# A COMPARISON OF CUE-WEIGHTING IN THE PERCEPTION OF PROSODIC PHRASE BOUNDARIES IN ENGLISH AND CHINESE

By

Xinting Zhang

A dissertation submitted in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy
(Linguistics)
in The University of Michigan
2012

Doctoral Committee:

Associate Professor Andries W. Coetzee, Co-Chair
Professor San Duanmu, Co-Chair
Professor Patrice Speeter Beddor
Associate Professor William H. Baxter III

## ACKNOWLEDGMENTS

The writing of this dissertation has been a long and difficult process, and its completion would not have been possible without the support and encouragement of many people, to whom I would like to express my sincere gratitude.

I would first like to thank my advisor, Dr. San Duanmu, for his continuous guidance and support throughout my graduate studies. I am very fortunate to have San as my advisor. I enjoyed the discussions on many topics during our regular meetings, and I learned from him the importance of the big picture and critical thinking in doing research. His patience and support helped me overcome many difficult situations and finish this dissertation.

I would also like to thank my co-advisor, Dr. Andries Coetzee, for always being there to give advice and the push I needed to continue working. Andries introduced me to the experimental approaches and his suggestions helped shape the topic of the dissertation. He is not only a mentor but also a friend. His support and encouragement helped me go through many difficult times.

I am also grateful to the other members of my dissertation committee: Dr. Patrice Beddor and Dr. William Baxter. They provided valuable comments and feedback on my dissertation. I am especially indebted to Dr. Beddor, who offered invaluable advice on experimental design, data interpretation, and theoretical implication, and guided me

through every stage of my research. Her insightful comments along with technical and editorial advice were essential to the completion of this dissertation

I thank all the members of the Phondi group: Robert Felty, Michael Marlo, Anthony Brasher, Susan Lin, Miyeon Ahn, Kevin McGowan, Jonathan Yip, Yan Dong, Harim Kwon, Michael Opper, and Cameron Rule, for their technical support and helpful comments.

I want to thank my department, especially the Graduate Chair, Robin Queen, for the financial and emotional support when I most needed them.

Finally, I would like to thank my family for their love and support.

## TABLE OF CONTENTS

## LIST OF FIGURES

FIGURE

## LIST OF TABLES

TABLE

## LIST OF APPENDICES

APPENDIX

## ABSTRACT

Prosodic phrasing plays an important role in language comprehension and processing. Although prosodic boundaries are known to be marked by a variety of acoustic cues that involve pitch change, pauses, and pre-boundary lengthening, there is no consensus on the relative importance of these cues in perception. The present study investigates the acoustic correlates used in the production and perception of prosodic phrase boundaries. Specifically, it examines the perceptual weighting of these cues contributing to the marking of prosodic phrase boundaries differ in two languages, English and Chinese, with a focus on the difference in the perceptual reliance on pitch information by speakers of languages with and without lexical tone.

A production study examined the realization of pause duration, pre-boundary lengthening, and F0 change in syntactically ambiguous utterance pairs contrasting in the presence and absence of prosodic boundaries (e.g. *coffee, cake* vs. *coffee cake*) in English and Chinese. Results showed that speakers of both languages utilized durational (pause and pre-boundary lengthening) and pitch cues to signal phrase boundaries. Speakers of these languages differ, however, in the type of pitch information they employed for boundary categories: in English, F0 slope (representing dynamics of the pitch contour) was found to be an effective predictor; whereas in Chinese, pitch information was conveyed by a reset of the pitch declination.

A perception study investigated the relative weighting assigned by native English and Chinese speakers to these temporal and spectral properties in prosodic boundary perception. Responses to an identification task showed that both English and Chinese listeners use pause, pre-boundary lengthening, and pitch in perceiving prosodic boundaries in their native language. However, the two groups of listeners weight these cues differently, with English listeners attending more to pause than the other two cues, while Chinese listeners weight pitch reset most heavily.

These differences in perceptual weighting indicate an effect of language experience on the relative importance of perceptual cues. Language experience modulates the listener's attention to cues that are particularly relevant in the native language. Native speakers of a tone language attend to pitch information more than do native speakers of a non-tonal language because of the phonemic status of pitch in their native language.

# CHAPTER I

# INTRODUCTION

Spoken utterances are not just sequences of words, but always provide prosodic information such as rhythm, stress, and intonation (Ladd & Cutler, 1983). In addition to conveying linguistic information, such as the syntactic, semantic, and pragmatic structure of a sentence, prosodic cues also provide information such as emotion and attitude of the speaker. For example, higher pitch can signal excitement or urgency (Ladd, 1996). The same sequence of words can convey different meanings with variation in prosody. For example, the utterance *you went to the store* can be conveyed as a statement, or a question depending on the intonation used. *You went to the store* said with a high or rising pitch at the end generally implies a question, while a falling final pitch is usually associated with a statement.

The focus of the present study concerns one aspect of prosody: prosodic phrasing, or prosodic boundary (PB) in particular. A PB is a perceptible break that marks the grouping of words in an utterance. The three main cues that speakers across languages use to signal a prosodic boundary are acoustically realized as a lengthening of the word before a prosodic boundary, a change in the fundamental frequency, and/or the presence of a silent pause. In speech, adequate use of such cues can help the listener interpret the speaker's message because PBs often coincide with boundaries of syntactic constituents,

e.g. boundaries between phrases, clauses, or utterances (Scott, 1982; Wightman, Shattuck-Hufnagel, Ostendorf & Price, 1991).

The importance of prosodic boundaries is clearly illustrated when potentially ambiguous utterances need to be disambiguated (Lehiste, Olive, & Streeter, 1976; Price, Ostendorf, Shattuck-Hufnagel, & Fong, 1991; Scott, 1982). For example in the utterance: "*John and Paul or Steve will go to the party*" (Lehiste, 1973), it could mean that John and Paul will go to the party, or Steve, or John will go to the party with either Paul or Steve. To get the first meaning, a break after Paul is essential: "[John and Paul] [or Steve]", whereas with the second meaning, there would be a break directly after John: "[John] [and Paul or Steve]". If these boundaries are not realized appropriately, listeners will find it difficult to interpret the meaning of the utterance. In contrast, if prosodic boundaries do not match the expected syntactic structure, the processing is impeded. Sanderman and Collier (1997) have shown that an inappropriately phrased utterance (that is, an utterance with prosodic boundaries in inappropriate positions and with inappropriate realizations) slowed down processing compared with an utterance with appropriate phrasing.

As discussed above, PBs are marked by three major acoustic cues: silent pause, pre-boundary lengthening, and pitch change. While it is generally agreed that these cues play an important role in the realization of PBs, there is no consensus on the relative importance of these cues. Previous studies on cue-weighting in phonetic categorization have found that when there are multiple cues to a particular linguistic category, listeners weight certain cues more than others in speech perception (Boersma, 2005; Cho & McQueen, 2006; Escudero, 2005; Gottfried & Beddor, 1988; McGuire, 2007). However, studies using well-designed stimuli with respect to cue-weighting in the perception of

PBs are lacking. Moreover, it is not precisely clear whether boundaries are signaled and perceived differently in different languages. The present study investigates through production and perception experiments whether the perceptual weighting of the prosodic cues contributing to the marking of prosodic boundaries differ across two structurally different languages: American English (English hereafter) and Standard Chinese (Chinese hereafter).

The remainder of this dissertation is structured as follows: Chapter 2 assesses the prosodic phrasing literature with regards to acoustic correlates of prosodic boundaries in the production and perception of speech, and cue-weighting in speech perception. In Chapter 3, I present the methods and results of production experiments that investigated the phonetic cues speakers use in the realization of prosodic boundaries of both English and Chinese. Chapter 4 presents a perception experiment that investigated listeners' perception of utterances with manipulated prosodic cues, with the analysis focusing on the relative importance of these cues within and across the two languages. Finally, Chapter 5 contains a discussion and conclusion of this study, including ideas for future research.

# CHAPTER II

# LITERATURE REVIEW

## 2.1 Introduction

This chapter provides an overview of previous research on acoustic correlates of PBs and their interaction in the production and perception of speech. These areas will be reviewed in sections 2.2. In addition, previous research on cue-weighting, and language-specific cue-weighting in speech perception, will be reviewed in sections 2.3 and 2.4 in order to motivate hypotheses concerning the perception of PBs. Based on the findings from previous research, the purpose, research questions, and predictions of the current study will be discussed in section 2.6. A summary of this chapter will be presented in section 2.7.

## 2.2 Acoustic correlates of prosodic boundaries

As noted above, previous studies have established three major acoustic correlates of prosodic boundaries: silent pause, pre-boundary (or final) lengthening, and changes in fundamental frequency (F0). Each of the three acoustic correlates is briefly described below.

The presence of a pause after a prosodic boundary has long been considered an important acoustic correlate of prosodic phrasing (Cooper, Paccia, & Lapointe, 1978;

Scott, 1982; Streeter, 1978). There has been much research on the relation between the presence/absence of a silent pause and prosodic phrasing, and the role of silent pauses in boundary perception (e.g., Carlson & Swerts, 2003; Strangert & Heldner, 1995 in Swedish; Krivokapic, 2007 in English; Lin and Fon, 2009; Yang, 2007 in Mandarin Chinese. The presence of a pause was found to be highly correlated with the perception of a boundary in these studies.

Pre-boundary lengthening, often also called phrase-final lengthening, refers to the phenomenon in which the duration of the syllable preceding a prosodic boundary is longer than it is in the no-boundary case (Berkovits 1993; Crystal & House, 1988; Klatt, 1975; Ladd & Campbell, 1991; Lehiste, 1973; Lehiste et al., 1976; Scott, 1982; Streeter, 1978; Wightman et al., 1992). The lengthening is attributed to the effect of a slowing down of the articulation rate before the phrase boundary. Pre-boundary lengthening has been reported as a prosodic boundary marker in a variety of languages, like English (Price et al., 1991, Turk & Shattuck-Hufnagel 2007; Wightman et al. 1992), Chinese (Duanmu, 1996; Shen, 1992), Korean (Cho & Keating, 2001), French (Martin, 1982), Dutch (Cambier-Langeveld, 1997), and Swedish (Lindblom & Rapp, 1973), to name just a few.

While temporal cues (such as pauses and pre-boundary lengthening) are important acoustic correlates used in the realization of prosodic boundaries, studies have shown that pitch change is also important in the production and perception of these boundaries. The main pitch characteristics of prosodic boundaries are pitch movement in the nucleus and in unstressed syllables following it (Cruttenden, 1997), and phrase-initial pitch reset

related to declination. (de Pijper, 1994 for Dutch; see also Wagner & Watson, 2010 for a review).

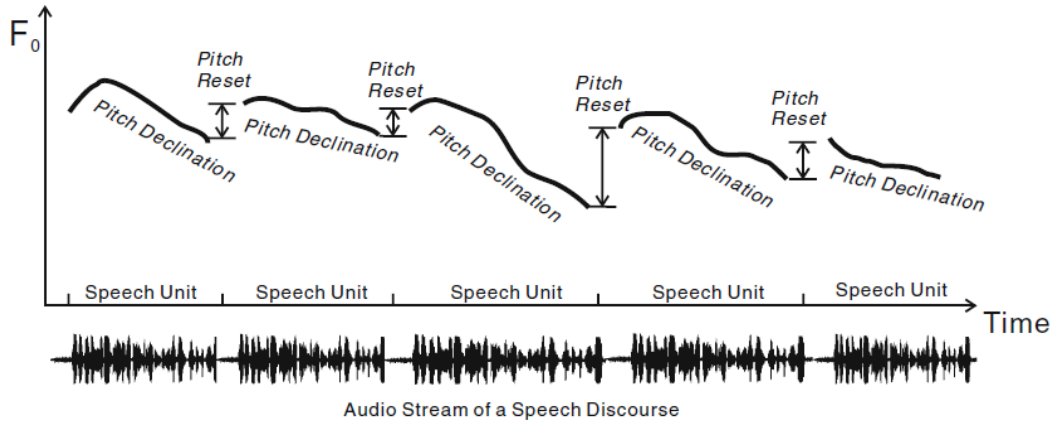



Figure 2.1 Pitch declinations and reset (Xie, 2008)

Pitch tends to decline across the course of an utterance, known as pitch declination. Pitch reset refers to the readjustment of the pitch height to a higher value in the course of the F0 declination at junctures (see Figure 2.1).

Cooper and Sorensen (1977) examined the relationship between F0 contours and clausal boundaries using read utterances. Final declination and pitch reset were consistently observed within and across a clausal boundary respectively. In acoustic studies of English and Dutch, researchers have found that, when a major boundary follows a certain target word, the word tends to have a steeper fall, whereas the stressed syllable in the following word has a steeper rise, relative to the no-boundary case. These studies also show that the pitch of the post-boundary syllable tends to be higher than in the no-boundary case (Lehiste, 1973). Yang and Wang (2002) investigated the acoustic correlates of prosodic boundaries based on a large labeled corpus of Mandarin Chinese.

The results showed a significant degree of pitch reset at phrase boundaries and intonational phrase boundaries. Moreover, the higher the prosodic boundary is, the larger the extent of the pitch reset.

Because most of the studies involved more than one acoustic correlate, a more detailed discussion on the production and perception of these cues will be presented in the section 2.4.1, which focuses on the interaction between these prosodic cues.

## 2.3 Cue-weighting in speech perception

### 2.3.1 Cue-weighting at the segmental level

Just like multiple acoustic cues comprising a prosodic boundary, many linguistic categories a listener encounters in everyday life contain multiple acoustic cues. For example, in addition to voice onset time (VOT; Lisker, 1975), acoustic cues to the voicing distinction include F1 transition, vowel length, F0 adjacent to the closure, stop closure duration, amplitude of release burst etc. (Lisker, 1975; Repp, 1979; Stevens & Klatt, 1974; Summerfield, 1981).

Perceptual experiments have shown that when there are multiple cues to a particular linguistic category, listeners pay attention to certain cues more than others in speech perception (Boersma, 2005; Cho & McQueen, 2006; Escudero, 2005; Gottfried & Beddor, 1988; Harnsberger, 2001; McGuire, 2007). That is, listeners do not give equal importance to all cues available to them and weight these cues differently. For example, both spectral and temporal acoustic cues differentiate English tense and lax vowels like /i/ and /ɪ/. Adult native American-English listeners, however, rely more on formant

frequency than vowel duration in categorizing tense and lax vowels; they identified a great majority of vowels in isolation and CVC syllables correctly irrespective of the manipulated duration of the vowel (Hillenbrand, Clark, & Nearey, 2001). This perceptual bias has also been found in studies on stop-consonant recognition in which the relative importance of the release burst and formant transitions are investigated (Ohde & Haley, 1997; Walley & Carrel, 1983). In Walley and Carrel (1983), children and adults were asked to identify synthetic stop CV stimuli in which formant transition information and the onset spectrum specified conflicting place of articulation of the stop consonant. The results showed that the contribution of the release burst and the formant transitions is similar for adults and children of 6 years and older: they can identify stops with only the burst or only transition information, but formant transitions determine place of articulation when transition and burst cues are in conflict.

### 2.3.2 Cue-weighting at the supra-segmental level

Cue weighting studies at the supra-segmental level have mainly focused on the perception of stress. In a series of classic studies, Fry (1955, 1958) used synthesized stimuli to study the relative importance of F0, duration, intensity, and formant structure in native English speakers' perception of stress position. In Fry (1958), the word *subject* was manipulated to have eight unequally-sized steps of F0 in combination with five levels of vowel duration, first on the first syllable and then on the second syllable. The results showed a consistent effect of F0 at each duration level, with syllables with higher F0 being more likely to be perceived as stressed. He thus concluded that "the fundamental frequency cue may outweigh duration" (Fry, 1958: 151).

Work on how listeners weight cues to prosodic boundaries has been scarce, and unlike the general agreement regarding acoustic cues used by speakers and listeners to mark prosodic boundaries, there is little consensus on the relative importance of these cues. Although silent pause has been found to be more salient to the perception of boundary and of degree of boundary strength than other cues, most findings are based on studies involving temporal cues (pre-boundary lengthening and pausing) only.

For example, Scott (1982) investigated the role of pause and lengthening by using temporally manipulated sentences. She found that both the duration of the pause alone and the combined duration of pause plus lengthening provided listeners with a sufficient cue for prosodic boundaries. She thus claimed that prosodic boundaries could be marked differently and boundary features can occur together, separately, or not at all. In a study of the relation between temporal and syntactic structures in Mandarin, Shen (1992) used ambiguous sentences in literary Mandarin Chinese to demonstrate that both pause and final syllable lengthening are robust cues in signaling sentential boundary location. Experimental results indicated that, in the speech production of Mandarin, speakers use both pauses and final lengthening to convey syntactic boundaries. For example, the same syllable occupied about 13-14% of total duration of the utterance in a boundary position while it only occupied about 9-11% when it is at a non-boundary position. However, in perception, listeners relied predominantly on pauses rather than final lengthening as boundary markers; the duration of the phrase-final syllable had to be increased by a significant amount for it to be a reliable cue for syntactic boundaries.

Studies that investigated both durational and pitch cues produced different results regarding the relative importance of cues. Streeter (1978) tested the separate influences of

duration, pitch and amplitude on the perception of prosodic boundaries by conducting a listening experiment using ambiguous algebraic expressions such as (1) "[A plus E] times O" vs. (2) "A plus [E times O]". In one experiment, she exchanged the values of duration, pitch and amplitude, either individually or in various combinations, of bracketing structure (1) onto bracketing structure (2). Experimental results showed that duration is more important than pitch contour and amplitude in parsing ambiguous algebraic expressions. The cue 'duration' and the combinations of the cues 'duration and amplitude', 'duration and pitch' and the combination of all three cues, led the listeners to interpret the utterance they heard as having the meaning in (1). Amplitude and pitch alone did not lead to the first reading; instead, most of the listeners still disambiguated the utterance as having the reading (2). Moreover, the effects of duration pattern and pitch contour were not interactive and they cumulatively produced more correct responses, because Streeter (1978) found that the combination of cues 'duration and pitch' was more effective in disambiguating the utterance than the cue 'duration' alone.

Beach (1991) used synthesized speech materials to study the interaction between phrase-final lengthening and a pitch cue in the perception of syntactically ambiguous sentences such as in *Jay believed the gossip about the neighbors {right away/wasn't true}*. The main verb (*believed* in the example) of the sentence was manipulated in its duration and the extent of F0 fall from transitive realization (*right away* case in the example) to complement realization (*wasn't true* case in the example). Subjects were more likely to choose the complement interpretation with greater durations and more marked falls. The results showed that both duration and pitch are important in perception of prosodic boundaries and, more importantly, they are processed interactively. This interaction is

seen as cue-trading relations, i.e. duration and pitch cues are perceived together as one integrated percept; the influence of one cue is greater when the other cue is weaker. This trading relationship was also observed in other studies. For example, Horne, Strangert, and Heldner (1995) proposes that there is a trading relationship between segment duration and following pause duration, showing that segment duration is negatively correlated with silent interval duration at lower ranked boundaries.

### 2.4 Language-specific cue-weighting in speech perception

Numerous studies suggest that the way acoustic cues are weighted in speech perception and production is language-specific, i.e., speakers have learned to pay more attention to acoustic cues that signal contrasts in the phonological category in their native language (Escudero, Benders, & Lipski, 2009; Kluender, Lotto, Holt, & Bloedel, 1998). In this view, native-language input is crucial to the formation of cue weighting strategies (Holt, Lotto, & Kluender, 2001; Jusczyk, 1993). For example, in a study investigating Dutch and English listeners' cue-weighting of vowel duration as a perceptual cue for nonword-final fricative voicing in English, language background was found to have an effect on the categorization of final fricatives: Dutch listeners used vowel duration, but less than English listeners did for final /v-f/ and /z-s/ contrasts. The results of L1-Dutch L2-English listeners can be explained by their language background, because Dutch listeners have native language experience with the use of vowel duration as a perceptual cue for vowel length contrasts and for intervocalic consonant voicing, but not for final voicing contrasts (Broersma, 2010).

There is also evidence from developmental studies that showed linguistic experience can have an influence on acoustic cue weighting strategies in children's speech perception. Previous studies have discovered that children weight certain acoustic cues differently from adults, and the weighting changes as they gain linguistic experience. Nittrouer (1992, 2002), for example, found that when identifying /s/ and /ʃ/ contrasts, young children relied more on vowel formant transitions, and relatively less on the fricative noise spectrum, as compared to adults. They gradually reached adult-like weighting of fricative noise as more important than transition around 7-8 years old. Differences between children and adults in their relative weighting of acoustic cues have also been found in other studies (Greenlee, 1980; Morrongiello, Robson, Best, & Clifton, 1984).

In the following, I will discuss the studies that investigated the effect of linguistic experience on the perception of pitch information. The literature has shown that there are differences among languages in the reliance on this cue – especially between languages with and without lexical tone.

Effects of linguistic experience have also been reported in some studies on suprasegmental features, such as stress patterns, lexical tone identification, and sentence-level prosodic patterns. It was found that listeners' L1 prosodic systems have a profound effect on their perception of the suprasegmental features. For example, Gandour (1983), using multidimensional scaling, investigated the perceptual weights of the tonal dimensions (i.e. pitch height and pitch contour) by listeners of four tonal languages, including Mandarin, Cantonese, Taiwanese, and Thai, as well as by those of a nontonal language, English. He found that English listeners tended to focus on pitch height, while

listeners from tone languages focused on both pitch height and pitch contour when discriminating tones. Gandour (1983) attributed this difference to the lack of contrastive tones in English, which arguably led English listeners to direct their attention almost exclusively to the F0 height of the stimuli.

Previous studies also distinguished the contribution of pitch information at the lexical level and the sentence level. Liang and Van Heuven (2007) compared the perception of Chinese tone and intonation (question vs. statement) by native Chinese and L2 Chinese learners where L1 was either a tonal or a non-tonal (Uyghur) language. They found that L2 learners who speak a tonal language were more sensitive to lexical tones but were less sensitive to F0 information at the sentence level (intonation meaning) compared to L2 learners from a non-tonal language. They suggested that listeners of a tonal language had to face two competing tasks in the use of pitch cues: listen for pitch information at the word level and monitor pitch change at the sentence level. Their processing priority given to lexical tones reduced the sensitivity to pitch cues at the sentence level. A similar finding was reported in more recent work by Braun and Johnson (2011). They tested Mandarin and Dutch listeners' performance in speeded ABX tasks on CVCV nonsense words. The stimuli were manipulated to have either a rising pitch contour on the first syllable (signaling tone 2 in Mandarin, but non-linguistic in Dutch), or a rising contour on the second syllable (signaling an interrogative contour in Dutch, and tone 2 in Mandarin). The results showed that Mandarin listeners were more attentive than Dutch listeners to pitch movements as these signaled potential lexical contrasts in Mandarin. Dutch listeners were more attentive to stimuli that represent the linguistically meaningful pitch contrasts (interrogative question) than to the non-linguistic pitch

contrasts. They thus concluded that listeners should be particularly attentive to any pitch information that signals meaningful information in the native language. This includes pitch movements signaling lexical contrasts as well as postlexical contrasts.

However, in some other tasks, it has been shown that listeners with tone language experience do not differ in their performance in the processing of tone information compared to listeners without tone language experience (Bent, Bradlow, & Wright, 2006; Cutler & Chen, 1997; Francis, Ciocca, Ma, & Fenn, 2008). Francis et al. (2008) compared the recognition of Cantonese lexical tones by English and Chinese learners. Results showed that both groups performed similarly on the pretest. They claimed that the mere presence or absence of lexical tone contrasts in the native language is not sufficient to determine cross-language perception of lexical tones. Instead, the findings of these studies suggest that what matters are the F0 patterns that listeners have been exposed to in their native language irrespective of their function as cues to tone as opposed to intonational categories. This was evidenced by different performance on tone 25[1] (high rising) vs. tone 21 (low falling) by English listeners. They were quite good at identifying the 25 tone, possibly due to its contour similar to English question intonation, but their performance on the 21 tone, not similar to any native intonational category, was poor.

In summary, listeners tend to weight cues differently based on their linguistic experience or on the salience of the cue itself. Furthermore, linguistic properties vary across languages, and languages that differ substantially in certain linguistic aspects can result in different cue-weighting strategies used by their speakers in perception. Cue-

[1] The numbers represent pitch movement on a 5-point scale, with 1=lowest, 5=highest.

weighting and cue-weighting strategies are relatively well-studied in the perception of native and non-native consonants and vowels. Similar studies on suprasegmentals are scarce.

## 2.5 Present study

### 2.5.1 Research questions

This dissertation investigates the acoustic cues speakers use to convey prosodic boundaries using acoustic analysis, as well as an evaluation of listeners' perception of the cues, in order to determine the perceptual weighting of the prosodic cues contributing to the marking of prosodic boundaries within different languages (English and Chinese), and to compare them across languages. In particular, the following research questions are asked:

(1) Which acoustic cues do speakers use to convey prosodic boundaries? To answer this question, a production experiment examining the realization of pause duration, pre-boundary lengthening, and F0 change in syntactically ambiguous utterance pairs contrasting in presence and absence of prosodic boundaries (e.g. coffee, cake vs. coffee cake) is carried out. The minimal pair construction of the test utterances enables us to directly compare the acoustic features under two prosodic conditions.

(2) Are prosodic cues observed in the production experiment employed in perception by listeners? This is tested by manipulating the prosodic cues observed in the production experiment and testing them in an identification task.

(3) How important are these cues relative to each other within each language? To answer this question, the relative importance of cues will be analyzed in logistic regression models, and cue-weighting is determined by the cue weight analysis.

(4) Does the perceptual weighting of the cues differ across languages? Logistic regression models are built for each language, and the relative weights obtained from each model are compared for their values in cue-weighting.

### 2.5.2 Predictions

Previous research on the perception of prosodic boundaries has shown that pause, pre-boundary lengthening, and pitch change are all well-established acoustic correlates for different languages. It is thus predicted that both English and Chinese speakers will use them to convey the presence of a prosodic phrase boundary.

Based on the findings on the influence of language experience on speech perception in segments and suprasegmentals, different perceptual patterning is expected to hold for the two languages.

Chinese is a tone language. Pitch is used in Chinese both at the word level to differentiate between four lexical tones and at the sentence level to signal differences in intonation (such as representing focus, distinguishing between statements and questions). The magnitude of the pitch differences for signaling the contrasts at the sentence level is less than that at the word level. Arguably, pitch in Chinese is tied up at the word level and can thus no longer be used as freely at the level of prosodic patterning. Chinese listeners therefore are expected not to be sensitive to the pitch information at the sentence level, as has been found in previous studies (Braun & Johnson, 2011; Liang & Van Heuven, 2007).

Previous studies on the relationship between intonation and tone in Chinese also found that listeners find it difficult to identify question intonation on a sentence with a final rising tone (Yuan, 2006). This is just one example showing the interference between tonal and intonation contours. If pitch contours cannot change freely at the prosodic level in production, Chinese speakers might exploit other cues more in the realization of prosodic boundaries. If this finding is upheld for production, Chinese listeners might be expected to be relatively insensitive to pitch information at the sentence level.

However, an opposite prediction can also be made based on the influence of linguistic experience on cue-weighting. As tone language learners, Chinese listeners might be more sensitive to pitch information as F0 signals contrast in their native language, regardless of whether it is at the word level or sentence level. They might therefore weight pitch change more heavily than temporal cues.

In contrast, English listeners are expected to rely more on durational cues than pitch cues due to the lack of F0 as a phonological contrast in their native language. Based on findings from previous studies, it is predicted that pause will be weighted more heavily than pre-boundary lengthening in the identification of prosodic boundaries.

# CHAPTER III

# ACOUSTIC STUDY

## 3.1 Introduction

The implementation of prosodic boundaries by native speakers of English and Standard Chinese is examined through an acoustic study. The acoustic study aims to examine which prosodic cues speakers use to convey prosodic boundaries. In particular, the realization of pause duration, pre-boundary lengthening, and F0 change is investigated in syntactically ambiguous utterance pairs differing in presence or absence of prosodic boundaries (e.g. *coffee, cake* vs. *coffee cake*). The identical segmental and stress composition of these phrase pairs provides for a high degree of control, making it easier to compare a given prosodic correlate in boundary and no-boundary positions.

The organization of this chapter is as follows: section 3.2 will explain the methods used in this study, including a description of the speech materials used in the production experiments (section 3.2.1), the speakers (section 3.2.2), and the general procedure for the reading task (section 3.2.3). Section 3.2.4 will give a description of the acoustic measurements taken of the speech data. In Section 3.3, experimental results and data analysis will be presented, and section 3.4 will summarize the chapter.

## 3.2 Method

### 3.2.1 Materials

The materials for this experiment were designed following those used by Dankovicova, Pigott, Wells, and Peppé (2004), but with a more controlled word structure (which will be explained below). The stimuli consist of 10 pairs of syntactically ambiguous utterances, as shown in (1), in each language. Each pair was constructed using the same words, but with different meanings depending on the presence of the prosodic boundaries: the first two nouns form a compound noun in utterance (a), in which there is no prosodic boundary after the first noun, whereas they are two single nouns in utterance (b), separated by a prosodic boundary.

(1) English a. turkey-salad and coffee
b. turkey, salad, and coffee
Chinese a. mogu-shala he hongjiu 'mushroom-salad and red wine'
b. mogu, shala, he hongjiu 'mushroom, salad, and red wine'

In the examples above, although the utterances in a pair consist of identical syllables, two different syntactic structures can be interpreted. The location of the prosodic boundary determines the syntactic interpretation conveyed in a particular reading.

These utterance types were chosen, firstly because of their clear distinction between the boundary vs. no-boundary reading so that speakers and listeners are able to disambiguate these types of utterances easily and precisely, and secondly because a more natural perception task without directly referencing the location of boundaries can be

employed in the following perception experiment. Instead of asking the participants to determine where the boundaries are located in the utterances, they are asked to identify how many items there are in the stimuli they hear.[2] A 2-item identification response indicates that the first two words form a compound noun, and hence that there is no boundary between them, while a 3-item identification response indicates the first two words are separate nouns, suggesting the presence of a boundary.

Results of a pilot study showed that speakers tended to confuse the two types of utterances when they were presented with printed versions of the test utterances. In order to elicit clearer distinction between the two, colored pictures were used in the production experiment (see appendix for lists of colored pictures). As a result, the possibility of "picturable presentation" was another consideration of the test words.

In order to control for the influence on the F0 and duration measurement from adjacent sounds, the second word in the utterance was the same across all the utterances. *Salad* and *shala* ('salad') were chosen because 1) they have the same meaning in the two languages; 2) they start with fricatives, which facilitate segmentation and measurement; 3) most importantly, they can be combined with a variety of words to form Noun-Noun (NN) compounds.

Another constraint on the choice of test words in English is imposed by number marking in English nouns. The targeted nouns are in the singular form in the compound condition irrespective of whether they are count or mass nouns. However, they have to be in the plural form in the single noun condition if they are count nouns, which will make

[2] The task is not part of the production experiment and is relevant in Chapter 4 where the details of this task will be further discussed.

the utterance pairs differ in segment combination. For example, the count noun *potato* will have two different forms in these particular phrase types—*potato-salad and juice* vs. *potatoes, salad, and juice*. Therefore, only mass nouns were selected.

For the purpose of the cross-linguistic comparisons, the target words in the stimuli (the first noun) control for number of syllables and stress pattern in the two languages. They are disyllabic, having stress on the first syllable. Ideally, the syllable structures (heavy or light) of those words would also be controlled, but because of the many restrictions on the word selection stated above, it turned out to be impossible to control for the syllable structure of test words.

Most English disyllabic words with initial stress are unstressed on the second syllable. In Chinese all full syllables (syllables that carry lexical tones) are phonetically stressed and all weak syllables (syllables that do not carry lexical tones) are phonetically unstressed (Duanmu, 2007). Therefore, only weak syllables were used in the second syllable in order to match the English counterparts. As the tone of a weak syllable depends on the tone of the preceding syllable, only tone 1 and tone 2 were used in the first syllable of the first word. Pitch contours of the weak tone following tone 1 and tone 2 are both high falling— 41 and 51 respectively (Lin & Yan, 1980), similar to that of S(trong)W(weak) words in English.

Additionally, stress assignment in English NN compounds is variable. Although most NN compounds have the main stress on the first word, there are cases where the main stress is on the second word, rather than on the first, e.g. *Madison Avenue*, *silk tie*, etc. (Plag, 2003). Furthermore, there is also speaker variation in the way this type of

compound is stressed. To make sure that all the test compounds have the same stress pattern, an informal survey involving two participants was carried out to investigate the location of their main stress. Speakers for the production experiments were also interviewed about their stress assignment of those words. No speakers were found to assign main stress on the second word for the selected words.

Altogether, ten pairs of test utterances were created in each language. Pilot study results showed that some speakers tended to emphasize the first word, presumably because they were followed by the same word in all the test utterances. In order to avoid the effect of focus stress, ten pairs of filler utterances were created for each language. These filler utterances were also composed of food items, but they were not controlled for stress pattern, number of syllables, or number of nouns. A complete list of test utterances for each language is listed in Tables 3.1, and 3.2. See Appendices A-C for the list of filler words, and colored pictures used in the production experiment.

Table 3.1 Utterance list of the production task for English

| **No boundary** | **With boundary** |
|---|---|
| 1. Bacon-salad and wine | 1. Bacon, salad, and wine |
| 2. Chicken-salad and juice | 2. Chicken, salad, and juice |
| 3.Ginger-salad and tea | 3.Ginger, salad, and tea |
| 4.Kiwi-salad and yogurt | 4.Kiwi, salad, and yogurt |
| 5.Melon-salad and milk | 5.Melon, salad, and milk |
| 6.Pasta-salad and coffee | 6.Pasta, salad, and coffee |
| 7.Pepper-salad and juice | 7.Pepper, salad, and juice |
| 8.Salmon-salad and wine | 8.Salmon, salad, and wine |
| 9.Tuna-salad and wine | 9.Tuna, salad, and wine |
| 10.Turkey-salad and coffee | 10.Turkey, salad, and coffee |

Table 3.2 Utterance list of the production task for Chinese speakers (transcribed in Pinyin)

| No boundary | Gloss | With boundary | Gloss |
|---|---|---|---|
| 1. hetao shala he hongjiu | 1. Walnut-salad and red wine | 1. hetao, shala he hongjiu | 1. Walnut, salad, and red wine |
| 2. huanggua shala he chengzhi | 2. Cucumber-salad and orange juice | 2. huanggua, shala he chengzhi | 2. Cucumber, salad, and orange juice |
| 3. juzi shala he suannai | 3.Orange-salad and yogurt | 3. juzi, shala he suannai | 3.Orange, salad, and yogurt |
| 4. mogu shala he hongjiu | 4.Mushroom-salad and red wine | 4. mogu, shala he hongjiu | 4.Mushroom, salad, and red wine |
| 5. putao shala he niunai | 5.Grape-salad and milk | 5. putao, shala he niunai | 5.Grape, salad and milk |
| 6. qiezi shala he cha | 6.Eggplant-salad and tea | 6. qiezi, shala he cha | 6.Eggplant, salad, and tea |
| 7. shiliu shala he suannai | 7. Pomegranate-salad and yogurt | 7. shiliu, shala he suannai | 7. Pomegranate, salad, and yogurt |
| 8. xigua shala he niunai | 8.Watermelon-salad and milk | 8. xigua, shala he niunai | 8.Watermelon, salad, and milk |
| 9. yezi shala he kafei | 9.Coconut-salad and coffee | 9. yezi, shala he kafei | 9.Coconut, salad, and coffee |
| 10. yingtao shala he hongjiu | 10.Cherry-salad and red wine | 10. yingtao, shala he hongjiu | 10.Cherry, salad, and red wine |

### 3.2.2 Participants

Two groups of speakers participated in the study: native speakers of English and native speakers of Chinese; each group consisted of 10 speakers (5 male and 5 female). They were paid to participate in the study.

The English participants (age 18–22 years, median 20 years) were all undergraduate students at the University of Michigan. Most of them came from the Midwest, more specifically from Michigan. The Chinese speakers (age 18-38 years, median 26 years) were recruited from the University of Michigan; they were either students or faculty members at the University of Michigan. All of them were born and

brought up in Beijing, China. Beijing speakers were selected because the test words involve weak (unstressed) syllables. Although such syllables are absent in many Chinese dialects, they are characteristic of Beijing Mandarin.

### 3.2.3 Procedure

For each test item, there was a picture card showing the two or three items in the utterances. Test words were also printed under each picture as it was difficult to memorize the names of each picture in short periods of time. For example, the picture card for "chicken-salad and juice" involved two pictures, with first picture showing a food item (e.g. a bowl of chicken salad) designated by the compound noun, and second picture showing the other food-item (e.g. juice). The card for "chicken, salad, and juice" showed three pictures of the three food-items. Figure 3.1 illustrates a pair of picture cards used in the production experiment.


Figure 3.1 Illustration of picture cards used in the production experiment

During the recording, speakers were asked to read each utterance as naturally as possible in a hypothetical setting in which they were supposedly giving their order to a waitress (the researcher). Their pronunciations should be clear so that the waitress would know whether they ordered two or three items.

Each speaker was recorded separately in a sound booth at the University of Michigan. All recordings were conducted with Edirol UA25 Audio Capture recorder and AKG C 4000 B microphone, at a sampling rate of 44.1 kHz. In the experiment, the utterances were randomized for each speaker. The recordings were arranged in five blocks, each with one repetition of the stimuli, the order of presentation semi-randomized in each block, although the randomization had no adjacent minimal pair members. Also, test utterances were separated by fillers. There was an optional break between each block. Before the recording, a familiarization session of the test items was also carried out to make sure that subjects knew what was represented in each picture. Each participant made five recordings of the 20 test utterances and they typically completed the entire procedure in approximately 20-25 minutes.

### 3.2.4 Acoustic Measurements

All acoustic measurements were taken using Praat (Boersma &Weenink, 2010). For each utterance, each of the first three syllables was annotated as a measurement interval with the help of a Praat script "ProsodyPro" written by Xu (2005-2011). The script was also used to extract durations and F0 information of the specified interval. A total of 1,200 tokens were analyzed in this experiment (20 utterances × 3 repetitions (the

middle 3 of the 5 recordings) × 20 speakers). The following is a detailed description of the acoustic measurements.

Pause was defined as the interval between the offset of the first word and the onset of the second word. Pause duration was measured from the end of periodic voicing of the first word (they were controlled to have a sonorant ending) to the beginning of the second word. The second words are *salad* in English and *shala* (salad) in Chinese, the fricative onset ([s] and [ʂ] respectively) of which made the segmentation straightforward.

Pre-boundary lengthening was determined as follows: the total duration of the final syllable of the first noun was measured.[3] It was measured from the end of periodic voicing of the preceding syllable to the end of periodicity in the waveform for the second syllable itself.

Based on observation of the pitch contour of the test items, two types of measurements were taken to represent properties of F0 change. First, F0 slope of the pre-boundary syllable was employed to quantify the fall-rise patterns of F0 contours, because a continuation rise intonation (which typically occurs when speakers produce a list of items) was observed in the boundary condition of English utterances. In this study, pitch slope (in ST/s (semitones per second)) is derived as F0 Slope $= \frac{f_{max}-f_{min}}{t_{max}-t_{min}}$ (Plag, Kunter, & Schramm, 2011), where $f_{max}$ and $f_{min}$ are the F0 maximum and F0 minimum of the

[3] One of the common practices for measuring pre-boundary lengthening effect is to use proportional measurement (normally with reference to the duration of whole sequence) in order to take speech rate into consideration. This was not adopted here because test sentences were relatively short in this study, so there was a large proportional increase of the duration of the whole sequence under boundary condition. The percentage of segments undergoing a lengthening effect and the addition of silent pause in the boundary condition increased the duration of the sentence to a large extent, which may offset the lengthening effect of the pre-boundary syllable if the duration of the whole sentence was used as the denominator in the calculation.

target pitch contour in semitones, and $t_{max}$ and $t_{min}$ are the times at which the maximum and minimum pitches are observed. If $f_{max}$ occurrs before $f_{min}$, the value for $t_{max}$ - $t_{min}$ and slope will be negative, indicating a falling pitch contour; a positive slope, consequently, indicates a rising pitch contour.

Second, the reset of F0 declination (F0 reset) is used to represent F0 movement. Different measurements of F0 reset have been proposed in the literature. For example, Ladd (1988) defined F0 reset as the difference in F0 between the last pre-boundary peak and the first post-boundanry peak across two utterances. In Swerts (1997), however, pitch reset was derived by subtracting the pitch range of the post-boundary syllable from that of the pre-boundary syllable. Pitch range (measured in Hz) is measured in the vowel portion of the syllables before and after the boundary at the maximal point of vowel intensity. In this study, following Wang (2002), F0 reset is measured as the difference in Hz between the minimum F0 of the pre-boundary syllable and that of post-boundary syllable, because it is a more widely used method (Li, Yang, & Lu, 2010)

An example of the segmentation criteria is shown in Figures 3.2.

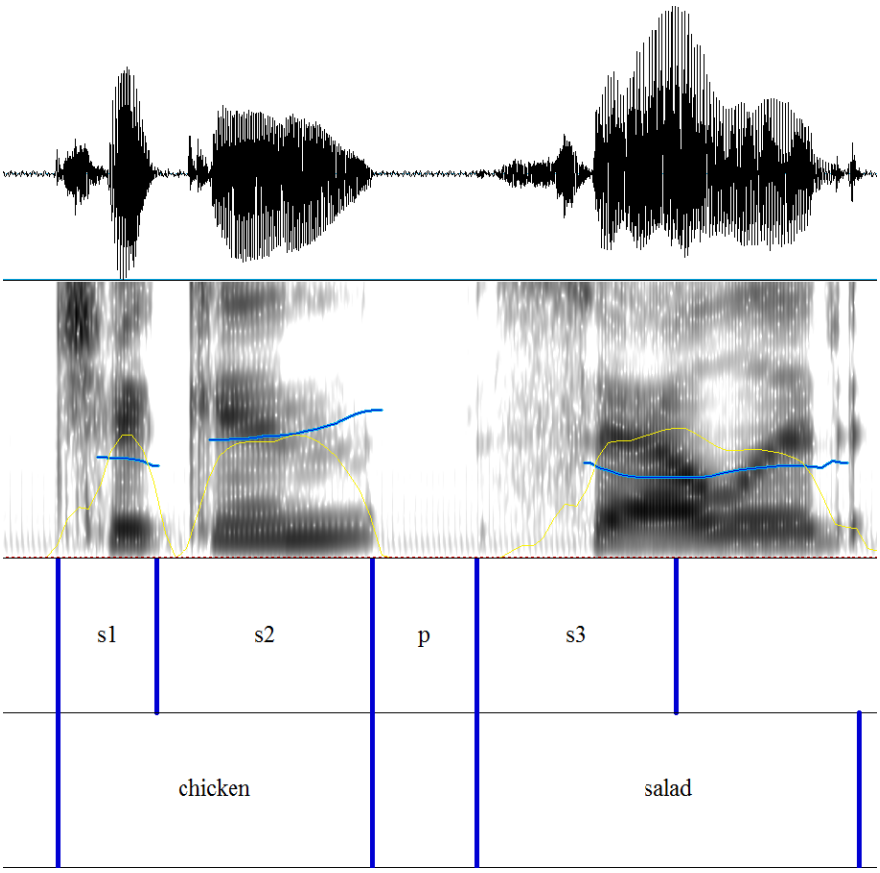



Figure 3.2 Illustration of measurements of acoustic correlates

In Figure 3.2, the interval marked as “s2” is the measurement for pre-boundary lengthening, indicated by the duration of the pre-boundary syllable; and the interval marked as “p” is the measurement for pause duration. F0 values are obtained from the pitch contour produced by the Praat pitch tracker, represented by the blue lines in the picture. F0 slope is taken at the pre-boundary syllable “s2”, where the maximum and minimum F0 and the times at which they occur are recorded automatically by the Praat script. F0 reset is measured at the pre-boundary syllable “s2” and post-boundary “s3”, where the minimum F0 values are taken.

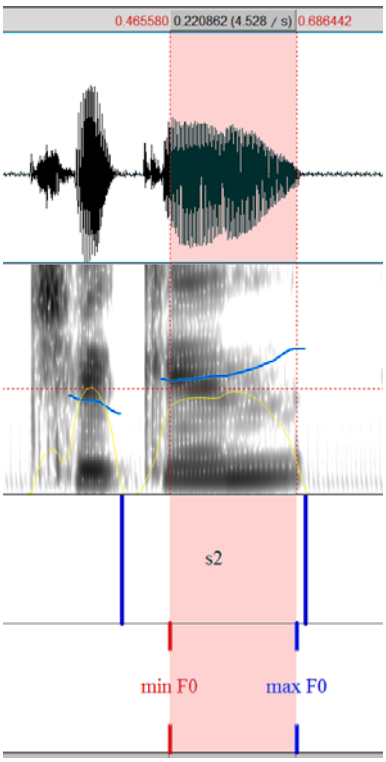



Figure 3.3 Illustration of F0 slope calculation

The calculation of the F0 slope for the example in Figure 3.2 is as follows. First, for the target syllable s2, we obtained the minimum and maximum F0, which is 284.38 Hz and 343.90 Hz (marked as min F0 and max F0 in Figure 3.3), and the times at which they occurred, which is 0.466 and 0.686 as shown in Figure 3.3. Next, we derived the F0 Slope by using the formula introduced above: F0 Slope $= \frac{f_{max}-f_{min}}{t_{max}-t_{min}} = \frac{343.9\ Hz-284.38Hz}{0.686-0.466}$ $= \frac{3.29\ ST}{0.22} = 14.95$ ST/s.

Figure 3.4 gives an example of F0 reset calculation. The minimum F0 values were taken for the syllables before and after the pause (s2 and s3), and the locations of the

measurement were marked as min F0 in Figure 3.4. The resulting F0 reset value = the difference between the two minimum F0 = 248.79 – 200.84 = 47.95 Hz.

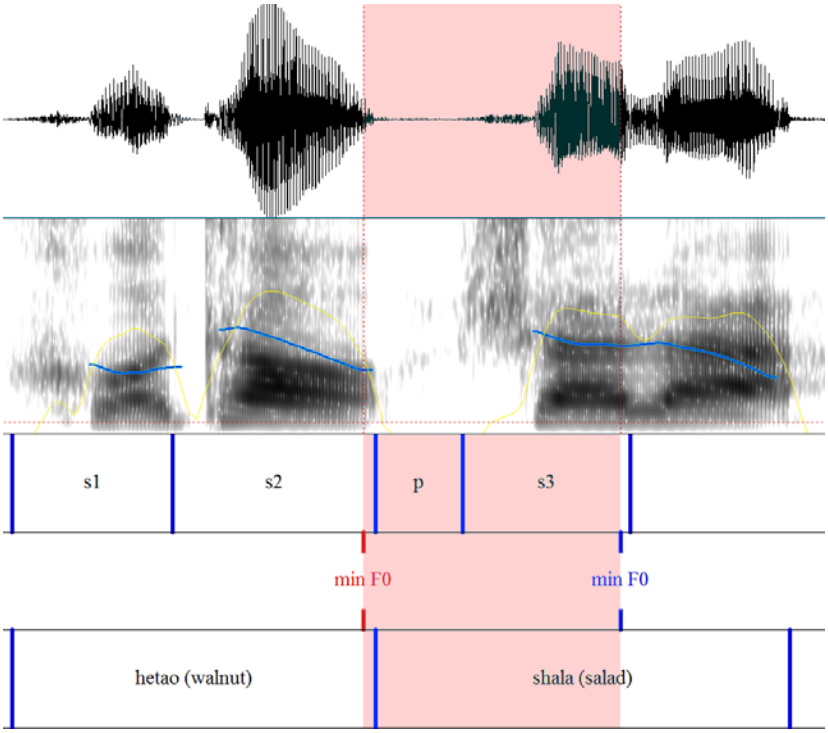



Figure 3.4 Illustration of F0 reset calculation

### 3.3 Results

In the following sections, each acoustic parameter, i.e. pause, pre-boundary syllable duration, F0 slope, and pitch reset are first assessed with mixed effects models (Jaeger, 2008; Baayen, 2008) for statistical significance. Then, a mixed effects logistic regression model (Jaeger, 2008; Baayen, Davidson, & Bates, 2008) is applied to determine which acoustic parameters are effective in predicting the boundary categories.

### 3.3.1 Acoustic parameters

In this section I present the results of the mixed effects models for the four acoustic parameters. Mixed-effects models were chosen because both participants and materials in this study were sampled from a larger population and should be treated as random effects. The inclusion of speaker as a random effect also controls for individual differences between speakers with regard to speech rate, F0 range, and other speaker-specific variability. The random effect of item is used to account partially for variation introduced by vowel-intrinsic and syllable structure differences. It is well-known that different vowel phonemes have different intrinsic F0 and duration—high vowels such as [i] and [u] tend to have higher F0 and shorter duration than low vowels such as [a] (Lehiste, 1970; Whalen & Levitt, 1995). Different syllable structures (i.e. heavy vs. light syllables) also differ in duration, with heavy syllables longer than light ones.

In the following analyses, gender was found to be a significant predictor for all four acoustic parameters, but the effect of language and condition was in the same direction for both genders, with the effect being significantly stronger for females. Given that in these cases the direction of the effect was the same for both genders, and given that we are not primarily interested in gender differences, the inclusion of gender as a main effect only served to control for the effect of gender.

In these models we employed condition (no boundary, with boundary) and language (English, Chinese) as fixed predictors, and gender as a main effect with no further interactions to control for gender-specific differences. Subjects and words were treated as random effects. Because no effect of repetition was found in either of the

groups, I collapsed data across the three repetitions in the final models. In the analyses, data points greater than two standard deviations above or below the mean value of a measurement for a subject were removed from the analyses. These discarded data were treated as missing data points and played no part in the following analyses. Altogether, 5.7% of the trials were discarded.

Overall, we found significant effects for all four parameters, i.e. pause, pre-boundary lengthening, pitch reset, and F0 slope. In these models, the effects of subject and item were also significant, suggesting that there was large variation in subjects and items. Figures 3.5 through 3.8 illustrate the effects of condition and language on the four parameters by means of each parameter as predicted by the models. In what follows we will discuss in more detail the results for each parameter in turn.

*3.3.1.1 Pre-boundary lengthening*

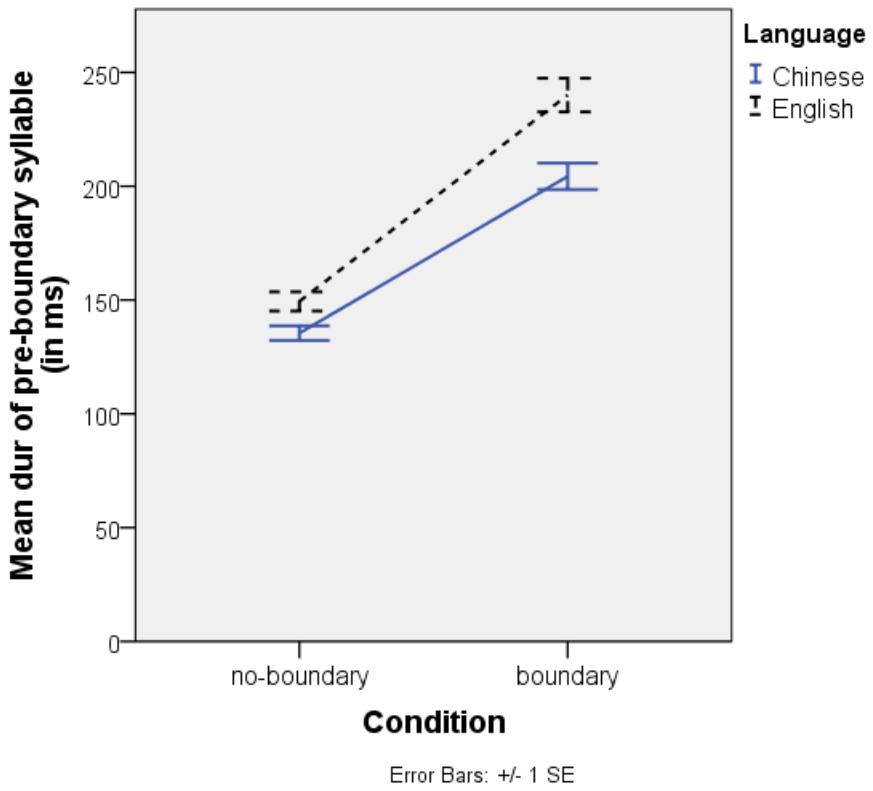



Figure 3.5 Interaction between condition and language in the prediction of pre-boundary syllable duration. Error bars indicate 1 standard error.

The results of the fixed effects analysis on pre-boundary lengthening (Table 3.3) show a significant main effect of Condition, $F(1, 195.622) = 491.930$, $p < .001$, suggesting that the mean syllable duration was significantly longer in boundary ($M = 222.5$, $SD = 67.95$) than in no-boundary ($M = 142.4$, $SD = 38.24$) positions. There was also a significant effect of language, $F(1, 197.364) = 14.546$, $p < .001$, with syllables being significantly longer in English ($M = 194.5$, $SD = 75.09$) than in Chinese ($M = 169.2$, $SD = 57.28$). The significant interaction $F(1, 195.629) = 9.223$, $p = .003$ between Language and Condition (as is shown in Figure 3.5) suggests that the difference between

the two languages in mean syllable duration was greater in the boundary condition than in the no-boundary condition. To summarize, native English speakers utilize pre-boundary lengthening to a significantly greater extent than do Chinese speakers.

Table 3.3 Fixed-effect coefficients in a mixed-effects model for pre-boundary syllable duration. Values in bold are significant at the p<.05 level.

| **Parameter** | **Coefficient** | **Std.Error** | **t** | **Sig.** |
|---|---|---|---|---|
| Intercept[4] | 230.975 | 6.202 | 37.243 | **<.001** |
| Condition [no-boundary] | -91.187 | 5.083 | -17.938 | **<.001** |
| Condition [boundary] | 0 | | | |
| Language [C] | -35.925 | 7.507 | -4.785 | **<.001** |
| Language [E] | 0 | | | |
| Condition [no] * language[C] | 21.964 | 7.232 | 3.037 | **.003** |
| Condition [no] * language[E] | 0 | | | |
| Gender [female] | 19.247 | 6.539 | 2.943 | **.004** |
| Gender [male] | 0 | | | |

Note: The parameters with a coefficient value of 0 are the default reference levels.

*3.3.1.2 Pause*

The results of the fixed effects analysis on pause (Table 3.4) show a significant main effect of Condition, $F(1,394) = 6.683$, $p < .001$, suggesting that pause duration was significantly longer in the boundary (M = 105.53, SD = 84.416) than in the no-boundary (M = 0, SD = 0) positions. There was also a significant effect of Language, $F(1, 394) = 6.683$, $p = .01$, such that silent pause was significantly longer in Chinese (M = 115.64, SD = 87.776) than in English (M = 94.3, SD = 79.5) in boundary condition. The significant interaction $F(1, 394) = 6.746$, $p = .01$ between Language and Condition (as is

[4] The intercept represents the mean duration value for English for the boundary condition, which is set as baseline by SPSS. This is true for the intercept term in the following 3 mixed effects models.

shown in Figure 3.6) suggest that the difference between the two languages in pause duration was greater in the boundary condition than in the no-boundary condition.

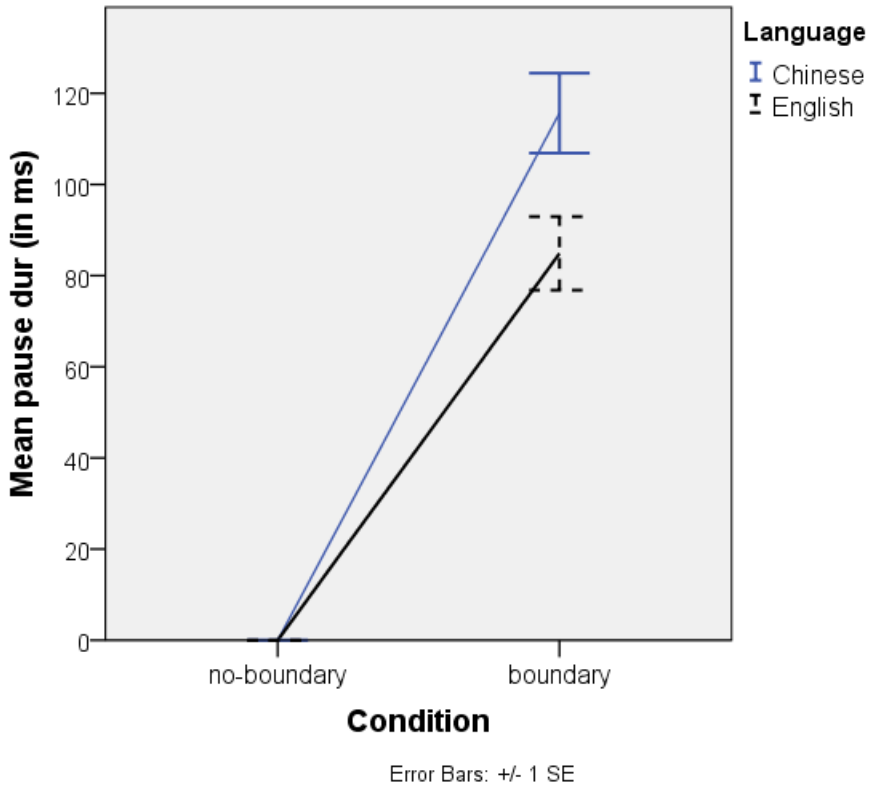



Figure 3.6 Interaction between condition and language in the prediction of pause duration. Error bars indicate 1 standard error of the means.

Table 3.4 Fixed-effect coefficients in a mixed-effects model for pause duration. Values in bold are significant at the p<.05 level.

| Parameter | **Estimate** | **Std.Error** | **t** | **Sig.** |
|---|---|---|---|---|
| Intercept | 92.027 | 6.630 | 13.880 | **<.001** |
| Condition [no-boundary] | -84.800 | 8.405 | -10.089 | **<.001** |
| Condition [boundary] | 0 | | | |
| Language [C] | 30.763 | 8.384 | 3.669 | **<.001** |
| Language [E] | 0 | | | |
| Condition [no] × language[C] | -30.835 | 11.872 | -2.597 | **.010** |
| Condition [no] × language[E] | 0 | | | |
| Gender [female] | -14.309 | 5.936 | -2.411 | **.016** |

Note: The parameters with a coefficient value of 0 are the default reference levels.

*3.3.1.3 F0 Slope*

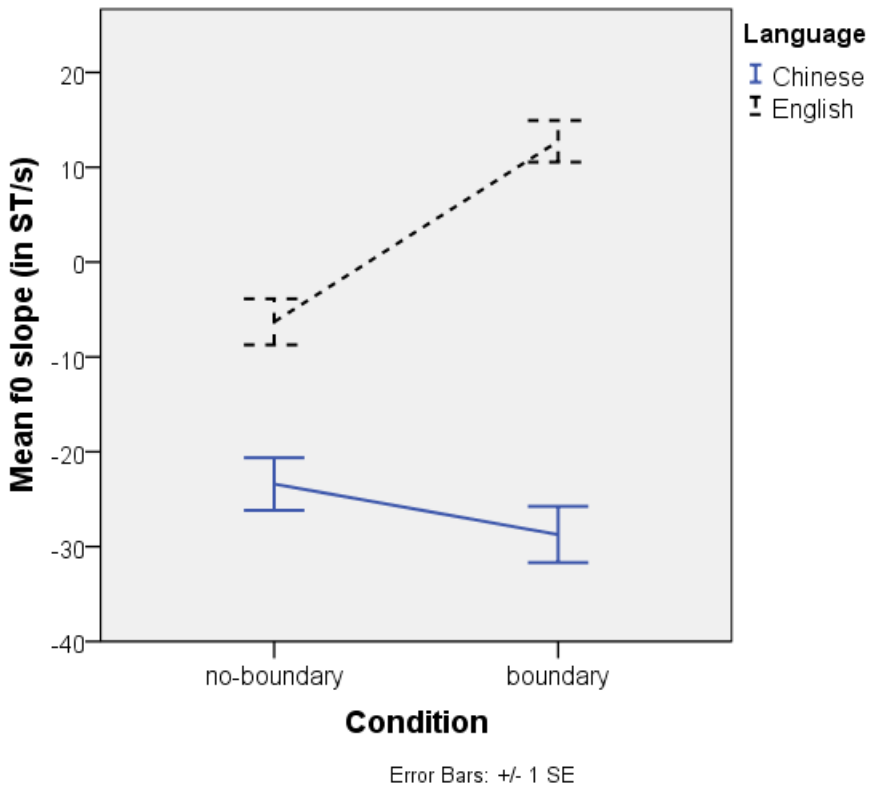



Figure 3.7 Interaction between condition and language in the prediction of F0 slope. Error bars indicate 1 standard error of the means.

Figure 3.7 gives the mean F0 slope in the two languages and conditions. Unlike the above two durational measures, F0 slope showed different patterns for the two languages. In Chinese, the values for both boundary and no-boundary conditions were negative, indicating a falling contour under both conditions, with little difference between the two values. In English, however, F0 slope was negative in no-boundary condition, indicating a slightly falling contour, whereas it was positive in boundary condition, indicating a rising contour.

The results of the fixed effects analysis (Table 3.5) show a significant main effect of Condition, F(1,372) = 6.879, p = .009, with F0 slope value was significantly higher in the boundary (M = -6.84, SD = 31.95) than in the no-boundary (M = -14.89, SD = 27.29) position. There was also a significant effect of Language, F(1, 372) = 129.277, p =<.001, suggesting that F0 slope values were significantly smaller in Chinese (M = -25.86, SD = 27.585) than in English (M = 3.073, SD = 24.657). The significant interaction F(1, 372) = 22.2716, p < .001 between Language and Condition (as is shown in Figure 3.5) is due to the difference between the two languages in pause duration being greater in the boundary condition than in the no-boundary condition.

Table 3.5 Fixed-effect coefficients in a mixed-effects model for F0 slope. Values in bold are significant at the p<.05 level. Note: The parameters with a coefficient value of 0 are the default reference levels.

| Parameter | Coefficient | Std.Error | t | Sig. |
|---|---|---|---|---|
| Intercept | 9.033 | 2.873 | 3.143 | **.002** |
| Condition [no-boundary] | -18.926 | 3.596 | -5.262 | **<.001** |
| Condition [boundary] | 0 | | | |
| Language [C] | -41.475 | 3.729 | -11.121 | **<.001** |
| Language [E] | 0 | | | |
| Condition [no] * language[C] | 24.33 | 5.155 | 4.719 | **<.001** |
| Condition [no] * language[E] | 0 | | | |
| Gender [female] | 7.338 | 2.573 | 2.852 | **.005** |
| Gender [male] | 0 | | | |

*3.3.1.4 Pitch reset*

The results for F0 reset are shown in Figure 3.8. The negative values for the boundary condition in English indicate that no F0 reset was observed. On the contrary, a substantial reset was found under the boundary condition in Chinese, as represented by a positive value. As expected, there was no F0 reset for either language under the no-boundary condition. The results of the fixed effects analysis (Table 3.6) show that there is a significant main effect of Condition, $F(1,194.823) = 17.644$, $p < .001$, suggesting that pitch reset values were significantly greater in boundary ($M = -6.601$, $SD = 36.623$) and in no-boundary ($M = -16.83$, $SD = 23.85$) positions. There was also a significant effect of Language, $F(1, 195.271) = 88.412$, $p =< .001$, suggesting that pitch reset values was significantly greater in Chinese ($M = 1.256$, $SD = 29.29$) than in English ($M = -24.98$, $SD = 27.31$). The significant interaction $F(1, 194.803) = 27.061$, $p = <.001$, between Language and Condition (as is shown in figure 3.4) is due to the difference between the two languages in pitch reset being greater in the boundary position than in the no-boundary position.

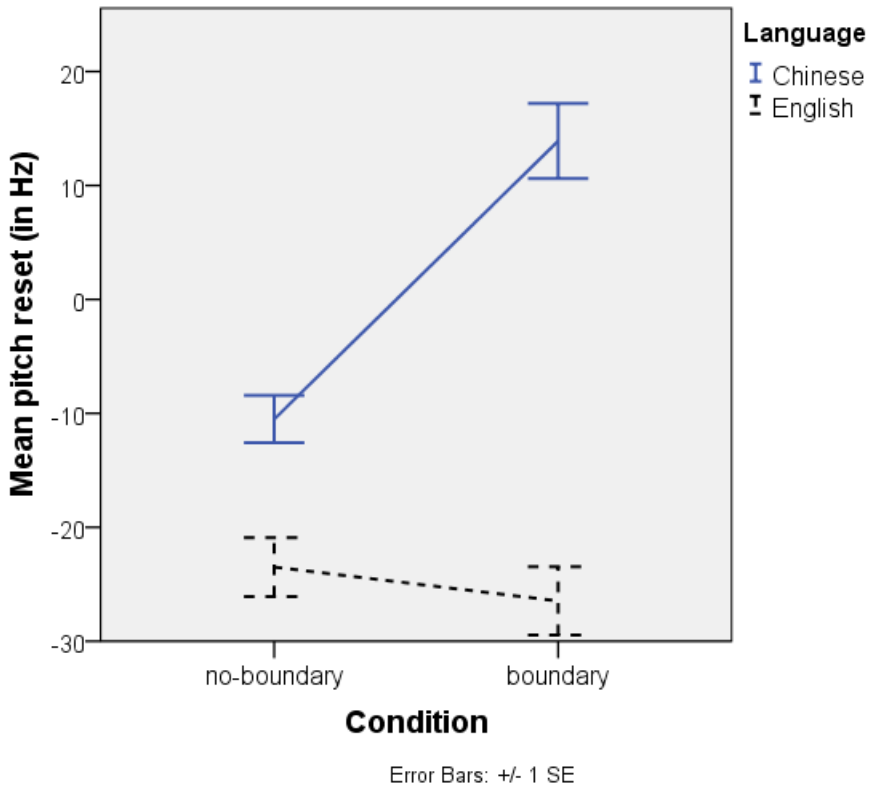



Figure 3.8 Interaction between condition and language in the prediction of F0 reset. Error bars indicate 1 standard error of the means.

Table 3.6 Fixed-effect coefficients in a mixed-effects model for pitch reset. Values in bold are significant at the p<.05 level.

| Parameter | Coefficient | Std.Error | t | Sig. |
|---|---|---|---|---|
| Intercept | -21.963 | 3.073 | -7.147 | **<.001** |
| Condition [no-boundary] | 2.597 | 3.671 | 17.644 | **<.001** |
| Condition [boundary] | 0 | | | |
| Language [C] | 40.289 | 3.881 | 10.379 | **<.001** |
| Language [E] | 0 | | | |
| Condition [no] * language[C] | -26.95 | 5.181 | -5.202 | **<.001** |
| Condition [no] * language[E] | 0 | | | |
| Gender [female] | -8.95 | 2.851 | -3.139 | **.002** |
| Gender [male] | 0 | | | |

Note: The parameters with a coefficient value of 0 are the default reference levels.

To summarize, the two durational parameters, namely pause and pre-boundary lengthening, showed significant effects of both condition (with boundary or no-boundary) and language type. Both English and Chinese utterances show larger durational values in the boundary condition than in the no-boundary condition. In contrast, the two pitch parameters, F0 slope and F0 reset were found to be significantly affected by one of the languages. English production showed large differences in F0 slope between the two boundary conditions, while Chinese production showed differences only in F0 reset.

### 3.3.2 Predicting boundary categories on the basis of acoustic parameters

In this subsection we determine how well we can predict the boundary categories on the basis of the four acoustic parameters inspected in the previous section. For each language, a mixed effects logistic regression model (Jaeger, 2008; Baayen et al., 2008) was applied to determine which acoustic parameters are effective in predicting the prosodic boundary categories by specifying individual speakers and test items as random variables. Mixed-effects models were chosen because both participants and materials in this study were sampled from a larger population and should be treated as random effects.

Logistic Regression is a type of predictive model that can be used when the dependent variable is binary (such as in the case of this study where there were two boundary categories) and the independent variables can be continuous, categorical, or both.

The logistic regression model explains the probability of a single event as a function of one or more independent predictor variables as in:

$$Prob(event) = \frac{1}{1 + e^{-(\beta_0 + \beta_1 X_1 + \beta_2 X_2 + \cdots \beta_n X_n + \varepsilon_i)}}$$

where

- Prob (event) is the probability that a single event may happen (i.e. in this study, a utterance has a prosodic boundary);
- $e$ is the base of the natural logarithm;
- $\beta_0$ is a constant;
- $\varepsilon$ is a residual term;
- $X_1$ , $X_2$… $X_n$ are the predictor variables. These predictor variables can be either categorical or continuous;
- $\beta_1$ , $\beta_2$ … $\beta_n$ are coefficients attached to the predictor variables. These coefficients indicate the weight of each predictor variable's contribution to the probability of the event. The sign on a coefficient $\beta_n$ indicates the direction of the effect, so that a positive $\beta_n$ coefficient increases the probability of the event, while a negative $\beta_n$ coefficient decreases the probability of the event. Instead of simple β, exponential β (Exp (β)) is always used in logistic regression as the independent coefficient. Exponential β provides an odd ratio for the dependent variable based on the independent variables.

*3.3.2.1 Logistic regression results*

In these models the dependent variables were the two boundary conditions, and we employed the four prosodic correlates, i.e. pause, pre-boundary lengthening, pitch

reset, and F0 slope as fixed predictors. Gender was found not to be a significant predictor and was removed from the model. Speakers and words were treated as random effects. Because no effect of repetition was found in either of the groups, we collapsed data across the three repetitions in the final models.

*English*

Table 3.7 describes the results for the logistic regression of English productions, with the dependent variables being boundary condition, coded as 0, and no-boundary condition, coded as 1.

Table 3.7 Summary of the mixed effects logistic regression model for English productions. Values in bold are significant at the p<.05 level.

| **Parameter** | **Coefficient** | **Std.Error** | **t** | **Sig.** | **Exp(β)** |
|---|---|---|---|---|---|
| Intercept | 4.853 | 1.054 | 4.605 | **<.001** | 128.110 |
| Pause | -0.021 | 0.007 | -2.893 | **.004** | 0.980 |
| Pre-boundary | -0.018 | 0.005 | -3.456 | **.001** | 0.983 |
| F0 reset | 0.015 | 0.009 | 1.598 | .112 | 1.015 |
| F0 slope | -0.034 | 0.011 | -3.133 | **.002** | 0.967 |

Three of the four acoustic parameters, pause duration, pre-boundary lengthening and F0 slope, are significant in predicting the boundary categories in English. The negative coefficients in the table indicate a negative correlation with no-boundary condition, which was expected since the no-boundary condition should have lower values for pause duration, pre-boundary syllable duration, and F0 slope than does the boundary condition, as illustrated in Figures 3.5-3.8. F0 reset was not a significant predictor.

*Chinese*

Table 3.8 Summary of the mixed effects logistic regression model for Chinese productions. Values in bold are significant at the p<.05 level.

| **Parameter** | **Coefficient** | **Std.Error** | **t** | **Sig.** | **Exp(β)** |
|---|---|---|---|---|---|
| Intercept | 3.063 | 0.894 | 3.426 | **.001** | 21.401 |
| Pause | -0.024 | 0.006 | -4.064 | **<.001** | 0.976 |
| Pre-boundary | -0.012 | 0.005 | -2.239 | **.026** | 0.998 |
| F0 reset | -0.035 | 0.017 | -2.033 | **.044** | 0.965 |
| F0 slope | -0.002 | 0.008 | -0.281 | .779 | 0.998 |

Table 3.8 presents the results for the logistic regression of Chinese productions. The acoustic parameters of pause duration, pre-boundary lengthening and F0 reset are significant in predicting boundary categories in Chinese. Since pause duration, pre-boundary syllable duration, and F0 reset all have lower values in the no-boundary condition than in the boundary condition, their coefficients are negative. F0 slope was found not to be a significant predictor.

To summarize, speakers of both languages produce significant distinctions in different aspects of the acoustic correlates of prosodic boundaries. Both English and Chinese speakers produce longer pre-boundary pauses and longer pre-boundary syllables. However, these speakers' productions differ in pitch dimensions—English speakers produce consistent differences between the two boundary categories in F0 slope, while Chinese speakers mainly use F0 reset as the pitch device to distinguish the two categories.

*3.3.2.2 Relative weight analysis*

The above analyses mainly investigated what phonetic cues made significant contribution to the production of prosodic boundaries in both languages. In this section, a relative weight analysis was used to determine the relative importance of these predictors within each language (Johnson & LeBreton, 2004; Tonidandel & LeBreton, 2011). A relative weight analysis examines "the proportionate contribution each predictor makes to $R^2$ considering both its individual effect and its effect when combined with other variables in a regression equation" (Johnson & LeBreton, 2004). A relative weight analysis supplements a logistic regression analysis and takes into account collinearity issues. It examines the comparative usefulness of new variables, and determines which variable or variables are primarily driving the $R^2$. Relative weights are calculated by creating a new set of uncorrelated predictors that are maximally related to the original set of correlated predictors and both sets of variables are used to estimate importance (Johnson, 2000).

We conducted Relative Weight Analysis using SAS and publicly available macros (Tonidandel & LeBreton, 2011) that produce relative weights from given data. A relative weight analysis was performed to determine the importance of the phonetic cues in predicting the presence of a prosodic boundary in both English and Chinese (see Table 3.9 and 3.10). Results showed that, for English speakers, pause had the highest relative weight (.65) and is the most important in predicting the presence of a prosodic boundary (accounting for 78% of explained variance). Next was pre-boundary lengthening (.15), accounting for 18 % of the $R^2$. F0 slope explains little variance (4%) in the production of a prosodic boundary, with a relative weight of 0.028.

Table 3.9 Relative weight analysis of boundary production in English

| Relative Weights Analysis of boundary production in English (Criterion = boundary condition) | | |
|---|---|---|
| | *Raw relative weights* | *Relative weights as percentage of* $R^2$ |
| Pause | 0.65 | 0.78 |
| Lengthening | 0.15 | 0.18 |
| F0 slope | 0.028 | 0.04 |

Table 3.10 Relative weight analysis of boundary production in Chinese

| Relative Weights Analysis of boundary production in Chinese (Criterion = boundary condition) | | |
|---|---|---|
| | *Raw relative weights* | *Relative weights as percentage of* $R^2$ |
| Pause | 0.59 | 0.75 |
| Lengthening | 0.11 | 0.14 |
| Pitch reset | 0.083 | 0.11 |

For Chinese speakers, pause also had the highest relative weight (.59) and is the most important in predicting the presence/absence of a prosodic boundary (accounting for 75% of explained variance). Pre-boundary lengthening in Chinese is also more important than pitch cue (pitch reset), but the difference between the two (a 3% difference in terms of the variance explained) is smaller than in English (a 14% difference).

### 3.3.3 Individual speaker analysis

The results presented above are based on the average values across all speakers. A closer look at each individual's data revealed that different speakers appear to make different use of phonetic cues to mark prosodic boundaries. Some speakers made more extensive use of all three acoustic cues than others in a systematic fashion. Some speakers, in contrast, employed just two of the three cues to signal prosodic boundaries. It

is therefore interesting to examine the production pattern across speakers in each language. In what follows we will discuss the individual variation in the realization of the prosodic boundary with regard to the employment of the three acoustic cues in each language.

*3.3.3.1 Pause*

The presence of a silent pause has been considered the most salient cue in both production and perception studies. However, the individual data revealed that not all speakers make use of pause to mark a prosodic boundary. As shown in Figure 3.9, which illustrates the difference between pause duration in the boundary and no-boundary conditions,[5] three English speakers produced little difference in pause duration between the two boundary conditions. For those who did use pause, the range of the average pause duration is large (56.3ms–176.3ms). Moreover, the three speakers who did not use pause used different strategies to repair for the lack of this cue. In the boundary condition, Speaker 2's production displayed a very sharp F0 rise of the pre-boundary syllable, while Speaker 5 tended to lengthen the pre-boundary syllable. [6]

A cross-speaker variation in the use of silent pause can also be observed in the Chinese data (see Figure 3.10). The length of the duration again covered a wide range across 10 speakers, ranging from 4.7 ms to 254.3 ms. High variability both within and across speakers has also been supported in previous studies (Cooper & Paccia-Cooper,

[5] Note that pause duration in the no-boundary condition is always 0 ms, so the pause difference is basically the duration of silent pause in the boundary condition.

[6] Speaker E10 was unusual in that he did not make a distinction in the production of the utterances in the two boundary conditions, which could be seen as nearly zero difference between the values of his pause difference and pre-boundary syllable difference in Figure 3.9 and 3.11. However, he confirmed that he understood the task and produced utterances according to the pictures. He was therefore not excluded from the statistical analysis, but he will not be discussed in the individual analysis.

1980). Thus there seems to be a wide range of acceptable pause durations for phrase boundaries.

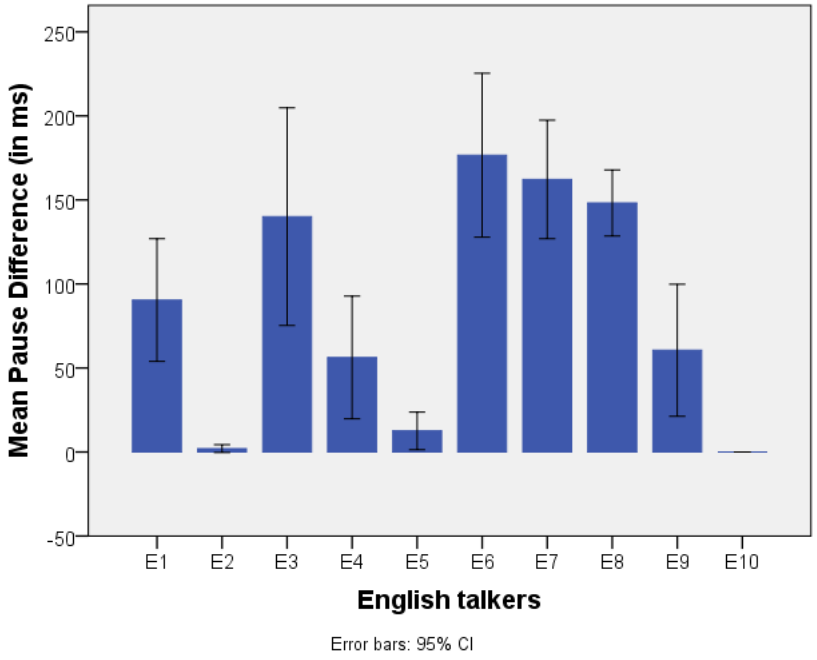



Figure 3.9 Mean duration of pause difference (boundary condition – no-boundary condition) under the two boundary conditions for ten native English speakers.

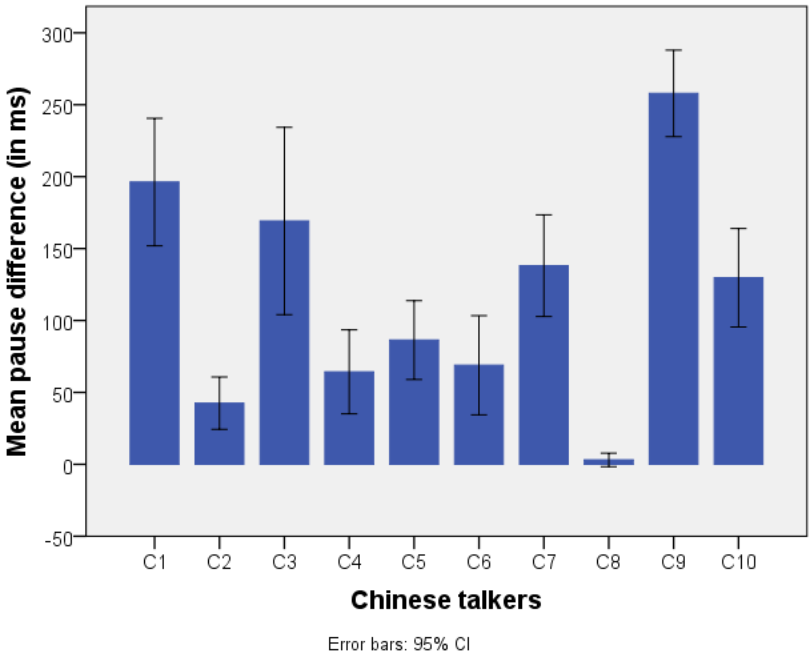



Figure 3.10 Mean duration of pause difference (boundary condition – no-boundary condition) under the two boundary conditions for ten native Chinese speakers

### *3.3.3.2 Pre-boundary lengthening*

Figures 3.11 and 3.12 display the difference in the mean duration of pre-boundary syllables, which was derived by subtracting the duration of the pre-boundary syllable in the boundary condition from that of the same in boundary position. It is therefore a measure of the degree of the lengthening effect. Unlike pause duration, pre-boundary lengthening was consistently found across English and Chinese speakers' production (except E10, see footnote 5), although the extent of the lengthening differed across speakers. In English, the lengthening could be as long as 163 ms, or as short as 39 ms. In Chinese, three out of 10 speakers lengthened pre-boundary syllable by less than 50 ms in the boundary condition as compared to the syllable in the no-boundary condition.

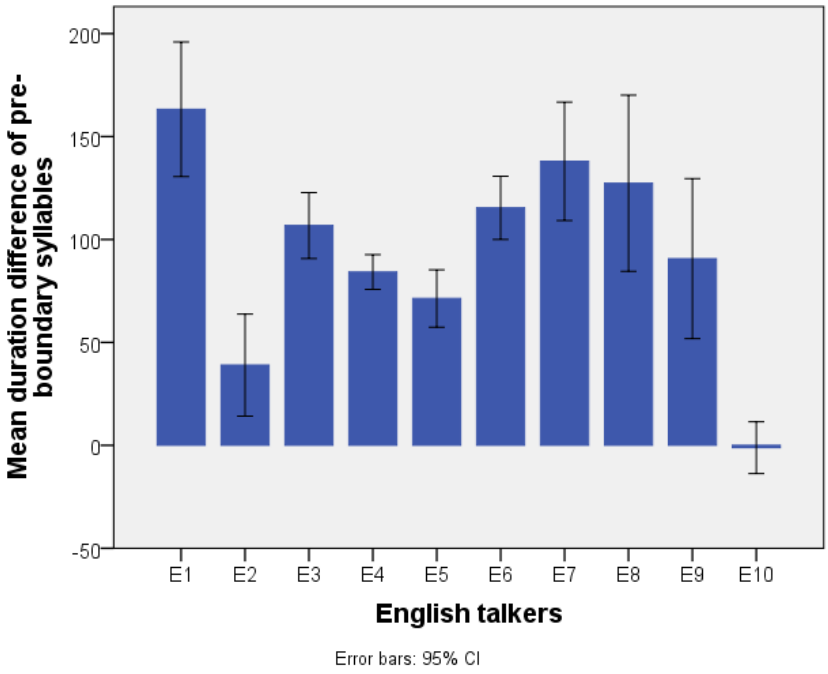



Figure 3.11 Mean duration difference (boundary condition – no-boundary condition) of pre-boundary syllable for ten native English speakers

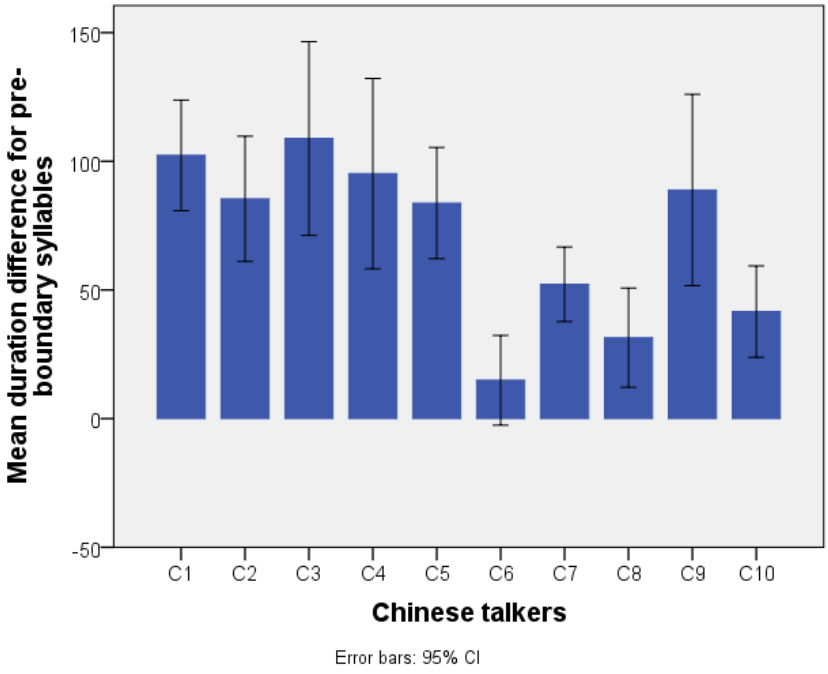



Figure 3.12 Mean duration difference (boundary condition – no-boundary condition) of pre-boundary syllable for ten native Chinese speakers

*3.3.3.3 Pitch cues*

The statistical analyses in section 3.3.1 showed that pitch information is represented differently in English and Chinese: English speakers produce consistent differences between the two boundary categories in F0 slope, while Chinese speakers mainly use F0 reset as the pitch device to distinguish the two categories. Figures 3.13 and 3.14 show the mean F0 slope values across English speakers and F0 reset values across Chinese speakers, respectively. Again, there is substantial variation across speakers. English speakers E4 and E9 barely utilized rising pitch contour in their productions (Figure 3.13). Similarly, some Chinese speakers did not employ F0 change as represented by pitch reset in their realization of the phrase boundary (Figure 3.14).

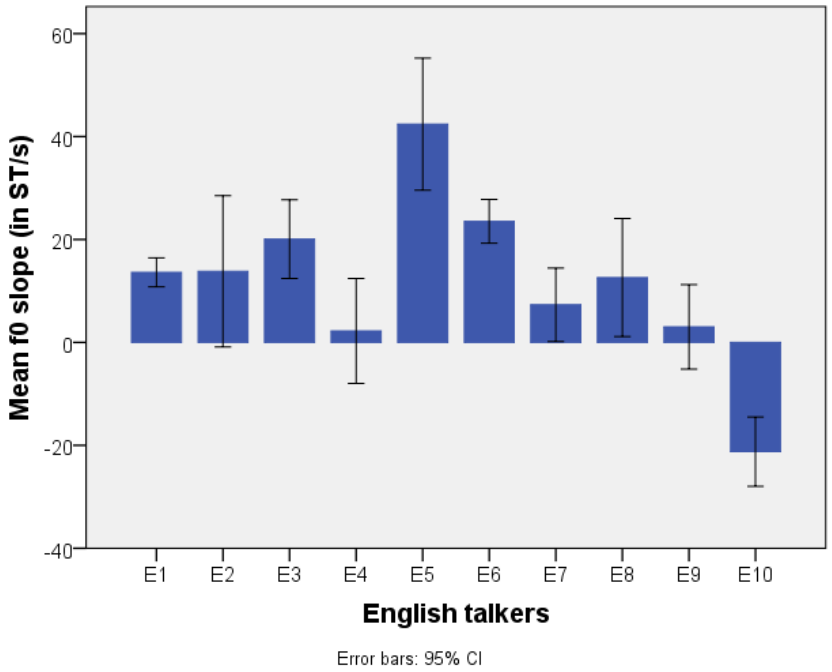



Figure 3.13 Mean F0 slope in the boundary condition for ten native English speakers

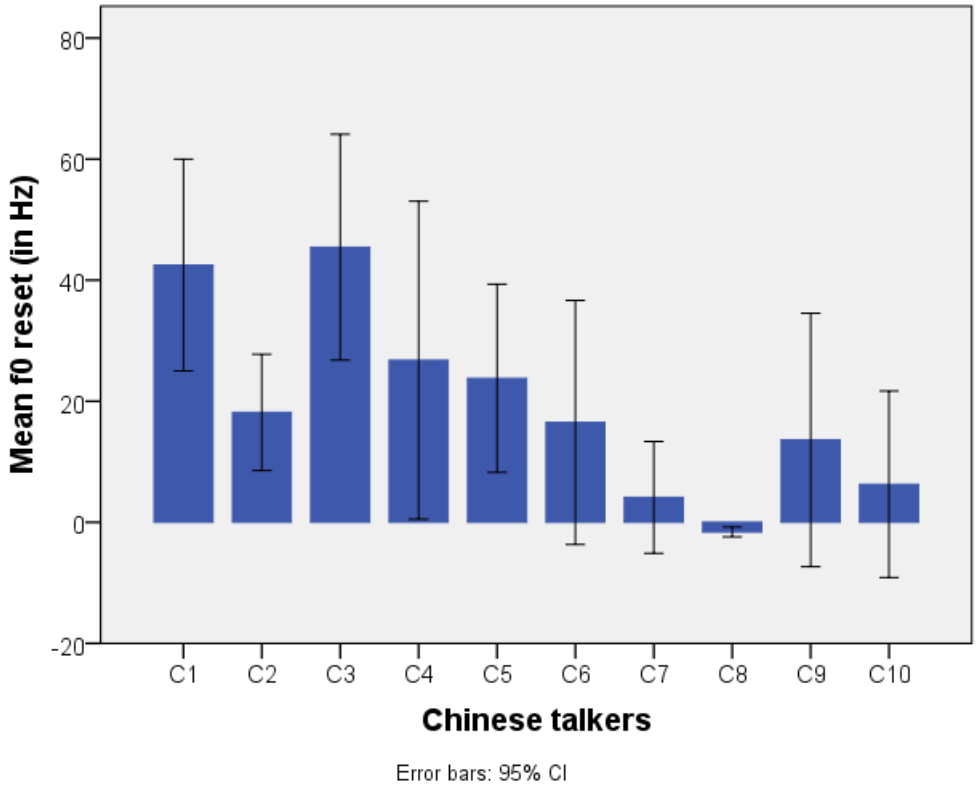



Figure 3.14 Mean F0 reset in the boundary condition for ten native Chinese speakers

*3.3.3.4 Relationship between cues*

The above analysis of speaker variation suggests that combinations of the three cues can be considered as possible phonetic strategies of the speakers to mark prosodic boundaries, with speakers differing in their selection of cue combinations. It seems that the lack of one cue is sometimes compensated for by the extensive use of other cues. For example, neither of English speakers E2 and E5 used silent pause in their realization of a prosodic boundary (Figure 3.9). They also use pre-boundary lengthening to a lesser degree compared with other speakers (Figure 3.11). For at least speaker E5, the use of the pitch cue may be compensating. Speaker E5 displayed a very sharp rise of the pre-

boundary syllable (the largest F0 slope among all the speakers; Figure 3.12). For speaker E2, although the average value of her F0 slope was not especially high, the large variation in her data (as represented by the 95% CI) indicated that she realized prosodic boundaries with a sharp rise for at least some of the tokens.

However, not all speakers used this compensation strategy for cues in their productions. Some speakers made limited use of all the cues (e.g. Chinese speakers C6 and C10), and other speakers made extensive use of all the cues (e.g. English speakers E3 and E6; Chinese speakers C1 and C3).

### 3.4 Summary and discussion

The present study examined the production of the prosodic characteristics that serve as acoustic correlates (pause duration, pre-boundary lengthening, and pitch change) of a specific type of prosodic boundaries (list vs. non-list) in English and Chinese. The results of the present study demonstrated that speakers of both languages utilized durational and pitch cues to signal this type of phrase boundary. As expected from previous research, significant effects of durational cues (pause and pre-boundary lengthening) were found for speakers of both languages, above the speaker and item variation. Pitch cues were also significant predictors of prosodic boundaries. However, the two languages displayed different F0 patterns in the distinction of boundary categories. In English, F0 slope was found to be an effective predictor for boundary categories, which is determined by the particular pitch pattern of the tested utterance type. Specifically, a rising tone is normally used when speakers produce a list of items. In Chinese, pitch contour is tied up at the word level and can thus no longer be modified as

freely at the level of prosodic patterning. Therefore, F0 slope (representing the rising and falling pitch contours) will not make a difference in the two boundary conditions, and rather it is represented by a reset of the pitch declination.

The relative importance of the acoustic cues in English and Chinese was investigated in a relative weight analysis. The results showed that both English and Chinese speakers considered pause as the most important cue in producing a prosodic boundary, and the relative importance assigned to pause was larger in English (78%) than in Chinese (75%). They also weighted pre-boundary lengthening more heavily than pitch cues (F0 slope and pitch reset in English and Chinese respectively), but the difference between the two cues was larger for English than for Chinese speakers, with Chinese speakers relying more on pitch (11%) than English listeners (4%)

Analyses on the individual speakers revealed large speaker variation with respect to the use of the three acoustic cues. It was found that while some speakers made extensive use of all available cues, there are also speakers who only used a subset of available cues. Speakers could choose to compensate for the lack of a certain cue by increasing the use of other cues.

# CHAPTER IV

# PERCEPTION STUDY

## 4.1 Introduction

This chapter describes two perception experiments designed to explore the relative importance of pause, pre-boundary lengthening, and pitch contour in the perception of prosodic boundaries by native speakers of Chinese and English. These particular target cues were chosen based on their roles in the realization of prosodic boundaries in the acoustic study. Pause and lengthening of the pre-boundary syllable were realized similarly by native Chinese and English speakers in terms of both range and magnitude. The role of F0 cues in the perception of prosodic boundaries by native speakers of Chinese and English is of particular interest, due to its phonemic status in Chinese, and the different realizations found in the production experiments.

This chapter presents the methods and results of the perceptual investigations. Section 4.2.1 outlines the manipulation of the three cues, F0, pause duration and the lengthening of the pre-boundary rime on the target words to create the 100 test stimuli. Section 4.2.2 introduces the two groups of participants, native English speakers and native Chinese speakers. Section 4.2.3 outlines the detailed procedure of the perception experiments. In Section 4.3, experimental results and data analysis are presented, and section 4.4 summarizes the chapter.

## 4.2 Method

### 4.2.1 Stimuli construction

The stimuli were manipulated versions of the naturally produced utterances from the production experiment to ensure that they sound natural while making systematic variations of the intended acoustic cues possible.

One female speaker from each language was chosen based on the criterion that their production was representative production among all speakers and that they employed all the three cues clearly. In the choice of appropriate original utterances, the criterion was that target syllables (second syllable in the first word) are similar in syllable structure in the two languages. As a result, the English pair *Turkey salad and coffee* and *Turkey, salad, and coffee*, and the Chinese utterance pair *Mogu shala he hongjiu* ('Mushroom salad and wine') and *Mogu, shala, and hongjiu* ('Mushroom, salad, and wine') were chosen, because both target syllables /ki/ and /gu/ have a vowel rime. Although the vowels are different, they are the best match among the available rimes.

Two series of stimuli from each language were created: the first series starts with the original two-item reading (hence having a no-boundary pitch pattern), with gradually lengthened pause duration and pre-boundary syllable duration; the second series starts with the original three-item reading (hence having a boundary pitch pattern), with gradually shortened pause duration and pre-boundary syllable duration.

The manipulation procedure for each series is described below.

*4.2.1.1 Manipulation of pause duration*

The manipulation values for pause duration were determined based on the measurements obtained from the natural production data in the production experiments. See Table 4.1 for details.

Table 4.1 Min, Max and Average pause durations (in ms) under the boundary condition in English and Chinese.

| | **English** | | **Chinese** | |
|---|---|---|---|---|
| | Dur for all speakers | Dur for the target speaker | Dur for all speakers | Dur for the target speaker |
| **min** | 0 | 46 | 0 | 0 |
| **max** | 304 | 158 | 326 | 147 |
| **average** | 83 | 121 | 98 | 86 |

Note: the pause durations under the no-boundary condition are not included in the table as they were 0 in the majority of trials.

Because the longer the pause is, the more salient the presence of a boundary, the maximum value for the pause duration was set at 80 ms based on the average data in the two languages, in order that the pause cue would not be too strong and override the contribution of the other two cues. Five different levels of pause duration were created. Rather than distributing the five levels evenly, the concentration was denser in the region of ambiguity (when the pause was especially short), and was sparser in the more salient region (at longer pause durations). This uneven distribution of pause durations was used in order to keep the total number of tokens at a reasonable number. The resulting five levels are 0, 10, 20, 40, and 80 ms.

*4.2.1.2 Manipulation of pre-boundary lengthening*

In the production experiments, lengthening of the pre-boundary syllable was shown to be a significant acoustic correlate of boundary presence. However, the amount of lengthening was not constant across the syllable; rather the rime lengthened more than the onset. This unequal lengthening has also been found in previous studies of the temporal scope of the pre-boundary lengthening effect. Berkovits (1994) examined lengthening of phrase-final disyllabic words with initial stress in Hebrew and he found that the lengthening in the segments increases as the boundary is approached. Turk (1999) also reported that in English significant lengthening mainly affects the rimes, not the onsets. As a result, lengthening or shortening the entire syllable sounded unnatural. Consequently, the duration manipulation was applied only to the rime.

The manipulation values for pre-boundary lengthening were determined based on the results of previous studies and also measurements obtained from the natural production data in the production experiments.

Table 4.2 gives the min, max and average rime duration of the productions of the selected speakers in the two languages. These data are based on the measurements taken from the monosyllabic rimes only (i.e. “i” in ‘juzi’, ‘qiezi’, and ‘yezi’, and “u” in ‘mogu’; [i] in ‘turkey’and ‘kiwi’, [ɚ] in ‘ginger’, ‘tuna’, and ‘pepper’(*pasta* was not included due to its extremely short duration). Measurements from all speakers are not included because they were not available from the production experiments, where the duration of the syllable (rather than rime) was measured.

Table 4.2 Min, Max and Average rime duration (Dur., in ms) in English and Chinese.

| | English | | Chinese | |
|---|---|---|---|---|
| | 2-item | 3-item | 2-item | 3-item |
| **min** | 71 | 175 | 77 | 158 |
| **max** | 98 | 214 | 109 | 181 |
| **average** | 87 | 197 | 89 | 169 |

The endpoints of the rime duration were determined based on the typical duration of the 2-item and 3-item readings. The resulting five levels are 80, 105, 130, 155, and 180 ms.

*4.2.1.3 Manipulation of F0*

Ideally, F0 movement would be manipulated in a 5-step continuum to investigate the contribution of F0 change to boundary perception. However, as was shown in the production experiment, F0 movement was realized differently in English and Chinese in this particular utterance structure – English has different pitch contours as represented by different F0 slopes under two versus three-item conditions, while Chinese relies on pitch reset. Pitch cues were therefore represented by only two steps, corresponding to the naturally produced utterances of the 2-item and 3-item conditions, representing the no-boundary and boundary conditions, respectively.

*4.2.1.4 Manipulation of the post-boundary part*

Previous studies have also shown that post-boundary lengthening and post-boundary pitch change was also possible cues of prosodic boundaries. According to Strangert (1990, 1992), both pre- and post-boundary information are important cues when perceiving prosodic phrase boundaries in Swedish. She also showed that it is possible to

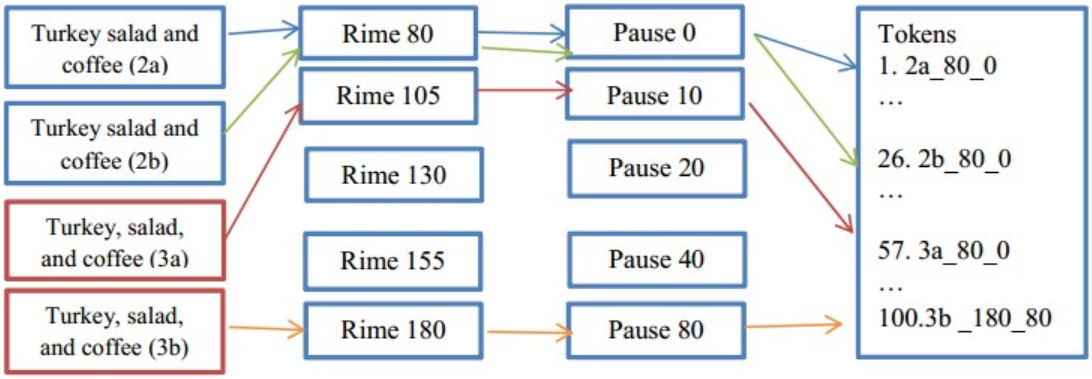
differentiate between different types of syntactic boundaries on the basis of the pre- and post-boundary cues alone (Strangert, 1992).

Although post-boundary cues are not the focus of this study, I chose to control for possible post-boundary influences. Every manipulated stimulus had a counterpart that differed in the post-boundary elements: stimuli from the original 2-item reading were replaced with the post-boundary portion from the 3-item reading, and stimuli from the original 3-item reading were replaced with the post-boundary portion from the 2-item reading.

*4.2.1.5 Manipulation process*

One sample utterance from each condition (2-item and 3-item) was selected for further duration manipulation. Firstly, rime duration was altered to create 5 steps using the Time-Domain Pitch-Synchronous Overlap-and Add (TD-PSOLA) manipulation method implemented in Praat. This step produced 10 stimuli: 5-step rime durations for 2-item (no-boundary) pitch condition, and 5-step rime durations for 3-item (boundary) pitch condition. Each resulting stimulus was then manipulated to create a 5-step pause duration continuum by adding or deleting silence from the pause duration in the target area. The resulting 50 stimuli were then manipulated for the post-boundary control by replacing the post-boundary portion with their pitch-condition counterpart (i.e. post-boundary sections from the original 2-item condition were replaced with those from the original 3-item condition and vice versa). This produced the 100 stimuli for the perception experiment. The following chart shows the steps to construct the stimuli.

Table 4.3 Construction of the 100 test tokens




(Note: "2" indicates that the pitch contour of the token is based on 2-item (no-boundary) reading, while "3" indicates that the pitch contour is based on 3-item (boundary) reading. "a" indicates the post-boundary parts are from the original utterance, and the post-boundary parts of "b" are from the counterpart pitch condition.)

The following two figures illustrate the endpoint tokens in each language, i.e. the token with the shortest rime (80 ms) and pause duration (0 ms) in the no-boundary pitch condition vs. the token with the longest rime (180 ms) and pause duration (80 ms) in the boundary pitch condition. We can see that the pitch movement differs between the manipulated tokens in each language. Figure 4.1 shows that F0 change in Chinese was represented by pitch reset after a prosodic boundary. The reset value after the prosodic boundary is 36.9 Hz, which was calculated by subtracting the minimum F0 of the pre-boundary syllable (210.9 Hz) from that of the post-boundary syllable (247.8 Hz). In comparison, F0 change in English was represented by pitch contour change: falling contour in the no-boundary condition, and rising contour in the boundary condition due to listing tone. The difference between the maximum and minimum F0 values in the no-boundary condition is 23.5 Hz (note that $f_{max}$ occurrs before $f_{min}$), and this difference in the boundary condition is 15.4 Hz ($f_{max}$ occurrs before $f_{min}$ instead).

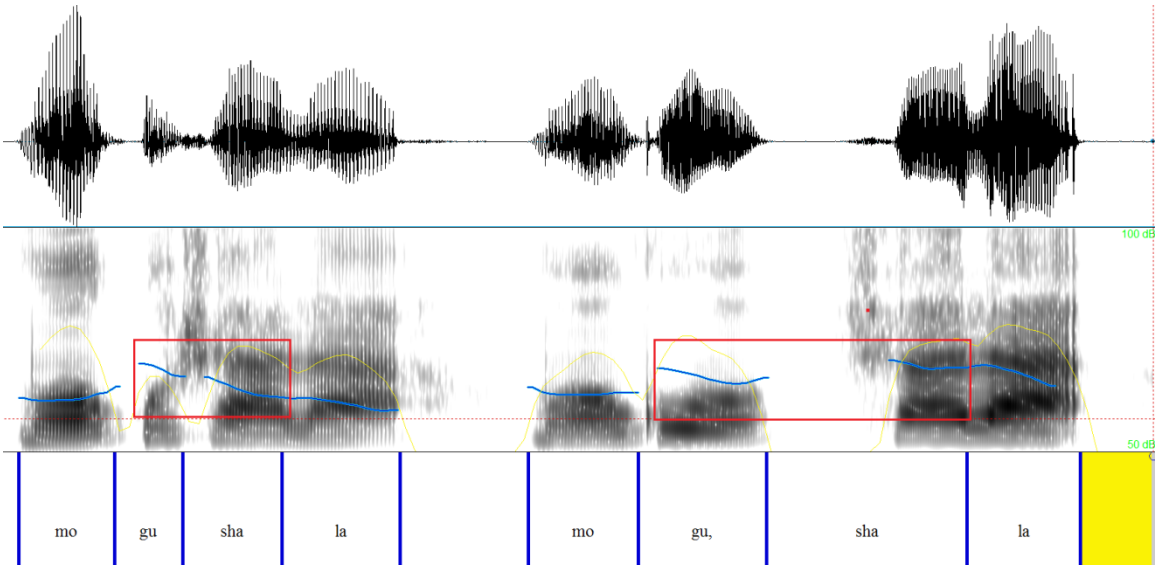



Figure 4.1 Illustration of F0 patterns in two boundary contexts in Chinese. The utterance on the left corresponds to the no-boundary condition, and that on the right to the boundary condition.

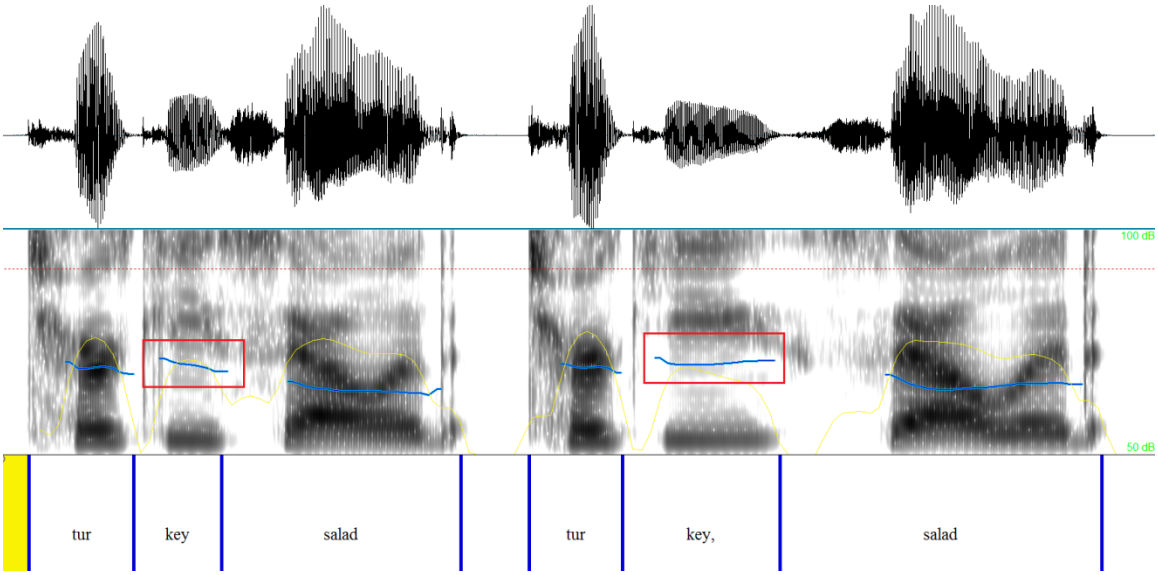



Figure 4.2 Illustration of F0 patterns in two boundary contexts in English (represented by F0 contour change). The utterance on the left corresponds to the no-boundary condition, and that on the right to the boundary condition.

### 4.2.2 Participants

Two groups of listeners participated in the study. The English group consisted of twenty (ten male and ten female) native English speakers. The Chinese group had twenty (ten male and ten female) native Chinese (specifically Beijing Mandarin) speakers. They were paid to participate in the study.

All English participants were students at the University of Michigan and were between 18 and 22 years of age. Most of them came from the Midwest, more specifically from Michigan. All Chinese listeners were students at the Central University of Finance and Economics in Beijing, China and were between 18 and 20 years of age. All of them were born and brought up in Beijing, China.

### 4.2.3. Procedure

The perception experiment was a forced choice identification task, run on a Mac book via Superlab software 4.5 (Cedrus Corporation). Participants were tested individually in sound-attenuated booths. The English experiment was conducted in the sound room at the Department of Linguistics in the University of Michigan, and the Chinese experiment was done in a language lab in the Central University of Finance and Economics in Beijing, China. The apparatus (headphones, response pad, and the laptop), listening environment, and procedure were the same for the two experiments and the experiments were conducted by the researcher.

Stimuli were auditorily presented to participants over AKG headphones. The order of the stimulus presentation was differently randomized for each participant.

Stimulus presentation and response recording were controlled using a RB 620 response pad (SuperLab Pro, Cedrus Corporation).

Before the experiment started, the experimenter explained to the participants that he or she would listen to a series of utterances that might contain two or three items. This was done by showing the participants pictures containing a utterance pair used in the production experiment, but not in the perception test. They were instructed to listen to the utterances carefully, and then decide whether the utterance contained two or three items. They recorded their judgments by pressing one of the labeled buttons ("2" or "3") on a response pad. It was stressed that these responses were to be made as quickly and accurately as possible. Reaction times were also collected for future analysis.

Every experimental trial for the identification experiment had the following structure. Listeners heard a single stimulus drawn from the stimuli set. When the stimulus finished playing, a visual prompt (a blank page) appeared on the screen, to prompt listeners to answer. Listeners pressed one of two labeled buttons to indicate how many items they thought the utterance they heard contained. The entire interval during which listeners could enter their response was 3000 ms. If no response was collected during this interval, the software automatically recorded an incorrect response and presented the next trial. After the listener responded or after the 3000 ms interval had elapsed, the software waited an additional 500 ms before presenting the next stimulus.

Prior to the experimental trials, there was a practice session consisting of 10 practice trials. The practice trials were used to familiarize the participants with the task, and the data collected from them were not included in the final analysis. In the

experimental session, the 100 stimuli were repeated 5 times in 5 blocks. After completing each test block and at the end of the practice section, participants received a message telling them to take a short break, and to press a button when they were ready to continue. Testing time was approximately 30-40 minutes per participant.

After the experiment, each listener also completed a simple questionnaire about demographic information and language background (age, gender, native language).

**4.3 Results**

The targets of data analysis in the present study are the participants' responses to the 100 manipulated utterance tokens in each language in the perception task. We expect participants to provide systematic responses to the manipulated tokens. For example, when the three cues are manipulated in the same direction for a given stimulus, e.g. longer in pause duration and rime duration, and having an original 3-item pitch contour, then a participant should be especially likely to perceive it as containing 3 items. When the three cues provide conflicting boundary information, then the participants' judgments may vary according to the importance they assign to each cue or cue combination in boundary perception. In analyzing participants' responses to different tokens with different combinations of cue configurations, of particular interest is to 1) explore the weight assigned to each cue in the perception of prosodic boundaries; 2) determine the difference in cue weighting between the two languages by comparing the difference between English and Chinese participants' responses. One thing should be noted is that different stimuli are used in the two languages. Although special care has been taken to ensure that test words used in both languages have the same syllable structure and stress

pattern, segmental structures of words are different. The F0 change between the boundary and no-boundary condition is similar (about 40 Hz) in the two languages, but the F0 manipulation is not the same. F0 change is in pitch reset in Chinese, whereas in English, it is the contour change.

In the following sections, the results are first presented as mean percentage of boundary categorization as a function of pitch (Fig. 4.1), pause (Fig. 4.2), and pre-boundary rime duration (Fig. 4.3). Then, the relation between boundary categorization and the three main variables and comparison between the two languages are presented using the results of the logistic regression analyses and relative weight analyses.

### 4.3.1 Descriptive statistics

A total of 20,000 responses from the perception test (40 subjects * 100 tokens * 5 replications) were subjected to analysis. Overall, 39.6% of the responses favored 2-item, 59.8% favored 3-item, and 114 responses (0.057%) were missing as no responses were received before the time-out period (after 3000 ms).

For the English data, 37.26% of the responses favored 2-item, 61.8% favored 3-item, and 94 responses (0.094%) were missing.

For the Chinese data, 42.01% of the responses favored 2-item, 57.79% favored 3-item, and 20 responses (0.02%) were missing.

The pattern of responses is summarized in the following figures. Figure 4.3 shows the mean percentage of 2-item identification as a function of pitch categories. The X-axis represents the two pitch levels.  In English, they refer to a falling F0 contour in the no-

boundary condition and a rising contour in the boundary condition, and in Chinese, they refer to the absence of pitch reset (represented as continuous pitch declination) in the no-boundary condition and pitch reset in the boundary condition. The Y-axis is the mean percentage of 2-item responses.

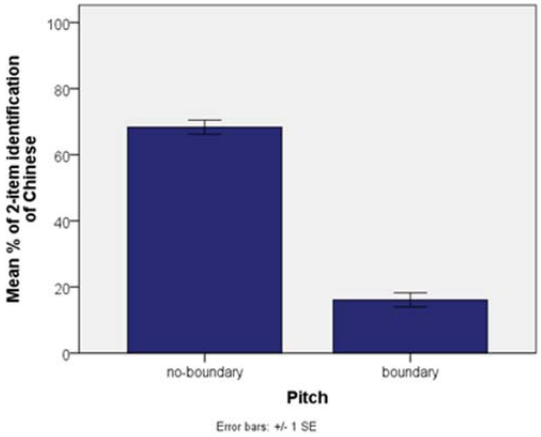



Figure 4.3a Mean Percentage of 2-item identification as a function of pitch categories of Chinese listeners. Error bars indicate 1 standard error of means.

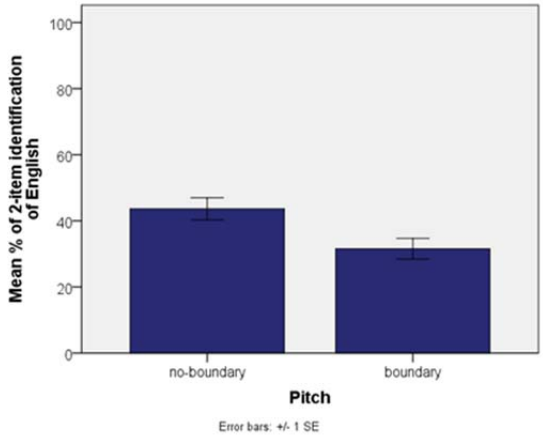



Figure 4.3b Mean Percentage of 2-item identification as a function of pitch categories of English listeners. Error bars indicate 1 standard error of means.

Figures 4.3a and 4.3b show that pitch, as manipulated here, had a greater effect on Chinese than on English listeners' identification of prosodic boundaries. For Chinese listeners, the no-boundary pitch pattern (absence of pitch reset) resulted in 68% of 2-item identifications, and the boundary pitch pattern (presence of pitch reset) resulted in 84% of 3-item identifications. However, English listeners' responses showed a small shift in response to the change in pitch contour, with 43% of no-boundary pitch stimuli (represented by a falling contour) being identified as containing 2 items, and 68% of boundary pitch stimuli (represented by a rising contour) resulting in 3-item identification.

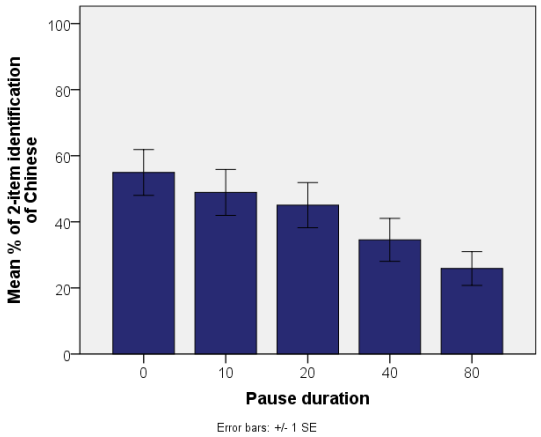



Figure 4.4a Mean Percentage of 2-item identification as a function of pause duration of Chinese listeners. Error bars indicate 1 standard error of means.

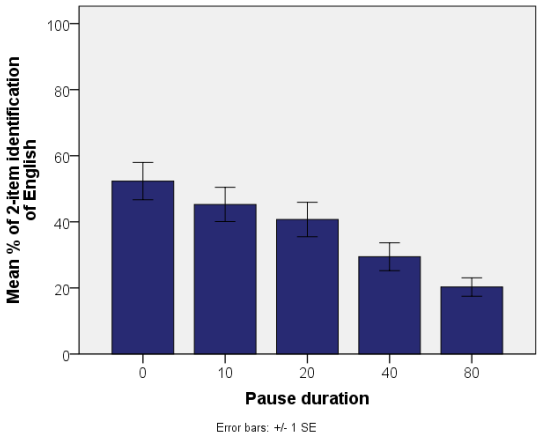



Figure 4.4b Mean Percentage of 2-item identification as a function of pause duration of English listeners. Error bars indicate 1 standard error of means.

Figures 4.4a and 4.4b show the mean percentage of 2-item identification as a function of pause manipulations. The X-axis represents the 5 pause steps, and the Y-axis is the mean percentage of 2-item responses. For both Chinese and English, an increase in pause duration caused a decrease of 2-item identification, and a similar magnitude of the change could be observed in the two languages. For Chinese, the change of pause duration from 0 ms to 80 ms caused a decrease of 2-item identification from 55% to 26%, and same change of duration resulted in a decrease from 52% to 20% in English.

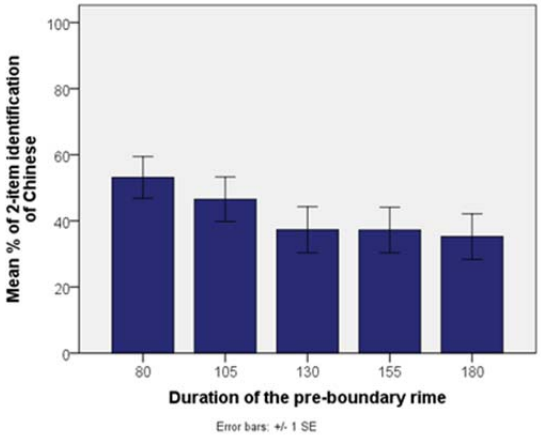



Figure 4.5a Mean Percentage of 2-item identification as a function of the duration of pre-boundary rimes of Chinese listeners. Error bars indicate 1 standard error of means.

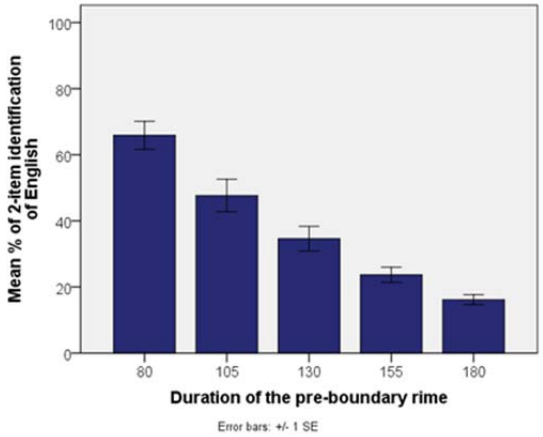



Figure 4.5b Mean Percentage of 2-item identification as a function of the duration of pre-boundary rimes of English listeners. Error bars indicate 1 standard error of means.

Figures 4.5a and 4.5b shows the mean percentage of 2-item identification as a function of the duration of pre-boundary rimes. The X-axis represents the five steps of pre-boundary syllable duration, and the Y-axis is the mean percentage of 2-item responses. Similar to the effect of pause duration, an increase in the duration of the pre-boundary rime led to a decrease of 2-item identification in both Chinese and English listeners. The two languages differ, however, in the pattern and magnitude of the effect. For English, each decrement in duration influenced identification, and the average identification difference between the shortest and longest rime duration was 49%. This identification difference was only 18% in Chinese. Moreover, increasing rime duration beyond 130 ms in this task had almost no effect on 2-item versus 3-item identification.

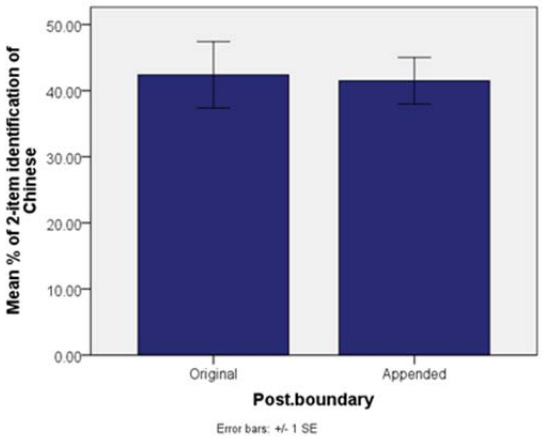



Figure 4.6a Mean Percentage of 2-item identification as a function of post-boundary categories of Chinese listeners. Error bars indicate 1 standard error of means.

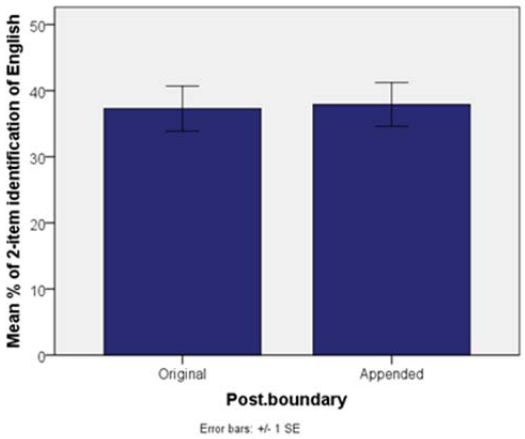



Figure 4.6b Mean Percentage of 2-item identification as a function of post-boundary categories of English listeners. Error bars indicate 1 standard error of means.

Figures 4.6a and 4.6b show that post-boundary factor did not have an effect on the boundary identification in both languages. For English listeners, the original post-boundary stimuli resulted in 37.3% of 2-item identifications, and the appended post-boundary stimuli resulted in 37.9% of 2-item identifications. Similarly, Chinese listeners' responses showed little difference in the identification rate, with 42.4% of the original post-boundary stimuli being identified as containing 2 items, and 41.5% of appended post-boundary stimuli resulting in 2-item identification.

To better illustrate the different effect of pitch in the two languages, the following four figures present 2-item classifications as a function of pause duration (Figures 4.7 and 4.9) and pre-boundary rime duration (Figures 4.8 and 4.10) based on the distinction between languages and pitch patterns.

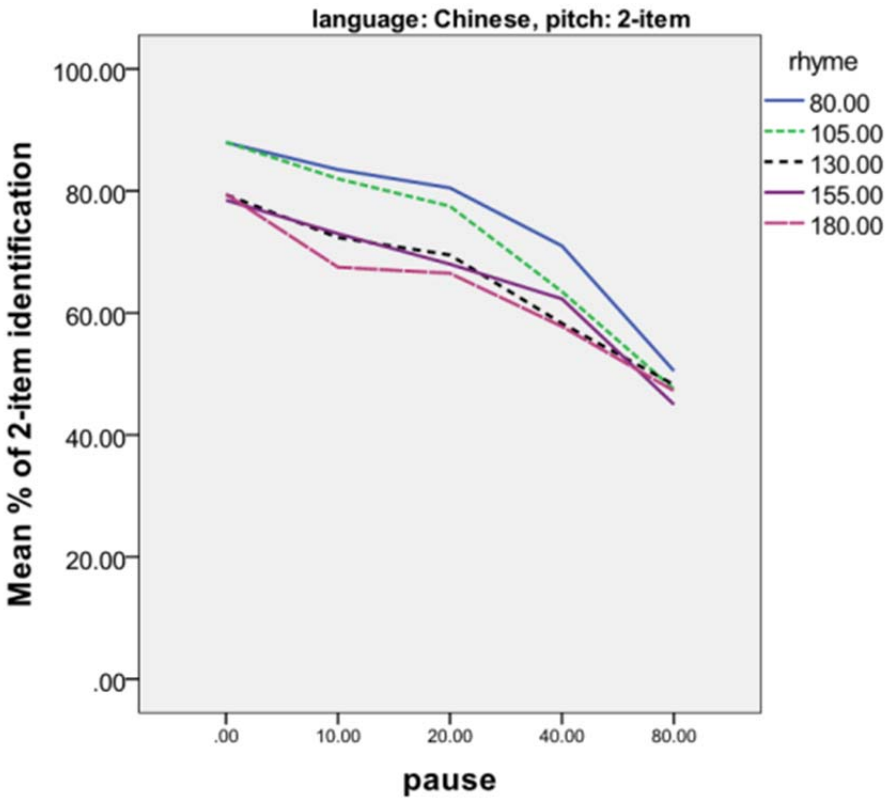



Figure 4.7 Classification of prosodic boundary according to pause and rime for the no-boundary pitch condition in Chinese

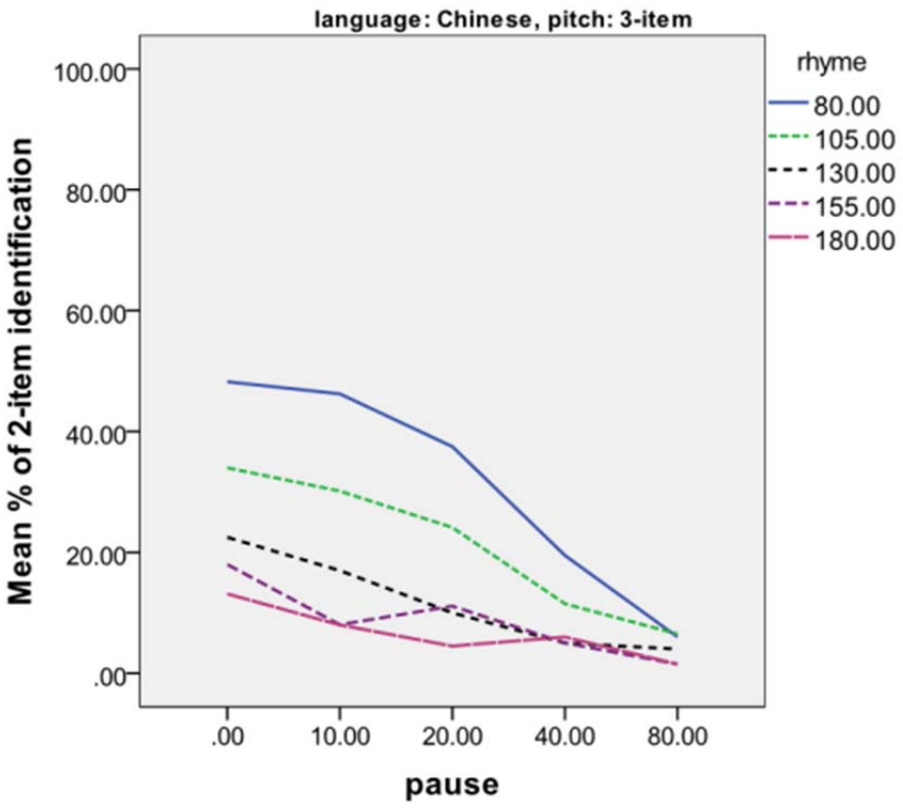



Figure 4.8 Classification of prosodic boundary according to pause and rime for the boundary pitch condition in Chinese

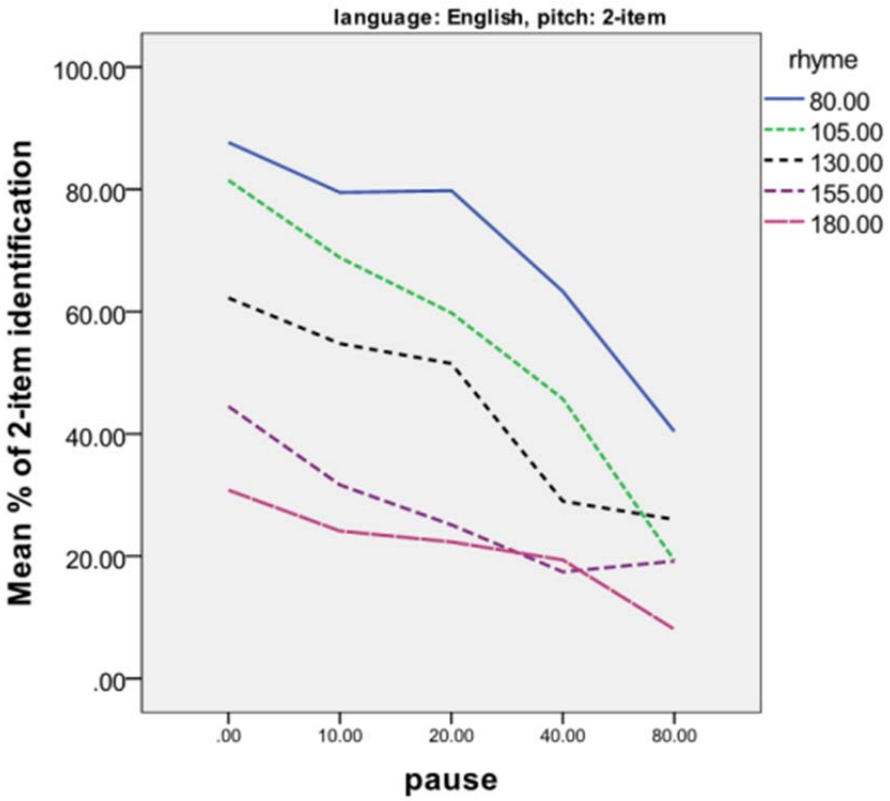



Figure 4.9 Classification of prosodic according to pause and rime for the no-boundary pitch condition in English

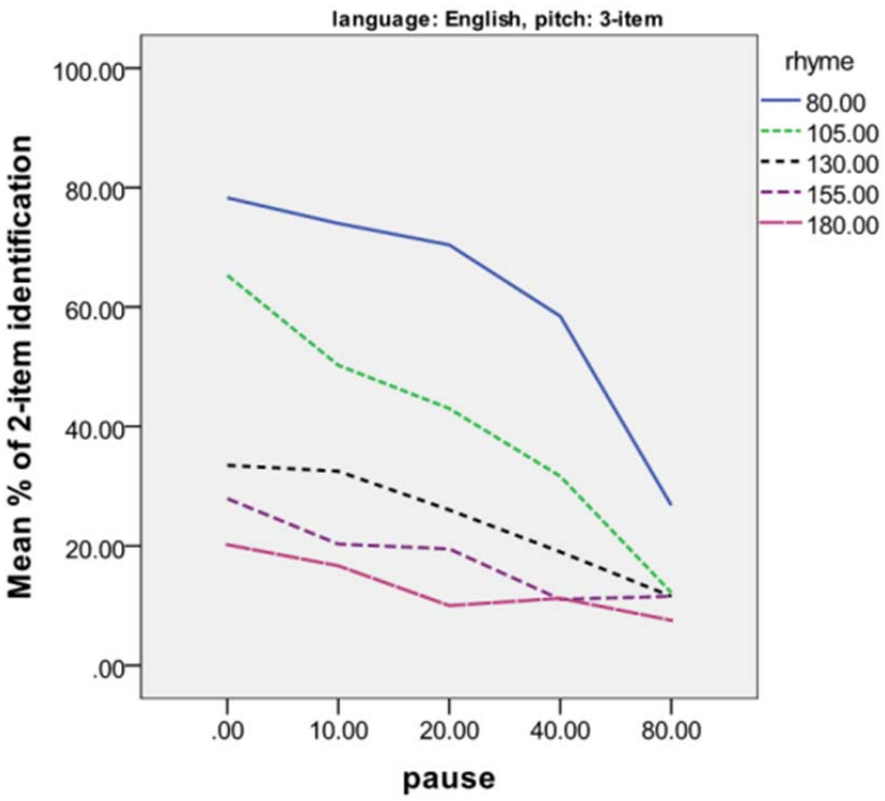



Figure 4.10 Classification of prosodic boundary according to pause and rime for the boundary pitch condition in English

Figures 4.7 and 4.8 clearly show the effect of pitch on boundary identification in Chinese. As was seen earlier, 2-item pitch pattern favored 2-item identification and 3-item pitch pattern favored 3-item identification, resulting in semi-categorical differentiation across the pause and pre-boundary lengthening manipulations. Identification functions cluster above 50% 2-item identification under the no-boundary pitch condition, and below 50% 2-item identification in the boundary pitch condition. In contrast, the range of 2-item identification responses across the duration and pre-boundary lengthening manipulation are similar under the two pitch conditions in English, with both occupying the same region on the chart.

The figures also show the interaction between pause duration and pre-boundary lengthening. For Chinese listeners, the rime duration of 80, 105 and 130 (in ms) generally received the same identification rate across five levels of pause duration under the 2-item pitch condition (Figure 4.7). Under the 3-item pitch condition (Figure 4.8), the effect of pre-boundary lengthening was strongest at the 0 pause duration, and the magnitude of the effect gradually reduced with increasing pause duration. The five rime duration levels nearly converged at the pause duration of 80 ms both in 2-item and 3-item conditions, indicating that the effect of pre-boundary lengthening can be overridden when the pause is sufficiently long (80 ms for the Chinese case).

For English listeners' responses, the effect of duration and pre-boundary lengthening appeared similar under the two pitch conditions, with longer pause and longer pre-boundary rime causing more 3-item classifications. Unlike Chinese listeners' identifications, the effect of pre-boundary lengthening could be observed across five levels of pause duration, although this effect appears to be the smallest at the pause duration of 80 ms, with all the rime duration except the 80 ms one receiving similar identification rates.

### 4.3.2 Mixed-effects Logistic Regression analyses

The observed patterns were further examined in a mixed-effects logistic regression model to determine whether the three acoustic manipulations are predictive of the boundary categorization in each language, and whether there are differences in the relative importance of the three properties in each language. A mixed-effects model was used because subjects were sampled randomly from a large population and therefore

should be considered as a random effect, which enables us to control for the variability introduced by individual subjects. As discussed in Chapter 3, a logistic regression model is suitable for predicting a binary outcome (in this case, whether there is a prosodic boundary or not) from a series of predictors.

The main question being asked in this experiment is whether listeners from different language backgrounds are differentially sensitive to the three acoustic correlates of prosodic boundaries. For this purpose, three logistic regression models were built in SPSS, one for the Chinese listeners, one for the English listeners, and another one incorporating the two language groups collectively. The three predictors: pre-boundary lengthening, pause duration, and pitch pattern (pitch reset for Chinese and pitch contour change for English) were entered into the language-specific model to examine whether a statistically reliable model could be built for each language and if each predictor made a unique and significant contribution to the model. An overall model incorporating the two languages was then built to specifically test the interaction between the language and the three acoustic correlates, which enables us to investigate whether the odds ratio for a particular cue was different across the two groups of listeners. Finally, a relative weight analysis was conducted as a supplement to the logistic regression analysis to investigate the relative importance of each cue within each language.

For all the logistic regression models, the initial fixed effects were Language (English and Chinese, only for the model that included both languages), Pitch (boundary vs. no-boundary), Pause (5 levels), Pre-boundary lengthening (5 levels), Repetition (5 levels), and Post-boundary elements (original vs. appended). For the model incorporating both languages, as we are interested in the influence of language on listeners' use of the

three major acoustic properties, the following three interaction terms were also included in the model as fixed effects: Duration * Language, Pitch * Language, and Pre-boundary lengthening * Language. The effect terms that were found not to be significant were removed from the model and the models were then refitted using the remaining terms. The results reported are based on the reduced models. The details of the final model for each analysis will be reported below.

In all the analyses, all responses provided by the group participants were submitted for analysis. As mentioned earlier, in the Chinese group, 9,980 out of 10,000 possible responses were used and, in the English group, 9,906 out of out of 10,000 possible responses were used in the construction of the model. The model incorporating the two languages therefore included 19,886 responses.

*4.3.2.1 Logistic Regression Analysis for Chinese*

The final model for the logistic regression analysis of speech perception in Chinese includes the following three predictors: Duration of the pre-boundary rime (5 levels), Pause duration (5 levels), and Pitch (2 levels). The Post-boundary factor was removed from the analysis because it did not contribute significantly to the prediction of the outcome variable ($p > .05$). Repetition was shown to have a significant effect with $p = .025$, and the percentage variance (79.6%) explained by the model including the repetition term was 0.1% higher than the one without it. We then decided to remove it from the final model.

Table 4.4 describes the results of the logistic regression analysis with the dependent variables being the absence of a prosodic boundary, coded as 0, and the

presence of a boundary, coded as 1. The reference category was set at 1, thus the model predicted the probability of the absence of a prosodic boundary.

Table 4.4 Summary of Logistic Regression Analysis for Variables Predicting the Identification of a Prosodic Boundary in Chinese (reference category:1). The bold values are significant at the p<.05 level.

| **Parameter** | **Coefficient** | **Std.Error** | **t** | **Sig.** | **Exp(β)** |
|---|---|---|---|---|---|
| Intercept | -3.595 | 0.195 | -18.468 | **<.001** | 0.027 |
| Pre-boun Dur=80 | 1.208 | 0.084 | 14.342 | **<.001** | 3.348 |
| Pre-boun Dur=105 | 0.769 | 0.083 | 9.217 | **<.001** | 2.157 |
| Pre-boun Dur=130 | 0.241 | 0.083 | 2.884 | **.004** | 1.272 |
| Pre-boun Dur=155 | 0.131 | 0.084 | 1.569 | .117 | 1.140 |
| Pre-boun Dur=180 | 0 | | | | |
| Pause=0 | 1.983 | 0.088 | 22.563 | **<.001** | 7.263 |
| Pause=10 | 1.577 | 0.086 | 18.260 | **<.001** | 4.839 |
| Pause=20 | 1.321 | 0.086 | 15.428 | **<.001** | 3.746 |
| Pause=40 | 0.720 | 0.085 | 8.478 | **<.001** | 2.054 |
| Pause=80 | 0 | | | | |
| Pitch=2 | 2.918 | 0.059 | 49.655 | **<.001** | 18.504 |
| Pitch=3 | 0 | | | | |

Note: The parameters with a coefficient value of 0 are the default reference levels.

The results show that all the three acoustic cues, pre-boundary lengthening, pause duration and pitch pattern, are significant predictors of prosodic boundary identification for Chinese listeners. The positive coefficients of pre-boundary lengthening and pause duration indicate that the shorter the pause and the pre-boundary syllables are, the more likely the stimuli are to be perceived as not having a prosodic boundary. This is consistent with the production pattern. For example, the odds (represented by Exp(β) in the figure) of a no-boundary percept are three times greater for a pre-boundary rime duration of 80 ms compared to a pre-boundary rime duration of 180 ms. There are no

model differences between the pre-boundary rime of 155 ms and 180 ms, indicating that a 155 ms pre-boundary rime duration suffices for the identification of a prosodic boundary. The odds of a no-boundary percept are 7.3 times greater for a 0 ms duration compared to a duration of 80 ms. The large perceptual difference elicited the two pitch patterns (Figures 4.6 and 4.7) was also manifested by the coefficients and odds ratio of the pitch variable. The odds of a no-boundary percept are 19 times greater when it is a no-boundary pitch pattern compared to a boundary pitch pattern.

*4.3.2.2 Logistic Regression Analysis for English*

The final model for the logistic regression analysis of identification by English listeners includes three predictors: duration of the pre-boundary rime (5 levels), pause duration (5 levels), and pitch (2 levels). The Post-boundary and repetition factors were removed from the analysis because they did not contribute significantly to the prediction of the outcome variable ($p > .05$).

Table 4.5 describes the results of the logistic regression analysis with the dependent variables being the absence of a prosodic boundary, coded as 0, and the presence of a boundary, coded as 1. The reference category was set at 1.

Table 4.5 Summary of Logistic Regression Analysis for Variables Predicting the Identification of a Prosodic Boundary in English (reference category:1). The shaded values are significant at the p<.01 level.

| Parameter | Coefficient | Std.Error | t | Sig. | Exp(β) |
|---|---|---|---|---|---|
| Intercept | -3.358 | 0.153 | -21.910 | **<.001** | 0.035 |
| Pre-boun Dur=80 | 2.616 | 0.084 | 31.221 | **<.001** | 13.680 |
| Pre-boun Dur=105 | 1.722 | 0.081 | 21.366 | **<.001** | 5.595 |
| Pre-boun Dur=130 | 1.078 | 0.081 | 13.277 | **<.001** | 2.938 |
| Pre-boun Dur=155 | 0.409 | 0.085 | 4.825 | **<.001** | 1.505 |
| Pre-boun Dur=180 | 0 | | | | |
| Pause=0 | 2.027 | 0.083 | 24.336 | **<.001** | 7.592 |
| Pause=10 | 1.620 | 0.083 | 19.623 | **<.001** | 5.053 |
| Pause=20 | 1.389 | 0.083 | 16.801 | **<.001** | 4.010 |
| Pause=40 | 0.828 | 0.084 | 9.862 | **<.001** | 2.288 |
| Pause=80 | 0 | | | | |
| Pitch=2 | 0.764 | 0.048 | 11.995 | **<.001** | 2.147 |
| Pitch=3 | 0 | | | | |

Note: The parameters with a coefficient value of 0 are the default reference levels.

Similar to the results obtained for the Chinese model, all three acoustic cues are shown to be significant predictors of prosodic boundary identification by English listeners. It was also found that the shorter the pause and the pre-boundary rimes are, the more likely the stimuli are to be perceived by listeners as not having a prosodic boundary. However, the magnitude of the importance of the rime duration differed across the two languages. For example, for English, the odds of a no-boundary percept are 14 times greater for a pre-boundary rime duration of 80 ms compared to a pre-boundary rime duration of 180 ms (cf. Chinese where the same duration difference only resulted in an odds ratio of 3 times). The contribution of pause was similar in the two languages. The odds of a no-boundary percept are 7.6 times greater for a 0 ms duration than for an 80 ms duration (cf. 7.3 for Chinese for the same duration difference). Pitch was also found to be

a significant predictor for boundary identification, with the change from a no boundary pitch pattern to a boundary pitch making the percept of a boundary two times more likely. In comparison, the pitch factor (keeping in mind the different pitch manipulation in the two languages) rendered the boundary percept 19 times more likely in Chinese.

*4.3.2.3 Logistic regression analysis of the model incorporating two languages*

The results of the logistic regression analysis for each language showed that the three acoustic properties are significant predictors of the presence versus absence of a prosodic boundary, and the comparison in the coefficients and odds ratio of the predictor variables revealed different contribution of pitch and pre-boundary lengthening in the two languages. In this section, a logistic regression model incorporating data from the two languages was built to specifically investigate the interaction between the three cues and language.

The final model for the analysis included the following fixed main effects: Language (English and Chinese), Duration of the pre-boundary rime (continuous), Pause duration (continuous)[7], Pitch (2 levels), Repetition (continuous), and two interaction terms: Language * Pre-boundary lengthening and Language * Pitch. The Language * Pause interaction term was found not to be significant and was removed from the model. The post-boundary factor was also removed from the model due to its non-significant contribution to the outcome variable.

[7] Pre-boundary lengthening and duration each has five levels. For the convenience of presenting and interpreting the interaction between them and the language, they are considered in this model as scale variables.

Table 4.6 describes the results of the logistic regression analysis with the dependent variables being the absence of a prosodic boundary, coded as 0, and the presence of a boundary, coded as 1. The reference category was set at 1.

Table 4.6 Summary of Logistic Regression Analysis for Variables Predicting the Identification of a Prosodic Boundary. The bold values are significant at the p<.01 level.

| **Parameter** | **Coefficient** | **Std. Error** | **t** | **Sig.** | **Exp(β)** |
|---|---|---|---|---|---|
| Intercept | 0.499 | 0.118 | 4.231 | **<.001** | 1.648 |
| Language=Eng | 2.656 | 0.150 | 17.687 | **<.001** | 14.237 |
| Language=Chi | 0 | | | | |
| Repetition | -0.058 | 0.014 | -4.249 | **<.001** | 0.944 |
| Pre-boun Dur | -0.011 | 0.001 | -13.905 | **<.001** | 0.989 |
| Pause | -0.023 | 0.001 | -30.851 | **<.001** | 0.977 |
| Pitch=2 | 2.699 | 0.060 | 44.946 | **<.001** | 14.872 |
| Pitch=3 | 0 | | | | |
| Duration*[lang=E] | -0.014 | 0.001 | -12.079 | **<.001** | 0.986 |
| Duration*[lang=C] | 0 | | | | |
| [lang=E]*[pitch=2] | -2.051 | 0.080 | -25.750 | **<.001** | 0.129 |
| [lang=C]*[pitch=2] | 0 | | | | |

Note: Parameters with a coefficient value of 0 are the default reference levels.

The results shown in Table 4.6 indicated main effects of pause, pre-boundary lengthening, and pitch, which was consistent to the findings from the individual language models. The negative coefficients for pre-boundary lengthening and pause indicate that an increase in the two variables reduced the probability of the percept of no prosodic boundary. There is also a significant effect of language: when the pitch pattern is at the boundary condition, the odds for English listeners to identify a token as having no-boundary is 14 times higher in relation to Chinese listeners. The interaction term of Language * Pitch indicates that pitch effect changes substantially for Language = English

relative to Language = Chinese, and the change is significant ($p < .001$). The net effect of pitch for Language = English reduced to 0.648 as compared to 2.699 when Language = Chinese. The interaction between Language and Pre-boundary lengthening indicates that change in the lengthening effect when Lang = English relative to Language = Chinese is significant ($p < .001$), and the net effect of Pre-boundary lengthening for Language = English is -0.025 as compared to -0.011 when Language = Chinese.

The effect of the interactions is further displayed in the following three figures, which show the mean probability of a no-boundary percept as predicted by our overall model. Figure 4.11 shows that pitch is modeled as having a substantially smaller effect on boundary perception for English than for Chinese listeners. In the latter case, the probability of a no-boundary percept is 52.3% lower in the boundary pitch condition than in the no-boundary pitch condition.

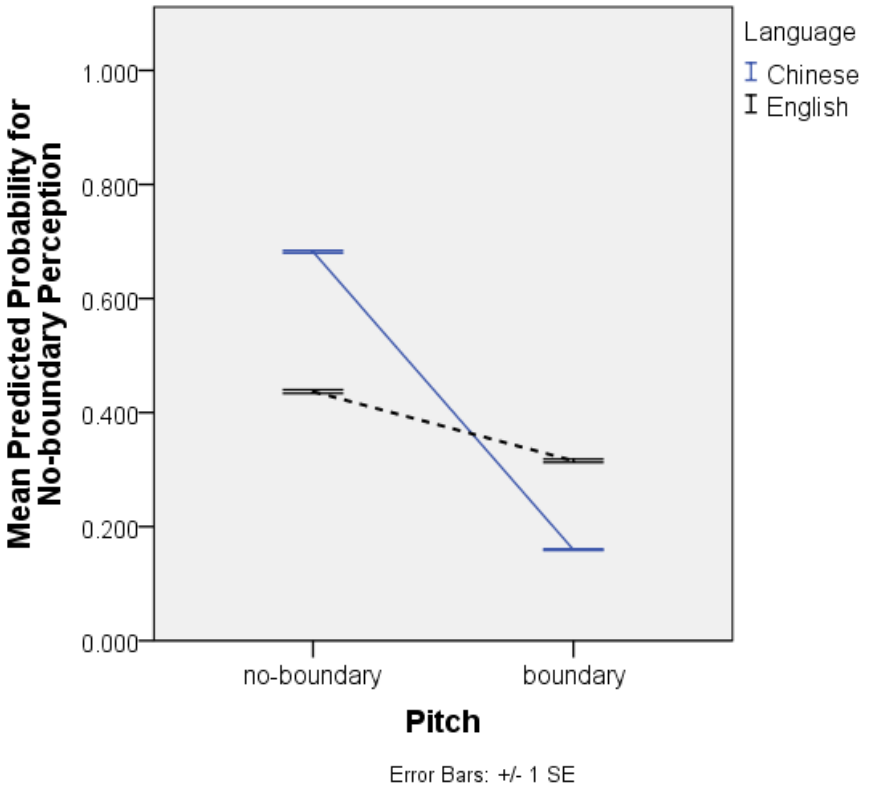



Figure 4.11 Probability of the no-boundary percept as a function of Language and Pitch

The mean probability as predicted by the function of language and pre-boundary lengthening is shown in Figure 4.12. The probability of a no-boundary percept dropped from 63.3% to 14.9% at the two end points of the durational levels in English, whereas in Chinese the probabilities at the same levels were 51.1% and 33.1% respectively.

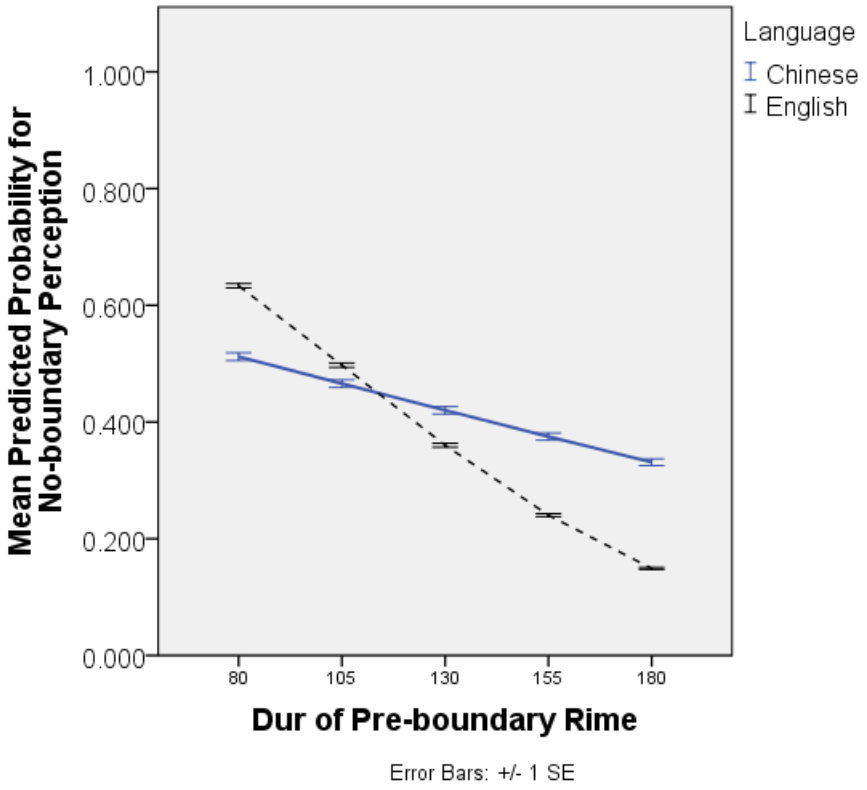



Figure 4.12 Probability of the no-boundary percept as a function of Language and Pre-boundary lengthening

As discussed earlier, the interaction between Language and Pause was found not to be significant, which can be visually observed in Figure 4.13, which shows that probability of a no-boundary percept across the five pause levels in the two languages showed a parallel pattern.

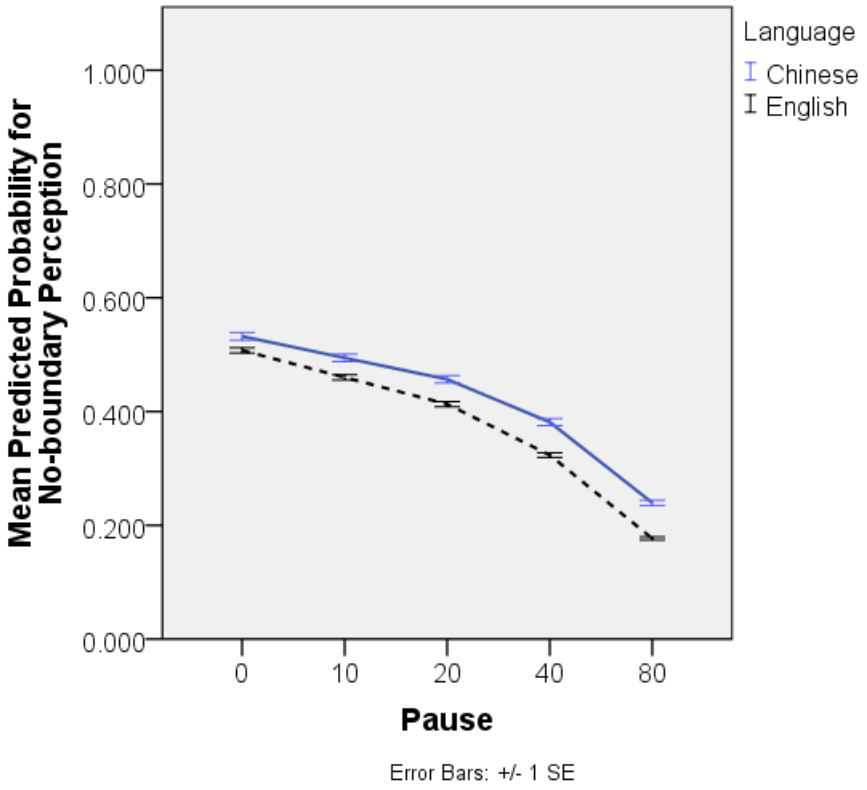



Figure 4.13 Probability of the no-boundary percept as a function of Language and Pause

*4.3.2.4 Relative weight analysis*

The above analyses investigated the difference in the contribution of acoustic cues to the perception of prosodic boundaries for Chinese and English listeners. In this section, a relative weight analysis was performed to determine the importance of the prosodic cues in predicting the presence of a prosodic boundary for both English and Chinese (see Table 4.7 and 4.8).[8] Results showed that, for English listeners, pause had the highest relative weight (.063) and is the most important in predicting the presence of a prosodic boundary (accounting for 78.5% of explained variance). Next was pitch (.015),

[8] Comparison of coefficient values for the relative importance analysis is not appropriate in this study because the three predictors are not distributed on the same scale and unit—there are two levels for the pitch condition and five levels for the effects of pre-boundary lengthening and pause.

accounting for 18.1% of the $R^2$. Pre-boundary lengthening explains little variance (3.4%) in the perception of a prosodic boundary, with a relative weight of 0.003.

Table 4.7 Relative weight analysis of prosodic cue perception in English

| Relative Weights Analysis of prosodic cue perception in English (Criterion = identification of a boundary) | | |
|---|---|---|
| | *Raw relative weights* | *Relative weights as percentage of* $R^2$ |
| Pause | 0.063 | 0.785 |
| Pre-boundary lengthening | 0.003 | 0.034 |
| Pitch | 0.015 | 0.181 |

For Chinese listeners, pitch had the highest relative weight (.264) and is the most important in predicting the presence/absence of a prosodic boundary (accounting for 81.5% of explained variance). Similar to English, pause in Chinese is also relatively more important than pre-boundary lengthening.

Table 4.8 Relative weight analysis of prosodic cue perception in Chinese

| Relative Weights Analysis of prosodic cue perception in Chinese (Criterion = identification of a boundary) | | |
|---|---|---|
| | *Raw relative weights* | *Relative weights as percentage of* $R^2$ |
| Pause | 0.059 | 0.181 |
| Pre-boundary lengthening | 0.001 | 0.003 |
| Pitch | 0.264 | 0.815 |

## 4.4 Summary

The results of the perception experiments show an expected pattern of difference between English and Chinese listeners' use of the three main acoustic cues in the perception of prosodic boundaries. The statistical analyses showed that both English listeners and Chinese listeners use pause, pre-boundary lengthening, and pitch change in

perceiving prosodic boundaries in their native language. However, the two groups of listeners weight these cues differently, with English listeners paying more attention to pause than pitch information in their perception, while Chinese listeners weight pitch (pitch reset) more heavily than pause. Listeners of both languages assign the least to pre-boundary lengthening. As pitch manipulation differed in the two languages, the conclusion that can be drawn here is that the importance of the specific type of pitch change (pitch contour change for English and pitch reset for Chinese) relative to the durational cues in the two languages is different.

The cue-weighting difference for speakers of the two languages is further exemplified in Figures 4.14 and 4.15, which show the probability of the boundary percept based on the three acoustic cues in each language.

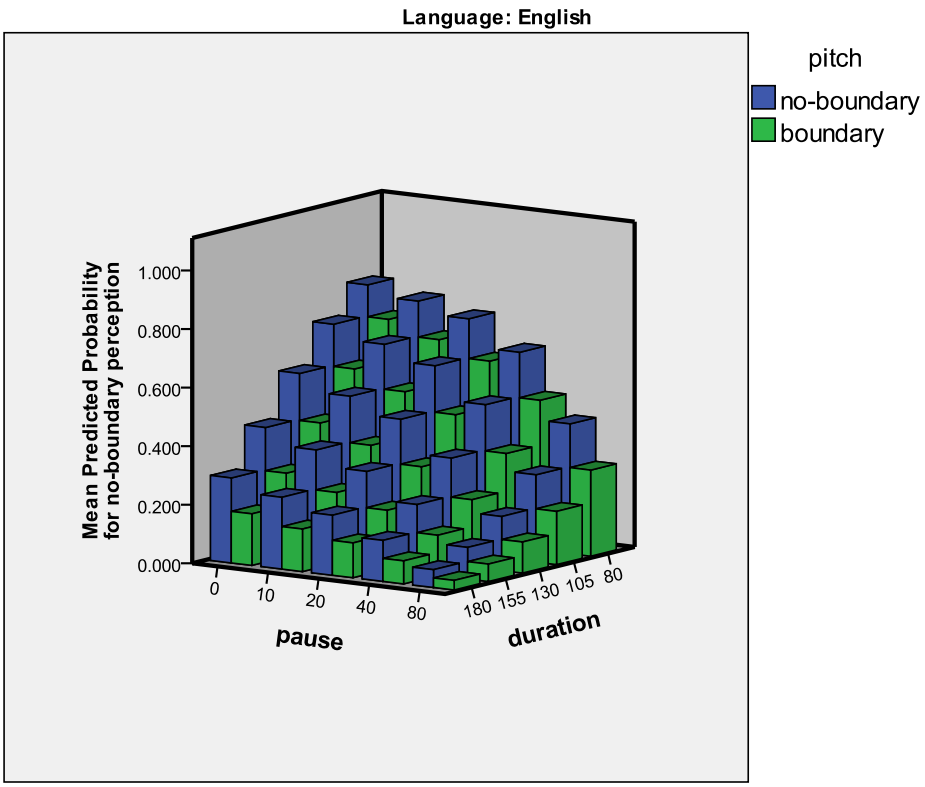



Figure 4.14 Probability of the no-boundary percept as a function of pause, duration and pitch in English

Figure 4.14 shows that for English listeners, the probability of the no-boundary percept is negatively correlated with pause and pre-boundary lengthening: the probability of no-boundary identification decreases with increasing pause duration and lengthening of the pre-boundary rime. The effect of pitch contour change can also be seen; tokens with same pause and pre-boundary duration values consistently received a higher probability of no-boundary percepts when the pitch is in the no-boundary condition.

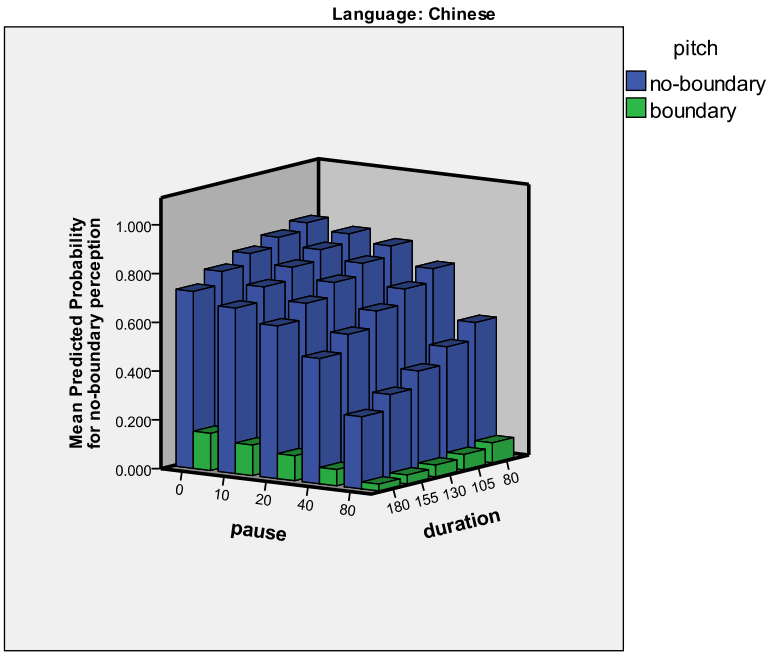



Figure 4.15a Probability of the no-boundary percept as a function of pause, duration and pitch in Chinese

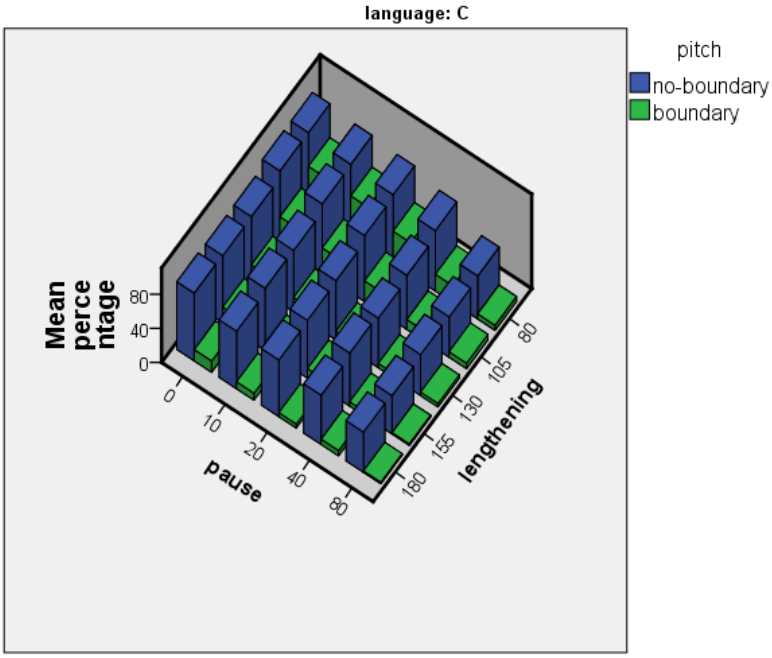



Figure 4.15b Probability of the no-boundary percept as a function of pause, duration and pitch in Chinese (rotated)

Figures 4.15a and 4.15b for Chinese show the predominant role played by pitch reset in predicting the probability of the no-boundary percept. Tokens with same pause duration and pre-boundary rime duration are perceived different under the two pitch conditions. The effect of the two durational cues, pause and pre-boundary lengthening can also be observed. The probability rate of the no-boundary percept increases with decreasing duration of pause and pre-boundary rime. However, the magnitude of the change is smaller than was observed for the English.

The different weighting of the three cues by English- and Chinese-speaking listeners was further confirmed in a relative weight analysis. The results showed that Chinese listeners considered pitch reset to be the most important cue in predicting the presence/absence of a prosodic boundary, while English listeners relied more heavily on pause duration.

# CHAPTER V

# CONCLUSION AND DISCUSSION

## 5.1 Acoustic correlates of prosodic phrase boundaries

This study investigated which acoustic cues are used in the production and perception of prosodic boundaries in English and Chinese, and whether the relative importance of these cues differs between the two languages. Of particular interest was, whether pitch information is weighted differently by native listeners of a tone language in which pitch information signals lexical contrasts in the phonology, compared to native listeners of a non-tonal language in which pitch only signals contrast at the postlexical level (Braun & Johnson, 2011).

It was shown that pause, pre-boundary lengthening and pitch change all significantly influence the production and perception of prosodic boundaries in both English and Chinese, but English and Chinese speakers use different pitch cues and weigh these cues differently in their native language. In agreement with previous studies (Klatt, 1975; Swerts et al., 1994), English listeners relied predominantly on pause rather than on pitch and pre-boundary lengthening, although they still used all three to some degree. However, the finding that Chinese listeners relied predominantly on pitch as a cue to boundary perception rather than on pause and pre-boundary lengthening was not consistent with some of the previous findings reported in the literature. Some earlier

studies distinguished the processing of pitch at the word level and the sentence level, showing that listeners who spoke a tonal language were more sensitive to lexical tones and less sensitive to F0 information at the sentence level than were listeners who spoke a non-tonal language (Braun & Johnson, 2011; Liang & Van Heuven, 2007). The result of this study showed, however, that listeners whose native language has lexical tones were also more sensitive to postlexical pitch information than listeners who spoke a non-tonal language. A possible explanation for these conlicting findings is presented in 5.3.

The perceptual salience of pause as information for the presence of a boundary has been found in many previous studies. Our findings for English substantiated this pattern. The result for Chinese is not compatible with previous findings, however. For Chinese-speaking listeners, the pitch information determines the boundary perception to such a large degree that there is very little freedom left to use pause. The effect of pause is largely overridden by the effect of pitch.

### 5.2 Relation between Production and Perception

The results of the relative weights analyses of the acoustic cues in both production and perception of native Chinese speakers show a discrepancy between production and perception. Chinese listeners were more sensitive to pitch reset than pause and pre-boundary lengthening in identifying prosodic boundaries, but Chinese speakers weighted pitch reset the least heavily, and pause the most heavily in their production of a prosodic boundary.

Although this result is unexpected, discrepancies between production and perception have also been found in previous studies. Gottfried and Beddor (1988)

reported a production-perception discrepancy for the French vowel contrast /o/-/ ɔ/. In differentiating [o:]-[ɔ], French listeners were insensitive to duration differences even though these vowels reliably differ in both temporal and spectral properties in production. Idemaru and Holt (2007) reported that Japanese listeners showed great individual variation in weighting absolute duration (stop duration) and relational duration (ratio of stop duration to preceding mora duration) cues in categorizing stop length, with some favoring one cue, some favoring the other, and some favoring both. In production, however, the relational cue was found to be the more reliable one. They thus concluded that a highly reliable cue in speech production may not necessarily be prominent in speech perception.

The results of the current study do not allow us to offer a satisfactory explanation of the discrepancy between production and perception in Chinese because all the cues play a significant role in the production and perception of prosodic boundaries; it is difficult to explain their relative importance in terms of cue reliability. The phonemic status of pitch cues in Chinese phonology can explain why pitch is weighted the most heavily in perception, but cannot explain why its role is reduced in production. However, it is tempting to speculate that this discrepancy might result from experimental design. In production, the distinction in pause duration under the two conditions ranges from zero ms in the no-boundary condition to over 300 ms in the boundary condition. However, in the perception task, the maximum pause duration was set at 80 ms. This setting quite possibly reduced the contribution of pause, making pitch reset a more prominent predictor for Chinese listeners. In contrast, for English listeners, pause remained the most important predictor in prosodic boundary percept.

### 5.3 Pitch Reset and Pitch Slope

Due to the specific type of utterances used in the study, two different types of pitch cues were employed by native speakers of Chinese and English in the production and perception of prosodic boundaries: pitch reset by Chinese speakers and pitch contour change by English speakers. The findings that Chinese listeners relied more on pitch as a cue to perceiving prosodic boundaries than did English listeners could mean, that pitch cues (no matter what type) are weighted more heavily by Chinese listeners than English listeners due to the phonemic status of pitch in the language's phonological system. Alternatively, the results might simply indicate that pitch reset and pitch slope are weighted differently.

The two types of pitch movement clearly have different consequences in Chinese. Pitch slope, realized by pitch contour change, has the same dimension as lexical tones. Presumably, it cannot change freely because the change could possibly result in a different word. Lexical tone interference reduces the sensitivity to pitch cues at the sentence level. Pitch reset, on the other hand, utilizes a different phonetic dimension (pitch height) that is used for lexical tones, thus having more freedom to show its effects.

This distinction helps explain some apparent contradictions in the literature. Previous studies distinguished pitch processing at the word and sentence level (Braun & Johnson, 2011; Liang & Van Heuven, 2007), but not of pitch reset and pitch slope change. Liang & Van Heuven (2007) proposed that listeners who spoke a tonal language were more sensitive to lexical tones but were less sensitive to F0 information at the sentence level compared to listeners who spoke a non-tonal language because the contrast induced

by pitch was at the word level. The results of the current study seemed contradictory in that listeners of a tonal language were also very sensitive to post-lexical pitch information. This contradiction can be resolved if we differentiate pitch reset and pitch slope. Liang & Van Heuven (2007) examined processing of sentence intonation (statement vs. question) which is realized by pitch contour change. As discussed above, lexical tone interference resulted in insensitivity to the pitch contour change at the post-lexical level. Pitch information was realized as pitch reset in the current study, which is free of lexical tone interference. Interference in the case of pitch contour change but not pitch reset may be responsible for the different outcomes. In comparison, both pitch reset and pitch slope change can be used freely in English, it is therefore speculated that their weighting in perception is similar. A direct comparison between pitch reset and pitch slope change is not possible as they are used in different contexts.

### 5.4 Contributions and Limitations

A contribution of this research is that it is the first study using systematically manipulated stimuli to examine cue-weighting differences for speakers of a tonal language and speakers of a non-tonal language. Previous studies of cue weighting in the perception of prosodic boundaries either manipulated durational cues only, or investigated three cues in one language. The comparison of the cue weights in boundary perception for speakers of tonal and non-tonal languages helps us to gain a better understanding of the use of pitch information used at the sentence level, about which there are contradictory findings in previous studies.

One of the limitations of the study is that different test tokens were used for the two groups of listeners. Although special care was taken to make the two sets of tokens as parallel as possible (yet still be compatible with the production data), the different results obtained from two groups of listeners could possibly be due to the difference in the test tokens themselves. This problem, however, could be addressed by extending the research to a cross language study on L2 perception or using synthesized stimuli comprised of same sets of nonsense words for the two languages under investigation. For example, Braun and Johnson (2011) used same sets of CVCV nonwords produced with different pitch fall and pitch rise on the first or second syllable to resemble both Dutch intonation and Chinese tones, in the comparison of pitch processing between Dutch and Mandarin listeners.

### 5.5 Future studies

This study investigated cue-weighting in the production and perception by native speakers and listeners of English and Chinese. It is desirable to expand the study to L2 learners so that same test tokens can be used for speakers of different languages. It would be especially interesting to examine the sensitivity to pitch cues of L2 learners whose native language is tonal (e.g. Thai) and learners whose native language is non-tonal (e.g. English). In this way, we could examine the influence of language experience on cue-weighting using the same set of data, thus increasing the validity of the study. It is expected that L2 learners whose L1 is tonal will pay more attention to pitch cues, while learners whose L1 is non-tonal will rely more on pause.

# APPENDICES

A. Experiment 1: English Production lists

| | **No boundary** | **With boundary** |
|---|---|---|
| test utterances | 1. Bacon-salad and wine | 1. Bacon, salad, and wine |
| | 2. Chicken-salad and juice | 2. Chicken, salad, and juice |
| | 3.Ginger-salad and tea | 3.Ginger, salad, and tea |
| | 4.Kiwi-salad and yogurt | 4.Kiwi, salad, and yogurt |
| | 5.Melon-salad and milk | 5.Melon, salad, and milk |
| | 6.Pasta-salad and coffee | 6.Pasta, salad, and coffee |
| | 7.Pepper-salad and juice | 7.Pepper, salad, and juice |
| | 8.Salmon-salad and wine | 8.Salmon, salad, and wine |
| | 9.Tuna-salad and wine | 9.Tuna, salad, and wine |
| | 10.Turkey-salad and coffee | 10.Turkey, salad, and coffee |
| Fillers | 1. Grape jam and bread | 1. Grapes, jam, and bread |
| | 2. Peanut butter and bread | 2. Peanuts, butter, and bread |
| | 3.Cheese cake and coffee | 3.Cheese, cakes, and coffee |
| | 4. Lemon pudding and tea | 4. Lemons, pudding, and tea |
| | 5. Peach yogurt and cookies | 5. Peaches, yogurt, and cookies |
| | 6. Orange juice and sandwiches | 6. Oranges, juice, and sandwiches |
| | 7. Almond cookies and milk | 7. Almond,, cookies, and milk |
| | 8. Beef sandwiches and coke | 8. Beef, sandwiches, and coke |
| | 9. Cherry pie and coffee | 9. Cherries, pie, and coffee |
| | 10. Chocolate ice-cream and honey | 10. Chocolate, ice-cream, and honey |

B. Experiment 1: Chinese Production lists

| No boundary | Gloss | With boundary | Gloss |
|---|---|---|---|
| **Test utterances** | | | |
| 1. hetao shala he hongjiu | 1. Walnut-salad and red wine | 1. hetao, shala he hongjiu | 1. Walnut, salad, and red wine |
| 2. huanggua shala he chengzhi | 2. Cucumber-salad and orange juice | 2. huanggua, shala he chengzhi | 2. Cucumber, salad, and orange juice |
| 3. juzi shala he suannai | 3.Orange-salad and yogurt | 3. juzi, shala he suannai | 3.Orange, salad, and yogurt |
| 4. mogu shala he hongjiu | 4.Mushroom-salad and red wine | 4. mogu, shala he hongjiu | 4.Mushroom, salad, and red wine |
| 5. putao shala he niunai | 5.Grape-salad and milk | 5. putao, shala he niunai | 5.Grape, salad and milk |
| 6. qiezi shala he cha | 6.Eggplant-salad and tea | 6. qiezi, shala he cha | 6.Eggplant, salad, and tea |
| 7. shiliu shala he suannai | 7. Pomegranate-salad and yogurt | 7. shiliu, shala he suannai | 7. Pomegranate, salad, and yogurt |
| 8. xigua shala he niunai | 8.Watermelon-salad and milk | 8. xigua, shala he niunai | 8.Watermelon, salad, and milk |
| 9. yezi shala he kafei | 9.Coconut-salad and coffee | 9. yezi, shala he kafei | 9.Coconut, salad, and coffee |
| 10. yingtao shala he hongjiu | 10.Cherry-salad and red wine | 10. yingtao, shala he hongjiu | 10.Cherry, salad, and red wine |
| **Fillers** | | | |
| 1. mangguo dangao he kafei | 1. Mango cake and coffee | 1. mangguo, dangao he kafei | 1. Mango, cake, and coffee |
| 2. huasheng jiang he mianbao | 2. Peanut butter and bread | 2. huasheng, guojiang he mianbao | 2. Peanut, butter, and bread |
| 3. Pipa guantou he mianbao | 3. canned Pipa and bread | 3. Pipa, guantou he mianbao | 3. Pipa, canned food, and bread |
| 4. ningmeng dangao he hongcha | 4. Lemon cakes and black tea | 4. ningmeng, dangao he hongcha | 4. Lemon, cakes, and black tea |
| 5. caomei suannai he binggan | 5. Strawberry yogurt and cookies | 5. caomei, suannai he binggan | 5. Strawberry, yogurt, and cookies |
| 6. huangtao suannai he binggan | 6. Peach yogurt and cookies | 6. huangtao, suannai he binggan | 6. Peach, yogurt, and cookies |
| 7. xingren binggan he niunai | 7. Almond cookies and milk | 7. xingren, binggan he niunai | 7. Almond, cookies, and milk |
| 8. niurou sanmingzhi he kele | 8. Beef sandwich and coke | 8. niurou, sanmingzhi he kele | 8. Beef ,sandwich, and coke |
| 9. yingtao pai he kafei | 9. Cherry pie and coffee | 9. yingtao, pai he kafei | 9. Cherry, pie, and coffee |
| 10.Qiaokeli bingqilin he fengmi | 10.Chocolate ice-cream and honey | 10.Qiaokeli, bingqilin he fengmi | 10.Chocolate, ice-cream, and honey |

C. Picture strips used in the production experiments (arranged in the order in the word lists in appendix 1 and 2)


Tuna


salad

(and)


wine


Turkey


salad

(and)


coffee


Tuna salad

(and)


wine


Turkey salad

(and)


coffee


Grapes


jams

(and)


bread


Peanuts


butter

(and)


bread


Grape jam

(and)


bread


Peanut-butter

(and)


bread


Cheese


cakes

(and)


coffee


Lemons


pudding

(and)


tea


Cheese-cake

(and)


coffee


Lemon pudding

(and)


tea


Peaches


yogurt

(and)


cookies


Oranges


juice

(and)


sandwiches


Peach yogurt

(and)


cookies


Orange juice

(and)


sandwiches


Almonds


cookies (and)


milk


Beef


sandwiches (and)


coke


Almond cookies (and)


milk


Beef sandwiches (and)


coke


Cherries


pie (and)


coffee


Chocolate


Ice-cream (and)


honey


Cherry pie (and)


coffee


Chocolate ice-cream (and)


honey


核桃


沙拉

(和)


红酒


黄瓜


沙拉

(和)


橙汁


核桃沙拉

(和)


红酒


黄瓜沙拉

(和)


橙汁


桔子


沙拉

(和)


酸奶


蘑菇


沙拉

(和)


红酒


桔子沙拉

(和)


酸奶


蘑菇沙拉

(和)


红酒


葡萄


沙拉

(和)


牛奶


茄子


沙拉

(和)


茶


葡萄沙拉

(和)


牛奶


茄子

(和)


茶


石榴


沙拉

(和)


酸奶


西瓜


沙拉

(和)


牛奶


石榴沙拉

(和)


酸奶


西瓜沙拉

(和)


牛奶


椰子


沙拉

(和)


咖啡


椰子沙拉

(和)


咖啡


樱桃


沙拉

(和)


红酒


樱桃沙拉

(和)


红酒


柠檬


蛋糕

(和)


红茶


芒果蛋糕

(和)


红茶


花生


果酱

(和)


面包


花生酱

(和)


面包


枇杷


罐头


面包


枇杷罐头

(和)


面包


柠檬


蛋糕

(和)


红茶


柠檬蛋糕

(和)


红茶


草莓


酸奶

(和)


饼干


草莓酸奶

(和)


饼干


黄桃


酸奶

(和)


饼干


黄桃酸奶

(和)


饼干


杏仁


饼干

(和)


牛奶


牛肉


三明治

(和)


可乐


杏仁饼干

(和)


牛奶


牛肉三明治

(和)


可乐


樱桃


派

(和)


咖啡


巧克力


冰激凌

(和)


蜂蜜


樱桃派

(和)


咖啡


巧克力冰激凌

(和)


蜂蜜

## REFERENCES


Baayen, R. H. (2008). Mixed-effects modeling with crossed random effects for subjects and items. *Journal of Memory and Language*, 59, 390-412.

Baayen, R. H., Davidson, D. J., & Bates, D. M. (2008). Mixed-effects modeling with crossed random effects for subjects and items. *Journal of Memory and Language*, 59, 390-412.

Beach, C. (1991). The interpretation of prosodic patterns at points of syntactic structure ambiguity: Evidence for cue trading relations. *Journal of Memory and Language*, 30(6), 644-663.

Bent, T., Bradlow, A. R., & Wright, B. A. (2006). Influence of linguistic experience on the cognitive processing of pitch in speech and nonspeech sounds. *Journal of Experimental Psychology: Human Perception and Performance*, 32, 97-103.

Berkovits, R. (1993). Utterance-final lengthening and the duration of final-stop closures. *Journal of Phonetics*, 21, 479-489.

Berkovits, R. (1994). Durational effects in final lengthening, gapping, and contrastive stress. *Lang Speech*, 37, 237-250.

Beddor, P. S. (2009). A coarticulatory path to sound change. *Language*, 85(4):785-821.

Braun, B., & Johnson, E. K. (2011). Question or Tone 2? How language experience and linguistic function guide pitch processing. *Journal of Phonetics*, 39, 585-594.

Boersma, M. (2005). Perception of familiar contrasts in unfamiliar positions. *Journal of the Acoustical Society of America*, 117(6), 3890-3901.

Broersma, M. (2010). Perception of final fricative voicing: Native and nonnative listeners' use of vowel duration. *Journal of the Acoustical Society of America*, 127, 1636–1644.

Boersma, P., & Weenink, D. (2010). Praat: doing phonetics by computer [Computer program].

Byrd, D. (2000). Articulatory vowel lengthening and coordination at phrasal junctures. *Phonetica*, 57, 3-16.

Byrd, D., Krivokapic, J., & Lee. S. (2006). How far, how long: On the temporal scope of phrase boundary effects. *Journal of the Acoustical Society of America*, 120, 1589-1599.

Cambier-Langeveld, T. (1997). The domain of final lengthening in the production of Dutch. In J. Coerts & H. de Hoop (Eds.), *Linguistics in the Netherlands* (pp. 13-24). Amsterdam: John Benjamins.

Carlson, R., & Swerts, M. (2003). Perceptually based prediction of upcoming prosodic breaks in spontaneous Swedish speech materials. In *The proceedings of the International Congress of Phonetic Sciences*, Barcelona, Spain.

Cho, T., & Keating, P. (2001). Articulatory strengthening at the onset of prosodic domains in Korean. *Journal of Phonetics*, 28,155-190.

Cho, T., & McQueen, J. M. (2006). Phonological versus phonetic cues in native and non-native listening: Korean and Dutch listeners' perception of Dutch and English Consonants. *Journal of the Acoustical Society of America*, 119(5), 3085-3096.

Cooper, W. E., Paccia, J. M., & Lapointe, S. G. (1978). Hierarchical coding in speech timing. *Cognitive Psychology*, 10, 154-177.

Cooper, W. E., & Paccia-Cooper, J. (1980). *Syntax and speech*. Cambridge, MA: Harvard University Press.

Cooper, W. E., & Sorensen, J. M. (1977). Fundamental frequency contours at syntactic boundaries. *Journal of the Acoustical Society of America*, 62, 683-692.

Cruttenden, A. (1997). *Intonation* (2nd edn). Cambridge: Cambridge University Press.

Crystal, T., & House, A. (1990). Articulation rate and duration of syllables and stress groups in connected speech. *Journal of the Acoustical Society of America*, 88, 349-371.

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

Cutler, A., Dahan, D., & Donselaar, W. van. (1997). Prosody in the comprehension of spoken language: a literature review. *Language and Speech*, 40, 141-201.

Dankovicova, J., Pigott, K., Wells, B., & Peppé, S. (2004) Temporal markers of prosodic boundaries in children's speech production. *Journal of International Phonetic Association*, 34, 17-36.

Duanmu, S. (1996). Pre-juncture lengthening and foot binarity. *Studies in the Linguistic Sciences*, 26.1/2, 95-115.

Duanmu, S. (2007). *The Phonology of Standard Chinese. 2nd Edition.* Oxford University Press.

Escudero, P. (2005). Linguistic Perception and Second Language Acquisition: Explaining the attainment of optimal phonological categorization. (Doctoral dissertation). Utrecht University.

Escudero, P., Benders, T., & Lipski, S.C. (2009). Native, nonnative and L2 perceptual cue weighting for Dutch vowels: The case of Dutch, German, and Spanish listeners. *Journal of Phonetics*, 37, 452-465

Francis, A. L., Ciocca, V., Ma, L., & Fenn, K. (2008). Perceptual learning of Cantonese lexical tones by tone and non-tone language speakers. *Journal of Phonetics*, 36, 268-294.

Fry, D. B. (1955). Duration and intensity as physical correlates of linguistic stress. *Journal of the Acoustical Society of America*, 27, 765-768.

Fry, D. B. (1958). Experiments in the perception of stress. *Language and Speech*, 1, 126-152.

Gandour, J.T. (1983). Tone perception in Far Eastern languages. *Journal of Phonetics*, 11, 149-175.

Gottfried, T. L., & Beddor, P. S. (1988). Perception of temporal and spectral information in French vowels. *Language and Speech*, 31, 57-75.

Greenlee, M. (1980). Learning the phonetic cues to the voiced voiceless distinction: A comparison of child and adult speech perception. *Journal of Child Language*, 7, 459-468

Harnsberger, J. D. (2001). The perception of Malayalam nasal consonants by Marathi, Punjabi, Tamil, Oriya, Bengali, and American English listeners: A multidimensional scaling analysis. *Journal of Phonetics*, 29, 303-327.

Hillenbrand, J.M., Clark, M.J., & Nearey, T.M. (2001). Effects of consonant environment on vowel formant patterns. *Journal of the Acoustical Society of America*, 109 (2), 748-763.

Holt, L. L., & Lotto, A. J. (2006). Cue weighting in auditory categorization: Implications for first and second language acquisition. *Journal of the Acoustical Society of America*, 119, 3059-3071.

Holt, L.L., Lotto, A. J., & Kluender, K. R. (2001). Influence of fundamental frequency on stop-consonant voicing perception: A case of learned covariation or auditory enhancement? *Journal of the Acoustical Society of America*, 109, 764-774.

Horne, M., Strangert, E., & Heldner, M. (1995). Prosodic boundary strength in Swedish: Final lengthening and silent interval duration. In *The proceedings of the International Congress of Phonetic Sciences*, 170-173, Stockholm, Sweden.

Idemaru, K., & Holt, L. L. (2007). Relational timing or absolute duration? Cue weighting in the perception of Japanese singleton vs. geminate stops. *Proceedings of the 16th International Congress of Phonetic Sciences*, Saarbruecken, Germany, 753-756.

Jaeger, T. F. (2008). Categorical data analysis: Away from ANOVAs (transformation or not) and towards logit mixed models. *Journal of Memory and Language*, 59, 434-446.

Johnson, J.W. (2000). A heuristic method for estimating the relative weight of predictor variables in multiple regression. *Multivariate Behavioral Research*, 35, 1-19.

Johnson, J. W., & LeBreton, J. M. (2004). History and use of relative importance indices in organizational research. *Organizational Research Methods*, 7, 238-257.

Jusczyk, P. W. (1993). From general to language-specific capacities: The WRAPSA model of how speech perception develops. *Journal of Phonetics*, 21, 3-28.

Katz, W.F., Beach, C.M., Jenouri, K., & Verma, S. (1996). Duration and fundamental frequency correlates of phrase boundaries in productions by children and adults. *Journal of the Acoustical Society of the America*, *99,* 3179-3191.

Klatt, D. (1975) Vowel Lengthening is Syntactically Determined in a Connected Discourse. *Journal of Phonetics* 3:129-140.

Kluender, K.R., Lotto, A.J., Holt, L.L., & Bloedel, S.B. (1998). Role of experience in language-specific functional mappings for vowel sounds as inferred from human, nonhuman, and computational models. *Journal of the Acoustical Society of America*, 104, 3568-3582.

Krivokapic, J. (2007). Prosodic planning: Effects of phrasal length and complexicity on pause duration. *Journal of Phonetics*, 35(2), 162–179.

Ladd, D. R. (1988). Declination 'reset' and the hierarchical organization of utterances. *Journal of the Acoustical Society of the America*, 84, 530-544.

Ladd, D. R. (1996). *Intonational phonology*. Cambridge: University Press.

Ladd, D. R., & Campbell, W. N. (1991). Theories of prosodic structure: evidence from syllable duration. In *Proceedings of the 12th International Congress of Phonetic Sciences* (vol. 2), 290-293. Aix-en-Provence.

Ladd, D. R., & Cutler, A. (1983). Models and measurements in the study of prosody. In Cutler & D. R. Ladd (Eds.), *Prosody: Models and measurements* (pp.1-10). Berlin: Springer-Verlag.

Lehiste, I. (1970). *Suprasegmentals*. Cambridge, MA: MIT Press.

Lehiste, I. (1973). Phonetic disambiguation of syntactic ambiguity. *Glossa*, 7, 107-122.

Lehiste, I., Olive, J. P., & Streeter, L. A. (1976). Role of duration in disambiguating syntactically ambiguous utterances. *Journal of the Acoustical Society of America*, 60(5), 1199-1202.

Liang, J., & VanHeuven,V. J. (2007). Chinese tone and intonation perceived by L1 and L2 listeners. In C.Gussenhoven, & T.Riad (Eds.), Tones and tunes, Vol.2: Experimental studies in word and utterance prosody (pp. 27-61). Berlin, New York: Moutonde Gruyter.

Lin, H. Y., & Fon, J. (2009). Perception of temporal cues at discourse boundaries. In *The Proceedings of Interspeech*, Brighton, UK.

Lin, M., & J. Yan. (1980) Beijinghua qingsheng de shengxue xingzhi [Phonetics of neutral tone in Beijing mandarin]. *Fangyan* 3: 166-178

Lindblom, B., & Rapp, K. (1973). Some Temporal Regularities of Spoken Swedish. *Papers from the Institute of Linguistics (PILUS)* 21. University of Stockholm, Sweden.

Lisker, L. (1975). Is it VOT or a First formant detector? *Journal of the Acoustical Society of America*, 57, 1547-1551.

Liu, Y., & Li, A., (2003). Cues of prosodic boundaries in Chinese spontaneous speech. In The *Proceedings of 15th ICPHS* (pp. 1269-1272), Barcelona, Spain..

Martin, P. (1982). Phonetic realizations of prosodic contours in French. *Speech Communication*, 1(3,4), 283-294.

McGuire, G. L. (2007). Phonetic Category Learning. (Doctoral dissertation). The Ohio State University.

Morrongiello, B. A., Robson, R. C., Best, C. T., & Clifton, R. K. (1984). Trading relations in the perception of speech by five-year-old children. *Journal of Experimental Child Psychology*, 37, 231-250.

Nittrouer, S. (1992). Age-related differences in perceptual effects of formant transitions within syllables and across syllable boundaries. *Journal of Phonetics*, 20, 351-382.

Nittrouer, S. (2002). Learning to perceive speech: How fricative perception changes, and how it stays the same. *Journal of the Acoustical Society of America*, 112(2), 711-719.

Ohde, R. N., & Haley, K. L. (1997). Stop-consonant and vowel perception in 3- and 4-year-old children. *Journal of the Acoustical Society of America*, 102, 3711-3722.

Pijper, J. R. (1994) On the perceptual strength of prosodic boundaries and its relation to suprasegmental cutes. *Journal of the Acoustical Society of America*, 96 (4), 2037-2047.

Plag, I. (2003) *Word-formantion in English*. Cambridge: Cambridge University Press.

Plag, I., Kunter, G., & Schramm, M. (2011). Acoustic corrlates of primary and secondary stress in North American English. *Journal of Phonetics*, 39 (3), 362-374.

Price, P. J., Ostendorf, M., Shattuck-Hufnagel, S., & Fong, C. (1991) The use of prosody in syntactic disambiguation. *Journal of the Acoustical Society of America*, 90(6), 2956-2970.

Repp, B. H. (1979). Relative amplitude of aspiration noise as a voicing cue for syllable-initial stop consonants. *Language and Speech*, 27, 173-189.

Sanderman, A., & Collier R. (1997). Prosodic phrasing and comprehension. *Language and Speech*, 40(4), 391-409.

Scott, D. R. (1982) Duration as a cue to the perception of a phrase boundary. *Journal of the Acoustical Society of America*, 71(4), 996-1007.

Shen, X. (1992) A pilot study on the relation between the temporal and syntactic structures in Mandarin. *Journal of the International Phonetic Association*, 22: 35-43.

Stevens, K. N. & Klatt, D. H. (1974). Role of formant transitions in the voiced–voiceless distinction for stops. *Journal of the Acoustical Society of America*, 55(3), 653-659.

Strangert, E. (1992). Prosodic cues to the perception of syntactic boundaries. ICSLP 92.

Strangert, E., & Heldner, M. (1995). The labeling of prominence in Swedish by phonetically experienced transcibers. In *The proceedings of 8th ICPhS* (pp.13-19), Stockholm, Sweden.

Streeter, L. A. (1978) Acoustic determinants of phrase boundary representation. *Journal of the Acoustical Society of America*, 64, 1582-1592.

Summerfield, Q. (1981). Articulatory rate and perceptual constancy in phonetic perception. *Journal of Experimental Psychology: Human Perception and Performance*, 7(5), 1074-1095.

Swerts, M. (1997). Prosodic features at discourse boundaries of different strength. *Journal of the Acoustical Society of America*,101(1), 514-521.

Swerts, M., Strangert, E., & Heldner, M. (1996). F0 declination in spontaneous and read-aloud speech. In *Proceedings of the International Conference on Spoken Language Processing,* Philadelphia, October 1996, pp. 1501-1504.

Tonidandel, S., & LeBreton, J. M. (2011). Relative importance analysis – A useful supplement to regression analyses. *Journal of Business and Psychology*, 26, 1-9.

Turk, A. E. (1999). Structural influences on boundary related lengthening in English. In *Proceedings of the 14th International Congress of Phonetic Sciences* (pp. 237-240). San Francisco.

Turk, A. E., & Shattuck-Hufnagel, S. (2007). Phrase-final lengthening in American English. *Journal of Phonetics*, 35(4), 445-472.

Wagner, M., & Watson, D. G. (2010). Experimental and theoretical advances in prosody: A review. *Language and Cognitive Processes*, 25, 7, 905-945.

Walley, A. C., & Carrell, T. D. (1983). Onset spectra and formant transitions in the adult's and child's perception of place of articulation in stop consonants. *Journal of the Acoustical Society of America*, 73, 1011-1022.

Whalen, D. H., & Levitt, A. G. (1995). The universality of intrinsic F0 of vowels. *Journal of Phonetics*, 23, 349-66.

Wightman, C., Shattuck-Hufnagel, S., Ostendorf, M., & Price. P. (1992). Segmental durations in the vicinity of prosodic phrase boundaries. *Journal of the Acoustical Society of America*, 91, 1707-1717.

Xie , L. (2008). Discovering salient prosodic cues and their interactions for automatic story segmentation in Mandarin broadcast news. *Multimedia Systems*, Springer, 14(4), 237-253.

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

Xu, Y. (2005-2011). ProsodyPro.praat. Available from: http://www.phon.ucl.ac.uk/home/yi/ProsodyPro/.

Yang, L. C. (2007). Duration, pause and the temporal structure of mandarin conversational speech. In *The proceedings of the International Congress of Phonetic Sciences*, Saarbrucken, Germany.

Yang, Y. & Wang, B. (2002). Acoustic Correlates of Hierarchical Prosodic Boundary in Mandarin. *Proc. of Speech Prosody 2002*, Aix-en-Provence

Yuan, J. (2006) Mechanisms of question intonation in Mandarin. In Q Huo, B. Ma, E-S. Chng, & H. Li (Eds.), *Chinese Spoken Language Processing* (pp. 19-30). Springer.