PHONOLOGICAL PROCESSING AND ENCODING OF TONES IN MANDARIN CHINESE

SIYING CHEN

PhD

The Hong Kong Polytechnic University

2024

The Hong Kong Polytechnic University

Chinese and Bilingual Studies

PHONOLOGICAL PROCESSING AND ENCODING OF TONES IN MANDARIN CHINESE

Siying Chen

A thesis submitted in partial fulfillment of the requirements for the degree of Doctor of Philosophy

August 2023

CERTIFICATE OF ORIGINALITY

I hereby declare that this thesis is my own work and that, to the best of my knowledge and belief, it reproduces no material previously published or written, nor material that has been accepted for the award of any other degree or diploma, except where due acknowledgement has been made in the text.

(Signed)

___Siying Chen______ (Name of Student)

ABSTRACT

The present study examines the perception and production of lexical tone in Mandarin Chinese. Lexical tone, which adds pitch contours to the sonorant part of a syllable, is a lexically contrastive phoneme similar to how consonants and vowels (also known as “segments”) distinguish between words in Indo-European languages. This study uses behavioral methods to test two aspects of tone in spoken language in an attempt to answer the question: are tones perceived and produced in a qualitatively different way from segments?

The first part of the study looks at the status of tones in phonological processing. The first series of 4 experiments uses a lexical selection task to examine how willing Mandarin speakers were to change tones as opposed to segments. The results support previous research which claims that tones are less lexically binding than onsets and vowels; in the present study, participants were more likely to change a tone than to change an onset or a vowel.

However, segments include not only onsets and vowels, but also codas, of which Mandarin has two: /n/ and /ŋ/. Given that there are only two items in the Mandarin coda category, they provide significantly less lexical information than onsets and vowels. Depending on the model, onsets and vowels are counted and categorized differently, with onsets being anywhere from 19 to 23 items, and vowels classically analyzed as anywhere from 5 to 7 distinct items, but up to 24 items when counting complex vowels, such as diphthongs and triphthongs.

More importantly, the coda category also has fewer items than the tone category, which has 4 items. Experiments 2-3 showed that speakers were more willing to change codas than to change tones, which lends support to the idea that tones do not behave intrinsically differently from segments as a whole. Instead, tones behave in a way similarly to segments - they constrain word access in the same way as segments do, and the amount of that constraint is determined by how much information they provide.

In the second part of the study, seeing as how tones don't behave qualitatively differently than segments in perception, I turn to see whether tones behave like segments in production, using a tongue twister paradigm to examine the role of tones in speech production. Speech production involves a process called phonological encoding, the process by which a speaker builds the articulatory plan for an intended utterance. Certain phonemes can be viewed as having early or active phonological encoding, which shows that they play an active role in the speech preparation process. Speech errors have often been examined as a way to shed light on the role of different phonemes in speech production, and the general consensus is that if a type of phoneme is actively encoded, it would incur a substantial amount of errors in speech, due to the active mapping of a phoneme to its position in a lexical sequence. Previously, there has been conflicting research supporting both frequent tone error and infrequent tone error in natural speech. Frequent tone error lends support to the theory that tone is actively encoded in speech, encoded similarly to segments. Infrequent tone error supports the theory that tone is inactively processed, or processed later, as compared to segments.

As with the first set of experiments on phonological process, my second experiment focusing on phonological encoding also breaks segments down into their components: onsets, vowels, and codas. My experiment uses a tongue twister paradigm designed to elicit equal amounts of error in 4 phoneme categories: tone, onset, vowel, and coda. If tone error rates are significantly lower than that of the 3 segmental categories, then I would conclude that tones are not actively encoded. However, if tone error rates were similar to those of the 3 segmental categories, I would assume that there is insufficient evidence to conclude that tones are encoded in a way unsimilar to segments. The results showed that codas showed the most errors, followed by onsets and tones at around the same error rates, and with vowels showing the least error rates. Additionally, tone errors were modulated by context in the same way as

were segmental errors, a key indicator of active speech encoding. These results show substantial evidence arguing for the active encoding of tones in Mandarin speech.

Overall, the present study provides a great deal of new data showing how tone behaves in the perception and production of Mandarin Chinese. Although tone is a suprasegmental phoneme, the results suggest that tones and segmental phonemes are perceived similarly and that tones and segments are encoded similarly once different classes of segmental phonemes are considered separately.

PUBLICATIONS ARISING FROM THESIS

Chen, J.S., Politzer-Ahles, S. Acceptance of tonal and segmental variability correlates to inventory size in Mandarin Chinese. Paper accepted at: *The 13th International Symposium on Chinese Spoken Language Processing*; December 2022; Singapore.

Chen, J.S., Politzer-Ahles, S. Tones encoded similarly to segments in Mandarin speech production: a tongue twister study. Poster presented at: *The 12th International Workshop on Language Production*; June 2022; Pittsburgh, PA.

Chen, J.S. Acceptance of tonal and segmental variability correlates to inventory size in Mandarin Chinese. Poster presented at: *The 8th International Conference on Phonology and Morphology*; June 2022; Seoul, South Korea (virtual).

Chen, J.S. Tonal versus segmental perception in Mandarin Chinese speakers. Talk presented at: Postgraduate Research Symposium on Linguistics, Language, and Speech; June 2021; Hong Kong (virtual). Won Best Oral Presentation Award.

ACKNOWLEDGEMENTS

First and foremost, I would like to thank my dissertation supervisors, Dr. Stephen Politzer-Ahles and Dr. Yao Yao, for their time, their generosity, and their wisdom throughout the entire process of my doctoral program. It would not be an overstatement to say that without their help, this dissertation would not exist today.

Additionally, I would like to thank the Research Grants Council (RGC) of Hong Kong for their sponsorship of my studies at the Hong Kong Polytechnic University over the past three years. Being a recipient of the Hong Kong PhD Fellowship Scheme is an honor that I will cherish forever, as it gave me the freedom to pursue my research and academic interests outside of what may have normally been afforded to me.

I would also like to thank the professors involved in the confirmation, Dr. Caicai Zhang and Dr. Si (Sarah) Chen, for their advice and encouraging words as I moved into the latter half of my doctoral program.

At PolyU, not only did I spend the past three years as a doctoral student; but I also had the pleasure of working as a research assistant for both Dr. Politzer-Ahles and Dr. Yu Yin Hsu. I would like to thank Dr. Politzer-Ahles and Dr. Hsu for the opportunity to learn about and experience being a part of the exciting research projects currently underway in the department.

While working as a research assistant, I made the acquaintance of Leon Ka Keung Lee, who has been working as a research assistant in the Department of Chinese and Bilingual studies for the past 4 years. Leon has been an irreplaceable part of my experience at PolyU - he taught me to use much of the equipment at CBS, and has many a time let me into the studios when I didn’t have access to a key.

While writing my dissertation, I attended several academic conferences that opened my eyes to ways to improve both my research methods and perspective. I am very grateful to the reviewers and colleagues at the International Workshop on Language Production (2022), the International Conference on Phonology and Morphology (2022), and the International Symposium on Chinese Spoken Language Processing (2022). Their feedback has been invaluable to me, and many of the improvements made in the latter half of my program was a result of their suggestions.

I mentioned Dr. Politzer-Ahles at the onset of this section, but I cannot stress enough how much Dr. Politzer-Ahles has shaped my academic career thus far. Coming to Hong Kong for the first time in 2018, as the only American graduate student at CBS, meeting Dr. Politzer-Ahles and attending his Comparative Analysis course was a watershed experience for me. Without his generous support, it's hard to imagine that I would one day write my own doctoral dissertation. Although he's no longer a part of CBS, he remains a very large part of my PolyU experience, and I could not be more grateful for the role he's played in my academic life. Thank you from the bottom of my heart.

I cannot end this section without expressing my gratitude to everyone at CBS that has guided and supported me in my journey for the past 5 years. It has truly been an indelible 5 years, and I cannot be more thankful for the knowledge, wisdom and experience I've gained.

Last but not least, I would not have been able to come to Hong Kong were it not for the constant support and love of my parents, Jinwen and Xiaoyan. Throughout the years, you have been my rock in everything I've done. From Changchun to Philadelphia to Hong Kong, I've come a long way, but I've been content throughout it all, knowing that I have the best parents in the world.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

CHAPTER 1: INTRODUCTION

Mandarin Chinese (henceforward “Mandarin”) is a language in the Sino-Tibetan family, a language family consisting mainly of Chinese, Tibetan and Burmese languages. For the purposes of this dissertation, I will be discussing Standard Mandarin, the official language of China, which was adapted from the Beijing variation of Mandarin.

Chinese has many features that set it apart from Indo-European languages, including a logographic orthography, a lexicon consisting of monosyllabic morphemes as the foundational units, and perhaps most notably, a lexical tonal system. As this dissertation focuses on the phonological side of Mandarin, I will briefly discuss the structure of Chinese syllables, as well as its tonal system, which some linguists claim affects the phonological perception and production of Chinese.

A Chinese syllable has the following possible structures:

1. **C**V
2. **C**VC
3. V
4. VC

In this system, V stands for monophthongs, diphthongs, and triphthongs. Some systems also view diphthongs and triphthongs as having a glide plus a main vowel or a vowel plus an on-glide before and an off-glide after (Lin, 1989). The onset, or initial consonant, can hold any consonant in Mandarin other than *ng*. The coda, or final consonant, can hold only nasal consonants *n* and *ng*.

The Chinese syllable structure is not unique in the above regards, in which the Chinese syllable typically consists of minimally a vowel, with a consonant either in the onset or coda positions, or both. However, what sets Chinese apart is the way that it is typically taught in schools - the sounds of Chinese are not usually taught in its consonant and vowel constituents, but rather as initial and final combinations (Zhang, 1992; Yip, 2000). In this system, "initials" (声母, *sheng1mu3*) consists of all initial consonants, and "finals" (韵母, *yun4mu3*) consists of everything else: the vowel (or vowel combination), with or without a coda (see Figure 1). As you can see, the final/rime can have up to 4 phonological units. Thus, Chinese speakers' metalinguistic knowledge of the sounds of their language tends to be in terms of initials and finals, not consonants and vowels.

The focus on initials and finals may have an effect on the way Chinese speakers perceive the sounds in their language, as there is research showing that Chinese speakers do not use segments (consonants and vowels) individually to prime words in implicit spoken word planning the same way as do speakers of Indo-European languages (O'Seaghdha et al., 2010; Zhao et al., 2011). Additionally, since many psycholinguistics experiments split initials and finals when testing for phonological priming (rather than splitting up each segment individually), I believe that it is important to talk about this system as a foundation for discussing Chinese phonology.

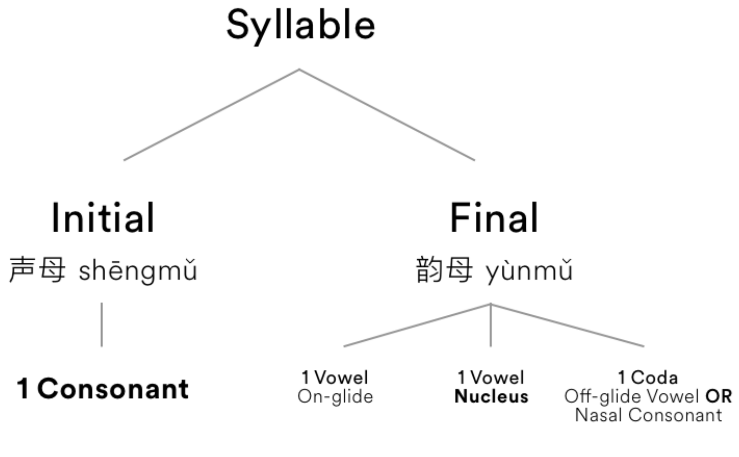


Figure 1: Mandarin syllable structure, as traditionally taught in schools. *Adapted from "A diachronically-motivated segmental phonology of Mandarin Chinese" by W. Li, 1999.*

According to this system, Mandarin has 22 initials and 35 finals (See Tables 1 and 2). It is through these 22 initials and 35 finals that form all the possible syllable combinations in Mandarin. In theory, 22 initials multiplied by 35 finals would yield 770 syllables, but as with all naturally formed languages, there are exceptions and systemic gaps in the syllable structure. For example, the initials *j*, *q*, *x* (in Chinese pinyin) are in complementary distribution with *zh*, *ch*, *sh*, *r* and *z*, *c*, *s*, and can only precede the vowels *i* and *ü* (Yip, 2000). In total, there are about 410 syllables in active use today (Commercial Press, 2020).

Initials

bilabial/labiodental	**p** [p^{h}], **b** [p], **m** [m], **f** [f]
alveolar	**t** [t^{h}], **d** [t], **n** [n], **l** [l]
alveolar	**c** [ts^{h}], **z** [ts], **s** [s]
retroflex	**ch** [tʂ^{h}], **zh** [tʂ], **sh** [ʂ^{h}], **r** [ʐ]
palatal (dorsal)	**q** [tɕ^{h}], **j** [tɕ], **x** [ɕ]
velar	**k** [k^{h}], **g** [g], **h** [x]
no initial	**ø**

Table 2: A list of Mandarin initials, organized by place of articulation.

Finals

a	**a** [a], **an** [an], **ang** [aŋ], **ai** [ai], **ao** [ɑu]
o/e	**o** [o], **e** [ɤ], **en** [ən], **eng** [əŋ], **ei** [ei], **ou** [ou], **ong** [uŋ], **er** [ɚ]
u	**u** [u], **ua** [ua], **uo** [wo], **uai** [wai], **ui** [wei], **uan** [wɑn], **un** [wən], **uang** [wɑŋ]
i	**i** [i]/[ɹ̩], **in** [in], **ia** [ja], **iao** [jɑu], **ie** [je], **iu** [jou], **ian** [jɛŋ], **iang** [jɑŋ], **ing** [iəŋ], **iong** [juŋ]
ü	**ü** [y], **üe** [ɥe], **üan** [ɥɛŋ], **ün** [ɥn]

Table 3: A list of Mandarin finals, organized by the first vowel/glide that appears.

Although traditional Chinese phonology is organized in this manner, modern linguists have mainly found this organization problematic, as the rhyming unit in the syllable, the nucleus and coda, does not form its own unit (the final in this system includes the *glide*, the nucleus, and the coda). Many theories have been proposed for the organization of the syllable structure in Chinese, but there has not been a consensus. Due to the on-glide vowel being a problem, many theories have separated the glide from the nucleus and coda. Figure 2 summarizes the main theories of the Chinese syllable structure.

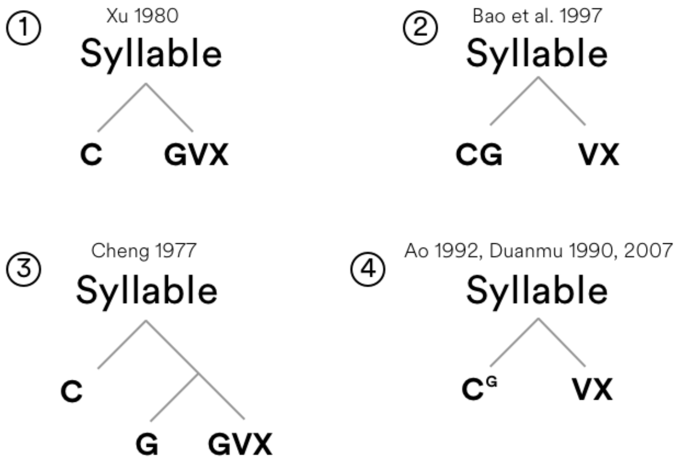


Figure 2: Chinese Syllable Structure Theories. Compiled by Duanmu (2011); the graphic representations are inspired by Duanmu's compilation. *Note: the 4th theory presented here is the simplified CVX model, in which the initial C can become a complex sound, with a glide as a secondary articulation.*

As the focus of this study is on the behavior and status of tones in Mandarin, the underlying structure of the Chinese syllable will not be an area in which I make any predictions or analyses. However, through these various theories presented over the years, it is important to note that the status of the initial-final structure in Chinese is challenged by many linguists, an idea that will be explored further in the phonological perception study.

In addition to the initial and final syllable structure, Mandarin makes the use of lexical tones. Lexical tone is defined as a distinctive pitch level that accompanies a syllable, carrying crucial lexical information - tones in nature can be of varying heights and varying contours, and Mandarin has both level (flat) pitch and contour (rising, falling or a combination) pitch tones. Lexical tone is considered to be *suprasegmental*, a type of phoneme that exists beyond the segmental realm of consonants and vowels. Other items in the suprasegmental category include stress, intonation, and rhythm.

As mentioned before, the basic building blocks of Chinese are monosyllables, and each monosyllable in Chinese has an accompanying tone. Modern Chinese phonology has 4 distinct tones, each of different height, duration and pitch contours. Tones in Mandarin are lexically distinctive, which means the height and pitch associated with each syllable is used to distinguish different words. For example, the syllable /ma/ has different meanings depending on its tone: a high, flat tone (tone 1) /ma/ means "mother;" a rising tone (tone 2) /ma/ means "numb" or "hemp;" a low, dipping tone (tone 3) /ma/ means "horse," and a falling tone (tone 4) /ma/ means "to scold." Throughout this dissertation, I will be reporting the monosyllable-tone combinations in the following way: *ma1* to mean the Chinese pinyin *ma*, accompanied by tone 1. In the instance that a syllable-tone combination is not a real word in Mandarin, an asterisk will accompany that combination (i.e. *ca2**).

Chinese only has about 410 syllables, but with the addition of the 4-tone system, the total distinctive monosyllable pool in Chinese rises to upwards of 1640 syllable-tone combinations. However, not every syllable-tone combination in Chinese exists today, and the combination of syllable-tones that do not exist in the language are called **accidental gaps**. These are phonotactically sound words, but do not exist with 1 or more of the 4 tones in Chinese. An example of a syllable with an accidental gap is *zhen*. *Zhen1* ('real,' 'needle,' etc), *zhen3* ('rash,' 'examination,' etc), and *zhen4* ('town,' 'battle position,' etc) all exist, but *zhen2** does not contain any words, making *zhen2** an accidental gap. Accidental gaps like *zhen2** will come in handy for the first set of experiments, which explores the role tones and segments play in Mandarin lexical access.

It’s important to note that while most of the syllable-tone combinations (monosyllables) mentioned so far are individual morphemes, a large portion of the modern Chinese lexicon consists of compound words - mostly disyllabic and longer combinations of monosyllables. In a review of the 3000 most frequently used words in Chinese, 27% were monosyllabic, 70% were disyllabic, and 3 or more syllable words made up around 3% (He & Li, 1987). Within this research, I will use monosyllabic morphemes as the word stimuli in the phonological perception study, as single syllables are easier to manipulate in lexical decision tasks, and match the stimuli of experiment the study is patterned off of.

Although Mandarin is the official language of the People’s Republic of China, another commonly spoken Chinese language family is Cantonese, heavily spoken in the southern parts of Mainland China, Hong Kong and Macau. Cantonese, like Mandarin, uses Chinese characters, but has a distinct inventory of initials and finals, as well as its own distinct tone system. Instead of 4 tones, Cantonese has 6 tones; and instead of codas being limited to the Mandarin nasals *n* and *ŋ*, Cantonese codas also include *m*, *t*, *p* and *k*. Although this dissertation primarily investigates the phonological processing and production of Mandarin, Cantonese will also be briefly mentioned, along with some other tonal languages, as points of comparison and contrast with the non-tonal languages I will introduce and discuss.

Mandarin is only one of the over 60 percent of the world’s languages that are tonal (Yip, 2002), but there still remains a lack of understanding among psycholinguists regarding the actual role tone plays in the perception and production of spoken language. When phonological perception and production were first studied by linguists, most studies were conducted on Indo-European languages. The languages from these studies do not make use of lexical tone, so studies of this nature mainly investigated consonants and vowels, and how the two differed in contributing to lexical access and during speech production. During this time, dozens of theories and models of spoken language perception and production were made based on the foundation of these studies, the majority of which do not mention tone at all. In the past 30 years, there has been a sincere effort to replicate the aforementioned studies on tonal languages, to try to understand where tone fits within the existing theories of spoken language.

However, there is a great deal that remains unknown about tones. Due to the nature of Mandarin's tonal system, as well as its phonetically opaque logographic system consisting mainly of monosyllabic morphemes, there has long been debate over whether segments and suprasegmentals, such as tone, behave the same way in Mandarin as they do in non-tonal languages and in languages without alphabetic writing systems.

As mentioned before, there is some research showing that Chinese consonants and vowels may not behave similarly to consonants and vowels in Indo-European languages. Most of the focus has been due to Chinese as a logographic orthography, its lexical tonal system, its relatively small syllable inventory, and its initial and final internal structure. However, these differences may or may not impact the way Chinese behaves in a predictable, monolithic way. This dissertation aims to address some of the common assumptions about how tones work and provide new evidence that challenges those assumptions.

1.1: Tones in Phonological Perception

Phonological processing reflects the brain's ability to identify sounds in a language and make use of the sounds in comprehension. This process requires a listener's correct discrimination of different sounds, some of which are acoustically similar. This identification could involve anything from a segment-sized sound to full sentences in the language. For understanding spoken language, this is an integral component of the equation, and understanding how this process works has been a focus of psycholinguistics.

For example, when an English speaker hears /s/, they activate all words that begin with /s/, such as *sand*, *soft*, *said*, etc. As the word unfolds, and they hear /sæ/, they can eliminate *soft* and *said*, because these words do not start with /sæ/. *Sand* would still be in contention, as well as other words that begin with /sæ/. It is through this process by which a listener reaches the proper identification of the words in their language. In a language like Chinese, which has both segments (such as consonants and vowels), and suprasegmentals (such as lexical tones), how segments and tones contribute to this proper identification of words is still up for debate.

Early research using classification of phonemes in same-different and lexical decision tasks (Cutler & Chen, 1997; Taft & Chen, 1992; Repp & Lin, 1990) show that for Chinese speakers, discriminating different tones is consistently more difficult than discriminating different segments. This led to an early conclusion among linguists that tone is processed in a different way, and contributes less to **lexical access**, the process by which a listener retrieves lexical items for a heard spoken utterances.

In recent years, the proliferation of neuroimaging and EEG technology have given researchers an additional way to study phoneme processing, and research in this area has shown that although tonal information tends to contribute less to the correct identification of a word when compared to segmental information, tone information is used as soon as it becomes available to constrain lexical access (Malins & Joanisse, 2012; Zhao et al., 2011, Schirmer et al., 2005). Eye-tracking has shown that tone information actually takes priority over vowel information in a recent study (Deng et al., 2022).

As you can see, over time, the research has taken a different direction, showing evidence for a more nuanced distinction between tone and segment contribution to lexical access. However, the majority of the above studies compared tones and segments not as simple abstractions of a single tone or a simple segment, but rather tones, onsets, and rimes/finals, the last of which we know can be up to 4 segments in Mandarin (please refer to Figure 1). So far, tones have been compared mainly against onsets and rimes, which has shown how individual initial consonants and (sometimes) individual vowels behave, but to date, no research has focused on final consonants, also known as codas. As a fully functioning category of segment in Mandarin that contributes to lexical access and speech perception, codas should be considered a segmental category separate from onsets and vowels. It is my intention to abstract the individual components (segments) of a syllable and compare them individually with tones, to see if tones truly behave differently than every other segment. If tone behavior is entirely different than that of all segments, including onsets, vowels, and codas; then I will accept that as evidence that tone does behave in a qualitatively different way.

The method I will use as a starting point is the **word reconstruction task** (van Ooijen, 1996), a task in which listeners hear a nonword stimulus that differs from a real word by a single segment. The listeners are then asked to change either a consonant or vowel to turn that stimulus into a real word. For example, one critical stimulus was *kebra*, which can be turned into a word by switching a consonant (/k/ to /z/) to make *zebra*, or by switching a vowel (/i/ to /o/) to make *cobra*. This task allows us to know if listeners give priority to consonants or vowels when processing spoken words.

Linguists know that there are fundamental differences between consonants and vowels: consonants involve more constriction of the vocal and nasal tracts than vowels, and therefore consonants aren’t usually syllabic, whereas syllables usually are. As a consequence of consonants involving more constriction of the vocal and nasal tracts, consonants tend to be shorter in duration and consonants tend to outnumber vowels in languages. Sometimes the ratio is very unevenly skewed, such as the case in Castilian Spanish (20 consonants to 5 vowels; Maddieson, 1984; Stockwell & Bowen, 1965); sometimes the ratio is more evenly balanced, as in Dutch (19 consonants and 16 vowels; Booij, 1995). However, how these differences contribute to the correct identification and discrimination of words remained relatively unknown until the word reconstruction task.

The results of the word reconstruction task showed that despite vowels being smaller in number in the languages, and lasting a longer time within words, vowels were found to play a weaker role in lexical selection than consonants do. In this and similar tasks, vowels were changed more frequently, more quickly and more accurately than consonants.

As the task was done on Indo-European languages only (Van Ooijen, 1996; Cutler, Sebastian-Galles, Soler-Vilageliu, Van Ooijen, 2000), not much was known about how the task fares with a lexical tonal language, such as Mandarin. To this end, Wiener and Turnbull (2015) replicated the study with Mandarin, and manipulated the stimuli so that participants could choose to switch either the consonant, vowel or tone to switch the nonword into a real word.

The results were inconsistent with the results found from Indo-European languages: with Mandarin, participants chose to switch the tone most frequently, followed by the consonant, with vowels remaining most resistant to change. Although there were no prior assumptions about whether tone should be more or less resistant to change than segments, the results of consonant and vowel change contradicted the results of previous word reconstruction task studies, which were conducted on Indo-European languages (Van Ooijen, 1996; Cutler, Sebastian-Galles, Soler-Vilageliu, Van Ooijen, 2000). Due to this, Wiener and Turnbull argued that consonants and vowels may also behave differently in tonal languages than in the previously tested non-tonal languages. Additionally, the overwhelming tone change in the study as compared to consonants and vowels prompted Wiener and Turnbull to conclude that tone behaves differently from segments in spoken word recognition.

These results beg the question: are tones inherently processed differently from segments in spoken word recognition, as suggested by this previous study? Additionally, are vowels processed differently in Mandarin compared to the previously studied Indo-European languages? In chapter 2, I will argue that the task used by Wiener & Turnbull (2015) was unable to capture the way that vowels are processed in Mandarin. Rather than asking participants to switch either a consonant or a vowel, Wiener & Turnbull instead asked them to switch either an onset or a final (in this case, a phoneme chain containing up to 4 segments). Due to this change, individual segments could not be broken down and analyzed individually against tones. Thus, the results do not provide strong support for the idea that tone is inherently processed differently from segments as a whole. Instead, the task should be changed to bypass the necessity of asking participants to switch a consonant or vowel, because such concepts are not readily accessible for most speakers of Mandarin.

Additionally, Wiener & Turnbull (2015), as well as prior studies in Mandarin speech processing, approach the tone vs. segment issue by comparing tone with consonants and vowels. However, analyzing tone vs. segment using a different method might be more apt, due to the following two reasons. One, because tone has a time-bound position in a syllable (it is typically carried over the nucleus or final), it would be more apt to compare tone to segments in their time-bound positions within a syllable: onsets, vowels, and codas. Two, the inventory size of a phonemic category, not whether it's segmental or suprasegmental, may be the real reason behind different speech processing behaviors. In Mandarin, although theoretical debates surrounding the nature of underlying phonemes muddies the debate on the actual number of phonemes in different categories, the 3 categories of segments can be viewed like this: onsets have upwards of 20+ items in its category, vowels have anywhere from 5 to 22 items in its category, whereas codas are limited to two items: /n/ and /ŋ/. Compared to tonal information, which has 4 items, codas are the one segmental category which has fewer items compared to tones. If inventory size of a phonemic category truly has a negligible impact on acceptance of its variability in spoken language processing, comparing coda with tone variability acceptance will provide a definitive answer to this question.

The first part of this dissertation continues the work done previously by Wiener & Turnbull (2015), using much of the same stimuli but introducing a new behavioral paradigm that improves upon limitations in the implementation of the traditional word reconstruction task in Chinese. In the end, my goal is to see if a spoken word recognition task suited for Mandarin will clearly tell us if tone is processed in a qualitatively different way than segments.

1.2: Tones in Phonological Encoding

The other major aspect of spoken language is speech production, a process that allows speakers to articulate their ideas aloud. Speech production is normally viewed as a three-step process: first, the speaker conceptualizes what they plan to say; second, the speaker generates the linguistic code that expresses the concept(s) they prepared; and lastly, the code is articulated through a flow of air from the lungs to the oral and nasal cavities. For the purposes of this dissertation, I am focusing on the second part of this process, called **phonological encoding** or word form encoding (Schiller, 2006), a set of processes by which the brain generates the form of an utterance based on semantic and syntactic information (Meyer, 1992). The focus of this dissertation will be this process, the process by which the brain pieces together phonemes and the code for speech articulation.

Because this process cannot be observed directly, linguists have primarily relied on behavioral evidence to form theories and models of the speech planning process. One source of this evidence lies in controlled behavioral studies, using tasks that prompts participants to produce specific utterances, usually facilitated by implicit priming and picture-word interference contexts. An early study of Mandarin speech production involving implicit priming (Chen et al., 2002) showed that priming only occurred when entire segmental syllables were identical between the prime and the target. As they did not see priming occur when primes and targets shared only an initial consonant (onset) or tone, Chen et al. argued that the syllable (without tone) is an important phonological unit at the speech planning level for Mandarin.

Another way to investigate how similarity of one word may facilitate or inhibit the speed of uttering another word is through the picture-word interference task (PWI). In the PWI, participants utter the names of pictures while ignoring distractor words that appear at the same time. The PWI is another task that linguists have frequently used to investigate the speech planning process. Contrary to the results reported by Chen et al. (2002), a picture-word interference task conducted in Cantonese (Wong & Chen, 2008) showed that facilitation effects were found when the distractor item shared the entire syllable, the same onset and vowel (with or without the same tone), or the same vowel and coda (with or without the same tone). However, facilitation effects were not found when the distractor item only shared a vowel and a tone. Thus, Wong and Chen asserted that the syllable in Chinese does not warrant a special status in speech planning above and beyond a combination of their individual parts. Nonetheless, both Chen et al. (2002) and Wong and Chen (2008) argue that tones are not processed or represented in the same way as segments. Based on their studies, tone did not contribute to speech facilitation, unlike the way that either whole syllables or combinations of segments did.

Another source of behavioral evidence that allows us to make hypotheses about the speech planning process is speech errors: unplanned deviations from the speaker's intended targets, which can vary in size, from (arguably) features to entire syllables or words. The existence of speech errors gives credible support for the idea that speech planning happens in stages: that is, when a person intends to say the phrase “Santa comes through the chimney,” the entire phrase does not enter the encoding process at the same time. With speech errors such as “Kanta comes,” we can see that the word “Santa” is not encoded completely before “comes" is encoded - this intrusion of the initial segment in the previous word shows us that there is an ***incremental*** activation of the phonemes within an intended utterance. This means that phonemes are incrementally activated from one word to the next, rather than being contained entirely within word boundaries. Thus, the context of the intended utterance (the fact that a subsequent word intruded upon the utterance of “Santa,” a type of error known as *anticipation*) plays a factor in phonological encoding.

With the knowledge that contextual factors affect segmental encoding, there has been increased interest in how tones in tonal languages like Mandarin are encoded, and whether or not they are subject to the same contextual errors. In addition to consonant or vowel swapping as found in non-tonal languages, tone adds a layer of complexity, as it is a suprasegmental feature. Researchers have long been interested in how rates of tone error (a suprasegmental feature) compared to consonant and vowel errors (segmental features). As with non-tonal languages, the majority of studies investigating the encoding of tones primarily uses speech error corpora, annotated by the interested parties for the sake of finding the different types of speech slips in spontaneous speech.

The first of these studies, conducted by Wan and Jaeger (1998), looked at 788 speech errors in Mandarin and concluded that tone errors were relatively common in comparison to segments, and exhibited the same kinds of errors modulated by context, such as perseverations, anticipations, exchanges, etc. Because these kinds of behaviors are similar to those of segments, this suggests that tones are **encoded actively, or early**, in the speech planning process.

An alternative theory of tonal encoding came from the results of another large-scale Mandarin speech error study, conducted by Chen (1999). In Chen's study of 987 speech errors, only 24 were deemed to be tonal, and he argued that the tone errors previously cited as movement errors by other linguists (Wan & Jaeger, 1998; Gandour, 1977; among others) were in fact errors of a different nature, such character blending, haplology, malapropism, and misapplication of tone sandhi. Thus, he argues that tonal errors do not behave the same way as do segmental errors, being that they typically are not modulated by context (perseverations, anticipations, exchanges, etc). With these discoveries, Chen concluded that tone is not actively selected in word-form retrieval, but rather that tone is a part of the word form structure, selected late in the encoding process, after segments have already been selected. According to Chen, **late encoding** of tones explains the lack of tone errors.

However, speech errors in spontaneous speech can be difficult both to identify and categorize properly, and fortunately there is a better way to find the speech error information needed to make proper theories and conclusions about the nature of tone encoding. In this dissertation, I take an alternative approach to analyze speech errors. Instead of analyzing spontaneous speech errors from recordings of casual conversation, I created a tongue twister task that is designed to elicit equal amounts of tone and segment error. Similar to the perception study earlier, I break segments down into 3 individual categories: onset, vowel, and coda. The tongue twisters were designed to elicit tone error, error in one of the 3 segmental categories, or a combination of both. As I mentioned earlier, the inventory size of the phonemic category may affect its behavior in speech perception and production, so the breakdown of segmental categories allows us to come closer to understanding how different segmental categories compare to tonal behavior in speech production.

I discuss in-depth in chapter 3 why this tongue twister paradigm is an improvement over not only the methods discussed thus far, but also over existing tongue twister studies.

Through this new tongue twister study, I hope to provide strong evidence showing a nuanced portrayal of tone error in comparison to segment error in speech. To this end, I hope to contribute to the existing debate: if tone errors really are much rarer than segment error, including onsets, vowels and codas viewed individually; then I will conclude that tone is not actively encoded. Conversely, if tone errors are produced at a similar rate as segments, then I will conclude that tone is actively encoded.

1.3: The Structure of this Dissertation

This paper uses psycholinguistic and behavioral methods to investigate the perception and production of spoken Mandarin, comparing the behavior of segments and tones. The first section, consisting of four experiments, uses a forced-choice word selection task that tests the acceptance of tonal versus segmental variability in monosyllabic words. Data from these four experiments shows that tone's contribution to lexical access is nuanced: it contributes less compared to onset and vowel, but it contributes more compared to coda. In a word, tone's contribution to lexical access is the same as that of segments - the difference is that previous research only analyzed the positions where segments contribute more information, whereas in my studies, all positions of segments are analyzed in comparison to tonal information, including the positions of segments that contribute less information than tones.

The second half of the dissertation focuses on the phonological encoding process that occurs when Mandarin speakers speak aloud. This section uses a tongue twister task that elicits an equal amount of errors in tones, onsets, vowels and codas, and the errors were analyzed for their frequency in comparison to those of the other categories. The results show that codas had the highest error rate, followed by onsets and tones with similar error rates, and with vowels having the lowest error rates. Based on these results, there is strong evidence to show that tones are not encoded qualitatively differently compared to segments.

Through the careful examination of tones and segments in both speech perception and production tasks, I show that tones behave similarly to segments once segments are not viewed as a monolith of equal features and elements, but rather once different classes of segmental phonemes are considered separately.

CHAPTER 2: TONAL PERCEPTION

2.1: Background

2.1.1: Introduction

Speech perception is the process by which the sounds of a language are heard, processed, and used to interpret meaning. Psycholinguistics research in the domain mainly aims to figure out how listeners recognize speech sounds and use this information to comprehend the sounds in a meaningful way. Classically, the speech perception model is viewed in the following way:

1. The sound signal is transferred from speaker to listener.
2. The speech sounds are processed: acoustic cues and phonetic information are extracted.
3. The phonetic information is then used for higher-level meaning processing: word recognition, phrase recognition, sentence comprehension.

However, the way that this process actually occurs within the listeners' minds is still up for debate. Language comprehension is a difficult process, and spoken word recognition requires the listener to intake speech and map the sounds onto words, a process that unfolds over time as the speech signal is heard. An important part of the speech perception process is **lexical access**, the method by which individuals access and retrieve words in their mental lexicon in order to comprehend or produce a word. Because every language has words that sound similar to each other, the listener must have a mental system that is able to disentangle minimally similar words to arrive at the correct word. Given the fact that a listener has to overcome distractions, such as background noise, individual speaker differences, as well as regional and foreign accents, this system is thought to be as complex as it is efficient.

One of the first models proposed to explain spoken word recognition was the **cohort model** (Marslen-Wilson & Walsh, 1978), which proposed that the brain starts processing audio input as soon as the input is heard, instead of waiting until a word is finished to start processing. For example, in the recognition of the word “candle,” the first sound the listener hears is /k/. At this point, all words starting with /k/ are activated, but as the word unfolds with the subsequent vowel /ae/, words that don’t begin with /kae/ would be kicked out, leaving items such as “candy,” “canopy,” “cattle,” along with the intended word “candle.” Once the phoneme /n/ is added, “cattle” would be discarded from the equation, and once /d/ is added, only “candy” and “candle” would remain. Once we reach the final phoneme /l/, the word “candle” would finally be recognized.

Rivaling the cohort model is a model called **TRACE** (McClelland & Elman, 1986), a connectionist model of speech recognition that views speech recognition as a more interactive process. Contrary to the cohort model, which views speech recognition as a one-directional funneling process, TRACE views speech recognition as a process in which a phoneme sequence (the input) is simultaneously mapped onto 3 different layers: the feature layer, the phoneme layer, and the word layer. As the acoustic input unfolds over time, nodes on the feature, phoneme and word layers compete for recognition. As a feature node is selected, the other feature nodes on the layer are inhibited, and that information is relayed to the corresponding phoneme node on the upper-level phoneme layer, where the same process occurs before reaching the word layer.

What sets the TRACE model apart is its ability for feedback from higher levels (word, phoneme) to influence lower levels (phoneme, feature) by using a listener’s lexical awareness. This model takes into account speaker differences, such as foreign accents; as well as external acoustic influences, such as background noise. For example, the phonemic sequence a listener hears might have been “gat,” but once the sequence is processed at the word layer, the listener recognizes that “gat” is probably not the word the speaker intended to say, the word layer will relay feedback back down to the phoneme layer and correct the /g/ to a /k/, and further down to the feature layer and correct the voiced phoneme to an unvoiced one. Because the TRACE model has this distinctive feedback feature, it is able to explain phenomena such as the Ganong effect, in which an ambiguous spoken sound is usually perceived as one that completes a real word and not a nonword (Ganong, 1980). For example, a listener may have difficulty hearing a sound that is approximately halfway between /t/ and /d/ - identical sounds save that the former is unvoiced and the latter is voiced. However, when the ambiguous sound is placed before /utor/, listeners overwhelmingly perceive the ambiguous sound as /t/, due to “tutor” being a real word and “dutor” being a nonword.

A more recent model of speech recognition is the **Neighborhood Activation Model** (NAM) (Luce & Pisoni, 1998). Like the previous two models, NAM believes that lexical selection is a process by which various phonological competitors must be incrementally eliminated before arriving at the correct lexical entry of an uttered speech sequence. However, NAM believes that factors such as the amount of phonologically similar words to the uttered speech sequence (neighborhood density), as well as the frequency of the lexical entry and that of its neighborhood items, influence a listener’s ability to discriminate the word correctly and quickly. Specifically, NAM believes that competitors to the word are all words that diverge from the word in one phoneme - for example, neighbors of *cap* would include *cop*, *cape*, and *clap*, as well as rhymes, such as *tap*, *zap*, and *map*.

Due to this, NAM does not account for the temporal course of processing words, which some view as a limitation (Zhao et al., 2011). In contrast, a model of speech recognition that strongly accounts for the temporal course of processing words is the aforementioned cohort model (Marslen-Wilson & Walsh, 1978), which eliminates lexical competitors as the speech sequence unfolds, starting from words that diverge at the onset.

However, Luce & Pisoni (1998) conducted a series of auditory word recognition experiments that yielded results that countered the strict temporal competition model espoused by Marslen-Wilson and Walsh, leading them to arrive at their own model, one that takes into account neighborhood density, as well as neighborhood frequency.

All aforementioned models have been very influential in informing how we think about speech perception and lexical access, with TRACE even being implemented in computation models, such as using C and JavaScript. Overall, it seems that all above models offer convincing arguments that support their version of how speech recognition occurs, but with none being completely fool-proof to counterarguments.

In the next section, I will discuss behavioral experiments that take the premise of these speech recognition models to examine the role different phonemes play in the recognition of different languages.

2.1.2: Behavioral Experimental Research on Speech Perception

2.1.2a: The Lexical Decision Task

Since the 1970s, a popular method of testing lexical access has been the **lexical decision task** (Meyer & Schvaneveldt, 1971). The task has a simple procedure: present participants with word-like items and ask them to classify the items as words or non-words. Accurate lexical decision would be the correct identification of the stimulus being a word (if the stimulus is a word) or a non-word (if the stimulus is a non-word). The lexical decision task is commonly paired with another experimental technique known as **priming**, in which exposure to one stimulus impacts one's response to a subsequent stimulus. In the realm of linguistics, priming has commonly been used to investigate orthographic and phonological activation in word recognition. Along with priming, the lexical decision task has often been used to investigate semantic similarity between items.

The two data points that can be derived from the lexical decision task are 1) accuracy of the lexical decision, and 2) the reaction time – the time it took the participant to make the choice. Items that are very clearly words and items that are very clearly not words in the language tested generally receive the highest degree of accuracy and the fastest reaction times (Meyer & Schvaneveldt, 1971). An example of a clearly non-word item could be "sfjsbf" in English; although English allows for more complicated consonant clusters than most languages, "sf," "js" and "bf" are not acceptable consonant clusters in the language – these items thus elicit very quick responses of "no."

The items that generally take longer and have less accuracy are items that do not break the phonotactic rules of that language, ones that could technically exist. An example of this might be the item "blick" – the phonemes are phonotactically supported by the English language (/bl/ is a common consonant cluster at the beginning of words, and /ick/ ends many words; there are also many words that have the same phoneme sequence, minus the first phoneme - "slick," "click," etc.), but for some reason that combination of phonemes does not currently exist as a word. Speakers of the language have to go through their mental lexicon and rule out similar-sounding words (including the aforementioned "slick" and "click") before deciding that this word does not exist. As a result, this category of items elicits more hesitation, longer reaction times and less accuracy rates.

Thus, this simple word recognition task can be used to explore a wide variety of underlying language processing systems. Originally devised as a written task (with stimuli presented visually on a computer screen), the lexical decision task has also been used as an auditory task (with stimuli presented auditorily through headphones or speakers). Unlike the written task, which investigates the reading process, the auditory task studies speech perception, the language process we're interested in. Similar to the written task, in which we can investigate how participants react differently to nonwords of different types, the auditory lexical decision task can shed light on how nonwords that are nonwords in specific ways may lead participants to react faster or slower in their judgment of whether the combination of sounds they just heard is a word or nonword. For example, would a nonword that is a nonword based on an incorrect tone or an incorrect segment cause more difficulty in judging lexicality? This is a question that can be answered by the oral lexical decision task.

In the next section, I will talk about how the auditory presentation of the lexical decision task can be used to help us understand how tones and segments are perceived in Cantonese, a Chinese language with 6 tones, a larger tone inventory as compared to that of Mandarin.

2.1.2b: Behavioral Experiments on Cantonese Tone Perception

One use of the lexical decision task to test the processing differences between tones and segments was the work of Cutler & Chen (1997). Based on previous work showing that in Cantonese, the processing and correct identification of tones seem to cause difficulties for native speakers when listening to spoken full sentences (Tsang & Hoosain, 1979), Cutler and Chen wanted to use the lexical decision task to verify the results in disyllabic, single-word stimuli.

According to Cantonese corpus *Jyutdin*, 66%[1] of the words in the Cantonese lexicon are disyllabic (*Jyutdin*, 2023), meaning that the majority of Chinese words are a pairing of two monosyllabic characters. A monosyllabic word such as *fo2* ("fire") can combine with the monosyllabic word *gai1* ("chicken") to create *fo2 gai1* ("turkey"). The Cutler and Chen study

[1] Out of 34508 words in the Cantonese corpus, 22733 are disyllabic.

employed this disyllabic word structure to create non-words by altering the second syllable of each word.

Cutler and Chen used the auditory lexical decision task to examine native Cantonese speakers' accuracy responses and reaction time to spoken disyllabic Cantonese non-words. The critical stimuli took disyllabic Cantonese words, kept the 1st syllable intact and altered the 2nd syllable, turning the word into nonwords that erred in either tone, onset, or vowel, or several combinations involving two or more of these categories. One example would be the word 博士 *bok3 si6* ("doctorate"), with the tone nonword being *bok3 si2**, the vowel nonword being *bok3 sy6**, onset nonword being *bok3 ji6**, and combinations of tone, vowel and onset alterations that resulted in other nonwords: vowel-tone nonword *bok3 sy2**, onset-tone nonword *bok3 ji2**, onset-vowel nonword *bok3 jy6**, and onset-vowel-tone nonword *bok3 jy2**. Of course, real Cantonese words were also included as controls. As with all lexical decision tasks, participants were asked to decide if the spoken stimuli presented were words or nonwords.

The results showed that the reaction time differences of the selections were statistically insignificant across conditions. However, in the critical conditions, when only the tone differed from a real word, the ***accuracy*** rate was the lowest. The second lowest accuracy rates were when only the vowel differed. Based on these results, it seems to suggest that tone contributes to lexical access the least in Cantonese, followed by vowels.

However, Cutler and Chen then closely examined the Cantonese tone space and found a potential reason for the tone-only nonword items' high error rates. The Cantonese tone space has 6 categories, in which Tone 1, a tone that starts high and dips toward the latter half of the syllable, is substantially different from the other 5, which all start at considerably lower, similar F0s. Tones 2, 3, 4, 5 and 6 start with relatively similar F0s and do not differ significantly from each other until the latter part of the syllable. This means that Tone 1 is different from the other 5 tones during the first half of the syllable, and is thus easier to differentiate, if it appears in a nonword in which the intended tone of the real word is tones 2, 3, 4, 5 or 6.

As the task asks for participants to make a decision on whether the presented stimulus is a real word or nonword, and participants were encouraged to decide as soon as possible, the late differentiation of tones 2-6 would have made the classification of those stimuli more difficult, as they might have classified the stimuli as a real word before they heard the tone difference, according to Cutler and Chen. This means that items that involve Tone 1 and one of the other 5 tones would be objectively easier to tease apart compared to items involving Tones 2, 3, 4, 5 and/or 6, due to the greater acoustic differences in the former category and fewer acoustic differences in the latter category.

An example of an easy distinction would be between the word 茶杯 *tsa4 bui1* (“teacup”), whose second syllable is Tone 1, and its nonword counterpart, *tsa4 bui3**. *Tsa4 bui3**, in which the second syllable is Tone 3, would be relatively easy to discriminate against. As mentioned earlier, Tone 1 starts at an F0 substantially higher than that of tones 2-6, so *tsa4 bui3**, at Tone 3, would be relatively easy to distinguish from 茶杯 *tsa4 bui1* at Tone 1. An example of a difficult distinction would be the aforementioned 博士 *bok3 si6* (“doctorate”), whose tone-changed nonword stimuli was *bok6 si2**. As the real word ends in Tone 6 and the nonword tonal-only stimuli ends in Tone 2, two tones that start with low F0, the discrimination of the two tones would be relatively difficult.

Thus, Cutler and Chen then looked at only the nonword tone items that involved Tone 1 stimuli and whose real word counterpart was one of the other five tones (and vice versa: stimuli involving one of the five other tones and whose real word counterpart was Tone 1) and found that the error rates were very similar to the error rates of the onset-only difference items. The high error for tone-only items thus primarily came from the items involving discrepancies between Tones 2, 3, 4, 5 and/or 6. From this analysis, they concluded that the results of the high level of error when only tone differed might actually be an artifact of how late the acoustic differences between 5 of the 6 tones emerge in Cantonese, rather than reflecting the possibility that tone does not contribute substantially to lexical access.

From this experiment, it seems clear that the difficulty of correctly identifying the tone-only difference items was a result of the fact that the tone differences themselves were objectively difficult and take a longer time to perceive, as compared to segments. However, the studies do not make statements about the intrinsic phonological qualities of tones themselves, such as their contribution to lexical access.

Several other studies throughout the 1990s, using similar same-different judgment tasks or phoneme categorization[2] tasks, found similar results for Mandarin Chinese (Taft & Chen, 1992; Repp & Lin, 1990), showing both slowed response times for tones as compared to segments, in both same-different judgment and phoneme categorization tasks. In particular, Repp and Lin (1990) tested Tones 1 and 4 in Mandarin, which start at the same F0 and only start differing in contour in the second half of the tone, so they acknowledge that the slowed tonal categorization responses may have been an artifact of later tonal differential information, similar to the results from Cutler and Chen (1997). Although these studies show that tonal differentiation is difficult in acoustically similar tones, they don't provide evidence that shows tones are perceived differently from segments.

Next, I will be looking at how tone and segment priming, a commonly-used psycholinguistic task, informs our understanding of how tone and segment may be perceived differently.

[2] Phoneme categorization tasks ask participants to categorize a given phoneme sequence (e.g. CV-structure syllable) and classify it by different types of phonemes, such as by consonant, vowel, or tone.

2.1.2c: Priming Research on Mandarin Spoken Word Perception

Priming, a task commonly used to show how one item facilitates or inhibits recognition of a similar subsequent item, is often used in linguistics to show how certain linguistic features are processed in the mind. As mentioned before, priming is often used with the aforementioned lexical decision task to glean an additional level of information - whether the first item facilitates or inhibits recognition of the second item may tell us more about the shared (or diverging) phonemic information between the two. Sereno & Lee (2015) used this paradigm to test whether or not tones are processed differently from segments in a spoken single-word priming task.

Sereno & Lee created a study that contrasted 4 types of prime-target pairs: identical (*ru4-ru4*), segment-only overlap (*ru**3**-ru4*), tone-only overlap (***sha**4-ru4*), and unrelated (*qin1-ru4*). Participants hear the first word (*ru3*), and then they hear a second word (*ru**4***) and are asked to judge if the second word (*ru**4***) is a real word or not. In this task, the critical data in the priming task is the difference between the reaction times of the related conditions and unrelated conditions - the bigger the difference between the reaction times of the related conditions and the unrelated conditions, the more the target item was primed by the previous word in the related condition. However, in this case, Sereno and Lee wanted to use this paradigm to investigate if tonal information and segmental information would be used immediately to block incorrect lexical candidates in the same manner. According to Sereno and Lee, if tonal and segmental information were equally important to lexical access in Mandarin, segment-only overlap and tone-only overlap items would show similar amounts of priming. However, if tonal information was less crucial to lexical access (which is what they predicted), segment-only overlap items (*ru3-ru**4***) would show more priming compared to those of tone-only overlap items (*sha4-**ru**4*). The logic is that when someone hears *ru3*, they activate *ru* in all 4 tones: *ru1*, *ru2*, *ru3*, and ru4, so when they hear the target word, *ru4*, they would respond to it faster, as it has already been activated. This means that the mismatching tone information in *ru3*-***ru4*** would not lead to substantial issues in lexical access. On the converse, when a participant hears *sha4*, they would not activate all monosyllabic morphemes with Tone 4, so when they hear the target *ru4*, they would not respond to it faster.

Their results showed strong priming effects when both segments and tones overlapped, and to a lesser extent, when only segmental information overlapped, and actually inhibition when only tones overlapped. Because segment-only overlap (e.g., *ru3*-***ru4***) still showed priming/facilitation relative to the unrelated condition (e.g., *qin1*-***ru4***), Sereno and Lee concluded that despite incorrect tone information, segmental information was effective in facilitating lexical access to spoken word recognition. Conversely, when only tone was overlapped (e.g., *sha4*-***ru4***), the reaction speeds of correctly responding to the second item were actually inhibited - the opposite of priming. These results led Sereno and Lee to conclude that although tonal mismatch is not detrimental to native speakers' recognition of a word, segmental mismatch is. In other words, the results suggest that tonal information is not key to facilitating lexical access, while segmental information is.

They followed up this first study to see which specific segmental mismatches between the prime and the target would result in more or less priming. For this experiment, they kept the tone identical between all critical prime-target items (so there was no tone mismatch), and created 4 new types of prime-target pairs: identical (segments and tone matching), onset & tone overlap (with rime mismatching), coda & tone overlap (with onset mismatching), and unrelated (with segments and tone mismatching). They repeated the experiment procedure on a new set of participants and found that although the identical items showed priming over the unrelated items, neither partial segment mismatching items showed priming, with onset & tone overlap (rime mismatching) showing some inhibition. This inhibition showed that any amount of partial segment mismatch was used by participants to block incorrect lexical candidates, unlike the way that tone mismatch did not seem to prevent priming from occurring.

Due to these results, the conclusion that Sereno and Lee came to at the end of these two experiments is that segments play a strong role in facilitating lexical access, whereas tones play a weak, secondary role. Combined with the results from the previous lexical decision and same-difference tasks, there is a large amount of evidence from behavioral studies showing that segments play a much stronger role in Mandarin speech recognition compared to tones.

However, priming, lexical decision and same-different judgment tasks are not the only way we can examine phonological processing, so in the next section, I will summarize some recent neurophysiological research on tone perception, and see if the results there can give us more insight on tone processing.

2.1.3: Eye-tracking and EEG Research on Chinese Spoken Word Perception

As we have seen with the above behavioral research on Chinese tone perception and processing, the time course of speech signal is crucial to the proper identification of tones. However, in many of the tasks, participants were asked to respond to the task as soon as possible, which may have clouded some of the results pertaining to the identification of tones. That’s why I believe that it is important to look into some tasks in which the fine-grained physiological and/or neurological behavior can be analyzed to glean more insight on the way tone is used to process language.

In the last 15-20 years, there has been a greater use of more time-sensitive measures, such as eye-tracking and event-related potentials (hereafter referred to as ERP) in comparison to the purely behavioral measures used in the past. Eye-tracking is often used to measure and investigate where a participant’s attention is focused on during an experiment involving written, oral, or multimedia stimuli. Eye-tracking is often recorded using an eye-tracker, which tracks where on a screen the eye focuses attention on for an extended period of time during the unfolding of stimuli. ERP, on the other hand, is accomplished by placing electrodes on the human scalp and recording brain waves during the time course of a sensory, motor, or cognitive event. The waveforms recorded therein are compared with well-known ERPs in order to analyze how the brain responds to those stimuli. The benefits of online measurements include greater access to fine-grained incremental behavioral and physiological data as a speech sound unfolds, as well as access to brain processes that simply cannot be seen in button presses resulting from lexical decision tasks, priming tasks, etc.

Starting with eye-tracking, a task often used by psycholinguists when conducting eye-tracking experiments involves the "visual world paradigm." In the visual word paradigm, a target word is played auditorily, and the target word, along with competitor items, are displayed visually on a screen, which participants are instructed to freely fixate on. Often, the competitor words are phonologically related to the target word, but vary in the positions in which they diverge from the target word. For example, Allopenna et al. (1998) used the visual world paradigm in the following way: one of the pictures was of the target word (e.g., *beaker*), and competitors shared the onset and initial vowel (e.g., *beetle*) or rimes (e.g., *speaker*). In the beginning of the trial, participants produced eye movements to both the target word and the onset & vowel competitor, as the segments in the beginning part of the word are identical. As the trial went on, participants produced eye movements to both the target word and the rime competitor, as the segments in the later part of the word became identical. From this study, we can see that phonological competition between the target and competitor items can be seen through the amount of fixations to both items.

Borrowing this paradigm, Malins and Joanisse (2010) decided to look at how Mandarin tonal and segmental cues contribute to recognition of a spoken word. Their study consisted of auditory monosyllabic words (e.g., *chuang2*, ‘bed’), with tonal mismatch competitors (e.g., *chuang1*, ‘window’) and segmental mismatch competitors - also called cohort competitors (e.g., *chuan2*, ‘boat’). They focused on the tonal mismatch competitors and segmental mismatch competitors because the words are identical in the first half, and only diverge in the second half of the duration - this setup is identical to the *beaker-beetle* study mentioned earlier (Allopenna et al., 1998). If these two competitors receive the same amount of fixations, that would be evidence supporting concurrent access of tonal and segmental information, as well as the idea that tonal and segmental information play similar roles in facilitating lexical access.

The results of the study support this idea, as eye movements to the target word were slowed by both tonal mismatch and segmental mismatch competitors, at similar rates. A similar study conducted using the visual word paradigm (Deng et al., 2022) also found that tonal information is used to distract from target words as much as segmental information does.

Moving onto ERP, studies on Chinese phonology of this nature usually involve a behavioral task, such as lexical decision tasks, paired with EEG, and measure both reaction times as well as the wavelengths in the brain resulting from the stimuli and task provided. Based on their later time course in a syllable (as compared to onsets and vowels), tones are shown to be processed as soon as they become available, giving further evidence to Cutler and Chen (1997)’s study that showed tonal discrimination differences were eliminated once the difficult-to-perceive tone pairs were eliminated.

One such study, Malins and Joanisse (2012), is based on the idea that tonal information being recognized later than segmental information doesn’t necessarily mean that the processes underlying tonal processing are different than those underlying segmental processing. Malins and Joanisse (2012) tested ERP components PMN and N400 on a simple picture-auditory word matching task with monosyllabic Mandarin words to investigate whether brain waveforms resulting from the task would show different patterns of activation between tonal and segmental mismatch.

PMN, or phonological mapping (mismatch) negativity, is an ERP component most commonly associated with phonological mismatch between an expected target and an actual target. It typically occurs in the range of 200-400 ms post stimulus onset. The N400, on the other hand, is an ERP component usually associated with semantic mismatch. Like its name suggests, it typically occurs around 400 ms post stimulus onset. Both are used in this experiment to see how Mandarin speakers use tonal and segmental information to process word recognition.

In this study, a participant would be shown a simple picture of a flower (*hua1*), followed by one of the following auditory words: *hua1* (identical word; match), *hua4* (tone mismatch), *hui1* (rime mismatch), *gua1* (onset mismatch), *jing1* (complete segmental mismatch), and *lang2* (complete mismatch). They would then have to decide if the picture they saw and the speech sound they heard were the same. The different types of stimuli are categorized by mismatch because this task focuses on phonetic mismatches between the picture and sound, whereas the previously mentioned Sereno and Lee (2015) study focused on phonetic overlaps between two sounds. Because the time course of brain activation can be easily accessed through the use of EEG data, the tone mismatch and rime mismatch items were analyzed in conjunction with identical match items as the baseline to analyze whether tone and segmental processing were different.

In their analysis, ERP components were time-locked to the onset of the auditory stimuli. Results showed that the tone mismatch items showed strong expectancy violation in the PMN window, whereas the rime mismatch items did not show a significant expectancy violation until the N400 window. As tone starts at the onset of the rime, with rime mismatch items possibly not showing divergence until the latter part of the rime (e.g., *hu**i**1* versus *hu**a**1*), Malins and Joanisse concluded that these results are in line with the idea that both tonal and segmental information are used to constrain lexical access as soon as they become available.

These results were replicated in another ERP study focusing on how mismatching tonal and segmental information impacts lexical access in Mandarin. Zhao, Guo, Zhou & Shu (2011) conducted a similar experiment in which written words and matching drawings were presented visually, followed by a spoken word that diverged from the written word and matching drawing in either onset mismatch, rime mismatch, tone mismatch or syllable mismatch (with an identical item for control). Then, they would be shown another drawing and asked to judge whether that drawing belonged to the same semantic category as that of the first drawing/written word.

An example of this would be the visual presentation of the first drawing/written word being *bi2* ("nose") with a drawing of a nose, followed immediately by the spoken word *bi2* (in the identical item) or *bi3* (in the tone mismatch item) or *bo2* (in the rime mismatch item), etc. After a 3000 ms gap, the visual presentation of another drawing - an ear - appears, and at this point, participants were asked to judge whether the ear in the second drawing belongs in the same semantic category as that of the first drawing. In this case, the answer would be yes, since they are both body parts. If the second drawing was that of a chair, the correct answer would be no.

According to Zhao et al. (2011), the additional drawing, accompanied by the task of judging whether the two drawings belong to the same semantic category, helps avoid the explicit matching task during spoken word recognition that previous picture-word tasks elicited. The results showed that spoken words with tone mismatch elicited comparable amplitudes and onsets of the N400 as spoken words with rime mismatch. Despite the subtle differences with the Malins and Joanisse (2012), which found that a later effect for rime than tone, both studies show that tone information is used as soon as it becomes available to constrain lexical access.

Additional ERP evidence showing that tones play a similar role as segments in word processing exists in Cantonese as well (Schirmer et al., 2005). In a study in which participants listened to either semantically correct sentences or semantically incorrect sentences with one word erring in either tone, segment, or both; Schirmer et al. found that both types of semantically incorrect sentences produced a N400-like negativity between 200 and 450 ms, showing no significant differences in waveforms between the two.

However, Shirmer et al. acknowledge that their study examined words within a sentence, which might have led to the participants' lexical knowledge of the context having an influence on their reactions to the incorrect word, leading to a semantic bias that may show equal amounts of processing given to both tones and segment. This effect was shown in an earlier study by Ye and Connine (1999), in which participants were asked to monitor for specific target tones or vowels in syllable sequences. When the syllables were presented in isolation, participants showed faster response times when asked to monitor for vowels as compared to tones, but when the syllables were presented in the context of an idiom, the effect disappeared.

Regardless of whether the semantic context effect truly exists, the above ERP studies show that a later processing time for tones does not necessarily mean that tones are processed in a qualitatively different way from segments, nor does it mean that tone information is secondary to segmental information in constraining lexical access. In fact, in many of the studies we've discussed so far, tonal information seems to be processed at approximately the same time as rime information.

However, the majority of studies seem to believe that tone and rimes are phonemic components with comparable features and structures, often comparing tone match/mismatch to rime match/mismatch. The issue with this is that the rime item is often not controlled for the number of phonemes, unlike how onset is controlled to be 1 phoneme (1 consonant at the beginning of a word). In the Deng et al. (2022) study above, some cohort competitors were only 1 phoneme different from the target word ("qian2" to "qiang2"), but some cohort competitors were 3 phonemes different from the target word ("kuang4" to "ke4"). However, this kind of comparison is not an accurate way to investigate tonal and segmental contribution to lexical access in Chinese. Investigating the different contribution of tones and segments to lexical access should compare individual tones with individual segments, not whole rimes.

To this end, the next section will introduce a new type of experimental paradigm not previously discussed, one that helps us get closer to how tones and segments may or may not provide different levels of contribution to Mandarin lexical access.

2.1.4: The Word Reconstruction Task & Wiener & Turnbull study

Another experimental paradigm that investigates how different phonemes contribute to lexical access is the "word reconstruction task" (van Ooijen, 1996). The original word reconstruction task tests English native speakers' sensitivity to consonant and vowel change in the processing of a word. In the word reconstruction task, participants hear a spoken nonword, and then are asked to choose either a consonant or vowel to change in the nonword to turn it into a real word. For example, one critical stimulus was *kebra*, which can be turned into a word by switching a consonant (/k/ to /z/) to make *zebra*, or by switching a vowel (/i/ to /oʊ/) to make *cobra*. There were two conditions for the task: the first condition forced the participants to change either a consonant *or* a vowel; and the second condition was a free choice condition, which allowed participants to switch any one segment of their choosing.

The accuracy of creating words from nonwords (sometimes the produced word were nonwords - these would count against the accuracy rating), the reaction speed of saying the new words aloud, as well as the ratio of choosing to switch either a consonant or a vowel in the free choice condition, were recorded. In this task, the roles of consonants and vowels are analyzed from the perspective of both **lexical access** (the retrieval of the lexical information appropriate for a heard auditory stimuli) and **lexical selection** (the selection of a specific word in a specific context). The first part of the task, in which participants hear an auditory nonword, they're using lexical access to find the item within their mental lexicon that corresponds to that audio. Next, they are to select a lexical item from their mental lexicon to match the condition they're instructed to follow – either by changing a consonant or vowel to change the nonword to a real word – this process uses lexical selection. Because the task's output is the participants' real word response to the nonword stimuli, the results can primarily be viewed from the perspective of lexical selection. Due to this, it is uncertain whether or not this task can provide information regarding consonants and vowels' contribution to lexical access.

The results showed that participants overwhelmingly chose to change the vowel over the consonant in the free choice condition. In the other condition, the forced vowel changes had much higher accuracy and faster reaction speeds compared to the forced consonant changes. These results led van Ooijen to conclude that English speakers treat vowels as more mutable than consonants, and that they constrain lexical selection less than consonants.

Although the word reconstruction task varies in methodology from the previously discussed studies studying lexical decision, discrimination, and priming, it provides substantial evidence arguing for the difference between two types of phonemes, consonants and vowels, in how they contribute to lexical access. By asking participants to listen to a nonword and switch either a consonant or vowel to turn the nonword into a real word, van

Ooijen created a task that shows whether speakers tend to disregard incorrect consonants or vowels more in a situation in which the acoustic sequence does not produce a lexical item. For example, in the *kebra* example shown above, the fact that participants overwhelmingly chose to switch the vowel (/i/ to /o/) shows that participants preferred to hear *kebra* as *k*br** over **e**a*, meaning that vowels /e/ and /a/ didn't constrain lexical access as much as did /k/ and /br/, the consonants in the nonword.

Because the results of this study provided strong evidence for the diminished importance of vowels in lexical selection, several follow-up studies replicated the experiment in several different languages, in attempts to rule out possible confounding factors that may explain the experimental data. For example, English has a skewed vowel-to-consonant ratio, with 8 vowels and 21 consonants. In addition, the vowel space in English is crowded, with many vowels similar to each other in terms of height and rounding. Thus, there were concerns over whether the overwhelming vowel change was a result of either the 1) smaller vowel inventory, leading to simpler lexical search; or 2) perceptual confusion stemming from English's crowded vowel space. Follow-up studies were conducted in Dutch, a language with a more balanced vowel-consonant ratio; and Spanish, a language with only 5 vowels that are highly distinct from each other. If either of these possible confounding factors interfered with the results from the original 1996 English study, these follow-up studies would easily provide evidence for them.

On the contrary, the results from both Dutch and Spanish (Cutler, Sebastian-Galles, Soler-Vilageliu, and van Ooijen, 2000) showed that vowels were changed more quickly and accurately in both the free choice and the forced condition. The results of these studies led Cutler and colleagues to conclude that the original conclusion – that vowels intrinsically are more mutable and contribute less to lexical access than consonants – was the most likely correct one. However, there are possible issues with this conclusion. According to Cutler et al., a smaller vowel inventory and a crowded vowel space might lead to more vowel change, whereas a bigger vowel inventory and a less crowded vowel space might lead to less vowel change. Unfortunately, the languages tested here, Dutch - which has a bigger vowel inventory but more crowded vowel inventory - and Spanish - which has a smaller vowel inventory but a less crowded vowel inventory - cancel each other out, based on this theory. Thus, it is more likely that the studies done in Spanish and Dutch end up similar to the study done in English. It is difficult to make conclusions about how the inventory size of a phoneme category leads to mutability acceptance of that category.

Moreover, this conclusion was made based on experimentation only with European languages and provided fertile ground for exploration in other language families. Subsequently, Wiener and Turnbull (2016) replicated the study in Mandarin Chinese, a language with lexical tone, in an attempt to see if vowels are more mutable than consonants in Chinese, as well as seeing how mutable tones would be. In addition to being presented with a spoken nonword, the native Mandarin-speaking participants were asked to change either a consonant, vowel, *or* tone, to turn the nonword into a real word[3]. The results showed that instead of being the most mutable phoneme category, vowels were the most stable category, with the least amount of selection in the free choice condition, and the slowest and least accurate reconstruction was in the forced vowel change condition. Consonants, which were the most stable category in the English, Dutch and Spanish studies, were the second most mutable category, with tone being the category with the quickest and most accurate changes. Wiener and Turnbull thus concluded that the universal vowel mutability hypothesis is not compatible with tonal languages, and that tonal information contributes least to lexical access, with vowels contributing most to lexical access.

[3] The experiment aimed to look at consonants, vowels and tones, but in execution, the category presented as "consonants" were actually onsets, and "vowels" were rimes. This is explained further in the next paragraph.

These results seem to mark a turn in our understanding of the universal vowel mutability hypothesis, but there are some important differences between the Wiener & Turnbull study and the original word reconstruction task, which complicate the interpretation of these results. As explained by Wiener and Turnbull in their Discussion section, modern Chinese education does not break phonemes down to consonants and vowels. When Chinese children start phonics education, they're taught that words are broken down into *sheng1mu3* (initials/onsets) and *yun4mu3* (finals/rimes). They are not taught about the concepts of consonants and vowels. Correspondingly, Wiener and Turnbull also asked the participants to change *sheng1mu3* (initials/onsets), *yun4mu3* (finals/rimes), or tone (Refer to Figure 1 in Chapter 1). This is where the Wiener and Turnbull implementation of the word reconstruction task breaks down.

In the original study, the key aspect of the word reconstruction task focuses on one-phoneme changes, either as a consonant or as a vowel. But because Chinese speakers typically do not have metalinguistic awareness of vowels as an independent phoneme category, *yun4mu3* (finals) was substituted for vowels, which means that participants were not actually asked to change a vowel, but instead were asked to potentially change a sequence of multiple segments. As we have seen in Figure 1, the *yun4mu3* in Mandarin includes the majority of the syllable in most cases, and thus, the majority of the word. The *yun4mu3* includes monophthongs, diphthongs (sometimes analyzed as a vowel with a glide), triphthongs (sometimes analyzed as diphthongs with approximants), as well a combination of these vowels attached with one of 2 nasal codas: /n/ and /ŋ/. Instead of a task that asks a participant to change a consonant or a vowel to turn a nonword into a word, the task here asks a participant to change either the initial consonant or the final, which could include up to 3 vowels (or 1 vowel and 2 glides, depending on interpretation) and a coda.

For example, one of the presented oral stimuli was _tian4*_. When asked to change either a _sheng1mu3_ (initial/onset), _yun4mu3_ (final/rime), or tone, the participant has the following options (partial list):

1. **Initial/onset change**: mian4 (“flour; noodles”), bian4 (“change, convenient, excrement”), dian4 (“electricity, store”), etc.
2. **Final/rime change**: ti**4 (“replace, shave”), tu**4 (“rabbit”), ta**4 (“step”), t*an4 (“carbon”), etc.[4]
3. **Tone change**: tian1 (“sky”), tian2 (“field”)

As we can see from the above example, final/rime change is the only category in which there is a possibility of changing more than just one phoneme. In this change category, participants can change from as few as one phoneme to as many as 3 phonemes. As multiple vowels and codas (which are consonants) are included in final/rime change, the “vowel” results from the Wiener and Turnbull do not reflect single vowel change results. Not only does the “vowel” results include consonant change, they also include multiple vowel change, as we can see in the example above.

Thus, due to this change in experimental method, we can’t be sure that the results of the Wiener and Turnbull study show concrete proof that vowels are most resistant to change, despite the fact that vowels may indeed be least mutable in Mandarin. The seemingly very strong resistance to vowel change in Wiener and Turnbull’s results may instead be due to the fact that the “vowel” in the study is actually a string of vowels (and sometimes a coda). The nature of the task allows participants to assume that every nonword can be made into a real word by changing any of the following: an onset, a final, or a tone, so it makes sense that participants would choose the option that uses the least mental load - one phoneme (such as

[4] Underlined portions show a phoneme that was replaced with another one. Asterisks (*) show where a phoneme was removed.

one onset or one tone) is a lot less to process than the final, which contains 1-3 vowels and possibly a coda at the end. However, due to the nature of Chinese speakers' awareness of Chinese phonology, the change the researchers made to the research design was well-motivated given the nature of Chinese, but it also introduces this unavoidable issue.

Additionally, a potential problem with the word reconstruction task is the inevitability of participants doing explicit lexical search during the task. In this situation, because participants are given a nonword and instructed to change one phoneme to make the nonword into a real word; **explicit lexical search**, the process by which participants reach into their lexicon and individually go through each of the phoneme categories they were instructed to change and arrive at a correct real word, is unavoidable. As mentioned in the original van Ooijen (1996) study, as well as the follow-up studies (Cutler, Sebastian-Galles, Soler-Vilageliu, and van Ooijen, 2000), the results showing that vowels are universally mutable were tempered by the idea that vowels have a much smaller inventory as compared to consonants in each of the three languages they investigated: English, Dutch and Spanish. However, despite the researchers' attempt to control the ratio of consonants and vowels in the languages studied, vowel inventory remained smaller than consonant inventory in all cases, so the inventory size being a factor in the results cannot be ruled out.

In the case of Mandarin, the three categories have the following inventory sizes, from largest to smallest: 38 finals, 21 initials, and 4 tones. Given the nature of the task, for any nonword presented, a participant knows that switching the tone of the nonword to one of three others in the category would suffice for the task. This is in stark contrast to them having to search among 37 other finals or 20 other initials to arrive at a solution to the task. Given that the Wiener & Turnbull results show the exact reverse order of tendency to change (tones were changed most, followed by initials, and then finals), the imbalance of these inventory sizes might also have contributed to the Wiener & Turnbull (2016) results. In subsequent research, it would be a worthwhile task to avoid tasks that involve explicit lexical search, as it would instantly put focus on the inventory sizes of the different phonemic categories featured in the task.

That is why I have edited the word reconstruction task into a task that is able to analyze the exact same questions, without explicitly asking participants to change a consonant, vowel, or tone.

2.2: Creation of a new behavioral task: Forced-Choice Word Selection Task

Because Chinese speakers are not familiar with the concept of consonants and codas, I cannot use a behavioral task that explicitly states these concepts throughout the course of the experiment. Instead, I must create a task that forces a participant to subconsciously change one phoneme category over another.

The task I created for this study is an update on the two-choice forced choice task, which I call the **forced-choice word selection task**, starts off each trial in the same way as does the word reconstruction task: participants hear either a word or a word-like nonword through headphones. The nonwords, which are the critical stimuli in the experiment, differ from real words by just one tone and either one consonant or vowel, for example, *zai2** which does not map to any real words in Mandarin. After hearing *zai2**, participants are then presented two real characters on the computer screen in front of them: one word differs from the audio stimuli in tone (e.g., 灾 *zai1* “disaster”); the other word differs from the audio stimuli in either one consonant or vowel (e.g., 才 *cai2* “just; ability; talent”). The participant is then asked to select the word that most closely resembles the sound they heard. If they select the word with the tone mismatch, it means they view the tone as more variable and mutable when compared to the segment. If they select the word with segment mismatch, it means they view the segment as more variable and mutable when compared to the tone. This task, which does not involve participants selecting a word actively from their mental lexicon, does not involve lexical selection. Thus, the results can be viewed exclusively from the perspective of lexical access.

To make sure participants are actively listening to the audio stimuli and processing the sounds to match to lexical items, I decided to include control stimuli that matched the format of the critical stimuli exactly. As with the critical stimuli, participants hear a monosyllable-tone combination, and two characters appear on the screen. There are 3 versions of control stimuli: easy, hard, and impossible. In the easy control stimuli, participants would be given two characters to choose from, one of which was correct and one which was different in both tone and a segment. For example, in one of the stimuli, the audio heard would be *diu1*, and the choices would be 丢 (*diu1*, 'lose,' correct answer) and 流 (*liu2*, 'flow,' incorrect answer). In the difficult control stimuli, the participants would be given two characters to choose from, one of which was correct and one which was different in tone. For example, the audio heard would be *bu4*, and the choices would be 步 (*bu4*, 'step,' correct answer) and 补 (*bu3*, 'mend,' incorrect answer). From the easy and difficult control items, each participant's accuracy score would be calculated. With a high accuracy score, a participant's active listening and processing of the stimuli can be assumed.

Lastly, in the impossible control stimuli, the participants would be given two characters to choose from, one of which was incorrect in tone and one of which was incorrect in a segment. For example, in one of the stimuli, the audio heard would be *dan4*, and the choices would be 胆 (*dan3*, 'guts,' incorrect in tone) and 探 (*tan4*, 'flow,' incorrect in segment). The impossible control stimuli are meant to mimic the critical stimuli in which there is no correct answer from the choices given. These impossible control stimuli are different from the critical stimuli in that the audio stimuli for the former are real words, whereas in the critical stimuli, the stimuli are nonwords.

The easy, hard, and impossible control stimuli items combined is approximately 2.5 times the amount of critical stimuli. This kind of disproportionate stimuli distribution is key, as the critical task itself is a difficult one, and it is important that the participant is actively paying attention to the task. The easy and hard control items, which test for listening and reading skills, also prevent the need for a pre-experiment listening/reading test.

This task also eliminates the need for explicit lexical search, as I mentioned in section 2.5. In this task, participants do not have to search for an open-ended answer; instead, they are given two choices and told to select one. This allows the task to reveal the variability acceptance of consonants and vowels without the complication introduced by putting all non-onset segments into one category (the aforementioned final/rime).

This task not only allows the basic theory of the word reconstruction task to be replicated in Mandarin, but can also be used in all languages with a written form. This task eliminates the issue of languages whose phonological systems do not make explicit references to the concepts of consonants and vowels. In addition to the priming-lexical decision tasks and same-different tasks mentioned earlier, which all provide insight into how different phonemes constrain or aid lexical access, the forced-choice word selection task can be used when a language does not typically break down phonemes into consonants and vowels.

The study took place over the span of 4 experiments: the first experiment focused on tones and onsets, the 2nd and 3rd experiments focused on tones and codas, and the 4th experiment focused on tones and vowels.

For the first experiment, I pitted tones against onsets. Based on Wiener & Turnbull’s replication of the word reconstruction task, vowels were seen as most resistant to change, so I decided to focus on onsets for the first experiment. As onsets (as compared to codas) in Mandarin have a large number of items, I decided to use onsets as the “pilot” study to see how accurate the Wiener & Turnbull results are. If Wiener & Turnbull's conclusions are correct, then in this experiment, I expect to see that the tone mismatch items are selected at a higher rate than onset mismatch items. This would show that tone is less resistant to change compared to onsets when deciding to change one feature to turn a nonword to a real word.

2.2.1: Experiment 1

2.2.1a: Methods

Participants (N=72)

In Experiment, all 72 (75 tested in total; 3 were removed for poor accuracy on control items) participants self-reported as native Mandarin speakers, with 52 stating that they have a local dialect they speak with family at home. Among the participants, 24 self-reported as female, 32 self-reported as male, and 16 did not report their gender. Age information was not collected, but a majority of participants were university students in Hong Kong and Mainland China. Due to the rise of the Covid-19 pandemic at the time of the experiment, the participants were recruited online, mostly through PolyU research mailing lists. 3 participants were removed from the data because they scored less than 80% on the easy and difficult control items.

Materials

The experimental stimuli included 40 critical stimuli - 40 monosyllabic nonwords that could be changed to real words by changing either the tone or the onset (See Appendix A). The two word options presented on the screen were real words that differed from the nonword stimuli in either tone or onset. Because the task presents an impossible task for participants, it was important for each nonword's tonal or onset mismatch items to be as phonetically similar to the nonword as possible. Additionally, as the task is aimed to see if participants “mind” tonal differences or segmental differences more, the mismatch items were selected based on their comparable phonetic distance to the nonword stimuli using the methods explained below.[5]

For tones, I chose to split tones into two categories that include phonetically similar contours: relatively high (Tones 1 and 2) and relatively low (Tones 3 and 4). Although Mandarin tones are all acoustically distinct, I attempted to put relatively similar tones together. These groups were selected primarily because Tones 3 and 4 both have falling parts, and Tone 2 has a relatively level portion in the very first part of its duration within a syllable, similar to the entirely level Tone 1. Additionally, when Tone 3 is not at the end of the word, the rising part of the tone is omitted, making Tone 3 more acoustically similar to Tone 4 in these instances. Although the stimuli used in this study are all monosyllables, thus leading to no word-final positions, there may be underlying metalinguistic knowledge that allows Mandarin speakers to perceive Tone 3 similarly to Tone 4.

The nonword stimuli’s tone-mismatched real word option would be the other tone in the category the nonword stimuli’s tone fell in. For example, if the nonword stimuli was a Tone 2, then the tone-mismatched real word option would be a Tone 1. One example is the critical stimuli *ca2**. The tone-mismatched item was 擦 (*ca1*, “to wipe”). However, because all 4 tones in Mandarin occur in different fundamental frequency spaces for the majority of their duration, these categories were merely an attempt to offer the most acoustically similar options for the task. All 4 tones were present in the nonword items, as well as in the tone mismatch choices.

[5] The writer is not aware of any objective measure to compare the phonetic distance between two segments and two tones.

As for the onset mismatch options, because tone inaccuracy was widely accepted in the Wiener and Turnbull study, I took special effort to make sure the onset mismatch items were as acoustically close as possible to the nonword stimuli. For each onset mismatch option, the onset (being a consonant) was controlled to differ in only 1 feature from the nonword audio stimuli: either place of articulation, manner of articulation, or aspiration[6]. One example critical stimulus is as follows: the participant would hear *ca2** (/tsʰa/ with rising tone, Tone 2) and be given two written options on the screen: the onset mismatch item 杂 (*za2*: /tsa/ with rising tone, Tone 2), and tone mismatch item 擦 (*ca1*: /tsʰa/ with high tone, Tone 1). Note that the onset mismatch item only differs from the nonword stimulus in aspiration, and that the tone mismatch item is the other item in the relatively high tone category (Tone 1 for the Tone 2 nonword stimuli).

In terms of character selection, to make the mental workload as minimal as possible for participants, the characters used for each of the options was selected based on the most frequently used character of each syllable-tone pairing. For example, *bei1*'s most frequently-used character is 杯 (ZhTenTen, n.d.), which is why it was selected as the character to be displayed on the screen. Additionally, great care was taken to make sure that none of the characters in the critical stimuli had overlapping radicals.

[6] Stimuli showing these 1-feature mismatches and be found in Appendix A.

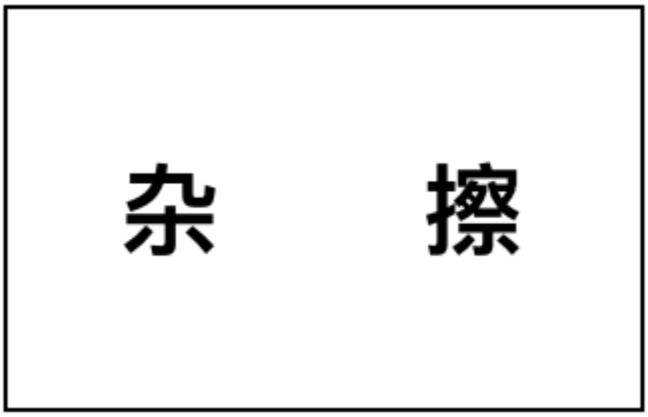


Figure 3: Example Stimuli Screen (Tone versus Onset Mismatch). *In this item, the audio was "*ca2" and the onset mismatch item on left "za2" and tone mismatch item on right "ca1". Participants are asked to press a key on the keyboard to select either the item on the left or the item on the right.*

In addition to the 40 critical stimuli, 60 real word stimuli that matched the format of the critical stimuli exactly (with the exception that there was a correct answer) were also included in the study. The 60 real word stimuli were mixed in with the 40 critical stimuli, so there was no break between the two. (The instructions were the same for all stimuli.) These real word stimuli served to provide an accuracy rating for each participant - any participant that scored less than 80% accuracy in these stimuli were excluded from analysis. The control stimuli were split into easy and difficult levels - easy control items had one correct answer and the other word differing in both tone and onset; difficult control items had one correct answer and the other word differing in either tone or onset, but not both. An example of the easy control items was the audio stimuli *diu1*, and the two words on the screen were 丢 (*diu1*, the correct answer) and 流 (*liu2*, the incorrect answer, with differed from the audio stimuli in both onset and tone). An example of the hard control items was the audio stimuli *bu4*, and the two words on the screen were 步 (*bu4*, the incorrect answer), and 补 (*bu3*, the correct answer). These 60 real word stimuli that doubled as control stimuli were also purposefully chosen to give participants a sense of purposeful accomplishment.

In addition, 20 real word stimuli were added that simulated the impossible choice of the 40 critical stimuli: although the 20 words were real words, the two choice options on the screen did not include the character of the real word, but rather one option that mismatched in tone and one option that mismatched in onset. For example, one such stimulus was the audio stimuli "*dan4"* (a real word), and the words on the screen were 胆 (*dan3*, which differs from the audio in tone only) and 探 (*tan4*, which differs from the audio in onset only). These were added in the experiment to create a balance of possible (a real answer existed) and impossible (no real answer existed) trials (60 possible: 60 impossible).

The spoken words were recorded by a native Mandarin speaker from northern China, and I confirmed the speaker produced all stimuli in the intended tone by listening to the stimuli and by visually inspecting the pitch tracks (see Figure 3) through Praat (Boersma & Weenick, 2023).

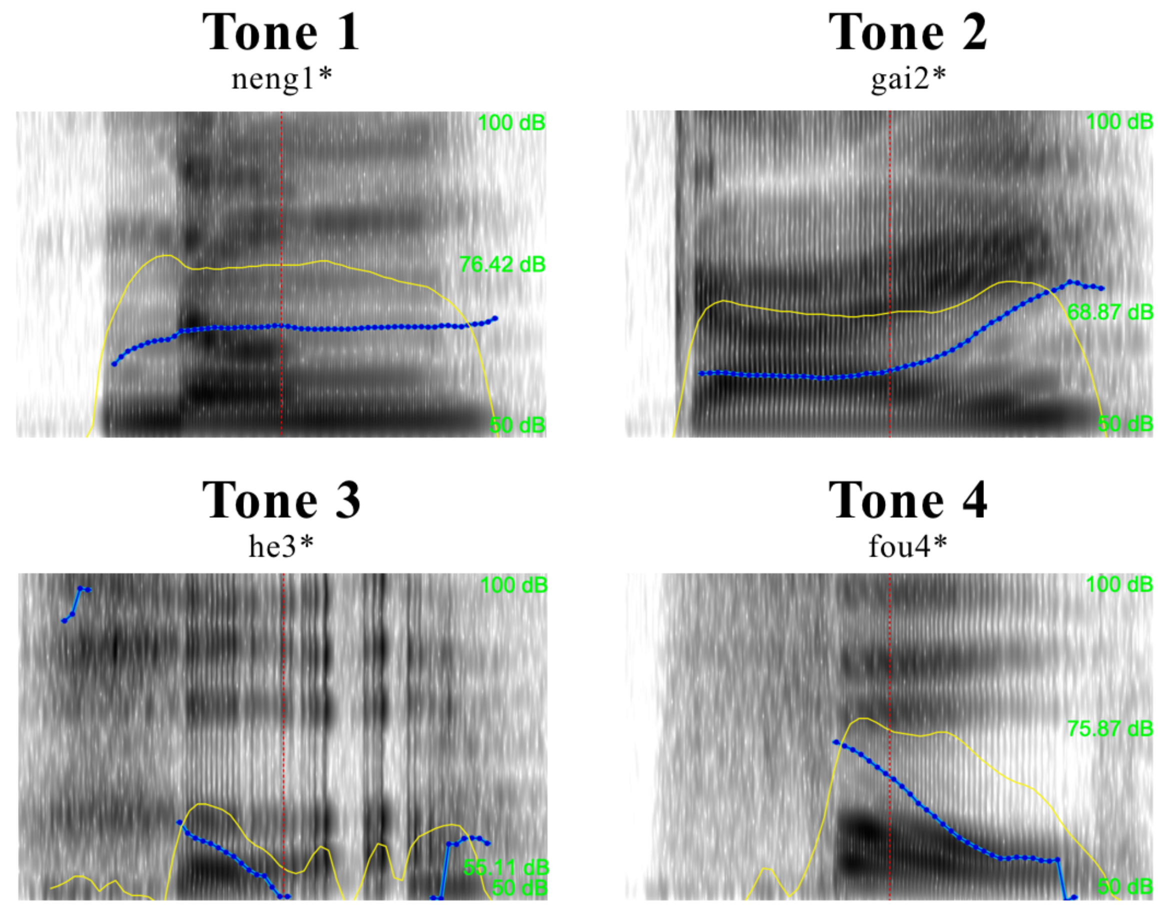


Figure 4: Praat spectrograms of critical stimuli. *The blue dots/lines represent the pitch contours. The asterisks* represent the fact that the stimuli are nonwords in Mandarin.*

Procedure

Due to the effects of the Covid-19 pandemic, it was difficult to recruit participants to come to a university campus to carry out the experiment, so Experiment 1 was conducted through the internet through PsychoPy (Pierce et al., 2019), a Python-based software for psychology experiments. All participants were asked to fill out a language background questionnaire, detailing all (if any) non-standard Chinese dialect and foreign language background, as well as their birthplace, their handedness, etc. Even though the experiment was conducted remotely, the participants were given extensive instructions on the need to be in an isolated room, to use headphones, and to have an uninterrupted time period of roughly

30 minutes. All communication between the researcher and the participants before, during and after the experiment was conducted using simplified Mandarin characters, and there was no pre-screen test prior to the experiment.

Through PsychoPy, the Chinese characters were presented in size 30 Songti font on two sides of the computer screen, and participants were instructed to press Left Shift for the character on the left, and Right Shift for the character on the right. The audio utterance played at exactly the same time the characters appeared on the screen, with no offset delay between the two. The segment mismatch and tone mismatch items' locations on the screen were randomized. There was no fixation or blank screens between trials, and trials proceeded directly one after another, with no time gap. There was no timeout period between trials.

The order of trials was completely randomized, with 4 practice items at the beginning of every experiment session to familiarize participants with the procedure. The practice items' results were excluded from the data.

Analysis

First, each participant received an accuracy score based on the percentage of correct answers in the control stimuli. Any participant who scored lower than 80% in the control stimuli was excluded from analysis. 3 participants were eliminated due to this proficiency check.

Next, each participant received a score for the percentage of choosing the tone mismatch item over the onset mismatch item. Reaction times for the responses were automatically collected, but they were not factored into the analysis because the paradigm was designed to see whether participants would choose tone or onset more often, just like the "free" choice in the word reconstruction task.

2.2.1b: Results

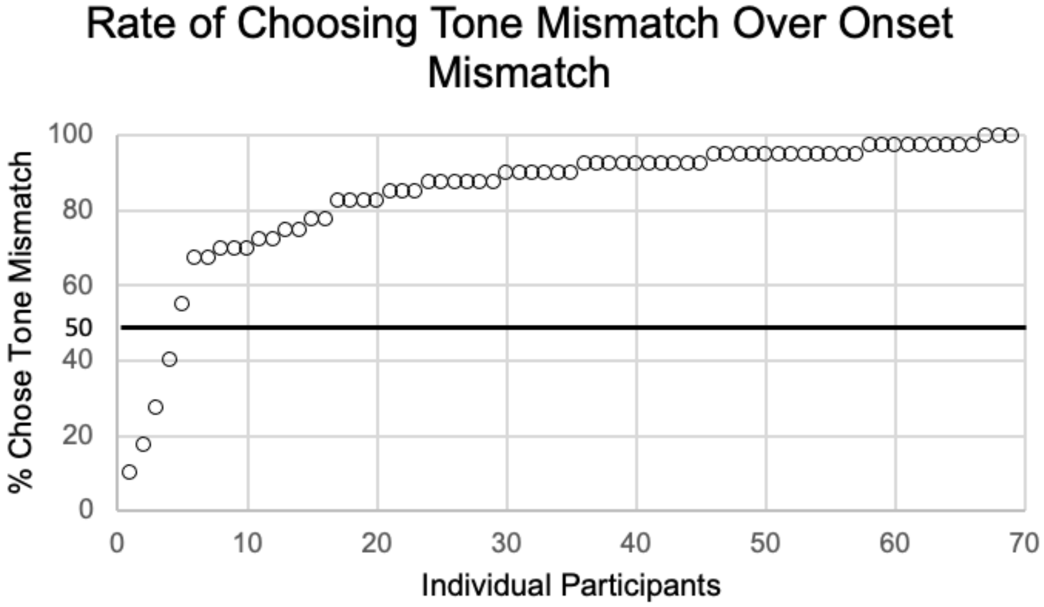


Figure 5: Rate of Individual Participants Choosing Tone Mismatch Over Onset Mismatch (N=72). *Note: The majority of participants chose the tone mismatch as more similar to the stimuli they heard. Only 4 participants chose the onset mismatch as more similar to the stimuli they heard.*

The results of Experiment 1 are as follows: collectively, participants chose to **switch the tone mismatch option** 85% of the time and chose to **switch the onset mismatch option** 15% of the time, for a strong difference statistically, *t*(71) = 16.61, *p* < .001. The effect size, measured by Cohen's d, was d = 1.96, a large effect. Among the 72 participants who passed the proficiency test, their average accuracy was 96%.

Regarding normality, a quick glance at the histogram tells us that the distribution does not follow a normal distribution curve (see Figure 6). Additionally, a Jarque-Bera test of normality yielded a p-value of <.001, further emphasizing the skewness of the data. Based on the Wilcoxon Signed Rank Test for non-normally distributed data, the test statistic is 60, and with the sample size being 72, the test statistic is smaller than the critical value, so we can confidently say that the two distributions are statistically different from each other.

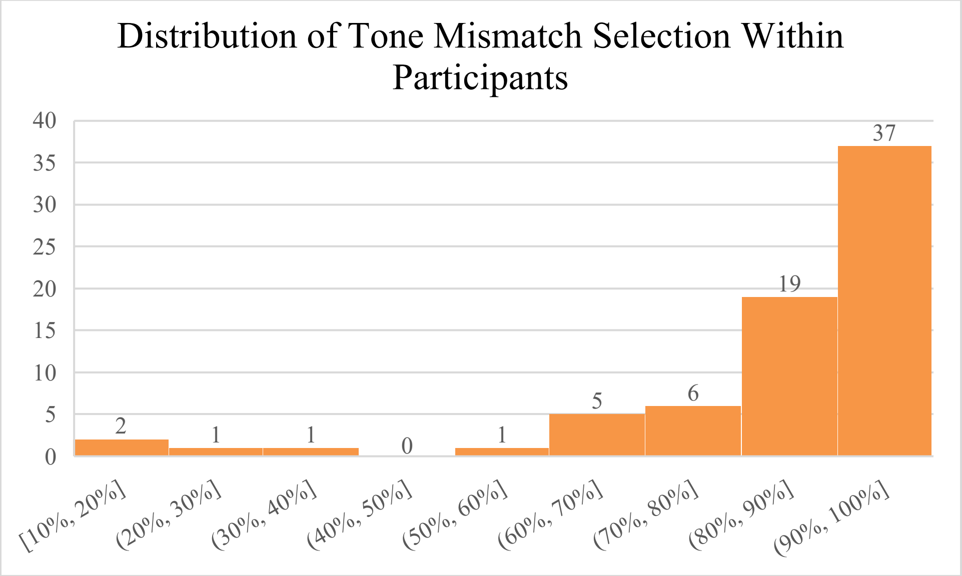


Figure 6: Histogram of the participants' relative selection of tone mismatch items (N=72)

2.2.1c: Discussion

The results of Experiment 1 were quite conclusive: participants overwhelmingly chose the tone mismatch items for the critical stimuli, which shows that they much more willingly accepted tone errors in the forced-choice word selection task. A majority of participants showed a lack of willingness to switch the onset, but there were 4 participants that selected the tone mismatch items less than 50% of the time. These 4 participants were more willing to switch the onset mismatch, which showed that they were more willing to overlook onset mismatch information over tone mismatch information. After looking into the geographic and dialectal background of these 4 participants, I found no concrete pattern that would be able to explain their selections.

These results so far seem to quite closely mirror the results from the Wiener & Turnbull (2016) study, perhaps showing even more extreme results for the general acceptance of the mutability of tonal information, as in Wiener & Turnbull's study, the tone change in the free choice condition was 60% and the consonant change was 27%, with vowel change being most infrequent, with change at 13%. Although the binary nature of the experiment (compared to the ternary nature of the Wiener & Turnbull study) may lead to more extreme phoneme change rates, it seems that the results still show a very skewed tendency for Mandarin listeners to ignore tonal misinformation when compared with segmental information.

Despite the fact that the forced-choice word selection task eliminated the explicit lexical search, there seem to be other factors playing into the results that lead to similar data patterns in both this forced-choice word choice task and Wiener & Turnbull's word reconstruction task. Is the answer as simple as the fact that tone is processed differently from segments? Or could there be another reason behind this tendency to overlook tone mismatch in the experiments thus far?

Seeing as how eliminating this explicit lexical search in the forced-choice word selection task did not make it any less easy to switch the tone (or accept tone mismatch), there must be something else behind this pattern. The next possibility is the inventory size of the respective phonemic categories. Under this theory, the inventory size of a phonemic category has an inverse relationship with how much information that phoneme contributes to lexical access.

For example, we can take a look at the syllable-tone combination *hao3* ("good"). If we look at the onset, knowing that the onset is /h/ means that we know that the word isn't *bao3*, *cao3*, *dao3*, *gao3*, *kao3*, *lao3*, *nao3*, *pao3*, *rao3*, *sao3*, *tao3*, or *zao3* (12 other possible words with the same rime and tone). If we look at the tone, knowing the tone is 3 means that we know the word isn't *hao1*, *hao2* or *hao4* (3 other possible words with the same onset and rime). As you can see, knowing that the onset is /h/ tells us that the word isn't 12 other potential words. However, knowing that the tone is 3 only tells us that the word isn't 3 other potential words. In this case, knowing the tone gives us 1 ⁄ 4 the amount of information about the identity of the word compared to knowing the onset. We can see that tone, a category with only 4 items, contributes to the lexical access of a word a lot less than the onset does.

Each combination of syllable-tone will give us slightly different ratios, but in Mandarin, tone has an inventory of four, which is much smaller than the inventory of 21 of onset consonants. Based on this, naturally, a single tone restricts lexical recognition more than 4 times less than that of a single onset. This may have had a direct effect on the results so far, seeing as participants overwhelmingly chose to switch the tone, which constricts lexical recognition much more than do onsets.

As Cutler, Sebastian-Galles, Soler-Vilageliu, and van Ooijen (2000) showed that with Dutch, a language with a relatively balanced vowel-to-consonant ratio, participants still showed preference to switch vowels over consonants, one might think that would be the end of the story. However, even though Dutch has a more balanced vowel-to-consonant ratio when compared to English or Spanish, there are still more consonants (19) than vowels (16) in Dutch. According to Cutler et al. (2000), there are no languages in which there is an exact balance of consonants to vowels, nor are there languages in which vowels outnumber consonants, so they were unable to replicate the experiment on a language with a truly balanced vowel-to-consonant ratio. Regardless of how hard Cutler et al. tried to balance out the inventory size of their two segmental categories, it was impossible from the start to control for inventory size while using this specific task.

Thus, we cannot rule out the inventory size of the phonemic categories as the primary cause of the accuracy and the tendency to switch those respective categories. For example, in the case of English, Spanish and Dutch, because vowels have smaller inventories than consonants in all three languages, the higher rate of change in the vowels could still be due to the smaller vowel inventory of those sizes, which leads to simpler lexical search. In Mandarin, because tones (4) have the smallest inventory, as compared to consonants and vowels, the results from the Wiener & Turnbull study may also be explained by this element. I also wonder, to what extent, this element affected the results in my Experiment 1 as well. Is there possibly a way to test a phoneme category with an inventory size even smaller than tones?

As it turns out, there is a phoneme category in Mandarin that is indeed smaller than tones. I briefly mentioned the fact that Mandarin has 35 finals/rimes, which include anything from 1-3 vowels, with the possible addition of 1 or 2 nasal codas: /n/ and /ŋ/.

As mentioned before, the way Chinese is traditionally taught does not separate phonemes into consonants and vowels, so analysis into Chinese phonology has consisted of the final/rime as one complete unit. However, the reason for the high rate of tone mismatch compared to segment mismatch may be precisely related to the fact that the final/rime has not been broken down into vowels and codas. If the reason for the high rate of tone mismatch so far is due to the fact that tone has the smallest inventory among the two phoneme categories tested in prior studies (onsets and finals/rimes), then when tone is pitted against coda, a phoneme category with an even smaller inventory, then coda mismatch should have a higher acceptance rate than tone mismatch acceptance.

Thus, the next set of two experiments will replicate the methodology from the first experiment, but instead of pitting tones against onsets, tones will be pitted against codas, a phoneme category with a smaller inventory size than tones.

2.2.2: Experiments 2 & 3

2.2.2a: Methods

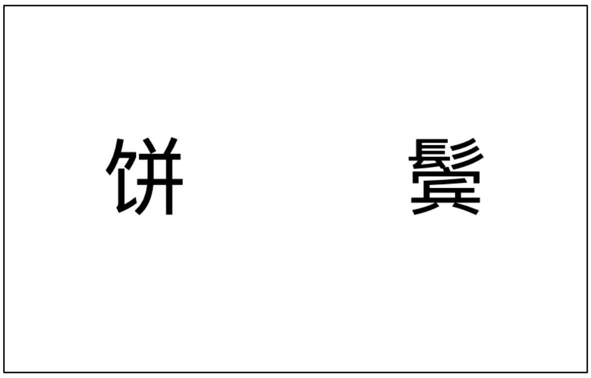
Participants (Experiment 2, N=62; Experiment 3, N=67)

Experiment 2 consisted of 38 females and 24 males, with the mean age being 25, age range 18-51. Experiment 3 consisted of 42 females and 27 males, with the mean age being 24, age range 18-53. By this time, the Covid-19 pandemic situation had alleviated somewhat, and students were allowed to come back to campus to conduct experiments, so both experiments were held on campus at The Hong Kong Polytechnic University. A combined total of 129 PolyU students and staff were recruited, all of whom were native Mandarin speakers, born and raised in Mainland China, with the exception of two participants, one of whom came from Taiwan, and the other from Malaysia. Most had been living in Hong Kong for 1-3 years prior to the experiment. They all reported normal hearing and vision.

Materials (Experiment 2)

The materials for experiments 2 and 3 included all the items from experiment 1, with the addition of a new category of critical stimuli: **tone mismatch versus coda mismatch**. The new category had 22 critical stimuli, consisting of accidental gap nonwords that can be turned into words by switching either a segment or a tone. Although I wanted the number of coda onset mismatch stimuli to match that of the critical onset mismatch stimuli, the limitations to the number of available tone and coda mismatch combinations allowed for a maximum of 22 critical items in the coda mismatch category.

An example critical coda stimuli is as follows: the participant hears the sound *bin3** and in the screen in front of them, they see the words 饼 (*bing3*) and 鬓 (*bin4*) - See Figure 7. As with experiment 1, they are told in advance to choose the word that most closely resembles the sound they heard.



饼　　　鬓

Figure 7: Example Stimuli Screen (Tone versus Coda Mismatch). *In this item, the audio was "*bin3" and the coda mismatch item on left "bing3" and tone mismatch item on right "bin4". Participants are asked to press a key on the keyboard to select either the item on the left or the item on the right.*

Materials (Experiment 3)

The reason for the 3rd experiment is due to an oversight made in the 2nd experiment: in the 2nd experiment, I added critical stimuli that prompted participants to select between mismatching tones or mismatching codas, but I did not include control items that insured that the participants could indeed hear the difference between the /n/ and /ŋ/ codas. Without the control items, even if participants showed more willingness to change the coda instead of the tone, it could be because they didn't hear the difference in codas at all, not because they could hear the difference and chose to switch it anyway. The reason for this is because some southern Chinese dialects do not differentiate between /n/ and /ŋ/, which could be an alternative explanation for elevated coda mismatch acceptance, if that turned out to be the case.

Thus, experiment 3 included 20 new control items that tested for the participants' ability to hear the difference between /n/ and /ŋ/ (See Appendix B). For example, one audio stimulus was the sound "*jiang1*," with the following words on the screen: 江 (*jiang1*, "river") and 间 (*jian1*, "between"). As the only difference in the two words was on the coda, accurate responses on these items would allow us to know that the participants can hear the difference between the /n/ and /ŋ/ coda. As with the first experiment, the data of participants who scored 80% or less on either the onset or coda control items were excluded from analysis. This proficiency check eliminated 2 participants in experiment 3.

Procedure

As experiments 2 and 3 were conducted on campus at PolyU, the participants were invited to an empty computer lab on campus and were staggered in their participation time slots. However, there were some participants whose participation slots overlapped, as some came later or earlier than their assigned time slots. However, this should not have affected the experiments, since all participants wore noise-canceling headphones for the duration of the experiment.

The experiment was carried out using psychological testing software DMDX (Forster and Forster, 2003). The settings were set to be the same as those from experiment 1, with all trials using completely randomized order and with 4 practice items at the start of the experiment.

Analysis

As with experiment 1, participants who scored less than 80% in the control items were first excluded from analysis. In experiment 2, no participants were excluded. In experiment 3, in which I added the coda control items, 2 participants were eliminated due to their poor performance in the coda control items.

Critical onset versus tone items were separated from critical coda versus tone items, and individual scores were calculated.

2.2.2b: Results

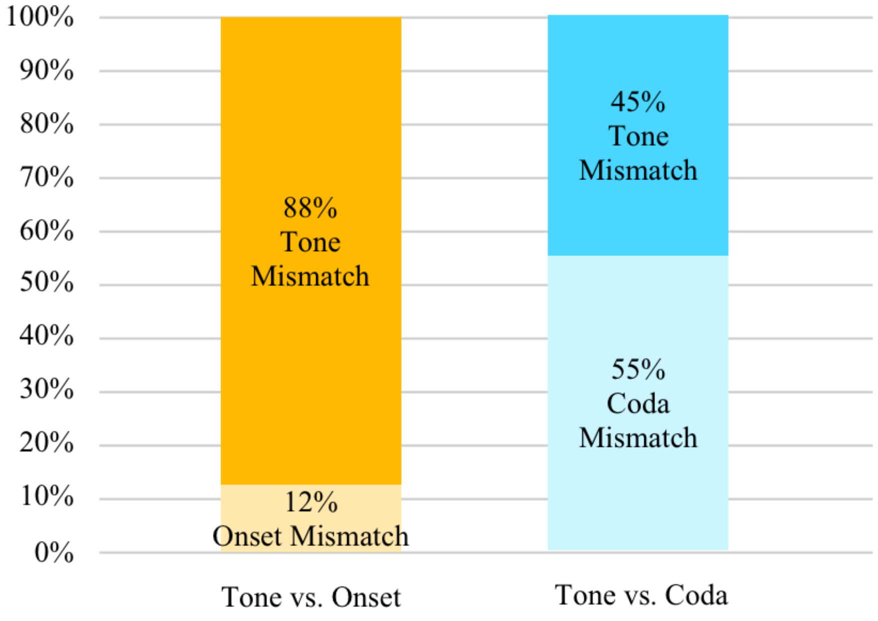


Figure 8: Rate of Tone Mismatch versus Onset and Coda Mismatch in Experiment 2 (N=62). *Note that onset change is very minimal, whereas coda change is more common than tone change.*

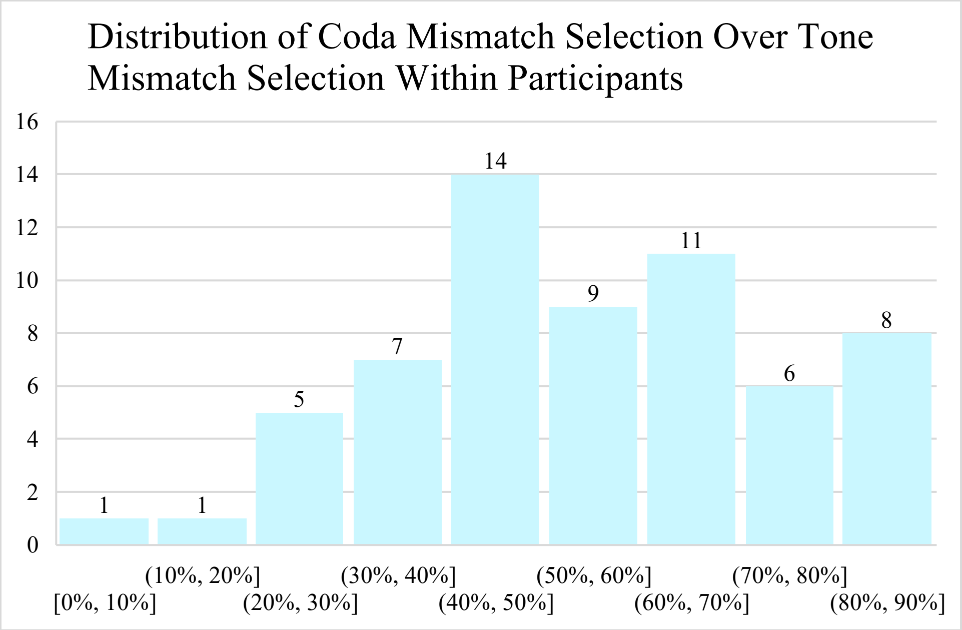


Figure 9: Histogram of Participants Selecting Coda Mismatch over Tone Mismatch in Experiment 2 (N=62)

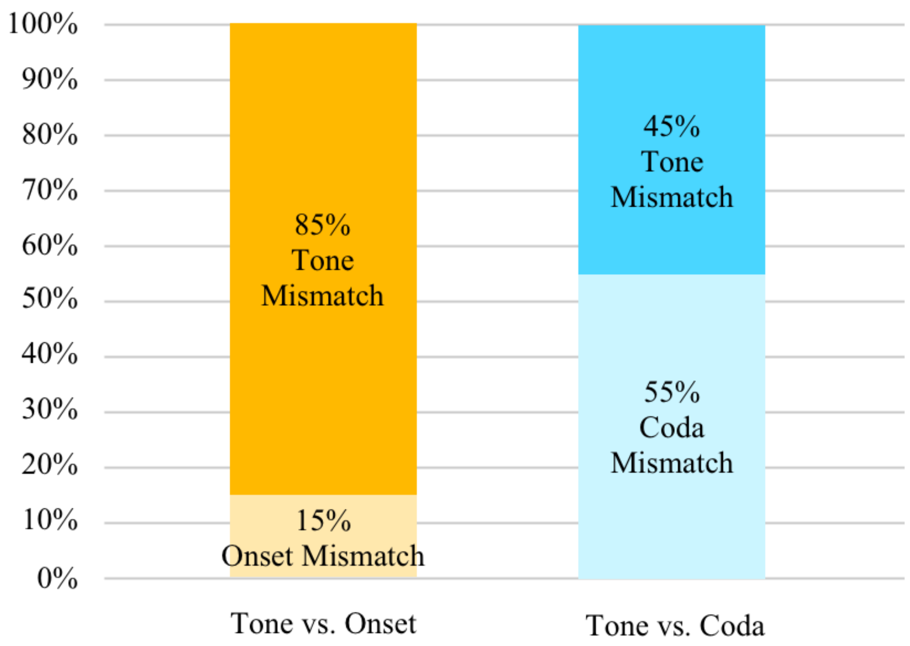


Figure 10: Rate of Tone Mismatch versus Onset and Coda Mismatch in Experiment 3 (N=67). *Note that onset change is very minimal, whereas coda change is more common than tone change.*

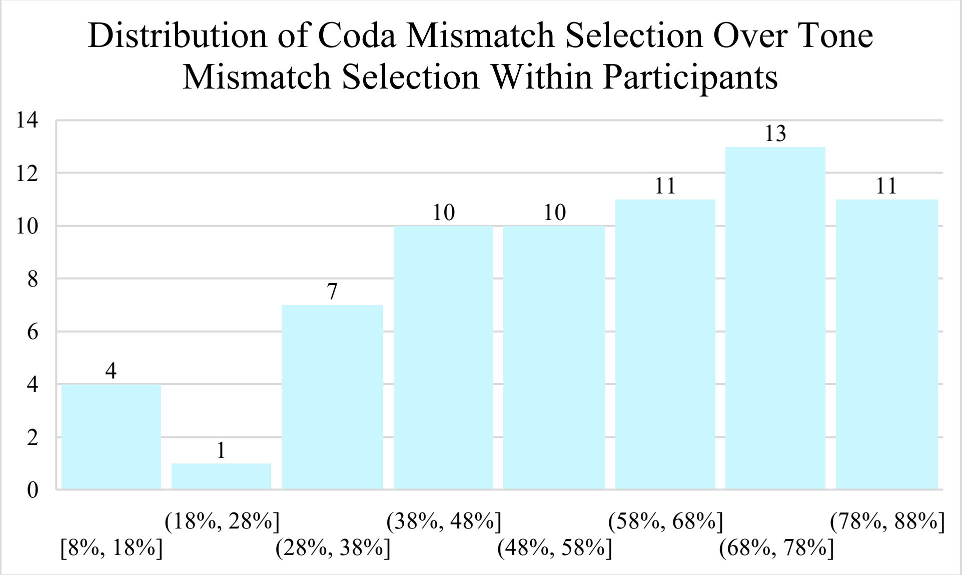


Figure 11: Histogram of Participants Selecting Coda Mismatch over Tone Mismatch in Experiment 3 (N=67)

In experiment 2 (see Figure 8), the average rate of choosing tone over onset mismatch was 88%, a significant difference from **onset mismatch over tone mismatch** at

12%, *t*(61) = -15.71, *p* < .001. The effect size, measured by Cohen's d, was d = 1.85, a large effect size. In contrast, the average rate of choosing **tone mismatch over coda mismatch** was 45%, a borderline significant difference from choosing **coda mismatch over tone mismatch**, *t*(61) = 1.94, *p* = .055. The effect size, measured by Cohen's d, was d = .23, a small effect size. The data of tone over coda mismatch (and vice versa) follows a normal distribution curve, as verified by the Jarque-Bera test, p=.55.

In experiment 3 (see Figure 10), the average rate of choosing **tone mismatch over onset mismatch** was 85%, which resulted in a significant effect when compared to choosing **onset mismatch over tone mismatch** (15%), *t*(66) = -16.73, *p* < .001. The effect size, measured by Cohen's d, was d = 1.97, a large effect size. In contrast, the average rate of choosing **tone over coda mismatch** was 45%, a significant effect, *t*(66) = 2.54, *p* = .013. The effect size, measured by Cohen's d, was d = .30, a small effect size. The data of tone over coda mismatch (and vice versa) does follows a normal distribution, with the Jarque-Bera test showing skewness, p = .17.

In other words, the rate of tone mismatch selection when pitted against onsets remained relatively constant, matching the rate found from experiment 1. However, when tone mismatch was pitted against the newly-added coda mismatch, people were much less willing to change the tone than they were when tone mismatch was pitted against onset mismatch. In experiment 2, people chose tone mismatch in the tone-vs-onset condition significantly more than they chose tone mismatch in the tone-vs-coda condition, *t*(61) = 17.83, *p* < .001. In experiment 3, the same results were found, *t*(66) = 11.31, *p* < .001.

Based on the rate of choosing tone mismatch over coda mismatch, this means that participants tended to overlook coda mismatch more than they overlooked tone mismatch. With the additional of the coda control stimuli in the 3rd experiment, the participants in that experiment proved that they were able to hear the difference between the codas, so the fact that they chose the coda mismatch cannot be because they didn’t notice the mismatch; instead, it must be that they noticed the mismatch but tolerated it.

2.2.2c: Discussion

The results of experiments 2 and 3 were a turning point in the research. These studies were the first to examine how coda variation acceptance contrasted with tone variation acceptance during Mandarin spoken word recognition. The fact that participants generally preferred to switch the coda mismatch over the tone mismatch appears to be a watershed moment for the theory that category inventory size has a direct correlation to how variable that category’s items are during phonological processing.

Additionally, experiments 2 and 3 solidified the results from experiment 1, which was conducted remotely over the internet. In all, the onset versus tone mismatch selection rates remained consistent through all 3 experiments, and the coda versus tone mismatch selection rates were almost identical in experiments 2 and 3.

Looking at the language backgrounds of the participants in experiments 2 and 3, I again did not spot any obvious patterns behind which participants chose coda mismatch over tone mismatch or vice versa. China has 6 language-region groups, and when the data is broken down by language-region groups, all groups are represented in both high coda and low coda selection rates.

However, as enticing as it may be to conclude with the inventory size theory, there is actually another possible alternative explanation for these results: **incremental processing**, a theory espoused by Marslen-Wilson’s cohort model (1978), which was previously mentioned in Chapter 2.1.1. In the cohort model, lexical activation is incremental, moving from beginning to end, in a bottom-up acoustics-first approach to speech recognition. In this approach, the first sound in a word activates the most amount of lexical items that are possible for selection and shows the strongest rate of activation. With each subsequent phoneme in the word produced, the number of lexical items the word could possibly be shrinks, with these new phonemes showing increasingly decreasing levels of activation.

The logic is as follows: since onsets are at the beginning of a word, they would get the strongest level of activation, and thus show more resistance to change. Tones, being in the middle part of the word (unfolding over the course of the nucleus/rime), would show middle-of-the-road levels of activation. Codas, being at the end of the word, would show the least amount of activation. The current experimental data completely fits with this reasoning. Thus, it's important to follow-up the current experiments with new data that might or might not challenge this assumption.

Originally, I had not planned to use the forced-choice word selection task on vowels, as vowels were the most resistant-to-change category in the Wiener & Turnbull study. However, due to the concern that incremental lexical activation could possibly be a major factor behind the data so far, I decided to use the forced-choice word selection task once again: this time, I was pitting tone mismatch against vowel mismatch.

Because tones and vowels unfold at around the same time course during the utterance of a Mandarin syllable, under the theory of incremental lexical activation, they should have similar rates of activation. If the incremental lexical activation theory is correct in this instance, tones and vowel mismatch items should show around a 50/50 rate of selection.

However, if inventory size is the correct explanation for the experimental data thus far, then tones, which has an inventory size of 4, should have higher rates of mismatch selection compared to vowels, which has an arguably larger inventory size.[7]

[7] Theories about the number of vowel phonemes in Mandarin can be found on page 67 (2.2.3c).

2.2.3: Experiment 4

2.2.3a: Methods

Participants (N=69)

For experiment 4, because the previous two experiments already accrued over 125 native Mandarin speakers currently attending or working at PolyU, there were concerns over the possibility that there would not be enough participants if the experiment was held in person. Thus, I decided to conduct the experiment using the same manner as experiment 1, and recruited 69 participants (30 females, average age: 24 years) over the internet, the majority of whom were living in Beijing at the time of the experiment, and who came from all over the country. 31 participants self-reported as speaking a local dialect with family, and no one dialect was spoken among over 15% of the participants. The participants were paid according to their time expenditure in the experiment.

Materials

Due to the consistent and robust results of the tone vs onset and tone vs coda stimuli from the previous 3 experiments, I decided to forgo those stimuli and create a shorter experiment with only tone vs vowel stimuli.

To stay in line with the critical stimuli of the previous experiments, I chose accidental gap nonwords that could be turned into real words by changing either a tone or a segment (in this case, a vowel). To keep conditions similar to previous experiments, all of the nonwords chosen for this experiment were words that contained monophthongs (with the nonword and vowel mismatch both being monophthongs). Due to the fact that Mandarin only has 7 monophthongs, plus the fact that the monophthongs do not occur in free rotation with onsets, this severely limited the amount of critical stimuli that could be chosen.

In the end, 12 critical vowel versus tone mismatch items were selected, and the control items were in the same ratio as the critical versus control items in the previous 3 experiments (2.5 control items to 1 critical item). The control items once again provided a means to tease apart the proficient speakers from those whose selections showed diminished language ability, inability to follow directions, or lack of focus during the experiment (See Appendix C).

Procedure

As with experiment 1, the 69 participants in experiment 2 were given extensive instructions on the need to be in an isolated room, to use headphones, and to have an uninterrupted time period of roughly 15 minutes (as this experiment was substantially shorter than the previous ones). All communication between the researcher and the participants was conducted using simplified Mandarin characters.

Analysis

As experiment 4 only had 1 category of critical stimuli, the control items were first used to eliminate any participants who scored less than 80% in the accuracy test items. Two participants were excluded through this process (these two participants are not included in the total participant number). Next, the rate of choosing tone mismatch over vowel mismatch was calculated for each individual participant.

2.2.3b: Results

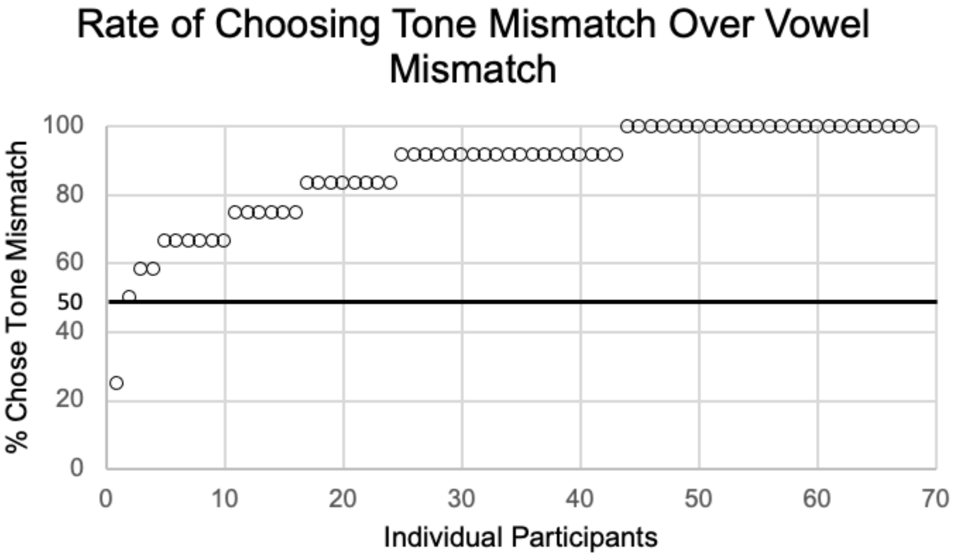


Figure 12: Experiment 4 - Rate of Overall Selection of Tone Mismatch versus Vowel Mismatch (N=69). *Note that vowel change is chosen about as often as onset change. This shows that listeners preferred tone mismatch to vowel and onset mismatch at about the same rate.*

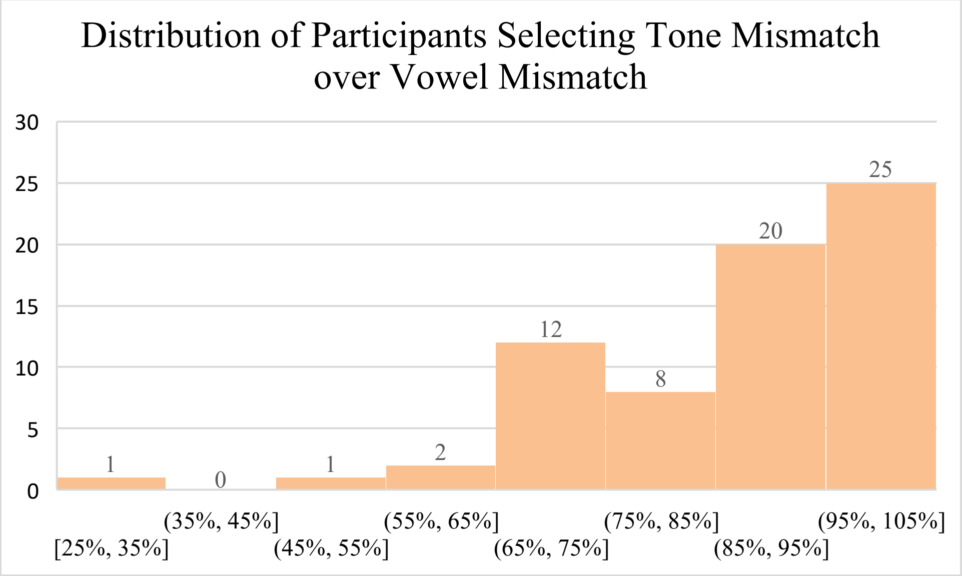


Figure 13: Histogram of Distribution of Participants Selecting Tone Mismatch over Vowel Mismatch in Experiment 4 (N=69)

The average rate of **choosing tone mismatch over vowel mismatch** was 88% across the 69 participants, showing a significant preference for selecting tone mismatch over vowel mismatch, *t*(68) = 18.70, *p* < .001. The effect size, measured by Cohen's d, was d = 2.2, a large effect size.

Additionally, the vowel mismatch selection rate (12%) was about as low as that of the onset mismatch selection rate (15%, 12% and 15%, respectively) in the first three experiments, and these distribution comparisons did not show significant differences through two-sample independent t-tests. Two-sample independent t-tests were conducted on the data from the 4th experiment in contrast with the previous three experiments: *t*(139) = 0.87, *p* = 0.384 for vowel mismatch and onset mismatch in experiment one, and *t*(129) = 0.04, *p* =.966 for vowel mismatch and onset mismatch in experiment 2, and *t*(136) = 1.73, *p* =.086 in experiment 3.

However, the vowel mismatch rate was slightly lower than that of the onset mismatch rate in experiment 3 (where onset mismatch rate was 15%), and it led to a marginally significant difference, *t*(134) =1.73, *p* = 0.085. There was not a clear relationship between the dialect background of the participants and their selections.

2.2.3c: Discussion

The results of experiment 4 show strong evidence against the idea that incremental lexical activation plays a strong role in the results of this series of forced-choice word selection tasks. Because tones and vowels unfold at around the same time course in a Mandarin syllable, the fact that tone variation showed an overwhelming amount of acceptance over vowel variation (88% to 12%) makes the incremental lexical activation explanation unlikely.

With regards to vowels, earlier I mentioned that in perception studies, tone is often compared with the rime (the final), on which the vowel is carried. As tone is carried over the sonorant part of the syllable, that tends to be the rime/final, so there is reason to suspect the two are connected. Some linguists even support the idea that tone is a property of the nucleus and that tonal distinctions are realized by manipulating properties of the vowel. However, other than the fact that shorter tones tend to correspond to shorter rimes, and longer tones tend to correspond to longer rimes, I have not seen strong evidence showing that tone distinctions can be produced by changing the vowel. However, as tone is inextricably connected to the tone, I suspect this theory will continue to propagate, and it is certainly a topic worthy of further research.

One element unmentioned until now is the precise inventory size of consonants and vowels in Mandarin. Starting with consonants, according to The Ministry of Education of the People's Republic of China (中华人民共和国教育部) lists 22 consonants in the language: *b, p, m, f, d, t, n, l, g, k, h, j, q, x, zh, ch, sh, r, z, c, s, ng* (2008)- these are the 21 initials from Chapter 1 (Introduction), plus the coda position-only *ng* [ŋ]. However, there is debate over linguists over the true number of consonants in Mandarin: some linguists that argue there are indeed 22 consonants (Cheng, 1973; Yip, 2000), the fact that 3 of the consonants listed above (palatal consonants *j, q, x*) are actually in complementary distribution with the dental consonants (*z, c, s*) and retroflex consonants (*zh, ch, sh*), has led some linguistics to believe that they should be viewed as part of the other consonants (Chao, 1934, 1968; Hartman, 1944; Hseuh, 1986: 34-6) or viewed not purely as consonants (Duanmu 2007 views them as consonant-glide combinations). In summary, there are 19 to 22 items in the consonant inventory, depending on interpretation.

Regarding vowels, there seems to be much greater debate over the nature of vowels in Mandarin. The aforementioned Ministry of Education of the People's Republic of China lists 10 vowels in the language: *a, o, e, i, u, ü, ê, -i* (front), *-i* (back), and *er* (2008). However, IPA for these vowels has not been provided, so it's difficult to infer exactly which sounds these letters refer to. From a research perspective, the number of vowels (monophthongs) proposed by linguists range from 5 to 9 (Yip, 2000; Shi, 2002; Zhang, 2002; Duanmu, 2007; Lin & Wang, 2013; Yang & Oh, 2020). One account argues that there are only two vowels in Mandarin (Wang, 1993), and another even argues that there are no vowels in Mandarin (Pulleyblank, 1984), although such theories will not be pursued further in this thesis.

Vowels in Mandarin, as in other languages, are typically analyzed in terms of frontedness, height, and spreadness/roundness. According to Yip (2000), Mandarin has 6 distinct vowels, but acknowledges that 4 of the 6 have allophonic variants, depending on the context.

	Front (*spreaded/rounded*)		Central	Back (*spreaded/**rounded***)	
High	-**i [i]:** -[ɹ̩] after *z*, *c*, *s* -[ɻ̩] after *zh*, *ch*, *sh*	**ü [y]**		**u [u]**	
Mid-high				-**e [ɤ]:** - [e] in -*ei* - [ə] in -*en*, -*eng* - [ɛ] in -*ie* - [ʌ] in -*eng*, -*ueng*	- **o [o]:** - [ɔ] in -*uo* - [u] in -*ao*
Mid					
Mid-low					
Low			- **a [A]:** - [a] in -*an*, -*uan* - [ɛ] in -*ian* - [ʌ] in -*uang* - [ɑ] in -*ang*, -*iang*		

Table 4: Mandarin's vowel inventory, according to Yip (2000). *The bolded and underlined items are unique vowel phonemes, and the list below each vowel phoneme shows how the sound changes depending on context. Italicized text represents pinyin.*

In the above chart, there are 17 unique vowel phones. Without getting into the specific breakdown of which of the phones in the above chart are vowel phonemes in Mandarin, the majority of theories accept that each vowel phonemes have *some* allophonic variants, although some argue for fewer distinctions (Duanmu, 2007) and others argue for more distinctions (Lin & Wang, 2013). The status of the retroflex vowel *er* [ɚ] is uncertain, with some linguists listing it alongside the vowel inventory, but not counting it as a contributing to the vowel inventory (Yip, 2000; Duanmu, 2007).

With regards to the earlier assertion that inventory size of a phoneme category has an inverse relationship with the amount of variance acceptance of that category, the arguments for the more conservative vowel inventory size (5 vowel phonemes) seems to be a challenge to this theory, as these arguments propose a modest number of unique vowel phonemes. In keeping with these above theories, our results suggest a larger inventory size for vowels as compared to tones, such as the underlying vowel inventory being closer to 9 items than 5 items.

However, regardless of whether Mandarin has 5 vowel phonemes or 9 vowel phonemes underlyingly, there is the possibility that acceptance of variation in the nucleus/vowel of Mandarin words isn't based on the number of vowel phonemes. One possibility is that it's based on the 17 specific phonetic realizations in context. Another possibility is that it's based on the nucleus inventory, including monophthongs, diphthongs and triphthongs, of which there are about 20 items (minus the coda). The results from the present study, which show that resistance to vowel variation is comparable to that of consonant variation, seem to support either of these latter two theories, but it also lends credence to the inventory size theory that vowels, with 5 or 9 inventory items, still receives more resistance than tones, with 4 inventory items, and codas, with 2 inventory items.

Conversely, the opposite argument may be made for the effects inventory size on the results of mismatch acceptance in the forced-choice word selection task. Instead of a small inventory leading to more mismatch acceptance, it could be the opposite: because a phoneme has a small inventory, the limited options within its confines make listeners cling more to the specific differences within that category. Although the idea has theoretical plausibility, the results of the current study do not support this theory. Presently, we can see that the phoneme inventories with the smallest number of items, codas (2) and tones (4), show much more mismatch acceptance than the phoneme inventories with the largest number of items, onsets

(19-22) and vowels (5-9). If the opposite theory is correct, further testing would need to be conducted to challenge the results from the current study.

Another theory that has been brought up to explain phonological processing of a certain phoneme is **phonological neighborhood density** (NAM; Luce & Pisoni, 1998). A phonological neighbor of a word is said to differ from said word in one phoneme, whether it's an addition, deletion, or subtraction. It has been found that in **visual word recognition**, words with high neighborhood density are recognized faster. For example, a word like *bed* (with at least the following neighbors: *head, said, led, bad, bud, bet*, etc.) would have faster lexical decision response rates compared to a word like *portable* (with only the following neighbors: *courtable*, etc.) (Ziegler, Muneaux, & Grainger, 2003). However, during **auditory word recognition**, it has been shown that words with high neighborhood density (i.e., *bed*) would have slower lexical decision response rates compared to those of low neighborhood density words (i.e., *portable*) (Garlock, Walley, & Metsala, 2001; Luce & Pisoni, 1998; Vitevitch & Luce, 1998; Ziegler et al., 2003). It is argued that auditory input is sequential where visual input is not, thereby causing lexical competition between words in a straight-forward sense (Ziegler et al., 2003).

Thus, it is important to look at the number of phonological neighbors for each of the tone and segmental mismatch options. Among the onset versus tone mismatch items (40 items), the average neighborhood density of onset mismatch items was 15.1 neighbors, compared to an average neighborhood density of tone mismatch items of 14.8. Among the coda versus tone mismatch items (22 items), the average neighborhood density of coda mismatch items was 13.9 neighbors, compared to an average neighborhood density of tone mismatch items of 14.6 items. Among the vowel versus tone mismatch items (12 items), the average neighborhood density of vowel mismatch items was 16.1 neighbors, compared to a neighborhood density of tone mismatch items of 14.6. The average tonal neighborhood density across the 3 experiment sets was 14.6. (See Appendix D for the neighborhood density of all critical stimuli). These numbers were calculated based on the Database of word-level statistics for Mandarin Chinese (Neergard et al., 2022), using the tonal fully segmented schema (C_G_V_X_T). For the stimuli that were homophones (the character 乐, for example, can be read *le4* or *yue4*), the neighborhood densities were combined for all homophones into one total neighborhood density for that stimuli.

On first glance, it looks like the neighborhood density numbers are a direct inverse to the relative mismatch acceptance of our tonal and segmental categories tested in the current study. Onsets and vowels, which experienced the lowest mismatch acceptance, also has stimuli with the greatest amount of phonological neighbors, at 15.1 and 16.1 items, respectively. Tone, with the second highest mismatch acceptance, has stimuli with the second smallest amount of phonological neighbors, at 14.6 items. Coda, which had the highest mismatch acceptance, has stimuli with the smallest amount of phonological neighbors, at 13.9 items.

Thus, it's reasonable to use phonological neighborhood density effects to explain the results of the current study. As phonological neighborhood density is a property of phoneme restrictions within a language, it's possible that its effects are linked to inventory size.

With tones showing much more acceptance to variation compared to both onsets and vowels, and with codas showing even more acceptance to variation, this last experiment provides strong evidence that the size of the inventory is inversely correlated to acceptance of variability in that respective inventory's lexical processing.

2.3: General Discussion

In a series of 4 forced-choice word selection tasks, I tested 270 participants' acceptance of tonal variation in contrast with onset, vowel, and coda variation. The task consisted of a nonword audio stimulus being played over headphones, followed with participants being prompted to choose one of two options that they feel most closely resembled the word they heard: one option differed in tone and one option differed in a segment. The results from the four experiments showed that tone variation was more acceptable than onset and vowel variation, but that tone variation was less acceptable than coda variation. In Mandarin, there are the highest number of onsets (19-22), followed by a debatable number of vowels (5-9), 4 tones, and 2 codas. Because the variation acceptance rate of the 4 experiments is in exactly the opposite order, with participants accepting coda variation the most, followed by tone acceptance, followed by onsets and vowels at around the same level, I showed that the inventory size of a phonemic category has an inverse correlation with the acceptance of variability of that category in spoken word processing.

The **forced- choice word selection task** was created with Mandarin's phonological system in mind, but the task can be used for any language with a writing system, to test levels of variability acceptability in a number of different ways. It is especially useful in languages, such as Mandarin or Cantonese, where non-linguist speakers of those languages do not tend to break down "consonants" and "vowels" in the language.

However, there are limits to this task, as there are to any behavioral task. First of all, it is not natural for a listener to be given an acoustic nonword and be told to choose between two items that are not the nonword they heard. It might be difficult to accept the results of these critical stimuli at first, but the high performance of participants on the control stimuli shows that they actively listened to each of the stimuli and performed the lexical selection task to the best of their abilities.

In the current study, tone was the phoneme category of interest, so the two-way mismatch items were tones compared to different categories of segments. However, in future studies, it would be meaningful to use the forced-choice word selection task to test comparative mismatch acceptance within segments themselves. As the tasks conducted in this study only pitted different categories of segments with tones, there are no tasks directly seeing how onsets behave in contrast to vowels, or how vowels behave in contrast to codas, etc. Thus, the statements made about different categories of segments are made only on indirect comparisons. In the future, the task could be used to test onset vs vowel mismatch acceptance, vowel vs coda mismatch acceptance, and onset vs coda mismatch acceptance. This research would further our knowledge of how different types of segments contribute to lexical access in Mandarin Chinese.

It would be worthwhile to use the forced-choice word selection task on the words used for the previous English, Dutch and Spanish experiments and see if the vowel mutability from their tasks would be replicated. Based on the inventory size theory I espoused earlier, I suspect that the results from the forced-choice word selection task of these items would indeed be quite similar, given that vowels are fewer than consonants in all 3 languages. This would provide further proof that inventory size of a phonemic category has a direct effect on the acceptance of variability of that category in spoken word processing.

Now that I’ve looked at the acceptance of variability of tones in Mandarin speech processing, there remains the question of how tones behave in speech production. Thus, the second part of my dissertation will focus on another aspect of Mandarin oral language: the opposite of my first focus, I will now look at how tone is utilized in the process of speech encoding, or the formation of speech codes in the mind.

CHAPTER 3: TONAL PRODUCTION

3.1: Previous research on speech production

3.1.1: Speech production in Indo-European languages

For this section, we will look at the other side of spoken language: speech production. The speech production process is usually broken down into 3 main sections:

(1) The speaker conceptualizes what they intend to say.

(2) The speaker generates the linguistic code for the intended utterance (also known as phonological encoding).

(3) The linguistic code is produced acoustically, in the form of air flow through the lungs and the oral and nasal cavities.

In the first part, the brain must first string together a series of words it intends to say, but without a specific plan of how it will be executed. This chapter will focus on the second and next part of this process, called **phonological encoding**: this involves preparing the sequence of phonemes that the speaker intends to say, which involves the selection of the proper morphemes, syllabification of the morpheme sequence, and the implementation of proper suprasegmental processes, such as stress. In the third and last part of the process, the speaker prepares the articulatory process, and then articulates the intended lexical item(s).

As we are concerned with phonological encoding, the second part of this process, it is important to know how the brain deals with the selection and execution of words in sequence. Are a word's individual phonemes selected first? Or does the selection work in a more granular manner: are the features of each phoneme selected first? Or perhaps, does the selection work in a broader manner: are the syllables of each word, rather than individual phonemes, selected first?

Let’s say you have selected the word you will say: *hamlet*. In the preparation of the production of ‘hamlet’, are the individual phonemes of /h/ /æ/ /m/ /l/ /ɪ/ /t/ (UK pronunciation) selected first, and then placed in order? Or are the features of each phoneme - voiceless glottal fricative (for /h/), near-open front unrounded vowel (for /æ/), voiced bilabial nasal (for /m/), etc., selected instead of the phonemes themselves? Or are the full syllables /hæm/ and /lɪt/ selected first, followed by the corresponding phonemes of each syllable then being filled in accordingly? In other words, below the word level, what is the fundamental phonological unit that makes up speech production - is it a phoneme, a syllable, or a feature? And how does the brain deal with the selection and execution of several, up to several dozen words, in a consecutive manner?

Fortunately, over the years linguists have developed two major methods with which we can study the process of phonological encoding: one is through behavioral experiments using a paradigm known as form preparation (described below), and the other is by studying speech errors. Behavioral experiments allow researchers to control stimuli in a way in which different phonological units, such as phonemes, strings of phonemes, or entire syllables, can be tested for speech facilitation or inhibition; while speech errors can tell us how natural breakdowns in speech give us insight into how the brain processes phonemes in the process of speech articulation.

3.1.1a: Behavioral research on Indo-European speech production

Starting with behavioral experiments, there has been the extensive use of the form preparation paradigm (also known as implicit priming) to tease out the first selectable phonological unit within a word or morpheme, referred to above as “the fundamental phonological unit,” but also commonly called a **proximate unit** (O’Seaghdha et al., 2010). Through the use of **implicit priming**, we have learned what the proximate unit is in different languages. In implicit priming, participants are asked to memorize pairs of words in which the targets overlap in various phonological ways – usually matching segments in either the onset, vowel, coda positions or a combination of these.

For example, a participant would be given a set of words like the following and told to memorize them together: single-*loner*, nearby-*local*, flower-*lotus* (with the former being the cue word and the latter being the target word). They would then be asked to name the target (latter) word upon seeing the cue (former) word. If they were shown "single," they would say "loner;" if they were shown "nearby," they would say "local," and if they were shown "flower," they would say "lotus." People respond faster to words in a set like this as compared to a set where the target words to be spoken aloud are not related (e.g., a set like happy-*paper*, worldly-*grass*, bored-*umbrella*). The reason why people are faster with a set that has related cue and target words is because before the participant even sees any cue, they already know that the word they have to say starts with "lo," so they can code that ahead of time. This shows that when the target words were phonetically related to each other, priming effects were elicited. Using this paradigm, we can test for the smallest phonological unit that can elicit priming effects between target words - that smallest phonological unit would be the smallest unit that can be used to plan speech in a given language (proximate unit).

Results from Dutch and English (Levelt, 1999; Roelofs & Meyer, 1998) show that significant priming occurred when the two words minimally shared an onset. However, there was no priming for two segments that shared all phonological features except for one ('**b**aby' did not prime '**p**atient' - the voicing on these onsets is different). These results provide evidence for the theory that the proximate unit in such Indo-European languages is the segment, and not something bigger, like a syllable (since whole-syllable overlap is not required to elicit implicit priming), or something smaller, like a feature (since overlap in a feature is not sufficient to elicit priming). This means that the segment is the first selectable phonological unit below the level of the word/morpheme, which means that it is also actively selected during the process of speech planning.

3.1.1b: Speech error research on Indo-European speech production

Converging evidence for this theory comes from research analyzing speech errors in these same languages. In Indo-European languages, speech error studies show that segments like consonants and vowels err not in random substitutions, but due to the language context. A large portion of errors are anticipatory (an error caused by a later word; intended utterance: Leaning Tower of Pisa, actual utterance: *Teaning Tower*), perseveratory (an error caused by an earlier word; intended utterance: pizza box, actual utterance: *pizza pox*), or exchanges (in which one element of two adjacent words are swapped; intended utterance: cookie jar, actual utterance: *jookie car*), etc. These errors give further evidence supporting modern models of speech planning, in which selection of the current word simultaneously activates the elements of the words in the immediate environment (Dell, 1986; Stemberger, 1982/1985).

In Indo-European languages, the most frequent unit of error is a single segment, which accounts for 60-90% of all naturalistic speech errors (Meyer, 1992). An example of this would be the perseveratory speech error of "gave the boy" being uttered as "gave the ***g***oy" (Meyer, 1992) - the speaker still had the word "gave" on their mind as they were uttering "boy," which led to the onset mistake in the latter word. This shows that the units that are selectable are indeed on the segmental level, and not a bigger unit, like the morpheme or syllable.

Other types of errors make up only a small proportion of all phonological errors. In the above example, "gave the ***g***oy", the intrusion of /g/ in "boy," could be analyzed not as a segment error, but a **feature error** - the velar feature from "gave" persevering to the next word. However, feature errors only make up less than 5% of all sound errors (Berg, 1985;

Fromkin, 1973; Shattuck-Hufnagel, 1983; Shattuck-Hufnagel & Klatt, 1979). As for **whole syllable errors**, they are also rare: they make up less than 5% of the speech error corpora (Shattuck-Hufnagel, 1983). An example of this would be the utterance of "journicle article," with the intended phrase being "journal article" (Shattuck-Hufnagel, 1993). **Stress errors** in lexical stress languages, such as English and Dutch, are also rare (Cutler, 1980). An example of this would the utterance "The noise sort of ENvelopes you," in which the EN represents stress on the first syllable of "envelopes," a verb which has lexical stress on the second syllable, "enVElops" (Fromkin, 1976). The incorrect stress on EN in "ENvelopes" is contrastive in this situation, as "ENvelop" is a noun referring to a container for letters, papers, etc.

Additionally, it seems that the same category of segment tends to interact with one another, usually maintaining their positions within syllables. Onset consonants tend to replace other onset consonants. For example, there are many instances of simple onset substitutions, such as "heft lemisphere" (for "left hemisphere"), or consonant cluster substitutions, such as "sloat thritter" (for "throat slitter") (Meyer, 1992). Some resources claim that syllable-onset positions are more error-prone than syllable-coda positions (McKay, 1970), but other resources claim that it's not necessarily the position of a segment that makes it more likely to be misspoken, but rather that less frequent phonemes tend to produce more speech errors (Shattuck-Hufnagel & Klatt, 1979).

Vowels also tend to replace other vowels. In particular, in an analysis on vowel errors, Shattuck-Hufnagel (1986) found that tense vowels (such as /i/ in "Pete") tended to replace other tense vowels, whereas lax vowels (such as /I/ in "lit") tended to replace other lax vowels. Additionally, it was found that there were more vowel errors based on place (back versus front) than in manner (Shattuck-Hufnagel, 1986).

On the contrary, there are also limitations to interactions within syllables: there are no recorded instances of the second consonant of a consonant cluster and the following vowel forming an error unit (Meyer, 1992). This shows that there's a pattern to the structure of speech planning, in which consonants and vowels seem to belong in separate categories.

Chapter 3.1.1c: Theories of phonological encoding in Indo-European languages

Faced with these different findings, linguists have developed different models of phonological encoding to explain them. Among these various models, Levelt's WEAVER++ (Levelt, Roelofs, & Meyer, 1999) model of speech production is perhaps the most influential. The WEAVER++ model states that once the lexical item has been selected, there is a parallel process by which phonemic segments and metrical/syllable frames are separately constructed. The phonemic segments contain the consonants and vowels for that lexical item, and the metrical/syllable frame dictates the number of syllables required, as well as where the lexical stress lies. Once both processes have occurred, there begins a matching process, by which the segments are then inserted into the metrical/syllable frame sequentially from left to right. This process of segment insertion into the metrical frame is when errors may possibly occur. Shattuck-Hufnagel's scan-copier (1979) and Dell's spreading activation models (1986) proposed similar parallel segmental and metrical frame processing mechanisms.

Part of the reason for the separate but parallel processes of segments and metrical frames in these models stems from the fact the languages studied (Dutch and English) make use of lexical stress, a type of prosody in which the addition or lack of stress on one or more syllables in a word can alter the meaning of the word. Stress is often linked to higher pitch, greater amplitude, and longer duration (Booij, 1995; Kenstowicz, 1994; Levelt, 1989; Roach, 2009). For example, in English, the word 'abstract' has two lexical entries: 'ab**STRACT**' (with stress on the second syllable) is a verb, whereas '**AB**stract' (with stress on the first syllable) is a noun.

Previous research on speech errors has consistently shown that lexical stress errors make up a very small portion of all speech errors (Culter, 1980). This led Levelt, in particular, along with other linguists, to argue that stress is not actively encoded in a way comparable to that of segments. The theory posits that the phonological frame is defined by syllables and the stress patterns within a lexical item. Because stress is attached to the metrical/syllable frame, it cannot be misselected, thus making them immune from errors. Segments, on the other hand, are *inserted* into this metrical frame, and this insertion pattern makes them prone to errors.

Faced with evidence from both behavioral studies and speech error compilations that support varying models of phonological encoding, linguists turned their focus on languages with different typological properties to determine whether the patterns seen in languages like English and Dutch are generalizable. Is the proximate unit for selecting a sound in these aforementioned Indo-European languages - the segment - the same in tonal languages? Are the separate phonemic and metrical frames applicable to all languages, or just Indo-European languages? Early studies in Mandarin and Thai seem to tell a different story.

3.1.2: Speech production in tonal languages

3.1.2a: The proximate unit in Chinese through behavioral studies

As mentioned above with Indo-European languages, linguists mainly focused on the proximate unit – whether it's a segment, a feature, or a syllable – and whether or not stress is actively encoded like segments. After conducting implicit priming experiments and analyzing naturalistic speech errors, linguists concluded that in these languages, the proximate unit is the segment and that stress is not actively encoded like segments, but rather that they are intrinsically part of the metrical frame of a word, thus being immune from speech errors.

However, would the results be mirrored in a tonal and orthographic language like Chinese? Linguists once again turned to the implicit priming paradigm to see if Chinese would also show that it minimally takes a segment to elicit priming in speech planning. So far, much of the evidence so far shows that Chinese does not behave in the same way: in the earlier-mentioned implicit priming paradigm that Meyer (1990) and Roelofs (1999) used on Dutch, facilitation for word planning was found when only the initial segment was identical between target words. However, in Chinese, no facilitation has been found when the only overlap is in the initial segments (Chen & Chen, 2002). In a study that tested if overlapping onset consonants would elicit priming like was found in Dutch (Meyer, 1990; Roelofs, 1999), Mandarin words whose response words were overlapping in an onset: 摸彩 (*mo1 cai3*), 麻雀 (*ma2 que4*), 牡丹(*mu3 dan1*), 蜜月 (*mi4 yue4*) did not show any priming. Overlapping tone between target words also did not show facilitation. Instead, Chen and Chen (2002) show that the overlapping segments between target words should minimally be a syllable (without tone) in order to show facilitation. In the same series of experiments from before, Chen and Chen (2002) saw priming when the response words shared the same syllable without sharing the same tone: 飞机 (*fei1 ji1*), 肥胖 (*fei2 pang4*), 翡翠 (*fei3 cui4*), 肺炎 (*fei4 yan2*).

However, in Cantonese, Wong and Chen (2009) showed that rather than having to be a syllable, *just two* overlapping phonemes between the prime and target words showed significant facilitation in the production of the target word. In two studies, which used a **picture-word interference task**, participants were asked to name pictures they are shown while ignoring a distractor accompanying each picture. The distractor items would differ from the picture names in different ways, to investigate what aspects of speech aid in speech processing. In the first experiment, the picture shown would be of a “star” 星 *sing1*, and the distractor items would be overlapping CVC 城 *sing4*, overlapping CV 食 *sik6*, overlapping VC 境 *ging2*, and unrelated 閣 *gok3*. In the second experiment, with the same picture of 星 *sing1*, the distractor items would be overlapping CVT 式 *sik1*, overlapping VCT 京 *ging1*, overlapping VT 必 *bit1*, and unrelated control. Different cohort and rime matches were investigated to examine the effect of those phonemes on speech planning.

Facilitation was only observed when two consecutive segments were identical between the picture and the distractor, and no additional facilitation was found when all three segments in the syllable were the same. No reliable effect was found when only one segment and one tone together was overlapping between the picture and distractor, such as in the overlapping VT (必 *bit1*) distractor items. More importantly, the results showed that there isn't convincing evidence to show that the syllable (without tone) is the essential phonological unit, because the entire syllable (without tone) did not show more facilitation than two overlapping segments (either onset and vowel or vowel and coda). Wong and Chen (2009) thus concluded that it would be safer to assume that the proximate unit is a slightly larger unit than the single segment, but a slightly smaller unit than the syllable.

Although the implicit priming paradigm and word-picture interference paradigms are quite different (the former uses cue words and the latter uses distractor words), the goal is the same: to see what phonological unit is minimally needed to prepare the speech encoding of a target word. In the Chen and Chen (2002) implicit priming experiment, cue words and target words were used in a duo set, but the key was that the target words shared either an onset, an initial syllable, or a tone. They found that the target words had to share an initial syllable for priming to be elicited (that means both the onset and the vowel). In the Wong and Chen (2009) picture-word interference experiment, target words were primed by either an onset, onset and vowel, vowel and coda (all with and without tone matching), and also the full syllable without matching tone. However, in this experiment, the key to facilitation was the vowel - both the onset and the vowel matching conditions, as well as the vowel and coda matching conditions showed facilitation. Because the entire syllable did not show more facilitation than just two consecutively matching segments, Wong and Chen (2009) shows that contrary to the results from Chen and Chen (2002), Chinese does not need an entire syllable to show priming.

3.1.2b: Tonal production in speech error studies

Next, we turn to speech error analysis in tonal languages. Previously, we have seen that stress errors are very rare in English and Dutch (Cutler, 1980), when compared to segmental error. Additionally, they do not seem to be conditioned by the language environment. Thus, in the analysis of speech errors in tonal languages – where or not tones behave like segments in speech encoding – linguists typically refer to two elements: the frequency of tone error as compared to segmental error, and the nature of tone error - whether or not tone error is due to context, namely: if it shows signs of anticipation, perseveration, and substitution of neighboring tones.

One of the earliest studies on the naturalistic speech errors in a tonal language was carried out by Gandour (1977) on Thai speakers. In a study of 350 tone errors, Gandour concluded that tone errors behave similarly to consonant and vowel errors: the majority of tone errors showed preservation, anticipation, and substitution (which Gandour calls “transpositions”), which are the classic signs of contextual speech error in phonemes (p.132). Additionally, Gandour found that tone errors tended to not occur with unstressed syllables, something that is consistent with the evidence from segmental errors (Nooteboom, 1969). However, he found an aspect of tone errors that seems to diverge from segmental errors: perseveratory tone errors outnumber anticipatory tone errors, with a ratio of about 2-to-1. This is in contrast with segmental errors, in which anticipatory errors tend to outnumber perseveratory errors (Fromkin, 1971). Gandour (1977) did not include a number for segmental errors, so we cannot make any conclusions about the frequency of tone errors in comparison to segmental errors.

In subsequent years, naturalistic speech errors were analyzed by linguists in other tonal languages: Mandarin (Moser 1991, Wan & Jaeger 1998, Chen 1999, Wan 2006), Cantonese (Alderete, Chan & Yeung 2019), and Taiwanese (Liu & Wang 2008). These studies calculated both the frequency of tone errors in comparison to segmental errors, as well as analyzed the nature of tone errors. However, these studies seemed to have conflicting results, with some finding few tone errors in comparison to segmental errors (Moser 1991, Chen 1999), and others finding tone errors that were comparable in number to segmental errors (Wan & Jaeger 1998, Wan 2006).

In the former camp, Chen (1999) found a substantial lack of tone errors in his Taiwan Mandarin speech error corpus and concluded that tones innately behave differently from segments. In his study of speech errors from radio recordings, he found 24 tone errors out of a total of 987 speech errors (tone errors were thus only 2.4% of all speech errors). However, out of the 24 errors, he believes that 19 of those are not actually tone errors, but errors of another nature: character blending, haplology, malapropism, and misapplication of tone sandhi. Thus, out of the 24 suspected tone errors, he only found four 5 true tone errors, 1 of which were of the anticipatory type, and 4 of which were of the perseveration type. This is in alignment with what Gandour (1977) found from his study of naturalistic tone errors in Thai, who also found more perseveration errors than anticipation errors.

Regarding the tone errors that are not true tone errors, Chen additionally alleges that tone errors previously cited as movement errors by other linguists (by Gandour, 1977; among others) were in fact errors of these types. An example of this is the utterance of "*nyu35** tong35 xue35" (*nü T2* + tong T2 + xue T3, meaning "*female* student"; correct pronunciation:

nü T3 + tong T2 + xue T3), which Wan (1996) claims is a error of anticipation (movement) of the tone of the following word: tong T2 (tong35). However, Chen (1999) argues that this tone error could be a simple blending of two competing words: the speaker may have intended to say both female 'nü T3' and male 'nan T2,' and ended up saying a blend of the two: nü T2.

As for segmental errors, Chen found 136, which is significantly higher than the number of tonal errors. However, although Chen subjects the suspected tone errors (of which there are 24) to extreme scrutiny, attributing them to anything but movement errors, he does not subject the segmental errors to the same scrutiny. In fact, he attributed all 136 of the suspected segmental errors to movement error due to context (anticipation, perseveration and exchange). Due to the much greater number of segmental errors as compared to tone errors found in the study, this seems unlikely, as previous studies in spontaneous segmental errors also found that a portion of the errors could not be attributed to movement (Gandour, 1977).

Regardless, facing the results that show both a significant lack of tone errors, as well evidence seeming to show that tone errors behave differently from segmental errors, Chen (1999) adopts the previously mentioned speech planning model of WEAVER++ (Levelt) and adapts it for tonal languages. According to Chen, in tonal languages, tone behaves similarly to stress and is encoded in a metrical frame context, not being a part of the phonological encoding process until after segments are first selected from mental storage and inserted into phonological frames. According to Chen (1999), when a tone is associated with a single syllable, like in Mandarin, the relationship between the two is unique and no mapping process occurs, "exempting Mandarin from possible tone errors" (p.295). Similar to previous research arguing that stress errors don't occur nearly as much as segment errors in Indo-European languages, in which stress defines the metrical/syllable frames, tones define the phonological frame in Chinese, according to Chen (1999), in which segments are inserted into. This insertion process, in which segments are matched to the syllable to which they are associated, renders them vulnerable to error. Tones, which define the frame, are thus not subject to errors.

However, a recent study on Cantonese speech errors (Alderete et al., 2019), the biggest of its kind for an Asian language, provides evidence that challenges Chen's conclusions. Using podcast episodes that totaled 1917 minutes of spontaneous speech, Alderete et al. found 2462 total speech errors, of which 432 were tonal and 1357 were segmental. Out of the 1789 tonal or segmental errors, approximately 24% were tonal. Compared to the 2462 speech errors found, which include errors such as lexical errors, tones make up about 18% of all errors found (compare this with 2.4% tonal error from the Chen 1999 study). Additionally, Alderete et al. (2019) found many instances of tone error due to the context (anticipation, perseveration, and substitution) and asserts this is additional evidence showing that tone is actively encoded during the speech planning process.

As the tone error ratio reported in this report is in strong contrast to the earlier study by Chen (1999), Alderete et al. counters that Chen (1999)'s study that contained only 24 tone errors had a low amount of segmental or tonal errors in general (16.2% of all errors). Out of 987 total slips of the tongue errors, only 160 were segmental or tonal errors (16.2% of all errors). Out of those 160 sound errors, 24 were tone errors, which makes up 15% of all segmental or tonal errors. According to Alderete et al., the ambiguous tone errors from that study that may have had alternative analyses not based on tone mis-encoding still make up a sizable portion of all sound errors. This, in addition to their findings from their own study shows that the amount of tone errors (18% of all speech errors) found is not a negligible amount of errors, and cannot be possible if tones are not actively selected in the speech planning process.

However, Alderete noted that although tones are actively selected alongside segments, tones are assigned to a syllable after segments are. Additionally, Alderete conceded that there’s a chance that tone encoding is not quite equivalent to segmental encoding, noting that tone errors have a higher chance of coexisting with segmental errors (⅓) versus segmental errors coexisting with tone errors (⅙).

3.1.2c: Theories of phonological encoding in Chinese

In studying tones in phonological encoding, linguists have turned to evidence arguing for the inclusion or exclusion of tone in spoken word planning. There's growing evidence showing that Chinese doesn’t behave similarly to Indo-European languages in spoken word planning.

As mentioned, Chen and Chen (2002) found that Mandarin needs minimally a syllable shared in implicit priming tasks to elicit facilitation. Due to these results, O’Seaghdha, Chen & Chen (2009, 2010) argued the proximate unit in Mandarin to be the segmental syllable. According to O’Seaghdha, Chen & Chen, the proximate unit in a language is the first selectable phonological unit below the level of the word/morpheme. They are selected early in the process of speech planning and are prone to selection errors. In Dutch and English, the proximate unit would be the segment - meaning that below the stage of word/morpheme access, the first selectable unit is the segment. Mandarin, a language whose orthography uses syllable-sized characters that do not break down into segments, has always been seen as being prone to different sublexical behavior, especially in reading (Zhou & Marslen-Wilson, 1999). If the results of these studies are true, and that the proximate unit in Chinese is the syllable, then this would be a major finding showing that speech production isn’t universal among languages. However, research by Wong and Chen (2008, 2009) seems to challenge this theory, at least for Cantonese.

The converging evidence from both speech error and behavioral studies then prompted the aforementioned speech production model WEAVER++ (Level, Roelofs, & Meyer, 1999) to be modified to fit Chinese (Roelofs, 2014). The WEAVER++ model, which originally named phonemic segmentals as the fundamental phonological unit paired with metrical stress frames in Germanic languages, was changed to syllables as the fundamental phonological unit paired with tonal frames in Chinese. Up until now, there is a near-consensus in the linguistics community that Chinese's fundamental phonological unit is the toneless syllable, and that tone is encoded in a separate tonal frame and inactively encoded (Chen et al., 2002; Chen et al., 2003; O'Seaghdha et al., 2009, 2010; Chen & Chen, 2012; Roelofs, 2014).

The previous research so far has used both spontaneous speech error rates and experimental studies to examine the role of tone in phonological encoding. Regarding speech error rates, there seems to be two competing camps: the former camp views tone error as infrequent and proposes that tone is inactively encoded, using a metrical frame context (Moser 1991, Chen 1999), and the latter camp views tone error as frequent and proposes that tone is actively encoded, similarly to segments (Wan & Jaeger 1998, Wan 2006, Alderete et al. 2019). Because the methods involved in spontaneous speech error curation and classification vary from study to study, there still remain questions surrounding this topic.

The next part of the chapter will introduce a specific subset of speech errors, one that I believe produces a more rigorously controlled set of data from which we can make judgments about whether tone is actively or inactively encoded.

3.2: Tongue twisters as an alternative paradigm for studying phonological encoding

Despite the numerous behavioral and speech error studies employed so far, it is difficult to make conclusions about the role of tone encoding based on these studies. Both implicit priming studies and picture-word interference tasks tell us about which phonemic units help us plan speech, but do not help us see the number of tone errors in comparison to segmental errors, which is a key indicator for evidence supporting either the active or inactive encoding of tone. Thus, we need to turn to speech error studies.

The previously mentioned naturalistic speech error studies are unfortunately subjective in nature. First of all, it's difficult to capture all the speech errors in any given large-scale corpus, as linguists may debate what is or is not considered a "slip of the tongue" error. First, the speech background of speakers may be diverse, with no reference as to if they speak a dialect other than Mandarin. There are instances in previous Chinese naturalistic speech error research that found errors, but were not sure if they were speech errors or non-standard speech due to a "speaker's accent," and were excluded from analysis. Additionally, other originally reported errors were excluded from analysis because they were concluded to be due to a speaker's "change of intention" (Chen, 1999). Another reason Chen (1999) decided to exclude errors from analysis was due to speakers' "memory lapses while quoting numbers" (p.292) - there is no provided transcript of their speech corpus, so we'd have to take his word that this is indeed the case. Clearly, there are several factors that a linguist may or may not take into account when even citing a non-standard spontaneous speech utterance as an error or not.

More importantly, it's difficult to properly categorize the nature of all speech errors, as different linguists may choose different ways to classify speech errors, leading to vastly different results. Some speech errors may be simple to classify (e.g., "environmental **en**cern" is clearly an example of preservation), but others may be less straight-forward (e.g., is "Let's do a final **RE**cording of the track?" a preservation of stress pattern from the previous word, or a lexical mistake due to the later word “track”? Record, or LP, and track are both music-related, which would be a mistake based on lexicality.). As mentioned earlier, Chen (1999) and Alderete et al. (2019) clashed on whether tonal errors observed were based on movement or not. Chen (1999) had alternative explanations for tone errors as cited by other linguists, but without knowing what the respective speakers had intended to say, it is impossible to know whose explanation is correct.

That’s why although the previous analyses on naturalistic Mandarin and Cantonese speech errors have been informative, but because they are unable to pinpoint what the speakers actually intended to say, it’s impossible to know if a “suspected tone error” is actually a tone error, let alone allow us to make conclusive statements about the similarities and differences between segmental and tonal errors.

3.2.1: The use of tongue twisters in speech research

The use of tongue twisters in experimental research has been around for decades. Tongue twister experiments conducted starting from the 1980s showed that although error rates from tongue twister studies tend to be higher compared to spontaneous speech error rates, the errors from tongue twisters are modulated by the same types of contexts as those stemming from spontaneous speech when speed was slowed down. Commonly, these experiments control the rate of speech using a metronome, set at a rate that is comfortable for natural speech production. Wilshire (1999) used 1.67 syllables per second, or 100.2 bpm, and was able to elicit errors.

Early tongue twister experiments forced participants to memorize a string of target words and then recite them from memory (Shattuck-Hufnagel, 1973), so there was concern that the tongue twister task relied on short-term memory and focused on recall errors, rather than pure “slips of the tongue.” However, subsequent studies that keep the tongue twister to be read aloud constantly on display have essentially eliminated this possible concern (Wilshire, 1999), and there is substantial evidence that tongue twister errors resulting from this task paradigm are shown to come from a pre-articulatory, phonological encoding origin (Wilshire, 1998). Instead of reading each word individually, anticipatory errors suggest that participants planned ahead to read target strings. All of this previous evidence shows that the tongue twister is an efficient yet effective way to test phonological encoding.

One method used to elicit speech error is the phenomenon known as the **repeated phoneme effect**. Supporters of the repeated phoneme effect argue that a shared sound in two neighboring words increases the rate of errors involving other sounds in those two words (Dell, 1984; MacKay, 1970). For example, in the phrase “deal beak,” the fact that the two words share the vowel [i] increases the error in production of the consonants in “deal” - there is the increased chance of the [d] being substituted by [b], as well as [l] being substituted by [k]. The reason for this is that by having the repeated phoneme [i] in the two words, the anticipatory activation of the second word increases, so it increases the flow of activation between the two words, which increases the likelihood that [b] in the second word will intrude on [d] in the first word, as well as increasing the likelihood that [k] in the second word will intrude on [l] in the first word.

This effect has been shown to be effective in tongue twister studies (Wilshire, 1999), which have used ABAB and ABBA patterns, such as *‘**s**oap dam **s**eam dip,’* which uses the ABAB pattern to alternate the onsets, as well as the classic ‘***S**he sells sea **s**hells,’* in which the onset follows an ABBA pattern. According to most modern models of speech production, active encoding involves the dynamic selection of a target word and its surrounding linguistic context - we can see that the alternating onsets of these words show evidence for active encoding for all the sounds in their respective tongue twisters.

Thus, if tongue twisters of this nature involving lexical tone were also shown to elicit a high rate of errors, then we have strong evidence to argue that tone is indeed encoded actively in the speech preparation process. I will use this foundation with which to test my hypothesis.

3.2.2: Tongue twisters in Mandarin Chinese

Previously, the majority of tongue twister studies have been done in English, but Kember et al. (2015) recently extended the tongue twister method to Mandarin Chinese, in a study that was methodologically sound and stringently carried out.

In order to see if segments and tones behave differently in phonological encoding, Kember et al. designed tongue twisters that either elicited onset error, tone error, or a combination of the two. Specifically, the study was stringent in that the tongue twisters afforded the same opportunities for error to both onsets and tones. Kember et. al used the classic ABAB and ABBA tongue twister pattern known to elicit speech error, in a 6-alternation design that included 4 "easy" conditions that only alternated either segment or tone and 2 "complex" conditions that alternated both segment and tone. Example stimuli from their study are shown in the following table (Table 4).

	Alternation type		**Example**	
	Segment	Tone	Character	Pinyin
Segment-alternating ABAB	ABAB	Constant	突苦突苦	tu1ku1tu1ku1
Segment-alternating ABBA	ABBA	Constant	突苦苦突	tu1ku1ku1tu1
Tone-alternating ABAB	Constant	ABAB	突土突土	tu1tu3tu1tu3
Tone-alternating ABBA	Constant	ABBA	突土土突	tu1tu3tu3tu1

ABAB(S)
ABBA(T)
	ABAB	ABBA	突苦土哭	tu1ku3tu3ku1

ABAB(T)
ABBA(S)
	ABBA	ABAB	突苦哭土	tu1ku3ku1tu3

Table 5: Kember et al. (2015) Tone and Onset Alternation and Format Conditions. *Table from Kember et al. (2015).*

Kember et al. (2015) hypothesized that if tone was actively encoded, then tone error would be comparable to that of onset error; and that if tone was not actively encoded in speech encoding (but more like lexical stress), then their errors would be a lot less frequent than that of onset error. In their study, they found that tones and onsets were given the same opportunities for error, they found that there were 3503 onset errors (72% of all phonetic errors) to 1372 tone errors (28% of all phonetic errors). Thus, they concluded that onsets (and thus segments) were more prone to error than tones, and that tones are encoded akin to that of lexical stress, associated to a metrical frame.

With such seemingly convincing evidence, it would make sense to shut the book on this debate. However, the overwhelming evidence showing that tones are not actively encoded may be rooted on a shaky phonological foundation. In the next section, we discuss why onset errors may not be the right segmental error against which to compare tone errors.

3.3: Motivation for current study: why focus on onsets?

3.3.1: Explanation of new methods & new scoring system

Much research on segmental speech errors focuses on onsets. According to Wilshire (1998), word onsets make up anywhere from 50% to 90% of spontaneous speech errors (based on speech corpora analysis by MacKay, 1970 and Shattuck-Hufnagel, 1987). This phenomenon, in which onsets of words tend to err more frequently than segments in other word positions, became known as the 'word onset effect.' Linguists have subsequently been trying to explain the reason for this phenomenon for the next 50 years. Remarkably, in the languages studied (English and Dutch), only the onset of a word, and not the onset of a syllable not at the beginning of a word, showed elevated levels of error. Additionally, Wilshire (1998) discovered that the word onset effect only occurs with real words, not word-like non-words. According to Wilshire, there seems to be something underlying spoken word formation that leads to high levels of word onset error, disassociated from the phonological features of onsets.

Thus, whether or not it's a coincidence that Kember et al. (2015) decided to use onsets to represent segments in their segment versus tone tongue twister study, one cannot know. However, approaching the study in this way may lead to spurious conclusions that tones behave differently from segments. By focusing on onsets only, linguists are not studying how tones behave differently from other types of segments, such as vowels or codas.

So far, tonal languages have mainly examined the effect of tones, consonants and vowels in spontaneous speech error. The Kember et al. (2015) study is the first one to look at segmental error specifically at the onset position. There remains more to be explored in other word position errors, such as vowels and codas.

Secondly, I have shown in the first part of my experiment that different segments behave differently in the phonological perception of Mandarin Chinese. In Kember et al. (2015)'s study, onsets were treated as the 'de facto' segment, to represent all segments. However, my previous set of 4 experiments has shown that onsets, vowels and codas behave quite differently in perception. This evidence gives us ample reason to believe that these 3 categories of segments may also behave differently in speech production.

Thus, my suggestion is that segments should be separated into 3 categories: onsets, vowels and codas; with tone being the last category; for a total of 4 categories of possible speech error. A score will be given for all 4 categories: an onset error rate, a vowel error rate, a coda error rate, and a tone error rate that averages the error rates from the 3 condition categories (tone alternating with onset, vowel and coda). Additionally, I will look at the tone error rates separately within each of the 3 categories. If tone truly is encoded inactively, its error rates should be the lowest of the bunch.

3.3.2: My Hypothesis

My hypothesis is as follows:

If tones are encoded actively, its error rates should not be the lowest of the 4-category set. If tones are encoded inactively, its error rates should be the lowest of the 4-category set.

3.4: Methods of Current Production Study (Experiment 5)

3.4.1: Participants (N=35)

The 35 participants (20 female, 15 male; aged between 18 and 35, average age: 25 years) for the tongue twister study were all native Mandarin speakers from Mainland China who resided in Hong Kong (or previously resided in Hong Kong) while attending university or working. Due to the 3rd wave pandemic in Hong Kong during the time of the data collection, participants recorded their tongue twisters in their home, but stringent instructions were given to participants and participants had to complete a pre-experiment test to make sure they were fluent Mandarin speakers, understood the experiment instructions properly, and had the proper equipment to carry out the recording successfully. 40 participants failed one or more of the above conditions and were excluded from the analysis.

The 35 participants came from different areas in China, and 30 self-reported as speaking a local dialect with family. They were reimbursed for their participation.

3.4.2: Materials

The tongue twisters used in this experiment followed the format of the tongue twisters from Kember et al. (2015). There were 6 condition types for each of the 3 categories: tone and onset alternation, tone and vowel alternation, and tone and coda alternation. For each category, there were 5 sets each, for a total of 90 different tone twisters.

Additionally, the stimuli from the Kember et al. (2015) study were kept in as a point of comparison: in the case that I get different results for tone and onset error, I would be able to see if it was due to the fact that I used different stimuli.

The method by which we selected the tongue twisters was simple - we wanted to select alternating words that would be difficult for native Mandarin speakers to say, while maintaining the rigor we needed with which to test my hypotheses. The following parameters were controlled when deciding on my tongue twister stimuli:

1. Kember et al. (2015) did not include Tone 2 in any of the materials. Because Mandarin only has 4 lexical tones, not including Tone 2 would be missing ¼ of the items in this phoneme category, so the results may not be representative of how tone works in Mandarin in general. Thus, I sought to include Tone 2 commensurate with

how often it occurs in the language (¼ of the time). Out of all the stimuli, 8 out of the 15 tongue twisters included Tone 2.

2. The Kember et al. (2015) study focused on onset contrasts that primarily contrasted in aspiration (3 out of 5 material sets) or place of articulation (2 out of 5 material sets). With my material set, I aimed to include a wider range of feature contrasts in my segment contrast items, including contrasts that tend to actually give native speakers pronunciation issues. I consulted real tongue twisters in Mandarin (list found in Appendix D), and a majority of the tongue twisters were inspired by the contrasts used in natural tongue twisters.
3. Mandarin Chinese has systematic tone change known as third tone sandhi, when two Tone 3 syllables are placed together sequentially - the first of the syllable pair turns into Tone 2. In other words, in a word like *ni3hao3*, the first of the 2-syllable pair turns into Tone 2, and the word is pronounced *ni2hao3*. However, tongue twisters may or may not be subject to the tone sandhi phenomenon, since the task is dissimilar to natural speech production. Due to this fact, because the goal of this experiment is to find tone speech error in comparison to segmental speech error, I tried my best to avoid the placement of two Tone 3's next to each other. However, due to the fact that I tried to place every tone in every position , I did end up having two tongue twisters that contained 1 or more conditions for the existence of Tone 3 sandhi change: 1. A set of ABAB(Segment)/ABBA(Tone) with 2 Tone 3's in the center of the tongue twister: 踢塔体他; and 2. a set of 4 Tone 3 characters: 访反访反 (tone and coda alternation ABAB pattern). In my analysis, I counted Tone 2 utterances correct in the first of a two-Tone 3 chain, and for the 4-character chain of Tone 3 characters, I counted Tone 2 utterances of the first 3 characters in each repetition as correct.

a. In the end, I believe that allowing some opportunities for tone sandhi in the tongue twisters was not a setback, but an opportunity - would tone sandhi surface in a situation where speech was not spontaneously planned? I calculate the instances of tone sandhi production in possible tone sandhi contexts in section **3.5.3**.

b. The Kember et al. (2015) study only found 5 instances of tone sandhi utterances in their entire study. We will use this number as a baseline by which to compare my results.

4. I tried my best to use characters that did not resemble each other in orthography, as that may add additional difficulty to an already difficult tongue twister task. However, a majority of Chinese characters are phono-semantic, which means that there is usually one radical that refers to pronunciation, with another referring to meaning. For the few that did contain the same radicals in two or more characters, they were not placed in the same position (left/right/top/bottom/inside/outside/solo).

The following tables outline the materials used for the experiment:
(Note: all items are written with the standard pinyin system, not with the IPA system.)

					Segment		Tone		Complex	
									ABAB(S)	ABBA(S)
					ABAB	ABBA	ABAB	ABBA	ABBA(T)	ABAB(T)
tu/	突	哭	土	苦	突哭突哭	突哭哭突	苦土苦土	突土土突	突苦土哭	突苦哭土
ku	tu1	ku1	tu3	ku3	tu1ku1tu1ku1	tu1ku1ku1tu1	ku3tu3ku3tu3	tu1tu3tu3tu1	tu1ku3tu3ku1	tu1ku3ku1tu3
					始市始市	始市市始				
shi/	始	市	死	四	shi3shi4	shi3shi4	四死四死	始死死始	始四死市	始四市死
si	shi3	shi4	si3	si4	shi4shi4	shi4shi3	si4si3si4si3	shi3si3si3shi3	shi3si4si3shi4	shi3si4shi4si3
					板办板办	板办办板	慢满慢满	板满满板	板慢满办	板慢办满
ban/	板	办	满	慢	ban3ban4	ban3ban4	man4man3	ban3man3	ban3man4	ban3man4
man	ban3	ban4	man3	man4	ban3ban4	ban4ban3	man4man3	man3ban3	man3ban4	ban4man3
					考靠考靠	考靠靠考	告搞告搞	考搞搞考	考告搞靠	考告靠搞
kao/	考	靠	搞	告	kao3kao4	kao3kao4	gao4gao3	kao3gao3	kao3gao4	kao3gao4
gao	kao3	kao4	gao3	gao4	kao3kao4	kao4kao3	gao4gao3	gao3kao3	gao3kao4	kao4gao3
					推退推退	推退退推	对堆对堆	推堆堆推	推对堆退	推对退堆
tui/	推	退	堆	对	tui1tui4	tui1tui4	dui4dui1	tui1dui1	tui1dui4	tui1dui4
dui	tui1	tui4	dui1	dui4	tui1tui4	tui4tui1	dui4dui1	dui1tui1	dui1tui4	tui4dui1

Table 6: The original Kember et al. stimuli.

					Segment		Tones		Complex	

					ABAB	ABBA	ABAB	ABBA	ABAB(S)
									ABBA(T)
										ABBA(S)
										ABAB(T)

	拔	爬	八	趴	拔爬拔爬	拔爬爬拔	八拔八拔	拔八八拔	拔趴八爬	拔趴爬八
ba/pa (2,1)	ba2	pa2	ba1	pa1	ba2pa2ba2pa2	ba2pa2pa2ba2	ba1ba2ba1ba2	ba2ba1ba1ba2	ba2pa1ba1pa2	ba2pa1pa2ba1
	起	洗	齐	习	齐习齐习	齐习习齐	齐起齐起	起齐齐起	起习齐洗	起习洗齐
qi/xi (3,2)	qi3	xi3	qi2	xi2	qi2xi2qi2xi2	qi2xi2xi2qi2	qi2qi3qi2qi3	qi3qi2qi2qi3	qi3xi2qi2xi3	qi3xi2xi3qi2
	湖	福	户	父	湖福湖福	湖福福湖	户湖户湖	湖户户湖	湖父户福	湖父福户
hu/fu (2,4)	hu2	fu2	hu4	fu4	hu2fu2hu2fu2	hu2fu2fu2hu2	hu4hu2hu4hu2	hu2hu4hu4hu2	hu2fu4hu4fu2	hu2fu4fu2hu4
	汁	资	制	字	汁资汁资	汁资资汁	制汁制汁	汁制制汁	汁字制资	汁字资制
zhi/zi (1,4)	zhi1	zi1	zhi4	zi4	zhi1zi1zhi1zi1	zhi1zi1zi1zhi1	zhi4zhi1zhi4zhi1	zhi1zhi4zhi1zhi1	zhi1zi4zhi4zi1	zhi1zi4zi1zhi4
	歌	科	格	壳	歌科歌科	歌科科歌	格歌格歌	歌格格歌	歌壳格科	歌壳科格
ge/ke (1,2)	ge1	ke1	ge2	ke2	ge1ke1ge1ke1	ge1ke1ke1ge1	ge2ge2ge2ge1	ge1ge2ge2ge1	ge1ke2ge2ke1	ge1ke2ke1ge2

Table 7: Tone and onset alternation stimuli. *The numbers (1,2,3,4) next to the item indicate the tones used in the tongue twister. E.g. ba/pa (2,1) means alternating ba Tone 1 and ba Tone 2 with pa Tone 1 and pa Tone 2.*

					Segment		Tones		Complex	

					ABAB	ABBA	ABAB	ABBA	ABAB(S)
									ABBA(T)
										ABBA(S)
										ABAB(T)

ba/bu (3,4)	把 ba3	补 bu3	爸 ba4	部 bu4	爸部爸部 ba4bu4ba4bu4	爸部部爸 ba4bu4bu4ba4	爸把爸把 ba4ba3ba4ba3	把爸爸把 ba3ba4ba4ba3	把部爸补 ba1bu4ba4bu3	把部补爸 ba3bu4bu4ba4
lu/lü (3,4)	卤 lu3	旅 lü3	鹿 lu4	绿 lü4	鹿绿鹿绿 lu4lü4lu4lü4	鹿绿绿鹿 lu4lü4lü4lu4	鹿卤鹿卤 lu4lu3lu4lu3	卤鹿鹿卤 lu3lu4lu4lu3	卤绿鹿旅 lu3lü4lu4lü3	卤绿旅鹿 lu3lü4lü3lu4
ti/ta (1,3)	踢 ti1	他 ta1	体 ti3	塔 ta3	踢他踢他 ti1ta1ti1ta1	踢他他踢 ti1ta1ta1ti1	体踢体踢 ti1ta1ti1ta1	体踢踢体 ti3ti1ti1ti3	踢塔体他 ti1ta3ti3ta1	踢塔他体 ti1ta3ta1ti3
mei/mai (3,2)	美 mei3	买 mai3	枚 mei2	埋 mai2	枚埋枚埋 mei2mai2meimai2	枚埋埋枚 mei2mai2mai2mei2	枚美枚美 mei2mei3meimei3	美枚枚美 mei3mei2mei2mei3	美埋枚买 mei2mai2mei2mai3	美埋买枚 mei3mai2mai3mei2
zhen/zhan (1,4)	真 zhen1	沾 zhan1	振 zhen4	战 zhan4	真沾真沾 zhen1zhan1zhen1zhan1	真沾沾真 zhen1zhan1zhan1zhen1	振真振真 zhen4zhen1zhen4zhen1	真振振真 zhen1zhen4zhen4zhen1	真战振沾 zhen1zhan4zhen4zhan1	真战沾振 zhen1zhan4zhan1zhen4

Table 8: Tone and vowel alternation stimuli. *The numbers (1,2,3,4) next to the item indicate the tones used in the tongue twister. E.g. ba/bu (3,4) means alternating ba Tone 3 and ba Tone 3 with bu Tone 3 and bu Tone 4.*

					Segment		Tones		Complex	

					ABAB	ABBA	ABAB	ABBA	ABAB(S)
									ABBA(T)
										ABBA(S)
										ABAB(T)

					音英音英	音英英音	银音银音	音银银音	音赢银英	音赢英银
	音	英	银	赢	yin1ying1	yin1ying1	yin2yin1	yin1yin2	yin1ying2	Yin1ying2
yin/ying (1,2)	yin1	ying1	yin2	ying2	yin1ying1	ying1yin1	yin2yin1	yin2yin1	yin2ying1	ying1yin2
					访反访反	房烦烦房	房访房访	访房房访	访烦房反	访烦反房
	访	反	房	烦	fang3fan3	fang2fan2	fang2fang3	fang3fang2	fang3fan2	fang3fan2
fang/fan (3,2)	fang3	dan3	fang2	fan2	fang3fan3	fan2fang2	fang2fang3	fang2fang3	fang2fan3	fan3fang2
					笨蹦笨蹦	笨蹦蹦笨	奔笨奔笨	笨奔奔笨	笨崩奔蹦	笨崩蹦奔
	笨	蹦	奔	崩	ben4beng4	ben4beng4	ben1ben4	ben4ben1	ben4beng1	ben4beng1
ben/beng(4,1)	ben4	beng4	ben1	beng1	ben4ben4	beng4ben4	ben1ben4	ben1ben4	ben1beng4	beng4ben1
					残藏残藏	残藏藏残	参残参残	残参参残	残仓参藏	残仓藏参
	残	藏	参	仓	can2cang2	can2cang2	can1can2	can2can1	can2cang1	can2cang1
can/cang (2,1)	can2	cang2	can1	cang1	can2cang2	cang2can2	can1can2	can1can2	can1cang2	cang2can1
					冰宾冰宾	冰宾宾冰	病冰病冰	冰病病冰	冰鬓病宾	冰鬓宾病
	冰	宾	病	鬓	bing1bin1	bing1bin1	bing4bing1	bing1bing4	bing1bin4	bing1bin4
bing/bin (1,4)	bing1	bin1	bing4	bin4	bing1bin1	bin1bing1	bing4bing1	bing4bing1	bing4bin1	bin1bing4

Table 9: Tone and coda alternation stimuli. *The numbers (1,2,3,4) next to the item indicate the tones used in the tongue twister. E.g. yin/ying (1,2) means alternating yin Tone 1 and yin Tone 2 with ying Tone 1 and ying Tone 2.*

3.4.3: Procedure

Due to the 3rd wave Coronavirus pandemic during the data collection phase of this study, I asked participants to take part in the experiment from their homes. Each participant had to make sure they had working personal computers with sound recording equipment. The operating system of the computer was inconsequential to the running of the experiment. Because the tongue twister task requires a specific procedure, as well as proper recording of the tongue twisters, I prepared an extensive instructions guide, as well as a pretest recording session, to make sure that the participants met the requirements necessary to contribute properly to this study.

First, I sent participants a demographic survey to fill out, where they filled out their age, birthplace, and if they spoke any other dialects. The participants had to then ensure that they were in a quiet room, with access to a computer outfitted with an adequate sound recording system. Then, they were asked to record a **practice test** using Microsoft Powerpoint. Participants were asked to play, in the background, an mp3 sound file consisting of a metronome beat of 160 bpm, a rate that was used in previous tongue twister studies and judged to be a proper speed to elicit natural speech (Kember et al., 2015; Croot et al., 2010). The PowerPoint consisted of 100 slides, with each of the 120 words used in the actual experiment on a single slide, plus 12 slides with four-word idioms on them, for them to get used to the four-word tongue twister format. The participants were told to keep on the 160 bpm metronome beat, and to repeat the words on each slide 3 times. The audio recordings were recorded directly into each of the 100 PowerPoint slides.

I then reviewed each file individually to make sure the articulation of the words was accurate, as well making sure the sound files were loud and clean enough for proper speech error transcription. As mentioned before, a total of 75 participants volunteered to take part in the experiment, but through the practice test, 40 participants were deemed to either not have the proper equipment for the experiment, could not follow the instructions precisely, or had inaccurate articulation of the practice test material.

Moving onto the actual experiment, the procedure was exactly the same as that of the practice test, except the actual experiment had a total of 132 tongue twisters[8], with each tongue twister presented in large font on individual separate PowerPoint slides (for a total of 132 slides). Participants were given very explicit directions on how to read aloud the tongue twisters:

1. Each tongue twister is to be read 3 times.
2. Each word (consisting of one syllable/Chinese character) must be on beat with the 160 bpm metronome beat.
3. If an error is made, do not stop and reread the error word. Continue to read the rest of the four-word tongue twister. Do not read each tongue twister more than 3 times.

Having participants read each tongue twister 3 times is a slight deviation from the Kember et al. (2015) study that this experiment is based on (as in that study, each tongue twister was read 6 times), but because the current experiment has so many more tongue twisters, reducing the number of repetitions per tongue twister would be a more efficient use of time and energy.

Additionally, each PowerPoint given to the participants had a randomly shuffled order of tongue twisters, to ensure there is no effect of a specific order. No participant received the same order of the tongue twisters.

[8] The actual experiment consisted of another condition, in which tone alternated with open vowels (vowels with no coda and no onset). This condition was later taken out of analysis, as it conflicted with the numbers of available segment to tone alternation conditions.

3.4.4: Speech Error Transcription and Coding

Speech errors were transcribed by the author, who is a native Mandarin Chinese speaker who was born and grew up in Changchun, Jilin. For each of the 35 participants, each of their 132 tongue twisters were listened to for a minimum of 3 times. After listening to the tongue twister 3 times, I transcribed the three repetitions of the tongue twisters into Pinyin and compared it to the intended tongue twister. Finally, the errors for each tongue twister were counted up.

Error Type	Description	Example	
		Intended	Actual Utterance
Tone	Intended tone substituted with wrong tone. Each tone substitution counts as 1 tone error.	fang3 fan2 fan3 fang2	fang3 fan2 fan**2** fang3
Segment (Onset, Vowel or Coda)	Intended segment substituted with wrong segment.	hu2 fu4 fu2 hu4	**f**u2 fu4 fu2 hu4
Hesitation	Improper break/hesitation in the flow of the reading. Marked with // or with –.	hu2 hu4 hu4 hu2	**h-h**u2 hu4 **//** hu4 hu2
Reading	Any 2+ consecutive misreadings of the same repetition. Each syllable with an error is counted as 1 reading error.	zhen1 zhan4 zhan1 zhen4	3x: **zhen4** zhan4 **zhan4** zhen4
Omission	Where a syllable or entire repetition is omitted. 1 syllable counts as 1 omission error; all 4 syllables count as 4 errors.	ti1 ta1 ti1 ta1	3x: ti1 __ ti1 ta2

Table 10: Types of Error Coded.

Errors were coded based on 2 overarching categories: **phonological errors** and **non-phonological speech errors**. Under the umbrella of phonological errors, there are 4 categories of errors: onset, vowel, coda and tone. Under the umbrella of non-phonological speech errors, there are 3 categories of errors: hesitation, omission, and repeated incorrect readings. This is a major change from the Kember et al. method of coding, which ignored all segments that weren't onsets (vowels and codas). An error is counted when there is any deviation from the intended target - the tongue twister at present on the screen at the time. For the purpose of coding, the errors of interest included only the phonetic ones.

The errors that were deemed to be speech errors, such as hesitation, omission and repeated incorrect reading errors, were not included in the overall phonetic error rates. A proper explanation of the classification of speech errors is as follows.

Hesitation errors are counted whenever someone makes a mistake in the middle of a tongue twister reading. Because the rate of reading is kept constant by the 160 bpm metronome beat, if a participant starts reading on a beat, they have to continue to read all 3 repetitions in one go. If they stop anywhere in the middle of a reading and take one or more beats before finishing a reading, a hesitation error is counted. However, as mentioned before, this is considered a speech error, and was not counted within the critical phonetic error rate calculation.

As for omission, since each tongue twister was read 3 times, there are a total of 4 x 3 = 12 words in each tongue twister. An omission error is counted whenever an entire syllable (or word) is omitted. For example, in the tongue twister (*can1 can2 can1 can2*), each repetition that is missing one of the syllables (*can1* or *can2* in any position) is counted as 1 omission error. Two syllables missing in a reading would be counted as two omission errors.

In the instance where a participant reading a tongue twister incorrectly produces two repetitions (or more) in exactly the same way, each incorrect repetition would be counted as a reading error. For example, in the tongue twister (*can1 can2 can1 can2*), if the first two repetitions were read as (*can1 can1 can1 can2*), then it would be counted as 2 incorrect reading errors (See Table 9). See Table 10 for the total amount of tone reading errors, total amount of segment reading errors, as well as the average rate of tone errors (computed for each tone and segment tongue twister pair.

The reason for counting incorrect reading errors is due to the fact that a sustained error that is repeated cannot be wholly attributed to a phonological origin; it's possible that the participant simply misread the entire character chain incorrectly and it would be safer to just discard the entire chain as incorrect readings, rather than try to extract information about the encoding/articulatory planning process that I'm interested in.

	Average Rate of Tone Errors	Total Tone Reading Errors	Total Segment Reading Errors
Onset vs Tone Tongue Twisters	59.7%	195	101
Vowel vs Tone Tongue Twisters	79.9%	315	67
Coda vs Tone Tongue Twisters	43.0%	342	661

Table 11: Summary of reading errors, broken down by tone-segment tongue twister type.

With all these speech errors in mind, an example of a portion of some coded tongue twister data can be found in Table 11.

Type of Tongue Twister	**Tongue Twister**	**Speech Transcription**	**Total Errors**

Complex
ABBA(S-Coda)
ABAB(T)
	访烦反房
	fang3 fan2 fan3 fang2
		1. fang3 fan2 fan2 fang3 (2 Tone Errors)
		2. fang3 fan2 fan2 fang2 (1 Tone Error)
		3. fan3 fang2 fan2 fang2 (1 Tone Error, 2 Coda Errors)
			Onset Errors: 0
			Vowel Errors: 0
			Coda Errors: 2
			Tone Errors: 4

Complex
ABBA(S-Onset)
ABAB(T)
	湖父福户
	hu2 fu4 fu2 hu4
		1. Correct
		2. fu2 fu4 //* fu2 hu4 (1 Onset Error, 1 Hesitation Error)
		3. fu2 fu4 fu2 hu4 (1 Onset Error)
			Onset Errors: 2
			Vowel Errors: 0
			Coda Errors: 2
			Tone Errors: 0
			Hesitation Errors: 1

Complex
ABBA(S-Coda)
ABAB(T)
	真战沾振
	zhen1 zhan4 zhan1 zhen4
		1. Correct
		2-3. zhen1 zhan4 zhan4 zhen4 (2 Reading Errors)
			Onset Errors: 0
			Vowel Errors: 0
			Coda Errors: 0
			Tone Errors: 0
			Reading Errors: 2

Simple ABBA(T)	湖户户湖 hu2 hu4 hu4 hu2	1. hu2 hu2 hu2 hu4 (3 Tone Errors) 2. Correct 3. hu4 // hu2 // hu4 hu2 (2 Tone Errors, 2 Hesitation Errors)	Onset Errors: 0 Vowel Errors: 0 Coda Errors: 0 Tone Errors: 5 *Hesitation Errors: 2*

Table 12: Example of Tongue Twister Speech Transcription & Error Coding. **Represents a slight break/hesitation that is noticeable in the flow of the reading.*

3.4.5: Statistical Analyses

The dependent variables of interest in this study were the number of errors in each phoneme category: onset, vowel, coda and tone. These are analyzed within each participant, and added up to arrive at a total score for each of the 4 categories. Due to the nature of the tongue twisters used in the study, there are 3 separate tone error rates (from the 3 types of segmental alternation). I report all 3 tone error rates, and average it for an overall tone error rate.

In addition, I then averaged error rates between individuals for each of the 4 dependent variables: onset error, vowel error, coda error, and tone error. For the distribution of errors in the different positions in the ABBA formats, both in simple and complex patterns, please refer to Appendix E. When appropriate, I will use t-tests to verify statistically significant differences or lack thereof between two groups.

3.5: Results of Current Production Study

3.5.1: Overall Errors

Error type	Tone	Tone (Avg)	Onset	Vowel	Coda	Hesitation	Omission	Reading
Total	874	291.33	334.00	135.00	614.00	171.00	1681.00	205.00
Mean number of errors per person	24.97	8.32	9.54	3.86	17.54	4.89	48.03	5.86
SD	16.8	5.60	6.70	4.26	12.31	6.72	38.61	8.88

Table 13: Number of errors, mean number of errors per person, and standard deviations broken down by error type. *'Tone (Avg)' is calculated by averaging the tone errors found in each segment-tone alternating pair (onset, vowel and coda).*

As the study was designed with alternating tone and each of the three segmental categories (onset, vowel and coda), there were three (3) times more possibilities for tone error as compared to each of the three segmental categories. That's why, when looking at the overall rates for each of the 4 phonemic categories, I decided to average the tone errors from each of the 3 segmental category conditions (leading to a tone error value that is not a whole number). For this section, both the total tone errors ("Tone" in the above table, Table 12), as well as the averaged tone errors from the 3 different conditions ("Tone(Avg)" in the above table), are reported. Out of all of the phonetic errors made (1957), 874 (45%) were in tone, 291.3 (21.2%) in tone after being averaged across the three conditions, 334 (24.3%) were in onset, 135 (9.8%) were in vowel, 614 (44.7%) were in coda (as seen in Table 12).

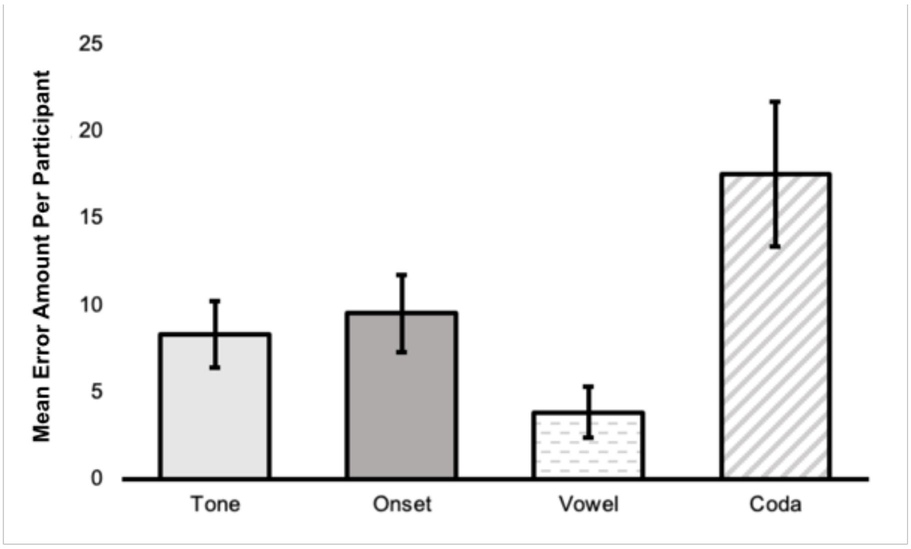


Figure 14: Mean Error Amount per Participant for Average Tone Error vs Onset, Vowel and Tone Error. *The left bar of each pair displays the tone error. Error bars represent ±2 SE of the mean.*

Figure 9 shows the visual representation of the information from Table 12. Tone error presented here is the average of the tone errors from all 3 alternation types. At a glance, you can see that coda error is quite high both in error rate and in standard error of the mean (SEM), whereas vowel error is similarly low in both error rate and in SEM. Tone and onset error are relatively similar in terms of error rate and SEM. Due to this similarity, I performed a paired t-test on tone and onset error, and did not find a statistically significant difference between the two groups, *t*(34)=-.904, *p*=.37. Comparing tone error with the lowest rate category of error, vowel error, a paired t-test revealed a statistically significant difference between the two groups, *t*(34)=4.43, *p*<.001. This means that tone error is not the category with the lowest rates of error, which leads us to believe that tone is actively encoded.

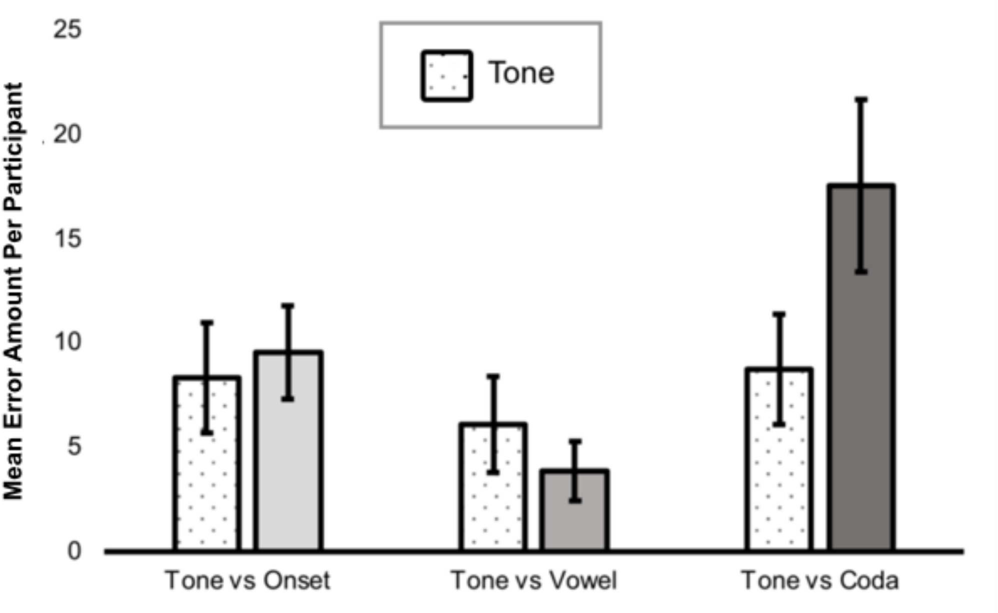


Figure 15: Mean Error Amount per Participant for Tone vs Onset Alternation, Tone vs Vowel Alternation, and Tone vs Coda Alternation. *Error bars represent ±2 SE of the mean.*

In Figure 10, you can see the tone error rates when compared to each segmental alternation: tone versus onset alternation, tone versus vowel alternation, and tone versus coda alternation. Again, you can see that coda error is quite high both in error rate and in standard error of the mean (SEM), whereas vowel error is similarly low in both error rate and in SEM. Tone and onset error are relatively similar in terms of error rate and SEM. To see if the results of individual error rates would be similar to the error rates when I averaged tone error across all 3 alternations, I performed a paired t-test on tone and onset error within that alternation type, and did not find statistical significance between the two groups, *t*(34)=.345, *p*=.72. Comparing tone error with the lowest rate category of error, vowel error, a paired t-test revealed a statistical significance between the two groups, *t*(34)=2.06, *p*=.047. Although this p-value is not as low as the p-value found when tone was averaged, the significance within tone and vowel alternation shows us that vowel is indeed the phoneme category with the lowest error rates, despite how you analyze the data. With vowel being the phoneme category with the lowest error rates, and not tone, we reaffirm the hypothesis that tone is actively encoded.

Within the non-phonological speech error category (hesitation, reading and omission errors), there was an overwhelming amount of reading errors, but depending on the way it gets coded, the error rate could have ended up being a lot lower. I decided to go with a more stringent way of counting the reading errors, for the sake of consistency of coding between the phonological errors, which were based on errors on a single syllable. The significant amount of reading errors shows that, in addition to the errors found in the phonological section, there were many speech errors that I could not attribute to either tone or a specific segmental category.

3.5.2: Results from Replication of Kember et al. (2015) Study

In order to see if the results from my tongue twister study would be reflective of the the Kember et al. (2015) study it was inspired by, I included the tongue twisters from the study into the experiment to see if the rates of tone and onset error would be comparable from the error rates found there. The reason this is important is because in the case that my results are significantly different from those of Kember et al.'s, there are a few possibilities that could explain the differences. One, there was something different with my stimuli. Two, there was something different with my methodology. Three, Kember et al.'s results do not replicate. In doing this replication, I am able to rule out possible confounding factors of my study.

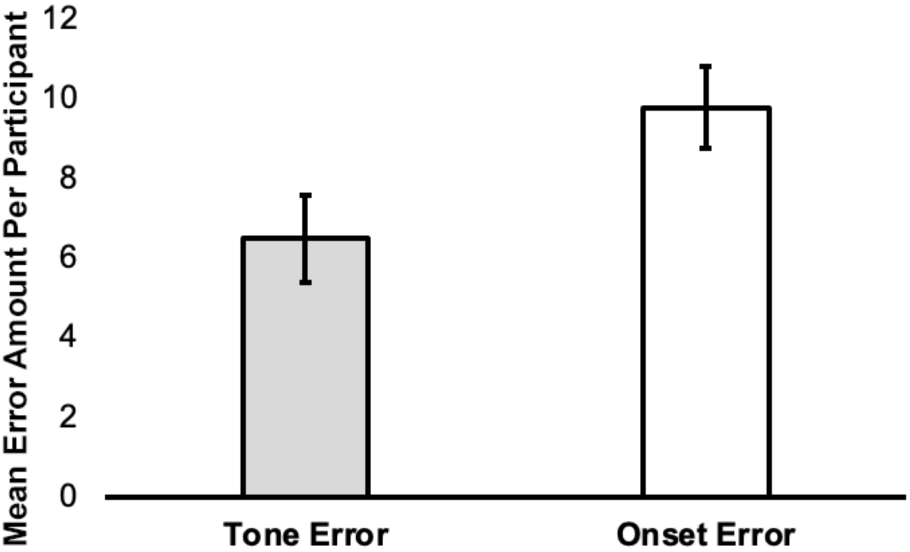


Figure 16: Tone and Onset Errors with Kember et al. (2015) Stimuli. *The figure above shows the average tone and onset error amount per participant. The error bars stand for standard error.*

As shown in Table 2, I included the tongue twisters used in the Kember et al. (2015) study, but excluded the data from those tongue twisters in my results section.

Overall, there were 227 (40%) tone errors and 342 (60%) onset errors among my 35 participants, for a total of 569 tone and onset errors. The difference between the distributions of tone and onset errors were found to be statistically significant, with *t*(34)=-2.65, *p*=.012.

Compare this with the errors found in the original Kember et al. (2015) results: 1372 tone errors (28.1%) and 3503 segment errors (72.8%). Since the specific data points for each participant were not provided, it is impossible to know the t-score and p-value.

Since my study got similar results as Kember's when using the same stimuli, this shows that the different patterns that showed up in the vowel and coda conditions in my study are likely due to differences in how vowel and coda encoding work, rather than due to methodological differences between the current study and Kember's.

3.5.3. Regarding Tone Sandhi

As mentioned in Chapter **3.4.2:** Materials, I took great efforts to avoid using tongue twisters that could elicit Tone 3 Sandhi, a systematic tongue change in Mandarin in which the first of two consecutive Tone 3 characters would turn into Tone 2. However, despite my efforts, my tongue twister stimuli still included 2 tongue twisters (out of 120) that had 2 or more consecutive Tone 3 characters, which could elicit the Tone 3 Sandhi.

Although tone sandhi encoding in tongue twisters is not the main research question I'm looking at, the issue did come up during the transcription of the speech errors, so I decided to include it in the results section.

The original Kember et al. (2015) stimuli, which I also included in my study, contained 5 tongue twisters that included two or more consecutive Tone 3 characters (out of 30). I examined each of these 7 tongue twisters in each participant's recordings to see if some/any instances of these Tone 3 Sandhi positions led to actual tone sandhi utterances.

I combine all instances of tone sandhi context from my tongue twisters and those from Kember et al.'s to form the following list of possible tone sandhi utterances, and compared each participants' actual utterances or lack of utterances in tone sandhi in each tongue twister.

For each 2-character Tone 3 Sandhi context, I counted a tone sandhi utterance if the first item of each 2-character pair is to be produced with a Tone 2. For each 4-character Tone 3 Sandhi context, I counted a tone sandhi utterance if any of the first 3 characters were produced with a Tone 2.

Experiment	Tone Sandhi Tongue Twister	Possible Tone Sandhi Positions
Kember et al. (2015)	突土土突	3
	突苦土哭	3
	始死死始	9
	板满满板	9
	考搞搞考	9
The present study	踢塔体他	3
	访反访反	9

Table 14: Tone Sandhi Context Tongue Twisters from Current Experiment & Kember et al. (2015). *The items that have 3 possible TS positions have 2 Tone 3 items in the center of the tongue twister, and the items that have 9 possible TS positions have Tone 3 items for all 4 characters in the tongue twister.*

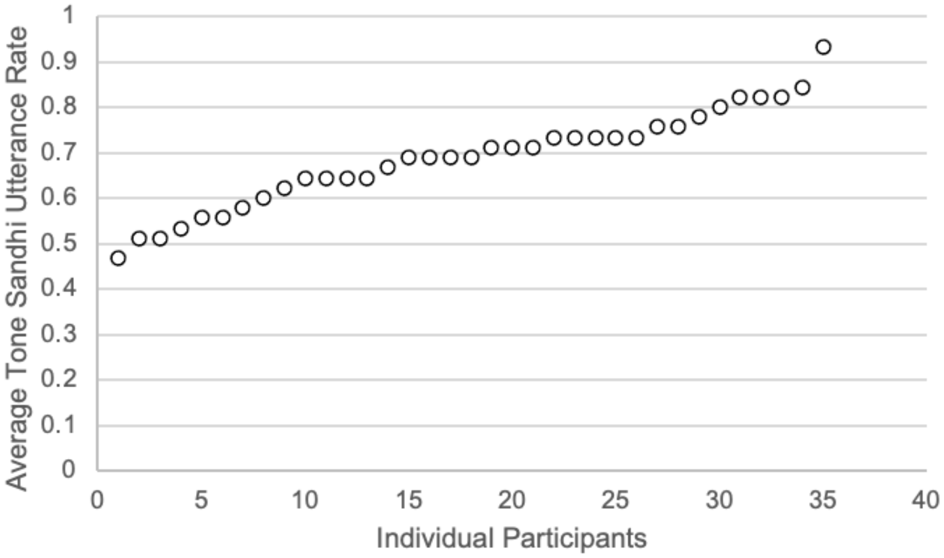


Figure 17: Individual Rates of Uttering Tone Sandhi in Tone Sandhi-Context Tongue Twisters. *Note: Only one participant uttered tone sandhi in less than half of tone sandhi-contexts.*

In my study, I included the Kember et al. (2015) tongue twister stimuli into the tone sandhi analysis since my experiment only had two tongue twisters that include tone sandhi contexts. With the 7 tongue twisters (see Table 13) that have tone sandhi contexts, there were 45 total tone sandhi possibilities, multiplied by 35 participants, for a total of 1575 possible positions of tone sandhi utterances. Out of 1575 possible tone sandhi positions, 1128 tone sandhi utterances were produced. The breakdown of the average tone sandhi rates per tongue twister can be seen in Table 14.

Experiment	Tone Sandhi Tongue Twister	Possible Tone Sandhi Positions	Average T.S. Utterance Rate
Kember et al. (2015)	突土土突	3	0.76
	突苦土哭	3	0.68
	始死死始	9	0.70
	板满满板	9	0.74
	考搞搞考	9	0.73
Chen (2023)	踢塔体他	3	0.45
	访反访反	9	0.64

Table 15: Tone Sandhi Context Tongue Twisters and Average T.S Production Per Tongue Twister. *Note that the Kember et al. (2015) tone sandhi context tongue twisters elicited more tone sandhi utterances than the Chen (2023) tone sandhi context tongue twisters.*

Based on the average tone sandhi utterance rate breakdown, you can see that the Kember et al. (2015) tongue twisters elicited higher rates of tone sandhi utterance compared to the tongue twisters in the current study. Although the reason behind these disparate tone sandhi utterance rates cannot be deduced from this evidence alone, the one pattern that arises is that the tongue twisters in the Kember et al. (2015) study are all tone and onset alternations, whereas the two tongue twisters in the current study that have tone sandhi context are vowel and tone alternation and coda and tone alternation, respectively.

Regardless of the reason behind the varying rates of tone sandhi production in tone sandhi context tongue twisters, what's clear is that the results are drastically different from those in the Kember et al. (2015) study. In the Kember et al. study, in all of the tongue twisters that had a tone sandhi context (5 tongue twisters in 66 possible tone sandhi positions, with 52 participants, for a total of 3,224 possible tone sandhi utterances), they stated that only 5 instances of tone sandhi were produced. I'm not sure what transcription or rating system they used to quantify an utterance as a tone sandhi production or not, but the results from this study are diametrically opposed to the results from their study. Due to this very distinct conflict of results, more experimentation should be done to verify whether tongue twisters do or do not provide the right circumstances for Tone 3 Sandhi production.

Chapter 3.6: Summary of Results

From the results in Chapter 3.5, I see that overall, tone error rates are not the lowest error rates among all 4 categories of phonemes I tested in this study: tone error rate is approximately the same (and slightly lower) than onset error rate, and much higher than vowel error rate, with coda error rate as the highest. Based on my hypothesis, these results suggest that tone is actively encoded in speech planning.

Chapter 3.7: Discussion

The present study used a classic tongue twister paradigm to examine the rate and location of errors within 4-character tongue twisters that alternated in either tone or segment. Tones included all 4 lexical tones in Mandarin Chinese, and segments included onsets, vowels and codas, in segment alternations that are commonly seen in Mandarin tongue twisters. With 90 novel tongue twisters as the basis of the study, participants read each tongue twister 3 times, with a total of 2160 chances for error.

The results showed that codas showed the highest error rates, with onsets and tones trailing with substantially lower error rates, and with vowels showing the lowest error rates. In contrast to the word onset effect that has been witnessed in English and Dutch (Wilshire, 1998), nasal codas in Mandarin shows the highest amount of speech errors.

In the replication of the stimuli from the original Kember et al. (2015) study, on which this study was inspired, the results showed slightly different, but relatively similar, patterns as compared to the original study, with tone errors making about 40% of the phonological errors, and with onsets making up about 60% of the phonological errors. Recall that in the Kember et al. study, tone errors comprised 28.1% of the phonological errors, while onset errors made up the remaining 72.8% of errors.

Whether or not this minor divergence from the original Kember et al. study is simply an artifact of repeated testing (in which results can vary slightly from experiment to experiment) or something else, such as participant background and condition, experimental procedure, or speech transcription and coding, etc, remains to be seen. However, with 35 participants in the present study from native Mandarin speakers residing in China at the time of the speech recordings, participant selection seems to not be the problem. The experimental procedure of the Kember et al. stimuli was the same as that as the present study, barring that the present study was conducted online and the Kember et al. (2015) study was done in-person. Speech transcription and error coding also followed the original procedure quite closely, so that should not be an issue.

Previous tone speech error in Mandarin, Cantonese, and Thai show that tone error tends to be of the perseveratory nature (Gandour, 1977; Chen, 1999; Alderete et al., 2019), whereas segment error tends to be anticipatory. The current study does not analyze whether tone errors and segment errors follow this same trend, but because I have the data, further data analysis to clarify the perseveratory nature of tone research can be done in the future.

Regarding speech errors, previous research has shown that syllable structure tends to have an impact on the position of speech errors. For example, branching onsets tend to be more susceptible to errors (Fudge, 1987). Nucleus and coda, which make up the rime, tend to undergo speech errors together, as opposed to an onset and a nucleus, which do not make up a unit within a syllable (Page et al., 2007). However, because the tongue twisters used for the current study alternated segments with tones, there were no particular opportunities presented for a nucleus and coda to undergo a speech error together, and the results reflected this: there were no instances of a nucleus and coda erring together in the same utterance. Regarding speech structure, none of the stimuli used in the experiment had branching onsets, so this would not be relevant to the current study.

Additionally, reviewers have brought up other potential sources of speech errors, such as temporal processing abilities and the varying information load of different components. As of this writing, I have not found any particular links between to these ideas and the current data, but they may certainly be topics for future research.

Based on my results, I have sufficient evidence to assert that tones are indeed actively encoded in speech encoding.

CHAPTER 4: CONCLUSION

The present study examined the phonological processing and production of lexical tone in Mandarin Chinese. Lexical tone, the use of varying pitch to accompany the sonorant portion of a syllable (which occurs in a majority of the world's languages), is a lexically contrastive phoneme similar to how consonants and vowels distinguish between words in Indo-European languages. However, due to the fact that tonal systems across languages greatly vary (some tonal languages have as few as 2 tones and some have up to 6 or even more), in addition to the fact that they cannot occur independently of segments, there is a large amount of literature arguing that tones do not behave the same way that segments do. This study used behavioral methods to test two aspects of tone in spoken language - phonological processing and production - in an attempt to answer the question: are tones processed and produced in a qualitatively different way from segments?

Previously, tones and segments have been seen as two opposing dichotomies, with tones occupying a solo space, and either onsets alone or consonants and vowels representing segments. However, I believe that segments cannot be viewed as one aggregate form; rather, I believe that segments should be viewed from the perspective of its 3 discrete temporal components: onsets, vowels and codas. In my project, lexical tone is viewed as an extension of segments, and occupies the 4th phonemic category. I took this approach in the execution of my research project, and at the end, reflected upon whether or not this approach is appropriate based on the results.

The first part of the study looked at the status of tones in phonological processing. Phonological processing in spoken language recognition is commonly seen as activating multiple lexical candidates based on phonetic similarity as the speech signal unfolds. In a word or a series of words, I'm interested in seeing if tones and segments contribute to the activation of lexical candidates to different extents.

I designed a series of 4 experiments that uses a forced-choice word selection task to examine native Mandarin speakers' acceptance of variation in tone in comparison to the three types of segments I previously mentioned: onsets, vowels, and codas. Previous research has shown that in a word reconstruction task, when native speakers are presented with a nonword and asked to switch either a consonant, vowel, or tone to turn the nonword into a word, tone was chosen as the most readily changed phoneme type, over consonants and vowels. Because of this high acceptance of tone variation, researchers were encouraged to claim that tones are the least lexically binding phoneme type in Mandarin, and that they innately behave differently than segments in phonological processing.

However, the paradigm used for the experiment, the word reconstruction task, leaves room for alternative explanations, so the present study uses a forced-choice word selection task to verify the results of the previous study. The results support the previous research that claims that tones are indeed less lexically binding than segmental onsets and vowels. However, when faced off with segmental codas (which has an inventory of 2 items in Mandarin), tones showed more resistance to change, which lends support to the idea that tones do not behave intrinsically differently from segments as a whole.

Rather, the willingness of native speakers to accept tone variation is just a result of its phonological inventory size: participants are less sensitive to deviations in phonemes that contribute less information to word recognition. Mandarin tone has an inventory size of 4 items. The segmental coda, which only has an inventory of 2 items in Mandarin, has an even smaller inventory, so thus contributes to word recognition even less than tone does.

The second part of the study looked at the status of tones in speech production, using a tongue twister paradigm to examine the role of tone in phonological encoding, the planning and articulation of speech. Researchers have primarily used implicit priming tasks, word form preparation tasks, and speech errors to shed light on this complex process. Speech errors, despite the heavy amount of work that goes into transcription and coding, have often been examined as a way to shed light on the roles of different phonemes in speech production.

Previously, there has been conflicting research supporting both frequent tone error and infrequent tone error in natural speech. Frequent tone error lends support to the theory that tone is actively encoded in speech, encoded similarly to segments. Infrequent tone error supports the theory that tone is inactively processed, or processed later, as compared to segments. Additionally, there are some claims that tone errors are not modulated by context, which further lends to the idea that tones are not actively processed.

My experiment uses a tongue twister paradigm designed to elicit equal amounts of error in 4 phoneme categories: tone, onset, vowel, and coda. If tone error rates end up being significantly lower than that of the 3 segmental categories, then we would conclude that tones are not actively encoded. However, if tone error rates were similar to those of the 3 segmental categories, I would assume that there is insufficient evidence to conclude that tones are encoded in a way unsimilar to segments. The results of my study showed that codas showed the most errors, followed by onsets and tones, with vowels showing the least error rates. I believe that this shows insufficient evidence for the theory that tones are encoded differently from segments as a whole. Additionally, I found that although tone errors did not behave in the way that many speech models assume segments behave, tone errors were indeed modulated by context, which provides additional evidence showing that tones are actively encoded during Mandarin speech production. However, this warrants further research, as it does seem that tone errors arise from a different phonological origin as compared to segments.

Overall, the results from my two studies provide strong evidence that tones play a similar role compared to segments within the Mandarin Chinese language. Despite the abundance of research that argues that tones, being suprasegmental in nature, do not behave similarly to segments, which are commonly seen as the “core” of oral language, I believe that tones occupy a role similar to that of a phoneme in Mandarin Chinese. Like other phonemes in the language that are modulated by inventory size, temporal location, and articulation constraints, tone provides lexical value and information proportional to its intrinsic phonological features and properties. To argue that tones, as a whole, behave differently from all segments would be to overlook the fine-grain differences between segments as well. My work hopes to elucidate the subtle differences and similarities between how tones and different categories of segments behave in the language.

REFERENCES

About Pleco. (n.d.). Retrieved from http://www.pleco.com/about/

Alderete, J. (2023). Cross-linguistic trends in speech errors: An analysis of sub-lexical errors in Cantonese. *Language and Speech*, *66*(1), 79-104.

Alderete, L., Chan, Q., & Yeung, H. H. (2019). Tone slips in Cantonese: Evidence for early phonological encoding. *Cognition*, *191*, 1-16.

Allopenna, P. D., Magnuson, J. S., & Tanenhaus, M. K. (1998). Tracking the time course of spoken word recognition using eye movements: Evidence for continuous mapping models. *Journal of Memory and Language*, *38*(4), 419–439.

Ao, B. (1992). The non-uniqueness condition and the segmentation of the Chinese syllable. *Working Papers in Linguistics*, *42*, 1-25. Ohio State University.

Bao, Z., Shi, J., & Xu, D. (1997). *Shengcheng yinxixue lilun jiqi yingyong* [Generative phonology: theory and usage]. Beijing: Zhongguo Shehui Kexue Chubanshe.

Berg, T. (1985). Is voice suprasegmental? *Linguistics, 23,* 883-915.

Boersma, P. & Weenink, D. (2023). Praat: doing phonetics by computer [Computer program]. Version 6.3.08, retrieved 10 February 2023 from http://www.praat.org/.

Chao, Y.-R. (1934). The non-uniqueness of phonemic solutions of phonetic systems. *Bulletin of the Institute of History and Philology, Academia Sinica, 4*(4), 367-97.

Chao, Y.-R. (1968). *A Grammar of Spoken Chinese*. Berkeley and Los Angeles: University of California Press.

Chen, J.-Y. (1999). The representation and processing of tone in Mandarin Chinese: Evidence from slips of the tongue. *Applied Psycholinguistics*, *20*, 289-301.

Chen, J.-Y. (2000). Syllable errors from naturalistic slips of the tongue in Chinese. *Psychologia*, *43*, 15-26.

Chen, J.-Y., Chen, T.-M., & Dell, G. (2002). Word-form encoding in Mandarin Chinese as assessed by the implicit priming task. *Journal of Memory and Language*, *46*(2), 751-781.

Chen, J.-Y., Lin, W.-C., & Ferrand, L. (2003). Masked priming of the syllable in Mandarin Chinese speech production. *Chinese Journal of Psychology, 45*(1), 107–120.

Cheng, C.-C. (1973). *A Synchronic Phonology of Mandarin Chinese* (Vol. 4). The Hague: Mouton.

Cheng, R. L. (1966). Mandarin phonological structure. *Journal of Linguistics*, *2*(2), 135-158.

Commercial Press. (2020). *Xinhua zidian* [Xinhua Dictionary] (12th ed.).

Costa, A., Cutler, A., & Sebastian-Galles, N. (1998). Effects of phonemic repertoire on phoneme decision. *Perception & Psychophysics*, *60*, 1022-1031.

Croot, K., Au. C., & Harper, A. (2010). Prosodic structure and tongue twister errors. In C. Fougerton, B. Kuehnert, & M. d'Imperio (Eds.), *Laboratory Phonology 10: Phonology and Phonetics*, 433-459. Germany: De Gruyter Mouton.

Crompton, A. (1982). Syllables and segments in speech production. In: A. Cutler (ed.), *Slips of the tongue and language production*. The Hague: Mouton.

Cutler, A. (1980). Errors of stress and intonation. In V. Fromkin (ed.). *Errors in Linguistic Performance: Slips of Tongue, Ear, Pen, and Hand* (pp. 67-80). New York: Academic Press.

Cutler, A. & Chen, H.C. (1997). Lexical tone in Cantonese spoken-word processing. *Perception & Psychophysics*, *59*(2), 165-79.

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

Dell, G. S. (1984). Representation of serial order of speech: Evidence from the repeated phoneme effect in speech errors. *Journal of Experimental Psychology: Learning, Memory and Cognition*, *10,* 222-233.

Dell, G. S. (1986). A spreading activation theory of retrieval in sentence production. *Psychological Review*, *93*, 283-321.

Dell, G. S. (1988). The retrieval of phonological forms in production: Tests of predictions from a connectionist model. *Journal of Memory and Language*, *27*(2), 124-142.

Deng, X., Farris-Trimble, A., & Yeung, H. H. (2022). Contextual effects on spoken word processing: An eye-tracking study of the time course of tone and vowel activation in Mandarin. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *49*(7), 1145-1160.

Duanmu, S. 1990. *A formal study of syllable, tone, stress and domain in Chinese languages* [Doctoral thesis, Massachusetts Institute of Technology]. MIT Working Papers in Linguistics. http://mitwpl.mit.edu/catalog/duan01

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

Duanmu, S. (2011). Chinese syllable structure. *The Blackwell Companion to Phonology* (eds M. Oostendorp, C.J. Ewen, E. Hume and K. Rice). John Wiley & Sons.

Fónagy, I., & Magdics, K. (1960). Speed of utterance in phrases of different lengths. *Language and Speech*, *3*, 179-192.

Forster, K. I., Forster, J. C. (2003). DMDX: A Windows display program with millisecond accuracy. *Behavior Research Methods, Instruments, & Computers, 35*, 116–124.

Fox, R. A., & Unkefer, J. (1985). The effect of lexical status on the perception of tone. *Journal of Chinese Linguistics, 13,* 69-90.

Fromkin, V. A. (1971). The non-anomalous nature of anomalous utterances. *Language, 47*, 27-52.

Fromkin, V. A. (1973). Introduction. In VA. Fromkin (Ed.), *Speech errors as linguistic evidence* (pp. 11-45). The Hague: Mouton.

Fromkin, V. A. (1976). Putting the emPHAsis on the wrong sylLABle. In L. Hyman (Ed.), *Studies in Stress and Accent*. Los Angeles: Univ. of Southern California, 15-26.

Fudge, E. (1987). Branching structure within the syllable. *Journal of Linguistics*, *23*(2), 359-377.

Gandour, J. (1977). Counterfeit tones in the speech of Southern Thai bidialectals. *Lingua*, *41*, 125-143.

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

Garlock, V. M., Walley, A. C., & Metsala, J.L. (2001). Age of acquisition, word frequency, and neighborhood density in spoken word recognition by children and adults. *Journal of Memory and Language*, *45*(3), 468-492.

Goldrick, M. (2006). Limited interaction in speech production: Chronometric, speech error, and neuropsychological evidence. *Language and Cognitive Processes*, *21*, 817-855.

Goldrick, M., & Blumstein, S. (2006). Cascading activation from phonological planning to articulatory processes: Evidence from tongue twisters. *Language and Cognitive Processes*, *21*, 649-683.

Goldstein, L., Pouplier, M., Chen, L., Saltzman, E., & Byrd, D. (2007). Dynamic action units slip in speech production. *Cognition*, *103*, 386-412.

Hartman, L. M. (1944). The segmental phonemes in Peiping dialect. *Language*, *20*, 28-42.

He, K. & Li, D. (1987). *Xiandai Hanyu sanqian changyong cibiao* [Modern Chinese Character 3000 Most Commonly Used Words]. Beijing: Beijing Normal University Press.

Hsueh, F. (1986). *An Anatomy of the Pekingese Sound System*. Taipei: Taiwan Xuesheng Publishing Co.

Jyutdin. [粵典|粵典 words.hk.] (2023). Retrieved January 16, 2024, from https://words.hk/.

Kember, H., Croot, K., & Patrick, E. (2015). Phonological encoding in Mandarin Chinese: Evidence from tongue twisters. *Language and Speech, 58*, 417-440.

Levelt, W. J. M., Roelofs, A., & Meyer, A. S. (1999). A theory of lexical access in speech production. *Behavioral and Brain Sciences*, *22*, 1-75.

Li, W. (1999). *A diachronically-motivated segmental phonology of Mandarin Chinese*. Peter Lang Publishing Inc.

Lin, T. & Wang, L. (2013). *Introduction to phonetics* (Revised Edition). Peking: Peking University Press.

Lin, Y. H. (1989). Autosegmental treatment of segmental processes in Chinese phonology. [Unpublished doctoral dissertation]. University of Texas at Austin.

Lin, Y. H. (2007). *The Sounds of Chinese*. Cambridge: University Press.

Liu, J. H. C., & Wang, H. S. (2008). Speech errors of tone in Taiwanese. *North American Conference on Chinese Linguistics, 20*, 189-203.

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

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.

Malins, J. G., & Joanisse, M. F. (2012). Setting the tone: An ERP investigation of the influences of phonological similarity on spoken word recognition in Mandarin Chinese. *Neuropsychologia, 50*(8), 2032-2043.

Marlsen-Wilson, W.D., & Welsh, A. (1978). Processing interactions and lexical access during word recognition in continuous speech. *Cognitive Psychology, 10*(1), 29-63.

MacKay, D. G. (1970). Spoonerisms: The structure of errors in the serial order of speech. *Neuropsychologia, 8*, 323-350.

Mathesius, V. (1929). *La structure phonologique du lexique du technique modern* [The phonological structure of the lexicon of modern technology]. *Travaux du Cercle linguistique de Prague, 1*, 67–84.

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

McMillan, C. T., & Corley, M. (2010). Cascading influences on the production of speech: Evidence from articulation. *Cognition, 117*, 243-260.

Meyer, A. (1992). Investigation of phonological encoding through speech error analyses: achievements, limitations, and alternatives. *Cognition*, *42*, 181–211.

Meyer, D. E., & Schvaneveldt, R. W. (1971). Facilitation in recognizing pairs of words: Evidence of a dependence between retrieval operations. *Journal of Experimental Psychology, 90*(2), 227–234.

Ministry of Education. (2008). *Xiandai hanyuzhong youmeiyou yuanyin, fuyin?* [Does modern Chinese have vowels and consonants?] *Ministry of Education of the People's Republic of China.* Retrieved from http://www.moe.gov.cn/jyb_hygq/hygq_zczx/moe_1346/moe_1364/tnull_39887.html

Modern Chinese Word Frequency Dictionary (1986). Beijing, China: Language Institute Press.

Moser, D. (1991). Slips of the tongue and pen in Chinese. *Sino-Platonic Papers, 22*, 1-45.

Neergaard, K. D., Xu, H., German, J. S., & Huang, C.-R. (2022). Database of word-level statistics for Mandarin Chinese (DoWLS-MAN). *Behavior Research Methods*, *54*, 897-1009.

Nooteboom, S.G. (1969). The tongue slips into patterns. A.G. Sciarone et al. (Eds.), *Nomen: Leyden Studies in Linguistics and Phonetics* (pp.114-132). The Hague: Mouton.

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

Page, M. P. A., Madge, A., Cumming, N., & Norris, D. G. (2007). Speech errors and the phonological similarity effect in short-term memory: Evidence suggesting a common locus. *Journal of Memory and Language, 56*(1), 49–64.

Peirce, J. W., Gray, J. R., Simpson, S., MacAskill, M. R., Höchenberger, R., Sogo, H., Kastman, E., Lindeløv, J. (2019). PsychoPy2: Experiments in behavior made easy. *Behavior Research Methods*, *51*, 195-203.

Pouplier, M., & Goldstein, L. (2010). Intention in articulation: Articulatory timing in alternating consonant sequences and its implications for models of speech production. *Language and Cognitive Processes, 25*, 616-649.

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

Roelofs, A., & Meyer, A. S. (1998). Metrical structure in planning the production of spoken words. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 24*(4), 922–939.

Schiller, N. O. (2006). Phonological encoding in speech production. In A. Botinis (Ed.), *Proceedings of ISCA Tutorial and Research Workshop on Experimental Linguistics* (pp. 53-60). University of Athens.

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.

Shattuck-Hufnagel, S. (1983). Sublexical units and suprasegmental structure in speech production planning. In: MacNeilage, P.F. (eds), *The Production of Speech*. Springer, New York, NY.

Shattuck-Hufnagel, S. (1986). The representation of phonological information during speech production planning: Evidence from vowel errors in spontaneous speech. *Phonology Yearbook, 3,* 117-149.

Shattuck-Hufnagel, S. (1992). The role of word structure in segmental serial ordering. *Cognition*, *42*, 213-259.

Shattuck-Hufnagel, S., & Klatt, D. H. (1979). The limited use of distinctive features and markedness in speech production: Evidence from speech error data. *Journal of Verbal Learning and Verbal Behaviour*, *18*, 41-55.

Shi, F. (2002). The vowel system of Beijinghua. *Nankai Yuyan Xuekan*, *1*, 30-36.

Sereno, J. A., & Lee, H. (2015). The contribution of segmental and tonal information in Mandarin spoken word processing. *Language and Speech, 58(2)*, 131–151.

Surendran, D., & Levow, G.-A. (2004): The functional load of tone in Mandarin is as high as that of vowels. *Proceedings of the International Conference on Speech Prosody 2004*, 99-102.

Wan, I.P. (1996). *Tone errors in Mandarin Chinese*. [Unpublished Master's thesis]. State University of New York, Buffalo.

Wan, I.P. (2006). Tone errors in normal and aphasic speech in Mandarin. *Taiwan Journal of Linguistics*, *4*, 85-112.

Wan, I.-P., & Jaeger, J. J. (1998). Speech errors and the representation of tone in Mandarin Chinese. *Phonology*, *15*, 417-461.

Wang, J. Z. (1993). *The Geometry of Segmental Features in Beijing Mandarin*. [Unpublished doctoral dissertation]. University of Delaware, Newark.

Wiener, S., & Turnbull, R. (2015). Constraints of tones, vowels and consonants on lexical selection in Mandarin Chinese. *Language and Speech*, *59*(1), 59-82.

Wilshire, C. (1998). Serial order in phonological encoding: An exploration of the ‘word onset effect’ using laboratory-induced errors. *Cognition*, *68*, 143-166.

Wilshire, C. (1999). The “tongue twister” paradigm as a technique for studying phonological encoding. *Language and Speech*, *42*, 57-82.

Wong, A. W.-K., & Chen, H.-C. (2008). Processing segmental and prosodic information in Cantonese word production. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, *34*, 1172-1190.

Wong, A. W.-K., & Chen, H.-C. (2009). What are effective phonological units in Cantonese spoken word planning? *Psychonomic Bulletin and Review*, *16*, 888– 892.

Van Ooijen, B. (1996). Vowel mutability and lexical selection in English: Evidence from a word reconstruction task. *Memory & Cognition*, *24*(5), 573-583.

Vitevitch, M.S., & Luce, P.A. (1998). When words compete: Levels of processing in perception of spoken words. *Psychological Science*, *9*(4), 325-328.

Xu, S. (1980). *Putonghua yuyin zhishi* [Phonology of Standard Chinese]. Beijing: Wenzi Gaige Chubanshe.

Yang, H. & Oh, M. (2020). Loanword adaptation of English coronal fricatives into Mandarin Chinese. *Linguistics Research*, *37*(1), 71-93.

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

Yi, T. L. (1920). *Lectures on Chinese Phonetics.* Shanghai: Commercial Press.

Yip, M. J. (2002). *Tone*. Cambridge University Press.

Yip, P.-C. (2000). *The Chinese Lexicon: A Comprehensive Survey*. Routledge.

Ziegler, J. C., Muneaux, M., & Grainger, J. (2003). Neighborhood effects in auditory word recognition: Phonological competition and orthographic facilitation. *Journal of Memory and Language*, *48*(4), 779- 793.

Zhang, B. (2002). *New Edition of Modern Chinese*. Shanghai: Fudan University Press.

Zhang, J.-X. (张静贤). (1992). *Xiandai Hanzi Jiaocheng* [Modern Chinese Character Teachings]. Beijing: Modern Publishing Co.

Zhao, J., Guo, J., Zhou, F., and Shu, H. (2011). Time course of Chinese monosyllabic spoken word recognition: evidence from ERP analyses. *Neuropsychologia*, *49*, 1761–1770.

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

ZhTenTen – Chinese corpus from the web. Sketch Engine. (2020, September 17). Retrieved April 10, 2023, from https://www.sketchengine.eu/zhtenten-chinese-corpus/.

APPENDIX A: Stimuli list for Experiment 1

Critical Stimuli:

Tone Mismatch	Onset Mismatch	Stimuli	Condition
擦	杂	ca2	critical
组	醋	cu3	critical
头	都	dou2	critical
干	寒	gan2	critical
森	坟	sen2	critical
价	下	xia3	critical
才	灾	zai2	critical
杯	陪	bei2	critical
贺	舍	he3	critical
色	舍	se3	critical
混	损	sun4	critical
他	答	ta2	critical
舔	店	tian4	critical
碟	贴	tie2	critical
允	俊	jun3	critical
锤	崔	cui2	critical
错	妥	cuo3	critical
呆	台	dai2	critical
后	否	fou4	critical
白	该	gai2	critical

给	被	gei4	critical
掊	抠	kou2	critical
同	空	kong2	critical
款	罐	kuan4	critical
滑	刷	shua2	critical
素	虎	su3	critical
太	歹	tai3	critical
特	渴	te3	critical
准	顺	zhun4	critical
词	滋	zi2	critical
平	冰	bing2	critical
车	折	che2	critical
叼	条	diao2	critical
葵	龟	gui2	critical
画	耍	hua3	critical
踏	卡	ka4	critical
图	哭	ku2	critical
蒙	能	neng1	critical
扁	片	pian3	critical
商	航	shang2	critical

Control Stimuli:

Option 1	Option 2	Stimuli	Condition
白	戴	bai2	easy

州	抽	chou1	easy
传	转	chuan2	easy
力	底	di3	easy
丢	流	diu1	easy
仔	改	gai3	easy
滚	顿	gun3	easy
坠	毁	hui3	easy
狂	黄	kuang2	easy
板	满	man3	easy
猫	包	mao1	easy
匹	米	mi3	easy
领	命	ming4	easy
脑	牢	nao3	easy
酒	牛	niu2	easy
排	来	pai2	easy
捞	袍	pao2	easy
陪	枚	pei2	easy
聊	票	piao4	easy
热	色	re4	easy
针	神	shen2	easy
事	日	shi4	easy
妆	双	shuang1	easy
多	脱	tuo1	easy
洗	李	xi3	easy

骂	亚	ya4	easy
摇	捞	yao2	easy
鲁	雨	yu3	easy
帐	棒	zhang4	easy
桌	坐	zuo4	easy
薄	包	bao2	hard
补	步	bu4	hard
藏	舱	cang2	hard
村	存	cun2	hard
赶	杆	gan3	hard
寡	挂	gua4	hard
货	火	huo4	hard
记	挤	ji4	hard
静	景	jing3	hard
橘	居	ju2	hard
控	恐	kong3	hard
扣	口	kou3	hard
剖	掊	pou1	hard
忍	认	ren4	hard
扔	仍	reng1	hard
勺	烧	shao1	hard
少	绍	shao3	hard
缩	锁	suo3	hard
替	体	ti4	hard

捂	无	wu2	hard
五	雾	wu3	hard
翔	香	xiang1	hard
凶	熊	xiong1	hard
艳	演	yan3	hard
音	银	yin1	hard
右	有	you3	hard
元	冤	yuan2	hard
远	院	yuan4	hard
正	整	zheng4	hard
汁	直	zhi2	hard
胆	探	dan4	impossible
订	挺	ding3	impossible
土	度	du3	impossible
逢	灯	feng1	impossible
落	伙	huo4	impossible
教	小	jiao3	impossible
搞	靠	kao3	impossible
棍	捆	kun4	impossible
闹	老	nao3	impossible
七	习	qi2	impossible
绳	挣	sheng1	impossible
汁	石	shi1	impossible
熟	周	shou1	impossible

打	踏	ta3	impossible
对	腿	tui4	impossible
位	美	wei3	impossible
路	五	wu4	impossible
星	零	xing2	impossible
盐	边	yan1	impossible
牛	优	you2	impossible

APPENDIX B: Stimuli list for Experiments 2 & 3

Same as Appendix A but with the additional coda stimuli.

Critical Stimuli:

Coda Mismatch	Tone Mismatch	Stimuli	Condition
喊	航	hang3	critical
咱	脏	zang2	critical
灿	舱	cang4	critical
染	让	ran4	critical
盆	捧	pen3	critical
猛	门	men3	critical
趁	逞	chen3	critical
忍	仍	reng3	critical
爽	涮	shuang4	critical
鬓	饼	bin3	critical
聘	乒	ping4	critical
敏	名	ming3	critical
您	拧	nin3	critical
拎	另	ling1	critical
款	狂	kuan2	critical
藏	残	cang3	critical
扔	认	ren1	critical
涮	爽	shuan3	critical
品	瓶	ping3	critical
命	敏	min4	critical

宽	矿	kuan4	critical
肯	坑	keng3	critical

Control Stimuli (only used in experiment 3):

Option 1	Option 2	Stimuli	Condition
韩	夯	han2	easy
恒	很	hen3	easy
放	翻	fan1	easy
粉	逢	feng2	easy
令	林	ling4	easy
懒	狼	lan3	easy
让	然	ran2	easy
认	扔	ren4	easy
干	港	gan1	easy
梗	跟	geng3	easy
囔	难	nang1	easy
丧	散	san3	easy
宾	病	bing4	easy
赢	音	ying2	easy
紧	静	jin3	easy
琴	请	qin2	easy
板	班	ban1	hard
饼	冰	bing3	hard
棒	绑	bang3	hard

坟	分	fen1	hard
碰	朋	peng4	hard
稳	文	wen2	hard
听	挺	ting3	hard
蹦	笨	beng4	hard
频	瓶	pin2	hard
残	藏	cang2	hard
您	宁	nin2	hard
缓	谎	huan3	hard
间	江	jiang1	hard
像	线	xiang4	hard
门	盟	men2	hard
亮	练	liang4	hard
闽	名	min2	impossible
桑	伞	san1	impossible
胖	攀	pan4	impossible
硬	阴	ying1	impossible
喷	棚	pen2	impossible
按	航	an1	impossible
党	蛋	dan3	impossible
勤	青	qin1	impossible
缝	芬	fen4	impossible
饭	访	fan3	impossible
银	影	yin3	impossible

APPENDIX C: Stimuli list for Experiment 4

Critical Stimuli:

Vowel Mismatch	Tone Mismatch	Stimuli	Condition
鲁	乐	le3	critical
咪	木	mu1	critical
乳	日	ri3	critical
虎	贺	he3	critical
死	色	se3	critical
俗	思	si2	critical
踢	特	te1	critical
则	资	zi2	critical
富	佛	fo4	critical
客	卡	ka4	critical
壳	哭	ku2	critical
度	德	de4	critical

Control Stimuli:

Option 1	Option 2	Stimuli	Condition
入	弱	ru4	easy
摸	马	ma3	easy
敌	大	di2	easy
你	拿	ni3	easy
热	如	re4	easy
无	蛙	wu2	easy

屁	破	pi4	easy
普	趴	pa1	easy
把	部	bu4	easy
腻	纳	na4	medium
坡	劈	po1	medium
佛	罚	fa2	medium
突	图	tu1	medium
西	习	xi2	medium
答	搭	da1	medium
沙	啥	sha1	medium
以	意	yi3	medium
墨	抹	mo3	medium
射	舍	she4	hard
恶	额	e3	hard
物	五	wu3	hard
皮	批	pi1	hard
厨	出	chu2	hard
木	母	mu3	hard
鹿	卤	lu3	hard
踏	塔	ta4	hard
八	拔	ba1	hard
打	肚	da4	impossible
呼	和	hu2	impossible
炸	猪	zha1	impossible

居	集	ju2	impossible
辣	里	la3	impossible
髮	幅	fa3	impossible

APPENDIX D: Neighborhood Density of Experiments 1-4's Critical Stimuli

Onset vs Tone Stimuli

Tone Mismatch Item	Tone Mismatch Phonological Density	Segment Mismatch Item	Segment Mismatch Phonology Density
他	26	下	11
价	13	俊	6
允	7	冰	15
准	11	刷	12
叼	12	卡	33
同	18	台	14
后	16	否	14
呆	21	哭	20
商	23	坟	12
图	19	妥	14
太	22	寒	20
头	16	崔	10
干	12	店	11
平	12	折	14
扁	12	抠	15
才	15	损	9
擦	23	杂	17
杯	7	条	8
森	13	歹	19
款	13	渴	11
混	16	滋	8
滑	9	灾	19
特	15	片	12
画	10	空	16
白	13	答	18

	碟	6	罐	16	
	素	22	耍	9	
	组	21	能	14	
	给	9	舍	13	
	舔	12	航	19	
	色	15	虎	20	
	葵	7	被	10	
	蒙	35	该	22	
	词	5	贴	14	
	贺	16	都	37	
	踏	22	醋	22	
	车	13	陪	9	
	错	19	顺	14	
	锤	6	龟	12	
	掊	11	刷	12	
	混	15	舍	15	
Average	Tone	14.8	Onset	15.1	

Coda vs Tone Stimuli

	Tone Mismatch Item	Tone Mismatch Phonological Density	Segment Mismatch Item	Segment Mismatch Phonology Density	
	乒	14	命	12	
	仍	15	咱	14	
	另	15	品	9	
	名	14	喊	23	
	坑	17	宽	15	
	拧	13	忍	12	
	捧	11	您	9	
	敏	9	扔	15	
	残	18	拎	22	
	涮	14	敏	9	
	爽	8	染	22	
	狂	5	款	13	
	瓶	12	涮	14	
	矿	10	灿	22	
	脏	22	爽	8	
	航	19	猛	11	
	舱	22	盆	11	
	认	15	聘	10	
	让	19	肯	11	
	逞	13	藏	18	
	门	13	趁	15	
	饼	13	鬓	11	
Average	Tone	14.6	Coda	13.9	

Vowel vs Tone Stimuli					
	Tone Mismatch Item	Tone Mismatch Phonological Density	Segment Mismatch Item	Segment Mismatch Phonology Density	
	乐	13	俊	6	
	佛	7	冰	15	
	卡	11	刷	12	
	哭	12	卡	22	
	德	14	台	14	
	思	16	否	14	
	日	21	哭	20	
	木	23	坟	12	
	特	19	妥	14	
	色	22	塞	20	
	贺	13	崔	10	
	资	12	店	11	
	佛	12	折	14	
	卡	12	抠	15	
	思	15	损	9	
Average	Tone	14.4	Vowel	16.1	

Average Phonological Neighborhood Density by Tone and Segment Type				
	Tone (Averaged)	Onset	Vowel	Coda
Neighborhood Density	14.6	15.1	16.1	13.9

APPENDIX E: Chinese tongue twisters (used for inspiration)

Type of Segment	Segment(s)	Tongue Twister
Onset	b/p	补破皮褥子不如不补破皮褥子（《补皮褥子》）
Onset	j/q/x	七巷一个漆匠，西巷一个锡匠，七巷漆匠偷了西巷锡匠的锡，西巷锡匠偷了七巷漆匠的漆（《漆匠和锡匠》）
Onset	h/f	一堆粪，一堆灰，灰混粪，粪混灰（《一堆粪》）。
Onset	z/zh	隔着窗户撕字纸，一次撕下横字纸，一次撕下竖字纸，是字纸撕字纸，不是字纸，不要胡乱撕一地纸（《撕字纸》）。
Onset	g/k	哥挎瓜筐过宽沟，快过宽沟看怪狗。光看怪狗瓜筐扣，瓜滚筐空哥怪狗（《哥挎瓜筐过宽沟》）。
Vowel	u/ü	军车运来一堆裙，一色军用绿色裙。军训女生一大群，换下花裙换绿裙（《换裙子》）。
Vowel	ei/ai	大妹和小妹，一起去收麦。大妹割大麦，小妹割小麦。大妹帮小妹挑小麦，小妹帮大妹挑大麦。大妹小妹收完麦，噼噼啪啪齐打麦（《大妹和小妹》）。
Vowel	ang/eng	长城长，城墙长，长长长城长城墙，城墙长长城长长（《长城长》）。
Coda	en/eng	陈庄程庄都有城，陈庄城通程庄城。陈庄城和程庄城，两庄城墙都有门。陈庄城进程庄人，陈庄人进程庄城。请问陈程两庄城，两庄城门都进人，哪个城进陈庄人，程庄人进哪个城？（《陈庄城和程庄城》）
Coda	an/ang	张康当董事长，詹丹当厂长，张康帮助詹丹，詹丹帮助张康（《张康和詹丹》）
Coda	uan/uang	那边划来一艘船，这边漂去一张床，船床河中互相撞，不知船撞床，还是床撞船（《船和床》）。

APPENDIX F: Tongue twister errors by position in tongue twister

Previous research has shown that when two or more consecutive words are held constant in a specific position in contrast to another position (e.g. in '*She* ***sells sea*** *shells*,' the onset /s/ /ʃ/ /s/ /ʃ/ changes in an ABBA pattern while the /i/ nucleus is held constant), there is a greater chance of speech error (Goldstein et al., 2007; Pouplier & Goldstein, 2010). Both ABAB (repeated alternating pattern, such as *soap* ***dam*** *seam* ***dip***; Whilshire, 1999) and ABBA patterns elicit errors, but in the ABBA format, the first and third words are subject to the greatest errors, See Figure 10 (Croot et al., 2010; Shattuck-Hufnagel, 1992).

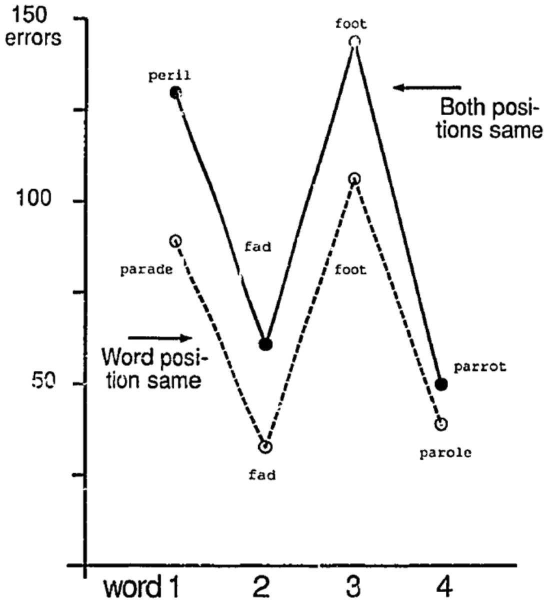


Figure 1. *Error rate by word position in the stimulus (Experiments 1 and 1(a) combined).*

Figure A: Word position errors in ABBA onset tongue twister pattern. From "The role of word structure in serial word processing," by S. Shattuck-Hufnagel, 1992, *Cognition*, 42 (1992) 213-259.

There are a few explanations for the phenomenon of tongue twister errors: some espouse the coupled oscillator model, based off of gestural phonology (Pouplier & Goldstein, 2010); while others prefer a general phonological competition model (Goldrick & Blumstein, 2006; McMillan & Corley, 2010). According to these theories, phonemes in positions 1 and 3 in the ABBA tongue twister format would be subject to the most error. As the research question is not about which theory of speech encoding is correct, I decided to relegate this analysis of the tongue twister results to the Appendix, for those interested in how the results look from a phonological competition model perspective.

Tone and Onset Errors

Alternation Type	T-1	T-2	T-3	T-4	O-1	O-2	O-3	O-4
ABAB	**16**	**10**	8	5	**12**	**12**	9	8
ABBA	7	**13**	**9**	7	**22**	12	**22**	17
Complex ABAB	9	18	**34**	**36**	9	4	**29**	**23**
Complex ABBA	13	12	**18**	**20**	**35**	16	**50**	33

Table A: Tone and Onset Errors Based on Alternation and Position. *T-1 means tone error that exists in the first position of a tongue twister. O-4 means onset error that exists in the fourth position of a tongue twister, etc. The 2 highest error positions are bolded in each category.*

As mentioned in Kember et al. (2015), the coupled oscillator theory (Pouplier & Goldstein, 2010), along with general phonological competition theory Goldrick & Blumstein, 2006; McMillan & Corley, 2010) both predict that the 1st and 3rd items on a 4-item tongue twister will be subject to the most errors in an ABBA pattern. As expected, onset errors in ABBA formats did indeed peak in the first and third positions, for both simple ABBA and complex ABBA onset alternation patterns.

Tone did not show the same patterns of error placement in the ABBA patterns: in the simple ABBA pattern, errors peaked at the 2nd and 3rd positions, and in the complex ABBA pattern, errors peaked at the 3rd and 4th positions.

Tone and Vowel Errors

Alternation Type	T-1	T-2	T-3	T-4	V-1	V-2	V-3	V-4
ABAB	**5**	**3**	**3**	2	**5**	3	**4**	2
ABBA	**7**	**7**	**8**	2	**5**	3	**7**	4
Complex ABAB	10	12	**33**	**24**	5	2	**31**	**12**
Complex ABBA	13	11	**28**	**21**	6	6	**13**	**13**

Table B: Tone and Vowel Errors Based on Alternation and Position. *T-1 means tone error that exists in the first position of a tongue twister. V-4 means vowel error that exists in the fourth position of a tongue twister, etc. The 2 highest error positions are bolded in each category.*

Looking at the errors based on alternation type and position in tone and vowel alternation, I see a slightly different pattern. For simple ABBA, I see that vowel error rates also peak in the 1st and 3rd positions, but tone error rates also peak in the 1st and 3rd positions (with 2nd position tying for 2nd place). In the complex ABBA tongue twisters, however, I see a different pattern altogether. For both complex ABBA patterns in tone and vowels, I see error rates peaking in the 3rd and 4th positions, essentially the end of the tongue twisters. The rates are not negligible as well; there is quite a big difference between the error rates at the end of the tongue twisters for both complex ABBA and complex ABAB patterns.

Looking at the simple ABAB pattern error rates, it is quite surprising to see that vowels showed the most errors in the 1st and 3rd positions; so far, there doesn't seem to be a speech model that explains this clearly. Tones also show the highest error rates in the 1st, 2nd and 3rd positions. However, the data in the simple alternation type tongue twisters may be too little to come to any strong conclusions.

Tone and Coda Errors

Alternation Type	T-1	T-2	T-3	T-4	C-1	C-2	C-3	C-4

ABAB	**21**	11	**13**	**13**	16	**18**	9	**21**
ABBA	**35**	**21**	20	11	**32**	31	**35**	25
Complex ABAB	11	14	**55**	**47**	31	32	**40**	**35**
Complex ABBA	19	13	**29**	**26**	41	33	**54**	**44**

Table C: Tone and Coda Errors Based on Alternation and Position. *T-1 means tone error that exists in the first position of a tongue twister. C-4 means coda error that exists in the fourth position of a tongue twister, etc. The 2 highest error positions are bolded in each category.*

The alternation pattern and positional distributions of tone and coda errors are quite similar to those of tone and vowel. Complex ABAB and complex ABBA tongue twisters for both tone and coda alternating patterns showed the highest error rates in the 3rd and 4th positions. Again, it is hard to explain the cause of these similar error distributions across ABAB and ABBA tongue twister patterns in both tone and coda.

Regarding simple ABBA, coda did maintain the segmental pattern of showing the highest error rates in the 1st and 3rd positions. Tone did not show the exact same pattern, but the 1st position did indeed show the highest error rates, with the 2nd and 3rd positions approximately tying in 2nd place.

In the simple ABAB alternation type, coda error distribution shows strong differences from those of its segmental counterparts, onset and vowel. Instead of errors falling mostly in the first half (onset) or in the 1st and 3rd positions (vowel), coda errors peaked in the 2nd and 4th positions. For tone in the simple ABAB alternation type, errors peaked in the 1st position.

Conclusion

The three segmental categories (onset, vowel and coda) indeed show the classic high-error prominence on the 1st and 3rd positions in the simple ABBA pattern, and replicate the

Kember et al. (2015) study in that the segments show influence by the coupled oscillator theory (Pouplier & Goldstein, 2010). Other than in the tone-vowel alternation, tone error did not show this distribution pattern. According to these theories, from these results, I can see that tone is not produced by the same mechanism that is responsible for producing segments.