotheses they tested and their results:

Table 8 Summary of hypotheses and corresponding results

Name of Analysis	Hypothesis	Result
Average looking time to objects	The target could be predicted by the context, perhaps also phonological competitors too	Target and competitor received more fixations than distractors in both contexts and even more to the target in high-constraint context
Average looking time to competitors	Tone competitors should be more activated than rime competitors	Tone competitors were more fixated at than rime competitors only in high-constraint context
First fixation to target	Tone competitors should be more distracting than rime competitors	Fixations to the target in both competitor conditions were not significantly different across both context types

2.6. Results

In this section, the results from the three pre-registered analyses are presented. The first sub-section outlines the results of participants' average looking time to the targets, the phonological competitors and the baseline during the time window of the classifiers. The second sub-section describes the results of average looking time to the tone competitors and the rime competitors. The last sub-section summarizes the results of latency to fixate the target in tone and rime competitor conditions.

2.6.1. Analysis One: Average looking time to objects

In order to answer the first research question, average looking time to the three types of objects, i.e. the targets, the phonological competitors and the baseline (i.e., averaged looking time to both distractors), were measured by averaging the fixations across trials during the time window of the preceding classifiers in the two types of contexts (i.e. high- and low-constraint contexts). The following figures present the curves of average fixations to different objects over time (Figure 2a and 3a) and specifically during the average classifier time window (Figure 2b and 3b), which illustrates differences in real-time processing in high-constraint context (Figure 2a and 2b) and low-constraint context (Figure 3a and 3b):

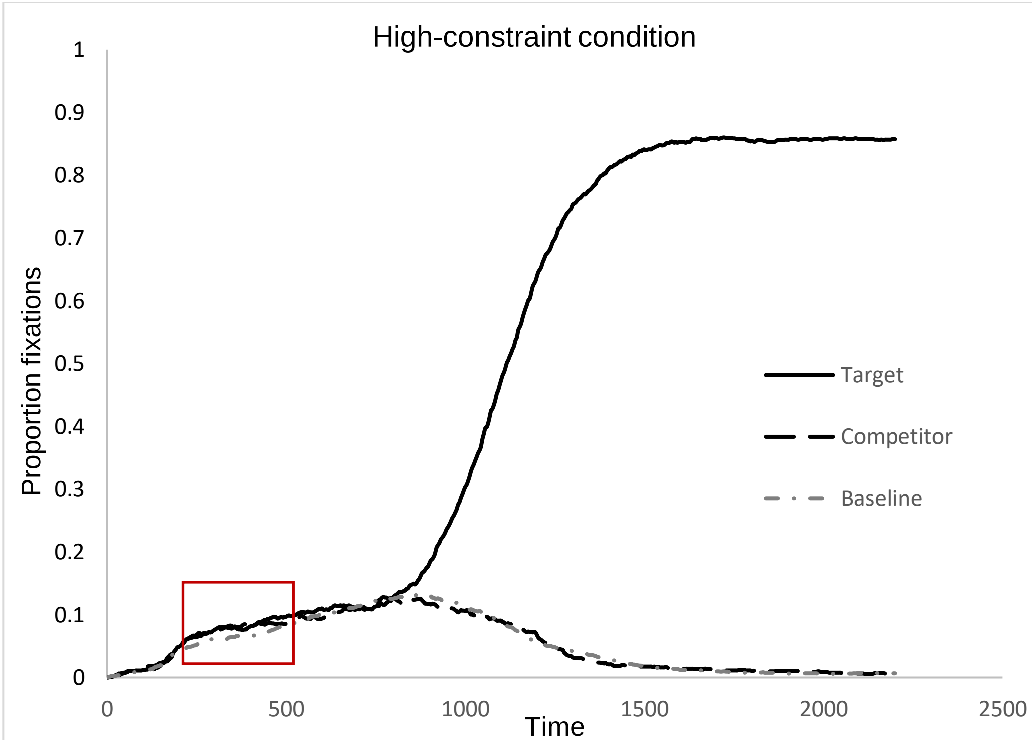

Figure 2a. **Fixations by object type in high-constraint context. The red box roughly equals to classifier time window, which is from 200ms to 536ms (averaged across classifiers).**

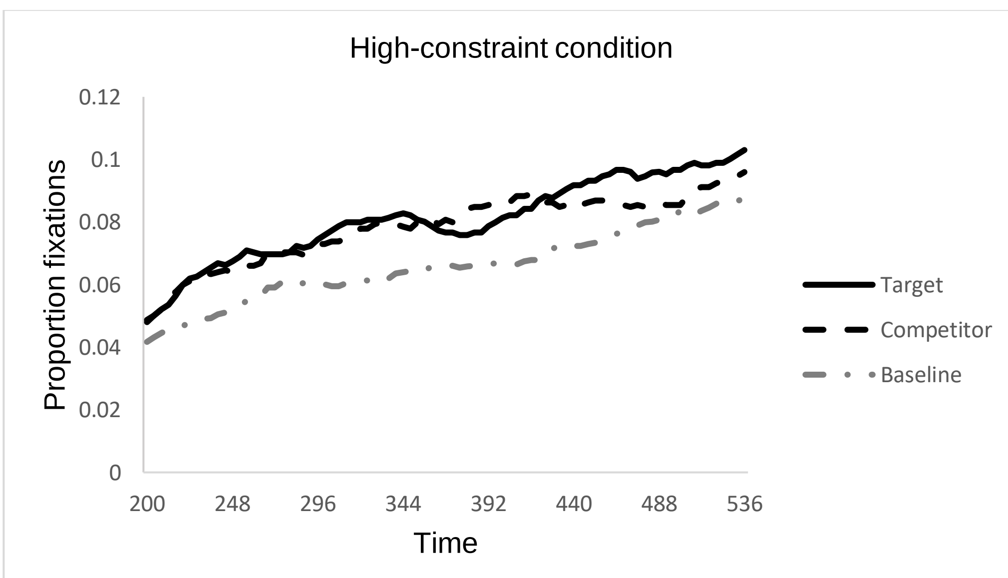

Figure 2b. **Fixations by object time in high-constraint context during the classifier time window.**

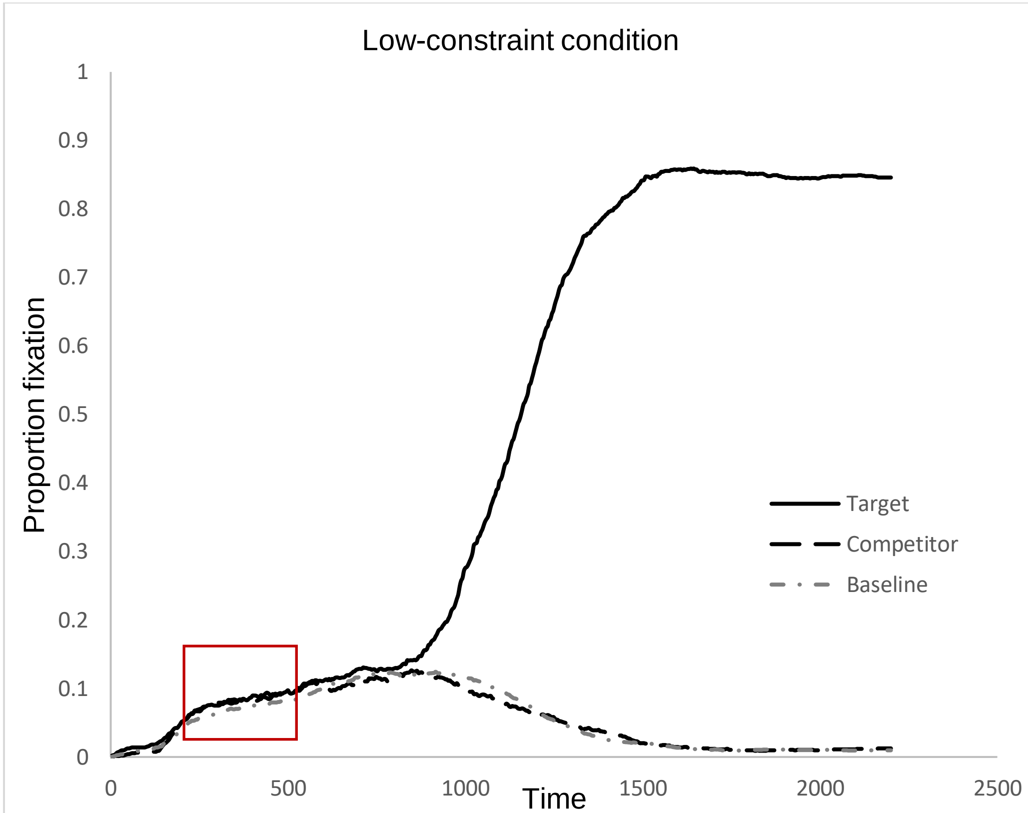

Figure 3a. **Fixations by object type in low-constraint context. The red box roughly equals to classifier time window, which is from 200ms to 536ms (averaged across classifiers).**

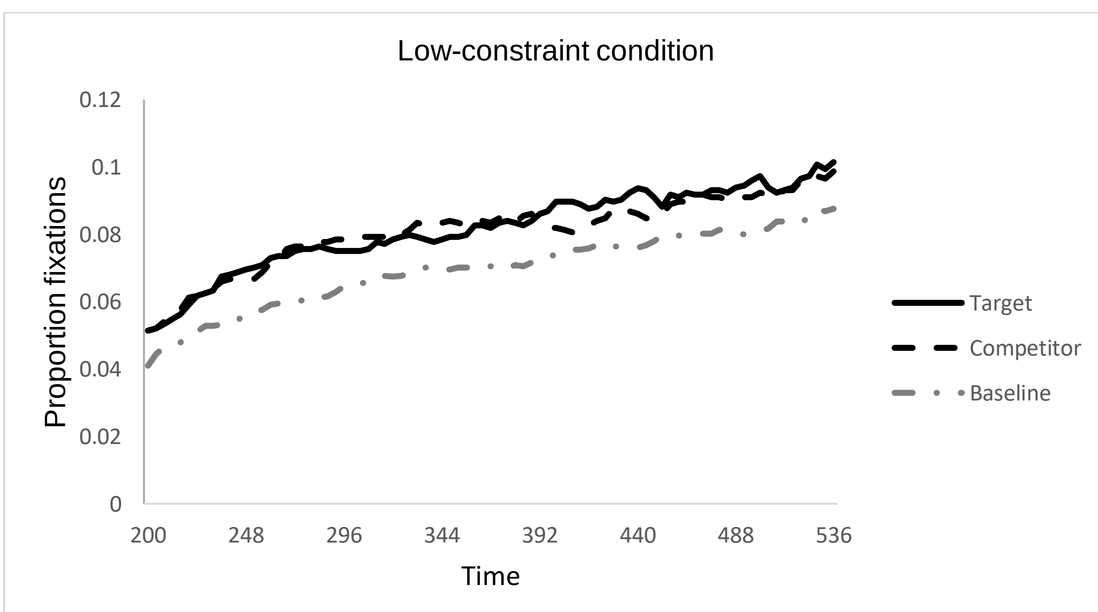

Figure 3b. **Fixations by object time in low-constraint context during the classifier time window.**

In the final analysis, however, each classifier has an individual time window of interest, depending on its duration, which starts from 200ms after the auditory onset of the classifier because it usually takes 200ms to plan and launch an eye movement (Viviani, 1990) and ends at time points ranging from 440ms to 656ms (end point of each classifier plus the delayed 200ms). Trial-by-trial results were analyzed with a linear mixed effects regression model where context type (whether it is a high-constraint or low-constraint context) and object type (whether it is a target, competitor or baseline) were fixed factors, with subject and item (which classifier) as random factors. Following the approach advocated by Baayen, Davidson, & Bates (2008), I started with a maximal model, which included the main effects of the two fixed factors and their interaction, as well as the maximal random effects structure. Following the syntax of the lme4 package in R https://cran.r-project.org/web/packages/lme4/citation.html, the original maximal model was shown in (12):

(12) AvgLook ~ Context * Object + (1 + Context * Object | Subject) + (1 + Object | Item)

This maximal model did not converge, and so the random effects structure (but not the fixed effects structure) was simplified in the following ways suggested in Kuznetsova, Brockhoff, & Christensen (2017) until the model converged. First, random slopes and intercepts were uncorrelated. Second, I removed the highest order random slope (i.e., the interaction first), or else the random slope that contributed the smallest explanation of variance, and then correlated the slopes and intercept again. Third, I iterated through the first two steps (de-correlating and then removing another random slope) until the model converged, which was the model, listed below Table 9, including the interaction between the two fixed factors and the two random intercepts.

The ANOVA method from the lmerTest package (Kuznetsova et al., 2017) was applied to this reduced model. Results were presented in Table 9. The denominator degrees of freedom (Den.df) for F statistics and p values are calculated using Satterthwaite's method of approximation (Satterthwaite, 1941). There was a significant overall effect of object type ($p < .0001$) and a significant interaction between context and object type ($p = .0001$), but no significant effect of context type ($p = .27$).

Table 9. Test of fixed effects of context and object type on average fixations

	Sum Sq	Mean Sq	Num. df	Den. df	*F*	*p*
Context	.01	.01	1	11	1.33	.27
Object	2.75	1.38	2	73261	173.26	<.0001
Context * Object	.14	.07	2.73	73261	8.90	.0001

Note*:* These are results from a model specified as following, using lme4 syntax in R: AvgLook ~ Context * Object + (1 | Subject) + (1 | Item)

Follow-up analysis using the difflsmean() function in lmerTest package was conducted to calculate differences of least square means for the fixed factors in the linear mixed effects model constructed previously, using Satterthwaite's method of approximation (Satterthwaite, 1941). The results in Table 10 show that there was a significant difference between target object and baseline in both high and low-constraint conditions, with larger average looking time fixating at the target object than the baseline ($p < .0001$). Likewise, there was significant greater average looking time to the competitors than the distractors in both context types ($p < .0001$). However, the difference between the looking time to the target and the competitor was only significant in the high-constraint condition ($p < .0001$), and not in the low-constraint condition ($p = .32$).

Table 10. Post-hoc tests comparing objects in different contexts

	Contrast	EST	*SE*	*df*	*t*	*p*
High	Competitor VS. Baseline	.01	.001	73261	9.41	<.0001*
	Competitor VS. Target	-.006	.001	73261	-5.35	<.0001*
	Baseline VS. Target	-.02	.001	73261	-14.76	<.0001*
Low	Competitor VS. Baseline	.01	.001	73261	9.42	<.0001*
	Competitor VS. Target	-.001	.001	73261	-.99	.32
	Baseline VS. Target	-.01	.001	73261	-10.41	<.0001*

2.6.2. Analysis Two: Average looking time to competitors

In order to test the second hypothesis, average looking time to the two types of competitors, i.e. the tone competitor and rime competitor, were measured by averaging the fixations across trials during the time window of the preceding classifiers in the two types of contexts. Figure 4a, 4b, 5a and 5b present the curves of average fixations to different competitors as time grew to demonstrate differences in real-time processing in high-constraint context (Figure 4a and 4b) and low-constraint context (Figure 5a and 5b):

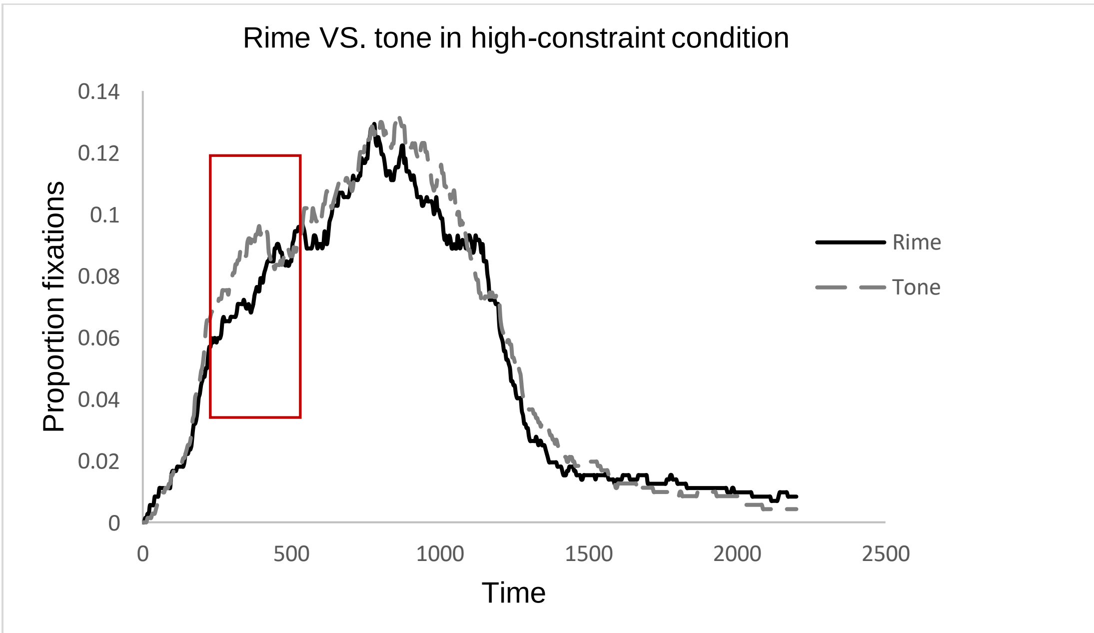

Figure 4a. **Competitor fixations by competitor type in high-constraint context. The red box roughly equals to classifier time window, which is from 200ms to 536ms (averaged across classifiers)**

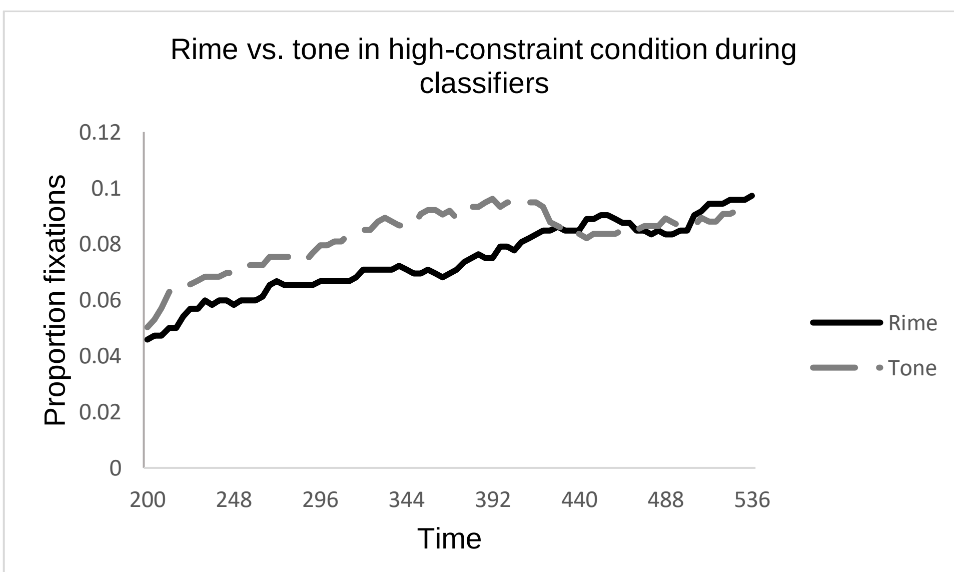

Figure 4b. **Competitor fixations by competitor type in high-constraint context during classifier time window.**

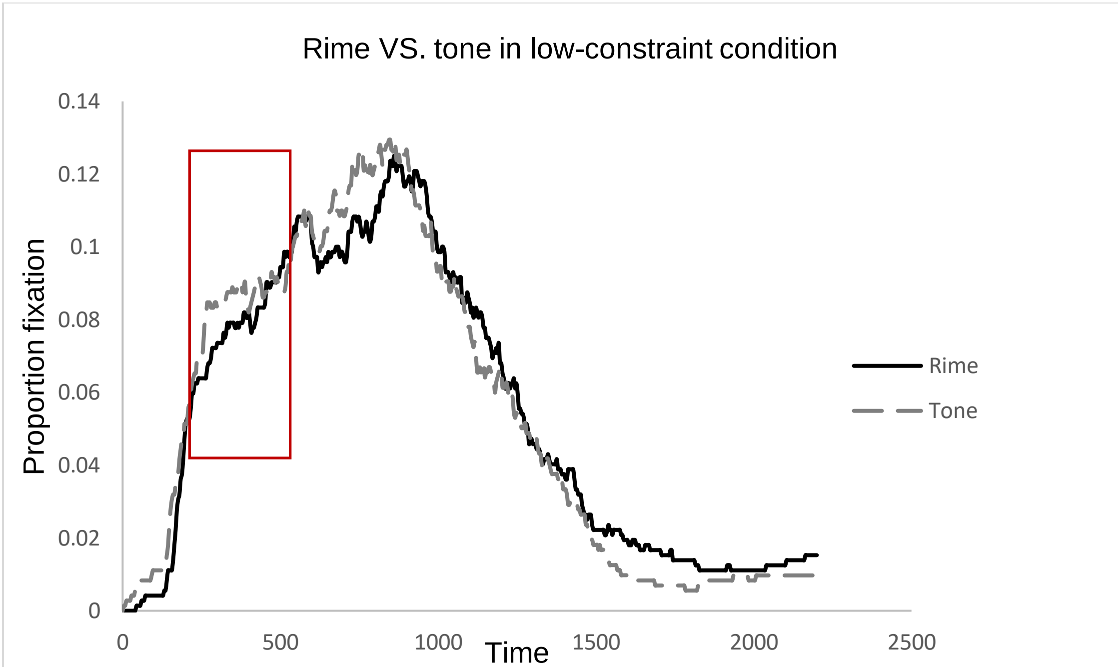

Figure 5a. **Competitor fixations by competitor type in low-constraint context. The red box roughly equals to classifier time window, which is from 200ms to 536ms (averaged across classifiers).**

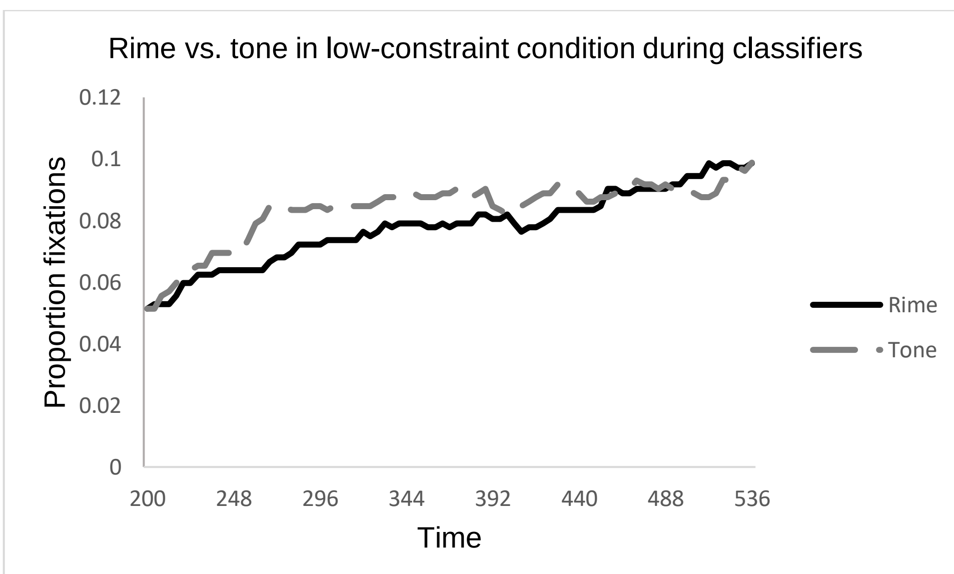

Figure 5b. **Competitor fixations by competitor type in high-constraint context during classifier time window.**

Trial-by-trial results were analyzed with a linear mixed effects regression model. Following the approach taken for the first analysis, the original maximal model was shown in (13):

(13) AvgLook.lmer = AvgLook ~ Context * Competitor + (1 + Context * Competitor | Subject) + (1 + Competitor | Item)

Again, this maximal model did not converge, and so the random effects structure was simplified in the same way as before, using identical methods to calculate effect significance. Results are presented in Table 11, and reveal that, again, there was a significant overall effect of competitor type but no significant effect of context type, but there was a significant interaction between these two factors.

Table 11. Test of fixed effects of context and competitor type on average fixations

	Sum Sq	Mean Sq	Num. df	Den. df	F	*p*
Context	.004	.004	1	23	.25	.62
Competitor	.33	.33	1	48854	21.17	<.0001
Context * Competitor	.06	.06	1	48854	4.09	.04

Note*:* These are results from a model specified as following, using lme4 syntax in R: AvgLook.lmer = AvgLook ~ Context * Competitor + (1 + Context | Subject) + (1 | Item)

Again, follow-up analysis using difflsmean() function in LmerTest package was conducted. As in Table 9, the post-hoc test shows that there was a significantly difference between tone and vowel competitor only in the high-constraint context (*p* < .0001), with larger average looking time fixating the tone competitor than the rime competitor. However, this effect was not found in low-constraint context (*p* = .06).

Table 12. Post-hoc tests comparing competitors in different contexts

	Contrast	EST	*SE*	*df*	*t*	*p*
High	Tone VS. Rime	.008	.002	48854	4.48	<.0001
Low	Tone VS. Rime	.003	.002	48854	1.92	.06

2.6.3. Analysis Three: Target latency in competitor conditions

A supplementary analysis compared another measure of looking time, which was essentially a measure of how distracting tone or rime competitors were in the two types of constraint contexts. Specifically, I analyzed the latency of participants' first fixations to the target, which measures how quickly the participant shifted their eye gaze from a certain competitor to the target after first hearing the critical stimulus. One of the major differences from the average looking time analyses above is that the present latency analysis is not strictly locked within the classifier window. As Table 13 shows, during this period, less than 10 % of the total fixations were made to the target (no matter whether this was the first, second fixation, etc.), and so analysis of first fixations to the target could not necessarily have included a lot of information made during the same classifier time window:

Table 13. Mean fixation by object type in each context condition

Context	Object	Proportion of fixation
High	Competitor	.079
	Baseline	.066
	Target	.081
Low	Competitor	.081
	Baseline	.070
	Target	.082

The latency of the first target fixation was measured from the onset of the classifier, but any trials where participants were already fixating the target at 200ms after the classifier onset were excluded (since this analysis, by default, excludes trials where participants happen to have begun fixating the target before auditory information could possibly inform them of the target). This results in an exclusion of 7.8% of the trials. Trials where participants directly fixated the target without initially fixating the competitor were also eliminated, since these cases of fixation may have meant that participants never had the opportunity to be distracted by the phonological competitors at all. Such action further yielded elimination of 16.9% of the trials.

Trial-by-trial first fixations were analyzed using linear mixed effects models on the remaining trials. Following the approach used in the first two analyses, the original maximal model is shown in (14):

(14) FirstFixation.lmer = FirstFixation ~ Context * Competitor + (1 + Context * Competitor | Subject) + (1 + Competitor | Item)

Again, this maximal model did not converge, and results of the reduced model are presented in Table 11 (17), showing that there were no significant effects of either context type (*p*=.19) or competitor type (*p*=.87), as well as no interaction between the two fixed factors (*p*=.76).

Table 14. Test of fixed effects of context and competitor type on first fixation

	Sum Sq	Mean Sq	Num. df	Den. df	F	*p*
Context	122611	122611	1	12.19	1.94	.19
Competitor	1741	1741	1	38.32	.03	.87
Context * Competitor	5788	5788	1	57.63	.09	.76

Note*:* These are results from a model specified as following, using lme4 syntax in R: FirstFixation.lmer = FirstFixation ~ Context * Competitor + (1 + Context + Competitor | Subject) + (1 + Competitor | Item)

The mean time when the first fixation was made in each condition is shown in Table 15:

Table 15. Mean time of first fixation in each competitor condition in different contexts

Context	Competitor	First Fixation (ms)
High	Rime	1102
	Tone	1096
Low	Rime	1159
	Tone	1142

The mean first fixations in all conditions were made after the classifier time window which starts from 200ms and ends at an average time point of 536ms (end points range from 440ms to 656ms). This suggests that on average, first fixations across all the participants were made when some acoustic information of the noun target was presented.

2.6.4. Interim Summary

This study found three key results. First, there are significant differences between the average looking time to the target, competitor, and to the distractors (across both context types) within the time window of the classifier. This suggests that the target and its phonological competitors were pre-activated before any acoustic exposure to the target. When comparing looking time to the competitor versus the target, however, there was no difference in low-constraint contexts, but a significant difference in high-constraint contexts, which indicates a contextual effect whereby target items were more fixated to phonological competitors in a high-constraint context than in a low-constraint context.

Second, I analyzed the role of different types of competitors by measuring average looking time during the same time window (during just the classifier). Here, there were significant differences between the average looking time to the tone competitor and to the rime competitor only in the high-constraint context. Such result suggests that there was a greater activation of tone information than rime information when the context is highly predictive.

Finally, in terms of target latency, there are no statistically significant differences between the time points of the first fixation to the target under the conditions where there was distraction of the two competitors. However, the mean time when the first fixations were made in the four conditions were all after the classifier window when some acoustic signals of the target were presented, making it different from the average looking time analyses. This surprising result will be further discussed in the next chapter.

Chapter 3.

Discussion and Conclusions

In this chapter, I will elaborate on how the results from Chapter 2 answer the research questions I proposed in Chapter 1. In Section 3.1, I will discuss fixation analysis of the three object types during the time window when participants had just heard the classifier. In Section 3.2, I will discuss the fixation analysis of the two competitor types in the same time window. Then, in Section 3.3, I will discuss the results of the target latency analysis, looking at comparisons between tone and rime competitors. Finally, in Sections 3.4 and 3.5, I will outline the limitations and future directions of the current study, and then summarize the overall conclusions.

3.1. Predicting upcoming information: An analysis of object type when hearing the classifier

The results of the current study have shown that there are significantly greater average fixations to the target compared to the baseline (i.e., the distractor objects) during the time window of the classifier. This held true in both high-constraint and low-constraint classifier contexts, and thus essentially confirms the first hypothesis proposed in the previous chapter: There was early pre-activation of the target word that was elicited by the semantic prediction of the classifier even before any acoustic signals were available, which supports the interactive view and argues against the integrative view. That is, participants actively used information at a specific time point to anticipate an upcoming word, since the target item was the only semantically plausible option for the classifier. This result would not have been predicted if, during lexical access, a prior semantic context only helps to choose among lexical candidates that were already activated by some limited sensory input of the target word (Traxler & Foss, 2000; Tyler & Wessels, 1983; Zwitserlood, 1989). Instead, the lexical access was initiated early in the prior context, which is consistent with the results found in Sereno, Brewer, & O'Donnell (2003). The result is also paralleled with the eye tracking result in Altmann & Kamide (1999), as the authors found that participants were able to fixate the target referent based on the semantic or thematic information ([+edible]) provided by the verb context even before the onset of target. Both of those prior studies show that listeners make use of any available information to pre-select an object that fits with the semantic requirement of the context. In summary, with regard to the locus of the contextual effect, my results support the idea that the activation of the target word in context can take place even before the availability of bottom-up acoustic information.

An important additional pattern in the results was that there were also significantly greater average fixations to the phonological competitors relative to the baseline in both context types. Such data is interesting, since this begs the question of what made the phonological competitors behave differently than the baseline, as the phonological competitors in this study violate the semantic requirements of relevant classifiers (and thus should not have been fixated). Under the interactive view, one possibility is that, at this early time stage in lexical processing, the semantically constraining classifiers were priming phonological information in the target word, which could have activated the tone and rime competitors as well. This is consistent with studies such as DeLong, Urbach, & Kutas (2005) and Liu, Shu, & Wei (2006). Recall that DeLong et al. (2005) had found that the phonological content of a specific word was pre-activated by the prior context: They had presented a preceding article that was not phonologically consistent with the anticipated word, which caused processing difficulty. Likewise, Liu et al. (2006) reported compatible results, showing that phonological information was fully activated by the preceding context in their study, where a received target word mismatching the onset of the expected word on a single phonological feature required less cognitive load than when the received word mismatched on multiple features. My results go beyond these previous reports, which had just focused on processing load (i.e., N400 activation and reaction time), and show that fixations are influenced by phonological predictions only based on semantic context, which has not been tested so far. Together, these results have supported the idea of contextual activation from the first chapter, which suggested that semantic constraints from the classifier not only facilitated the prediction of higher-level semantic content in the noun target, but also facilitated the activation of phonological detail in a top-down fashion.

Lastly, the factor of context type was found to influence fixations to the target and competitor. Specifically, in high-constraint contexts, there was greater activation to the target than to the competitors, whereas in low-constraint contexts, both objects were activated to the same extent. This is in line with all the previous studies finding different patterns of ERPs in high- and low-constraint contexts and thus supplements the argument for the interactive view (Dambacher, Kliegl, Hofmann, & Jacobs, 2006; Federmeier, 2007; Liu et al., 2006; Sereno et al., 2003; Van Petten & Kutas, 1990). The fact that there was greater activation of the target than the competitor only in the high-constraint context is evidence for the stronger predictability that high-constraint classifiers have for a very limited number of nouns. All the detailed phonological information of the target was thus more readily accessible in the high-constraint context, such that participants were capable of distinguishing the target from phonological competitors dissimilar in one component. However, for low-constraint classifiers, since they are able to activate a series of noun candidates, they have remarkably less bias towards a specific noun target. Here, all nouns that coarsely matched the target in phonological information were activated, but the target was not able to stand out from all the possible phonological candidates based on the activation they received at this point. This is consistent with Liu et al. (2006), where the authors found that the effect of mismatching the phonological features of the target word disappeared in low-constraint contexts, since such context provided less expectation for the phonological information of the following word.

Similar to what has been found in the low-constraint context in the current study, such related results have been observed in Zwitserlood (1989), which on the contrary, supported the integrative view. That author found, in both neutral and semantically biased contexts, that there was an equal activation of both the target word and its phonological competitor (that differed from the target word in its vowel and was semantically implausible in the biased context) at the pre-lexical stage. A context-effect was manifested later such that there was more activation of the target word only in the semantically biased context, but only after some acoustic information was accessible. Zwisterlood (1989) thus argued for the integrative view because the semantically constraining context did not pre-select a phonologically specified continuation, and thus participants could not make a distinction between the target and its phonological associates, at least not without the support of acoustic information. The current result indeed found that the target and its phonological competitor were not distinguished from each other, but it is still in favor of the interactive view. First, this latter result is found only in the low-constraint context. The stimuli and the strength of the constraining context in Zwitserlood (1989) are different from those in the current study, so it is possible that even the semantically biased context from Zwitserlood (1989) was not constrained enough to distinguish the target word from its phonological associates at the pre-lexical stage. Second, more importantly, by including items without any phonological relation to the target, it can be observed that both the target and its phonological competitors were still activated more than this baseline. If the target was not predicted at all by the context, such difference would not have been detected.

The current result also supports the validity of the experiment paradigm. There were four options on the screen with only one item that satisfied all the semantic requirements of the classifier, which forced participants to make a decision before hearing any sound. However, the degree of classifier constraint was still reflected in fixations. That is, the fact that high- and low-constraint classifiers had any effect shows that the auditory classifier context still likely activates of all legal noun candidates stored in their mental lexicon, instead of driving them to directly focus on the only plausible option on the screen. For high-constraint classifiers, only a couple specific nouns stored in the mental lexicon are available for selection (e.g. classifier “du3” requires the noun complement to specifically be [+wall]), whereas for low-constraint classifiers, a larger set of nouns that contain semantic features required by the classifiers can be legal candidates (e.g. classifier “tiao2” activates series of nouns that are [+long, +thin, +vertical], such as “rope”, “line”, “towel”). This is all to say that participants did not just check the semantic features of the four items on the screen to see whether they matched the requirements of relevant classifiers, but rather, activated a certain number of nouns in their mental lexicon when hearing the classifier, and then matched these candidates with what were shown on the screen to make a final decision. Since there were fewer plausible options in the mental lexicon for a high-constraint classifier, the matching process must have been faster and easier for the participants, and I found also that the phonological information of the target was more activated and specified in this high-constraint context, resulting in more fixations to the target than its phonological competitors. On the contrary, if it was the experiment paradigm that urged them to make the prediction, one might have predicted instead that there would not have been any differences between the two context types.

3.2. Pre-activation of tone and rime information: An analysis of competitor type when hearing the classifier

The second analysis described in the last chapter has compared fixations to different types of phonological competitors when participants had just heard the classifier. This tests how tone and rime information is processed in semantically constraining context. The results have demonstrated that participants were more likely to fixate the tone competitors than the rime competitors during this time window, but only in a high-constraint classifier context, which indicates that there was more distraction from the shared tone information than from shared rime information during this early activation process. However, this preference for tone over rime disappeared in low-constraint contexts. Such a result is in line with part of my second hypothesis on average looking time from Chapter 1, which predicts a greater average looking time to the tone competitor in high-constraint context. The greater distraction from the tone competitors in high-constraint contexts could have been caused by more tone activation at very early stages of lexical access. Recall that in Ye & Connine (1999), monitoring tone was found to be significantly faster than monitoring vowel in a strongly constraining context (four-syllable idioms that are very restricted in number) but not in a neutral context, and in Liu & Samuel (2007) vowel detection was impaired by tone mismatch in semantically-constraining contexts (strongly constraining sentences and four-syllable idioms), while tone detection stayed intact under vowel mismatch, with further results showing the tone advantage disappearing in isolated word contexts. However, since neither of these studies examined differences between pre-lexical, lexical and post-lexical stages, the mechanism of the contextual effect on tone and vowel processing had been unclear. The result found in the current study resonates with these studies such that tone information was given more perceptual salience than rime information, but clarifies that this happens, at least, in an early processing window. Another contribution is to show that this tone-over-rime context effect manifests in a new type of semantically constraining context, i.e., not just four-syllable idioms (Ye & Connine, 1999), and sentences (Liu & Samuel, 2007), but also in contexts of classifier-noun agreement as used in the current study.

While the previous analysis, which was described in Section 3.1, had already supported the notion that lexical access to both semantic and phonological aspects of an object could be initiated very early in a classifier-noun context, the current result further shows that there was an earlier access to tone information compared to the rime information during this early stage of prediction. Overall, this suggests that rime and tone information may be processed separately at different time stages in natural speech processing. Tone information may have been activated more at this pre-lexical stage, while rime information started to win back more perceptual salience at a later phase. A further question is how the semantic context managed to make the perceptual weight of tone more pronounced during this early phase.

There are several theoretical proposals about sensory prediction in lexical access that may help to explain these results. In a lexical recognition task with simultaneous EEG recording (Foucart, Ruiz-Tada, & Costa, 2015), participants falsely recognized expected words (when what had been actually presented was muted) significantly more than unexpected words in a semantically constraining context. That result suggests that a prior context provides semantic and syntactic constraints, which can generate hallucinations of upcoming words by creating a memory trace that resembles the upcoming actual sensory input. The current result also supports this implication. It is possible that the pre-activation of the target words and their phonological competitors happened because participants were themselves generating sensory input about the target words, which may also have facilitated the activation of competitors sharing phonological elements. Some researchers (Gambi & Pickering, 2013; Hickok, 2014; Pickering & Garrod, 2007, 2013) further propose that during the incremental process of speech comprehension, predictions of the upcoming word not only consists of the generation of sensory input, but also involves implicit imitation in production systems that can involve articulatory resources. They suggested that the basis of prediction shares commonalities between the perception and production systems, as the latter can function as an emulator during perception: Production systems might then compare the prediction against input, which can be used to incrementally improve subsequent predictions, which could in turn facilitate comprehending speech information rapidly.

This kind of proposal may also predict the current results. For instance, in the current experiment, when the participants heard and processed the classifier structure “na4-du3”, this context would be sufficient to generate the prediction of an imaginative noun target “qiang2” as if it had been actually perceived. At this moment, it is highly possible participants were also implicitly articulating their prediction of “qiang2”, which could have accelerated checking the prediction against the incoming auditory input to benefit comprehension. Evidence supporting the association between perception and production can be found in Heim, Opitz, Müller, & Friederici (2003), where the authors found an increase of motor-evoked potentials from tongue muscles when listening to sounds that greatly involved tongue movements, relative to non-speech, and in Watkins, Strafella, & Paus (2003), where the authors found that superior portion of Broca's area (responsible for speech production) along with left frontal and temporal areas (which support phonological processing in both speech production and perception) were activated in a phonological decision task based on perception. More specifically, there is evidence arguing for a remarkable role of production system in prediction: Wilson & Wilson (2005) found that the listeners were capable of predicting the beginning and ending point of a speaker's speech and synchronize their syllabic rate in conversations by entraining the production via the perception system.

One advantage of these production accounts of incremental perception is that they help to explain why tone might be weighted more heavily in lexical access under highly constraining conditions. That is, if lexical access in context is guided by prediction processes linked to articulatory speech planning, then the prediction of different phonological elements in perception, such as tone and rime information, should also be reflected at different levels in production. For example, evidence from speech errors has supported the idea that both tones and segments are selected at an early stage before the tone is assigned to a planned atonal syllable (Alderete, Chan, & Yeung, 2019), and tone is encoded separately from segments at an independent stage (Alderete et al., 2019; J. Chen, 1999; J. Chen, T. Chen, & Dell, 2002; Q. Zhang & Damian, 2009; Q. Zhang & Zhu, 2011). In J. Chen (1999), the majority of the segment errors were considered as movement errors (the nucleuses of "xun4-xi2" were swapped and it was mispronounced as "xin4-xu2"). Some tone errors such as pronouncing "pi1-pan4" as "pi1-pan2" was yielded by assigning the tone of the mistaken syllable "ping2" in "pi1-ping2" to the intended syllable without making changes to the original segments. More interestingly, in segment errors, tones stayed intact after the movement of segments and in movement errors with tone, segments stayed the same. Additionally, in speech errors in Cantonese, Alderete et al. (2019) found that most segment and tone errors did not involve the structure of each other and both in Mandarin (Wan & Jaeger, 1998) and Cantonese (Alderete et al., 2019), there was a greater than chance probability of contextual tone errors where the intended tones were substituted incorrectly by a source tone in the surrounding context, which suggests that tone is selected early before it is tightly linked to a syllable. If articulatory resources were indeed used in the prediction process, it is possible that listeners also selected tone and segment (rime) separately for upcoming perception, resulting in an atonal syllable, along with the phonemes and the tone at different processing levels (Gao et al., 2019). Although Alderete et al. (2019) were not conclusive about the timing of segmental and tonal access in production of single word, the current result at least supports the idea that tone and rime are intrinsically dissociative at a planning stage, not only in production, but also in preparation for perception.

If tone and rime are anticipated independently, a remaining question is how a high-constraint context may activate tone *ahead* of rime. In some previous studies (Gao et al., 2019; Shuai & Malins, 2017; Ye & Connine, 1999), scholars modified TRACE models in order to include lexical tone (or toneme node) in perception. For example, Ye & Connine (1999) and Shuai & Malins (2017) argued for a toneme level to be constructed in parallel with the phoneme level in the proposed perception model, as shown in Figure 6. Ye & Connine (1999) specified that in a neutral context, lexical nodes only provide activation for the phoneme node. The toneme node received activation via the connection with the phoneme node in these isolated contexts, whereas in a high-constraint context, the toneme node also received feedback from the lexical node and the feedback from lexical representation had a greater weight on the toneme node than the phoneme node.

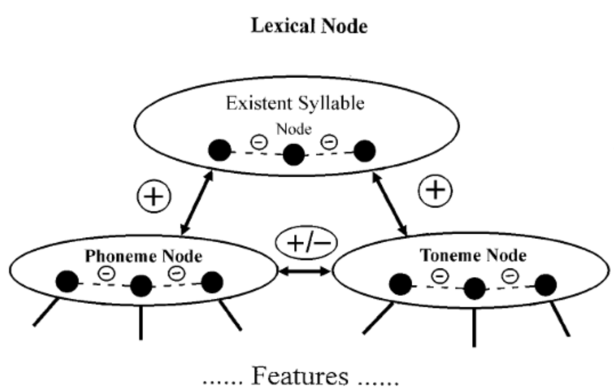

Figure 6. Phoneme, toneme and syllable nodes and connections.

Gao et al. (2019) proposed a Reverse Access Model (RAM), as presented in Figure 7, where two more new levels were added, i.e. the atonal syllable and lexical tone level. Information accessing was assumed to be initiated from the word level at the top and to proceed to the next level only if the information at the higher level is insufficient. Information at syllable and syllable-above levels is readily available (indicated by the solid eclipses) whereas information at lower tone and phoneme levels are "hidden" and can only be accessed to through a mental replay of the perceived word (indicated by the eclipses with dotted lines).

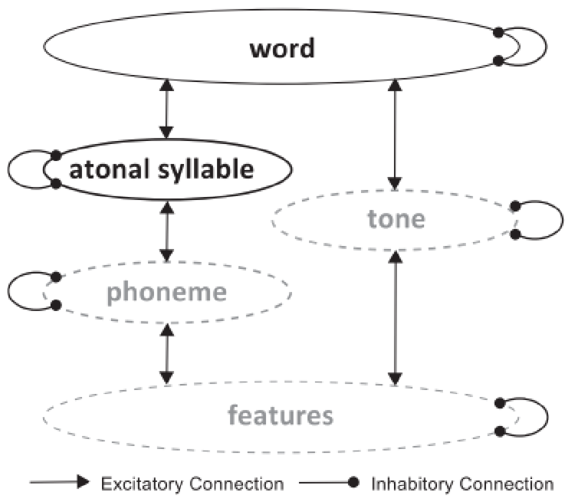

Figure 7. The Reverse Accessing Model for sub-lexical phonological processing of tonal speech.

In this model, lexical tone level was constructed one connection node closer to the word level compared to the phoneme level. According to the authors, when the word is embedded in a high-constraint context, the context provides semantic activation for the word level, which is strong enough to readily influence downstream tone level that is originally "hidden," but when such semantic activation proceeds as far as to the phoneme level, the effect of activation is much weaker. The phoneme information, instead, must wait till the mental replay of the perceived word after the acoustic signals are available. However, the 'rime competitor' in the current study resembles the atonal syllable in the RAM more. Gao et al. (2019) found that atonal syllables were processed more accurate and faster than component-level information (i.e. phonemes and lexical tones) in discrimination tasks, concluding that atonal syllable level must be an independent level where the information of syllables should be stored as basic units for perception. The current result that there was a perceptual advantage of tone competitor over atonal syllable competitor cannot fit in this model. Under the RAM, even in a high-constraint context, atonal syllables should still be more readily accessed than tone information because it is ranked closer to the higher word level. The reason for the contradiction between the RAM and the current study still remains unknown. One thing to note is that the experiment methods and conditions of the two studies are quite different, which might have led to different results. In Gao et al. (2019), the experiments were discrimination tasks for single words, which focused more on the lexical processing while my study includes preceding constraining context and focuses on the processing at the pre-lexical stage. A model that suits real-time processing of tone and segments in context is still needed.

Both of these models argue for a more equal role between lexical tone and phoneme, with lexical tone having an independent connection to the higher lexical level, which can only be activated under strong semantic influence. In the current study, I further propose that it is the mechanism of prediction that provide semantic activation of the word level, which further constrains the access to tone and phoneme level. I assume that a context does not directly predict a specific word but rather it predicts series of semantic features that constrain how many target words meet the semantic requirements of the context. Among these candidates, for each individual participant, there is a different probability for a certain word to be the completion of the context (Staub, Grant, Astheimer, & Cohen, 2015). For example, a classifier predicts that the upcoming word must have the semantic features of [+A, +B, +C] and words that contain these features (words that may be labeled D, E, F) all meet this requirement. However, if for Participant 1, the probability of D being the best completion is 60%, 30% for E, and 10% for F, it is highly possible that Participant 1 would anticipate D to a much stronger extent than the other words when the participant hears this classifier. For another Participant 2, the probabilities of Word D, E, F might be 30%, 60% and 10% respectively, and Word E is more likely to win out as the most expected word among the three. For a high-constraint classifier such as “duo3”, which refers to objects that are [+flower] and has only two plausible completions, the mean probability of “hua1” (flower) to be the most expected completion across subjects is 80% and that for the other noun “yun2” (cloud) to be the best completion is only 20% according to the current data (Appendix D). If tone level is indeed ranked closer to the word level in a perception model like the RAM (ignore the discussion of the existence of atonal syllable level for now), word-level information of “hua1” was more likely to be pre-activated across subjects and the activation was strong enough to proceed to influence the closer tone level but was not able to continue to activate the more distant phoneme level.

In addition, this assumption is also able to account for the absent tone advantage in low-constraint context. A low-constraint classifier such as “tiao2” refers to objects that are [+long, +thin, +vertical, (+able to be hung)]. The plausible noun completions have a broader semantic scale than the nouns of high-constraint classifiers and thus are greater in number. The mean probabilities of four of its legal continuation, “gou3” (dog), “she2” (snake), “he2” (river), “yu2” (fish) being the best completion are 13.3%, 6.7%, 6.7% and 20% respectively (Appendix D). Compared to the high-constraint condition, a specific noun candidate is anticipated to a lesser extent due to the smaller differences among the probabilities. All those nouns may receive a more equal activation, attenuating the amount of activation allocated to each individual noun. Therefore, this results in similar activation of the target and phonological competitors in low-constraint context (as discussed previously), as what the result of the first analysis shows. Take target noun “hua1” as an example, under the RAM, the atonal syllable “hua4” (rime competitor of “hua1”) and atonal syllable “hui1” (tone competitor) must also have been accessed to the same extent as the target, resulting in an equal activation between the rime competitor “hua4” and tone competitor “hui1”. This yields the disappearance of tone advantage in the low-constraint context.

Compared to the previous studies (Liu & Samuel, 2007; Ye & Connine, 1999) that are unclear about the mechanism of the context effect on constraining tone and rime processing, the current study has moved our understanding of tone processing forward by specifying the timing and stage of processing for tone and phoneme information: My results suggest that the tone advantage from a high-constraint context takes place as early as a pre-lexical stage, ahead of target acoustic input. Furthermore, I argue that it is through the mechanism of prediction that generates early activation of word information, which further constrains tone and rime processing. I have proposed a preliminary framework for thinking about the prediction of certain semantic features, and how the probability of a certain qualified word winning out as the best completion could yield activation of word level information that further spreads down to influence component-level information, in conjunction with the RAM model proposed by Gao et al. (2019).

3.3. Supplementary analysis for tone and rime processing: An analysis of first fixation latency

Recall that a supplementary analysis of first fixation was also conducted to test processing tone and rime in a larger time window, as described in Chapter Two. Instead of looking at the accumulation of fixations to different phonological competitors, this analysis investigates how quickly the participants could shift their eye gaze to the target when there was distraction from tone competitor versus rime competitor. The result showed that participants made their first fixation to the target at a comparable time point no matter whether the distraction was from a tone competitor or a rime competitor in either a high- or low-constraint context, which suggests that both tone and rime information were activated to a similar extent. Such result is inconsistent with the hypotheses that a) participants would make the first fixation to the target significantly later when the fixation was distracted by a tone competitor, compared to when being distracted by a rime competitor, and b) that this effect would be significantly greater in the high- than low-constraint context. The result also seems to contradict what has been found in prior analyses, where the effect of competitor type was observed.

One way to interpret this seemingly conflicting result is to look at when the average first fixation to the target was made, which was after the classifier window, once bottom-up acoustic input about the target started to exert its effect on eye-gaze. Thus, although there was a tone advantage in the classifier window when calculating the average number of fixations (Section 3.2), the average proportion of fixations made to the target in both contexts was relatively low at this pre-lexical stage, which indicated a relatively low level of pre-activation during the classifier. Because this current measure, the average latency for the first fixation to the target, was centered at a timepoint after the pre-lexical stage, it was increasingly affected by bottom-up sensory input interacting with top-level semantic information. During this period, the tone advantage in the high-constraint context seemed to have disappeared with the rime having started to re-gain its perceptual salience. This is consistent with Malins & Joanisse (2010) where the authors found that was a significant delay of looking at the target in both tone-competitor and rime-competitor condition relative to the baseline, and the delay in these two conditions happened at a comparable time point. In the current study, the first fixations made to the target under the two types of competitor conditions were also statistically similar. Both studies have suggested that, with the exposure to the acoustic signals of the target word, rime and tone information is processed along a similar time course and are equally important in lexical access.

The result of this first fixation analysis, combined the other two analyses in Sections 3.1 and 3.2, suggests that any tone-over-rime advantage in lexical access only occurs at a pre-lexical stage when there was only restricted activation of the abstract semantic concept (and an initial plan for constructing phonological information about the target word). When there was actual exposure to the target word, processing tone and rime in context tends to behave in agreement with processing words in isolation. The results from the three analyses appear to suggest different mechanism in the early stages of perception versus later stages actual perception. In Foucart et al. (2015), the authors found that, when perceivers were exposed to the target word, if acoustic input was presented that was inconsistent with a contextual prediction (i.e., when participants were presented with an unexpected article), they were able to instantly correct their prediction (reflected by more correct rejection when they heard an unexpected article). This suggests that listeners would experience two stages during word processing in semantic contexts, one of which includes generating prediction about the syntactic, semantic and phonological information of the upcoming word (based on the prior context) and another stage, which includes checking the prediction with actual exposure to the presented acoustic signals. During the early prediction stage, tone and rime are prepared at different levels in the memory template and there is an earlier tone selection than rime selection. But later, when the listeners were validating their different expectations for tone and rime, bottom-up acoustic information might have obscured the difference generated by that initial pre-activation, because when listeners were exposed to the target acoustic information directly, tone and rime were processed along similar time course and were equally salient in lexical access, just as what has been found in single word perception (Malins & Joanisse, 2010).

3.4. Limitations

The main limitation of the current study is that the experiment was designed under very restricted conditions, where classifier-noun agreement and phonological competitors of the noun targets were used as the stimuli, in order to meet the requirement of the visual world eye-tracking paradigm. The difficulty of experiment design can be demonstrated in three aspects: First, the target noun needed to be the best completion of the noun, had to be concrete, and had to have both a tone and rime competitor that were also concrete. This yielded a few unqualified best completions of some low-constraint classifiers, such as “niu2” of “na4-tou2”, as described in Chapter 2. Thus, the next best, or even the third best, completions of these classifiers were selected, resulting in all of the high-constraint classifiers being paired with their best completions, but only two of the low-constraint classifiers being paired with their best completions. Such discrepancies have left the target expectancy not fully controlled, which does not completely resemble the stimulus design in previous studies, where the expected targets always had the highest cloze probability following the context (Chou et al., 2014; Federmeier et al., 2007; Foucart et al., 2015). Nevertheless, it remains unlikely that the current stimulus selection caused any biased results here. This is because this design choice only accentuated the difference between high- and low-constraint classifiers: High-constraint classifiers, by definition, are strongly predictive of only one or two target nouns that are apparently their most expected continuation, whereas for low-constraint classifiers, there is more equal activation of series of nouns. In the current study, for the target nouns after the low-constraint classifiers, even if four of them were not the best completion, the classifiers might still have equivalent predictability of the next best completions, and such trivial differences of expectancy were unlikely to have affected the present results.

Second, the selection of the tone competitors of the target words had to simultaneously be concrete and have balanced lexical frequencies with each other, as well as be balanced in frequency with the targets, as stated in Chapter 2. This resulted in fact that one target word and the selected tone competitor also partially differ in their onset, as in (15).

(15) a. Target ke4 b. Tone competitor: kuang4 c. Rime competitor: ke2
d. Target xin4 e. Tone competitor: xian4 f. Rime competitor: xin1

In this case, “kuang4” was selected because and no other tone competitor without labialization of the initial consonant of could satisfy the standards for selection. Note that, while the traditional view of syllabification counts the medial glide as part of the rime structure, in the current study, it is syllabified into onset, following the method in Lin (2007). The proportion of this partially inconsistent onset only takes up 8.3% among all the experimental trials in one block. Also note that despite the trivial difference in the onsets, the rimes of this tone competitor (both the nucleus and coda) differed completely with the target. Yet, what consequence these trials might have brought to the final results remains unknown. In terms of the first two above-described analyses of average looking time (during the noun classifier), the time window captured was before any actual exposure to the acoustic information of the target noun, and so it is rather difficult to assume that these nuanced inconsistencies would have significantly influenced the results by bringing more fixations to one competitor type in a biasing way. With regard to the first fixation latency analysis, where contact with target acoustic information took place, such difference in onset could have made the fixations less distracted by the tone competitors in these trials, because the participants might have realized the difference between the target and tone competitor earlier. Nevertheless, the results show that the distraction from the two types of competitors was not significantly different.

In future research, one way to more closely investigate how tone and rime are processed after the actual contact with the target is to look at the noun time window to see whether the tone advantage will be obscured based on the average looking time to the two types of competitors. Doing so would more clearly show whether contextual effects of lexical processing could still occur in the presence of bottom-up acoustic activation from the target within the same time window, which would allow a clearer comparison of information from predictive, top-down versus acoustic, bottom-up information. Such an analysis could function as a companion to the current study to ask how exposure to acoustic information affects tone and rime activation, compared to when there is only constraining contextual information. For example, a similar experiment could be conducted, which uses a completely neutral context, such "possessive + noun" structure. Such experiment may function as a baseline to investigate the contextual effect at the pre-lexical stage, since little context information would be provided during this phase.

Finally, it is worth pointing out that distractors in any one experimental trial were selected from the targets and competitors in other experimental trials in order to have a closed stimulus set across all trials. Although in each trial, the most phonologically dissimilar distractors from the target were selected, there were, on some trials, some phonemes overlapping between the target and the distractors (the tones of distractors were ensured not to overlap with those of the target). The mean phonemic overlap between target and baseline distractors in the tone competitor condition is 7.5%, while that in the rime competitor condition is 1.7% (calculated by dividing the number of overlapping phonemes in the same position in a pair by the total number of the phonemes in the longer word), both of which are much lower than the degree of phoneme overlap reported in Malins & Joanisse (2010). Although this discrepancy does not seem to have influenced the average looking time to the two types of competitors (and because there was larger amount of phonemic overlap between the target and the tone-competitor, than between the target and any distractor), this design feature might have reduced the observable effect size from the tone competitor. Future research may need to more tightly control such phonemic overlap.

3.5. General conclusions

This study has investigated the eye gaze of native listeners of Mandarin Chinese, asking how vowel and tone information is used in online speech processing. Prior work has examined this question using isolated words, but the effects of context have not previously been explored in eye-tracking. In the current study, I investigated when and how top-down contextual effects constrained real-time processing in eye gaze, and whether it would have a different impact on activating tonal and vowel information in a type of semantically constraining context, i.e., classifier-noun agreement in Mandarin Chinese.

Results revealed that average looking time to both the target item and the phonological competitor were significantly greater than that to the distractors (which served as a baseline), which held in both classifier contexts. Moreover, fixations to the target were greater than to the phonological competitor only in the high-constraint context. In addition, there was more distraction from a tone than a rime competitor in the high-constraint context, but the tone advantage disappeared in the low-constraint context. Nevertheless, there were no significant effects of context or competitor types in the analysis of first fixation latency to the target.

The results indicate that there was an early top-down activation of the target noun and its phonological competitors, and that listeners actively used available contextual information to predict the upcoming word even before acoustic signals about the target became available. This provides evidence for an interactive model of lexical access. Listeners implicitly generate phonological activation of words before the exposure to actual acoustic signals, and during this stage, the specified phonological details of the target are also able to prime its phonological competitors. The interaction between context types and the activation of the objects suggests that a high-constraint context is predictive of very limited targets, making the target information more ready to stand out from its phonological associates. In low-constraint context, there is more equal activation of series of target words, resulting in a more coarse-grained activation of the target phonological information. Additionally, when the prior high-constraint context provides strongly predictive information for the following word, the tone information seems to be activated more than rime information. However, such tone advantage is not observed in the low-constraint context and when there is exposure to the actual acoustic signals, the rime tends to re-gain its perceptual salience. Such result suggests that tone and rime are processed at dissociative stages.

According to the RAM (Gao et al., 2019), the path between the tone level and the word level is shorter than that between the phoneme level and word level. However, the RAM has elicited problems of accounting for the current result since the rime competitors in the current study resemble the atonal syllables more in the RAM. Although, a more fine-grained model is still needed in the future study, the current study has specified the time stage of contextual activation of the word level representation in accessing further down tone and phoneme information, which takes place pre-lexically ahead of the target acoustic input. Furthermore, I propose that it is the mechanism of prediction that motivates the early activation of word information that further constrains tone and rime processing. Finally, it seems that exposure to acoustic input obscures the tone advantage: Future research needs to be conducted at the later noun stage and the model needs to be extended to include this stage.

References

Alderete, J., Chan, Q., & Yeung, H. H. (2019). Tone slips in Cantonese: Evidence for early phonological encoding. *Cognition*, *191*(April 2018), 103952. https://doi.org/10.1016/j.cognition.2019.04.021

Altmann, G. T. M., & Kamide, Y. (1999). Incremental interpretation at verbs: Restricting the domain of subsequent reference. *Cognition*, *73*(3), 247–264. https://doi.org/10.1016/S0010-0277(99)00059-1

Baayen, R. H., Davidson, D. J., & Bates, D. M. (2008). Mixed-effects modeling with crossed random effects for subjects and items. *Journal of Memory and Language*, *59*(4), 390–412. https://doi.org/10.1016/j.jml.2007.12.005

Chen, Jenn Y. (1999). The representation and processing of tone in Mandarin Chinese. *Evidence from Slips of the Tongue*, *20*(2), 289–301. Retrieved from http://www.scopus.com/inward/record.url?scp=0033239106&partnerID=8YFLogxK%5Cnhttp://www.scopus.com/inward/citedby.url?scp=0033239106&partnerID=8YFLogxK

Chen, Jenn Yeu, 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*(4), 751–781. https://doi.org/10.1006/jmla.2001.2825

Chou, C. J., Huang, H. W., Lee, C. L., & Lee, C. Y. (2014). Effects of semantic constraint and cloze probability on Chinese classifier-noun agreement. *Journal of Neurolinguistics*, *31*, 42–54. https://doi.org/10.1016/j.jneuroling.2014.06.003

Cutler, A, & Chen, H. C. (1997). Lexical tone in Cantonese spoken-word processing. *Perception & Psychophysics*, *59*(2), 165–179. https://doi.org/10.3758/BF03211886

Cutler, Anne. (1986). Forbear is a homophone: Lexical prosody does not constrain lexical access. *Language and Speech*, *29*(3), 201–220. https://doi.org/10.1177/002383098602900302

Cutler, Anne, & Clifton, C. (1984). The Use of Prosodic Information in Word Recognition. *Attention and Performance X: Control of Language Processes. Hillsdale, NJ: Erlbaum*, 183–196.

Cutler, Anne, & Donselaar, W. V. A. N. (2001). Voornaam is not (really) a Homophone : Lexical Prosody and Lexical A ccess in Dutch *, (24).

Cutler, Anne, Mehler, J., Norris, D., & Segui, J. (1987). Phoneme Identification and the Lexicon. *Cognitive Psychology*, *19*, 141–177.

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

Dambacher, M., Kliegl, R., Hofmann, M., & Jacobs, A. M. (2006). Frequency and predictability effects on event-related potentials during reading. *Brain Research*, *1084*(1), 89–103. https://doi.org/10.1016/j.brainres.2006.02.010

DeLong, K. A., Urbach, T. P., & Kutas, M. (2005). Probabilistic word pre-activation during language comprehension inferred from electrical brain activity. *Nature Neuroscience*, *8*(8), 1117–1121. https://doi.org/10.1038/nn1504

Elman, J. L., & McClelland, J. L. (1988). Cognitive penetration of the mechanisms of perception: Compensation for coarticulation of lexically restored phonemes. *Journal of Memory and Language*, *27*(2), 143–165. https://doi.org/10.1016/0749-596X(88)90071-X

Federmeier, K. D. (2007). Thinking ahead: The role and roots of prediction in language comprehension. *Psychophysiology*, *44*(4), 491–505. https://doi.org/10.1111/j.1469-8986.2007.00531.x

Federmeier, K. D., & Kutas, M. (1999). A Rose by Any Other Name: Long-Term Memory Structure and Sentence Processing. *Journal of Memory and Language*, *41*(4), 469–495. https://doi.org/http://www.idealibrary.com on A

Federmeier, K. D., Wlotko, E. W., De Ochoa-Dewald, E., & Kutas, M. (2007). Multiple effects of sentential constraint on word processing. *Brain Research*, *1146*(1), 75–84. https://doi.org/10.1016/j.brainres.2006.06.101

Foucart, A., Ruiz-Tada, E., & Costa, A. (2015). How do you know I was about to say “book”? Anticipation processes affect speech processing and lexical recognition. *Language, Cognition and Neuroscience*, *30*(6), 768–780. https://doi.org/10.1080/23273798.2015.1016047

Frauenfelder, U. H., Segui, J., & Dijkstra, T. (1990). Lexical Effects in Phonemic Processing: Facilitatory or Inhibitory? *Journal of Experimental Psychology: Human Perception and Performance*, *16*(1), 77–91. https://doi.org/10.1037/0096-1523.16.1.77

Gambi, C., & Pickering, M. J. (2013). Prediction and imitation in speech. *Frontiers in Psychology*, *4*(June), 1–9. https://doi.org/10.3389/fpsyg.2013.00340

Ganong, W. F. (1980). Phonetic categorization in auditory word perception. *Journal of Experimental Psychology: Human Perception and Performance*, *6*(1), 110–125. https://doi.org/10.1037/0096-1523.6.1.110

Gao, X., Yan, T. T., Tang, D. L., Huang, T., Shu, H., Nan, Y., & Zhang, Y. X. (2019). What Makes Lexical Tone Special: A Reverse Accessing Model for Tonal Speech Perception. *Frontiers in Psychology*, *10*(December). https://doi.org/10.3389/fpsyg.2019.02830

Gou, E,, Rao, G., Xiao, X., & Zang, J. (2016). Dashuju beijingxia BCC yuliaoku de yanzhi [The construction of the BCC Corpus in the age of Big Data]. Yuliaoku yuyanxue [Corpus Linguistics] (1).

Heim, S., Opitz, B., Müller, K., & Friederici, A. D. (2003). Phonological processing during language production: fMRI evidence for a shared production-comprehension network. *Cognitive Brain Research*, *16*(2), 285–296. https://doi.org/10.1016/S0926-6410(02)00284-7

Hickok, G. (2014). The architecture of speech production and the role of the phoneme in speech processing. *Language, Cognition and Neuroscience*, *29*(1), 2–20. https://doi.org/10.1080/01690965.2013.834370

Kamide, Y., Altmann, G. T. M., & Haywood, S. L. (2003). The time-course of prediction in incremental sentence processing: Evidence from anticipatory eye movements. *Journal of Memory and Language*, *49*(1), 133–156. https://doi.org/10.1016/S0749-596X(03)00023-8

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

Lee, C. (2007). Does Horse Activate Mother ? Processing Lexical Tone in Form Priming, *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. https://doi.org/10.1080/01690960600989600

Liu, Y., Shu, H., & Wei, J. (2006). Spoken word recognition in context : Evidence from Chinese ERP analyses, *96*, 37–48. https://doi.org/10.1016/j.bandl.2005.08.007

Liu, Z. (2013). Hanyu Liangci Dacidian [Chinese Classifier Dictionary]. Shanghai Cishu Chubanshe [Shanghai Lexicographic Publishing House].

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*, *62*(4), 407–420. https://doi.org/10.1016/j.jml.2010.02.004

McClelland, J. L., & Elman, J. L. (1986). The TRACE model of speech perception. *Cognitive Psychology*, *18*(1), 1–86. https://doi.org/10.1016/0010-0285(86)90015-0

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

Norris, Moss, H. E., & Marslen-Wilson, W. D. (1993). Access to Word Meanings During Spoken Language Comprehension: Effects of Sentential Semantic Context. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *19*(6), 1254–1276. https://doi.org/10.1037/0278-7393.19.6.1254

Norris, D. (1986). Word recognition: Context effects without priming. *Cognition*, *22*(2), 93–136. https://doi.org/10.1016/S0010-0277(86)90001-6

Paul D. Allopenna, James S. Magnuson, and M. K. T. (1997). Tracking the Time Course of Spoken Word Recognition Using Eye Movements: Evidence for Continuous Mapping Models. *Journal of Memory and Language*, *243*(38), 191–243.

Pickering, M. J., & Garrod, S. (2007). Do people use language production to make predictions during comprehension? *Trends in Cognitive Sciences*, *11*(3), 105–110. https://doi.org/10.1016/j.tics.2006.12.002

Pickering, M. J., & Garrod, S. (2013). An integrated theory of language production and comprehension. *Behavioral and Brain Sciences*, *36*(4), 329–347. https://doi.org/10.1017/S0140525X12001495

Rayner, K., & Frazier, L. (1989). Selection Mechanisms in Reading Lexically Ambiguous Words. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *15*(5), 779–790. https://doi.org/10.1037/0278-7393.15.5.779

Satterthwaite, F.E. Synthesis of variance. *Psychometrika* **6,** 309–316 (1941).

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), 1–12. https://doi.org/10.1162/0898929052880057

Segui, J., & Frauenfelder, U. H. (1986). The Effect of Lexical Constraints Upon Speech Perception, (Ld4ie 2013), 1127–1132.

Segui, J., Frauenfelder, U. H., & Mehler, J. (n.d.). Phoneme monitoring , syllable monitoring and lexical access Reference, *72*(4), 471–477.

Seidenberg, M. S., Waters, G. S., Sanders, M., & Langer, P. (1984). Pre- and postlexical loci of contextual effects on word recognition. *Memory & Cognition*, *12*(4), 315–328. https://doi.org/10.3758/BF03198291

Sereno, J. A., & Lee, H. (2015). The Contribution of Segmental and Tonal Information in Mandarin Spoken Word Processing. https://doi.org/10.1177/0023830914522956

Sereno, S. C., Brewer, C. C., & O'Donnell, P. J. (2003). Context Effects in Word Recognition: Evidence for Early Interactive Processing. *Psychological Science*, *14*(4), 328–333. https://doi.org/10.1111/1467-9280.14471

Shuai, L., & Malins, J. G. (2017). Encoding lexical tones in jTRACE: a simulation of monosyllabic spoken word recognition in Mandarin Chinese. *Behavior Research Methods*, *49*(1), 230–241. https://doi.org/10.3758/s13428-015-0690-0

Slowiaczek, L. M. (1990). Effects of lexical stress in auditory word recognition. *Language and Speech*, *33*(1), 47–68. https://doi.org/10.1177/002383099003300104

Soto-faraco, S., Sebastián-gallés, N., & Cutler, A. (2001). Segmental and Suprasegmental Mismatch in Lexical Access, *432*, 412–432. https://doi.org/10.1006/jmla.2000.2783

Staub, A., Grant, M., Astheimer, L., & Cohen, A. (2015). The influence of cloze probability and item constraint on cloze task response time. *Journal of Memory and Language*, *82*, 1–17. https://doi.org/10.1016/j.jml.2015.02.004

Traxler, M. J., & Foss, D. J. (2000). Effects of Sentence Constraint on Priming in Natural Language Comprehension. *Journal of Experimental Psychology: Learning Memory and Cognition*, *26*(5), 1266–1282. https://doi.org/10.1037/0278-7393.26.5.1266

Tweedy, J. R., Lapinski, R. H., & Schvaneveldt, R. W. (1977). Semantic-context effects on word recognition: Influence of varying the proportion of items presented in an appropriate context. *Memory & Cognition*, *5*(1), 84–89. https://doi.org/10.3758/BF03209197

Tyler, L. K., & Wessels, J. (1983). Quantifying contextual contributions to word-recognition processes. *Perception & Psychophysics*, *34*(5), 409–420. https://doi.org/10.3758/BF03203056

Van Petten, C., & Kutas, M. (1990). Interactions between sentence context and word frequencyinevent-related brainpotentials. *Memory & Cognition*, *18*(4), 380–393. https://doi.org/10.3758/BF03197127

Wan, I. P., & Jaeger, J. (1998). *Speech errors and the representation of tone in Mandarin Chinese*. *Phonology* (Vol. 15). https://doi.org/10.1017/S0952675799003668

Watkins, K. E., Strafella, A. P., & Paus, T. (2003). Seeing and hearing speech excites the motor system involved in speech production. *Neuropsychologia*, *41*(8), 989–994. https://doi.org/10.1016/S0028-3932(02)00316-0

Wiener, S., & Turnbull, R. (2016). Constraints of Tones, Vowels and Consonants on Lexical Selection in Mandarin Chinese. *Language and Speech*, *59*(1), 59–82. https://doi.org/10.1177/0023830915578000

Wilson, M., & Wilson, T. P. (2005). An oscillator model of the timing of turn-taking. *Psychonomic Bulletin and Review*, *12*(6), 957–968. https://doi.org/10.3758/BF03206432

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

Zhang, Q., & Damian, M. F. (2009). The time course of segment and tone encoding in Chinese spoken production: an event-related potential study. *Neuroscience*, *163*(1), 252–265. https://doi.org/10.1016/j.neuroscience.2009.06.015

Zhang, Qingfang, & Zhu, X. (2011). The temporal and spatial features of segmental and suprasegmental encoding during implicit picture naming: An event-related potential study. *Neuropsychologia*, *49*(14), 3813–3825. https://doi.org/10.1016/j.neuropsychologia.2011.09.040

Zwitserlood, P. (1989). The Locus of the Effects of Sentential-semantic Context in Spoken-word Processing, *32*(April 1985), 25–64.

Appendix A.

This table presents the 25 classifiers selected from *Chinese Classifier Dictionary* (Liu, 2013), their noun collocation verified in BCC corpus. Based on the number of the collocations, 12 of them were categorized into high-constraint classifiers and 13 of them were low-constraint classifiers.

Classifier	Number of nominal collocation (monosyllabic and disyllabic)	Context type	First collocation	Second collocation
(yi4) zhan3 (一)盏	3	High	Deng1 灯	Cha2 茶
(yi4) tang2 (一)堂	1	High	Ke4 课	NA
(yi4) du3 (一)堵	5	High	Qiang2 墙	Shi2-bi4 石壁
(yi2) shu4 (一)束	5	High	Hua1 花	Guang1 光
(yi4) feng1 (一)封	7	High	Xin4 信	Qing2-shu1 情书
(yi4) fu2 (一)幅	7	High	Hua4 画	Tu2 图
(yi4) duo3 (一)朵	5	High	Hua1 花	Qi2-pa1 奇葩
(yi4) zhong1 (一)盅	9	High	Cha2 茶	Jiu3 酒
(yi4) lü3 (一)缕	12	High	Yang2-guang1 阳光	Yan1 烟

Classifier	Number of nominal collocation (monosyllabic and disyllabic)	Context type	First collocation	Second collocation
(yi2) ye4 (一)页	2	High	Zhi3 纸	Shu1 书
(yi2) zhan4 (一)站	4	High	Lu4 路	Che1 车
(yi2) jian4 (一)件	1908	Low	Shi4 事	Yi1-fu 衣服
(yi2) li4 (一)粒	696	Low	Sha1 沙	Zhong3-zi 种子
(yi4) shuang1 (一)双	1936	Low	Yan3-jing1 眼睛	Shou3 手
(yi4) tiao2 (一)条	6553	Low	Lu4 路	Tui3 腿
(yi4) gen1 (一)根	2080	Low	Sheng2-zi 绳子	Yan1 烟
(yi4) zhang1 (一)张	2080	Low	Zhuo1-zi 桌子	Zhi3 纸
(yi4) bao1 (一)包	1006	Low	Dong1-xi 东西	Yan1 烟
(yi4) qun2 (一)群	3605	Low	Ren2 人	Hai2-zi 孩子
(yi4) di1 (一)滴	567	Low	Yan3-lei4 眼泪	Shui3 水
(yi4) tou2 (一)头	523	Low	Niu2 牛	Zhu1 猪
(yi4) ping2 (一)瓶	1066	Low	Jiu3 酒	Pi2-jiu3 啤酒

Classifier	Number of nominal collocation (monosyllabic and disyllabic)	Context type	First collocation	Second collocation
(yi4) xie1 (一)些	16713	Low	Ren2 人	Shi4-qing2 事情
(yi4) bei1 (一)杯	1305	Low	Jiu3 酒	Cha2 茶
(yi2) ge4 (一)个	35977	Low	Nü3-ren2 女人	Yue4 月

Appendix B.

The table presents the familiarity rating of the 25 classifiers.

Classifier	Familiarity	Context type
(yi4) zhan3 (一)盏	2.8	High
(yi4) tang2 (一)堂	3.07	High
(yi4) du3 (一)堵	2.47	High
(yi2) shu4 (一)束	3.4	High
(yi4) feng1 (一)封	3.27	High
(yi4) fu2 (一)幅	3.53	High
(yi4) duo3 (一)朵	3.73	High
(yi4) zhong1 (一)盅	2	High
(yi4) lü3 (一)缕	2.87	High
(yi2)ye4 (一)页	4.33	High
(yi2) zhan4 (一)站	3.4	High
(yi2) jian4 (一)件	4.73	Low

Classifier	Familiarity	Context type
(yi2) li4 (一)粒	3.53	Low
(yi4) shuang1 (一)双	4.47	Low
(yi4) tiao2 (一)条	4.4	Low
(yi4) gen1 (一)根	4.4	Low
(yi4) zhang1 (一)张	4.8	Low
(yi4) bao1 (一)包	3.53	Low
(yi4) qun2 (一)群	4.27	Low
(yi4) di1 (一)滴	3.67	Low
(yi4) tou2 (一)头	3.53	Low
(yi4) ping2 (一)瓶	4.47	Low
(yi4) xie1 (一)些	4.8	Low
(yi4) bei1 (一)杯	4.47	Low
(yi2) ge4 (一)个	4.93	Low

Appendix C.

The table shows the results of the norming task, including the constraining of the classifiers, context type they belong to, and their best and next best completion.

Classifier	Constraining strength	Context type	Best completion	Next best completion
(yi4) du3 (一)堵	1.13	High	Qiang2 墙	Men2 门
(yi4) tang2 (一)堂	1.33	High	Ke4 课	Jiang3-zuo4 讲座
(yi4) duo3 (一)朵	1.67	High	Hua1 花	Yun2 云
(yi4) zhan3 (一)盏	1.87	High	Deng1 灯	Cha2/jiu3 茶/酒
(yi2) shu4 (一)束	2.07	High	Hua1 花	Guang1 光
(yi4) feng1 (一)封	2.13	High	Xin4 信	Qing2-shu1 情书
(yi4) fu2 (一)幅	2.27	High	Hua4 画	Zi4 字
(yi4) lü3 (一)缕	2.47	High	Yan1/yang2-guang1/tou2-fa 烟/阳光/头发	Qing1-si1 青丝
(yi2)ye4 (一)页	2.47	High	Zhi3 纸	Shu1 书
(yi2) jian4 (一)件	2.67	High	Yi1-fu 衣服	Shi4 事
(yi2) li4 (一)粒	2.87	High	Mi3 米	Yao4 药

Classifier	Constraining strength	Context type	Best completion	Next best completion
(yi4) shuang1 (一)双	3.73	Low	Xie2 鞋	Wa4-zi 袜子
(yi4) tiao2 (一)条	3.73	Low	Yu2 鱼	She2 蛇
(yi4) gen1 (一)根	3.73	Low	Cong1 葱	Sheng2 绳
(yi4) zhang1 (一)张	3.80	Low	Zhi3 纸	Ming2-pian4 名片
(yi4) bao1 (一)包	3.80	Low	Yan1 烟	Tang2 糖
(yi4) qun2 (一)群	3.87	Low	Ren2 人	Yang2 羊
(yi4) di1 (一)滴	4.07	Low	Shui3 水	Lei4 泪
(yi4) tou2 (一)头	4.13	Low	Niu2 牛	Zhu1 猪
(yi4) ping2 (一)瓶	4.40	Low	Shui3 水	Jiu3 酒
(yi4) xie1 (一)些	4.60	Low	Ren2 人	Shi4 事
(yi4) bei1 (一)杯	4.67	Low	Shui3 水	Jiu3 酒
(yi2) ge4 (一)个	5.60	Low	Ren2 人	Shou3-ji1/jian4-zhu4/xue2-xiao4/ping2-guo3 手机/建筑/学校/苹果

Appendix D.

During the selection, the following classifiers were excluded and the next qualified classifier on the ranking was selected:

	Constraining type	Constraining score	Best completion	Next completion
Yi2 shu4 一束	High	2.07	Hua1 花 “flower”	Guang1 光 “light”
Yi4 fu2 一幅	High	2.27	Hua4 画 “painting”	Zi4 字 “calligraphy”
Yi2 ge4 一个	Low	5.60	Ren2 人 “person”	Shou3-ji1 “cell phone”
Yi4 bei1 一杯	Low	4.67	Shui3 水 “water”	Jiu3 酒 “wine”
Yi4 xie1 一些	Low	4.60	Ren2 人 “person”	Shi4 事 “things”
Yi4 ping2 一瓶	Low	4.40	Shui3 水 “water”	Jiu3 酒 “wine”
Yi4 bao1 一包	Low	3.80	Yan1 烟 “cigarette”	Tang2 糖 “candy”
Yi4 zhang1 一张	Low	3.80	Zhi3 纸 “paper”	Ming2-pian4 名片 “name card”

The reasons for the exclusion of each classifier are:

a. For “yi2 shu4”, the first completion “hua1” overlaps with the first completion of “duo3”, which was selected before “shu4” because of the higher constraining score of “duo3”. Since “shu4” belongs to the high-constraint group, the next classifier on the ranking was selected instead of keeping “shu4,” and using its second completion as the target noun.

b. For “yi4 fu2”, the first completion “hua4” overlaps with the only eligible rime competitor of “hua1”, which is the first completion of “yi4duo3”. In order to balance the times of occurrences of each object as a visual and auditory stimulus in the experiment, the same object cannot be both a target and a competitor across different experimental trials

c. “ge4”, “xie1” and “zhang1” were excluded because their first completions do not have either tone competitors or vowel competitors. The second completions of “ge4” and “zhang1” are both disyllabic, which would be inconsistent with the monosyllabic completions of other selected classifiers and may introduce more unwanted variabilities to the results. Finally, the second completion of “xie1” is too generic to be picturable.

d. “bei1” (cup), “ping2” (bottle) and “bao1”(bag) are classifiers measuring the amount of the target nouns and they have lexical meanings corresponding to different containers. In the final eye-tracking experiment, on the one hand, if the containers denoted by these classifiers show up in the visual arrays, along with the nominal objects that they contain, they would probably bias participants’ eye gaze directly towards the target picture during the time window of the classifier, not because the classifiers activate certain nouns, but because the classifiers named in the audio recording also appear in the pictures, along with the target nouns. On the other hand, after checking with several native speakers, an agreement was reached that if participants heard these classifiers but do not see the containers denoted by them in the pictures, as the containers of the target nouns, it would be intuitively odd and might cause a delay of processing. Thus, these three classifiers were also eliminated.

Appendix E.

The table shows the final 12 classifiers selected, their best and next best completion, and the nouns' cloze probability.

Classifier	Completion	Noun	Cloze probability of being the best or next best
(yi4) du3 (一)堵	Best completion	Qiang2 墙	1
	Next best completion	Men2 门	0.13
(yi4) tang2 (一)堂	Best completion	Ke4 课	0.93
	Next best completion	Jiang3-zuo4 讲座	0.13
(yi4) duo3 (一)朵	Best completion	Hua1 花	0.8
	Next best completion	Yun2 云	0.3
(yi4) zhan3 (一)盏	Best completion	Deng1 灯	1
	Next best completion	Cha2/jiu3 茶/酒	0.27
(yi4) feng1 (一)封	Best completion	Xin4 信	0.93
	Next best completion	Qing2-shu1 情书	0.3
(yi4) lü3 (一)缕	Best completion	Yan1/yang2-guang1/tou2-fa 烟/阳光/头发	0.27

Classifier	Completion	Noun	Cloze probability of being the best or next best
	Next best completion	Qing1-si1 青丝	0.27[4]
(yi4) shuang1 (一)双	Best completion	Xie2 鞋	0.4
	Next best completion	Wa4-zi 袜子	0.46[5]
(yi4) tiao2 (一)条	Best completion	Yu2 鱼	0.2
	Next best completion	She2 蛇	0.27
(yi4) gen1 (一)根	Best completion	Cong1 葱	0.27
	Next best completion	Sheng2 绳	0.2
(yi4) qun2 (一)群	Best completion	Ren2 人	0.73
	Third best completion	Ya1 鸭	0.13
(yi4) di1 (一)滴	Best completion	Shui3 水	0.93
	Next best completion	Lei4 泪	0.3

[4] The probability of "qing1-si1" being the next completion is the same as "yan1" being the best completion. However, the probability of "qing1-si1" being the best completion is only 0.13. The methods of calculating the probability being the best and next best completion were different. See the description in Section 2.1.2.1.

[5] The cloze probability of "wa4-zi" being the best completion is 0.3 while that for being the next best is 0.46, outweighing other nouns for being the next best completion of "yi4 shuang1". Same for "she2".

Classifier	Completion	Noun	Cloze probability of being the best or next best
(yi4) tou2 (一)头	Best completion	Niu2 牛	0.6
	Next best completion	Zhu1 猪	0.53

Appendix F.

The table presents the 12 classifiers, their selected noun target, and the tone and rime competitors of the target nouns.

Classifier	Context type	Target	Tone competitor (LF)	Rime competitor (LF)
(yi4) du3 (一)堵	High	Qiang2 (2.13) 墙	Qian2(2.77) 钱	Qiang1(2.31) 枪
(yi4) tang2 (一)堂	High	Ke4(1.95) 课	Kuang4 (2.08) 矿	Ke2 (1.61) 壳
(yi4) duo3 (一)朵	High	Hua1(2.53) 花	Hui1(1.31) 灰	Hua4 (2) 画
(yi4) zhan3 (一)盏	High	Deng1 (2.42) 灯	Dao1 (2.09) 刀	Deng4 (1.46) 凳
(yi4) feng1 (一)封	High	Xin4 (2.44) 信	Xian4 (2.36) 线	Xin1 (2.99) 心
(yi4) lü3 (一)缕	High	Yan1 (2.3) 烟	Ying1 (0.58) 鹰	Yan2 (1.73) 盐
(yi4) shuang1 (一)双	Low	Xie2 (2.18) 鞋	Xiong2 (0.58) 熊	Xie4 (0.79) 蟹
(yi4) tiao2 (一)条	Low	Yu2 (2.53) 鱼	You2 (3.21) 油	Yu3 (2.5) 雨
(yi4) gen1 (一)根	Low	Sheng2 (1.48) 绳	Shi2 (2.24) 石	Sheng3 (2.31) 省
(yi4) qun2 (一)群	Low	Ya1 (1.2) 鸭	Yao1 (2.04) 腰	Ya2 (1.83) 牙
(yi4) di1 (一)滴	Low	Lei4 (1.72) 泪	Lu4 (2.84) 鹿	Lei2 (1.74) 雷

Classifier	Context type	Target	Tone competitor (LF)	Rime competitor (LF)
(yi4) tou2 (一)头	Low	Zhu1 (1.8) 猪	Zhen1 (1.81) 针	Zhu2 (1.71) 竹