Phonological Variation and Word Recognition in Continuous Speech

DISSERTATION

Presented in Partial Fulfillment of the Requirements for
the Degree Doctor of Philosophy in the
Graduate School of The Ohio State University

By

Lei Xu, M.A.

* * * * *

The Ohio State University

2007

Dissertation Committee:
Shari Speer, Adviser
Professor Mary E. Beckman
Professor Marjorie K.M.Chan

Approved by

______________________________
Adviser
Graduate Program in
Linguistics
ABSTRACT

In natural, continuous speech, words are not always produced in the “canonical” way as described in the dictionary. For example, the phrase “right berry” can sound like “ripe berry”. The research question for the current study is: how native listeners resolve the ambiguity quickly. More specifically, how “words” undergo sound changes in connected speech are stored in the mind, and how contextual factors such as prosodic boundary information influence the processing of this kind of ambiguous words.

Two eye movement monitoring experiments investigated listeners’ processing of ambiguous words in connected speech. The ambiguities in the experiment stimuli resulted from two types of phonological sound changes in phonetic context - English coronal assimilation and Putonghua Tone2 sandhi. Three different kinds of prosodic boundaries (Prosodic Word Boundary, Intermediate Phrase Boundary, Intonation Phrase Boundary) were inserted in between the critical word and its following context word to elicit different levels of assimilated tokens in natural speech. In each trial of the experiment, the participants’ task was to choose the most appropriate one from four pictures on the computer screen upon hearing a pre-recorded auditory sentence that contains the ambiguous words.

The results are most consistent with the proposed modified exemplar processing account, in which both abstract categories such as “word” and “prosodic boundary”,
together with each episode of sound change is stored in the listener’s mental lexicon, and are called to act in concert to help with the mapping of the incoming acoustic signal to the word to be retrieved.

The results provide new evidence that listeners use prosodic boundary information in a very early stage of lexical processing and that contrastive stress causes a heavier load of processing. They also support previous claims that IP boundaries trigger semantic wrap-up.
Dedicated to my parents and my husband.
ACKNOWLEDGMENTS

I would like to thank my advisor, Shari Speer, for giving me the chance to study psycholinguistics with her. I am extremely grateful for her patience, her encouragement, and the countless hours she spent on revising my paper and experimental designs. In addition to being a source of wisdom and knowledge, she has been a great role model in my personal life.

I would also like to thank my committee members – Mary Beckman and Majorie Chan – for their encouraging and critical comments on written drafts. Thank you, Mary, for helping me to write Praat Script and make me fascinated by phonetics. Thank you, Ma Laoshi, for keeping me interested in Chinese phonology.

I would like to thank SpeerLab members and fellow students at the Linguistic Department: Ping Bai, Kiwako Ito, Allison Blodgett, Laurie Maynell, Soyoung Kang, Anouschka Burgman, Jeonghwa Shin, Abbey Southerland, Jianguo Li, Xiaofei Lu, Fangfang Li, EunJong Kong and Tsan Huang for their nourishing friendship, much-needed technical support, constructive comment and practical suggestions.

I would like to thank my parents and my husband. I couldn’t have made it without their continual encouragement, endless support, and abundant love. You are the best part of my life.
I would like to thank the Linguistics Department of Ohio State University for the financial support, and the faculty, staff, graduate students at Linguistics Department, who have inspired me with their hard work, their creativity, and their willingness to help.

I thank God for all these blessings.
VITA

March 20, 1972. . . . . . . . . . . . Born -Shanghai, P. R. China

1994. . . . . . . . . . . . . . . . . . . . . B.A. English literature and linguistics
   East China Normal University
   Shanghai, P.R. China

1999 - 2001. . . . . . . . . . . . . . . M.A. Teaching English as a second language
   English Department, University of Arizona

2001 - 2006. . . . . . . . . . . . . . . Graduate teaching and research associate
   The Ohio State University

PUBLICATIONS

   Proceedings of the international conference on speech prosody, Dresden, Germany.

FIELDS OF STUDY

Major field: Linguistics

Specialization: Psycholinguistics
# TABLE OF CONTENTS

<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Abstract</td>
<td>ii</td>
</tr>
<tr>
<td>Dedication</td>
<td>iv</td>
</tr>
<tr>
<td>Acknowledgments</td>
<td>v</td>
</tr>
<tr>
<td>Vita</td>
<td>vii</td>
</tr>
<tr>
<td>List of Tables</td>
<td>xii</td>
</tr>
<tr>
<td>List of Figures</td>
<td>xiv</td>
</tr>
<tr>
<td>Chapters:</td>
<td></td>
</tr>
<tr>
<td>1. Introduction</td>
<td>1</td>
</tr>
<tr>
<td>1.1 The Problem</td>
<td>1</td>
</tr>
<tr>
<td>1.2 Two Types of Assimilation</td>
<td>3</td>
</tr>
<tr>
<td>1.2.1 English Coronal Assimilation</td>
<td>3</td>
</tr>
<tr>
<td>1.2.2 Putonghua Tone2 Sandhi</td>
<td>10</td>
</tr>
<tr>
<td>1.2.3 Implication for the current study</td>
<td>18</td>
</tr>
<tr>
<td>1.3 Prosodic structures in English and Putonghua</td>
<td>19</td>
</tr>
<tr>
<td>1.3.1 Prosodic Structure of English</td>
<td>20</td>
</tr>
<tr>
<td>1.3.2 Prosodic Structure of Putonghua</td>
<td>21</td>
</tr>
<tr>
<td>1.3.3 The effect of prosodic boundaries on the perception of changed sounds</td>
<td>25</td>
</tr>
<tr>
<td>1.4 Processing models</td>
<td>28</td>
</tr>
<tr>
<td>1.4.1 The phonological inference account</td>
<td>29</td>
</tr>
<tr>
<td>1.4.2 The Feature Cue Parsing Account</td>
<td>31</td>
</tr>
<tr>
<td>1.4.3 The perceptual-integration account</td>
<td>35</td>
</tr>
<tr>
<td>1.4.4 A Revised Exemplar Account</td>
<td>37</td>
</tr>
</tbody>
</table>
5.2 Picture-Word Matching Experiment .............................. 93
  5.2.1 Introduction .............................................. 93
  5.2.2 Method .................................................. 94
  5.2.3 Results .................................................. 96
  5.3 Summary .................................................... 98

6. EYE TRACKING EXPERIMENT ON ENGLISH CORONAL ASSIMI-
   LATION ......................................................... 99
  6.1 Introduction ................................................ 99
  6.2 Why Eye Tracking .......................................... 99
  6.3 Prediction ................................................ 101
  6.4 Method ..................................................... 105
    6.4.1 Participants ........................................... 105
    6.4.2 Design ............................................... 105
    6.4.3 Auditory Stimuli ...................................... 107
    6.4.4 Visual Stimuli ........................................ 107
    6.4.5 Procedure ............................................ 108
  6.5 Results ................................................... 111
    6.5.1 Condition 1 Non-coronal, Word Boundary ............. 113
    6.5.2 Condition 2 Non-coronal, ip Boundary ............... 115
    6.5.3 Condition 3 Non-coronal, IP Boundary ............... 118
    6.5.4 Condition 4 Coronal, Word Boundary ................. 120
    6.5.5 Condition 5 Coronal, ip Boundary ................. 124
    6.5.6 Condition 6 Coronal, IP Boundary ................. 127
  6.6 Discussion and conclusion ................................. 129

7. PRETESTS FOR PUTONGHUA EYE TRACKING EXPERIMENT .... 132
  7.1 Listening Pretest ......................................... 132
    7.1.1 Introduction ......................................... 132
    7.1.2 Method ............................................... 133
    7.1.3 Results ............................................... 136
  7.2 Picture-Word Matching Experiment .......................... 138
    7.2.1 Introduction ......................................... 138
    7.2.2 Method ............................................... 138
    7.2.3 Result ................................................ 141
    7.2.4 Discussion .......................................... 142

8. PUTONGHUA EYE TRACKING EXPERIMENT ...................... 143
  8.1 Introduction ............................................... 143
8.2 Method .................................................. 144
  8.2.1 Participants ........................................ 144
  8.2.2 Design ............................................... 145
  8.2.3 Auditory Stimuli ................................... 146
  8.2.4 Visual Stimuli ...................................... 146
  8.2.5 Procedure .......................................... 147
8.3 Results ................................................. 148
  8.3.1 Condition 1 Tone1, Word Boundary ................. 149
  8.3.2 Condition 2 Tone1, ip Boundary .................... 152
  8.3.3 Condition 3 Tone1, IP Boundary .................... 154
  8.3.4 Condition 4- Tone2, Word Boundary ............... 157
  8.3.5 Condition 5- Tone2, ip Boundary .................. 159
  8.3.6 Condition 6- Tone2, IP Boundary .................. 161
8.4 Discussion ............................................... 163

9. CONCLUSION AND DISCUSSION .......................... 167
  9.1 Experiment Result .................................... 167
    9.1.1 Comparison between each condition ............... 167
    9.1.2 Comparison between boundary types and discussion on segmental vs. suprasegmental context effect .............. 174
    9.1.3 Comparison between critical word types and discussion on segmental vs. suprasegmental assimilation ............ 174
  9.2 Implication to Processing ............................ 174
    9.2.1 Previous accounts ................................ 174
    9.2.2 Current account ................................... 175
  9.3 Conclusion ............................................ 175

Appendices:

A. Complete Sentences for the English Experiments ............... 177

B. Sample Sentences for the Putonghua Experiments ............... 184

Bibliography ............................................... 228
# LIST OF TABLES

<table>
<thead>
<tr>
<th>Table</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>2.1 Mean durations of the Critical Word, the Vowel in the Critical Word, the Interval from Critical Word Onset to Context Word Onset in each condition</td>
<td>46</td>
</tr>
<tr>
<td>4.1 Comparison of English and Putonghua Mean Critical Word Duration</td>
<td>85</td>
</tr>
<tr>
<td>5.1 Experimental conditions for pre-testing auditory materials, English</td>
<td>90</td>
</tr>
<tr>
<td>6.1 Experimental conditions and auditory materials, English</td>
<td>106</td>
</tr>
<tr>
<td>6.2 Result for Planned Comparison between Target and Competitor means every 150 ms between 0-900 ms after the offset of the vowel in the critical word in Condition 1, English eye tracking experiment</td>
<td>115</td>
</tr>
<tr>
<td>6.3 Results from Planned Comparison between Target and Competitor means every 150ms between 0-900 ms after the offset of the vowel in the critical word. Condition 2, English eye tracking experiment</td>
<td>118</td>
</tr>
<tr>
<td>6.4 Planned Comparison results between Target and Competitor every means 150ms between 0-900 ms after the offset of the vowel in the critical word, Condition 3, English eye tracking experiment</td>
<td>121</td>
</tr>
<tr>
<td>6.5 Result for Planned Comparison between Target and Competitor every means 150ms between 0-1500 ms after the offset of the vowel in the critical word. Condition 4, English eye tracking experiment</td>
<td>123</td>
</tr>
<tr>
<td>6.6 Results of Planned Comparison between Target and Competitor means every 150 ms between 0-1500 ms after the offset of the vowel in the critical word. Condition 5, English eye tracking experiment</td>
<td>126</td>
</tr>
</tbody>
</table>
6.7 Results for Planned Comparison between Target and Competitor means every 150ms between 0-1500ms after the offset of the vowel in the critical word. Condition 6, English eye tracking experiment 129

7.1 Putonghua study experiment design 136

8.1 Experimental Conditions for the Putonghua Eye Tracking Experiment 145

8.2 Results for Planned Comparisons between Target and Competitor in 150 ms, 6 windows from vowel offset. Condition 1, Putonghua eye tracking study 151

8.3 Results for Planned Comparisons between Target and Competitor in 150 ms, 6 windows from vowel offset. Condition 2, Putonghua eye tracking study 154

8.4 Results for Planned Comparisons between Target and Competitor in 150 ms, 6 windows from vowel offset. Condition 3, Putonghua eye tracking study 157

8.5 Results for Planned Comparisons between Target and Competitor in 150 ms, 15 windows from vowel offset. Condition 4, Putonghua eye tracking study 165

8.6 Results for Planned Comparisons between Target and Competitor in 150 ms, 6 windows from vowel offset. Condition 5, Putonghua eye tracking experiment 166

8.7 Results for Planned Comparisons between Target and Competitor in 150 ms, 6 windows from vowel offset. Condition 6, Putonghua eye tracking experiment 166
# LIST OF FIGURES

<table>
<thead>
<tr>
<th>Figure</th>
<th>Description</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.1</td>
<td>Wave form and F0 contour for the syllable yu3(rain)</td>
<td>12</td>
</tr>
<tr>
<td>1.2</td>
<td>Wave form and F0 contour for the syllable yu2(sandhi) (rain)</td>
<td>12</td>
</tr>
<tr>
<td>1.3</td>
<td>Wave form and F0 contour for the syllable yu2(fish)</td>
<td>13</td>
</tr>
<tr>
<td>1.4</td>
<td>Wave form and F0 contour for the syllable sequence yu2(sandhi)hen3(fish very)</td>
<td>13</td>
</tr>
<tr>
<td>2.1</td>
<td>F1, F2 and F3 aligned at the end of the vowel for the critical words “back” and “bat” in all three boundary types (Word, ip, IP). The red lines are for the coronal word “bat” and the black lines are for its non-coronal counterpart “back”</td>
<td>53</td>
</tr>
<tr>
<td>2.2</td>
<td>F1, F2 and F3 aligned at the end of the vowel for the critical words “sum” and “sun” in all three boundary types (Word, ip, IP). The red lines are for the coronal word “sun” and the black lines are for its non-coronal counterpart “sum”</td>
<td>54</td>
</tr>
<tr>
<td>2.3</td>
<td>F1, F2 and F3 aligned at the end of the vowel for the critical words “comb” and “cone” in all three boundary types (Word, ip, IP). The red lines are for the coronal word “cone” and the black lines are for its non-coronal counterpart “comb”</td>
<td>55</td>
</tr>
<tr>
<td>3.1</td>
<td>Mean duration of critical word in Putonghua study</td>
<td>63</td>
</tr>
<tr>
<td>3.2</td>
<td>Mean duration of vowel in critical word in Putonghua study</td>
<td>64</td>
</tr>
<tr>
<td>3.3</td>
<td>Mean duration from critical word onset to context word onset in Putonghua study</td>
<td>65</td>
</tr>
</tbody>
</table>
3.4 Mean word initial F0 in Putonghua study ............................... 66
3.5 Mean word medium F0 in Putonghua study ................................. 67
3.6 Mean word final F0 in Putonghua study .................................... 68
3.7 “Tone1 Critical Word Tone1” sequence, F0 aligned at the end of the critical syllables “shi” “hu”, “hu”, and “yin” in all three boundary types (Word, ip, IP). The red lines are for the Tone2 word and the black lines are for its Tone1 counterpart .................................................. 69
3.8 “Tone1 Critical Word Tone1” sequence, F0 aligned at the end of the critical syllables “cheng” “jie”, and “zhi” in all three boundary types (Word, ip, IP). The red lines are for the Tone2 word and the black lines are for its Tone1 counterpart .................................................. 72
3.9 “Tone1 Critical Word Tone2” sequence, F0 aligned at the end of the critical syllables “chuan” “ge”, and “xian” in all three boundary types (Word, ip, IP). The red lines are for the Tone2 word and the black lines are for its Tone1 counterpart .................................................. 73
3.10 “Tone1 Critical Word Tone3” sequence, F0 aligned at the end of the critical syllables “fang” “qiang”, “tong” and “guo” in all three boundary types (Word, ip, IP). The red lines are for the Tone2 word and the black lines are for its Tone1 counterpart .................................................. 74
3.11 “Tone1 Critical Word Tone4” sequence, F0 aligned at the end of the critical syllables “yi” “chu”, and “fa” in all three boundary types (Word, ip, IP). The red lines are for the Tone2 word and the black lines are for its Tone1 counterpart .................................................. 75
3.12 “Tone1 Critical Word Tone4” sequence, F0 aligned at the end of the critical syllables “shou” and “hua” in all three boundary types (Word, ip, IP). The red lines are for the Tone2 word and the black lines are for its Tone1 counterpart .................................................. 76
3.13 “Tone2 Critical Word Tone1” sequence, F0 aligned at the end of the critical syllables “ti”, “yan” and “qin” in all three boundary types (Word, ip, IP). The red lines are for the Tone2 word and the black lines are for its Tone1 counterpart .................................................. 77
3.14 “Tone2 CriticalWord Tone3” sequence, F0 aligned at the end of the critical syllables “pi”, “shi”, “ting” and “xi” in all three boundary types (Word, ip, IP). The red lines are for the Tone2 word and the black lines are for its Tone1 counterpart. .............................................. 78

3.15 “Tone2 CriticalWord Tone4” sequence, F0 aligned at the end of the critical syllables “bo”, “huan”, “shi” and “zhi” in all three boundary types (Word, ip, IP). The red lines are for the Tone2 word and the black lines are for its Tone1 counterpart. .............................................. 79

3.16 “Tone2 CriticalWord Tone4” sequence, F0 aligned at the end of the critical syllables “tan”, “quan”, “fang” and “qiu” in all three boundary types (Word, ip, IP). The red lines are for the Tone2 word and the black lines are for its Tone1 counterpart. .............................................. 80

3.17 Relative minimum F0 in critical word in all 6 conditions. .................. 81

5.1 English listening pretest result. Listeners’ evaluation of the ambiguity level of the critical word with different types of coda (/n/, /t/, /k/, /m/, /ŋ/, /p/) followed by IP, ip, Word Boundary. ................................. 92

5.2 Sample web page of a trial ............................................................. 96

6.1 Sample visual stimuli for English eye tracking experiment. Upper left: apple (irrelevant word), Upper right: cut (coronal), Lower left: cup (non-coronal counterpart), Lower right: key (same initial consonant cohort) ................................................................. 109

6.2 English eye tracking Experiment Condition1 (Non-coronal, Word Boundary): Fixation proportions over time for the Target, Competitor, and the two Distracters. Data are aligned at the offset of the vowel (Point 0) in the critical word. The dashed red line on the left indicates the probable point at which listeners’ eyes move in respond to the onset of the vowel in the critical word. The dashed red line on the right indicates the probable point at which listeners eyes move in respond to the offset of the critical word. ......................................................... 114
6.3 English eye tracking Experiment Condition 2 (Non-coronal, ip Boundary): Fixation proportion over time for the Target, Competitor, and the two Distracters. Data are aligned (Point 0) from the offset of the vowel of the critical word. The dashed red line on the left indicates the probable point at which listeners’ eyes move in respond to the onset of the vowel in the critical word. The dashed red line on the right indicates the probable point at which listeners’ eyes move in respond to the offset of the critical word.

6.4 English eye tracking Experiment Condition 3 (Non-coronal, IP Boundary): Fixation proportion over time for the Target, Competitor, and the two Distracters. Data are aligned (Point 0) from the offset of the vowel of the critical word. The dashed red line on the left indicates the probable point at which listeners’ eyes move in respond to the onset of the vowel in the critical word. The dashed red line on the right indicates the probable point at which listeners’ eyes move in respond to the offset of the critical word.

6.5 English eye tracking Experiment Condition 4 (Coronal, Word Boundary): Fixation proportion over time for the Target, Competitor, and the two Distracters. Data are aligned (Point 0) from the offset of the vowel of the critical word. The dashed red line on the left indicates the probable point at which listeners’ eyes move in respond to the onset of the vowel in the critical word. The dashed red line on the right indicates the probable point at which listeners’ eyes move in respond to the offset of the critical word.

6.6 English eye tracking Experiment Condition 5 (Coronal, ip Boundary): Fixation proportion over time for the Target, Competitor, and the two Distracters. Data are aligned (Point 0) from the offset of the vowel of the critical word. The dashed red line on the left indicates the probable point at which listeners’ eyes move in respond to the onset of the vowel in the critical word. The dashed red line on the right indicates the probable point at which listeners’ eyes move in respond to the offset of the critical word.
6.7 English eye tracking Experiment Condition 6 (Coronal, IP Boundary): Fixation proportion over time for the Target, Competitor, and the two Distracters. Data are aligned (Point 0) from the offset of the vowel of the critical word. The dashed red line on the left indicates the probable point at which listeners’ eyes move in respond to the onset of the vowel in the critical word. The dashed red line on the right indicates the probable point at which listeners eyes move in respond to the offset of the critical word.

7.1 Putonghua listening pretest result. Listeners’ evaluation on the ambiguity level of the critical word in 6 conditions.

7.2 Sample trial in the Putonghua picture-word matching experiment.

8.1 Sample visual stimuli. Upper left: juan3zhi3 (rolled paper, Distracter 1), Upper right: feng1zheng1 (kite, Distracter 2), Lower left: Jiang1zhi2 (straight, Tone2 word), Lower right: Jiang1Zhi1 (ginger juice, Tone1 word)

8.2 Putonghua eye tracking experiment Condition 1 (Tone1, Word Boundary): Fixation proportion over time for the Target, Competitor, and the two Distracters. Data are aligned (Point 0) from the offset of the vowel of the critical word. The dashed red line on the left indicates the probable point at which listeners’ eyes move in respond to the onset of the vowel in the critical word. The dashed red line on the right indicates the probable point at which listeners eyes move in respond to the offset of the critical word.

8.3 Putonghua eye tracking experiment Condition 2 (T1, ip Boundary): Fixation proportion over time for the Target, Competitor, and the two Distracters. Data are aligned (Point 0) from the offset of the vowel of the critical word. The dashed red line on the left indicates the probable point at which listeners’ eyes move in respond to the onset of the vowel in the critical word. The dashed red line on the right indicates the probable point at which listeners eyes move in respond to the offset of the critical word.
8.4 Putonghua eye tracking Experiment Condition 3 (Tone1, IP Boundary): Fixation proportion over time for the Target, Competitor, and the two Distracters. Data are aligned (Point 0) from the offset of the vowel of the critical word. The dashed red line on the left indicates the probable point at which listeners’ eyes move in respond to the onset of the vowel in the critical word. The dashed red line on the right indicates the probable point at which listeners eyes move in respond to the offset of the critical word.  

8.5 Putonghua Eye Tracking Experiment Condition 4 (Tone2, Word Boundary): Fixation proportion over time for the Target, Competitor, and the two Distracters. Data are aligned (Point 0) from the offset of the vowel of the critical word. The dashed red line on the left indicates the probable point at which listeners’ eyes move in respond to the onset of the vowel in the critical word. The dashed red line on the right indicates the probable point at which listeners eyes move in respond to the offset of the critical word.  

8.6 Putonghua Eye Tacking Experiment Condition 5 (Tone2, ip Boundary): Fixation proportion over time for the Target, Competitor, and the two Distracters. Data are aligned (Point 0) from the offset of the vowel of the critical word. The dashed red line on the left indicates the probable point at which listeners’ eyes move in respond to the onset of the vowel in the critical word. The dashed red line on the right indicates the probable point at which listeners eyes move in respond to the offset of the critical word.  

8.7 Putonghua eye tracking Experiment Condition 6 (Tone2, IP Boundary): Fixation proportion over time for the Target, Competitor, and the two Distracters. Data are aligned (Point 0) from the offset of the vowel of the critical word. The dashed red line on the left indicates the probable point at which listeners’ eyes move in respond to the onset of the vowel in the critical word. The dashed red line on the right indicates the probable point at which listeners eyes move in respond to the offset of the critical word.  

156  

158  

160  

162  

xix
CHAPTER 1

INTRODUCTION

1.1 The Problem

As a second language speaker of English, one of the biggest problems I have encountered in understanding native speakers has been that many times, words are not pronounced in the “canonical” way as described in the dictionary. For example, the final /t/ in the word “right” does not seem to get pronounced when embedded in a phrase such as “right berry”. Examples from many other languages demonstrate that it is common for the sounds of speech to be pronounced differently depending on the spoken language context in which they occur. Such sound changes have been described as the application of phonological rules, which can result in lexical ambiguities, even for native speakers. In the example above, the sequence could also be heard as “ripe berry,” when a speaker’s lips close to pronounce the /b/ of “berry” before the /t/ in “right” has been fully articulated. However, native speakers usually can resolve this ambiguity with ease. The research question of this current study is: how do native listeners process ambiguous changed sounds in context?
The changed phonological feature that causes lexical ambiguity can be segmental or supra segmental. The context that causes the sound change can also be segmental or suprasegmental. In the above-mentioned example, the segment /t/ in the word “right” often is perceived as having the “place of articulation” feature of the immediately following context segment, e.g. /b/ in “berry”. This is because the /t/ is produced before or during the closure of the lips for the following /b/ sound, so that listeners might hear it as deleted, or as the labial voiceless stop /p/-thus the confusion between “right” and “ripe”. This process is usually called coronal place assimilation. It is an instance of segmental change caused by segmental context. Suprasegmental sound change caused by suprasegmental context can be illustrated by Putonghua Tone2 sandhi, a process in which the Tone2 on a syllable can be altered in the context of the tones of its surrounding context syllables. For example, in fast speech, the three-syllable word “shou1 yin2 ji1” (cash register) could sound like “shou1 yin1 ji1” (radio). Both types of sound change alter a phonological feature to make the sound more similar to its neighbor, but the size of “the sound” differs, with coronal place assimilation affecting a coda consonant unit and the Tone2 sandhi affecting a syllable unit. Proximity to Prosodic boundaries is another type of suprasegmental context that influences sound change in both languages. For example, both coronal place assimilation and tone sandhi are less likely to take place when the sound that is subject to change and its context are separated by a prosodic boundary.

The experiments presented here examine a well-studied type of segmental sound change - English coronal assimilation - and a less well-studied type of suprasegmental sound change - Putonghua Tone2 sandhi. Segmental and suprasegmental phenomena were compared as they occur in situ (in controlled prosodic phrasal contexts). Eye
tracking methodology was used to investigate native listeners’ online processing of English coronal assimilation and Putonghua Tone2 sandhi in different prosodic boundary conditions as they listened to auditory sentences and chose matching pictures on a computer screen.

The rest of this chapter is organized in the following way:

Section 1.2 gives a brief background of English coronal assimilation and Putonghua Tone2 sandhi. Section 1.3 gives a summary of prosodic structures in English and Putonghua. Section 1.4 reviews related psycholinguistic processing models. Section 1.5 motivates using eye tracking as the research tool.

1.2 Two Types of Assimilation

1.2.1 English Coronal Assimilation

In connected speech in English, consonants produced with coronal place of articulation (consonants articulated by raising the blade of tongue, such as /t/, /d/, /n/) tend to be coarticulated with the context segment that immediately follow it in production (mostly labials and dorsals). This process is usually called “coronal place assimilation” in English. One example is the production of the phrase “right berries”, during which the tip of the tongue approaches the alveolar ridge and moves away again for producing the word final coronal /t/ in “right.” Immediately following this, the lips close for the word initial bilabial /b/ in “berries”. These two gestures, when coarticulated, resulting in the temporally overlapping pronunciation of the alveolar /t/ in “right” and the bilabial /b/ in “berries,” and giving rise to the perception of a bilabial /p/, resulting in a phrase that can sound like “ripe berries” to listeners.
The Gradient Nature of English Coronal Assimilation

A primary issue in the literature on coronal place assimilation is the question of whether the process of assimilation in production is discrete or gradient. The answer to this question is essential to understanding how sound alteration is processed by the listener. If assimilation is discrete (the perceived articulation of the coronal segment is either assimilated or not assimilated to the place of the following segment). An assimilated sound should leave no perceivable sound “trace”, and listeners would need to depend entirely on the segmental context surrounding the changed sound, or on the semantic context to determine whether they heard a coronal consonant or not. However, if English coronal assimilation is gradient (there are varying degrees of articulatory gestural overlap of coronal place toward the following segment, and these map to different degrees of recoverability of the coronal specification on the part of listeners), listeners might be able to use acoustic cues to decipher those tokens that are not completely “assimilated”.

Phonetic studies in general have shown that English coronal assimilation is gradient. Articulatory accounts of English coronal assimilation describe the process as spatial and temporal gestural overlapping (Nolan, 1992; Browman and Goldstein, 1990), consistent with the gradient account for coronal assimilation, as the gestures could have different degrees of overlap. For example, X-ray microbeam data from a speaker producing the phrase “seven plus seven times” was closely examined at the part where the final [n] in the first “seven” became [m] before [p] in “plus.” It was found that the alveolar closure gesture was still present in the assimilated version of the production but aerodynamically “hidden” by the more anterior labial closure (Browman and Goldstein, 1990). Electropalatography (EPG) was used to
record tongue contact with electrodes embedded in an artificial palate, when subjects produced phrases such as “late calls” and “make calls”, “boat covered” and “oak cupboard”, tokens with complete alveolar occlusion (also called full alveolar), fronted tongue but no alveolar occlusion (also called residue alveolar and almost velar only (also called zero alveolar) were compared. Results for the sequences with the coronal place assimilation context (namely “late calls” and “boat covered”) showed clear evidence of gradient coronal place assimilation instead of a discrete change (Nolan, 1992). In addition, formant transition analyses on minimal triplets of assimilated coronals, coronals, and noncoronals showed that the formant movement pattern of the assimilated coronal was intermediate between those of coronals and noncoronals (Gow and Hussami, 1999). (for evidence against the gradient view, see the spectral envelope data presented by Holst (Holst and Nolan, 1995)).

Results from perceptual studies also support the gradient account, and indicate that listeners are sensitive to differences in the degree of assimilation produced by speakers. Listeners were asked to choose one of two words (e.g. road or rogue) according to what they heard (in sentences such as “when does the road collapse?”). The /d/ in “road” was manipulated to be a full alveolar, a residue alveolar, a zero alveolar or a nonalveolar). Results showed the percentage of the alveolar [d] response was the highest (almost 100%) with the full alveolar condition, lower (less than half) with the residue alveolar condition, and the lowest with zero alveolar and nonalveolar (Nolan, 1992). In a follow-up experiment, the sentences with zero alveolar and nonalveolar were played as a pair for discrimination, and the listeners’ percentage of
correct identification for the zero-alveolar conditions were above chance level. This result showed that zero alveolar (the most strongly assimilated alveolar) was perceived differently from the non-alveolar.

**Factors Influencing Coronal Assimilation**

Coronal assimilation in English does not always take place each time a coronal consonant is pronounced followed by a labial or dorsal consonant. Therefore, to generate appropriate stimuli for this study, it is important to identify factors that influence the production of English coronal assimilation. These include speech rate and the strength of any prosodic boundaries that might intervene between the potentially assimilated segment and the context segment.

*Speech rate* was found to be an important factor for the occurrence of coronal place assimilation in production. In Electropalatography (EPG) studies, less alveolar contact was found when speakers produced connected speech at a fast rate (Nolan, 1992). Research (Byrd and Tan, 1996; Byrd, 1996) has examined how speech rate affects articulation units - Do speakers shorten the duration of each articulatory unit, and reduce the spatial magnitude of them as a consequence, or do they instead increase the overlap of the units in order to shorten the overall unit duration? The EPG data demonstrated that regardless of the place and manner of articulation, or the position in the syllable, the duration of consonants shortened as the speech rate become faster. In addition, the consonant at the coda position had significant spatial reduction as the rate increased, and the overlap of the consonants increased as the rate increased. Among the three, duration reduction was the most frequent means employed by the speakers to increase their speech rate. Presumably, this was a result
of the speakers’ intended control over their speech so that more perceptual cues could be preserved for the listeners. This study confirmed that coronals reduced more than non-coronals, a finding that can serve as the basis for an explanation of the source of listeners’ facility in resolving the potential ambiguity created by coronal place assimilation.

Prosodic boundaries. Prosody (lexical stress, syllable structure, intonational phrasing) influences articulation in various ways. Studies on how prosodic boundaries influence English coronal place assimilation brought people’s attention to how prosodic structure affects articulatory coordination during the production of consonant sequences. Phonetic studies have revealed different types of cues that are associated with prosodic boundaries, including changes in F0 contour (Beckman and Pierrehumbert, 1986; Yoon, 2005), glottalization (Dilley et al., 1996), VOT (Jun, 1995), phrase final lengthening (Lehiste, 1970; Oller, 1973; Edwards et al., 1991; Wightman et al., 1992), changes in the length of pauses and speech rate (Cho, 2003), amplitude, and the size of the articulatory gesture (Fougeron and Keating, 1997).

More specifically, prosodic boundaries have been found to influence the spatial-temporal properties of individual articulatory gestures. In the spatial aspect, the magnitude of lingual gestures for an initial consonant increases as the prosodic domain it belongs to increases (word-initial < phonological/accentual phrase-initial < intonational phrase initial) (Byrd et al., 2000), and articulatory gestures are larger at domain initial than at medial or final positions at the same level (Fougeron and Keating, 1997). In the temporal domain, various kinds of units (including last vowel,
final consonants, VCs, syllables, and words) have been found to lengthen at the domain final position, and gestures also lengthen at both phrase edges (Beckman et al., 1992; Byrd and Saltzman, 1998; Byrd et al., 2000; Byrd and Saltzman, 2003).

Byrd et al. (2000) studied the relation between temporal overlapping of gestural coarticulation and phrase structure. Movement tracking with a magnetometer was used to study a Tamil speaker’s production of [n#m] and [m#n] sequences across a word boundary, a small phrase boundary and a large phrase boundary (# indicates word boundary location). Different “articulatory signatures” were found between segments on either side of large phrase boundaries vs. small phrase boundaries. For example, as the strength of the boundary increased from small to large, results showed lengthening of the opening of the preboundary consonant, gestural magnitude increase in the closing for the postboundary consonant, and less temporal coproduction between the two consonants.

These findings lead to the following general predictions about the effect of prosodic boundaries on the operation of English coronal assimilation:

*The larger the phrase boundary, the less likely the occurrence of coronal place assimilation.*

With other factors that affect English coronal place assimilation held constant, such as speaker, speech rate and speech style, it is predicted that:
The larger the prosodic separation between a coronal consonant and the consonant that follows it, the less the gestural overlap between the coronal and its following consonant.

Additional factors might modulate the operation of English coronal place assimilation. The influence of the intonation phrase boundary on the perception of coronal assimilation might differ from that of the intermediate phrase boundary for semantic and pragmatic reasons. The intonation phrase boundary has been associated with the edge of a “sense unit” in spoken sentence structure, while the weaker intermediate or prosodic phrase boundary is associated with the edge of a syntactic constituent (Selkirk, 1984). Sentence comprehension studies have demonstrated that the presence of an intonation phrase boundary (but not an intermediate phrase boundary) triggers the integration of incomplete semantic and pragmatic processes (Schafer, 1997; Welby, 2001). However, speech style may be a more important factor than speech rate, since more alveolar contact could be found in the data when the speakers in the EPG studies spoke quickly, but more carefully (Nolan, 1992).

Individual differences have also been shown to play a role in the realization of English coronal place assimilation. The speaking rate effects on assimilation could be different from individual to individual (Byrd and Tan, 1996; Fougeron and Keating, 1997).

Finally, the place of articulation of the consonant after the coronal has an effect on coronal place assimilation. The coronal-dorsal sequence was found to overlap at a higher degree than the coronal-labial sequence in a movement-tracking study (Chitoran et al., 2002).
To summarize, most previous studies have found the production of English coronal assimilation to be a gradient process. As such, different degrees of articulatory overlap in coronal assimilation create different degrees of ambiguity in listeners’ perception of relevant segments. The degree of assimilation present in natural speech is due to a variety of interacting factors, including speech rate, the presence and type of prosodic boundary, in the assimilation context, speech style, and speaker-level individual preferences. In order to produce representative stimuli for the perception experiments in this research, a phonetically savvy psycholinguist was asked to produce continuous speech at different speech rates, with three different boundary types inserted between the word that ended with a coronal and the following word, in a semi-formal speech style.

1.2.2 Putonghua Tone2 Sandhi

Suprasegmental assimilation can be found in tonal languages such as Putonghua, or Standard Modern Mandarin (a tone language spoken in P.R. China).

Tones in Putonghua

Putonghua has 4 phonologically distinctive lexical tones: “high-level” (also called Tone1), “high-rising” (Tone2), “low-dipping” (Tone3), “high-falling” (Tone4) (Chao, 1948, 1968). They are also notated as numeric strings using a 5 point scale where 5 is the highest and 1 the lowest relative pitch. On this system, Tone1 is transcribed as “55”, Tone2 “35”, Tone3 “214”, and Tone4 “51” (Chao, 1948). There is also a reduced tone (also called neutral tone), where the usual tone is dropped.
Tone Sandhi in Putonghua

Tone shapes in Putonghua are often drastically altered in continuous speech as compared to citation form. Pronounced tone shapes are influenced by many different factors, including tonal coarticulation, dialectal influences, focus, and sentence intonation (Chen, 2000; Shih, 2007). Tone sandhi is a general term that covers all kinds of connected-speech-related tone changes. In this study, the narrow meaning of tone sandhi is used, which is the tone-shape change caused by the juxtaposition of a tone with the tonal patterns of the surrounding context syllable(s) (Wang). This phenomenon frequently leads to lexical ambiguities in the language. In this study, the changed tone will be called the “sandhi tone”, and written toneTs.

A well-studied Putonghua tone sandhi type is the Tone3 sandhi. This is the case where a Tone3 (low-dipping 214) becomes a Tone2 (mid-rising 35) when it occurs before another Tone3. The Tone3 Sandhi rule could be expressed as:

\[ \text{Tone3} \rightarrow \text{Tone2/}_\text{Tone3} \]

For example, the underlying Tone3 word yu3 (rain) has a dipping tone as shown in Figure 1.1. When the Third Tone Sandhi rule applies, this tone will be changed to a rising Tone2 (sandhi Tone2, or, Tone2s) as indicated in Figure 1.2, which is very similar to a real Tone2, as shown in Figure 1.3. This happens in the example sequence yu2s hen3 (rain very) as shown in Figure 1.4, which is perceptually very similar to yu2 hen3 (fish very), thus leading to a tone-dependent lexical ambiguity between yu2s(rain) and yu2(fish) when the following syllable carries Tone3.
Figure 1.1: Wave form and F0 contour for the syllable yu3(rain)

Figure 1.2: Wave form and F0 contour for the syllable yu2(sandhi) (rain)
Figure 1.3: Wave form and F0 contour for the syllable yu2(fish)

Figure 1.4: Wave form and F0 contour for the syllable sequence yu2(sandhi)hen3(fish very)
Empirical Studies on Tone Sandhi

Many studies of this type of suprasegmental ambiguity have shown that in general, listeners cannot distinguish between Tone2 and Tone2s in identification tasks (e.g. Peng (2000); Speer and Shi (1989); but see also Xu and Speer (2004)), so that additional contextual information (syntactic, semantic or pragmatic) is needed in order for listeners to identify the intended word. However, production studies have demonstrated that Tone2s tends to have a shorter duration and a shallower pitch contour than citation Tone2, Peng (2000); Xu and Speer (2004). Therefore, it is interesting to study whether listeners are sensitive to the subtle difference, although they might not realize this consciously.

In addition, empirical studies have shown that the prosodic context may change the strength or even block the application of the Tone3 sandhi rule. As the prosodic break size following the Tone2 increases, listeners are less likely to identify a spoken Tone2 Tone3 sequence as containing a sandhi tone. For example, Speer and Shi (1989) found that when the Tone2s and its following context Tone3 were separated by a intonation phrase boundary, listeners (mostly Taiwanese) in an identification task seldom thought that the sandhi rule had been applied in production. In contrast, when there was only a word boundary between the sandhi Tone2s and its following context Tone3, listeners often thought sandhi had been applied.

Although these previous studies on Tone3 sandhi give us a good basis on which to study how context might influence listeners’ processing of suprasegmental yet lexical ambiguities, Tone3 sandhi is not the focus of this study. This is because it is usually considered to be a dissimilation process (a phonological process that involves one of two similar or identical sounds within a word becoming less like each other or even
disappearing entirely. An example is the “mission of n in government (Dictionaries)).

Tone3 sandhi dissimilation is commonly motivated by the articulatory effort and time necessary to produce two consecutive dipping tones. In addition, for many speakers, Tone3 sandhi may be a conscious process. In some areas of China, such as Hunan Province, children were explicitly taught this rule so that they could consciously apply it when speaking Putonghua. In order to have maximal comparability with English coronal assimilation, this study focuses on Tone2 sandhi, which has been considered a tonal assimilation process, and which takes place naturally and in many cases unconsciously in native speakers’ Putonghua (Chao, 1968; Hyman, 1975; Shih, 2007).

Putonghua Tone2 sandhi is the case in which a Tone2 (mid-rising 35) becomes a Tone1 (high-level 55) when it is preceded by a Tone1 or Tone2 and followed by a tone other than a neutral tone. The rule can be written as followed:

\[ \text{Tone2} \rightarrow \text{Tone1 / Tone1} \mid 2 \_ \text{Tone}^* \quad (\text{=}1,2,3,4) \]

However, Tone2 sandhi does not take place each time a Tone2 word follows a Tone1 or Tone2 word. This is because the speakers also need to consider the clarity of their speech for the listeners. Previous studies have found that Tone2 sandhi most frequently occurs in prosodically weak positions (Chao, 1968; Chen, 2000; Shih, 2007). To be more specific, it mostly applies on word-medial syllables in casual, fast conversation-style speech, where the pitch-drop at the beginning part of the Tone2 does not have sufficient time to reach the low target.
The raising of the F0 at the beginning of the tone that undergoes Tone2 sandhi is often described as a result of the assimilation of the low tone to the two high targets on either side of it. That is, since the F0 value at the beginning of a canonical Tone2 is much lower than both the F0 at the end of it and the F0 at the end of the preceding Tone1 word, the pitch drop necessary to articulate the low is elided. From the perspective of articulation, F0 changes as a consequence of articulator movement. Following the law of inertia, the most effortless movement of an articulator is when it continues in the same direction. In comparison, actions such as making turns, stopping, or accelerating all require more effort (Lindblom, 1963; Nelson, 1983; Shih, 2007). In the temporal realm, the shorter the time, the more effort the speaker needs to make the same degree of turn and reach the same target heights. If time is short, and the speaker would not make special effort to increase velocity to compensate for the inadequate time, short duration would lead to short articulatory distance. Therefore, ease of articulation motivates speakers to make fewer “turns” when producing the initial low target in the Tone2 word in a relatively shorter period of time, and this results in more accommodation of the low tone to the high targets that surround it.

With this phonetic implementation model, it follows logically that Tone2 sandhi, similar to English Coronal Assimilation, should be gradient, rather than categorical. That is, the amount of raising in the early portion of the sandhi Tone2 may be smaller as speech slows or becomes more precisely articulated.

The above mentioned model of Tone2 sandhi was supported by evidence from an empirical study conducted by Shih (2007). In this experiment, four native Mandarin speakers’ reading and conversation in various styles ranging from formal to casual
were recorded when they produced “Tone1-Tone2-Tone1” and “Tone1-Tone1-Tone1” sequences that were embedded in sentence frames. The drop in pitch (the difference between the highest F0 at the beginning of the word and the lowest F0 of the word) for both the Tone1 and Tone2 words in the middle of the tri-syllable sequence was measured and plotted. The result showed that only extreme cases of altered Tone2 fell into the acoustic space of Tone1. Gradient effects were found for all Tone2 words. The data also showed that there was a strong positive correlation between the extent of Tone2 pitch drop and syllable duration. This was interpreted as a result of speakers’ not making effort to “speed up” their articulatory movements at weak prosodic positions when duration was short.

On the whole, the results of this experiment suggested that Putonghua Tone2 sandhi is a reduction process that takes place in prosodically weak (word medial, casual style) positions, and it usually correlates with short syllable duration.

The gradient account is supported by the fact that some changed Tone2s were more “strongly assimilated” and fell into the acoustic space of Tone1, while some did not. In addition, the production results showed no bimodal distribution (a distribution where all tones were either Tone1 or Tone2, with no tokens in the middle ground). Thus, one can suspect that Tone2 sandhi is gradient, similar to English coronal assimilation. This gradient nature is related to the shortened syllable duration and strength of production gesture (reduction).
1.2.3 Implication for the current study

Since Putonghua Tone2 sandhi is gradient and usually takes place in fast casual speech, it is comparable to English coronal assimilation in that the prosodic context may change the strength or even block the application of the phonological process. As an intervening prosodic boundary becomes temporally longer, it allows a longer time for the completion of the articulatory movement. This means that as “turns, decelerations and accelerations” are articulated during a prosodic boundary, speakers are more likely to produce the canonic shape of a Tone2, and thus the application of sandhi is more likely to be blocked. To be more specific, Tone2 sandhi is unlikely to take place across an Intonation Phrase Boundary, and is less likely to happen across an Intermediate Phrase Boundary than a Word Boundary. Therefore, while “shou1 yin2 ji1” (cash register) is likely to be heard as “shou1 yin1 ji1” (radio), “shou1 yin2 ji1 qi4” (cash register) is less likely to be confused with “shou1 yin1 ji1 qi4” (radio) (“qi4” means “machine”, and it does not change the meaning of the word). This is because the following context word “qi4” makes it possible to form the phrase in two feet as “shou1 yin2” and “ji1 qi4,” with an Intermediate Phrase Boundary in between the critical word “yin2” and the following word “ji1.”

Tone2 sandhi can also result in lexical ambiguity in a similar way to the English coronal assimilation and Putonghua Tone3 sandhi. For instance, the underlyingly tone2 word “yin2” (silver) has a high-rising tone. When the sandhi rule applies, this tone is changed to a high-level surface tone1. The resulting sequence “shou1 yin1s ji1” (cash register), is perceptually very similar to the sequence “shou1 yin1 ji1 (radio)”. The lexical ambiguity caused by this phonological process also needs to be resolved.
by further contextual information, if there is no sufficient phonetic difference between the changed sound and the sound it is changed to. This is also similar to the English coronal place assimilation and the Putonghua third tone sandhi rule.

These similarities between Putonghua Tone2 sandhi and English Coronal Assimilation allow us to study both using a similar experimental design. That is, we can observe and analyze listeners’ processing of the key words that undergo the phonological process in the context of various boundary types. In addition, we will compare the processing of the phonologically changed words to their counterparts, namely, the real tone1 word and the non-coronal words that may be confused with the coronal ones.

Since Tone2 sandhi usually takes place when word duration is very short, in order to elicit it, the speaker in the production study was instructed to speak very rapidly. Also, the speaker needed to be instructed to produce the words in a casual style so that the Tone2 word in the sandhi environment would be in a prosodically weak position.

In terms of perception, we would expect listeners to perceive different levels of ambiguities when they hear Tone1 words in the context of different prosodic boundary types, consistent with the gradient nature of Tone2 sandhi.

1.3 Prosodic structures in English and Putonghua

The prosodic hierarchy can be different for different languages (Beckman, 1996). A general prosodic hierarchy proposed by Nespor and Vogel (1982, 1986) has seven phonological constituents, or domains: the syllable, the foot, the phonological word,
the clitic group, the phonological phrase, the intonational phrase, and the phonological utterance. The intonational boundaries are the junctures at domain edges. Linguistic theoretical characterizations of prosodic structure have mainly focused on how many levels of prosodic phrasing there should be in a prosodic hierarchy, the interaction between different prosodic levels, and the correspondence between prosodic phrasing and other types of linguistic constituency (word boundaries, syntactic boundaries, semantic units). Prosodic phrasing helps speakers to group information into meaningful units during production, and assists listeners in processing the information more efficiently during comprehension by providing them salient temporal and pitch cues at phrase edges (Byrd and Saltzman, 2003).

In the following sections, I will summarize relevant current views on English and Putonghua intonational phrasing systems, and discuss how boundary effects have been shown to influence the operation of English coronal place assimilation and Putonghua Tone3 sandhi application respectively.

1.3.1 Prosodic Structure of English

The intonational boundary system of English can be described by The MAE_ToBI — Mainstream American English Tones and Break Indices, a system for transcribing intonation patterns and other aspects of the prosody of English utterances (Beckman and Hirschberg, 1994). For Midwestern American English, the current English-ToBI system distinguishes five types of intonational boundaries in English:
0 – the value for cases of clear phonetic marks of “clitic groups”, prosodically grouped, compound-word-like constitutes formed during connected speech. For example, the flapping of word-final /t/ and /d/ before a following vowel-initial word as in “got it”, and the medial affricate in contractions of “did you”.

1 – the “default” break index for most phrase-medial word boundaries.

2 – when there is a mismatch between constituency prescribed by the tonal transcription and the sense of disjuncture due to pauses and pause-like phenomena

3 – intermediate (intonation) phrases (ip) (aligned with the L- or H- phrase accent).

4 – (full) intonation phrases (IP)(aligned with the L% or H% boundary tone).

We have already discussed in Section 1.2, the possibility of intonational boundary effects on English coronal assimilation. For the purposes of this dissertation, boundary types 1 (phrase-medial word boundary, or “Word Boundary”), 3 (intermediate phrase boundary, or “ip Boundary”), and 4 (intonation phrase boundary, or “IP Boundary”) will be inserted in between the critical word and its context word. This will allow for the creation of stimuli with different levels of coronal assimilation. In addition, it will allow the phrasal boundaries in the English stimuli to be maximally comparable with those in the Putonghua stimuli- (to be discussed in the next section).

1.3.2 Prosodic Structure of Putonghua

Different proposals on the Putonghua prosodic hierarchy (Beattie, 1985; Shih, 1988; Chen, 2000; Chu and Qian, 2001; Fon, 2002; Peng, 2000; Peng et al., 2007; Wang, 2003) have been posited. The primary disagreements between them concern
the difficulty in determining the lower level rhythmic units, to be more specific, the
definition of “foot” in Chinese languages, and (consequently) the Tone3 sandhi do-
main in Putonghua.

Beattie (1985) proposed an eight-level prosodic hierarchy for Putonghua based on
Nespor and Vogel (1982, 1986)’s model and the notion of stress foot. Levels included
Utterance, Intonation Phrase, two levels of Phonological Phrase, two levels of Word,
the stress foot, and the syllable.

In contrast, Chen (2000) argued that there is no convincing evidence for any
prosodic units in between the MRU (the minimal rhythmic unit, which could be
thought of as a tone sandhi domain within which tone sandhi usually takes place )and the IP. He used evidence from the operation of the tone sandhi process as a
criterion to analyze Chinese prosodic structure, and offered the following example to
argue that the prosodic phrase is not part of the hierarchy in Chinese-

\[
\text{Keep dog guard gate "keep a dog to guard the gate"}
\]
\[
\text{Yang3 gou3 shou3 men2 --- > yang3 % gou2 shou3 men2}
\]

If intonation phrases were built from prosodic phrases, then the intonation phrase
boundary in this phrase should locate only in between the two prosodic phrases (yang3
gou3) and (shou3 men2). However, tone sandhi operates among (gou3 shou3 men2),
leaving the intonation phrase boundary in between the first syllable (yang3) and the
rest of the phrase (gou3 shou3 men2), hence cutting a prosodic phrase in the middle.
However, although native speakers of Chinese might have tacit knowledge of this
MRU, since they produce sandhi tone without effort, the linguistic definition of MRU is vague. Further study is needed to better define the phonetic/phonological nature of the traditional Chinese versification.

Chu and Qian (2001) omitted the lower levels such as syllable and foot, proposing a simplified three tier prosodic hierarchy comprised of prosodic word, intermediate phrase, and intonational phrase to improve the quality of Mandarin text to speech systems. Although it serves its purpose well in application, this proposal lacks sufficient detail for a complete theoretical analysis of Putonghua prosodic boundaries.

For the purposes of this dissertation, prosodic phrasing will be characterized using Pan Mandarin ToBI. Similar to MAE_ToBI, the Pan Mandarin ToBI (Peng et al., 2007), was developed based on the AS system by Tseng and Chou (1999), and described prosodic structure with a hierarchy of break indices. It defined the following boundary types in Mandarin:

- **B0** reduced syllable boundary boundaries that have been “deleted” when a syllable is extremely reduced
- **B1** normal syllable boundary the “default” boundary between syllables
- **B2** minor phrase boundary when no pause could be perceived
- **B3** major phrase boundary when pause could be perceived
- **B4** breath group boundary Marked by a pitch reset between sentences/phrases.
- **B5** prosodic group boundary Accompanied by a prolonged pause.

The disadvantage of this boundary system is that some differences between certain levels are non-categorical and prone to inter-transcriber discrepancies. For example, in Peng et al. (2007); Tseng and Chou (1999)’s study, because of the gradient nature
of the pitch reset and the pitch lowering before a boundary, 21% of the breaks were labeled differently by the transcribers for B4 and B5, and 9% for B2 and B3, even after exchanging notes and re-transcribing. (Fon, 2002) noted that there are too many levels in the Pan Mandarin ToBI. It was proposed that B4 and B5 be collapsed into one boundary type in the future, while keeping major versus minor phrases for the study of tone sandhi domains. However, better criteria need to be found to distinguish B2 and B3.

There are several advantages of the Pan Mandarin ToBI system. Although it has no clear description of acoustic cues for different boundary types, it is based on the phonetic differences between prosodic units, rather than ambiguous terms (e.g. foot) and syntactic knowledge. It is not clear what the relation between minor phrase boundary and prosodic phrase or MRU is. However, since minor phrase is defined so as to describe the prosodic grouping in natural speech, it is consistent with the MRU, or prosodic phrase, whichever correctly portrays the rhythmic grouping in Putonghua.

Another advantage of the Pan Mandarin ToBI system is its similarity to the ToBI system for American English, which makes it particularly appropriate for use in this dissertation. Three types of prosodic boundaries comparable to the ones chosen for our English coronal assimilation studies were chosen to be inserted between the critical word and the following context word in the Putonghua studies. They were: B1-normal syllable boundary (call Word Boundary in this study), which is comparable to B1 in MAE, B2 - minor phrase boundary (called ip Boundary later on), which is comparable to B3 in MAE, and B4- breath group boundary (called IP Boundary in this study), which is similar to B4 in MAE.
1.3.3 The effect of prosodic boundaries on the perception of changed sounds

There is no previous direct study of the effect of the presence of a prosodic boundary between a coronal and its following context segment in English. However, we can use what we know about the influence of prosodic boundaries on the production of English coronal assimilation (detailed in section 1.2.1) to hypothesize the pattern of such perceptual effects. Similarly, there have been no direct studies of the influence of prosodic boundaries on either the production or the perception of Putonghua Tone2 sandhi. However, there have been several studies of boundary effects on Putonghua Tone3 sandhi that may be relevant to the operation of Tone2 sandhi. The following section provides a brief discussion of the possible influence of prosodic boundaries on the perception of English coronal assimilation, a brief review of previous studies of boundary effects on the operation of Putonghua Tone3 sandhi, and a brief discussion of the possible influence of prosodic boundaries on the perception of Putonghua Tone2 sandhi.

If listeners’ perception of coronal assimilation is sensitive to the presence and strength of prosodic boundaries between the segments that participate in the assimilation process, it is predicted that:

As the intonational boundary between a coronal consonant and the consonant that follows it gets stronger, listeners’ ability to recover the coronal place also increase because there is less assimilation in those cases.
In addition, experience with the perception of assimilation could have an impact on memory and lexical representation. As the boundary size increases, the likelihood of assimilation decreases, and listeners exposed to an ambiguous (assimilated) signal should be less likely to perceive it as a coronal.

According to these hypotheses, the degree of English coronal place assimilation should be:

\[ \text{prosodic word boundary} > \text{intermediate phrase boundary} > \text{intonational phrase boundary} \]

In ToBI notation, it should be:

\[ \text{break level 1} > \text{break level 3} > \text{break level 4}. \]

The listeners’ ability to recover the coronal should be:

\[ \text{prosodic word boundary} < \text{intermediate phrase boundary} < \text{intonational phrase boundary} \]

Speer and Shi (1989) employed Shih (1988)’s model and carried out two experiments to investigate how prosodic boundary information was used by Guoyu (standard Mandarin spoken in Taiwan) listeners to determine whether sandhi had been
applied. In the first experiment, participants listened to surface tone2 tone3 sequences embedded in phrasal context in three conditions and marked tones for the phrase in an answer sheet.

The results showed that participants consistently identified the underlying tone (Tone3 was identified when Tone2 was heard) in sandhi context. Even when the tones were heard in nonsense strings, 51% of the time listeners reported hearing tone3 instead tone2. Importantly, they did not confuse surface tone2 with tone3 in non-sandhi contexts.

In the second experiment, participants heard ambiguous surface tone2 tone3 sequences (namely, the first syllable in the sequence formed a word with either Tone2 or Tone3) embedded in sentence/phrase context. Four conditions varied the strength of a prosodic boundary inserted between surface tone2 and tone3. Listeners chose the most plausible English translation for the sentence in an answer sheet.

Results showed that listeners used the prosodic boundary information between the surface tone2 and tone3 to decide whether sandhi had applied. As the boundary strength increased, they were more unlikely to predict sandhi had applied.

Wang (2003) used the same prosodic hierarchy as Speer and Shi (1989), but replaced the foot with prosodic word, and conducted a production experiment in which four subjects read sandhi phrases at different rates. She compared sandhi applications in different prosodic domains and concluded that tone sandhi obligatorily applies within a prosodic word, but across breaks above the level of the intonation phrase, tone sandhi does not occur.
The empirical studies (Speer and Shi, 1989; Wang, 2003) showed clear boundary effects on Putonghua third tone operation. Speer and Shi (1989) provided fundamental empirical evidence on the perception of sandhi in relation to prosodic boundaries. More specifically, listeners used different strengths of boundary cues to determine whether Tone3 sandhi had applied. Therefore, we can expect the participants in the current study to use different strengths of prosodic boundary cues to when determining Tone2 sandhi application as well.

The English Word Boundary is similar to B2 of the Pan-Mandarin ToBI, the ip Boundary roughly corresponds to B3 in the Mandarin ToBI, and the IP Boundary matches B5 in the Mandarin ToBI.

1.4 Processing models

The focus of this study was to find out how native speakers process ambiguous words caused by phonetic context in connected speech, to be more specific, words that undergo variation because of assimilation processes.

This section discusses three major processing accounts for assimilation in the literature and then, I propose a new one. The three existing major accounts for assimilation are: the phonological inference account (Gaskell and Marslen-Wilson, 1996, 1998; Gaskell, 2003), the feature cue parsing account (Gow, 2002, 2003; Gow and Im, 2004; McMurray and Gow, 2004) and the perceptual-integration account (Mitterer and Blomert, 2003; Mitterer et al.).
1.4.1 The phonological inference account

The *phonological inference account* proposes that listeners process changed sounds by using language-specific mechanisms that develop on the basis of the frequency with which a listener has encountered instances of the relevant context. The model accounts for the gradient nature of assimilation by suggesting that listeners perceive and evaluate the level of assimilation. For weaker assimilation cases (from 0% to 80% assimilation), the system relies on statistics of previously encountered tokens to decide whether the segment is a coronal or not. The weaker the assimilation, the stronger the likelihood that the segment will be a coronal. For strong assimilation cases (from 80% to 100%), the system uses the context (the segment that follows the segment under consideration) to retroactively decide the “place” feature of the preceding segment. If the context is viable for coronal assimilation, e.g. a labial /b/, the system would assign a very high probability for the coronal option. Since, for example, coda /t/ has 90% probability vs only 10% for coda /p/ and /k/ together.

Because of its heavy reliance on statistic probability of the segment being a coronal, which depends on the listeners’ knowledge of the language, this account is also called a “language specific account”. Empirical support for this account comes from two lines of research. First, phoneme monitoring experiments showed that listeners responded coronal 60% of time when hearing surface labial or velar segments such as *frayp* in *frayp bearer*. The percentage of coronal response is higher for words than for nonwords. It is argued that the listeners’ percept of the form of speech is a product of a phonological inference process that recovers the underlying form of speech (Gaskell and Marslen-Wilson, 1998). Second, cross-linguistic studies show that identical stimuli led to different responses from speakers of different languages (e.g. Cutler (1986);
Cutler and van Donselaar (2001); Efremova et al. (1963); Soto-Faraco et al. (2001); Cutler and Otake (1999)). For example, some studies showed that suprasegmental information had little or no effect on lexical processing in certain languages while others showed that they were as important as segmental information in other languages.

The major problem with the phonological inference account is that it fails to specify how exactly listeners use the context to infer the intended coronal or non-coronal. As mentioned previously, when assimilation is strong, the semantic context is neutral, and the following context segment is viable for coronal assimilation, this account would infer “coronal” mistakenly when the input is non-coronal. e.g. when the intended speech was “ripe berries”, the system would incorrectly infer “ripe berries”. Evidence from Gow (2002) shows that when listeners hear a phrase such as “ripe berries”, they do not make the predicted phonological inference, but could access the potential underlying form of the word “right”, which would approximate “ripe” in the context of the labial-initial word “berries”.

For the current study, in which listeners listen to auditory sentences while choosing a matching picture from a set of 4 pictures on the computer screen (one corresponds to the target, one corresponds to the competitor, two are distracters) the phonological inference account would predict that: when listeners hear a weakly assimilated item (with IP boundary or ip boundary in between the critical word and its context word), the processing system would correctly treat the input as an assimilated item, and predict that it would be in a context that would trigger assimilation (e.g. when hearing “right”, predict it would be followed by a /b/ ). When the contextual information is confirmed, the context would facilitate processing. This means, when participants hear the assimilated word, they would start to look at the target, when they heard
the context word, the fixation proportion to the target should increase immediately. For strong assimilations, the account predict that when hearing the critical word, the system would not be able to distinguish the intended critical word (e.g. right) from its non-assimilation counterpart (e.g. ripe. However, upon hearing the context segment, the system would assign “assimilation” to the preceding item. Therefore, for the Word Boundary assimilation condition, the ambiguity should resolve right after the context segment is available, before the semantic context is available. However, for the Word Boundary non-assimilation condition, the system would also assign “assimilation” to it. That is, listeners would experience ambiguity first, and then infer the assimilated word (e.g. “right”) even when hearing the non-assimilated word (e.g. “ripe”).

1.4.2 The Feature Cue Parsing Account

Based on the notion that the speech sounds of all languages can be aligned with a group of “temporally distributed acoustic feature value cues”, Gow (2001, 2002); Gow and Im (2004) argued that listeners use feature representations unified from a group of feature cues for parsing the sound signal during speech perception. Gow (2002) found that listeners hearing tokens of potentially assimilated items, such as ripe berries, access only their surface forms, whereas listeners hearing spontaneously assimilated tokens of right berries access only their underlying forms. This is true even for strongly assimilated segments, that is, words rated 4 on a scale of 1-7, 1 and 7 being coronal and non-coronal, 4 being most ambiguous.
Gow (2002) proposed that although the particular sounds that undergo change are distinct in different languages, the cause for the change can be a similar phonological process (e.g. assimilation). Therefore, listeners might use similar processing mechanisms to resolve the ambiguity in different languages (Gow and Zoll, 2002). For example, using five phoneme monitoring tasks, Gow and Im (2004) tested both native and nonnative speakers’ performance upon hearing Hungarian voicing assimilation and Korean labial-to-velar place assimilation. They found that both native and nonnative speakers relied on acoustic characteristics of phonological modification, rather than language-specific knowledge, to recognize words. In Experiment 1, native speakers of Hungarian heard cross-spliced VC#CV sequences containing initial consonants in assimilated, voiced or unvoiced conditions (Hungarian has voicing assimilation). They showed faster detection of voiced target phonemes after segments with assimilated voicing, which is similar to the facilitation effects found in English coronal place assimilation. In Experiment 2, nonnative speakers of Hungarian performed the same task as in Experiment 1, and showed the same pattern of results as native speakers. Experiments 3 to 5 used same design and methods, but Korean stimuli with either coronal place or labial-to-velar assimilation. Again, results were the same for both native and nonnative speakers. Gow and Im claimed that listeners’ knowledge of language-specific aspects of the assimilation process did not determine how they coped with assimilation.

To use the Universal Processing Mechanism to account for his English coronal place assimilation data, Gow (2003) suggested that the final segment of right in right berries has acoustic cues that are consistent with both the coronal /t/ and the labial /p/.
On this account, listeners can use context information in the following segment /b/ to subtract the labial cues from the mixture of cues on the preceding consonant, thus inferring that the segment before /b/ was a coronal. In the first two experiments, in which context /b/ was available, although the strength of the coronal feature cues were different (stronger in Experiment 1, and weaker in Experiment 2), listeners were able to detect both the coronal and labial feature cues, and correctly associate the two contemporaneous places of articulation to the modified segment and the following segment (e.g. coronal to “right” and labial to “berry”). Thus there was no need for them to reanalyze the identity of the assimilated segment. In contrast, in the fourth experiment, where the following segment (/b/ in “berries”) was removed, the place feature of the context segment (labiality) could not find an association, but still activated lexical items that had the place feature. Therefore, listeners experienced ambiguity (both right and ripe were primed) when the following context was not available. In general, Gow and Im (2004) interpreted the result from these four experiments as supporting evidence that listeners were able to correctly identify coronal place without retroactively processing information from the following context segment, context segment. On this view, the context segment is used as a confirmation resource for the judgment of coronality in strongly assimilated cases.

The problem with the Universal Processing Mechanism is that its claim that speakers of different languages use the same mechanism to process ambiguous sounds in continuous speech is too general, and lacks support from empirical studies. Many cross-linguistic studies show that identical stimuli lead to different responses from speakers of different languages, e.g. Cutler et al. (1986); Cutler and van Donselaar (2001); Efremova et al. (1963); Soto-Faraco et al. (2001); Cutler and Otake (1999).
More importantly, it is not clear what the specific feature cues are that help listeners to detect coronal and labial features within one changed segment. If the mental representation and the lexical access processes are the same across languages, as the Universal Processing Mechanism claims, then in the case of Putonghua Tone2 sandhi, in which the Tone2 is assimilated to the preceding high tone (Tone1 or Tone2), the mental representation of Putonghua sandhi tones should have features of both the underlying tone and the context tone. During processing, listeners should be able to assign both features correctly to the sandhi form and its context when the context is available. That is, the listeners’ performance should be similar to Gow’s participants, and should show no ambiguity in this case. When context is not available, since the listeners have no mechanism to deal with the dangling feature cues that come from the context tone, they should find the sandhi tone ambiguous. However, it is difficult to determine the feature cues for the preceding tone and those for the intended tone. In addition, there is no direct evidence that the assimilated token both preserves underlying place information and provides information about the place of the subsequent segment. In addition, Gow does not directly address the lexical representation of the words with assimilated tone. If listeners are able to separate the two places encoded in the assimilated segment, and thus retrieve the original place of the assimilated segment, this implies that words are lexically represented with an abstract canonical tone.

According to this account, in the current studies on English coronal assimilation and Putonghua Tone2 sandhi, listeners should not experience ambiguity of any kind. When hearing non-coronal or Tone1 sounds, listeners should always access only the surface form, and when hearing coronal or Tone2 sounds, listeners should always
be able to use “feature cues” to distinguish the coronal place of articulation in the stimuli. Listeners should be able to distinguish even strongly assimilated cases, and recognition should take place even before the following segment is available. There should not be ambiguity in the recognition process, and prosodic boundary between the critical word and its following context word would not make any difference.

1.4.3 The perceptual-integration account

Assuming that listeners perceive assimilated sounds at an earlier stage than the previous two accounts, the perceptual-integration account argues that acoustic information in the following context “overwrites” the acoustic effects of assimilation before features or feature cues are extracted because of “perceptual contrast”. That is, the acoustic information linked to the production of the assimilated and the assimilating segments interacts during auditory processing, decreasing the saliency of the conflicting information. Therefore, the feature cues, and consequently the phonological features, extracted from an assimilated utterance right berries will be similar to the feature cues extracted from a canonical pronunciation ripe berries. This integration happens before phonological-feature extraction.

Evidence for this account comes from an experiment with Hungarian words and nonwords, in which a viable and an unviable liquid assimilation was applied. Words were presented to Hungarian and Dutch listeners in an identification task and a discrimination task. Results indicated that viably changed forms were difficult to distinguish from canonical forms, independent of experience with the assimilation rule applied in the utterances, as represented by the native language of the listener Mitterer and Blomert (2003).
Results were interpreted to indicate that auditory processing contributed to perceptual compensation for assimilation, while language experience had only a minor role to play when identification was required.

There are two major arguments against the perceptual integration account. First, Fowler et al. (2000); Speer and Xu (2004) have demonstrated that human listeners show compensation for coarticulation and Tone3 sandhi in the absence of auditory contexts. In (Speer and Xu, 2004), listeners responded differently in cross-modal priming study to sandhi vs. non-sandhi Tone2s when the following context tone was not presented. Since perceptual integration depends entirely on context, this suggests that contrast could not completely account for these effects. Second, perceptual contrast does not explain facilitory progressive effects. The formant pattern of word-initial labials generally provides relatively unambiguous place information. However, when such segments are preceded by labialized coronals, contrast should exaggerate or distort the perception of this formant pattern. Miller (2001) has shown that such exaggeration does not facilitate perception. This account did not show any gradient effect of coronal either. That is, it treated strongly assimilated sounds the same way as the weakly assimilated sounds, and indicated that listeners should not experience ambiguity when processing even strongly assimilated sounds because of the exaggerating effect of the perception system.

Mitterer et al. did not discuss the gradient nature of coronal assimilation, but it can be inferred that the closer temporally the context is to the assimilated segment (e.g. in Word Boundary conditions), the more likely the perceptual-integration would be applied, as it happens at a very early stage of processing, earlier than the
processing of phonetic features. For IP Boundary and ip Boundary conditions, context information would not be available at the very early stage, and only surface form would be processed.

For the current study, this account predicts that there should not be ambiguity when listeners hear Word Boundary stimuli. In this case, when the context is viable, the processor would always favor the assimilation option even when hearing a non-assimilated word. This account is similar to the phonological inference account in that it would make the same mistakes, and incorrectly take a non-assimilated word (e.g. “ripe”) as an assimilated word “right”. For ip and IP Boundary conditions, since the contextual information is not immediately available, listeners would not be able to use it for recognition. If an item under these conditions were perceptually ambiguous, then it would remain ambiguous.

1.4.4 A Revised Exemplar Account

These three major existing models all rely on the assumption that there is an abstract “canonical” representation in memory for each word in the language. Detailed acoustic features of any given production are not encoded in the mental representation. An alternative view of how words are stored in memory is the exemplar model (Hintzman, 1986; Goldinger, 1998). In the exemplar model, traces of each episodic speech event are stored in the mental lexicon, and an abstract category is generated based on the collection of traces similar to the input at the time of retrieval.
The advantage of this pure exemplar model was that it could account for the representation of gradual sound changes (Pierrehumbert, 2006). However, the model lacks a phonological coding level, which is necessary to explain listeners’ fast segmentation of speech stream, the lack of correlation between word form and meaning (which should otherwise become more and more correlated due to language evolution as argued in Beckman and Pierrehumbert (2003)), and children’s fast learning of words without any specific situation or voice-specific cues (Pierrehumbert, 2006). Therefore, certain level(s) of abstraction are needed in the exemplar model.

A new processing account based on the exemplar model is proposed here. In the revised exemplar account, abstract phonological categories of different levels (e.g. segment, word, lexical tones, and prosodic boundaries) also come to be stored in the mental lexicon in terms of distribution of similar traces, together with each individual episode of speech. For example, prolonged duration of a segment is a feature of a stronger prosodic boundary. At the same speech rate, the words with an IP Boundary on them have longer duration than those that carry ip Boundary, which in turn are longer than words with Word Boundary. Therefore, the duration of an incoming word during speech processing can evoke stored traces of similar duration, which can serve as a loose frame for the categorization of boundary type, and will thus facilitate retrieval by providing top-down knowledge. More specifically, each episode of speech with an IP Boundary on a word is stored in the mental lexicon, together with a more abstract category of “IP Boundary”. During the retrieval process, upon hearing a new word with an IP Boundary on it, the abstract IP Boundary category will link the IP Boundary traces in the incoming signal with all the stored episodes of words
with IP Boundary traces on them and thus facilitate processing. This account thus explains both effects consistent with statistical probabilities and those consistent with acoustic cue factors in processing.

This new account would predict that both assimilated coronal and non-coronal segments would be ambiguous, as many similar exemplars exist in the mental lexicon. Listeners should be able to distinguish weakly assimilated segments by using traces of gradient sound changes, but in the case of strongly assimilated segments, they would not be able to distinguish them. That is, for non-coronal segments followed by a Word Boundary, listeners would experience ambiguity for a short while, but the subsequent acoustic information of labial closure would help with disambiguation quickly. In the ip Boundary and IP boundary conditions, a similar process will happen. For the coronal Word Boundary condition, the segment will remain ambiguous until subsequent sentential context containing syntactic, semantic and pragmatic information becomes available to aid disambiguation. For coronal ip Boundary and IP Boundary conditions, listeners should be able to distinguish the target quickly. In the ip boundary condition in English, listeners should hear an ambiguous segmental cue (the deleted/obscured /t/ or /p/), but also the pitch information about the presence of a phrase accent. Looking at the exemplars in memory, they should see many cases of coronal with a phrase accent followed by labial in rapid speech where assimilation did not happen (there should be more of these than for a coronal with a word boundary followed by a labial). Alternatively, some stored instances may include things like a release of the /t/ or /p/ stop that can disambiguate (there should be more of these than for a coronal with an IP boundary followed by a labial).
The current study focuses on empirical investigation of the processing accounts for assimilation in continuous speech. In the following chapters, series of parallel experiments conducted for this research are presented. The first two experiments are a pair of production studies with detailed phonetic analyses for English coronal assimilation described in Chapter 2 and for Putonghua Tone2 sandhi described in Chapter 3. These are compared in Chapter 4. Then a set of listening and word-picture matching pretests for the English eye tracking study are described in Chapter 5, followed by a description of the English eye tracking study in Chapter 6. The analogous set of listening and word-picture matching pretests for Putonghua eye tracking study (Chapter 7), and the Putonghua eye tracking study (Chapter 8) follows. Finally, a comparison of the results of the two main experiments, and discussion of their indications for processing models are given in Chapter 9.

The production studies not only generate auditory stimuli for the eye tracking experiments, but also investigate the acoustic features of the interaction between prosodic boundary types (IP Boundary, ip Boundary, and Word boundary) and the operation of more localized segmental and suprasegmental processes.

The pretests evaluate the intelligibility of the auditory stimuli and the goodness of match between the target words and the pictures that are associated to them. They also elicit baseline perception data for the stimuli that will be used in the eye tracking experiments. The eye tracking experiments investigate native listeners’ online resolution of the lexical ambiguity that can arise due to the presence of changed sounds in different prosodic boundary contexts. They are intended to address our major question about the mechanism for processing lexical ambiguities that arise when words are pronounced in continuous speech. Specifying such a mechanism is critical to the
understanding of what listeners do in most typical language processing situations. The comparison of this kind of processing in two languages that differ substantially in the nature of the sound change can increase the generality of our models of the language processing system.
CHAPTER 2

PRODUCTION STUDY FOR ENGLISH CORONAL ASSIMILATION

2.1 Introduction

The main purpose of the English production study was to generate auditory stimuli for the English eye tracking experiment, in which participants listen to sentences with different levels of lexical ambiguities resulting from coronal assimilation and prosodic boundary strength. Based on these sentences, participants make decisions on which one out of a set of four pictures on the computer screen matches the best with the auditory sentence. Since the participants' eye movement patterns are directly influenced by subtle acoustic features in the auditory stimuli, in this chapter we also describe several of the more relevant acoustic features of the sentences. These include: the ToBI annotation of the boundary strength, the duration of the critical word, and the pattern of the formants of the vowel in the critical word.

The next section of this chapter describes the method used to elicit speech in the production experiment, including information about the speaker, materials and the recording procedure. Section 2.3 contains a description of the results and Section 2.4 summarizes the experimental findings.
2.2 Method

2.2.1 Speaker

A female native Midwestern American English speaking linguist with extensive ToBI experience and phonetic knowledge produced all the utterances.

2.2.2 Material

36 word pairs were selected from Gow (2002)’s stimuli. The two words (e.g. line/lime) in each pair differed only in the place of articulation of the final consonant – one ended with a coronal (either /t/ or /n/), while the other ended with a non-coronal. For the sake of convenience, the word that ended with a coronal consonant will be referred to as “the coronal” and the other word “non-coronal” later in this paper. These words were embedded in neutral context sentences adapted from Gow’s stimuli, so that both meanings were naturally plausible, e.g. “He drew the line/lime perfectly”. We have called this sentence “the ambiguous sentence”. A following context sentence was then generated for each ambiguous sentence to make it clear whether the intended word was the one ending with a coronal or not. We call this sentence “the disambiguating sentence”.

The auditory ambiguous sentence incorporated two independent variables: First, the place of articulation of the last consonant in the critical word: either non-coronal (as in examples 1, 2, 3), or coronal (examples 4, 5, 6), and second, the prosodic boundary type inserted between the critical word and the word that follows it: Word Boundary (examples 1 and 4), Intermediate Phrase Boundary (ip Boundary, examples 2 and 4), or Intonation Phrase Boundary (IP Boundary, examples 3 and 6).
Example sentences in 6 conditions:

1: He drew the lime ( ) perfectly. I almost want to pick it up and smell it.
2: He drew the lime (L-) perfectly. I almost want to pick it up and smell it.
3: He drew the lime (L-H%) perfectly. I almost want to pick it up and smell it.
4: He drew the line ( ) perfectly. I don’t know how he made it so straight.
5: He drew the line (L-) perfectly. I don’t know how he made it so straight.
6: He drew the line (L-H%) perfectly. I don’t know how he made it so straight.

The complete set of sentences is provided in Appendix A.

2.2.3 Recording

The speaker read each of the 2 versions (that is, coronal vs. noncoronal) of the 36 critical sentences in groups of the same prosodic boundaries twice. For example, she read the sentence “He drew the line ( ) perfectly. I don’t know how he made it so straight.” in “Word Boundary” condition, then read the sentence “He drew the lime ( ) perfectly. I almost want to pick it up and smell it.” also in “Word Boundary” condition. This was done to keep the speaking rate and any incidental recording conditions as similar as possible for pair members.

The speaker then repeated the above sentences. After all the sentences were read with the same prosodic boundary type, she then read all the sentences with another prosodic boundary type. This was done to keep boundary productions as similar as possible within the two boundary types.
Intermediate phrases were always pronounced with a low phrase accent, and intonation phrases were also always pronounced with a low phrase accent followed by a high boundary tone, so that the difference between these prosodic phrase edges would be unambiguous for listeners.

Digital recordings of complete auditory sentences were made in a sound-attenuated booth, sampling at 22,050 Hz with 16-bit resolution. Praat 4.0.35 was used to view and edit the audio files.

The speaker re-recorded utterances until they meet the specified conditions as needed. For example, if ToBI annotation did not verify intended boundary condition, or if phonetic measures showed drastic differences in length for coronal vs. non-coronal words in the same boundary condition, the utterance was re-recorded.

2.3 Results

2.3.1 ToBI Annotation

The detailed ToBI annotation showed that the “word” (BI=1), “ip” (BI=3) and “IP” (BI=4) boundaries were produced in the intended way 100% of the time. The difference between the IP and ip boundaries were salient in tonal contrast. For example, the Intonation Phrase Boundaries in this experiment were always produced with a low phrase accent followed by a high boundary tone (L-H%). The Intermediate Phrase Boundaries were always produced with a high phrase accent (H-). (There were no boundary tones in the Word Boundary conditions.)
### Table 2.1: Mean durations of the Critical Word, the Vowel in the Critical Word, the Interval from Critical Word Onset to Context Word Onset in each condition

<table>
<thead>
<tr>
<th>Condition</th>
<th>Critical Word</th>
<th>Vowel</th>
<th>Onset to Onset</th>
</tr>
</thead>
<tbody>
<tr>
<td>NonCoronal Word</td>
<td>189.2 ms</td>
<td>99.8 ms</td>
<td>254.7 ms</td>
</tr>
<tr>
<td>Std.Dev.</td>
<td>89 ms</td>
<td>29 ms</td>
<td>84 ms</td>
</tr>
<tr>
<td>NonCoronal ip</td>
<td>264.9 ms</td>
<td>149.1 ms</td>
<td>390.8 ms</td>
</tr>
<tr>
<td>Std.Dev.</td>
<td>93 ms</td>
<td>38 ms</td>
<td>68 ms</td>
</tr>
<tr>
<td>NonCoronal IP</td>
<td>421.4 ms</td>
<td>251.3 ms</td>
<td>543.0 ms</td>
</tr>
<tr>
<td>Std.Dev.</td>
<td>97 ms</td>
<td>41 ms</td>
<td>91 ms</td>
</tr>
<tr>
<td>Coronal Word</td>
<td>198.5 ms</td>
<td>109.5 ms</td>
<td>249.8 ms</td>
</tr>
<tr>
<td>Std.Dev.</td>
<td>60 ms</td>
<td>32 ms</td>
<td>48 ms</td>
</tr>
<tr>
<td>Coronal ip</td>
<td>314.0 ms</td>
<td>172.9 ms</td>
<td>406.2 ms</td>
</tr>
<tr>
<td>Std.Dev.</td>
<td>81 ms</td>
<td>44 ms</td>
<td>71 ms</td>
</tr>
<tr>
<td>Coronal IP</td>
<td>441.9 ms</td>
<td>259.5 ms</td>
<td>531.8 ms</td>
</tr>
<tr>
<td>Std.Dev.</td>
<td>85 ms</td>
<td>57 ms</td>
<td>100 ms</td>
</tr>
</tbody>
</table>

2.3.2 Duration analysis

Duration measurements were calculated using Praat. Points measured in the sound files include: the onset of the word ending in the target coronal or non-coronal segment, henceforth, “critical word”, onset of the vowel, offset of the vowel, onset of the coda, offset of the coda and the onset of the word that follows the critical word, henceforth, the “context word”. The durations of the critical word, the vowel in the critical word, and the duration from the onset of the critical word to the onset of the following context word were then calculated by subtracting the measurement of the previous time points from that of the later time points.
Duration of the critical word

With different boundary types, the mean length of the critical words was different as shown in Table 2.1. The mean duration of the critical words in the IP Boundary conditions was longer than those in the ip Boundary conditions \((F[1, 71] > 100)\), and also was longer in the ip Boundary conditions were also longer than those in the Word Boundary conditions \((F[1, 71] > 100)\).

The mean length of the critical words was also different when the “place” features of their last consonants were different. Planned Comparisons of Coronal vs. Non-coronal showed that words ended with a coronal consonant are in general longer than words ended with a non-coronal consonant \((F[1, 100] > 100)\). However, a more detailed planned comparison between coronals and non-coronal words across all three boundary types demonstrates that coronal words are longer than non-coronal words in the “IP Boundary” condition \((F[1, 71] = 4.25, p = 0.043)\) and the “ip Boundary” condition \((F[1, 71] = 24.2, p = 0.0001)\), but in the “Word Boundary” condition, they are at about the same length \((F(1, 73) < 1)\).

Duration of vowel in the critical word

The duration of the vowel in the critical word shows exactly the same pattern as the duration of the critical word. Main effects were found for both Boundary \((F[2, 70] > 100)\) and Place \((F[1, 35] = 8.7, p = 0.0056)\) factors. The mean length of the vowels in the critical words was longer in the IP Boundary conditions than in the ip Boundary conditions, which in turn were longer than the Word Boundary conditions \((F[1, 71] > 100)\). The vowels in the words ended with a coronal were
longer than words ended with a non-coronal in the IP Boundary ($F[1, 35] = 4.7, p = 0.032$) and ip Boundary conditions ($F[1, 35] = 17.7, p = 0.0001$), but not in the Word Boundary conditions ($F[1, 35] = 2.9, p = 0.92$).

**Duration from critical word onset to context word onset**

Table 2.1 also showed that although words with different types of place features in coda were different in length, the overall length from the beginning of the critical word till the beginning of the following context length were not statistically different for the coronal and non-coronal words. That is, when the boundary types were the same, the word-to-word intervals were about the same for coronal and non-coronal words ($F[1, 35] < 1$).

In general, the critical word, its vowel, the durations of silence between the critical word and the following context word, as well as the duration from the critical word to context word onset increased as the strength of the prosodic boundary increased. The different degrees of final lengthening and pause between words was in accordance with contrasts among BI=1, BI=3, BI=4. This is also in accordance with the different degrees of coronal assimilation. The only exception is the silence between the critical word and its following word for the non-coronal words in the ip boundary and the IP boundary conditions. This is because in the IP condition, there is usually a release of the stop while in the ip Boundary condition, there usually is no stop release.
2.3.3 Formant Analysis

Previously, it has been found that the formant movement for the assimilated items had a pattern that was intermediate between the patterns associated with unmodified coronal and non-coronal place (Gow and Hussami, 1999).

Formant Graphs

To further study the formant differences between coronal and non-coronal words, a Praat script was used to extract the F1, F2 and F3 values during the vowel of the critical words, and an R script that then reads those values in and plots them.

Following are graphs of several example word pairs with different vowels and codas under all three prosodic boundary conditions. Each word pair were aligned at the end of the vowel in the critical word and the drawing of the formants was over a 0.2 second interval before the alignment point and 0.05 second after.

Figure 2.4 shows the F1, F2 and F3 formants of the word “back” and “bat” followed by the word “grounded” in all three prosodic conditions. The black dotted lines are the formants of “back”, and the red dotted lines are the formants of “bat”. The vertical line at about -0.05s in the figure is the ending point of the vowel in the critical word.

The graphs showed that the formant pattern of the coronal word “bat” is not significantly different from the non-coronal word “back” in F1 and F2. However, in F3, only the most assimilated formant in the coronal word “bat” shows the same pattern as the non-coronal word “back”. This is the “Word Boundary” condition. However, in the ip Boundary and the IP Boundary conditions, the F3 of the coronal word “bat” rose higher than the other conditions about 0.1 second before the end
of the vowel. This suggests that when listeners heard “bat ground” in the Word Boundary condition, they are more likely to experience ambiguity. However, in ip Boundary and IP Boundary conditions, they might be able to distinguish the /t/ sound through the subtle acoustic difference in F3.

Figure 2.5 shows the formants of the coronal word “sun” (red lines) and the non-coronal word “sum” (black lines) followed by the word “proceeded”.

This graph shows that the F2 of the coronal word “sun” rose above the non-coronal word “sum” starting about 0.1 second before the end of the vowel in all prosodic boundary types when preceding a labial /p/. Therefore, when listeners hear the sequence “sun proceeded”, in all three boundary conditions, they might be able to use the subtle acoustic cues to distinguish the intended sound.

Figure 2.6 F1, F2 and F3 aligned at the end of the vowel for the critical words “comb” and “cone” in all three boundary types (Word, ip, IP). The red lines are for the coronal word “cone” and the black lines are for its non-coronal counterpart “comb”. The circle in the figure shows that the F2 of the word “cone” were higher than the word “sum” in the ip Boundary and IP boundary conditions when followed by a /b/ sound as in “between”.

The same patterns were found in another randomly selected pair of words ended with /n/ and /m/ (“cone” and “comb”) as shown in Figure 2.6, but the departure of the coronal F2 in from the non-coronal F2 came a little later in this pair.

The R script was also used to calculate a total difference over some region of a “critical formant” (defined arbitrarily as F2 for labial pairs and F3 for velar pairs). The region chosen was between the end of the vowel and a constant distance back from that alignment point, with the constant distance defined by item set as half the
length of the shortest vowel in the six utterances. The R script also makes a bar plot of the mean differences in the critical formant, for F3 when the non-coronal is velar, for F2 when the non-coronal is labial, and averaging over both for all of the items. There is support for positing more assimilation in the Word condition than for the other two, but less clear support for a difference between IP and ip. On average, the difference is 416 Hz bigger for the IP condition compared to the ip condition, and 935 Hz bigger for the ip condition compared to the Word condition, but only the second comparison was significant ($p = 0.001$) in a paired, one-tailed Wilcoxon test. There are several possibilities for the lack of difference between IP and ip conditions. First, the formant tracks were better for the oral context than for the nasal context, just as one would expect given the wider bandwidth and lower energy at higher regions when the nose is opening. Second, F2 for labial and F3 for velar may be too simplistic, and one may need to choose one or the other depending on what the vowel is, as well as on what the consonant is. In this study, only 36 tokens in each boundary types were available, and the phonetic contexts of the coronal codas among these 36 tokens were quite different. For example, the two types of coronal place /n/ and /t/ were preceded by 8 different types of vowels and followed by four different types of onsets. Therefore, we did not have enough tokens in the same phonetic context for comparisons that could lead to rigorous statistical inference.

In summary, Formant (either F2 or F3) differences between coronal and noncoronal words were found in the production data of the current study, especially in the ip Boundary and IP Boundary conditions, and this difference starts at around the middle of the vowel. Since different items have different levels of assimilations (which is inherit to natural speech), and in several items, the Word Boundary has similar
acoustic cues as the ip and IP Boundaries, listeners’ ability to detect the intended word might vary with individual items. However, since the Word Boundary conditions were more ambiguous than other two boundary types in general, as shown in the listening pretest presented in Chapter 5, there should be an overall pattern in the eye tracking data that the Word Boundary conditions are more ambiguous than the other conditions.

2.4 Summary

ToBI annotation confirmed that correct boundary types were produced. The duration and formants analyses suggest that the speaker was able to produce relatively consistent prosodic boundary types, and coronal places at different levels of assimilations.

There were significant differences between coronal and non-coronal critical words in ip and IP conditions (for example, the coronals were longer than non-coronals), but not in Word condition. However, even in Word condition, there were still subtle formant differences after the end of the vowel between coronal and non-coronal.

This suggests that listeners might be able to distinguish coronals from non-coronals with ease in ip and IP conditions, but not in Word condition. However, in Word condition, there were still subtle acoustic differences that might help listeners to discern coronal and non-coronal targets.
Figure 2.1: F1, F2 and F3 aligned at the end of the vowel for the critical words “back” and “bat” in all three boundary types (Word, ip, IP). The red lines are for the coronal word “bat” and the black lines are for its non-coronal counterpart “back”.

53
Figure 2.2: F1, F2 and F3 aligned at the end of the vowel for the critical words “sum” and “sun” in all three boundary types (Word, ip, IP). The red lines are for the coronal word “sun” and the black lines are for its non-coronal counterpart “sum”.

54
Figure 2.3: F1, F2 and F3 aligned at the end of the vowel for the critical words “comb” and “cone” in all three boundary types (Word, ip, IP). The red lines are for the coronal word “cone” and the black lines are for its non-coronal counterpart “comb”.
CHAPTER 3

PRODUCTION STUDY FOR PUTONGHUA TONE2 SANDHI

3.1 Introduction

The purpose of the Chinese Tone2 sandhi Production study is parallel to that of the English production study, to generate auditory stimuli for the main Chinese eye tracking experiment. In this chapter, the most important acoustic features in production, such as duration and F0 of the critical word in this chapter. Section 3.2 describes the method of the production experiment, including information about the speaker, materials and recording procedure. Section 3.3 contains a description of the results of this experiment, while Section 3.4 summarizes the experiment and its results.

3.2 Method

3.2.1 Participants

The requirements set for the speaker in this experiment were similar to those of the English experiment. A female native speaker of Beijing Putonghua was trained to read the sentences in the desired manner.
3.2.2 Material

34 syllable pairs composed of identical segmental strings with different tones were selected. There were several criteria in the selection of the strings. First of all, the segmental content of the syllable had to make a meaningful unit in Putonghua when combined with both Tone1 and Tone2. This was necessary because in Putonghua, some tone and segment combinations are meaningful while others are not. For example, “si1” means “silk”, but there is no “si2” in the language. Second, each syllable should be able to be preceded by a Tone1 syllable, to make a meaningful word in Putonghua. This criterion was set according to the Tone2 sandhi Rule, which is “the process whereby an underlying Tone 2 (a rising tone) surfaces, or appears to surface as a tone 1 (a high level tone) when it occurs in a weak syllable following a tone that ends in a high target” Shih (2007)(See Chapter 1, Section 1.2.2). Third, both the Tone1/2+Tone1 and Tone1/2+Tone2 sequences should be able to be followed by the same syllable to form a meaningful word or phrase in Putonghua. This is so that a tone2 sandhi environment could be formed, providing only a Word Boundary between the critical syllable (In order to be comparable to the English study and for simplification, the “critical syllable” will be called “critical word” later on) and its following context syllable (called “context word” later). Fourth, both the Tone1/2+Tone1 and Tone1/2+Tone2 sequences should be able to be followed by the same word in Putonghua. This is also so that a tone2 sandhi environment could be formed, and either an intermediate phrase boundary (ip) or an intonation phrase boundary (IP) could be inserted between the critical word and its following context word. Fifth, the Tone1/2+Tone1+Tone*(+Tone*) and the Tone1/2+Tone2+Tone*(+Tone*)
sequences should be able to be embedded into the same contextual sentences to form an ambiguous sentence, without a strong bias toward either tone. Only 34 syllable pairs were found to match all five of the above-mentioned criteria.

Since about half of the words could not be found in the entries of the XianDai HanYu PinLiü CiDian (Contemporary Chinese Word Frequency Dictionary, Wang (1990)), it was difficult to balance the frequencies of the tone1 and tone2 words, or control the homophones for each critical syllable. Fortunately, however, when preceded by the Tone1 or Tone2 syllable, and to the words turned out to be homophones with the Tone1/2+Tone1/2 sequences. Furthermore, we later evaluated listeners’ bias for tone1 or tone2 interpretation with a “listening” pretest, in order to remove any item that had a strong bias toward either Tone1 or Tone2 (See Chapter 7, Section 7.2.3).

Similar to the English experiment, for each of the 34 Tone1/Tone2 pairs, a simple context sentence was generated, in which the critical syllable was preceded by a tone1 or tone2 word, and followed by a word with one of the four lexical tones. Both meanings of the tone1 and the tone2 words were naturally plausible in each of these simple sentences. A second simple sentence was then constructed for both the Tone1 and the Tone2 syllable pairs, in which, at the sentence end, the ambiguity of the previous sentence was resolved (See the example sentence). The context sentences varied in semantic and syntactic predictability. A complete set of the sentences is provided in Appendix B.
Sample sentence:

听 到 那 京 胡 声, 他 皱 了 皱
Hear that instrument sound, he frowned particle frown
眉。
mei2。
eyebrow.

He frowned when hearing that sound of the instrument “jinghu”.

听 到 那 惊 吼 声, 他 皱 了 皱
Hear that scream sound, he frowned particle frown
眉。
mei2。
eyebrow.

He frowned when hearing that scream.

听 到 那 京 胡 声 音 剌 耳, 他 皱 了 皱 眉。
Hear that instrument sound pierce ear, he frowned particle frown
眉。
zhou4 le0 zhou4 mei2。
eyebrow.

He frowned when hearing that piercing sound of the instrument jinghu.

59
听那惊呼声刺耳，他
Hear that scream sound pierce ear, he

皱了皱眉。

frown particle frown eyebrow.

He frowned when hearing that piercing scream. Disambiguating part:

For **tone1** interpretation: and he knows his son is bullying his sister again.

知道儿子又在欺负妹妹
Know son again is bully sister

了。

le0. particle

Knowing that his son is bullying his sister again. For **tone2** interpretation: and he knows his son is practicing playing the jinghu (instrument) again.

知道儿子又在偷偷练
Know son again is secretly practice

琴了。

qin2 le0.

instrument particle

Knowing that his son is practicing the instrument secretly again.
3.2.3 Recording

Similar to the English experiment, the speaker in this experiment read each of the two versions, that is, Tone1 vs. Tone2, of the 34 critical sentences in groups of the same prosodic boundaries twice. For example, she read the sentence “ting1 dao4 na1 jing1 hul sheng1 yin1 ci4 er3, ta1 zhou4 le0 zhou4 mei2, zhi1 dao4 er2 zi0 you4 zai4 qi1 fu4 mei4 mei4 le0” in “Word Boundary” condition, and then she read the sentence “ting1 dao4 na4 jing1 hul sheng1 yin1 ci4 er3, ta1 zhou4 le0 zhou4 mei2, zhi1 dao4 er2 zi0 you4 zai4 tou1 tou1 lian4 qin2 le0” also in “Word boundary” condition. She then repeated the same sentence before continuing on to the second set of sentences. After all the sentences were read with the same boundary type, she then read all the sentences again with the other boundary type. The speaker was asked to use a contrastive stress on the critical word in order to induce the IP Boundary because for most of the stimuli sentences, a pause in between the critical word and its following context word would yield unnatural and even ungrammatical results.

Digital recordings of complete auditory sentences were made in a sound-attenuated booth, sampling at 22,050 Hz with 16-bit resolution. The files were saved in .wav format to a PC. Praat 4.0.35 was used to view and edit the audio files.

The speaker and the experimenter verified that the utterances were produced with the intended tones and prosodic boundary types and the speaker re-recorded utterances as needed. However, the speaker had difficulty producing sandhi Tone1 for several items even in Word Boundary conditions. This was noted as a phenomenon in which certain items were not sandhi-pro, and were unnatural to be produced, at least for this particular speaker.
3.3 Results

3.3.1 ToBI Annotation

The detailed ToBI annotation showed that the “word” (BI=1), “ip” (BI=2) and “IP” (BI=4) boundaries were produced in the intended way 100% of the time.

It also showed that in IP condition, Tone2 critical words sounds like Tone2 100% of the time; in ip condition, 29 out of the 34 items sounds like Tone2; in Word condition, 11 out of the 34 items sounds like Tone2. This is to say, certain Tone2 items still sounds like Tone2 in Word condition.

3.3.2 Duration Analysis

The durations of the critical word, of the vowel in the critical word, and of the interval from the beginning of the critical word to the beginning of the following context word were measured in the same manner as in the English experiment.

Duration of Critical Word

As demonstrated in Figure 3.1, the duration of the critical word in this experiment shows the same pattern as in the English auditory stimuli - main effects were found for both boundary and tonal factors. Namely, critical words in the IP conditions are longer than those in the ip conditions, which in turn, are longer than words in the Word Boundary conditions ($F[1,33] > 100$). Also, the Tone2 words were longer than the Tone1 words in the IP conditions ($F[1,33] = 16.5,\ p = 0.0001$) and the ip condition ($F[1,33] = 4.5,\ p = 0.038$), but not significantly different in the Word Boundary condition ($F[1,33] < 1$).
Figure 3.1: Mean duration of critical word in Putonghua study

*Duration of vowel in the critical word*

Main effects for both the boundary type ($F[2, 32] = 95.8, p = 0.0001$) and tone type ($F[1, 33] = 4.2, p = 0.047$) were found to be significant in the vowel part of the critical words, as shown in Figure 3.2. As one would expect, the vowels in the IP Boundary conditions were longer than those in the ip Boundary conditions ($F[1, 33] > 100$), and vowels in the ip Boundary conditions were longer than those in the Word Boundary conditions ($F[1, 33] > 100$). However, vowels in Tone2 words were found to be longer than vowels in Tone1 words only in the IP Boundary condition ($F[1, 33] = 6.96, p = 0.01$), and not in the ip Boundary and Word Boundary conditions ($F[1, 33] < 1$).
Figure 3.2: Mean duration of vowel in critical word in Putonghua study

*Duration from critical word onset to context word onset*

The mean interval from the onset of the critical word to the onset of the following context word was also different from the English experiment. Main effects were found for the boundary factor, namely that the mean word-to-word duration in the IP Boundary condition was longer than in the ip Boundary condition ($F[2, 32] > 100$), while the ip Boundary condition was longer than the Word Boundary condition ($F[1, 33] > 100$). There was no main effect for the tone factor. In the IP Boundary condition, Tone2 words were longer than Tone1 words ($F[1, 33] = 9.6, p = 0.02$), while in both the ip Boundary condition and the Word Boundary condition, there was no Tone1/Tone2 difference ($F[1, 33] < 1$).
3.3.3 Tone height analysis

F0 of each critical word was measured at 3 points: beginning, middle, end.

Word Initial Position

A two-way ANOVA (boundary type and tone) yields main effects for both the Boundary Type factor and Tone Type factor ($F[2, 64] = 36.9, p = 0.0001$). F0 value at the initial part of the word in the IP Boundary condition was higher than the ip Boundary conditions ($F[1, 33] = 8.9, p = 0.004$), while F0 value at the initial position in the ip Boundary condition was higher than the F0 value in the Word Boundary conditions ($F[1, 33] = 58.5, p = 0.0001$).

Detailed planned comparison shows that at the initial position, Tone2 was lower than Tone1 in the IP Boundary ($F[1, 33] > 100$) and ip Boundary ($F[1, 33] = 32.8, p = 0.0001$) conditions, but not in the Word Boundary condition ($F < 1$).
Figure 3.4: Mean word initial F0 in Putonghua study

**Word Medium and Final F0**

This pattern was repeated at the word-median and word-end locations as shown in Figure 3.5 and Figure 3.6. Both Boundary Type and Tone Type factors were significant in word medium position ($F[2,64] = 22.6, p = 0.0001$) for Boundary type, ($F[2,64] > 100$) for Tone type; and in word final position ($F[2,64] = 89.6, p = 0.0001$) for Boundary type, ($F[2,64] = 60.3, p = 0.0001$) for Tone type). In the IP Boundary conditions, the F0 was always lower than in the ip Boundary conditions ($F[1,33] = 12.9, p = 0.0006$) (Word Medial) and ($F[1,33] = 93.6, p = 0.0001$) (Word Final)), and in the ip Boundary conditions it was always lower than in the Word Boundary conditions ($F[1,33] = 35.9, p = 0.0001$) (Word Medial) and ($F[1,33] = 78.6, p = 0.0001$)).
Figure 3.5: Mean word medium F0 in Putonghua study

Tone2 was lower than Tone1 in the IP Boundary \((F[1, 33] > 100)\) and ip Boundary \((F[1, 33] = 79.2, p = 0.0001)\) conditions, but not in the Word Boundary condition \((F[1, 33] < 1)\).

Item by item graphic analyses

The above “Tone height analyses” provided statistics at different points (beginning, medium and final position) of the critical word. In order to investigate the entire tone shape of the critical words, and the possible influence of the preceding and following tones, item-by-item plots were made so that we can compare the differences between the surrounding tone patterns as well as the differences on the critical syllable itself. There were 3 panels in each plot. Each panel overlays the F0 contour for the utterance containing the (underlying) Tone2 target syllable on top of the F0 contour for the utterance containing the Tone1 target syllable for the same prosodic boundary
condition, aligning the two contours at the end of the critical word. Following are the graphs grouped according to their tonal environment:

The graphs for “Tone1 CriticalWord Tone1” sequences showed that Tone2 and Tone1 critical syllables had very different F0 values in IP and ip conditions, but the tone shapes were very similar in Word conditions. However, there were still very subtle differences between the two. More interestingly, the preceding Tone1 has slightly higher F0 value for the Tone2 critical syllable than for the Tone1 critical syllable. This pattern was repeated in the graphs for “Tone1 CriticalWord Tone2-4” sequences.

The graphs for “Tone2 CriticalWord Tone2-4” sequences showed that when the syllable that precedes the critical syllables were Tone1, there is a common pattern that the Tone2 and Tone1 critical syllables had very different F0 values in IP and ip conditions, but the difference was subtle in Word conditions. In most cases, the preceding Tone1 has slightly higher F0 value for the Tone2 critical syllable than for the
Figure 3.7: “Tone1 CriticalWord Tone1” sequence, F0 aligned at the end of the critical syllables “shi” “hu”, “hu”, and “yin” in all three boundary types (Word, ip, IP). The red lines are for the Tone2 word and the black lines are for its Tone1 counterpart.
Tone1 critical syllable. This is probably because of the tonal “anticipatory effects”, which is usually a relatively small effect, in which a low onset value of a tone (here, Tone2) raises the maximum f0 value of a preceding tone (Xu, 1997).

The graphs for “Tone2 CriticalWord Tone*” sequences showed that when the syllable that precedes the critical syllables were Tone2, there is also a common pattern that the Tone2 and Tone1 critical syllables had very different F0 values in IP and ip conditions, but the difference was subtle in Word conditions. In most cases, the preceding Tone2 has slightly higher F0 value for the Tone2 critical syllable than for the Tone1 critical syllable. This could be explained in the same way as when the critical syllables were preceded by a Tone1 word, that is, it is due to the dissimilatory anticipatory effects. Another possibility is that it is due to the observed “delayed high target” effect of Tone2 when it is followed by Tone1, that is, the end of Tone2 rarely reaches the high level when the following tone starts high (Shih, 1988; Xu, 1997). Here, the Tone2 followed by a Tone1 critical syllable is lower than the Tone2 followed by a Tone2 critical syllable because the high target of Tone2 is realized a little later when it is followed by the high tone, Tone1.

Minimum F0 Analyses

Based on the observation that there is a difference between Tone1 and Tone2 even in Word conditions, we measured the minimum F0 in the critical syllable, presuming that Tone2 syllables should have lower minimum F0 than Tone1 syllables. Since there are sometimes small differences in apparent F0 range between the utterances that we are pairing, we measured this minimum relative to some reference F0 in the preceding syllable. This reference is set as the maximum F0 if the preceding syllable is a Tone1 morpheme and the minimum F0 if the preceding syllable is a Tone2 morpheme.
Figure 3.17 showed that there is a difference between underlying Tone2 and Tone1 even in the Word condition. Although the difference is smaller there than in the ip condition. The difference is also smaller in the ip conditions than in the IP conditions. Wilcoxon tests showed that the differences in the effect size between the difference conditions are highly significant ([p = 0.000] for both IP vs. ip and ip vs. Word comparisons).

3.4 Summary

Similar to the English production study, ToBI annotation confirmed that correct boundary types were produced. The duration and tone height (F0) analysis suggest that the speaker was able to produce relatively consistent prosodic boundary types and tone types at different levels of assimilations.

There were significant differences between Tone1 and Tone2 critical words in ip and IP conditions (for example, the Tone2 words were longer than Tone1 words), but not in Word condition. However, according to ToBI annotation, even in Word condition, there were still F0-difference between Tone2 and Tone1 in certain items.

This suggests that listeners might be able to distinguish Tone2 words from Tone1 words with ease in ip and IP conditions, but not in Word condition. However, in Word condition, there were still subtle acoustic differences that might help listeners to discern Tone1 and Tone2 targets.
Figure 3.8: “Tone1 CriticalWord Tone1” sequence, F0 aligned at the end of the critical syllables “cheng”, “jie”, and “zhi” in all three boundary types (Word, ip, IP). The red lines are for the Tone2 word and the black lines are for its Tone1 counterpart.
Figure 3.9: “Tone1 CriticalWord Tone2” sequence, F0 aligned at the end of the critical syllables “chuan” “ge” , and “xian” in all three boundary types (Word, ip, IP). The red lines are for the Tone2 word and the black lines are for its Tone1 counterpart.
Figure 3.10: “Tone1 CriticalWord Tone3” sequence, F0 aligned at the end of the critical syllables “fang” “qiang” “tong” and “guo” in all three boundary types (Word, ip, IP). The red lines are for the Tone2 word and the black lines are for its Tone1 counterpart.
Figure 3.11: “Tone1 CriticalWord Tone4” sequence, F0 aligned at the end of the critical syllables “yi” “chu” , and “fa” in all three boundary types (Word, ip, IP). The red lines are for the Tone2 word and the black lines are for its Tone1 counterpart.
Figure 3.12: “Tone1 CriticalWord Tone4” sequence, F0 aligned at the end of the critical syllables “shou” and “hua” in all three boundary types (Word, ip, IP). The red lines are for the Tone2 word and the black lines are for its Tone1 counterpart.
Figure 3.13: “Tone2 CriticalWord Tone1” sequence, F0 aligned at the end of the critical syllables “ti”, “yan” and “qin” in all three boundary types (Word, ip, IP). The red lines are for the Tone2 word and the black lines are for its Tone1 counterpart.
Figure 3.14: “Tone2 CriticalWord Tone3” sequence, F0 aligned at the end of the critical syllables “pi”, “shi”, “ting” and “xi” in all three boundary types (Word, ip, IP). The red lines are for the Tone2 word and the black lines are for its Tone1 counterpart.
Figure 3.15: “Tone2 CriticalWord Tone4” sequence, F0 aligned at the end of the critical syllables “bo”, “huan”, “shi” and “zhì” in all three boundary types (Word, ip, IP). The red lines are for the Tone2 word and the black lines are for its Tone1 counterpart.
Figure 3.16: “Tone2 CriticalWord Tone4” sequence, F0 aligned at the end of the critical syllables “tan”, “quan”, “fang” and “qiu” in all three boundary types (Word, ip, IP). The red lines are for the Tone2 word and the black lines are for its Tone1 counterpart.
Figure 3.17: Relative minimum F0 in critical word in all 6 conditions.
The purpose of this study was to examine native listeners’ processing of assimilation in connected speech. The two types of assimilation studied are English coronal assimilation and Putonghua Tone2 sandhi. Although there are some similarities between these two, they also have some major differences. It is interesting to compare the production data and make predictions about how the phonetic commonalities and differences would affect listeners’ perception.

4.1 Commonalities

As discussed in chapter 1, Section 1.2.2, both English coronal assimilation and Putonghua Tone2 sandhi have been argued to be assimilation processes. Both are optional, and usually take place when the speech rate is fast and when the potentially assimilated word is followed by a relatively weak prosodic break or no break. Both English coronal assimilation and Putonghua Tone2 sandhi have been argued to be gradient processes. That is, there is no discrete distinction between “assimilated” or “unassimilated” forms, but there is a continuum of different levels of assimilation,
associated with other types of continua such as speech rate. For the purposes of this
dissertation, both English coronal assimilation and Putonghua Tone2 sandhi materials
were created in a set of similar experimental conditions. That is, both were embedded
in neutral semantic context, and were disambiguated by a following short sentence.
Both had three types of prosodic boundaries inserted between the critical word and its
following context. Critically, this common manipulation resulted in different degrees
of assimilation, and consequently, different levels of ambiguity.

Phonetic measures of duration showed that the effects of boundary strength were
quite similar across the two sets of materials. The mean length of the critical word in
the IP Boundary conditions were longer than the critical words in the ip Boundary
conditions, which in turn were longer than the critical words in the Word Boundary
conditions in both the English and Putonghua production data. Also, the critical
words that underwent sound change, that is, the coronal words and the Tone2 words,
were significantly longer in duration than their non-coronal or Tone1 counterparts in
IP Boundary and ip Boundary conditions, but not in Word Boundary conditions.

4.2 Differences

4.2.1 Segmental vs. suprasegmental

An obvious difference between English coronal assimilation and Putonghua Tone2
sandhi assimilation is that one involves consonant place and the other involves syllable
tone. Therefore, the articulatory gestures that caused the assimilation were different,
and as a result, the acoustic cues were different as well. Many coronal words in the
English production study had a release even in the context of weaker ip boundaries,
and for the Putonghua production study, 11 out of the 34 Tone2 items were still heard as Tone2 even in the Word Boundary condition (see pretest results reported in Chapter 7). As the phonetic analyses of the production data in Chapter 2 and 3 indicated, these two different types of phonological processes differed in the location of the change in sound within the word, and also differed across the three types of boundary conditions. More specifically, the coronal and non-coronal words in English had different F2 or F3 patterns starting from the medial position in the vowel, and this was especially true for IP Boundary and ip Boundary conditions. In contrast, the Putonghua sandhi Tone1 (Tone1s) and Tone1 differed from each other in their F0 values across the three boundary conditions. The F0 value for Tone1s was lower than Tone1 in the IP Boundary and ip Boundary conditions, but not in the Word Boundary condition. This was true for measurements taken in word initial, medial and final positions.

4.2.2 Boundary tones

Although same types of prosodic boundaries were inserted in between the critical word and its following context word in both English and Putonghua production studies, the specific boundary tones used were different because of the language difference. In particular, contrastive stress was used in the Putonghua study to realize the IP Boundary because of the many experimental constraints as mentioned in Chapter 3.
### Table 4.1: Comparison of English and Putonghua Mean Critical Word Duration

<table>
<thead>
<tr>
<th>Condition</th>
<th>Boundary</th>
<th>Critical Word</th>
<th>Duration (E)</th>
<th>Duration (P)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Word</td>
<td>Non-coronal/Tone1</td>
<td>0.189</td>
<td>0.136</td>
</tr>
<tr>
<td>2</td>
<td>ip</td>
<td>Non-coronal/Tone1</td>
<td>0.265</td>
<td>0.192</td>
</tr>
<tr>
<td>3</td>
<td>IP</td>
<td>Non-coronal/Tone1</td>
<td>0.421</td>
<td>0.271</td>
</tr>
<tr>
<td>4</td>
<td>Word</td>
<td>Coronal/Tone2</td>
<td>0.198</td>
<td>0.141</td>
</tr>
<tr>
<td>5</td>
<td>ip</td>
<td>Coronal/Tone2</td>
<td>0.314</td>
<td>0.205</td>
</tr>
<tr>
<td>6</td>
<td>IP</td>
<td>Coronal/Tone2</td>
<td>0.442</td>
<td>0.297</td>
</tr>
</tbody>
</table>

4.2.3 **Duration**

The average length of the English critical word was longer than the Putonghua critical word as shown in Table 4.1.

Word lengthening associated with the presence of prosodic boundaries differed for the two types of sound change processes in the Putonghua vs. the English materials. The English production data showed that the vowels in coronal critical word were longer than the vowels in non-coronal critical word in the IP and ip Boundary conditions. However, in Putonghua production data, the vowels in Tone1s critical words were longer than the vowels in Tone1 critical words only in the IP Boundary condition.

In the English production data, although the vowel and critical word duration were in general longer for coronal words than non-coronal words, the overall length from the onset of the critical word to the onset of the following context word was comparable for the word pairs across all boundary conditions. This is to say; the
silence that followed the coronal critical word was shorter than that following the non-coronal word. There were no comparable differences in following silent durations between word pairs in the Putonghua productions.

The two production data sets also differed in the way silence was associated with the prosodic boundary conditions. In the English production data, silence followed the critical word in the IP and ip Boundary conditions. However, the Putonghua production data contained only very short periods of silence between the offset of the critical word and the onset of the following context word in all conditions. This was due to the fast rate of speech necessary to produce a natural-sounding Tone2 sandhi, which is quite uncommon at slower speech rates.

4.2.4 Assimilation across boundaries

In English production data, prosodic boundaries do not necessarily block the assimilation. Even the Coronal IP is still possibly assimilated, although the assimilation might be very weak, and the acoustic cues for the coronal place were obvious. However, the Putonghua Tone2 sandhi could be completely blocked by a strong prosodic boundary. In the Tone2 IP condition, none of the 34 items had F0 contour that fell in the realm of Tone1.

4.2.5 Predictions

The experimental conditions were similar for both English and Putonghua. The phonological sound change processes for the two languages were both gradient and likely to be blocked by stronger prosodic boundaries. Thus, it was predicted that for both English and Putonghua listeners, sentences in the Word Boundary conditions
would be more ambiguous than those in the ip Boundary Conditions, while the IP Boundary conditions should be the least ambiguous. Such a result is predicted because sound change would less likely to take place when there was a strong prosodic boundary between the critical word and its following context word, and when the duration of the critical words were longer, thus providing more time for the speaker to execute articulatory targets.

In the English production, F2 and F3 of the coronal were not different from the non-coronal until the middle of the vowel in the critical word, while in the Putonghua production, the F0 of Tone2s was different from Tone1 starting from the beginning of the vowel. Therefore, it was predicted that when the prosodic conditions for inducing assimilation are the same, Putonghua listeners should be able to distinguish Tone2s from Tone1 earlier than English listeners are able to distinguishing coronal from non-coronal. However, overt cues in some items, such as a release for the coronal /t/ at the end of the critical word in the English materials should facilitate listeners’ distinguishing of the coronal.

English words were longer than Putonghua words, with longer vowels for coronal than non-coronal segments in the ip Boundary condition. Putonghua words were of shorter duration overall than the English words, and had comparable durations for Tone1s and Tone1 items. In addition, there was silence following the critical word in English study, but not in Putonghua study. The combination of these differences in duration across studies might mean that the English tokens were less ambiguous than the Putonghua tokens, and the Putonghua listeners might find the auditory stimuli more confusable than the English listeners.
CHAPTER 5

PRETEST FOR ENGLISH EYE TRACKING EXPERIMENT

Two pretests were conducted for the English eye tracking experiment to assess the intelligibility and strength of assimilation in the auditory stimuli and the acceptability of the visual stimuli.

5.1 Listening Pretest

5.1.1 Introduction

Since English coronals may have different levels of assimilation, and the result of acoustic analysis of ambiguity level might not match that of human perception, pretests are needed to ensure that the auditory sentences to be used in the main eye-tracking experiment are comprehensible by the listeners, and preferably show a pattern that critical words followed by an IP boundary is perceptually easier to recognize than words followed by an ip boundary, which is in turn easier to be comprehended than words followed by a word boundary.
In this intelligibility task, the listeners were asked to listen to the ambiguous first sentences from the pairs of sentences created in the English production experiment. They made a forced choice decision about which of two critical words that they heard in the sentence (corresponding to coronal or non-coronal items) and then rated the level of ambiguity of the critical word on a scale from 1 to 7, where 1 indicating definitely the non-coronal, 7 indicating definitely the coronal, and 4 meant a point that was ambiguous between the two words.

Since the phonetic analysis of the production data showed that the coronal words in the Word Boundary conditions were not much different from the non-coronal conditions, but the ip and IP conditions showed more physical differences, it was predicted that prosodic boundary types would influence how listeners perceived the critical words. When there was no boundary between the critical word and its immediately following context word, the critical word should be more likely to be perceived as ambiguous. As discussed in the English production study (Chapter 2), when the prosodic boundary becomes stronger, the ambiguity level of the critical words should be lower.

5.1.2 Method

Participants

48 students from undergraduate linguistics classes at Ohio State University received course credit in exchange for their participation. All were native speakers of Midwestern American English, with normal hearing by self-report.
Table 5.1: Experimental conditions for pre-testing auditory materials, English

<table>
<thead>
<tr>
<th>Condition</th>
<th>Place</th>
<th>Boundary</th>
<th>Example</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>non-coronal</td>
<td>Word</td>
<td>She was careful not to let her cup ( ) bang into the table.</td>
</tr>
<tr>
<td>2</td>
<td>non-coronal</td>
<td>ip</td>
<td>She was careful not to let her cup (H-) bang into the table.</td>
</tr>
<tr>
<td>3</td>
<td>non-coronal</td>
<td>IP</td>
<td>She was careful not to let her cup (L-H%) bang into the table.</td>
</tr>
<tr>
<td>4</td>
<td>coronal</td>
<td>Word</td>
<td>She was careful not to let her cut ( ) bang into the table.</td>
</tr>
<tr>
<td>5</td>
<td>coronal</td>
<td>ip</td>
<td>She was careful not to let her cut (H-) bang into the table.</td>
</tr>
<tr>
<td>6</td>
<td>coronal</td>
<td>IP</td>
<td>She was careful not to let her cut (L-H%) bang into the table.</td>
</tr>
</tbody>
</table>

Materials

Sentence pairs from the English production experiment (Experiment 1) were truncated with Praat, so that only the ambiguous first sentence was used. For example, for the following sentence pairs, listeners only heard the ambiguous sentence.

He drew the line ( ) perfectly. I don’t know how he made it so straight.

He drew the lime ( ) perfectly. I almost want to pick it up and smell it.

Design

The 2 critical word coda place types (coronal vs. noncoronal), together with 3 prosodic boundary types (word boundary, ip boundary, and IP boundary) yielded 6 conditions for the experiment, as shown in Table 5.1:

Six lists were generated. Each list contained 36 experimental (test) items. The 36 experimental items rotated in a Latin square design through the 6 conditions across
the 6 lists so that in different lists, the experimental items were in different conditions. Each list was presented to 8 participants. Each participant heard 36 sentence fragments - 6 items in each of the 6 conditions. Thus, half of the Targets were coronals, the other half non-coronals; one-third of the Targets occurred in each of the three prosodic boundary conditions. The presentation order of the items was randomized for each participant.

Procedure

An online experiment was used for this pretest. The online experiment tool Web-Exp2Corley and Keller, a toolbox for conducting psychological experiments over the World Wide Web, was used to present the experimental stimuli.

On each trial, the participants viewed a screen with a “Play Sound” button above a pair of Target words associated with a 7-point scale. They were instructed to click the “Play Sound” button and listen to a sentence fragment, and then to make a judgment on which of the two words presented on the screen was heard in the sentence.

Two words, the coronal word (e.g., mat) and its non-coronal counterpart (e.g. map), were presented on the left and right of the screen with 7 number buttons in-between them. The coronal word was always presented on the right, and the non-coronal was always on the left.

The experiment began after a single practice trial. Participants were allowed to do the experiments on their own pace, listening to the sound file as many times as they wished.
5.1.3 Results

To make data analysis easier, the numbers were converted so that 1 indicated “Target”, the intended word in the recording, while 7 indicated “Competitor”, the word other than the Target in the critical word-pair.

Overall, the data showed that for both coronal and noncoronal words, the IP Boundary conditions were perceptually less ambiguous than the ip Boundary conditions, and the ip Boundary condition were less ambiguous than the Word Boundary conditions as demonstrated in Figure 5.1.

However, Wilcoxon test results indicated that words with coronal and non-coronal codas were not significantly different in this off-line listening test ($p = 0.77$ in IP conditions, $p=0.08$ in ip conditions, and $p = 0.09$ in Word conditions).
The primary purpose of this pretest experiment was to examine the goodness of the auditory stimuli for the main eye tracking experiment. Results show that the stimuli were intelligible (no condition produced a mean rating above 3.5) and that the stimuli had different degrees of ambiguity due to different boundary types, as predicted.

What the result didn’t show was a significant different between coronal and non-coronal coda. However, the online Eye Tracking study that we conducted later yielded a different result. This contrast in the result between online and off line studies also demonstrates the importance of using eye movement monitoring method to study the listeners’ processing of ambiguous words in continuous speech.

5.2 Picture-Word Matching Experiment

5.2.1 Introduction

For the eye movement monitoring experiments, it was important to know what name(s) participants typically assign to the pictures stimuli that would be potentially associated with the spoken Targets. An online word-picture matching experiment was used to test the goodness of fit of the pictures to the critical words that would be tested in the main English eye tracking experiment.
5.2.2 Method

Participants

Two faculty members and four graduate students from the Linguistics Department of Ohio State University voluntarily participated this experiment online. Five of them were native speakers of American English. One was a native speaker of British English.

Materials

Simple line drawing were chosen for eye tracking experiment to avoid confounding elements such as the variable complexity of detail present in photographs (personal communication, M.K. Tanenhaus, 2006). Because of the limited number of available coronal-noncoronal word pairs appropriate for the experimental conditions, determining appropriate line drawings was a non-trivial task. For example, some of the critical words did not have stereotypical, easily imageable forms that would be recognizable for most people (e.g. right), while for others, it was hard to represent the named concept with simple black-and-white line-drawings (e.g. the concept of “ripe”).

The author first searched google for the most representative pictures for each word pair, then asked an artist to convert these pictures into line drawings. These line drawings were then screened by a group of psycholinguists and modified extensively on the basis of these recommendations.

36 four-image pencil drawings by the artist were scanned, digitized, and adjusted to 340 pixels in height and 438 pixels in width by using Microsoft Photo Editor. This created 36 slides, each with four simple line drawing pictures. One corresponded to the critical word, e.g. “loon”. One matched to the Competitor. E.g. “loom”. One was related to a word that shares the same onset with the critical word, e.g. “lamp”,
and the fourth one was irrelevant to the critical word. The four types of pictures rotated in position in different slides, so that the participants would not predict in which position of the slide the critical word would appear. The 36 trials were presented randomly during the experiment, so that different participants saw the slides in different order.

Procedure

The experiment was presented by WebExp2 Corley and Keller, which is a toolbox for conducting psychological experiments over the World Wide Web. The participants used a computer that was connected to Internet to access the experiment website.

In each trial of the experiment, participants saw a slide with 4 pictures on the web page, the one in the upper-left region was marked as “A”, the one in the upper-right region was marked as “B”, the one in the lower-left region was marked as “C”, the one in the lower-right region was marked as “D”. An example screen from the experiment is shown in Figure 5.2.

Below the slide were four questions. The first one asked the participants to match the pictures to four English words (that is, the critical word, the Competitor, the same-onset word, and the irrelevant word) by input the letter associated with the picture (either A, B, C, or D) to the box next to the written word. The second question asked the participants to rate how well each picture matched the word on a 5-point scale, where “1” meant not well matched at all, and “5” meant perfectly matched. The third question asked which picture could be named right away, as a rough index of the level of abstraction needed to associate the word and the picture. The fourth question asked which picture could be named the critical word that would
undergo coronal place assimilation in the continuous auditory sentence for the main experiment, e.g., “loon”. After finishing all the questions in a trial, participants proceeded to the next trial by clicking on the “next” button at the lower right corner of the web page.

5.2.3 Results

For Question 1, the picture-word matching task, there were only two errors out of the total of $36 \times 6 = 216$. The error rate was 0.9
For Question 2, the rating of the match between word and picture, the average rating for the pictures corresponding to the coronal word (e.g. right) was 3.2 on a 5-point scale, and the average rating for the pictures corresponding to the non-coronal counterpart (e.g. ripe) was 3.9 on a 5-point scale. This suggests that pictures were reasonably well matched, with slightly better matches between picture and word for the noncoronals.

For Question 3, asking about whether the pictures could be named right away in 11 out of the 36 slides, at least half (that was 3 people) of the participants (that was 3 people) reported that the picture corresponding to the coronal word could be named right away. In 13 out of the 36 slides, at least half of the participants could name the picture corresponding to the counterpart of the coronal word right away. This suggests that, as a set, coronal and noncoronal pictures were comparably abstract.

For Question 4, all participants were able to identify all pictures for the coronal words. However, in 7 out of the 36 slides, some answers showed confusion. 2 out of the 7 had one participant confused between the coronal word and one of the distracters, 2 had two participants confused between the coronal word and one of the distracters, 1 had two participants confused among all the pictures, 1 had 3 participants confused between the coronal word and it’s non-coronal counterpart, and 1 has 4 participants confused between the coronal word and one of the Distracters.
5.3 Summary

The result of the picture-word matching task suggested that although people might not be able to name the pictures immediately upon seeing them, and they might not feel the picture was the most typical representation of the word, in general, participants experienced no problem in matching the picture with the corresponding Target word correctly for either coronals or non-coronals. In the eye tracking experiment, given plenty of scanning time before hearing the auditory stimuli and with the help with the auditory stimuli itself, listeners should not have much problem in linking the correct picture with the intended word.
CHAPTER 6

EYE TRACKING EXPERIMENT ON ENGLISH CORONAL ASSIMILATION

6.1 Introduction

The purpose of this experiment was to test listeners’ recognition of spoken words that were ambiguous because of contextual factors. Specifically, we tested native listeners’ processing of English coronal assimilation and its interaction with different prosodic boundary types that occurred in between the coronal word and its following context word.

Four black-and-white line drawings were shown in the four corners of a computer screen. Participants’ eye movements toward these pictures were recorded as they detected which corresponded to the word they had just heard in the context of an auditory sentence pair.

6.2 Why Eye Tracking

Many previous studies of coronal assimilation used interrupted and truncated speech as stimuli (e.g. gating), involved metalinguistic judgment tasks (e.g. lexical decision), or involved more complex and unnatural paradigms such as cross modal
priming. These tasks have intrinsic disadvantages; while gating requires metalinguistic reflection to process speech sounds that would not occur were it not for the possibility of digital editing of the speech stream, lexical decision and cross-modal priming tasks record listeners’ processing at a single time point, usually more than 300 ms after the offset of the auditory stimuli. However, with the hypothesis that automatic, unconscious processing take place upon hearing the stimuli, and conscious linguistic judgment arises quickly afterwards, we need a more “online” method that could allow us to observe listeners’ processing at the earliest possible time, and as it unfolds. Also, contextual effects only occur in continuous speech, therefore, we need to monitor the processing continuously. In addition, listeners might experience ambiguity when listening to the auditory stimuli; therefore, we need to observe their response in a continuous manner so that we can learn their ambiguity resolution process. Eye movement monitoring is now routinely used for examining the time course of spoken word and sentence processing (Trueswell et al., 1993).

The main experiments in this research use eye movement monitoring methodology to investigate listeners’ processing of English assimilated items and Putonghua Tone2 sandhi items in the context of prosodic boundaries of varying strength. During the experiment, a light weight head mounted eyetracker was used to monitor listeners’ eye movement continuously while they heard uninterrupted, complete auditory sentences, and choose one line drawing picture from a set of four on the computer screen that best matched the auditory stimuli. Details of the method are provided in Chapter 6 and Chapter 8.
Even with the many advantages of using eye movement monitoring method for this research it needs to be noted that the limited visual choice (4 pictures) might lead to task specific strategies such as memorizing the location of each picture, and not looking at the picture while listening to the stimuli. Nevertheless, we predict that listeners’ eye movements will show consistent patterns regardless of the strategies they might use.

6.3 Prediction

We hypothesized that the stronger the prosodic boundary between the critical word and its following context word, the less coarticulation would be found in the production gestures for the coronal, and the less ambiguity the listeners would perceive. In addition, English listeners have experienced many sequences of coronal consonants followed by labial or dorsal ones, presumably with a higher likelihood of assimilation for cases when the coronal is separated from the context segment by a word boundary than for cases when the segments are separated by an ip boundary, with the lowest likelihood of assimilation when the segments are separated by an IP boundary. If listeners are sensitive to this past experience, when the input sound is temporarily ambiguous, they should be more likely to perceive lexical items with coronal codas in the word boundary condition than in the ip boundary condition, and more likely to perceive lexical items with coronal codas in the ip boundary condition than in the IP boundary condition.

Based on these assumptions, we predicted more rapid recognition of the critical word as the prosodic boundary gets stronger. This would be reflected by earlier fixations to the Target picture in the stronger boundary conditions than the weaker
boundary conditions. In addition, we predicted a longer period of looks to both the Target and the distracter for the more ambiguous weaker boundary conditions.

Different spoken language processing models make sharply different predictions for the outcome of this experiment. Specifically, as mentioned in Chapter 1, the phonological inference account would predict that: when listeners hear an unassimilated or only weakly assimilated item (with an IP boundary or ip boundary between the critical word and its context word), the processing system would correctly treat the input as an assimilated item rather than an underlyingly non-coronal, predicting an upcoming context that would have triggered assimilation. For example, when hearing a weakly assimilated token of *right*, the processor will predict it would be followed by a /b/). When the contextual information is confirmed, processing is facilitated. This means, when participants hear the assimilated word, they would start to look at the Target, and when they hear the context word, the fixation proportion to the Target should increase immediately. For strong assimilations, this account predicts that when hearing the critical word, the system would not be able to distinguish the intended word (e.g. a stronger assimilated token of *right*) from its non-assimilated counterpart (e.g. *ripe*). However, upon hearing the context segment, the system would be able to retroactively assign “assimilation” to the preceding item. Therefore, for the strongly assimilated items in the Word Boundary condition, the ambiguity should resolve and looks to the Target should increase right after the context segment becomes available, and before the disambiguating semantic context becomes available. However, for non-assimilated items in the Word Boundary condition, the
system should also assign “assimilation” to the Target. That is, listeners would experience ambiguity first, and then infer the assimilated word (e.g. “right”) even when hearing the non-assimilation word (e.g. “ripe”).

In contrast, according to the feature cue parsing account, listeners should not experience ambiguity of any kind. When hearing non-coronal items, listeners should always complete lexical access only the basis of surface forms, and when hearing coronal items, listeners will always be able to use “feature cues” to distinguish the coronal place in the stimuli. Even with the strongly assimilated items, as in the Word Boundary condition, listeners should be able to distinguish coronal from non-coronal forms. Recognition should take place even before the following triggering context segment become available. There should not be ambiguity in the recognition process, so that listeners should look to the correct Target picture as soon as they have processed the Target word, and before they encounter the onset of the context word. A comparison of processing in the three prosodic boundary conditions might show that the time to look to the Target picture is dependent on word length, with shorter target words (in the weaker boundary conditions) producing the fastest recognition times.

The perceptual integration account also predicts that there should not be ambiguity when listeners hear strongly assimilated stimuli in the Word Boundary condition. On this account, the processor should always favor the assimilation option even when hearing non-assimilated word, just in case the following context segment is viable, that is, when the context segment creates a potential environment for assimilation. This is similar to the phonological inference account. For ip and IP Boundary conditions, since the contextual information is not immediately available, listeners would not
be able to use it for recognition. If an item under these conditions is perceptually ambiguous, then it would remain ambiguous for a longer period of time, until the relevant context is encountered.

The *revised exemplar account* predicts that both assimilated coronal and non-coronal items would be ambiguous, as many similar exemplars exist in the mental lexicon. Listeners should be able to distinguish weakly assimilated segments on the basis of their acoustic form, but they would not be able to distinguish strongly assimilated segments. That is, for non-coronal items in the Word Boundary condition, listeners would experience ambiguity when hearing the critical word, but the context segment would provide further acoustic information for disambiguation (as the formant graphs in Chapter 2 showed, the formants for coronals and non-coronals were different after the end of the vowel in Word Boundary conditions). In the ip Boundary and IP boundary conditions, comparable process would occur. For strongly assimilated coronal items in the Word Boundary condition, the segment will remain ambiguous until semantic information comes to help. For coronal items in the more weakly assimilated ip Boundary and IP Boundary conditions, listeners should be able to distinguish the Target quickly because of the gradient nature of the segment that undergoes assimilation.

Section 6.2 of this chapter presents the experiment method. Section 6.3 describes the results and section 6.4 summarizes and describes the results.
6.4 Method

6.4.1 Participants

40 native speakers of American English participated in an eye movement monitoring experiment. They were recruited from the Linguistics Outside the Classroom at the Ohio State University, and received partial course credit (1.25% of their final score) for their participation. Since they were undergraduate students of introductory linguistics or psycholinguistic classes, and the experiment was conducted at the beginning of the quarter, they can be considered naive listeners. They all had normal eyesight, normal hearing, and were around 18-24 years of age. Most of them were from Ohio and had been living in this state most their life. Four of the participants’ data were discarded due to track-loss. Analysis of the experiment results were based on the data collected from the remaining 36 participants.

6.4.2 Design

This experiment studies the influence of both English coronal place assimilation and prosodic boundary types on listeners’ lexical processing.

The 6 auditory conditions were the same as those shown in Table 5.1. For convenience sake, the conditions with full example sentences (both the ambiguous and the disambiguating parts) are also listed in the following:

For each item, as mentioned in Chapter 5, there were 4 line drawings displayed on the screen: the Target (the same word as in the auditory stimuli, e.g. “cut”), the Competitor (the (non-)coronal counterpart, e.g. “cup”), Distracter 1 (a word that shares similar onset with the Target/competitor, e.g. “key”), Distracter 2 (a word
<table>
<thead>
<tr>
<th>Condition</th>
<th>Place</th>
<th>Boundary</th>
<th>Example</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>non-coronal</td>
<td>Word</td>
<td>She was careful not to let her cup ( ) bang into the table. So that it wouldn’t be broken.</td>
</tr>
<tr>
<td>2</td>
<td>non-coronal</td>
<td>ip</td>
<td>She was careful not to let her cup (H-) bang into the table. So that it wouldn’t be broken.</td>
</tr>
<tr>
<td>3</td>
<td>non-coronal</td>
<td>IP</td>
<td>She was careful not to let her cup (L-H%) bang into the table. So that it wouldn’t be broken.</td>
</tr>
<tr>
<td>4</td>
<td>coronal</td>
<td>Word</td>
<td>She was careful not to let her cut ( ) bang into the table. So that it wouldn’t hurt again.</td>
</tr>
<tr>
<td>5</td>
<td>coronal</td>
<td>ip</td>
<td>She was careful not to let her cut (H-) bang into the table. So that it wouldn’t hurt again.</td>
</tr>
<tr>
<td>6</td>
<td>coronal</td>
<td>IP</td>
<td>She was careful not to let her cut (L-H%) bang into the table. So that it wouldn’t hurt again.</td>
</tr>
</tbody>
</table>

Table 6.1: Experimental conditions and auditory materials, English

that is irrelevant to the Target or the Competitor, e.g. “apple”), which rotate in position across the items.

6 lists were generated. Each list contained 3 practice items, 36 auditory experimental (test) items, and 36 filler items. The order of all items was the same across the lists. The 36 test items were rotated in a Latin square design through the 6 conditions across the 6 lists, so that in different lists, the test items were in different conditions. The visual stimuli were the same for each item across the lists and conditions. Each list was seen by 6 participants. Thus, each participant heard 75 sentences in total, 3
of them were practice items, 36 were fillers, and 36 were test items, with 6 items in each condition.

6.4.3 Auditory Stimuli

The auditory stimuli for this experiment were the sound files generated in the production study. The sentences were generated based on stimuli from Gow (2002) as described in Section 2.2.2.

The auditory ambiguous sentences incorporated two independent variables: First, the place of articulation of the last consonant in the critical word: either coronal or non-coronal. Second, the prosodic boundary inserted between the critical word and the word that follows it: Word Boundary, Intermediate Phrase Boundary, or Intonation Phrase Boundary. Each ambiguous sentence was followed by a short disambiguating sentence. The complete set of sentence materials is presented in Appendix A. There were also 36 filler sentences that were not related to coronal assimilation. Since some of the critical item and filler sound files were recorded in different sessions, a computer program was used to normalize the intensity of all the sound files, so that they had approximately the same amplitude.

6.4.4 Visual Stimuli

All slides were selected from the picture-word matching pre-test (Section 5.2.2). These slides, as well as all the slides for fillers, were readjusted to 621 x 480 pixels using Micro Soft Photo Editor and saved as a bitmap file, as required in order to use E-Prime software (Schneider et al., 2002) to control the display of stimuli. Bitmap
files were constructed so that the four line-drawings were close to the four respective corners of the display, but not too close. This was done to ensure that participants would make sufficient eye movement toward different quarters of the computer screen, but at the same time, not overshoot to the outside of the screen. The four pictures in each file included: the Target and its Competitor, namely, the word ending with a coronal (e.g. cut), its non-coronal counterpart (e.g. cup), Distracter 1 (corresponding to a word that had the same initial consonant as the Target, e.g. key), and Distracter 2 (corresponding to an irrelevant word, e.g. apple), rotated in Latin square design in the list.

Thus, for different trials in the same list, these pictures appeared in different corners of the screen. This is to avoid the possibility of participants developing any expectations about where the Target picture may appear during the experiment. In filler slides, the Targets are also randomly located in different corners. In all item (critical) slides, and 50% of the filler slides, efforts were made to have the four pictures composed with approximately the same number of lines and roughly the same size. Therefore, they would have about the same level of visual complexity. This was to avoid the participants’ eye movement pattern from being overly influenced by a comparatively more complex picture.

Sample visual stimuli are shown in Figure 6.1.

6.4.5 Procedure

Upon entering the Eye Tracking Lab, subjects were asked to sit in front of a 17 inch LED monitor at a distance of approximately 20 inches.
Figure 6.1: Sample visual stimuli for English eye tracking experiment. Upper left: apple (irrelevant word), Upper right: cut (coronal), Lower left: cup (non-coronal counterpart), Lower right: key (same initial consonant cohort)
The experimenter then placed the headgear of the ASL 6000 Model Eye Tracker on the participant’s head, and adjusted the monocle of the headgear to calibrate the left eye of the participant. The average time for calibration was about 5 minutes. For a few participants, it took 20 minutes. Eprime software coordinated the timing of the presentation of the visual displays and the sound files, and also automatically started and stopped the recording of eye location data via a serial connection to the ASL eye tracing system. After calibration, the experimenter read the experiment procedure from a sheet of paper to the participant, and informed them that at the beginning of each trial, they would see four simple line drawings on the computer screen. After looking at these pictures carefully, they could click the mouse. Upon the clicking of the mouse, the cursor would appear at the center of the screen and at the same time, they would hear an auditory sentence. The participants’ task was to then try to decide as quickly as possible which of the four pictures most closely matched the sentence they had just heard, and to move the mouse to that picture. To ensure that they had heard the entire sentence (especially the disambiguation part, which usually occurred at the end) before clicking on the mouse to move on to the next trial, the E-prime program was set in such a way that if the participant clicked the mouse when the sound file was still being played, it would not stop the current trial and jump to the next one. Only after the sound file stopped would the mouse clicking action be effective, and the program proceed to the next trial. At the end of this explanation, the experimenter gave a simple example of the experiment trial and what the participant should do upon hearing that trial, such as the following:

For example, if you hear the sentence “She was wondering if she should eat an apple this morning”, and see four pictures “pear, apple, orange, banana” on the
screen, you should move your mouse to “apple” as quickly as possible, and click on it at the end of the sentence.

The participants had a chance to ask any questions they had before the experiment started, then they were instructed to start the three practice trials by clicking the left button of the mouse once. If there were no problems during the practice session, they could proceed to the main session of the experiment.

During the main session, the experimenter closely observed the three monitors of the eye tracking system: 1) the monitor for the scene camera, which showed where the participant’s eye was looking, so that she could take notes on any odd behavior of the participants; 2) the monitor for eye calibration, to make sure that the eye was well calibrated during the entire experiment; and 3) the main computer with ASL software window to monitor “eye head integration” data, the raw data for the experiment.

6.5 Results

Fixations to all four areas-of-interest (AOI)s (Target, Competitor, Distracter 1, Distracter 2) were automatically generated by the ASL E6000 software. A Perl program was then used to transfer the raw data to arcsine transformed fixation proportions for each AOI for all 6 conditions across participants and items, as shown in Figures 6.2 through 6.7. Fixation proportions were calculated based on the proportion of total possible looks to the visual stimuli for all items and participants (denominator = 36 participants x 6 items/condition = 216). The fixation proportions were calculated for every 17 ms, from 500 ms before the offset of the vowel of the critical word to 3000 ms after the offset of the vowel. Data were aligned at the offset of the
vowel in the critical word for each spoken sentence. We chose the end of vowel as the alignment point because, as discussed in the production study, the place information of the coda consonant is shown in the formant information of its preceding vowel on a spectrogram, not the consonant part itself. This is especially true for those “heavily assimilated” items. In addition, the backward alignment of the data allows us to observe any earlier effect, e.g. coarticulatory information present in the vowel, or any bias due to factors other than those in auditory signal. For the statistical analyses of time intervals before the vowel offsets, we used different-sized time windows for the different conditions because the duration of the critical word- and the silent durations following the word varied across conditions.

A global examination of the pattern of fixations across all 6 conditions was used to determine the extent of analyses beyond the offset of the vowel. In all conditions except Condition4 (Coronal, Word Boundary), the pattern of eye movements stabilized by about 900ms after vowel offset, as participants settled on the identity of the Target picture. Therefore, to examine the time course of significant differences in fixating the Target, Competitor and Distracters, we analyzed the first 900 ms after vowel offset, using a relatively small (150ms) fixed window to closely examine any changes in the pattern of effects. For Condition 4, we analyzed the first 1500ms after vowel offset, using the same 150 ms fixed window frame. According to previous literature, 150 ms is also the minimum latency to plan and launch a saccade in simple tasks like the one used in the current experiment (Fischer, 1992; Saslow, 1967; Dahan et al., 2001). For all conditions discussed below, statistics were performed on arcsine-transformed mean fixation proportions for each 150 ms window for each Area of Interest, while plotted data reflect untransformed proportions.
6.5.1 Condition 1 Non-coronal, Word Boundary

Figure 6.2 shows listeners’ fixation proportions to the four areas of interest: Target (e.g. cup), Competitor (e.g. cut), Distracter 1 (e.g. key), Distracter 2 (e.g. apple) for non-coronal critical words followed by a Word Boundary. Data are aligned (Point 0) from offset of the vowel (e.g. ’cup’ in the sentence ‘She was careful not to let her cup() bang into the table.’). The figure shows that listeners began to look to the Target picture immediately after the offset of the vowel of the critical word (e.g. cup) when it was followed by a Word Boundary.

In this condition, and in all later conditions, listeners excluded Distracter 2 as a possible answer very early, since it did not share the same initial consonant with the Target. In contrast, listeners considered the Target, Competitor, and Distracter 1 when they were listening to initial portion of the critical word. This observation is confirmed by One-way ANOVA and planned comparisons conducted on the mean fixation proportion to the Target picture, the Competitor, and the two Distracters over the time window extending from 100ms (mean length of vowels in this condition) to 0ms before the offset of the vowel. The statistics show a difference between Target and Distracter 2 [$F(1, 35) = 11.4, p = 0.001$], but not for Target and Competitor [$F(1, 35) = 0.1, p = 0.75$], Target and Distracter 1 [$F(1, 35) = 2.3, p = 0.13$], or Competitor and Distracter 1 [$F(1, 35) = 1.4, p = 0.23$].

More importantly, Figure 6.2 shows that the fixation proportion to the Target started to increase at about 150 ms after the offset of the vowel, while looks to the Competitor remained basically constant, and those to Distracter 1 decreased quickly. As mentioned earlier, the minimum latency to plan and launch a saccade is 150 ms. This suggests that listeners’ looks to the Target began as soon as they
Figure 6.2: English eye tracking Experiment Condition1 (Non-coronal, Word Boundary): Fixation proportions over time for the Target, Competitor, and the two Distracters. Data are aligned at the offset of the vowel (Point 0) in the critical word. The dashed red line on the left indicates the probable point at which listeners’ eyes move in respond to the onset of the vowel in the critical word. The dashed red line on the right indicates the probable point at which listeners eyes move in respond to the offset of the critical word.
<table>
<thead>
<tr>
<th>Time Window Frame</th>
<th>F-Value</th>
<th>p-Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>0-150ms</td>
<td>2.8</td>
<td>0.09</td>
</tr>
<tr>
<td>150-300ms</td>
<td>1.5</td>
<td>0.2</td>
</tr>
<tr>
<td>300-450ms</td>
<td>8.6</td>
<td>0.004</td>
</tr>
<tr>
<td>450-600ms</td>
<td>17.9</td>
<td>0.0001</td>
</tr>
<tr>
<td>600-750ms</td>
<td>17.3</td>
<td>0.0001</td>
</tr>
<tr>
<td>750-900ms</td>
<td>19.9</td>
<td>0.0001</td>
</tr>
</tbody>
</table>

Table 6.2: Result for Planned Comparison between Target and Competitor means every 150 ms between 0-900 ms after the offset of the vowel in the critical word in Condition 1, English eye tracking experiment

had processed the completed information about vowels in the Non-coronal Word Boundary condition. However, according to the statistics listed in Table 6.1, the fixation proportions to Target and Competitor did not differ significantly until about 300 ms after the offset of the vowel. Around 900ms, looks to the Target drop off a little, and then they pick up again around 1400 ms. This was probably due to a “checking effect” as the disambiguating sentence became available.

### 6.5.2 Condition 2 Non-coronal, ip Boundary

Figure 6.3 again shows fixation proportions to the four AOIs: Target (e.g. cup), Competitor (e.g. cut), Distracter 1 (e.g. key), Distracter 2 (e.g. apple), for non-coronal critical words followed by an ip boundary. Data are aligned (Point 0) from the offset of the vowel of the critical word (e.g. 'cup' in the sentence 'She was careful not to let her cup(H-) bang into the table.') The figure demonstrates that listeners fixated the Target drawing before the offset of the vowel when hearing a non-coronal critical word (e.g. cup) followed by an ip Boundary.
Figure 6.3: English eye tracking Experiment Condition 2 (Non-coronal, ip Boundary): Fixation proportion over time for the Target, Competitor, and the two Distracters. Data are aligned (Point 0) from the offset of the vowel of the critical word. The dashed red line on the left indicates the probable point at which listeners’ eyes move in respond to the onset of the vowel in the critical word. The dashed red line on the right indicates the probable point at which listeners’ eyes move in respond to the offset of the critical word.
In this condition, similar to Condition 1, Distracter 2 was excluded from considerations very early, but there were comparable fixation proportions to Target, Competitor and Distracter 1 before the offset of the vowel in the critical word. This claim is supported by the results of a One-way ANOVA and planned comparisons conducted on the mean fixation proportions to the Target, the Competitor, and the two Distracters pictures over the time window extending from 160 ms (mean length of vowels in this condition) to 0ms before the offset of the vowel, which show a difference between Target and Distracter 2 \( [F(1,35) = 13.4, p = 0.0003] \), and Target and Distracter 1 \( [F(1,35) = 7.0, p = 0.009] \), but not for Target and Competitor \( [F(1,35) = 3.47, p = 0.06] \), or Competitor and Distracter 1 \( [F(1,35) = 0.618, p = 0.43] \).

More importantly, Figure 6.3 demonstrates that the fixation proportion to the Target started to increase at about the alignment point, while those to the Competitor remained constant, and those to Distracter 1 decreased quickly. According to statistics listed in Table 6.2, the fixation proportions to Target and Competitor already differed significantly between 0 ms and 150 ms after the offset of the vowel. Considering the 150 ms minimum latency to plan and launch a saccade, this suggests that listeners’ fixations to the Target began even before they had processed the completed vowel information. This is probably because the H- boundary tone caused rising F0 in the vowel, and listeners were able to make use of this acoustic cue to detect the ip boundary type following the word even before the vowel was completed. Since assimilation is less likely across an ip boundary than at a word boundary, the listeners predicted that what they heard was a non-coronal word based on the low probability of strong assimilation across an ip boundary in English.
<table>
<thead>
<tr>
<th>Time Window Frame</th>
<th>F-Value</th>
<th>p-Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>0-150ms</td>
<td>10.417</td>
<td>0.0017</td>
</tr>
<tr>
<td>150-300ms</td>
<td>11.9</td>
<td>0.0008</td>
</tr>
<tr>
<td>300-450ms</td>
<td>30.7</td>
<td>0.0001</td>
</tr>
<tr>
<td>450-600ms</td>
<td>38.8</td>
<td>0.0001</td>
</tr>
<tr>
<td>600-750ms</td>
<td>31.6</td>
<td>0.0001</td>
</tr>
<tr>
<td>750-900ms</td>
<td>37.1</td>
<td>0.0001</td>
</tr>
</tbody>
</table>

Table 6.3: Results from Planned Comparison between Target and Competitor means every 150ms between 0-900 ms after the offset of the vowel in the critical word. Condition 2, English eye tracking experiment

### 6.5.3 Condition 3 Non-coronal, IP Boundary

Figure 6.4 shows fixation proportions to the four areas of interest: Target (e.g. cup), Competitor (e.g. cut), Distracter 1 (e.g. key), Distracter 2 (e.g. apple), for non-coronal critical words followed by an IP boundary. Data are aligned (Point 0) from the offset of the vowel of the critical word. (e.g. 'cup' in the sentence 'She was careful not to let her cup (L-H*) bang into the table.) The figure demonstrates that listeners fixated the Target before the offset of the vowel, similar to Condition 2, when they heard a non-coronal critical word (e.g. cup) followed by an IP Boundary.

In this condition, Distracter 2 was again excluded from considerations very early, but there were comparable fixation proportions to Target, Competitor and Distracter 1 before the end of the vowel of the critical word. One-way ANOVA and planned comparisons conducted on the mean fixation proportion to the Target picture, the Competitor, and the two Distracters over the time window extending from 250 ms (mean length of vowels in this condition) to 0 ms before the offset of the vowel
Figure 6.4: English eye tracking Experiment Condition3 (Non-coronal, IP Boundary): Fixation proportion over time for the Target, Competitor, and the two Distracters. Data are aligned (Point 0) from the offset of the vowel of the critical word. The dashed red line on the left indicates the probable point at which listeners’ eyes move in respond to the onset of the vowel in the critical word. The dashed red line on the right indicates the probable point at which listeners eyes move in respond to the offset of the critical word.
show a difference between Target and Distracter 2 \( F(1, 35) = 6.5, p = 0.012 \), and Target and Distracter 1\( F(1, 35) = 2.2, p = 0.13 \), but not for Target and Competitor \( F(1, 35) = 0.405, p = 0.5 \), or Competitor and Distracter 1 \( F(1, 35) = 2.2, p = 0.13 \).

Figure 6.4 also shows that the fixation proportion to the Target started to increase at about 80ms after the alignment point while those to the Competitor remained constant, and those to Distracter 1 decreased quickly. According to the statistics in Table 6.3, the fixation proportions to Target and Competitor did not differ between 0 ms and 150 ms after the offset of the vowel\( F(1, 35) = 1.88, p = 0.17 \), but they diverged between 150 ms and 300 ms \( F(1, 35) = 12.6, p = 0.006 \). Again considering the 150 ms minimum latency to plan and launch a saccade, this suggests that listeners’ looks to the Target began even before they had processed the completed information about the vowel. Similar to Condition 2, this is probably because the L-H% boundary tone information available in the vowel caused listeners to be able to detect the IP boundary type following the word even before the vowel was completed. Since the duration of the vowel in this condition was much longer than the vowel in Condition 2 (ip Boundary) \( F(1, 35) = 326.6, p = 0.0001 \), and Condition 1 (Word Boundary) \( F(1, 35) = 717.0, p = 0.0001 \), it was more likely that listeners used the length of the vowel information to determine that there would be an IP Boundary following the word.

### 6.5.4 Condition 4 Coronal, Word Boundary

Figure 6.5 shows fixation proportions to the four AOIs: Target (e.g. cut), Competitor (e.g. cup), Distracter 1 (e.g. key), Distracter 2 (e.g. apple), for coronal critical words followed by a Word Boundary. Data are aligned (Point 0) from the offset of
the vowel of the critical word. (e.g. ‘cut’ in the sentence ‘She was careful not to let her cut ( ) bang into the table.) The figure shows that listeners fixated the Target only after the semantic disambiguation information in the context sentence became available, and could not distinguish the Target from the Competitor when hearing a coronal critical word (e.g. cup) followed by a Word Boundary.

Again in this condition as in previous conditions, Distracter 2 was excluded from consideration very early, while there were comparable fixation proportions to Target, Competitor and Distracter 1 before the end of the vowel of the critical word. One-way ANOVA and planned comparisons conducted on the mean fixation proportion to the Target, the Competitor, and the two Distracter pictures over the time window extending from 100 ms (mean length of vowels in this condition) to 0 ms before the offset of the vowel showed a difference between Target, Competitor and Distracter 1 vs. Distracter 2 \( F(1,35) = 5.1, p = 0.02 \), but not for Target and Competitor

<table>
<thead>
<tr>
<th>Time Window Frame</th>
<th>F-Value</th>
<th>p-Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>0-150ms</td>
<td>1.8</td>
<td>0.1732</td>
</tr>
<tr>
<td>150-300ms</td>
<td>12.7</td>
<td>0.0006</td>
</tr>
<tr>
<td>300-450ms</td>
<td>23.0</td>
<td>0.0001</td>
</tr>
<tr>
<td>450-600ms</td>
<td>48.4</td>
<td>0.0001</td>
</tr>
<tr>
<td>600-750ms</td>
<td>23.8</td>
<td>0.0001</td>
</tr>
<tr>
<td>750-900ms</td>
<td>28.1</td>
<td>0.0001</td>
</tr>
</tbody>
</table>

Table 6.4: Planned Comparison results between Target and Competitor every means 150ms between 0-900 ms after the offset of the vowel in the critical word, Condition 3, English eye tracking experiment
Figure 6.5: English eye tracking Experiment Condition 4 (Coronal, Word Boundary): Fixation proportion over time for the Target, Competitor, and the two Distracters. Data are aligned (Point 0) from the offset of the vowel of the critical word. The dashed red line on the left indicates the probable point at which listeners’ eyes move in respond to the onset of the vowel in the critical word. The dashed red line on the right indicates the probable point at which listeners eyes move in respond to the offset of the critical word.
<table>
<thead>
<tr>
<th>Time Window Frame</th>
<th>F-Value</th>
<th>p-Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>0-150ms</td>
<td>3.9</td>
<td>0.04</td>
</tr>
<tr>
<td>150-300ms</td>
<td>1.8</td>
<td>0.18</td>
</tr>
<tr>
<td>300-450ms</td>
<td>0.55</td>
<td>0.45</td>
</tr>
<tr>
<td>450-600ms</td>
<td>0.65</td>
<td>0.4</td>
</tr>
<tr>
<td>600-750ms</td>
<td>0.04</td>
<td>0.83</td>
</tr>
<tr>
<td>750-900ms</td>
<td>0.463</td>
<td>0.49</td>
</tr>
<tr>
<td>900-1050ms</td>
<td>3.5</td>
<td>0.06</td>
</tr>
<tr>
<td>1050-1200ms</td>
<td>9.1</td>
<td>0.003</td>
</tr>
<tr>
<td>1200-1350ms</td>
<td>21.2</td>
<td>0.0001</td>
</tr>
<tr>
<td>1350-1500ms</td>
<td>17.3</td>
<td>0.0001</td>
</tr>
</tbody>
</table>

Table 6.5: Result for Planned Comparison between Target and Competitor every means 150ms between 0-1500 ms after the offset of the vowel in the critical word. Condition 4, English eye tracking experiment

\[ F(1, 35) = 2.1, p = 0.14 \], Target and Distracter 1 \[ F(1, 35) = 0.0002, p = 0.98 \], or Competitor and Distracter 1 \[ F(1, 35) = 2.09, p = 0.15 \].

Also, as Figure 6.5 shows and according to statistics (see Table 6.4) the fixation proportion to the Target was comparable with those to the Competitor until about 1000 ms after the alignment point. This indicates that listeners experienced ambiguity when listening to a coronal word following by a Word Boundary. They were unable to use information from the coronal or the context segment, and instead had to use semantic information available much later in the context sentence to help with the disambiguation task. Also, this clearly shows the difference between “clarity” and “ambiguity”, since separation between Target and Competitor in non-coronal Target in Word condition clearly already emerging at 200 ms.
6.5.5 Condition 5 Coronal, ip Boundary

Figure 6.6 shows fixation proportions to the four areas of interest: Target (e.g. cut), Competitor (e.g. cup), Distracter 1 (e.g. key), Distracter 2 (e.g. apple), for non-coronal critical words followed by an ip boundary. Data are aligned (Point 0) from the offset of the vowel of the critical word. (e.g. ‘cut’ in the sentence ‘She was careful not to let her cut (H-) bang into the table.’) It demonstrates that listeners fixated the Competitor (e.g. cup) before the offset of the vowel, but quickly switched to the Target within 300 ms after the end of the vowel.

In this condition, there were comparable fixation proportions to Target, Competitor and Distracter 1 before the end of the vowel of the critical word. One-way ANOVA and planned comparisons conducted on the mean fixation proportion to the Target picture, the Competitor, and the two Distracters over the time window extending from 160 ms (mean length of vowels in this condition) to 0 ms before the offset of the vowel showed no difference for Target and Competitor \[F(1,35) = 0.328, p = 60.5\], or Target and Distracter 1 \[F(1,35) = 2.3, p = 0.13\], but difference were found between Competitor and Distracter 2 \[F(1,35) = 5.8, p = 0.02\].

As Figure 6.6 shows, fixation proportions for the Competitor were greater than those to the Target for the first 150 ms after the alignment point, which seems to suggest that listeners incorrectly perceived the non-coronal word as the Target. This may have been because, when hearing the vowel, the listeners could hear ip boundary information “H-” in this condition), suggesting that a closure would follow the word. The boundary tone in the vowel suggest that assimilation is unlikely to take place and that a non-coronal is more likely following the vowel. However, after 150 ms, when the entire information about the vowel, and, usually, a very subtle release following
Figure 6.6: English eye tracking Experiment Condition 5 (Coronal, ip Boundary): Fixation proportion over time for the Target, Competitor, and the two Distracters. Data are aligned (Point 0) from the offset of the vowel of the critical word. The dashed red line on the left indicates the probable point at which listeners’ eyes move in respond to the onset of the vowel in the critical word. The dashed red line on the right indicates the probable point at which listeners eyes move in respond to the offset of the critical word.
Table 6.6: Results of Planned Comparison between Target and Competitor means every 150 ms between 0-1500 ms after the offset of the vowel in the critical word. Condition 5, English eye tracking experiment.

<table>
<thead>
<tr>
<th>Time Window Frame</th>
<th>F-Value</th>
<th>p-Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>0-150ms</td>
<td>5.4</td>
<td>0.02</td>
</tr>
<tr>
<td>150-300ms</td>
<td>2.9</td>
<td>0.08</td>
</tr>
<tr>
<td>300-450ms</td>
<td>0.01</td>
<td>0.9</td>
</tr>
<tr>
<td>450-600ms</td>
<td>1.8</td>
<td>0.18</td>
</tr>
<tr>
<td>600-750ms</td>
<td>7.4</td>
<td>0.007</td>
</tr>
<tr>
<td>750-900ms</td>
<td>9.8</td>
<td>0.002</td>
</tr>
</tbody>
</table>

it, became available, the fixation proportion for the Competitor stopped increasing. The fixation proportion for the Competitor started to decrease at about 300ms after the offset of the vowel of the critical word. This is when additional acoustic cues such as silence after the critical word (average 90 ms in this condition), and the onset of the following context word became available. The fixation proportions for the Target started to increase at about the alignment point, increased steadily, and surpassed the fixation proportions to the Competitor at about 400 ms. Statistics in Table 6.5 show that fixation proportions for the Competitor were larger than those to the Target between 0ms and 150 ms, but were comparable to the Target between 150 ms and 600 ms. In addition, fixation proportions to the Target were greater than those to the Competitor after 600 ms (See Figure 6.11). In general, the data show that listeners considered the Competitor first when hearing coronal word followed by an ip boundary, but quickly used other acoustic cues such as formants in the vowel and the initial consonant of the context word to detect the Target.
6.5.6 Condition6 Coronal, IP Boundary

Figure 6.7 shows fixation proportions to the four AOIs: Target (e.g. cut), Competitor (e.g. cup), Distracter 1 (e.g. key), Distracter 2 (e.g. apple), for non-coronal critical words, followed by an IP boundary. Data are aligned (Point 0) from the offset of the vowel of the critical word. (e.g. 'cut' in the sentence “She was careful not to let her cut (L-H%) bang into the table.”) The figure shows that listeners fixated both the Target (e.g. “cut”) and Competitor (e.g. “cup”) before the offset of the vowel, but quickly fixated the Target only when hearing a coronal critical word (e.g. cut) followed by an IP Boundary.

In this condition, there were comparable fixation proportions to Target and Competitor before the end of the vowel of the critical word and the fixation proportions to Target and Competitor were always higher than those for Distracters 1 and 2. One-way ANOVA and planned comparisons conducted on the mean fixation proportions to the Target, the Competitor, and the two Distracter pictures over the time window extending from 250 ms (mean length of vowels in this condition) to 0ms before the offset of the vowel showed no difference for Target and Competitor \[ F(1, 35) = 0.17, \ p = 0.6 \], but difference were found between Target and Competitor vs. Distracters 1 and 2 \[ F(1, 35) = 22.3, \ p = 0.0001 \].

As Figure 6.7 shows, fixation proportions for the Competitor were comparable to the Target for the first 300ms after the alignment point, which seems to suggest that listeners were not sure if they heard coronal or non-coronal even a short while after the entire information about the vowel of the critical word was revealed. The fixation proportion for the Competitor decreased after 300ms, when the silence (average 90ms for this condition), any release for the coronal after the vowel (e.g. when it was a
Figure 6.7: English eye tracking Experiment Condition 6 (Coronal, IP Boundary): Fixation proportion over time for the Target, Competitor, and the two Distracters. Data are aligned (Point 0) from the offset of the vowel of the critical word. The dashed red line on the left indicates the probable point at which listeners’ eyes move in respond to the onset of the vowel in the critical word. The dashed red line on the right indicates the probable point at which listeners’ eyes move in respond to the offset of the critical word.
<table>
<thead>
<tr>
<th>Time Window Frame</th>
<th>F-Value</th>
<th>p-Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>0-150ms</td>
<td>0.003</td>
<td>0.9</td>
</tr>
<tr>
<td>150-300ms</td>
<td>1.1</td>
<td>0.29</td>
</tr>
<tr>
<td>300-450ms</td>
<td>2.9</td>
<td>0.08</td>
</tr>
<tr>
<td>450-600ms</td>
<td>19.2</td>
<td>0.0001</td>
</tr>
<tr>
<td>600-750ms</td>
<td>41.7</td>
<td>0.0001</td>
</tr>
<tr>
<td>750-900ms</td>
<td>36.5</td>
<td>0.0001</td>
</tr>
</tbody>
</table>

Table 6.7: Results for Planned Comparison between Target and Competitor means every 150ms between 0-1500ms after the offset of the vowel in the critical word. Condition6, English eye tracking experiment

/statistics shown in Table 6.6 indicate that fixation proportions for the Target were not higher than the Competitor until the 450 ms - 600 ms time window. In general, the data demonstrates that listeners considered both the Competitor and the Target when hearing a coronal word followed by an IP boundary, and waited to use acoustic cues from later part of the critical word, such as the release of the coronal /t/, the initial consonant in the following context word to detect the Target to resolve the ambiguity. Considering the 160 ms saccade planning time, and the silence after the critical word, which is about 100 ms, this data suggest that listeners were able to disambiguate upon hearing the onset of the following word.

6.6 Discussion and conclusion

In all conditions, listeners always excluded Distracter 2 very early, presumably because it did not share the same initial consonant as the rest of the object names. The fixation proportions for Distracter 1 also dropped quickly, usually before the
end of the vowel of the critical word. The fixation proportions for the Competitors usually became much lower than that to the Target once the ambiguity was resolved, but these never dropped to as low a level as the Distracters, indicating listeners’ awareness of the ambiguity throughout the trials.

In comparing the conditions where the critical word ended with a non-coronal (conditions 1-3), listeners showed no evidence of a perceived ambiguity for the Target word, even when a potential assimilation trigger was nearby as in the Word Boundary condition. Listeners were able to fix on the Target before the end of the vowel in the ip and IP Boundary conditions, while they made their decision by the end of the vowel in the Word Boundary condition. We suggested that the timing of these effects was probably due to the fast rate of speech and shorter length of the vowel in the Word Boundary condition.

In the conditions where the critical word ended with a coronal (conditions 4-6), the listeners experienced different levels of ambiguity. The faster the speech rate and the smaller the prosodic boundary following the critical word, the more confused the listeners were.

In Condition 4, with a Word Boundary, listeners were confused until about 1000 ms after the alignment point. In Condition 5, with an ip boundary, listeners started to fixate the Target exclusively at about 600 ms. Finally, in Condition 6, with an IP boundary, the fixation proportion for the Target became significantly higher than those to the Competitor at about 450 ms. Figure 6.14 shows that the slope of the fixation proportion line for the IP boundary condition is the sharpest among the three, indicating least ambiguity in this condition. The slope for the ip Boundary condition is intermediate between the IP Boundary and Word Boundary conditions, and the
overall trend is compatible with our contention that the Word Boundary condition is
the most confusing one, while the IP boundary condition is the least confusing.

When the coronal consonant is strongly assimilated, such as when it is followed by
a Word Boundary, which means, when it sounds more like a non-coronal, the listeners
could not distinguish it from non-coronal, and had to rely on semantic information
for disambiguation. When the coronal consonant undergoes some change, such as
when followed by an ip Boundary, listeners could use subtle acoustic cues to help
with disambiguation.
CHAPTER 7

PRETESTS FOR PUTONGHUA EYE TRACKING EXPERIMENT

In parallel with the English experiments, two pretests were conducted to prepare materials for the Putonghua eye tracking experiment to assess the strength of assimilation in the auditory stimuli and the acceptability of the visual stimuli.

7.1 Listening Pretest

7.1.1 Introduction

A pretest was conducted to test Chinese listeners’ perception of the ambiguity levels of the auditory sentences recorded from the production study. The result of this pretest also provided information about the intelligibility of the auditory stimuli for the Chinese eye tracking experiment. We also predicted that listeners would find sentences in the IP boundary conditions to be less ambiguous than those in the ip boundary conditions, which, in turn, should be less ambiguous than those in the word boundary condition.
The task for the Chinese listeners was exactly the same as that for the English listeners. They were asked to listen to the initial ambiguous sentence from the pairs in the Chinese production experiment, and then indicate whether they heard a tone1 or tone2 word, giving a rating of their certainty.

7.1.2 Method

Participants

36 native speakers of Putonghua with normal hearing voluntarily participated in this online experiment without payment. Their age was between 22 and 36. They came from different regions of China, and therefore, were exposed to different Chinese dialects. They live in different parts of the U.S. currently, and have different professions. Most of them were students or engineers.

Materials

Similar to the English experiment, auditory sentences recorded in the Putonghua production experiment were truncated with Praat at the end of the first sentence, and used as the stimuli for the current experiment.

For instance, the original sentences were:
听到那惊呼声音刺耳，他

Hear that scream sound pierce ear, he

皱了皱眉。

frown frown eyebrow

Hearing that piercing screaming sound, he frowned.

Disambiguation part

知道儿子又在欺负妹妹

Know son again be bully sister

了。

le0.

Knowing that his son is bullying his sister again.

For Tone2 ip Boundary and IP Boundary

听到京胡声音刺耳，他

Hear that instrument sound pierce ear, he

皱了皱眉。

frown frown eyebrow

Hearing that ear-piercing sound of Jing1Hu2, he frowned.
Disambiguation Part

Know son again be secretly practice instrument particle.

Knowing that his son is practicing playing that instrument secretly again.

The underlined words were the critical words, and the Word Boundary, ip Boundary and IP Boundary were inserted after the critical word. The disambiguating sentence was removed, so that the listeners only heard the potentially ambiguous sentence. *Design*

The combination of the 2 kinds of tone on the critical word and the 3 prosodic boundary types (word boundary, ip boundary, IP boundary) yielded 6 conditions for this experiment, shown in Table 7.1.

Six lists were generated. Each list contained 34 experimental (test) items. The 34 experimental items rotated in a Latin square design through the 6 conditions across the 6 lists so that in different lists, the experimental items were in different conditions. Each list was seen by 6 participants. Thus, each participant heard 34 sentence fragments - 5 or 6 items in each of the 6 conditions. WebExp2 randomized the presentation order of the items for each participant.
Table 7.1: Putonghua study experiment design

<table>
<thead>
<tr>
<th>Condition</th>
<th>Critical Word Type</th>
<th>Boundary Type</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Tone1</td>
<td>Word Boundary</td>
</tr>
<tr>
<td>2</td>
<td>Tone1</td>
<td>ip Boundary</td>
</tr>
<tr>
<td>3</td>
<td>Tone1</td>
<td>IP Boundary</td>
</tr>
<tr>
<td>4</td>
<td>Tone2</td>
<td>Word Boundary</td>
</tr>
<tr>
<td>5</td>
<td>Tone2</td>
<td>ip Boundary</td>
</tr>
<tr>
<td>6</td>
<td>Tone2</td>
<td>IP Boundary</td>
</tr>
</tbody>
</table>

Procedure

The experimental procedure was identical to that of the English listening pretest. Again, because the current version of WebExp2 does not support Chinese characters, we had to use PinYin, which contains more homophones than characters. However, given the context sentence, and the combination of the syllables, possible alternate meanings due to homophones could be ruled out easily by the listeners.

7.1.3 Results

In the raw data, “1” meant a clear tone 1, “7” meant a clear tone 2, and 4 meant unclear. To make data analysis easier, the numbers were converted as for the English experiment. For example, if what the listener heard in the recording was the sentence “ting1 dao4 na4 jing1 hu2 sheng1 yin1 ci4 er3 (Hearing that ear-piercing sound of Jing1Hu2)”, in which the intended word was T2, and he/she clicked the number “7”, which means it was a clear token of “Hu2”, in the data sheet, the number was converted to “1” (Target). However, if the listener clicked the number “1”, the number would be converted to “7” (Competitor).
Figure 7.1: Putonghua listening pretest result. Listeners’ evaluation on the ambiguity level of the critical word in 6 conditions.

As we predicted, the data showed that for both Tone1 and Tone2 words, IP Boundary conditions were perceptually less ambiguous than ip Boundary conditions, and the ip Boundary conditions are less ambiguous than the Word Boundary conditions.

In addition, in the IP boundary condition, the Tone1 word was slightly more ambiguous than the Tone2 word, while in ip Boundary and Word Boundary conditions, tone2 word were perceptually more ambiguous. This was because, in the IP Boundary condition, the Tone2 word the listeners heard were realized as a Tone2, which was perceptually not ambiguous. The Tone1 word was a little bit more ambiguous in this condition because a surface tone1 could have two underlying representations, namely, both Tone1 and Tone2. Therefore, a Tone1 is also somewhat ambiguous. However, in the ip Boundary conditions, and the Word Boundary conditions, tone2 words were
assimilated to Tone1, and were therefore ambiguous for the listeners. It seems that the listeners perceived a sandhi Tone1 as being different and more ambiguous than a real Tone1.

7.2 Picture-Word Matching Experiment

7.2.1 Introduction

Similar to the English Picture Word Matching Task, the purpose for this experiment was to pretest the visual stimuli for the eye tracking experiment. The only differences between these experiments were the subjects and the material.

7.2.2 Method

Participants

Eight native Putonghua speakers who live in different parts of the U.S. participated in this online experiment. There was no incentive for this experiment, and the participants completed the experiment out of the goodness of their hearts. Their ages were between 22 and 35, and they were of different professions, some were students, and some were computer engineers. They came from different regions of China, and spoken various dialects. Their age and dialect background were similar to the participants of the main eye tracking experiment.
Materials

Simple line drawings were made with procedures similar to those for the English visual stimuli. The size of the pictures and the organization of pictures corresponding to the critical word, its Competitor and the Distracters in the slides were also the same. An example experimental slide is shown in Figure 7.2.

![Sample trial in the Putonghua picture-word matching experiment.](image)

**Figure 7.2:** Sample trial in the Putonghua picture-word matching experiment.

Procedure

The experiment procedure was the same as that used in the English experiment.

For each trial in the experiment, there were four simple line-drawing pictures at size 700x500 pixels. Four letters “A”, “B”, “C”, “D”, were used to indicate the upper left, upper right, lower left and lower right pictures respectively. One problem
in the experiment design and implementation process was that WebExp2 currently did not support Chinese character presentations. Therefore, there was no way to present the characters in the questions. Instead, Pinyin (Romanized Chinese) and tone numbers (i.e. 1, 2, 3, 4) were used to indicate the Chinese words. However, this might create difficulty in reading for the participants, as some people might not be good at “Pinyin”, or, there might be homophones that share the same Pinyin, which might potentially confuse people. Therefore, we provided Chinese characters as part of the pictures to help participants determine which words we meant in the questions.

Chinese characters associated with each picture were presented at the bottom of the screen, and were always in a fixed order, which is Target (the Tone2 word that undergoes sandhi in the production experiment), Competitor (the Tone1 word that shares the same syllables with the Target, but a different tone on the second syllable), Distracter 1 (the second syllable of which is the same as the second syllable as the Target, but has a different tone), and Distracter2 (a word that is not associated to the Target in any sense). The fixed order of presenting the words was meant to prevent participants from predicting which tone would occur at which corner of the screen, as the Target, Competitor, Distracter1, Distracter 2 pictures rotated in positions throughout the experiment, and WebExp2 presented them randomly.

For example, in the picture shown in Figure 7.2, the characters “卷纸 (juan zhi2)” was associated to picture “A”, “风(feng1zheng1)” were associated to picture “B”, “僵直(jiang1zhi2)” was associated to “C”, “姜汁(jiang1zhi1)” was associated to letter “D”. The Chinese characters were always presented in the order of Tone2 word “僵直” (straight), Tone1 word “姜汁” (ginger juice), Distracter “卷纸” (paper roll), Distracter2 “” (kite).
Question 1 provided the four words that corresponded to the pictures in Pinyin, and asked the participants to match the pictures to the word. The participants needed to put the letters A, B, C or D in the box next to the Pinyin words. Question 2 asked the participants to use the numbers 1 to 5 to rate how well the picture matched the word, “1” means not well matched at all, “5” means perfectly matched. The participants needed to put the numbers in the box next to each of the four Pinyin words. Question 3 asked the question “Which of these could be named right away using one of the provided word?” Question 4 asked the question: Which picture could be named “xx” (the Target word), e.g. “jiang1zhi1”. After filling out all of the boxes, participants were allowed to click the “next” button to proceed to the next trial.

7.2.3 Result

For Question 1: Out of the 34 tested items, only 1 item had 2 participants who had mismatches \[1/(34 \text{ items} \times 8 \text{ subjects}=288 \text{ tokens})\], the error rate is 0.3%. 4 items had 1 participant who had mismatches; the error rate is 1.4%.

For Question 2, the average rating for the pictures corresponding to the Tone2 word (e.g. Jing1Hu2) was 3.5 on a 5-point scale, and the average rating for the pictures corresponding to the Tone1 counterpart (e.g. Jing1Hu1) was 3.4 on a 5-point scale.

For Question 3, only 4 of the slides were chosen as the words that would be named right away.

For Question 4, out of the 34 items, 5 have 4 or more than 4 wrong answers (eight “none” or a letter other than the Target letter). In terms of tokens, its 21 wrong answers out of 288 total tokens (7.3%). 8 have 3 wrong answers, this is 8.3% of the
total tokens. 7 have 2 wrong answers, which is 4.9% of the total tokens, none has 1 wrong answers. On the whole, there are 59 wrong tokens, which is 20% of the total tokens. This shows, without a sentence context, 1/5 of the chances, the subjects might experience ambiguity or uncertainty when looking for the Target picture, if they know the correct word for it.

7.2.4 Discussion

Similar to the English picture-word matching task, the result of the Chinese picture-word matching task suggested that although people might not be able to name the pictures immediately upon seeing them, and they might not feel the picture was the most typical representation of the word, in general, participants experienced no problem in matching the picture with the corresponding word correctly. Therefore, in the eye tracking experiment, given plenty of scanning time before hearing the auditory stimuli and with the help of the auditory stimuli itself, listeners should not have much problem in linking the correct picture with the intended word.
CHAPTER 8

PUTONGHUA EYE TRACKING EXPERIMENT

8.1 Introduction

Similar to the main English eye tracking experiment, the purpose of this experiment was to test listeners’ recognition of spoken words that were ambiguous because of contextual factors. Specifically, we tested native listeners’ processing of Putonghua Tone2 Sandhi and its interaction with different prosodic boundary types that were inserted between the Tone2 word that underwent sandhi and its following context word.

Four black-and-white line drawings were shown in the four corners of a computer screen. Participants’ eye movements toward these pictures were recorded as they detected which corresponded to the word they had just heard in an auditory sentence context.

We made the same hypothesis and prediction as the English experiment, except the IP Boundary conditions, where contrastive stress was used to induce the IP Boundary. We predict that the contrastive stress would cause more complicated processing and would slow down listeners’ response to the tonal information of the critical word.
Section 8.2 of this chapter presents the experiment method. Section 8.3 describes the results and section 8.4 discusses the results.

8.2 Method

8.2.1 Participants

36 native speakers of Putonghua, who had normal eyesight, normal hearing, and were between 18 and 35 years of age participated in the eye movement monitoring experiment on Putonghua Tone2 Sandhi processing. They were recruited from the Chinese community at the Ohio State University, and received $20 for participating in the experiment. Participants came from different part of China, including Beijing, Shangdong province, Hubei provising, Zhejiang province, Jiangsu province. The different dialects that each participant might have been exposed to could influence the result of the experiment. This is different from the English participants, for whom the “standard” Midwest English was their native language. However, the Chinese participants’ educational background ensured that they were at least native-like speakers of Putonghua– the modern standard Chinese spoken in Mainland China. It needs to be noted that the result of this experiment might be different from previous experiments, which used Guoyu (standard Chinese spoken in Taiwan) speakers as participants.
Table 8.1: Experimental Conditions for the Putonghua Eye Tracking Experiment

<table>
<thead>
<tr>
<th>Condition</th>
<th>Place</th>
<th>Boundary</th>
<th>Example</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>tone1</td>
<td>Word</td>
<td>ting1 dao4 na4 jing1 hu1(Word) sheng1, ta1 zhou4 le0 zhou4 mei2</td>
</tr>
<tr>
<td>2</td>
<td>tone1</td>
<td>ip</td>
<td>ting1 dao4 na4 jing1 hu1(ip) sheng1, ta1 zhou4 le0 zhou4 mei2</td>
</tr>
<tr>
<td>3</td>
<td>tone1</td>
<td>IP</td>
<td>ting1 dao4 na4 jing1 hu1(IP) sheng1, ta1 zhou4 le0 zhou4 mei2</td>
</tr>
<tr>
<td>4</td>
<td>tone2</td>
<td>Word</td>
<td>ting1 dao4 na4 jing1 hu2(Word) sheng1, ta1 zhou4 le0 zhou4 mei2</td>
</tr>
<tr>
<td>5</td>
<td>tone2</td>
<td>ip</td>
<td>ting1 dao4 na4 jing1 hu2(ip) sheng1, ta1 zhou4 le0 zhou4 mei2</td>
</tr>
<tr>
<td>6</td>
<td>tone2</td>
<td>IP</td>
<td>ting1 dao4 na4 jing1 hu2(IP) sheng1, ta1 zhou4 le0 zhou4 mei2</td>
</tr>
</tbody>
</table>

8.2.2 Design

This experiment studies the influence of both Putonghua Tone2 sandhi and prosodic boundary types on listeners’ lexical processing. We chose to investigate two tone types—tone1 and tone2 and the three boundary types—IP, ip and Word boundaries. This yielded 6 conditions for the experiment, as illustrated in Table 8.1.

6 lists were generated. Each list contained 3 practice items, 34 auditory experimental (test) items, and 34 filler items. The order of all items was the same across the lists. The 34 test items rotated in a Latin square design through the 6 conditions across the 6 lists, so that in different lists, the test items were in different conditions. The visual stimuli were the same for each item across the lists and conditions. Each
list was seen by 6 participants. Thus, each participant heard 71 sentences in total, 3 of them were practice items, 34 were fillers, and 34 were test items, 6 items in each condition. The complete list of sentences can be found in Appendix B.

8.2.3 Auditory Stimuli

The auditory stimuli of this experiment were the sound files from the Putonghua production study. Since some of the critical item files and filler files were recorded in different sessions, a Matlab program was used to normalize the intensity of all the sound files, so that they had approximately the same amplitude.

8.2.4 Visual Stimuli

The visual stimuli were prepared in the same way as the English eye movement monitoring experiment (described in Chapter 6). The four pictures in each file included: the Target and its Competitor, namely, the Tone2 word (e.g. jiang1zhi2), its Tone1 counterpart (e.g. jiang1zhi1), Distracter 1 (corresponds to a word that has the same syllable as the Target, but has either Tone3 or Tone4, e.g. juan3zhi3), Distracter 2 (corresponds to an irrelevant word, e.g. feng1zheng1), rotated in Latin square design in the list. Thus, for different trials in the same list, these pictures appeared in different corners of the screen. This was done to avoid the possibility of participants developing any expectations about where the target picture might have appeared during the experiment. In filler slides, the targets were also randomly located in different corners. To equate visual complexity, in all item (critical) slides, and 50% of the filler slides, the four pictures were chosen to have approximately the
same number of lines, and to be of roughly equal size. This was done to avoid the participants’ eye movement patterns from being influenced by the relative complexity of the line drawings.

![Sample visual stimuli](image)

Figure 8.1: Sample visual stimuli. Upper left: juan3zhi3 (rolled paper, Distracter 1), Upper right: feng1zheng1 (kite, Distracter 2), Lower left: Jiang1zhi2 (straight, Tone2 word), Lower right: Jiang1Zhi1 (ginger juice, Tone1 word)

### 8.2.5 Procedure

Participants were asked to sit in front of a 17 inch monitor at a distance of approximately 20 inches. Equipment setup and calibration procedures were identical to those described for the English experiment in Section 6.2.5.
8.3 Results

Similar to the main English experiment, for the main Putonghua experiment, fixations to all four area of interest (AOI) (Target, Competitor, Distracter 1, and Distracter 2) were automatically generated by the ASL E6000 software. A Perl program was then used to transfer the raw data to arcsine transformed fixation proportions for each AOI for all 6 conditions across participants and items, as shown in figures 1 through 6. The fixation proportions were calculated based on the proportion of total possible looks to the visual stimuli for all items and participants. The fixation proportions were calculated for every 17 ms time window, from 500 ms before the offset of the vowel to 3000 ms after the offset of the vowel in the critical word. For the word items in this experiment, all vowels were word final, so the alignment point was both the end of the vowel and the end of the word. As in the English experiment, different sized time window were used statistical analyses on data before the alignment point, because word lengths differed in the different conditions. As in the English experiment, a 150 ms fixed window size was used for on data after the alignment point, to provide a detailed picture of the patterns in the data over the time course of comprehension.

Also similar to the English experiment, a global examination of the pattern of the fixation across all 6 conditions was used to determine the extent of analyses beyond the offset of the vowel. In all conditions, listeners were able to fixate the target by 500 ms after the alignment point. However, unlike the results of the English eye tracking experiment, in which listeners' pattern of eye movements stabilized by 900 ms after offset of vowel in the critical word in all conditions except in Condition 4(Tone2, Word Boundary), the fixation proportion for the target in the Putonghua experiment
had a falling pattern around the time the first ambiguous sentence was finished and
disambiguation information became available. This is probably a reflection of partici-
pants’ realization of ambiguity, and their strategy of checking back and forth between
the target and the Competitor for the right answer constantly when listening to the
sentence. Therefore, we chose to use the same method as in the English experiment
and analyze the first 900 ms after the offset of the vowel of the critical word for all
conditions, but add another 2100 ms for Tone2 Word Boundary condition (Condition
4).

8.3.1 Condition1 Tone1, Word Boundary

Figure 8.2 shows fixation proportions to the four AOIs: Target (e.g. zhi1) Com-
petitor (e.g. zhi2), Distracter1 (e.g. zhi3), Distracter 2 (e.g. zheng1), for Tone1
critical words followed by a Word Boundary. Data are aligned (Point 0) from offset
of the vowel of the critical word. (e.g. 'zhi’ in the sentence 'ta1 xiang3 fa1 ming2 yi4
tai2 jiang1 zhi1(WB) ji1.’) The figure demonstrates that listeners fixated the target
picture after the context word that followed the critical word (e.g. “ji1” that follows
“zhi1”) when there was a Word Boundary between the two words.

Table 8.2 shows Planned Comparisons for means over the time window from -140
ms (average length of vowels in this condition) to 0 ms. The fixation proportions
for Target, Competitor, and the Distracters were comparable before the alignment
point. There were no significant differences between the Target and Competitor
\[F(1,35) = 2.1, p = 0.14\], or for the Target and Distracter 1, or Target and Distracter
2, or Competitor and Distracter 1 (all \[F(1,35) < 1\]), or Distracter 1 and Distracter
2 \[F(1,35) = 2.7, p = 0.1\].
Figure 8.2: Putonghua eye tracking experiment Condition 1 (Tone1, Word Boundary): Fixation proportion over time for the Target, Competitor, and the two Distracters. Data are aligned (Point 0) from the offset of the vowel of the critical word. The dashed red line on the left indicates the probable point at which listeners’ eyes move in respond to the onset of the vowel in the critical word. The dashed red line on the right indicates the probable point at which listeners eyes move in respond to the offset of the critical word.
The fixation proportion for the Target and Competitor continued to be statistically comparable until 600 ms (See Table 8.1). Considering the 150 ms saccade planning time, and the 134 ms average length of the following context word, the data suggest that listeners were able to distinguish the target when the context word that followed the critical word was already available. The Tone1 word was ambiguous to the listeners probably because it occurs as the result of Tone2 sandhi very often, which makes the listeners to consider both Tone1 and Tone2 possibilities although what they heard was a Tone1 word. After 600 ms, the fixation proportion for the target (Tone1 word) increased significantly, while those for the Competitor (Tone2 word) stayed at the same level. This suggests that listeners were still able to use acoustic information to distinguish the Target, only with longer processing time.

<table>
<thead>
<tr>
<th>Time Window Frame</th>
<th>F-Value</th>
<th>p-Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>0-150ms</td>
<td>2.4</td>
<td>0.11</td>
</tr>
<tr>
<td>150-300ms</td>
<td>1.9</td>
<td>0.16</td>
</tr>
<tr>
<td>300-450ms</td>
<td>0.23</td>
<td>0.6</td>
</tr>
<tr>
<td>450-600ms</td>
<td>1.45</td>
<td>0.22</td>
</tr>
<tr>
<td>600-750ms</td>
<td>7.3</td>
<td>0.0079</td>
</tr>
<tr>
<td>750-900ms</td>
<td>20.9</td>
<td>0.0001</td>
</tr>
</tbody>
</table>

Table 8.2: Results for Planned Comparisons between Target and Competitor in 150 ms, 6 windows from vowel offset. Condition 1, Putonghua eye tracking study.
8.3.2 Condition 2 Tone1, ip Boundary

Figure 8.3 shows fixation proportions to the four areas of interest: Target (e.g. zhi1), Competitor (e.g. zhi2), Distracter1 (e.g. zhi3), Distracter2 (e.g. zheng1), for Tone1 critical words followed by an ip Boundary. Data are aligned (Point 0) from the offset of the vowel of the critical word (e.g. ‘zhi1’ in the sentence ‘ta1 xiang3 fa1 ming2 yi4 tai2 jiang1 zhi1 ji1 qi4.’) The figure shows that listeners fixated the Target immediately after hearing the context word that followed the critical word (e.g. “zhi1”) when there was an ip Boundary between the two.

As Figure 8.3 and planned comparison over the time window from -200 ms (average length of vowels in this condition) to 0 ms show, the fixation proportions for Target and Competitor were comparable \[F(1, 35) < 1\] before the alignment point, and those for Distracter 1 and Distracter2 were comparable as well \[F(1, 35) < 1\], but there were a significant difference between Target and Competitor vs. Distracter 1 and 2 \[F(1, 35) = 14.4, p = 0.0002\].

The fixation proportion for the Target and Competitor continued to be comparable until 450 ms (See Table 8.3), but as shown in the figure, the fixation proportion for the Competitor started to drop at around 300 ms. (The Planned Comparison for the 300-450 ms window yields the marginal result of \[F(1, 35) = 3, p = 0.08\] )This is also when the fixation proportion for the Target rose above those for the Competitor. An alternative analyses using 300 ms as window size shows that the fixation proportion to the Target were already significantly more than those to the Competitor \([F(1, 35) = 5.6, p = 0.002]\). Considering the 150 ms saccade planning time, the data suggest that listeners were able to distinguish the Target even before hearing the end of the context word that followed the critical word. The processing time was shorter
Figure 8.3: Putonghua eye tracking experiment Condition 2 (T1, ip Boundary): Fixation proportion over time for the Target, Competitor, and the two Distracters. Data are aligned (Point 0) from the offset of the vowel of the critical word. The dashed red line on the left indicates the probable point at which listeners’ eyes move in respond to the onset of the vowel in the critical word. The dashed red line on the right indicates the probable point at which listeners eyes move in respond to the offset of the critical word.
<table>
<thead>
<tr>
<th>Time Window Frame</th>
<th>F-Value</th>
<th>p-Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>0-150ms</td>
<td>0.003</td>
<td>0.95</td>
</tr>
<tr>
<td>150-300ms</td>
<td>1.125</td>
<td>0.72</td>
</tr>
<tr>
<td>300-450ms</td>
<td>3.0</td>
<td>0.08</td>
</tr>
<tr>
<td>450-600ms</td>
<td>8.11</td>
<td>0.005</td>
</tr>
<tr>
<td>600-750ms</td>
<td>14.4</td>
<td>0.0002</td>
</tr>
<tr>
<td>750-900ms</td>
<td>20.9</td>
<td>0.0001</td>
</tr>
</tbody>
</table>

Table 8.3: Results for Planned Comparisons between Target and Competitor in 150 ms, 6 windows from vowel offset. Condition 2, Putonghua eye tracking study

in this condition that in Condition 1. This suggests that Tone1, when followed by an ip Boundary, was still somewhat ambiguous, and listeners used information such as word length (see Chapter 4) and boundary tone to help with disambiguation. That is, when the listeners heard an ip Boundary tone, they predicted that Tone2 sandhi was unlikely to be implemented, therefore, the surface Tone1 that they heard had a very big probability of being the intended tone.

### 8.3.3 Condition 3 Tone1, IP Boundary

Figure 8.4 shows fixation proportions to the four AOIs: Target (e.g. zhi1), Competitor (e.g. zhi2), Distracter 1 (e.g. zhi3), Distracter 2 (e.g. zheng1), for Tone1 critical words followed by an IP Boundary. Data are aligned (Point 0) from the offset of the vowel of the critical word. (e.g. ’zhi1’ in the sentence ’ta1 xiang3 fa1 ming2
The data demonstrated that listeners fixated the Target immediately after hearing the context word that followed the critical word (e.g. “zhi1”) when there was an IP Boundary between the two.

As Figure 8.4 and Planned Comparison over the time window from -350 ms (average length of vowels in this condition) to 0 ms show, the fixation proportions for Target and Competitor, Target and Distracter 1, Target and Distracter 2, and Distracter 1 and Distracter 2 were comparable (Target vs. Competitor $F(1, 35) = 3.7, p = 0.056$, all other $F(1, 35) < 1$) before the alignment point.

The fixation proportion for the Competitor was higher than that for the Target in the first 150 ms. Taking into consideration of the 150 ms saccade planning time, the first 150 ms after the alignment point reflect listeners’ processing when they were listening to the critical word. Remember that the speaker used a contrastive stress on the critical word in order to produce the IP Boundary. Thus, we hypothesize that the high initial fixation proportion for the Competitor might be due to listeners’ response to the contrastive stress. After 150 ms, the fixation proportion for the Target and Competitor became comparable until 450 ms (See Table 8.3) according to statistical analysis. However, as shown in Figure 8.4, the fixation proportion for the Competitor started to drop at around 200 ms, and the fixation proportion for the Target rose above those for the Competitor at about 300 ms. Considering the 200 ms average length of the following context word, the data suggest that listeners were able to distinguish the Target immediately after hearing the context word that followed the critical word. This is very similar to the results of Condition 2, but the fixation proportion to the Competitor in this condition decreased about 100 ms earlier than that in Condition 2. This suggests that Tone 1, even when followed by
Figure 8.4: Putonghua eye tracking Experiment Condition 3 (Tone1, IP Boundary): Fixation proportion over time for the Target, Competitor, and the two Distracters. Data are aligned (Point 0) from the offset of the vowel of the critical word. The dashed red line on the left indicates the probable point at which listeners’ eyes move in respond to the onset of the vowel in the critical word. The dashed red line on the right indicates the probable point at which listeners eyes move in respond to the offset of the critical word.
Table 8.4: Results for Planned Comparisons between Target and Competitor in 150 ms, 6 windows from vowel offset. Condition 3, Putonghua eye tracking study.

<table>
<thead>
<tr>
<th>Time Window Frame</th>
<th>F-Value</th>
<th>p-Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>0-150ms</td>
<td>4.06</td>
<td>0.046</td>
</tr>
<tr>
<td>150-300ms</td>
<td>1.3</td>
<td>0.25</td>
</tr>
<tr>
<td>300-450ms</td>
<td>3.3</td>
<td>0.069</td>
</tr>
<tr>
<td>450-600ms</td>
<td>11.4</td>
<td>0.001</td>
</tr>
<tr>
<td>600-750ms</td>
<td>20.6</td>
<td>0.0001</td>
</tr>
<tr>
<td>750-900ms</td>
<td>30.0</td>
<td>0.0001</td>
</tr>
</tbody>
</table>

an IP Boundary, was still somewhat ambiguous, and listeners still used information such as word length and boundary tone to help with disambiguation.

8.3.4 Condition 4- Tone2, Word Boundary

Figure 8.5 shows fixation proportions to the four AOIs: Target (e.g. zhi2), Competitor (e.g. zhi1), Distracter 1 (e.g. zhi3), Distracter 2 (e.g. zheng1), for Tone2 critical words followed by a Word Boundary. Data were aligned (Point 0) from the offset of the vowel of the critical word. (e.g. 'zhi2' in the sentence 'ta1 xiang3 fa1 ming2 yi4 tai2 jiang1 zhi2 ji1 qi4.') The data demonstrated that listeners fixated the Target immediately when hearing the critical word (e.g. “zhi2”), but experienced ambiguity at the end of the critical word, and had to rely on the semantic context in the second sentence in the auditory stimuli to help with disambiguation.

As Figure 8.5 and Planned Comparison over the time window from -140 ms (average length of vowels in this condition) to 0 ms show, the fixation proportions for Target and Competitor, Target and Distracter 1, Competitor and Distracter 1, Competitor
Figure 8.5: Putonghua Eye Tracking Experiment Condition 4 (Tone2, Word Boundary): Fixation proportion over time for the Target, Competitor, and the two Distracters. Data are aligned (Point 0) from the offset of the vowel of the critical word. The dashed red line on the left indicates the probable point at which listeners’ eyes move in respond to the onset of the vowel in the critical word. The dashed red line on the right indicates the probable point at which listeners eyes move in respond to the offset of the critical word.
and Distracter 2, and Distracter 1 and Distracter 2 were comparable \(F(1, 35) = 1.6, p = 0.2\), \(F(1, 35) = 3.2, p = 0.07\), \(F(1, 35) < 1\), \(F(1, 35) = 1.1, p = 0.29\), \(F(1, 35) < 1\), respectively, before the alignment point.

As the statistics in Table 8.4 indicate, the fixation proportion for the Target were higher than those for the Competitor in the first 150 ms time window (or, in the first 600 ms time window, if use 300 ms as the fixed window size for analyses), but then fell back to be comparable until 1650 ms after the offset of the vowel of the critical word. The fixation proportion to the Target started to be significantly higher than those to the Competitor at about 2100 ms after the offset of the vowel. Considering the mean length of the first sentence after the alignment point (987 ms), and the mean length of the second sentence after the alignment point (2661 ms), the data suggests that listeners had to rely on semantic information in the second sentence to help with disambiguation. It is difficult to explain why the fixation proportion for the Target was briefly higher than those for the Competitor in the first 150 ms window after the alignment point. It might simply be a random effect, but could also be that listeners were actually able to detect the slightly rising F0 in the stimuli and sense that the Target was a Tone2, however, the acoustic cue was not strong enough for them to be sure that it was not a Tone1.

8.3.5 Condition 5- Tone2, ip Boundary

Figure 8.6 shows fixation proportions to the four area-of-interest: Target (e.g. zhi2), Competitor (e.g. zhi1), Distracter 1 (e.g. zhi3), Distracter 2 (e.g. zheng1), for Tone2 critical words followed by an ip Boundary. Data are aligned (Point 0) from the offset of the vowel of the critical word. (e.g. 'zhi2' in the sentence 'ta1 xiang3
Figure 8.6: Putonghua Eye Tacking Experiment Condition5 (Tone2, ip Boundary): Fixation proportion over time for the Target, Competitor, and the two Distracters. Data are aligned (Point 0) from the offset of the vowel of the critical word. The dashed red line on the left indicates the probable point at which listeners’ eyes move in respond to the onset of the vowel in the critical word. The dashed red line on the right indicates the probable point at which listeners eyes move in respond to the offset of the critical word.
The data demonstrated that listeners fixated the Target immediately after hearing the critical word (e.g. “zhi2”).

As Figure 8.6 and Planned Comparisons over the time window from -140ms (average length of vowels in this condition) to 0 ms showed, the fixation proportions for Target and Competitor, Target and Distracter 1, Competitor and Distracter 1, and Distracter 1 and Distracter 2 were comparable \[ F(1, 35) < 1, F(1, 35) = 2.9, p = 0.09, \]
\[ F(1, 35) = 2.7, p = 0.1, F(1, 35) = 1.1, p = 0.28, \] respectively, before the alignment point. The fixation proportion for the Target were comparable with those for the Competitor in the first 150 ms time window, but was significantly higher than the fixation proportion for the Competitor immediately after 150 ms (See Table 8.6). Considering the 150 ms planning time, and the 193 ms average vowel length of this condition, the data indicated that the listeners started to fixate the Target immediately upon hearing the end of the vowel in this condition. This seems to suggest that Tone2 word, when following by an IP Boundary, was not ambiguous for the listeners.

### 8.3.6 Condition 6- Tone2, IP Boundary

Figure 8.7 shows fixation proportions to the four AOIs: Target (e.g. zhi2), Competitor (e.g. zhi1), Distracter 1 (e.g. zhi3), Distracter 2 (e.g. zheng1), for Tone2 critical words followed by an IP Boundary. Data were aligned (Point 0) from the offset of the vowel of the critical word. (e.g. 'zhi2' in the sentence 'ta1 xiang3 fa1 ming2 yi4 tai2 jiang1 zhi2 ji1 q4.') The data demonstrated that listeners fixated the Target immediately after hearing the critical word (e.g. “zhi2”).

As Figure 8.7 and Planned Comparisons over the time window from -350 ms (average length of vowels in this condition) to 0 ms show, the fixation proportions for Target
Figure 8.7: Putonghua eye tracking Experiment Condition 6 (Tone2, IP Boundary): Fixation proportion over time for the Target, Competitor, and the two Distracters. Data are aligned (Point 0) from the offset of the vowel of the critical word. The dashed red line on the left indicates the probable point at which listeners’ eyes move in respond to the onset of the vowel in the critical word. The dashed red line on the right indicates the probable point at which listeners eyes move in respond to the offset of the critical word.
and Competitor, Target and Distracter 2, Competitor and Distracter1, Competitor and Distracter 2, and Distracter 1 and Distracter 2 were comparable \( [F(1,35) < 1] \), \( [F(1,35) = 2.5, p = 0.11] \), \( [F(1,35) = 2.7, p = 0.1] \), \( [F(1,35) = 1.1, p = 0.28] \), \( [F(1,35) < 1] \), respectively, before the alignment point. The fixation proportion for the Target was comparable with those for the Competitor in the first 300 ms time window, but was significantly higher than the fixation proportion for the Competitor immediately after 300 ms (See Table 8.7). Considering the 150 ms planning time, and the 193ms average vowel length of this condition, the data indicates that the listeners started to fixate the Target immediately upon hearing the end of the vowel in this condition. This seems to suggest that Tone2 word, when following by an IP Boundary, was not ambiguous for the listeners.

### 8.4 Discussion

The data shows that when the critical word is followed by a Word Boundary, listeners always experience ambiguity, no matter the critical word is Tone1 or Tone2. Tone1 is ambiguous when followed by a Word Boundary because this is the environment that Tone2 Sandhi most likely applies. However, listeners could still distinct the Target without the help of semantic context in the second sentence in this condition. Tone2 words followed by a Word Boundary is much more ambiguous for the listeners, although they sensed the correct Target initially, but the weak acoustic information in the auditory stimuli was not strong enough for them. Listeners had to rely on the semantic information in the second sentence to decide the correct Target. When hearing a critical word followed by an ip Boundary (Condition 2 and 5), listeners were able to distinct the Target very quickly. When they heard Tone1 in Condition
2, the likelihood of what they heard could be an underlying Tone2 made them hesitated and prolonged their processing. As a result, they made their decision until after hearing the following context word. However, when they heard Tone2 in Condition5, listeners made their decision very quickly, immediately after the critical word is heard. A critical word followed by an IP Boundary was expected to be the least ambiguous conditions (Condition3 and Condition 6). However, the contrastive stress on the critical word cued listeners to consider the Competitor a little longer, which caused the disambiguation time point for Condition 3 to be the same as Condition 2, and Condition 6 to be a little longer than Condition 5, which was the least ambiguous condition among the 6.

On the whole, the data suggests that listeners can use F0 (as shown in Condition 4) and boundary information (as shown in Condition 3 and 6) very early in processing, and language specific information, such as the likelihood of sandhi also have a large influence on their processing.
<table>
<thead>
<tr>
<th>Time Window Frame</th>
<th>F-Value</th>
<th>p-Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>0-150ms</td>
<td>7.1</td>
<td>0.009</td>
</tr>
<tr>
<td>150-300ms</td>
<td>1.3</td>
<td>0.25</td>
</tr>
<tr>
<td>300-450ms</td>
<td>3.3</td>
<td>0.069</td>
</tr>
<tr>
<td>450-600ms</td>
<td>2.4</td>
<td>0.12</td>
</tr>
<tr>
<td>600-750ms</td>
<td>0.057</td>
<td>0.875</td>
</tr>
<tr>
<td>750-900ms</td>
<td>0.094</td>
<td>0.75</td>
</tr>
<tr>
<td>900-1050ms</td>
<td>1.8</td>
<td>0.17</td>
</tr>
<tr>
<td>1050-1200ms</td>
<td>0.094</td>
<td>0.75</td>
</tr>
<tr>
<td>1200-1350ms</td>
<td>2.85</td>
<td>0.09</td>
</tr>
<tr>
<td>1350-1500ms</td>
<td>2.0</td>
<td>0.15</td>
</tr>
<tr>
<td>1500-1650ms</td>
<td>2.98</td>
<td>0.08</td>
</tr>
<tr>
<td>1650-1800ms</td>
<td>6.26</td>
<td>0.01</td>
</tr>
<tr>
<td>1800-1950ms</td>
<td>5.4</td>
<td>0.02</td>
</tr>
<tr>
<td>1950-2100ms</td>
<td>0.9</td>
<td>0.33</td>
</tr>
<tr>
<td>2100-2250ms</td>
<td>4.7</td>
<td>0.03</td>
</tr>
<tr>
<td>2250-2400ms</td>
<td>12.3</td>
<td>0.0005</td>
</tr>
<tr>
<td>2400-2550ms</td>
<td>22.9</td>
<td>0.0001</td>
</tr>
<tr>
<td>2550-2700ms</td>
<td>14.1</td>
<td>0.0003</td>
</tr>
<tr>
<td>2700-2850ms</td>
<td>23.5</td>
<td>0.0001</td>
</tr>
<tr>
<td>2850-3000ms</td>
<td>38.9</td>
<td>0.0001</td>
</tr>
</tbody>
</table>

Table 8.5: Results for Planned Comparisons between Target and Competitor in 150 ms, 15 windows from vowel offset. Condition 4, Putonghua eye tracking study
<table>
<thead>
<tr>
<th>Time Window Frame</th>
<th>F-Value</th>
<th>p-Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>0-150ms</td>
<td>0.19</td>
<td>0.66</td>
</tr>
<tr>
<td>150-300ms</td>
<td>4.3</td>
<td>0.03</td>
</tr>
<tr>
<td>300-450ms</td>
<td>11.7</td>
<td>0.0009</td>
</tr>
<tr>
<td>450-600ms</td>
<td>18.9</td>
<td>0.0001</td>
</tr>
<tr>
<td>600-750ms</td>
<td>12.1</td>
<td>0.0007</td>
</tr>
<tr>
<td>750-900ms</td>
<td>18.1</td>
<td>0.0001</td>
</tr>
</tbody>
</table>

Table 8.6: Results for Planned Comparisons between Target and Competitor in 150 ms, 6 windows from vowel offset. Condition 5, Putonghua eye tracking experiment

<table>
<thead>
<tr>
<th>Time Window Frame</th>
<th>F-Value</th>
<th>p-Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>0-150ms</td>
<td>2.5</td>
<td>0.11</td>
</tr>
<tr>
<td>150-300ms</td>
<td>3.8</td>
<td>0.05</td>
</tr>
<tr>
<td>300-450ms</td>
<td>11.7</td>
<td>0.0009</td>
</tr>
<tr>
<td>450-600ms</td>
<td>23.9</td>
<td>0.0001</td>
</tr>
<tr>
<td>600-750ms</td>
<td>32.8</td>
<td>0.0001</td>
</tr>
<tr>
<td>750-900ms</td>
<td>20.3</td>
<td>0.0001</td>
</tr>
</tbody>
</table>

Table 8.7: Results for Planned Comparisons between Target and Competitor in 150 ms, 6 windows from vowel offset. Condition 6, Putonghua eye tracking experiment
CHAPTER 9

CONCLUSION AND DISCUSSION

This chapter summarize and compare the two eye tracking experiments, and discusses this study’s implication to lexical processing.

9.1 Experiment Result

9.1.1 Comparison between each condition

Condition1- Non-Coronal or Tone1 Critical Word, Word Boundary

In the English experiment, the fixation proportion to the Target and the Target and Competitor did not differ until about 300 ms after the offset of the vowel according to the statistical analysis (See Table6.1). In the Putonghua experiment, the fixation proportion for the Target and Competitor continue to be comparable until 600ms (See Table8.1). Considering the 150ms saccade planning time, and the 134ms average length of the following context word, the data suggest that listeners were able to distinguish the Target when the context word that follows the critical word was already available.
Both the non-coronal and the Tone1 word seems to be temporarily ambiguous to the listeners when followed by a Word Boundary, although they did not went through coronal or tonal assimilation. This is probably because the traces of assimilated sound in the mental lexicon is very similar to these non-assimilated sounds. During retrieval, both options were active. However, after the following contextual information become available, listeners were able to use statistic probability information, together with subtle acoustic information in the production (more details were presented in Chapter 4) to distinguish the Target, only with longer processing time.

**Condition 2: Non-Coronal or Tone1 Critical Word, ip Boundary**

In the English experiment, the fixation proportions for Target and Competitor already differ between 0 ms and 150 ms after the offset of the vowel (see Table 6.2). This suggests that the fixation to the Target began even before they had processed the completed information about vowels in the critical word in the Non-coronal ip Boundary condition. This is probably because the H- boundary tone caused rising F0 in the vowel, and listeners were able to make use of this boundary cue to detect the ip boundary following the word even before the vowel was completed. Since assimilation is less likely across an ip boundary, the listeners predicted that what they heard was a non-coronal word based on this probability in English and the acoustic information for the non-coronal word. In the Chinese experiment, the fixation proportion for the Target and Competitor continue to be comparable until 450 ms (see Table 8.2). The data suggest that listeners were able to distinguish the Target immediately after hearing the context word that follows the critical word. This is similar to the results of Condition 1, but the processing time was much shorter. This suggests that Tone1,
when followed by an ip Boundary, was still somewhat ambiguous, and listeners used information such as word length (See Chapter 4) and boundary tone to help with disambiguation.

It seems that in Condition 2, the Putonghua listeners experienced more ambiguity than the English listeners. This is probably due to the shorter length of the critical word in the Putonghua stimuli, which not only shortened the processing time, but also reduced the articulation gesture in temporal realm, and made the stimuli resembles assimilated Tone2. In both experiments, the ip Boundary tone helped listeners to process the critical word faster than in Condition 1.

**Condition 3- Non-Coronal or Tone1 Critical Word, IP Boundary**

In the English experiment, according to statistical analyses listed in Table 6.3, the fixation proportions to Target and Competitor diverged between 150 ms and 300 ms. This suggests that listeners’ fixation to the Target began even before they processed the completed information about vowels in the Non-coronal IP Boundary condition. Similar to Condition 2, this is probably because the L-H* boundary tone caused listeners to be able to detect the IP boundary type following the word even before the vowel was completed. Since the duration of the vowel in this condition was much longer than the vowel in Condition 2 and Condition 1, it was more likely that listeners used the length of the vowel information to determine that there would be an IP Boundary following the word, and predicted that assimilation would not take place, and therefore, recognized the Target correctly. In this condition, listeners fixated the Target a little later than in Condition2. This is probably due to “semantic wrap up at the IP Boundary position” as suggested by Schafer (1997). In the Chinese
experiment, the fixation proportion for the Competitor was higher than those for the Target in the first 150 ms, which reflects listeners’ processing when they were listening to the critical word. The high initial fixation proportion for the Competitor might due to listeners’ response to the contrastive stress—they were not sure why there is a stress there, and looked at the Tone2 Competitor. This can serve as an evidence for them to actually notice the correct Target at the very beginning of processing. After 150 ms, the fixation proportion for the Target and Competitor became comparable until 450 ms (See Table 8.3) according to statistical analysis. The data suggests that listeners were able to distinguish the Target immediately after hearing the context word that follows the critical word. This is very similar to the results of Condition2, but the fixation proportion to the Competitor in this condition decreased about 100 ms earlier than that in Condition 2. This suggests that Tone1, even when followed by an IP Boundary, was still somewhat ambiguous, and listeners still used information such as word length and boundary tone to help with disambiguation.

In condition 3, the critical word seems not to be ambiguous for the listeners, but they were slowed down than the ip Boundary condition because of the IP Boundary tone. The Putonghua listeners seem to be able to distinguish the correct Target, but confused by the contrastive stress from where the IP Boundary tone derived, and had prolonged processing time.

**Condition 4- Coronal or Tone2 Critical Word, Word Boundary**

In the English experiment, according to statistical analyses (see Table 6.4) the fixation proportion to the Target was comparable with those to the Competitor until about 1000 ms after the alignment point. This indicates that listeners experienced
ambiguity when listening to a coronal word following by a Word Boundary. They had to use semantic information available much later in the context sentence to help with the disambiguation task. In the Putonghua experiment, the statistical analyses summarized in Table 8.4 indicates that the fixation proportion for the Target were higher than those for the Competitor in the first 150 ms time window, but then fall back to be comparable until 1650 ms after the offset of the vowel of the critical word. The fixation proportion to the Target started to be significantly higher than those to the Competitor at about 2100 ms after the offset of the vowel. It is difficult to explain why the fixation proportion for the Target was briefly higher than those for the Competitor in the first 150 ms window after the alignment point. It might simply be a random effect, but could also be that listeners were actually able to detect the slightly rising F0 in the stimuli and sense that the Target was a Tone2, however, the acoustic cue was not strong enough for them to be sure that it was not a Tone1.

In this condition, both English and Putonghua experienced ambiguity and had to rely on semantic information in the context for disambiguation.

**Condition 5- Coronal or Tone2 Critical Word, ip Boundary**

In the English experiment, as Table 6.5 demonstrates, fixation proportions for the Competitor was larger than the Target between 0ms and 150 ms, but was comparable to the Target between 150 ms and 600 ms, and the fixation proportions for the Target was greater than those for the Competitor after 600 ms (See Figure 6.11). The data demonstrates that listeners considered the Competitor first when hearing coronal word followed by an ip boundary, but quickly used other acoustic cues such as formants in the vowel, the initial consonant in the context word to detect the
Target. In the Putonghua experiment, Table 8.5 shows that the fixation proportion for the Target were comparable with those for the Competitor in the first 150 ms time window, but was significantly higher than the fixation proportion for the Competitor immediately after 150 ms (See Table 5). The data indicates that the listeners started to fixate the Target immediately upon hearing the end of the vowel in this condition. This seems to suggest that Tone2 word, when following by an ip Boundary, was not ambiguous for the listeners.

In this condition, the results for the two experiments seem to diverge. In the English experiment, listeners relied on the ip Boundary cue to make the top-down hypothesis that the critical word was non-coronal first, but quickly used acoustic information toward the end of the word to correct the wrong hypothesis. However, Putonghua listeners seemed to be able to distinguish the Target immediately. This is probably because in the English experiment, the acoustic information for the “place” of the last segment was available later than the suprasegmental boundary information, while Tonal information in the Putonghua study was available earlier.

**Condition 6- Coronal or Tone2 Critical Word, IP Boundary**

In the English experiment, as Figure 6.7 shows, fixation proportions for the Competitor were comparable to the Target for the first 300 ms after the alignment point, which seems to suggest that listeners were not sure if they heard coronal or non-coronal even a short while after the entire information about the vowel of the critical word was revealed. However, after 300 ms, when the silence (average 90 ms for this condition), and maybe the release for the coronal after the vowel when it is a /t/
sound, and even the initial consonant of the following word became available, the fixation proportion for the Competitor decreased. Statistically, as illustrated by Table 6.6, fixation proportions for the Target was not higher than the Competitor until the 450 ms - 600 ms time window (See Table 6.6). In general, the data demonstrates that listeners considered both the Competitor and the Target first when hearing coronal word followed by an IP boundary, but quickly used acoustic cues from later part of the critical word, such as the release of the coronal /t/, the initial consonant in the following context word to detect the Target. In the Putonghua experiment, the fixation proportion for the Target was comparable with those for the Competitor also in the first 300 ms time window, but was significantly higher than the fixation proportion for the Competitor immediately after 300 ms (See Table 8.6). Considering the 150 ms planning time, and the 193ms average vowel length of this condition, the data indicates that the listeners started to fixate the Target immediately upon hearing the end of the vowel in this condition. This seems to suggest that Tone2 word, when following by an IP Boundary, was not ambiguous for the listeners.

In this condition, listeners of both languages waited until the end of the word, or immediately at the onset of the context word, to fixate the Target. This is not faster than condition 5. This could be explained by the same reason as condition 3, that more semantic information were involved in processing.
9.1.2 Comparison between boundary types and discussion on segmental vs. suprasegmental context effect

We predicted that critical words in IP Boundary condition should be less ambiguous than those in the ip Boundary condition, which was in turn less ambiguous than the Word Boundary condition. However, we only considered the acoustic information, but not the semantic information involved in the boundary tone. It seems that since prosodic boundary information was available earlier than the segmental information in the critical word and its context word in the English experiment, it was also used earlier in processing.

9.1.3 Comparison between critical word types and discussion on segmental vs. suprasegmental assimilation

It seems that in most of the conditions, if not consider the duration difference in the stimuli, the segmental and suprasegmental assimilation yield very similar results. For example, both changed and unchanged sounds are ambiguous (in contrast to all other previous accounts). The ambiguity was resolved sooner, as the prosodic strength is bigger. However, similar to the boundary information, the tonal assimilation information could be available earlier for processing than the segmental assimilation information, as indicated by results of condition 5.

9.2 Implication to Processing

9.2.1 Previous accounts

The phonological inference account could not explain why non-coronal and Tone1, the utterances contained were ambiguous for a short period of time. The feature cue parsing account and the perceptual integration account would predict that listeners
won’t experience ambiguity for coronal Word Boundary and Tone2 Word Boundary conditions, and they won’t need the semantic context for disambiguation, which is not true for the current experiment results.

### 9.2.2 Current account

Our revised exemplar model should be able to explain the result: First of all, the non-coronal and Tone1 were ambiguous, especially in the Word Boundary condition, for a short period of time, because they were very similar to the changed sound (assimilated coronal or Tone2) that are stored in the memory. Second, upon hearing coronal and Tone2 in Word Boundary conditions, since the input were similar to both coronal/noncoronal and Tone1/Tone2 that are stored in memory, listeners were not able to distinguish them without semantic context. In the case of Tone2 Word Boundary condition, there were some tokens that were different from Tone1, and maybe that was why listeners were able to look at the Target at the very beginning. However, since most of the tokens were similar to Tone1, and almost all of them were short and reduced, it could be similar to Tone1 as well. That was why listeners were still confused. The fact that listeners relied on boundary cues to disambiguate non-coronal ip Boundary condition, and coronal ip Boundary condition at a very early stage was also evidence of exemplar comparison and extractions.

### 9.3 Conclusion

The two eye tracking experiments yield a clear pattern of results. For non-coronal and Tone1 Targets, the Word Boundary conditions were ambiguous for the listeners when they first encounter the critical word, probably due to the short duration of
the word and the reduction of the stimuli. However, after the context segment is revealed, the ambiguity was resolved. The ip Boundary conditions were less ambiguous than the Word Boundary condition. Listeners seem to be able to use the boundary information to infer that phonological change is not likely to take place across the ip Boundary, and take the surface form of the stimuli as its underlying form. The IP Boundary conditions took a little longer for listeners to process than the ip Boundary conditions, either because of “end of prosodic group wrap up” (in the English case) or the contrastive stress we put on the critical word to produce the IP Boundary (in the Putonghua case).

For the coronal and tone2 Targets, the Word Boundary was the most ambiguous. Listeners usually need semantic context from the second sentence in the stimuli to help with disambiguation. Listeners still experienced a certain degree of ambiguity in the ip Boundary conditions. In the IP Boundary conditions, the critical word themselves were not ambiguous, but listeners spent a little longer time than we expected before they fixated the Target for the same reason as non-coronal/Tone1 IP Boundary conditions as mentioned previously.

The proposed revised exemplar account could explain the result from the experiment in a reasonable way, and it combines the strong points of the previous models.
APPENDIX A

COMPLETE SENTENCES FOR THE ENGLISH EXPERIMENTS

1. He drew the line perfectly. (LINE-LIME)(D1-light, D2-glasses)
   Line: I don’t know how he made it so straight.
   Lime: I almost want to pick it up and smell it.

2. He picked up the cone between his fingers.(CONE-COMB) (D1-kite, D2-bottle)
   Cone: and he put some ice cream in it.
   Comb: and started to detangle her hair.

3. She didn’t know how to fix the hen properly.(HEN-HEM)(D1-heart, D2-triangle)
   Hen: since she doesn’t know how to cook.
   Hem: since she doesn’t know how to sew.

4. He wanted to ask what section of the natural history museum the loon belonged in,(LOON-LOOM)(D1-lamp D2-beast)
   Loon: since he didn’t know much about birds.
   Loom: since he didn’t know much about weaving.
5. She wasn’t sure if she could ever do the scene properly, (Scene-Seam) (D1-street, D2-newspaper)

   Scene: therefore, she really appreciated the director’s encouragement.

   Seam: therefore, she really appreciated the tailor’s help.

6. The teen patiently awaited the start of the game, (Teen-Team) (D1-tiger, D2-computer)

   Teen: while she ate her popcorn.

   Team: and they listened carefully to the coach.

7. She put the phone beside her tools on the workbench, (Phone-Foam) (D1-fire, D2-gazebo)

   Phone: so that she can hear it ring.

   Foam: as she prepared to install the insulation.

8. The man watched as the Dane passed him on the street, (Dane-Dame) (D1-donkey, D2-monkey)

   Dane: and was wondering if he would bark.

   Dame: and tried to get her number.

9. The sun proceeded to rise at a steady pace, (Sun-Sum) (D1-sausage, D2-purpose)

   Sun: scattering all the clouds around it.

   Sum: the numbers seem to be adding up.

10. He wasn’t sure who the mat belonged to, (Mat-Map) (D1-mail, D2-box)
Mat: but he took it to Yoga class anyway.

Map: but he used it to find his way around the city anyway.

11. The neighbor knew who the cat belonged to. (CAT-CAP) (D1-cake, D2-loop)

Cat: but he fed it for several days before returning it.

Cap: but he wore it for several days before returning it.

12. She was careful not to let her cut bang into the table. (CUT-CUP) (D1-key, D2-apple)

Cut: so that it wouldn’t hurt again.

Cup: so that it wouldn’t get broken.

13. This time she tried to get the right berries for her pie. (RIGHT-RIPE) (D1-rob, D2-pear)

Right: not the wrong ones that she used last time.

Ripe: not the sour ones that she used last time.

14. The heat blew in the kitchen when they opened the door. (HEAT-HEAP) (D1-hair, D2-pear)

Heat: so they turned on the fan to cool down the room.

Heap: so they spent the next two hours picking up the trash.

15. She didn’t like the beat behind the lyrics of the song. (BEAT-BEEP) (D1-bus, D2-chair)

Beat: which was too fast to dance to.

Beep: which was high pitched and annoying.
16. She stood next to the cot beside the window. (COT-COP) (D1-coke, D2-fish)
   Cot: wondering if she should fold it up and put it away.
   Cop: wondering if she should talk to him or not.

17. His hit became a big topic of conversation. (HIT-HIP) (D1-ham, D2-plant)
   Hit: and everyone thinks he is a great musician now.
   Hip: because he had to undergo replacement surgery soon.

18. His rat became well-known. (RAT-RAP) (D1-rice, D2-socks)
   Rat: since it liked apples instead of cheese.
   Rap: since it got a lot of radio air time.

19. They hadn’t had any sleet before tonight. (SLEET-SLEEP) (D1-swim, D2-trashcan)
   Sleet: people went home early to avoid the dangerous driving condition.
   Sleep: the baby had cried every night since she coming home from the hospital.

20. His fan glimmered in the moonlight. (FAN-FANG) (D1-flag, D2-cissors)
   Fan: and it was moving slowly, squeaking.
   Fang: as the wolf smiled in anticipation.

21. She continued to clean gallantly until other people arrived. (CLEAN-CLING) (D1-king, D2-football)
   Clean: everyone was amazed by how much work she has done.
   Cling: everyone was amazed by how long she held on.
22. His bat ground into the dirt. (BAT-BACK)(D1-baby, D2-frame)

Bat: after he threw it down and ran for first base.

Back: it took him a long time to stand up.

23. Unfortunately he broke his net going for the ball. (NET-NECK)(D1-nose, D2-elephant)

Net: The tear may be hard to fix.

Neck: He had to wear a brace for 6 months.

24. He had a nightmare that the debt grew larger by the minute. (DEBT-DECK)(D1-duck, D2-window)

Debt: and someone threatened to kill him if he couldn’t pay the money back.

Deck: which quickly overtook his backyard and surrounding area.

25. They had the lot guarded by a high tech security system. (LOT-LOCK)(D1-lake, D2-rock)

Lot: since it’s private property that belonged to the president.

Lock: so that no one could break in to the safe of the bank.

26. The knot grew frustrating. (KNOT-KNOCK)(D1-note, D2-lamb)

Knot: She had forgotten how to undo it.

Knock: so she reluctantly opened the door.

27. His height gave him hope. (HEIGHT-HIKE)(D1-hand, D2-door)

Height: he might be able to make it into the NBA.
Hike: exercising helps him think more positively.

28. The bite gouged her leg. (BITE-BIKE)(D1-bee, D2-notebook)

   Bite: She couldn’t tell what bit her though.

   Bike: but she still got back on and rode home.

29. As a rule the wait could go on for hours. (WAIT-WAKE)(D1-waistcoat, D2-tree)

   Wait: but people stood in line patiently, since they were used to it.

   Wake: but people still went to comfort the family of the deceased.

30. The state gets pretty hot. (STATE-STEAK)(D1-stick, D2-leaf)

   State: since it is almost as far south as Florida.

   Steak: you will burn it if you aren’t careful.

31. She doesn’t light gas stoves. (LIGHT-LIKE)(D1-weave, D2-camp)

   Light: because she’s afraid to use matches.

   Like: she prefers electric stoves instead.

32. She was sure she had seen the kit go in. (KIT-KICK)(D1-closet, D2-blinder)

   kit: but the tools were nowhere to be found when she needed them.

   kick: but the judge somehow didn’t see it.

33. She almost let the spite kill her. (SPITE-SPIKE)(D1-spoon, D2-house)

   Spite: but then she decided to forgive him.

   Spike: but she quickly moved out of its way and avoided being impaled.
34. The clot grew harder to see. (CLOT-CLOCK)(D1-clown, D2-net)

Clot: It seemed that the treatment had been pretty effective.

Clock: As it became too dark in the room to see the time.

35. They asked him to seat grandfather outside. (SEAT-SEEK)(D1-sea, D2-plane)

Seat: because the fresh air would be good for the old man.

Seek: because they needed to ask him a question and they couldn’t find him in the house.

36. She should have a pat go on her back. (PAT-PACK)(D1-pig, D2-lamp)

Pat: She has been working very hard on this project.

Pack: It is so inconvenient to carry the paper bags by hands.
APPENDIX B

SAMPLE SENTENCES FOR THE PUTONGHUA EXPERIMENTS

1. 关于他有鼻炎平，这事很
   guan1 yu2 ta1 you3 bi2 yan2 ping2, zhe4 shi4 hen3
   about he has rhinitis pill, this thing very
   少有 人知道。他不希望别
   shao3 you3 ren2 zhi1 dao4 xi1 wang4 bie2
   few has person know, he not hope other
   人知道他有这病。
   ren2 zhi1 dao4 ta1 you3 zhe4 bin4
   people know he has this sickness.

   Few people know that he has rhinitis pills. He doesn’t want other people to
   know he has this problem.

2. 关于他有鼻烟瓶，这事很
   guan1 yu2 ta1 you3 bi2 yan1 ping2, zhe4 shi4 hen3
   about he has snuff bottle, this thing very
   少有 人知道。他不希望别
   shao3 you3 ren2 zhi1 dao4 bu4 xi1 wang4 bie2
   few has people know, he not wish other
   人知道他爱收藏。
   ren2 zhi1 dao4 ta1 ai4 shou1 cang2
   people know he like collection.

   Few people know that he has a snuff bottle; he doesn’t want other people know
   he likes collecting them.
3. 关于他有鼻炎，平常这事
关于 he has rhinitis, in general this thing
很少有人知道。他不希望
很少有人知道。他不希望
很多人知道他有这病。
其他人知道他有这病。
很少有人知道。他不希望
很少有人知道。他不希望

Few people know that he has rhinitis. He doesn’t want other people to know he has this problem.

4. 关于他有鼻烟瓶子这事
关于 he has snuff bottle this thing
很少有人知道。他不希望
很少有人知道。他不希望
其他人知道他爱收藏。
其他人知道他爱收藏。

Few people know that he has a snuff bottle; he doesn’t want other people to know he likes collecting them.

5. 看来必须要等到加时期，
看来必须等到加时期，
才能够解决问题了。比赛双方
才能够解决问题了。比赛双方

It seems that we will have to wait until overtime (before we can tell who will win). The two sides in the game are well matched in strength.
6. 看来必须要等到加湿期，

It seems one has to wait until the humidifying period before the problem can be solved. Static electricity will be reduced when the humidity level is higher than 50 degrees.

7. 看来必须要等到加时期

It seems that we have to wait until overtime (before we can tell who will win). The two sides in the game are well matched in strength.

8. 看来必须要等到加湿期

It seems one has to wait until the humidifying period before the problem can be solved. Static electricity will be reduced when the humidity level is higher
than 50 degrees.

9. 我叫她别提它，她偏不听

I told her not to lift it, but she just wouldn’t listen. As a result, she sprained her back.

10. 我叫她别踢它，她偏不听

I told her not to kick it, but she just wouldn’t listen. As a result, she hurt her foot.

11. 我叫她别提，她偏不听话。

I told her not to lift it, but she just wouldn’t listen. As a result, she sprained her back.

12. 我叫她别踢，她偏不听话。

I told her not to kick it, but she just wouldn’t listen. As a result, she sprained her back.
I told her not to kick it, but she just wouldn’t listen. As a result, she hurt her foot.

13. 前面好像 是兵船群，而不是
   3. qian2 mian4 hao3 xiang4 shi4 bing1 chuan2 qun2, er2 bu4
   ahead seems is warship group but not
   是小岛。应该 是总部派来
   shi4 xiao3 dao. ying1 gai shi4 zong3 bu4 pai4 lai2
   be island must be headquarter send
   救援我们的。
   jiu4 yuan2 wo3 men2 de0.
   rescue us particle

It seems to be a group of warships ahead, not an island. They should be sent by the headquarters to rescue us.

14. 前面好像 是冰川群，而不是
   3. qian2 mian4 hao3 xiang4 shi4 bing1 chuan1 qun2, er2 bu4
   ahead seems be iceberg group but not
   是小岛。小心别撞到它
   shi4 xiao3 dao3. xiao3 xin1 bie3 zhuang4 dao4 ta1
   be island careful not hit it
   的水下部分。
   de0 shui3 xia4 bu4 fen.
   's water lower part

It seems to be a group of icebergs ahead, not an island. Be careful not to hit the part of it that is underwater.

15. 前面好像 是兵船群体，而不是
   3. qian2 mian4 hao3 xiang4 shi4 bing1 chuan1 qun2 ti3, er2
   ahead seems be warship group but not
   不是小岛。应该是总部派
   bu4 shi4 xiao3 dao3. ying1 gai1 shi4 zong3 bu4 pai4
   not be island must be headquarter send
   来救援我们的。
   lai2 jiu4 yuan2 wo3 men0 de.
   here rescue us particle

188
It seems to be a group of warships ahead, not an island. They should be sent by the headquarters to rescue us.

16. 前面好像是冰川群体，而 ahead seems be iceberg group but
不 是小岛。小心别撞到它
not be island careful not hit it
们的水下部分。
pl. ‘s water below part

It seems to be a group of icebergs ahead, not an island. Be careful not to hit the part of it that is underwater.

17. 这么多选择中，他唯独对
pl. ‘s water below part

Of all the choices, he was only interested in the shipping department. Ever since childhood he dreamed of designing 10-thousand-ton ships.

18. 这么多选择中，他唯独对
pl. ‘s water below part

Of all the choices, he was only interested in the communications department. He dreamed of becoming a broadcaster.
Among all the choices, he was only interested in the communications department. He always wanted to be an announcer.

19. 这么多选择中，他唯独对
船船系科感兴趣。他从小
就梦想设计万吨轮。
then want design 10thousandship

Of all the choices, he was only interested in the shipping department. Since childhood he dreamed of designing 10-thousand-ton ships.

20. 这么多选择中，他唯独对
传播系科感兴趣。他从小
就想当播音员。
then want be anouncer

Of all the choices, he was only interested in the communications department. He always wanted to be an announcer.

21. 这块地会被分成两
部分，当中垒一道土墙。
part middle build one earthenwall

This land will be separated into two parts, by an earthen wall in the middle.
22. 这块地会被分割成两部
zhe4 kuai4 di4 hui4 bei4 fen1 ge2 cheng2 liang3 bu4
this land will be separated into 2 parts
分。大部分归给邻国。
fen4. da4 bu4 fen4 gui1 gei3 lin2 guo2.
piece big part give neighboring country

This land will be separated into two parts. The bigger part will be given to the neighboring country.

23. 这块地会被分隔，成为两
zhe4 kuai4 di4 hui4 bei4 fen1 ge2 cheng2 wei2 liang3
this land will be separate become 2
部分。当中垒一道土墙。
bu4 fen4. dang1 zhong1 lei3 yi2 dao4 tu3 qiang2.
piece middle build one earthen wall

This land will be separated into two parts, by an earthen wall in the middle.

24. 这块地会被分割，成为两
zhe4 kuai4 di4 hui4 bei4 fen1 ge2 cheng2 liang3
this land will be separated become 2
部分。大部分归给邻国。
bu4 fen4. da4 bu4 fen4 gui1 gei3 lin2 guo2.
piece big part give neighboring country

This land will be separated into two parts. The bigger part will be given to the neighboring country.

25. 在一阵高呼声中她登场
zai4 yi1 zhen4 gao1 hu2 sheng1 zhong1 ta1 deng1 chang3
in one gauge string instrument enter stage
了。清亮的琴声称得她英
le0. qing1 liang4 de0 qin2 sheng1 chen4 de2 ta1 ying1
dashing and spirited
姿飒双。
z1 sa4 shuang3.
she valiant and spirited
She entered the stage to the sound of Gaohu (a kind of stringed instrument). The clear and bright sound of the instrument made her look dashing and spirited.

26. 在一阵高呼声中她登场 
zhai4 yi1 zhen4 gao1 hu1 sheng1 zhong1 ta1 deng1 chang3
in one scream sound in she enter stage

She entered the stage to the sound of screaming. The audience got more and more excited.

27. 在一阵高胡，生歌唱中她登场 
zhai4 yi1 zhen4 gao1 hu2 sheng1 ge1 zhong1 ta1 deng1
in one gaohu , sheng song in she enter stage

She entered the stage to the sound of Gaohu (a kind of stringed instrument), sheng (another kind of stringed instrument) and song. The clear and bright sound of the instrument made her look dashing and spirited.

28. 在一阵高呼，生歌唱中她登场 
zhai4 yi1 zhen4 gao1 hu2 sheng1 ge1 zhong1 ta1 deng1
in one scream sound in she enter stage

She entered the stage to the sound of screaming. The audience got more and more excited.
She entered the stage to the sound of screaming, sheng (a kind of string instrument) and song. The audience got more and more excited.

29. 你要是为提琴来, 我这里

ni3 yao4 shi4 wei4 ti2 qin2 lai, wo3 zhe4 li3
you if for violin come i here

就实话告诉你, 我修不了。
jiu4 shi2 hua4 gao4 su4 ni3, wo3 xiu1 bu4 liao3.
them truth tell you i repair not particle

If you came for the violin, I would like to tell you the truth; I can’t repair it.

30. 你要是为提亲来, 我这里

ni3 yao4 shi4 wei4 ti2 qin1 lai, wo3 zhe4 li3
you if for matchmaking i here

就实话告诉你, 我不结婚。
jiu4 shi2 hua4 gao4 su4 ni3, wo3 bu4 jie2 hun1.
them truth tell you i not marry

If you are here for matchmaking, I would like to tell you the truth; I don’t want to get married.

31. 你要是为提琴, 来我这里

ni3 yao4 shi4 wei4 ti2 qin2, lai2 wo3 zhe4 li3
you if for violin come i here

就实话告诉你, 我修不了。
jiu4 shi2 hua4 gao4 su4 ni3, wo3 xiu1 bu4 liao3.
them truth tell you i repair not particle

If you came for the violin, I would like to tell you the truth; I can’t repair it.

32. 你要是为提亲, 来我这里

ni3 yao4 shi4 wei4 ti2 qin1, lai2 wo3 zhe4 li3
you if for matchmaking come i here

就实话告诉你, 我不结婚。
jiu4 shi2 hua4 gao4 su4 ni3, wo3 bu4 jie2 hun1.
them truth tell you i not marry
f you are here for matchmaking, I would like to tell you the truth; I don’t want to get married.

33. 她 觉 得 那 司 仪 秀 很 有 意
   ta1 jue2 de2 na4 si1 yi2 xiu4 hen3 you3 yi4
   she feel that host show very have interest
思。 5 号 选 手 很 有 智 慧。
   si1. 5 hao4 xuan3 shou3 hen3 you3 zhi4 hui4.
   interest5 number participant very has wisdom

She thought that host show was very interesting. Participant number five had a lot of wisdom.

34. 她 觉 得 那 丝 衣 秀 很 有 意
   ta1 jue2 de2 na4 si1 yi1 xiu4 hen3 you3 yi4
   she feel that silk clothes show very has interest
思。 5 号 衬 衣 非 常 飘 逸。
   si1. 5 hao4 chen4 yi1 fei1 chang2 piao1 yi4.
   interest5 number shirt very elegant

She thought that show about silk clothes was very interesting. Shirt number five was very elegant.

35. 她 觉 得 那 司 仪, 秀 美, 而 且
   ta1 jue2 de2 na4 si1 yi2 xiu4 mei3, er2 qie3
   she feel that host elegant and
   hen3 you3 zhi4 hui4.
   very has wisdom

She thought that host was very elegant and had a lot of wisdom.

36. 她 觉 得 那 丝 衣, 秀 美, 而 且
   ta1 jue2 de2 na4 si1 yi1, xiu4 mei3, er2 qie3
   she feel that silk clothes elegant and
   he2 shen1 you3 piao1 yi4.
   fit body and elegant

194
She thought that silk clothing was very elegant and fit the body well.

37. 听到那些京胡声音，他皱了皱眉。知道儿子又在偷偷练习。

He frowned upon hearing the piercing sound of the JingHu (a kind of stringed instrument), knowing his son was practicing it secretly.

38. 听到那些惊呼声，他皱了皱眉。知道儿子又在欺负妹妹。

He frowned upon hearing the sound of screaming, knowing his son was bullying his sister again.

39. 听到那些京胡声音刺耳，他皱了皱眉。知道儿子又在偷偷练习琴。

He frowned upon hearing the sound of piercing, knowing his son was secretly practicing the violin again.
He frowned upon hearing the piercing sound of the JingHu (a kind of stringd instrument), knowing his son was practicing it secretly.

40. 听到 那 惊呼 声音 刺耳，他
他
皱了 皱眉。知道 儿子 又 在
欺 负 妹妹 了。

He frowned upon hearing the sound of screaming, knowing his son was bullying his sister again.

41. 你来 负责 这 次 的 连环 计，
你来 负责 这 次 的 连环 计，
You will be in charge of this set of interlocking stratagems. Assign a job to each individual person. Make sure there is no mistake in any link.

42. 你来 负责 这 次 的 联欢 计，
你来 负责 这 次 的 联欢 计，

You will be in charge of this “get-together” stratagem. Assign a job to each individual person. Make sure there is no mistake in each program.

43. 你来负责这次的连环计
ni3 lai2 fu4 ze2 zhe4 ci4 de0 lian2 huan2 ji4,
you come in charge this time particlechain ring stratagem
划。一个环节都不能有错。
hua4. yi2 ge4 jie2 mu4 dou1 bu4 neng2 you3 cuo4.
strategem one ring all not can have wrong

You will be in charge of this set of interlocking stratagems. Assign a job to each individual person. Make sure there is no mistake in any link.

44. 你来负责这次的联欢计
ni3 lai2 fu4 ze2 zhe4 ci4 de0 lian2 huan1 ji4
you come in charge this time particleget together stratagem
划。一个节目都不能有错。
hua4. yi2 ge4 jie2 mu4 dou1 bu4 neng2 you3 cuo4.
strategem one program all not can have wrong

You will be in charge of this “get-together” stratagem. Assign a job to each individual person. Make sure there is no mistake in each program.

45. 他对那次毛皮展，不太满意
ta1 dui4 na4 ci4 mao2 pi2 zhan3, bu2 tai4 man3
he to that time fur exhibition very satisfied
意。说没见到紫雕皮。
yi4. shuo1 mei2 jian4 dao4 zi3 diao1 pi2.
satisfied not see sable fur

He was not happy with that fur exhibition, saying that he didn’t see sable fur.

46. 他对那次毛坯展，不太满意
ta1 dui4 na4 ci4 mao2 pi1 zhan3, bu2 tai4 man3
he to that time semi finished exhibition very satisfied
意。说砖块质量太差。
yi4. shuo1 zhuan1 kuai4 zhi4 liang4 tai4 cha4.
satisfied break quality too poor
He was not happy with that exhibition of the semi finished product, saying that the quarlity of the break was too poor.

47. 他 对 那 次 毛 皮 展 览, 不 太
he to that time fur exhibition not very
满 意。 说 没 见 到 紫 雕 皮。
satisfied say not see sable fur

He was not happy with that fur exhibition, saying that he didn’t see sable fur.

48. 他 对 那 次 毛 坯 展 览, 不 太
he to that time semi finished exhibition not very
满 意。 说 砖 块 质 量 太 差。
satisfied say break quality too poor

He was not happy with that exhibition of the semi finished product, saying that the quarlity of the break was too poor.

49. 那 个 收 银 机 坏 了。 难 怪 怎
that cash register bad particleno wonder how
么 也 算 不 准 余 款。
how particlecalculate not accurate remaining money

That cash register is not working. No wonder it cannot calculate correctly.

50. 那 个 收 音 机 坏 了。 难 怪 怎
that radio bad particleno wonder how
么 也 听 不 到 音 乐。
how particlehear not music

198
That radio was broken. No wonder we couldn’t hear the music.

That cash register is not working. No wonder it cannot calculate correctly.

That radio was broken. No wonder we couldn’t hear the music.

The problem now is that, unlike what we thought previously, the bookcase is bigger. Originally, I thought the closet was bigger.
The problem now is that, unlike we thought previously, the output is larger. Originally, I thought the input was larger.

55. 现在的问题是书橱大小
now particle problem be bookcase size

和我们预想的不一样。原
and we predict particle not same original

以为衣柜更大一些。
thought closet more big a little

The problem now is that, the size of the bookcase is unlike what we thought previously. Originally, I thought the closet was bigger.

56. 现在的问题是输出大小
now particle problem be output size

和我们预想的不一样。原
and us predict particle not same original

以为输入更大一些。
thought input more big a little

The problem now is that, the output amount is unlike we thought previously. Originally, I thought the input was larger.

57. 这条街拐角那边那家熟食
this street corner side that cooked food
The bistro at the corner of this street is very well known, their smoked fish is very tasty.

58. 这 条 街 拐 角 边 那 家 师
this street corner side that teacher
馆，非 常 有 名。开 餐 厅 的 老 师
restaurant very has name teacher
很 严 格。
very strict

The tutoring place at the corner of this street is very well known, the owner is very strict.

59. 这 条 街 拐 角 边 那 家 熟 食
this street corner side that cooked food
馆，非 常 有 名。做 的 煙 魚
restaurant very has name smoked fish
t 非 常 好 吃。
very good eat

The bistro at the corner of this street is very well known, their smoked fish is very tasty.

60. 这 条 街 拐 角 边 那 家 师
The tutoring place at the corner of this street is very well known, the owner is very strict.

The picture of the rock was drawn so vividly that I can’t help wanting to touch the moss on it.

The picture of the strict teacher in front of us was drawn so vividly that I can’t help remembering the math teacher at the primary school.
The picture of the rock was drawn so vividly that I can’t help wanting to touch the moss on it.

64. 面前的严肃的图画，画得维妙维肖。我不住想起了一面墙上绿色的青苔。

The picture of the strict teacher in front of us was drawn so vividly that I can’t help remembering the math teacher at the primary school.

65. 那个偏房里，没有什么秘密。警察都已经搜查过了。

There should not be any secrets in that side room. The police have already searched it.

66. 那个偏方里，没有什么秘密。

There should not be any secrets in that side room. The police have already searched it.
There should not be any secret in that folk prescription. Pharmacists have already done research on it.

67. 那个偏房，里面没有什么

na4 ge0 pian1 fang2, li3 mian4 mei2 you3 shen2 me0
that side room in no any

秘密。警察都已经搜索过了。
mi4 mi4. jing3 cha2 dou1 yi3 jing1 sou1 cha2 guo4le0.
secret police all already search has particle

There should not be any secrets in that side room. The police have already searched it.

68. 那个偏方，里面没有什么秘密

na4 ge0 pian1 fang2, li3 mian4 mei2 shen2 me0 mi4
that folk prescription in no secret

秘密。药学家都已经研究过了。
mi4. yao4 xue2 jia1 dou1 yi3 jing1 yan2 jiu1 guo4le0.
secret pharmacist all already research has particle

There should not be any secret in that folk prescription. Pharmacists have already done research on it.

69. 他在前庭里，缓缓地足度着

ta1 zai4 qian2 ting2 li3. huan3 di0 duo2 zhe0
he at front patio in slowly particlewalk -ing

步。随手摘了朵梅花。
bu4. shun4 shou3 zhai1 le0 duo2 mei2 hua1.
step convenient pick particle plum blossom

He walked slowly on the front patio, picking a plum blossom as he went.

70. 他在前厅里，缓缓地足度着

ta1 zai4 qian2 ting2 li3. huan3 huan3 di0 duo2 zhe0
he at front room in slowly particlewalk -ing

步。随手关上了窗户。
bu4. shun4 shou3 guan1 shang4 le0 chuangu4.

204
He walked slowly in the living room, closing a window as he went.

He walked slowly on the front patio, picking a plum blossom as he went.

He walked slowly in the living room, closing a window as he went.

Everyone knows that, after looking for a job, that one needs to ask other people for information, as each company’s requirements are different.
74. 每个人在求知识过后都
mei3 ge4 ren2 zai4 qiu2 zhi1 guo4 yi3 hou4 dou1
了解要多向别人请教。
zh1 dao4 yao4 duo1 xiang4 bie2 ren2 qing3 jiao4.
每一个人都有自己的领域。
ge4 ren2 dou1 you3 zi4 ji3 shan4 chang2 de0 lin3
yu4
dou1 you3 zi4 ji3 shan4 chang2 de0 lin3
field

Everyone knows that, after seeking knowledge, that one needs to ask other people’s advice. Everyone has his/her own field of specialty.

75. 每个人在求职过程中都
mei3 ge4 ren2 zai4 qiu2 zhi1 guo4 yi3 hou4 dou1
需要更多向他人请教。每个公司
yao4 duo1 xiang4 bie2 ren2 qing3 jiao4. mei3 ge4 gong1
都要有不同的要求。
si1 zha0 ren2 de0 yao1 qiu2 dou1 bu4 tong2.
compani
tive person particlerequirement all not same

Everyone knows that, after looking for a job, that one needs to ask other people for information, as each company’s requirements are different.

76. 每个人在求知过程中都
mei3 ge4 ren2 zai4 qiu2 zhi1 guo4 yi3 hou4 dou1
需要更多向他人请教。每个人
yao4 duo1 xiang4 bie2 ren2 qing3 jiao4. mei3 ge4 ren2
duoyoudou1 you3 zi4 ji3 shan4 chang2 de0 lin3 yu4.
all has self good at particlefield

206
everyone knows that, after seeking knowledge, that one needs to ask other people's advice. Everyone has his/her own field of specialty.

77. 我 在 你 们 的 言 谈 上，发 现
wo3 zai4 ni3 men0 de0 yan2 tan2 shang4,fa1 xian4
I at your pl. particletalk on find
了 很 多 问 题。最 重 要 的 是
le0 hen3 duo1 wen4 ti2. zui4 zhong4 yao4 de0 shi4
particlevery lot question most important particlebe
dui4 lao3 shi1 bu4 gou4 zun1 zhong4.
to teacher not enough respect

I found a lot of problems in your wording. The most important one was that you did not show enough respect to your teacher.

78. 我 在 你 们 的 盐 滩 上，发 现
wo3 zai4 ni3 men0 de0 yan1 tan2 shang4,fa1 xian4
I on your pl. particlesalt beach on find
了 很 多 问 题。最 重 要 的 是
le0 hen3 duo1 wen4 ti2. zui4 zhong4 yao4 de0 shi4
particlevery lot question most important particlebe
有 海 水 带 来 的 垃 圾。
you3 hai3 shui3 dai4 lai2 de0 lai1 ji1
has see water bring come particle trash

I found a lot of problems in your sea salt beach. The most important one was the trash brought by the sea.

79. 我 在 你 们 的 言 谈 上 面，发
wo3 zai4 ni3 men0 de0 yan2 tan2 shang4 mian4, fa1
i at your pl. particletalk on find
现 了 很 多 问 题。最 重 要 的
xian4 le0 hen3 duo1 wen4 ti2. zui4 zhong4 yao4 de0
find particlevery lot problem most important particle
是 对 老 师 不 够 尊 重。
shi4 dui4 lao3 shi1 bu4 gou4 zun1 zhong4.
to teacher not enough respect

be to teacher not enough respect
I found a lot of problems in your wording. The most important one was that you did not show enough respect to your teacher.

80. 我 在 你 们 的 盐 滩 上 面, 发 现 了 很 多 问 题。 最 重 要 的 是 有 海 水 带 来 的 垃 圾。

I found a lot of problems in your sea salt beach. The most important one was the trash brought by the sea.

81. 请 注意 分 辨 每 个 音 节 间

Please try to discern the differences between the syllables. Listen, some are plosives, some are fricatives.

82. 请 注意 分 辨 每 个 音 阶 间

Compare that particle sound high particle half degree
Please try to discern the differences between the scales. Listen, this sound is a half degree higher than that one.

83. 请 注意 分辨 每个 音节 间
qing3 zhu4 yi4 fen1 bian4 mei3 ge4 yin1 jie2 jian1
please careful discern every particlesytable between

隔 不同 的 地方。 你 听，有
ge2 bu4 tong2 de0 di4 fang1. ni3 ting1, you3
 seperanot same particleplace you listen some

的 是 爆 破 音， 有的 是 摩 擦
de0 shi4 bao4 po4 yin1, you3 de0 shi4 mo2 ca1
particlebe plosive sound some be fricate

音。
yin1.
sound

Please try to discern the differences between the syllables. Listen, some are plosives, some are fricatives.

84. 请 注意 分辨 每个 音阶 间
qing3 zhu4 yi4 fen1 bian4 mei3 ge4 yin1 jie1 jian1
please careful discern every particlescale between

隔 不同 的 地方。 你 听，这 个
ge2 bu4 tong2 de0 di4 fang1. ni3 ting1, zhe4 ge4
 seperanot same particleplace you listen this particle

音 比 那 个 音 高 了 半度。
yin1 bi3 na4 ge4 yin1 gao1 le0 ban4 du4.
sound comparthat particlesound high particlehalf degree

Please try to discern the differences between the sound. Listen, this sound is a half degree higher than that one.

85. 传说 中 的 神秘 源 泉 照 就
chuan2 shuo1 zhong1 de0 shen2 mi4 yuan2 quan2 zha04 jiu4
legend in particlemysterious spring picture right

在这里。 你看，泉水 比 地面
zai4 zhe4 li3. ni3 kan4, quan2 shui3 bi3 di4 mian4
at here you see spring water compare earth face

高 了 1 个 公分。

209
Here is the picture of the mysterious spring in the legend. Look, the spring is 1 centimeter higher than the ground.

86. 传说中的神秘圆圈照片就在这里。你看，发光的圆形物体好像飞碟。

Here is the picture of the mysterious circle in the legend. Look, the round object that gives off light looks like a flying disk (UFO).

87. 传说中的神秘源泉，照片就在这里。你看，泉水比地面上高了1个公分。

Here is the picture of the mysterious spring in the legend. Look, the spring is 1 centimeter higher than the ground.

88. 传说中的神秘圆圈，照片就在这里。发光的圆形物体好像飞碟。

Here is the picture of the mysterious spring in the legend. Look, the spring is 1 centimeter higher than the ground.
Here is the picture of the mysterious circle in the legend. Look, the round object that gives off light looks like a flying disk (UFO).

89. 他开了个酒吧叫“休闲林”

He opened a pub called “Leisure Woods Inn”. Lots of busy people love to go there to rest.

90. 他开了个酒吧叫“修仙林”

He opened a house called “Immortality Cultivation Wood Inn”. Lots of busy people long for the leisurely life on the world of the immortals.

91. 他开了个酒吧叫“休闲林”

He opened a pub called “Leisure Woods Inn”. Lots of busy people love to go there to rest.
He opened a pub called “Leisure Woods and Garden Inn”. Lots of busy people love to go there to rest.

He opened a house called “Immortality Cultivation Wood Inn”. Lots of busy people long for the leisurely life on the world of the immortals.

He suddenly found there was something special on that wall of flowers. What he found was that there were seven colors among the flowers.
He suddenly found there was something special in that coloratura. What he found was that there were seven complicated changes in the tone.

He suddenly found there was something special on that wall of flowers. What he found was that there were seven colors among the flowers.
He suddenly found there was something special in that coloratura. What he found was that there were seven complicated changes in the tone.

97. 你好 好 想 想，这 么 多 的 平

你 hard hard think think this many particle flat
房 数，一天里 清 点 完 太 难
方 数，一天里 清 点 完 太 难了。

思 考 还 是 先 把 高 楼 数 点 完 吧。

first paticle high building count finish

Think about it, there are so many ranch-style buildings. It would be difficult to finish counting them in one day. It is better to count high rise buildings first.

98. 你好 好 想 想，这 么 多 的 平

方 数，一天里 清 点 完 太 难了。

还 是 写 个 程 序 来 处 理 吧。

still be write one program come process particle

Think about it, there are so many square numbers. It would be difficult to finish counting them in one day. It is better to write a program to process it.

99. 你好 好 想 想，这 么 多 的 平

你 hard hard think think this many particle flat
房 数 目，一天里 清 点 完 太 难

还 是 先 把 高 楼 数 点 完 吧。

first particle high building count finish

Think about it, there are so many square numbers. It would be difficult to finish counting them in one day. It is better to write a program to process it.
Think about it, there are so many ranch-style buildings. It would be difficult to finish counting them in one day. It is better to count high rise buildings first.

Think about it, there are so many square numbers. It would be difficult to finish counting them in one day. It is better to write a program to process it.

What is the use of spending so much time on knitting it. The child has never likes wearing scarf.

What is the use of spending so much time on knitting it. The child has never likes wearing scarf.
What is the use of spending so much time on supporting it. The child never likes to live in the tent.

103. 你 花 了 那 么 多 的 时 间 来
你 spend particle that many particle time come
ni3 hua1 le0 na4 me0 duo1 de0 shi2 jian1 lai2

What is the use of spending so much time on knitting them? The children have never liked wearing scarves.

104. 你 花 了 那 么 多 的 时 间 来
you spend particle that many particle time come
ni3 hua1 le0 na4 me0 duo1 de0 shi2 jian1 lai2

What is the use of spending so much time on supporting it. The children have never liked living in the tent.

105. 他 一 生 的 大 多 数 时 间 都
ta1 yi4 sheng1 de0 da4 duo1 shu4 shi2 jian1 dou1
花 花 在 杀 伐 上。我不 喜 欢。我
he one life particle big more numbertime all

216
He spent most of his life on killing and warring. I don’t like it. I like those great people who promote world peace.

He spent most of his life on the sofa. I don’t like it. I like those people who love outdoor activities.
He spent most of his life on killing and warring. I don’t like it. I like those great people who promote world peace.

107. 他 一 生 的 大 多 数 时 间 都 
ta1 yi4 sheng1 de0 da4 du01 shu4 shi2 jian1 dou1 
He one life particle big more numbertime all
花 在 沙 发 上 面。 我 喜 欢 那 
hua1 zai4 sha1 fa1 shang4 mian4. wo3 xi3 huan1 na4 
spend at sofa on i like that
些 喜 欢 室 外 活 动 的 人 们。 
xie1 xi3 huan1 shi4 wai4 dong4 de0 ren2 men0. 
pl. like room out activity particle people

He spent most of his life on the sofa. I don’t like it. I like those people who love outdoor activities.

108. 我 听 说 这 些 红 薯 是 蒸 熟 
wo3 ting1 shuo1 zhe4 xie1 hong2 shu3 shi4 zheng1 shu2 
I heard talk this pl. red yam is steam cooked
过， 还 被 卖 过 后， 剩 下 来 的。 
guo4, hai3 bei4 mai4 guo4 hou4, shen4 xia4 lai2 de0. 
has and be sell has after remaining particle
难 怪 已 经 凉 了。 
nan2 guai4 yi3 jing1 liang2 le0. 
no wonder already cold particle

I heard that these red yams were the remaining few of a bunch that was steamed and sold. No wonder they are cold now.

109. 我 听 说 这 些 红 薯 是 收 
wo3 ting1 shuo1 zhe4 xie1 hong2 shu3 shi4 zheng1 shou1 
I hear talk this pl. red yam be collect
过， 还 被 卖 过 后， 剩 下 来 的。 
guo4, hai3 bei4 mai4 guo4 hou4, shen4 xia4 lai2 de0. 
has and be sell has after remaining particle
难 怪 已 经 发 芽 了。 
nan2 guai4 yi3 jing1 fa1 ya2 le0. 
no wonder already sprout particle
I heard that these red yams are the remaining few of a bunch that was collected and sold. No wonder I can see sprouts on them.

110. 我 听 说 这 些 红 薯 是 蒸 熟 
wo3 ting1 shuo1 zhe4 xie1 hong2 shu3 shi4 zheng1 shou2
I hear talk this pl. red yam be steam cooked
过 后，还 被 卖 过 后，剩 下 来 
guo4 hou4, hai2 bei4 mai4 guo4 hou4, shen4 xia4 lai2
has after and be sell has after remain
do1. nan2 guai4 yi3 jing1 liang2 le0.
particle no wonder already cold particle

I heard that these red yams are the remaining few of a bunch that was collected and sold. No wonder I can see sprouts on them.

111. 我 听 说 这 些 红 薯 是 征 收 
wo3 ting1 shuo1 zhe4 xie1 hong2 shu3 shi4 zheng1 shou1
i hear talk this pl. red yam be collect
过后，还 被 卖 过 后，剩 下 来 的。 
guo4 hou4, hai2 bei4 mai4 guo4 hou4, shen4 xia4 lai2
de0. nan2 guai4 yi3 jing1 liang2 le0.
particle no wonder already sprout particle

I heard that these red yams had been steamed and were the remaining few after sales. No wonder they are cold now.

112. 是 他 说 要 书 童 好，我 可 没 
shi4 tal shuo1 yao4 shu1 tong2 hao3, wo3 ke3 mei1
be he say want page boy good i particleno
说过。你看 跟 读 没 好 处 吧。
shuo1 guo4 ni3 kan4 pei2 du2 mei2 hao3 chu4 ba0.
say has you see companionly no benifit particle

It was he who said it was better to have a page boy, I have never said that. Now you see, it is not good to have a companion for studying.
113. 是他说要疏通好的，我可没

shi4 ta1 shuo1 yao4 shu1 tong1 hao3, wo3 ke3 mei2
be he say want mediate good i particleno
说 过。你 看 这 下 犯 法 了 吧。
shuo1 guo4. ni3 kan4 zhe4 xia4 fan4 fa3 le0 ba0.
say has you see this time violate law particle

It was he who said it was better to mediate, I have never said that. Now you see, we have violated the law.

114. 是他说要书童好的，我可

shi4 ta1 shuo1 yao4 shu1 tong2 hao3 de0, wo3 ke3
be he say need page boy good particleno particle
没 说 过。你 看 陪 读 没 好 处
mei2 shuo1 guo4. ni3 kan4 pei2 du2 mei2 hao3 chu4
not say particle you see compani study no benifit
吧。
ba0. particle

It was he who said it was better to have a page boy; I did not say it. Now you see, it is not good to have a companion for studying.

115. 是他说要疏通好的，我可

shi4 ta1 shuo1 yao4 shu1 tong1 hao3 de0, wo3 ke3
be he say need mediate good particleno particle
没 说 过。你 看 这 下 犯 法 了 吧。
mei2 shuo1 guo4. ni3 kan4 zhe4 xia4 fan4 fa3 le0 ba0.
not said particle you see this time violate law particle

It was he who said it was better to mediate, I have never said that. Now you see, we have violated the law.

116. 你是这边的这只泥球

ni3 shuo1 shi4 zhe4 bian1 de0 zhe4 zhi1 ni2 qiu2
you say be this side particle this particulmud ball
大，还是那边的大？不过这边
da4, hai2 shi4 na4 bian1 de0 da4? bu2 guo4 zhe4
dbig or be that side particle big but this
Tell me, which one is bigger, the mud ball on this side, or the one on that side? But neither of them is bigger than that iron ball over there.

117. 你 说 是 这 边 的 这 只 泥 鱼 秋
you say be this side particle this particle loach
da4, hai2 shi4 na4 bian1 de0 da4? bu2 guo4 zhe4
big or be that side particle big but this
俩 都 没 那 只 碧 虎 大。
lia3 dou1 mei2 na4 zhi1 bi4 hu3 da4.
two both not that particle house lizard big

Tell me, which one is bigger, the loach on this side, or the one on that side? But neither of them is bigger than that house lizard.

118. 你 说 是 这 边 的 这 只 泥 球
you say be this side particle this particle mud ball
da4 dian3, hai2 shi4 na4 bian1 de0 da4? bu2 guo4
big particle or be that side particle big but this
这 俩 都 没 那 只 铁 球 大。
zhe4 lia3 dou1 mei2 na4 zhi1 tie3 qiu2 da4.
this two both not that particle iron ball big

Tell me, which one is bigger, the mud ball on this side, or the one on that side? But neither of them is bigger than that iron ball over there.

119. 你 说 是 这 边 的 这 只 泥 鱼 秋
you say be this side particle this particle loach
da4 dian3, hai2 shi4 na4 bian1 de0 da4? bu2 guo4
big particle or be that side particle big but
这两都没那只壁虎大。

Tell me, which one is bigger, the loach on this side, or the one on that side? But neither of them is bigger than that house lizard.

120. 你现在应该搞清楚是

You should first decide if you want to use a guardian, or use some other method. I heard the empress dowager wanted to hold court from behind a screen.

121. 你现在应该搞清楚是

You should first decide if you want to use a frying pan, or use some other thing. I heard an oven would produce the same results.

122. 你现在应该搞清楚是

You should first decide if you want to use a frying pan, or use some other thing. I heard an oven would produce the same results.
使用 监 国 好 呢， 还 有 别
shi3 yong4 jian1 guo2 hao3 ne0, hai2 shi4 you3 bie2
use guardian good particle or be has other
的 办 法。 听 说 太 后 想 垂 帘
de0 ban4 fa3. ting1 shuo1 tai4 hou4 xiang3 chui2 lian2
particleway hear say empress dowager want lower curtain
听 政。
ting1 zheng4
liste report

You should first decide if you want to use a guardian, or use some other method. I heard the empress dowager wanted to hold court from behind a screen.

123. 你 现 在 先 应 该 搞 清 楚 是
ni3 xian4 zai4 xian1 ying1 gai1 gao3 qing1 chu3 shi4
you now first should make clear be
使用 煎 锅 好 呢， 还 有 别
shi3 yong4 jian1 guo1 hao3 ne0, hai2 shi4 you3 bie2
use fry pan good particle or has other
的 办 法。 听 说 烤 箱 也 是 一
de0 ban4 fa3. ting1 shuo1 kao3 xiang1 ye3 shi4 yi1
particlemethod hear say oven also be one
样 的。
yang de0.
kind particle

You should first decide if you want to use a frying pan, or use some other thing. I heard an oven would produce the same results.

124. 他 想 发 明 一 台 僵 直 机， 实
tai1 xiang3 fa1 ming2 yi4 tai2 jiang1 zhi2 ji1, shi2
he want invent one particle stiff machine realize
现 好 朋 友 的 梦 想， 把 坏 人
xian4 hao3 peng2 you3 de0 meng4 xiang3 ba3 huai4 ren
realize good friend particledream particlebad person
变得 浑 身 僵 直。
bian4 de0 hun2 shen1 jiang1 zhi2.
change particleall body stiff
He wanted to invent a “stiff machine”, and realize the dream of his best friend: to make bad guys’ bodies become stiff all over.

125. 他 想 发 明 一 台 姜 汁 机, 实 现 好 朋 友 的 梦 想。 把 姜 自 动 削 皮 榨 汁。

126. 他 想 发 明 一 台 僵 直 机 器, 实 现 好 朋 友 的 梦 想。 把 姜 自 动 削 皮 榨 汁。

127. 他 想 发 明 一 台 姜 汁 机 器, 实 现 好 朋 友 的 梦 想。 把 姜 自 动 削 皮 榨 汁。

He wanted to invent a “ginger juice machine”, and the dream of his best friend: to peel and squeeze the juice out of ginger automatically.

He wanted to invent a “stiff machine”, and realize the dream of his best friend: to make those bad guys’ body become stiff all over.
He wanted to invent a “ginger juice machine”, and realize the dream of his best friend: to peel and squeeze the juice out of ginger automatically.

I don’t know if I should inherit it or not. The king, my father, asked me if I can take the heavy burden or not.

I don’t know if the western part of the city is good or not, but the eastern part is agreed to be a good place.
I don’t know if I should inherit it or not. The king, my father, asked me if I can take the heavy burden or not.

131. 我 也 不 知 到 城 西 好 呢, 还
wo3 ye3 bu4 zhi1 dao4 cheng1 xi1 hao3 ne0, hai2
or not good city west good

I also not know city west good particle or

132. 他 做 了 个 繁 华 梦, 里 面 他
ta1 zuo4 le0 ge4 fan2 hua2 mien4 ta1
he have particle flourishing dream in he

133. 他 做 了 个 繁 花 梦, 里 面 他
ta1 zuo4 le0 ge4 fan2 hua2 mien4 ta1
he had particle full bloom dream in he

He had a dream of (the town) flourishing, in which he saw a completely different view outside of the window. There were high rise buildings everywhere.
He had a dream of full blooms, in which he saw the view outside the window completely changed. There was full blooms everywhere, looking like a piece of brocade.

He had a dream of (the town) flourishing, in which he saw a completely different view outside of the window. There were high rise buildings everywhere.

He had a dream of full blooms, in which he saw the view outside the window completely changed. There were full blooms everywhere, looking like a piece of brocade.
BIBLIOGRAPHY


Mitterer, H., Csépe, V., and Blomert, L. The role of perceptual integration in the perception of assimilated word forms. *Quarterly Journal of Experimental Psychology, Section A: Human Experimental Psychology*.


