Acceptance of tonal and segmental variability correlates to inventory size in Mandarin Chinese

Julie Siying Chen[1], *Stephen Politzer-Ahles*[2]

[1]The Hong Kong Polytechnic University
[2]University of Kansas
julie-siying.chen@connect.polyu.hk, sjpa@ku.edu

Abstract

Lexical access in spoken language recognition is commonly seen as activating multiple lexical candidates based on phonetic similarity as the speech signal unfolds. How different types of phonemes contribute to lexical access has been a question for quite some time. In Mandarin Chinese, a language with lexical tone, research has been sparse on the different levels of lexical contribution of tones and segments.

The present study uses a novel binary lexical selection task to test whether or not tones are more lexically binding than three categories of segments: onsets, vowels, and codas. Our results showed that tones are more susceptible to change compared to onsets and vowels, but more stable when compared to codas. These results are in line with the respective size of these categories in the Chinese phoneme inventory. Thus, we believe that tones are not processed in a qualitatively different way from segments, but that all categories are subject to the same restraints.

Index Terms: Mandarin Chinese, lexical tone, speech perception, lexical access, lexical selection, segments and suprasegmentals

1. Introduction

In Indo-European languages, studies have shown that when faced with a nonword and asked to change it into a real word by changing either a consonant or a vowel, speakers overwhelmingly chose to change the vowel. Furthermore, speakers were faster and more accurate when instructed to change the vowel than when instructed to change the consonant. Created by Van Ooijen in 1996 [7], the results of this word reconstruction task led to the universal vowel mutability theory, which posits that speakers cope with vowel variability because vowels constrain lexical access less tightly than consonants. Because the task was originally tested on English speakers, researchers wondered if the results are a product of the skewed vowel-to-consonant ratio in English. Since English has fewer candidates for vowel substitution, the higher accuracy and faster response times could be a result of simpler lexical search.

For example, one of the critical stimuli from the experiment was the nonword "kebra." Participants would most likely recall the consonant space (which has 24 items) and/or the vowel space (which has 17 items) to determine to change either a consonant or vowel in "kebra" to make it into a real word. In the participant's mind, then might go through the alphabet and start with consonants: "bebra," "debra," "febra"; similarly, they may do so with vowels: "kabra," "kibra," "cobra," etc. Because there are fewer items in the vowel category, it's speculated that it will take a participant less time to find a "correct" replacement – thus, simpler lexical search.

Another hypothesis came from the idea that vowels in English have a high amount of acoustic overlap and offer more perceptual confusion when compared to consonants. To test these hypotheses, the word reconstruction task was replicated in Dutch (which has a more balanced consonant-to-vowel ratio) and Spanish (which has a very high consonant-to-vowel ratio and more acoustically distinctive vowels), and the results surprisingly matched the results from the original English study, leading researchers to reject both alternative hypotheses and reaffirm the universal vowel mutability hypothesis [3].

These studies provide a persuasive argument for the unequal distribution of strength of consonants and vowels in spoken word recognition. However, it's important to note that although the Dutch phoneme system has a more balanced consonant-to-vowel ratio compared to English and Spanish, there are still more consonants (19) than vowels (16) in Dutch. Thus, all 3 languages do indeed have more consonants than vowels in their phoneme inventories, so we cannot rule out the inventory size effect on the results of the word reconstruction task.

Additionally, suprasegmental information, especially in languages that have lexically distinctive suprasegmental information, has not been explored with much frequency. Mandarin Chinese is one such language, in which tones (the pitch accompanying the segmental elements of the word) are lexically distinctive in the same way that consonants and vowels are.

Wiener and Turnbull [6] replicated the word reconstruction task in Mandarin Chinese and found results that challenged the universal vowel mutability hypothesis. Using accidental gaps in Mandarin, which are phototactically legal syllables but whose syllable and tone combination do not exist in the current language, Wiener and Turnbull found that speakers chose to change the tone most frequently, followed by consonants, with vowels most resistant to change. In contrast to the results from the Indo-European languages, consonants were changed with faster response times and with higher accuracy. This finding challenges the universality of the vowel mutability hypothesis, showing that different types of segments may have varying contribution to lexical access in Mandarin Chinese. In addition, the finding that tone, a suprasegmental phoneme category, was changed more often than both consonants and vowels is in line with other studies in the field that argue that tones are processed in a qualitatively different way compared to segments (Chen et al., 2002).

However, similar to Cutler et al. [3]'s idea that the results may be modulated by vowel inventory size, we believe the results of this study may be a result of there being fewer tones

979-8-3503-9796-3/22/$31.00

(4) than there are consonants (20)[1] in the phoneme inventory—meaning that they contribute less to word recognition—, not because tones are processed in a different way. To test the effect of inventory size of a phonemic category on speakers' ability to accept variation, we decided to extend the bounds of the Wiener and Turnbull study by comparing tones to onsets and vowels, in addition to comparing tones and codas, which has an even smaller phoneme inventory in Mandarin (2). We created a new behavioral task to better measure this variation.

2. Methods

The new behavioral task we created for this research question is called the binary lexical selection task. In this task, a recording of an accidental gap nonword is played, and then the participant is presented with 2 words on the computer screen, both of which differ from the spoken nonword in exactly 1 phoneme, each from a different phoneme category. The participant is then asked to select the word that most closely resembles the utterance they just heard. This method eliminates explicit lexical search, which was a critique of the word reconstruction task. Instead of asking participants to be productive, this task gives them straightforward options akin to the "yes, it's a word" or "no, it's not a word" options in the classic lexical decision task and provides a more direct answer to the research question: is acceptance of variability modulated by the inventory size of that category?

The experiment uses much of the same stimuli from Wiener and Turnbull [6], but instead of asking participants to choose a phonemic item to turn a nonword into a real word, we asked participants to choose one of two real words that mismatched the nonword in either a tone or a segment.

Specifically, we were interested in pitting tone against onset (the second-most lexically binding phoneme category in the Wiener & Turnbull's study) and coda (a phoneme category unexplored in the Wiener & Turnbull study). For example, a participant would hear the utterance **ca2* (/tsʰa/ with rising tone) and be given two options to choose from: the onset mismatch item 杂 (/tsa/ with rising tone), and tone mismatch item 擦 (/tsʰa/ with high tone). To allow for realistic responses, the mismatch options were chosen to be as phonetically similar as possible to the spoken utterance. All onset mismatch options were chosen to be only **one** feature difference of the spoken utterance: they differed in either manner of articulation, place of articulation, or voicing. Tones were also chosen to be in pairs that are deemed to be phonologically similar: Tone 1 (high level) and 2 (mid rising) were paired together, whereas Tones 3 (low dipping) and 4 (high falling) were paired together. Participants were told to choose the word (written in simplified Chinese characters) that most closely resembles the utterance they heard.

2.1. Experiment 1 (N=69)

Our first experiment, which was conducted over the internet using PsychoPy [5], recruited 69 native Mandarin-speakers living around the world. The experiment used stimuli such as the **ca2* example shown above, and included 40 critical tone versus onset mismatch items, in addition to 40 control stimuli of matching pattern, to control for language proficiency. Participants who scored less than 80% accuracy in the control items were excluded from the study (for a total of 3 excluded participants). The results showed that participants chose to choose the tone mismatch item 85% of the time, which means that participants accepted tone variability more than onset variability 85% of the time.

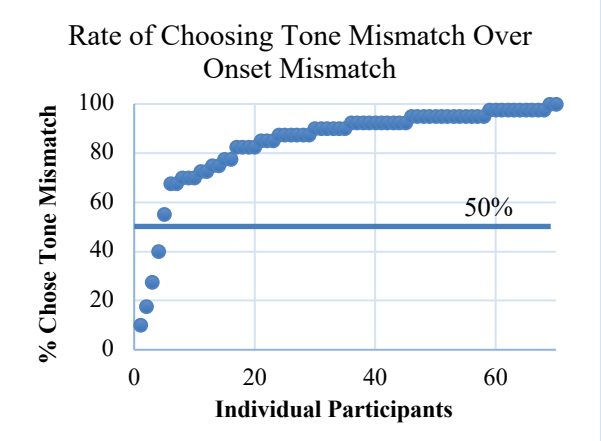

Figure 1: *This chart shows the rate the participants in Experiment 1 chose to choose the tone mismatch item over the coda mismatch item. The clear trend is that a majority of participants chose the tone mismatch item.*

2.2. Experiment 2 (N=62) and 3 (N=69)

The 2nd and 3rd experiments were conducted in-person. A total of 131 participants, the majority of whom were international students from Mainland China, were recruited, and all experiments were run on Lenovo workstations with noise-cancelling headsets, using DMDX software. The 2nd and 3rd experiments included the tone and onset mismatch items, in addition to 22 new tone and coda mismatch items, as well as 20 new control items.

In Mandarin Chinese, the coda has only two overt possibilities: /n/ and /ŋ/ (in addition to the lack of coda). The two coda-position phonemes form minimal pairs in many lexical items, but because there are only two phonemes, there are substantially fewer accidental gaps that can be made for these items, leading to only 22 critical coda mismatch stimuli. An example of the critical tone and mismatch item is shown in Figure 2.

Because some southern Chinese dialects do not differentiate between /n/ and /ŋ/, the 20 new coda control items were vital to ruling out participants who could not differentiate between /n/ and /ŋ/ in Mandarin Chinese. Similar to the first experiment, we eliminated participants whose accuracy on the control coda stimuli was 80% or lower. We did not eliminate any participants in the 2nd experiment, but using this criterion 2 participants were eliminated in the 3rd experiment.

2.3. Experiment 4 (N=69)

The 4th experiment was conducted over the internet using PsychoPy [5], much like the first experiment. Previously,

[1] There are classically 4 tones in Mandarin, but there is also potentially a 5th tone, which is a phonetic lack of tone. We could also consider consonants to have an inventory of 21, since a lack of onset can also be considered a consonant. However, no matter how we count them, there are many more consonants than tones in Mandarin.

because the phoneme category most resistant to change in Wiener & Turnbull's study was vowels, we decided not to include vowels in the experiment, due to the fact that we did not think that there would be much point in testing vowels.

However, due to the idea that lexical activation may be incremental, it might be argued that onsets are most lexically binding, that tones are middle-of-the-line binding, and that codas are the least lexically binding based on their locations in a word, we decided to do one last experiment pitting tone with vowels (which exists in approximately the same time space as tones).

Like the stimuli from the previous 3 experiments, we selected accidental gap nonwords whose options would differ in exactly tone or 1 vowel. For the purposes of consistency, we chose monophthong vowels only. Due to the nature of these limitations, we were only able to select 12 critical stimuli in this category. We also created new control items that tested the reading and listening skills of the participants and eliminated participants who scored less than 80% on these items. Under this criterion, 2 participants were excluded.

Figure 2: *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". This stimulus was used for experiments 2 and 3.*

3. Results

From the 1st experiment to the second two experiments, we added the tone versus coda mismatch items, but the results of the tone versus onset mismatch items did not change. Participants forced to choose between changing the onset or changing the tone chose to change the tone most of the time, with 88% tone change in the 2nd experiment and 85% tone change selection in the 3rd experiment (see Figure 3 and 4). This gives strong evidence that Mandarin speakers were able to accept tone variability a lot more than onset variability. These results are in line with results from Wiener & Turnbull's study [6].

In experiment 4 (see Figure 5), when we pitted tone against vowels, the results were quite similar to the results of tones being pitted against onsets. Participants selected the tone change items 88% of the time. This shows that similar to onsets, vowels were strongly resistant to change when compared to tones.

However, when forced to choose between changing the *coda* or changing the tone, participants chose to switch the coda slightly more often than tone, changing the coda 55% of the time. This shows that although tone is more subject to variability compared to onsets, tone is more resistant to change compared to codas. Given that coda has a smaller inventory (2) compared to tones (4), this gives support for the idea that acceptance of variability of a particular category is based on the inventory size of that category.

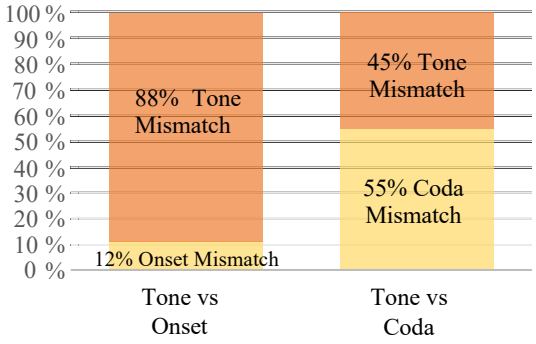

Figure 3: *Results from Experiment 2: Binary Lexical Selection. Note that onset change is very minimal, whereas coda change is more common.*

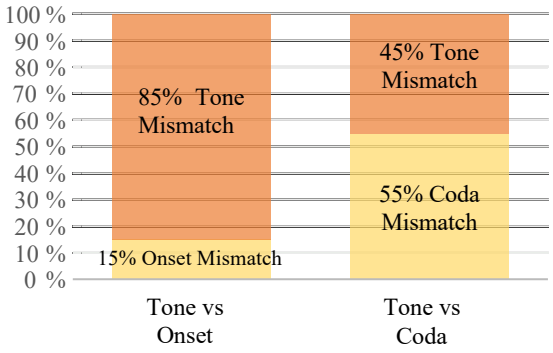

Figure 4: *Results from Experiment 3: Binary Lexical Selection. Note that onset change is very rare, whereas coda change is more common.*

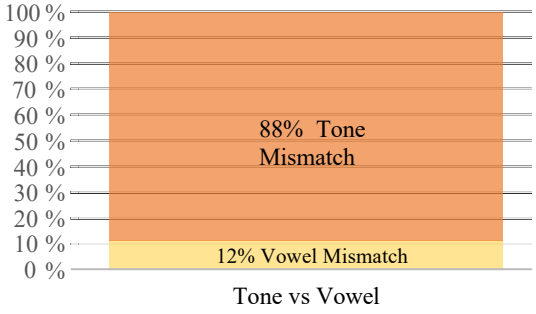

Figure 5: *Results from Experiment 4: Binary Lexical Selection. Similar to the results from experiments 1-3, vowel change is very rare, comparable to onset change.*

4. Discussion

This series of experiments aimed to find a better way to test tonal variability in comparison to segment variability in spoken language processing. Participants were more tolerant of

variability in tones than onsets and vowels, as seen in previous research. However, when we used the binary lexical selection task on a new phonological category (codas) so far not studied in similar studies, we found that this difference is not due to a qualitative difference between tones and segments, but just a result of its phonological inventory size: participants are more sensitive to deviations in phonemes that contribute more information to word recognition in that language.

Additionally, when revisiting the results from the Cutler et al. study [3], which showed that Dutch, a language with a relatively balanced consonant-to-vowel ratio, also showed higher rates of vowel change, it's uncertain that this is really a case of vowels behaving differently than consonants. In a previous study, Costa, Cutler and Sebastian-Galles (1998) showed that Dutch speakers showed relatively similar amounts of attention when presented with a task that asked for vowel uncertainty on consonant deduction and consonant uncertainty on vowel deduction. The same results were not found for Spanish, which has approximately 4 times more consonants than vowels. The imbalance of consonants to vowels in Spanish was seen as the key contributor of the high amount of consonant unpredictability in the task. We believe a similar thing is happening in both the word reconstruction task and the binary lexical selection task, which we invented to test the effect in Mandarin Chinese. Because the Mandarin coda has only two possibilities, when compared to the 4 possibilities in tone, participants derive less information from the coda and are thus more able to accept variability for the coda.

However, there is an alternative explanation for the results from this study. Several psycholinguistic theories provide evidence showing that the information in the earlier part of a word contributes more to lexical access than the latter part of the word. In the top-down model of Marlsen-Wilson's cohort model, lexical activation is incremental, and in some cases, one does not need to hear the end of a word in order to retrieve the word [5]. If this theory affects the binary lexical decision task, that may be a possible explanation for the results, seeing as the category most prone to variability in our results was the coda, the last phoneme in a word.

Knowing this, we conducted a 4th and final experiment in attempt to control for the effects of incremental lexical activation by testing vowel variability. Vowels exist in approximately the same time space as lexical tones in Mandarin but offer a much larger inventory size. If the experimental data could be simply explained by the position of a phoneme in a word, then vowels should be subject to the same variability acceptance as tones. However, this was not the case, as the vowel variability acceptability was almost identical to that of onsets, which has an inventory size of 21. The complex nature of the concept of vowels in Mandarin (there are 7 monopthongs, with an additional 30 dipthongs and compound vowels, some of which include the nasal codas we mentioned earlier) makes it hard to specify the size of the vowel inventory, but we are fairly confident that it's the relatively large inventory size of vowels in Mandarin that explains the minimal vowel variability acceptance in this experiment.

Barring the alternative explanation, we firmly believe that the binary lexical selection task invented for this study works as a better alternative for testing phoneme variability acceptance in a language like Chinese, whose phonological system traditionally does not divide segments into consonants and vowels, but rather "initials" and "finals," which are approximately represented by onsets and rimes, respectively. In the Wiener & Turnbull study, the word reconstruction task asked participants to change "initials" and "finals," which are not equivalent to consonants and vowels. Due to the constraints of the Chinese phonological system, the word reconstruction task is not the best way to test phoneme variability acceptance.

5. References

[1] 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, 751-781.

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

[3] 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.

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

[5] 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.* 10.3758/s13428-018-01193-y

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

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