Perception of English stress by Mandarin Chinese learners of English: An acoustic study

by

Qian Wang

B.A., Guangdong University of Foreign Studies, China, 1999
M.A., Tsinghua University, China, 2002

A Thesis Submitted in Partial Fulfillment of the
Requirements of the Degree of
DOCTOR OF PHILOSOPHY
in the Department of Linguistics

SUPERVISORY COMMITTEE

Perception of English Stress

by Mandarin Chinese Learners of English:

An Acoustic Study

By

Qian Wang
B.A., Guangdong University of Foreign Studies, China, 1999
M.A., Tsinghua University, China, 2002

Supervisory Committee

Dr Hua Lin (Department of Linguistics)

Supervisor

Dr John Esling (Department of Linguistics)

Departmental Member

Dr Sonya Bird (Department of Linguistics)

Additional Member

Dr Richard King (Department of Pacific and Asian Studies)

Outside Member

ABSTRACT

Supervisory Committee

Dr Hua Lin (Department of Linguistics)
Supervisor

Dr John Esling (Department of Linguistics)
Departmental Member

Dr Sonya Bird (Department of Linguistics)
Additional Member

Dr Richard King (Department of Pacific and Asian Studies)
Outside Member

Second language learners of English often experience difficulties in English lexical stress perception. This has traditionally been attributed to transfer of prosodic unit or settings from their first language (L1). Similarly, the problem of Chinese learners with English stress perception was assumed to arise from tonal transfer. However, little research has been devoted to the investigation of the phonetic details of second language (L2) stress perception. The present research focuses on the perception of English lexical stress by Chinese learners of English. The purpose of this study is to reveal the use of acoustic cues in stress perception by Chinese learners of English.

In the experiment, F0, duration and intensity were manipulated, each with five steps, on three disyllabic nonsense words to result in a total of 375 nonsense tokens. A group of native speakers of English (NE) and a group of Chinese learners of English (CE) participated in the study and judged whether the stress was on the first or second syllable in the test stimuli. The responses of Chinese learners of English in stress judgment were compared against the baseline of native English speakers. The statistical tests of reliance measures and logistic regression models were used in data analysis. Results indicated that, similar to NE participants, performance by CE participants showed systematic variation as a result of the manipulation of the three acoustic cues. However, CEs were different from NEs in their reliance on the three cues. CE had significantly lower duration and intensity reliance scores but significantly higher F0 than NE. In logistic regression analysis, compared to the NE group, F0 contributed most to the CE models, while the contribution of duration and intensity was minimal.

It is concluded from this study that while all three cues have significant effects on stress perception for native English speakers, only F0 has a decisive effect on stress judgments by Chinese learners of English. This study reveals that, rather than transfer of tone at the phonological level, there is transfer of reliance on F0 in the acquisition of L2 English stress. It is suggested that the investigation of phonetic details of learners' problems with L2 stress acquisition is necessary for L2 speech learning theories and also for L2 stress teaching.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

ACKNOWLEDGEMENTS

The completion of this dissertation and this degree would never have happened without the contribution of many people. I would like to take this opportunity to express my sincere gratitude.

First of all, I would like to thank my supervisor, Dr. Hua Lin, for her intellectual and emotional support throughout the years. She guided me into the wonderful field of linguistics and into the academic world. As an excellent researcher herself, and as a caring supervisor, she showed me the magic of doing research and helped me to start on my own topic. Her expertise in Chinese Linguistics and her insights into second language phonology and phonetics helped me significantly in my research for this dissertation as well as on other projects during my Ph.D. career. She has never hesitated to share her research experience, her collection of resources and her time when I needed them, and she has never hesitated to sit down with me to go over my conference papers, write reference letters for me or just talk about my problems, whatever the problems might be. She has always encouraged me to strive to be the best and has always supported me in doing so. I am sure I will continue to do so after my doctoral study. Dr. Lin is not only an academic advisor for my dissertation but also an advisor for life.

I have been most fortunate to work as a research assistant for Dr. John Esling. It is in his lab that I have broadened my outlook and learnt to be a diligent and precise researcher.

He is generous in providing not only comments and challenges but also financial support for me to be involved in academic activities. I am deeply indebted to Dr. Richard King for his unfailing support. I thank him for his quick replies with insightful feedback and for advice from a different angle. I also thank him for writing recommendation letters for me in the wee hours of the night. Dr. Sonya Bird has worked with me as a boss, recorded for me as a native speaker of English, and met with me as my committee member. Most importantly, she is a dear friend. She always shares her experience and provides her encouragement when I need it. My sincere thanks also goes to Dr. Yue Wang, who graciously agreed to be my external examiner despite her busy schedule.

I would also like to thank those people who are not on my committee but nevertheless helped me in one way or another. Thank you Carolyn Pytlyk, Shu-min Huang, Allison Benner, Scott Moisik, Jun Tian, Laura Hawkes, who helped me at various stages of my writing, with the experiment, in preparing to defend, and to be there when I needed to talk. I also want to express my sincere thanks to all the participants, native English speakers and Chinese learners of English, for their precious time and effort.

I would also like to acknowledge the financial support I received from the University of Victoria throughout my Ph.D. program. Thank you to the Department of Linguistics for employing me as a Teaching Assistant and sessional instructor. Thank you to the University of Victoria for the Ph.D. fellowship, Graduate Travel Grants, and various other awards. Thank you to Dr. Esling for employing me as a research assistant from the very beginning of my program to the last day of my program.

I would like to thank the Linguistics staff, Chris Coey, Gretchen McCulloch and Maureen Kirby. Without the help of Chris, the experiment would never have been possible. Without the support of Maureen and Gretchen, I wouldn't have been able survive my Ph.D. program.

Finally, I have to say thank you to my parents for their support and their belief in me. And, last but not least, I would like to thank Win for being there for me. It helped a lot to know that there is always someone there for you even if you fail.

DEDICATION

for Erik and my IBM x31

Chapter One

Introduction

1.1 Background

Lexical stress plays an important role in native speakers' perceptions and processing of speech (Field 2005, Culter & Clifton 1984). However, researchers have found that learners from a non-stress language background may not possess a system of stress in the same way that native speakers do. Learners may apply stress placement in L2 learning according to their native strategy, or they may indicate stress in a position where it would not fall in either the native or in the second language (Archibald 1993). French learners of English, with a fixed-stress background, could be said to be 'stress deaf' (Pepperkamp & Dupoux 2002). In other words, they have difficulties perceiving stress contrasts. Studies have found that Chinese learners of English are no exception. Juff (1990) found that Chinese speakers used a high level tone with extra length to indicate lexical stress. For Cantonese speakers, Chao (1980) indicated that they associated high and low tones with stressed and unstressed syllables. The Chinese participants in Archibald's study (1997) did not seem to have acquired principles of English stress assignment. They appeared to treat stress as a purely lexical phenomenon in the same way they treat tone.

Learners' problems in L2 stress acquisition have been attributed to L1 prosodic transfer. The Stress Deafness model (Dupoux et al. 2001, Pepperkamp & Dupoux 2002) proposes that whether learners can successfully perceive stress differences in an L2 depends on the regularity of stress assignment in their native language. The more regular the L1 stress system is, the more difficulties learners would have in L2 stress acquisition. The Stress Typology Model (STM), proposed by Vogel (2000) and Altmann & Vogel (2002), is similar to the Stress Deafness model. The STM predicts that learners from languages without a 'predictable stress' setting would have no problems in learning stress. These languages include non-predictable stress languages such as Spanish and also tone languages such as Chinese.

Despite the recognition of possible prosodic transfer from learners' L1 background, few acoustic analyses have been devoted to the investigation of the phonetic realization of L2 stress, especially in terms of perception. As Archibald (1997, 177) suggested, learners' problems in stress acquisition may result from their incapability "to utilize the cues for stress (vowel quality, heavy syllables, etc.)". Research on L2 *segment* development has already taken phonetic details into account to gain a better understanding of the possible transfer of L1. Different speech acquisition models, the Perception Assimilation Model (Best, 1994a, b, 1995, 2001), the Native Language Magnet model (Kuhl, 1991, 1993, 2000) and the Speech Learning Model (Flege, 1986, 1995), have proposed to explain L1 transfer in terms of the phonetic correlates of segments and segmental contrasts. In order to expand the understanding of the acquisition of stress by L2 learners and to offer insights into the more general area of L2 prosodic acquisition, detailed phonetic and instrumental studies are necessary.

1.2 The Perception Experiment

Studies on the acoustic correlates used by native speakers in stress perception have shown that they rely on F0, duration, and intensity in stress perception. This study focuses on the use of these three cues by Chinese learners of English in stress perception. The goals of the research are to find out:

- First, are Chinese learners 'deaf' to English stress?
- Second, if they can perceive English stress, how is their perception affected by the three cues?
- Finally, how will the heavy use of F0 for tonal perception in their native language affect their use of F0 in English stress perception?

To investigate the weightings of the three acoustic cues in L2 English stress perception, this study borrows Fry's (1955, 1958, 1964) classic design in studying English stress correlates. Adjustments were made so that the weightings of the cues are comparable between two groups of participants, native English speakers as a baseline, and Chinese learners of English. Nonsense words were constructed to reduce the difference in word familiarity between native speakers and Chinese learners of English. Factors other than the three acoustic correlates that were believed to affect stress judgments were minimized or controlled for. The three cues were manipulated, each at five levels of difference, and realized on the nonsense words. Previous research in L1 stress studies were used as a reference in the setting of the range and the manipulation of step size. The synthesized tokens were organized into a forced choice perception test. Chinese learners of English (CE) and native English speakers (NE) were asked to make stress judgments on the nonsense tokens. By looking at the stress judgment patterns recorded from the participants, we can compare the two groups in terms of the exact correlates they use in stress judgments and how they rank the these correlates in their stress judgment. The analyses of the data not only help us to evaluate whether Chinese learners are 'deaf' or not in stress judgments but also to clarify what correlates they use and rely on most heavily in stress perception.

1.3 Organization

The organization of the dissertation is as follows: Chapter 1 provides a brief introduction to the background of the study and the experiment used to achieve the research goals. Chapter two contains an overview of relevant literature. The importance of stress in L2 learning is discussed first. Stress acquisition models and general speech learning models are presented. A detailed description of the acoustic correlates for stress in English and for tone in Chinese is provided next., followed by an introduction to studies on the phonetic details in L2 stress acquisition. Chapter two concludes by providing the rationale of the current study, as well as the research questions and hypotheses. Chapter three discusses the construction of the stimuli and the detailed steps involved in the synthesis process. Participants involved in the study and the experimental procedure are also introduced in this chapter. Chapter four explains the reasons for the choice of statistical analyses. Results concerning the weightings of each cue in stress perception are compared within each group and then compared across the two groups. Chapter five presents the discussion of the results by referring to the hypotheses raised in chapter three. The implications of this study for L2 stress acquisition are also offered. The final chapter summarizes the research and concludes the dissertation by listing the contributions and limitations of this study and pointing out possibilities for future studies.

Chapter Two

Literature Review

Suprasegmental properties, including stress, play an important role in second language acquisition. They are shown to be closely related to foreign accent perceived in L2 production and to difficulties in L2 perception. Researchers have attributed the problems with stress to the influence of the L1 prosodic system. However, these studies are inadequate, as their discussion of stress acquisition mainly relies on the comparison of the phonological systems of L1 and L2. As Flege (1987) pointed out in research on L2 speech development at the segmental level, it is important to take phonetic details into account in order to gain a better understanding of the possible transfer of L1. The same is true for studies of prosody. It is possible that the influence of L1 lies in the difference between L1 and L2 in the employment of relevant phonetic correlates. In order to expand our understanding of the acquisition of L2 prosody, or more specifically, lexical stress, detailed phonetic and instrumental studies are necessary. Thus, in analyzing the possible influence of Chinese learners' L1 tonal background on their perception of English stress, we should refer to the acoustic correlates of stress and tone.

In this chapter, we first elaborate on the importance of suprasegmentals and stress in L2 acquisition in section 2.1. Section 2.2 reviews studies that have discussed the influence of L1 phonological systems on L2 stress acquisition and assess the inadequacy of such studies. In section 2.3, we present three speech learning models to offer more insight into the study of L1 influence in L2 acquisition, arguing that an analysis of the effects of L1 at the acoustic level is necessary in studies of L2 stress perception. Section 2.4 offers a detailed comparison between the acoustic features used in English stress and Chinese tone. The final section, section 2.5, describes some research that has studied the acoustic correlates of L2 stress perception and production.

2.1 The Importance of Suprasegmentals and Stress in L2 Acquisition

2.1.1 The importance of suprasegmentals

To learn a second language, pronunciation is always a difficult step, especially in adult years. Learners may have acquired perfect reading and writing skills while still being unable to communicate functionally in L2. Problems in pronunciation can be traced to segmental as well as suprasegmental difficulties. Although most previous research has been conducted on the segmental level, recent studies show that suprasegmentals may play a more important role than segmentals in the acquisition of a second language phonological system (Anderson et al. 1992, Derwing et al 1998).

Anderson, Johnson & Koehler (1992) investigated the nonnative pronunciation deviance at three different levels: syllable structure, segmental structure and prosody. The correlation between the actual deviance at the three levels and nonnative speakers' performance on the Speaking Proficiency English

Assessment Kit (SPEAK) Test was calculated. It was shown that while all three areas had a significant influence on pronunciation ratings, the effects of the prosodic variable were the strongest.

In Derwing, Munro & Wiebe's (1998) study, native speakers were invited to evaluate the final results of three types of instruction, i.e. segmental accuracy, general speaking habits and prosodic factors, and no specific pronunciation instruction, after a 12-week pronunciation course. Treated in three different ways, three groups of ESL learner were recorded reading sentences and narratives at the beginning and end of the course. Both the first and second groups, who received pronunciation instruction, showed significant improvement in sentence reading. However, only the second group, where prosodic factors were included in the instruction, showed improvement in accentedness and fluency in the narratives.

Researchers have also investigated learners of English from different language backgrounds and have found similar results (Johansson 1978, Palmer 1976, Anderson-Hsieh & Koeher's 1988, Munro 1995).

In Johansson's (1978) study of Swedish-accented English speech, segmental and non-segmental errors were compared in terms of accentedness scores. Native English judges were presented with two kinds of production, those with native English intonation but segmental errors on the one hand, and those with nonnative intonation (Swedish-accented) but no segmental errors on the other. Higher scores were assigned to productions with native-like suprasegmental characteristics but poor segmentals. Similarly, for French learners, Palmer (1976) found that the frequencies of suprasegmental errors were more correlated with intelligibility than were segmental errors in their production of English.

Studies with Chinese speakers have yielded the same results. In an earlier study, Anderson-Hsieh & Koeher, (1988) compared three male Chinese speakers' production of English on three levels: segmentals, syllable structure and prosody, including stress, rhythm and intonation. Although all three aspects were found to correlate with comprehensibility of the speakers' production, there was evidence that they didn't weigh the same in affecting comprehension. At a faster speaking rate, "prosodic deviance may more adversely affect comprehension". (Anderson-Hsieh & Koeher 1988: 585). In a more recent study, Munro (1995) used low-pass filtered English speech produced by Mandarin speakers for accent judgment. Untrained native English listeners were invited to rate the speech samples. It was found that non-segmental factors such as speaking rate, pitch patterns and reduction contribute to the detected foreign accent in Mandarin speakers' production and that their foreign accent can be detected based solely on suprasegmental information.

2.1.2 The importance of stress

Of the different components of suprasegmentals, lexical stress is one of the most important factors, yet the most complicated and least investigated one. "Lexical stress plays a central role in determining the profiles of words and phrases in current theories of metrical phonology" (Hogg & McCully 1987 in Field 2005: 403). Furthermore, word stress may also influence the intonation and rhythm of sentence production. Bond (1999) found that in processing speech, native speakers put more emphasis on the stressed syllables than the unstressed ones. In other words, they tend to ignore mistakes in unstressed syllables. In addition, misplacement of stress in a word is more likely to affect the processing of speech by native speakers than mispronunciation of a phoneme. In a study on the processing of lexical stress, Cutler and Clifton (1984) found that misplacement of stress in disyllabic words has detrimental effects in speech processing. A shift of the stress from the left syllable to the right syllable seriously hampered intelligibility. This can be illustrated with an example of WAllet, where capital letters represent the stressed syllable[1]. If the word is mispronounced as waLLET, native listeners have much lower efficiency in word recognition. One other interesting finding of the study is that if vowel quality change is also involved in the stress misplacement, greater effects on word recognition are observed. In other words, changing a full vowel into a reduced vowel or vice versa can compromise intelligibility severely, i.e. in the case of waLLET, [`wɒlɪt] → [wɒ`let]. In other words, incorrect placement of primary stress in L2 words may lead to miscommunication since the misplacement of lexical stress can "precipitate false recognition, often in defiance of segmental evidence" (Cutler 1984: 80). L2 learners, on the other hand, may not pay attention

[1] In the rest of the paper, a stressed syllable in a word will always be shown with capital letters.

to stress placement in listening to a stress language and may not use stress as a cue in lexical processing. In production, their stress mistakes can cause severe problems for a native speaker who may rely primarily on stress. Thus, to study second language learners' problems in lexical stress may lead to overall improvements in second language perception and production. Pedagogically speaking, it is pointed out by Dalton and Seidlhofer (1994) that, in pronunciation instruction, lexical stress is easier to teach than intonation but has greater communicative value than the phoneme. It is thus worthwhile to study in greater detail what learners' problems are with English lexical stress perception. Furthermore, better performance in stress perception and production may lead to overall improvement in intonation and rhythm. As a result, this research is designed to study the difficulties faced by Chinese learners of English in stress perception.

2.2 Studies in L2 Stress Acquisition

Although native English speakers rely heavily on lexical stress in speech perception and speech processing, learners of English may not be aware of the importance of stress. Speakers of a fixed-stress language or a tone-language may lack perceptual sensitivity to stress and stressed syllables. Researchers found that learners with a fixed-stress language background or a tone-language background have problems with stress acquisition (e.g. Archibald 1997, Peperkamp & Dupoux 2002, Altman & Vogel 2002,).

Archibald conducted a series of systematic studies of stress acquisition in L2 (1991, 1992, 1993a, 1993b, 1995, 1998, 2000). He focused on the phonological aspects of stress acquisition and studied learners with different language backgrounds. Archibald (2000: 152) addressed the question of L2 learners' acquisition of stress by suggesting that "their interlanguages are a combination of UG principles, correct L2 parameter settings (from resetting), and incorrect L1 parameter settings (from transfer)". From his studies with learners from three different language backgrounds, Spanish (a variable fixed-stress language) (Archibald 1993b) and Polish and Hungarian (fixed-stress languages) (Archibald 1992, Archibald 1993a), it seems that the metrical parameters of L1 are transferred in L2 stress acquisition. For example, Polish speakers uniformly stress the penultimate syllable, while Hungarian learners stress the initial syllable. In talking about the reasons for L1 influence on L2 parameter setting, Archibald mentioned the possible "mismatch between production and perception" (1993a, 46) and suggested that if learners can't perceive native-stress placement the way native speakers do, then the input will not "act as triggering data" for correct L2 parameter setting. This suggestion points to the need for more detailed phonetic study of how learners perceive stress.

In more recent years, two L2 stress perception models have been proposed independently to predict learners' performance in stress perception, based on their L1 phonological system. Both models were tested with learners from different language backgrounds. Peperkamp and Dupoux (2002) proposed the

“Stress Deafness” Model, which suggests that languages can be ranked according to the degree of stress predictability. The more predictable the stress is in a language, the harder it is for learners with this language background to discriminate stress differences in a second language. Dupoux et al. (2001) compared French learners with Spanish learners, and Peperkamp and Dupoux (2002) studied Finnish, Hungarian, and Polish speakers. It was concluded that French and Finnish speakers, with regular stress always at an utterance edge in their L1, have the poorest performance in stress discrimination. Hungarian speakers are better, because the stress is regular in their native language except for unstressed function words. Polish speakers are better than Hungarian speakers, because stress correlates less regularly with boundaries compared to the languages above. Spanish speakers, on the other hand, have varying stress assignment rules in their L1 and have the best performance in stress discrimination. However, the stress-deafness model doesn’t include non-stress languages. It is hard to predict the performance of learners from tone languages such as Chinese based on this model.

The second stress perception model is the Stress Typology Model, proposed in Vogel (2000) and Altmann and Vogel (2002). This model is based on a typology of stress phenomena. One of the advantages of the Stress Typology Model is that it attempts to include tone languages such as Chinese. The stress typology model divides languages into two groups, stress languages and non-stress languages. Stress languages can be further divided into predictable and non-predictable languages (similar to that of the Stress Deafness Model) and non-stress languages can be divided into pitch languages and non-pitch languages. Tone languages such as Chinese and pitch accent language such as Japanese are both examples of pitch languages, or what Hirst and Di Cristo called ton languages, where Japanese is classified as an accentual tone language and Swedish is classified as a tonal accent language (Hirst & Di Cristo 1998). Korean is argued by Altmann and Vogel (2002) to lack contrastive stress and also lack tone or pitch accent, and is thus classified as a "no pitch" language. Learners with L1 backgrounds of Spanish, French, Arabic, Turkish, Japanese, Chinese, and Korean were selected for participation in the experiments. These seven languages, together with English, were chosen as examples of language types, as shown in.

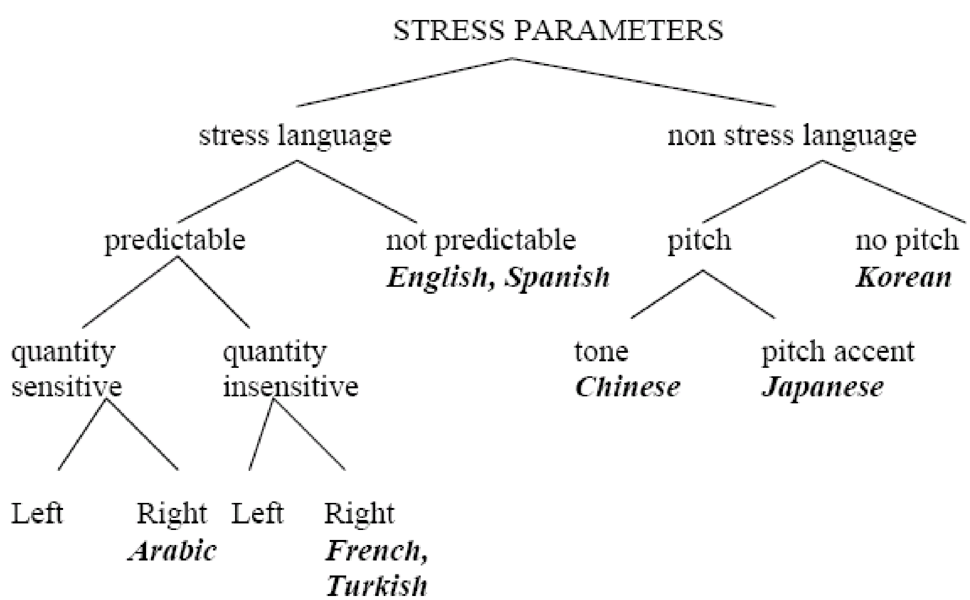


Figure 2.1 Proposed Stress Typology, taken from Altman 2006: 38.

In this experiment, nonce words with two, three, or four syllables were used as stimuli, and learners judged the stress position on the nonce words. Learners with L1 French, Arabic and Turkish performed significantly worse than learners from other language backgrounds. The results supported the two-way division between languages: languages with a positive parameter setting of predictable stress, and those with a negative setting of predictable stress. Languages with predictable stress have a positive setting and all other languages have a negative setting. The researchers posit that the positive parameter of predictable stress has "a detrimental influence on the listeners' ability to identify the location of primary stress in a word" (Altman 2006: 95). According to this model, Chinese learners have no problems with lexical stress perception.

However, according to many other studies with Chinese learners, different conclusions have been drawn. The Chinese subjects in Archibald's study (1997b, 175) do seem to have problems with lexical stress. Archibald suggested that they did not seem to be "acquiring the principles of English stress assignment with regards to such things as the influence of syllable structure or grammatical category on stress assignment," but they seemed to treat stress "as a purely lexical phenomenon" just the way they treated tone, as a part of the phonological representation of words that had to be memorized. In this study, Chinese learners were requested to complete both a production and a perception test, where they first read a list of English words with stress transcription on them and then listened to the same list of words read by native speakers and marked the stressed syllable. The tasks were repeated again 4 months later. It was shown that perception test results were worse than production test results, and that the results didn't improve after 4 months. Archibald (1997b: 177) suggested that this may result from their incapability "to utilize the cues for stress (vowel quality, heavy syllables, etc.)".

In other studies with Chinese learners, Cheng (1968) found that Chinese speakers of English interpret English stress as tones when they insert English words in their Chinese productions. Chao (1980) indicated that Cantonese speakers associate high and low tones with stressed and unstressed syllables. Similarly, Juffs (1990) studied nineteen undergraduate students' production of a passage in English. He found that Chinese learners produced English stress as tone. It was suggested that one of their problems may be that they used pitch movement instead of pitch height to indicate English stress. Also, they used tone one in Chinese, which is a high level tone, with an inordinate degree of length to indicate lexical stress.

Different from Altman and Vogel (2002), these studies about Chinese learners indicate that stress perception is problematical for them. Furthermore, the results of these studies imply that the root of this problem may lie in the influence of tone on their L1.

Why would the results of these studies contradict the prediction of the Stress

Typology Model? Both types of studies are based on the fact that Chinese is a tone language and assume that the L2 stress perception and production are influenced by the L1 tone system. The Stress Typology Model predicts that as Chinese is a tone language and stress is not predictable in Chinese, Chinese learners should not have problems with stress perception. However, studies focused on the production and perception of stress by Chinese learners found that they mixed English stress with tone in their native language. There is a short-coming of all the studies mentioned above. That is, all of them treated stress as a surface form without considering rules governing stress assignment and without understanding the processes involved in stress perception and production. Stress is phonetically defined by at least three acoustic cues, F0, duration and intensity. Thus, stress is actually the employment of these acoustic cues in speech perception and production. Rather than the question of whether learners can perceive stress or not, the key question is whether learners can employ these acoustic cues in a native way or not (Montero 2007). The contradiction found by the previous studies may be explained if we take this acoustic dimension into consideration. On the one hand, Chinese learners can perceive stress as a surface form in certain contexts, but on the other hand they differ from native speakers in their realization of stress and use a non-native strategy in stress perception and production. In the following section, three speech learning models are reviewed to provide insights into the ways L1 acoustics influence L2 speech learning.

2.3 Speech Learning Models

Linguists have long believed that perception and production of foreign speech are influenced by a listener's native language (Sapir 1921; Polivanov 1974; Abramson & Lisker 1970). The influence of the native language system on the segmental level has been widely studied (Goto 1971; Best & Strange 1992; Best 2001; Werker et al. 1981; Werker & Tees 1984). The three important models proposed in the area of segmental acquisition are Best's perceptual assimilation model (PAM) (Best, 1994a, b, 1995, 2001), Kuhl's Native Language Magnet model (NLM) (Iverson and Kuhl, 1996; Kuhl, 1991, 1993, 2000), and Flege's Speech Learning Model (SLM) (Flege, 1986, 1995).

2.3.1 Native language magnet

Kuhl (1993, 2000) proposed the Native Language Magnet model. NLM holds that infants are equipped with a discriminative ability at birth to categorize phonetic units. They make use of the pattern information and of the statistical properties of the language input in speech learning. Through language development, an individual's perception is gradually distorted by his/her language experience (Iverson & Kuhl, 1996) and the acoustic dimension underlying, speech is warped (Kuhl, 2000). With more input from their native language, they gradually develop acoustic prototypes for native phonemic categories. However, in L2 speech learning, such acoustic prototypes for non-native categories are not created, due to insufficient relevant acoustic experience. Our native language acts as a filter and changes what we attend to in speech perception. The acoustic space is expanded or shrunk to highlight the contrasts in the native language. This language-specific filter makes L2 speech learning much more difficult because we may not be aware of dimensions of speech that are not important in L1 learning.

Iverson and Kuhl (2003) used synthesized speech stimuli to study the perception of L2 sound contrasts. Japanese speakers were compared to native English speakers in the perception of syllables begin with /r/ and /l/ in English. The stimuli were systematically manipulated for two acoustic cues, F2 and F3. There were three steps for F2 change and six steps for F3 change, producing a total of eighteen stimuli (3 steps of F2 × 6 steps of F3). Native American listeners perceived the 18 stimuli as either instances of /r/ or instances of /l/. The 18 stimuli, spaced equally in terms of the two acoustic cues, were not spaced equally in the perception map of American listeners. The so-called magnet effects and boundary effects were observed for them. The magnet effect refers to the shrinking of perceptual space for the good instances of either /r/ or /l/ categories. In other words, allophonic or free variations of either /r/ or /l/ are perceived to be close to the prototypical /r/ or /l/ in the language, although their actually acoustic distances to the prototypes are further. The boundary effect, on the other hand, refers to the stretching of perceptual space at the division area of the two categories. This means that an instance of /r/ and an instance of /l/ could be perceived to be two totally different segments (perceptually further apart) but acoustically closer to each other than their respective distances to the prototypes. Thus, around the boundary of the two segments, there is a perceptual division that is exaggerated despite the acoustic similarity. Japanese listeners showed a perceptual map that was totally different from native American speakers. First, they seemed to differ from American listeners in the acoustic dimensions that they attend to in perception. While native listeners were most sensitive to F3 cues, Japanese listeners were sensitive to F2 cue variation. Second, no magnet or boundary effect was observed for Japanese listeners on the dimension of F3. Only one category of sound emerged from the Japanese perceptual map.

This study shows that the symmetric acoustic dimensions are distorted by native speakers in speech learning to maximize the difference between contrasts in their native language. The distorted perceptual map, once formed, makes L2 speech learning more difficult. In other words, if there is an L2 contrast that doesn't exist in L1, then it is very hard to create a perceptual map for this contrast, as in what Japanese listeners have experienced. It can be inferred from this model that even if L1 and L2 share a same or similar contrast, the perceptual map may not be the same in the two languages because speakers may rely on different acoustic dimensions in perception which still make it difficult for a L2 learner to form accurate categorizations.

2.3.2 Perception assimilation model

Best (1995, 2001) proposed an L2 speech learning model, the Perception Assimilation Model (PAM). This model explicitly draws on Articulatory Phonology and argues that listeners discriminate the speech signal based on information about articulatory gestures (e.g. Best 1995k Fowler et al. 1990, Browman & Goldstein 1992). These gestures, in turn, "are defined by the articulatory organs (active articulator, including laryngeal gestures, constriction locations place of articulation), and constriction degree (manner of articulation) employed" (Best 2001: 777). PAM proposes that the listeners' native knowledge, whether implicit or explicit, has a strong effect on the perception of non-native speech, and listeners have a strong tendency to assimilate non-native sounds to a native phoneme or category which is similar in terms of its articulatory gestures (Best 1995, 2001). PAM (Best, 2001) predicts that a non-native phone can be assimilated to the native phonological system in one of the three ways: as a categorized phone, an uncategorized sound, or a nonassimilable nonspeech sound. More importantly, PAM not only predicts the assimilation of a single non-native phone but also the assimilation of a non-native contrast. Depending on the assimilation pattern of the two non-native phones involved in the contrast, six types of assimilation are predicted for non-native contrasts. They are Two Category assimilation (TC), Single Category assimilation (SC), Category Goodness difference (CG), Uncategorized-Categorized pair (UC), Uncategorized assimilation (UU), Non-Assimilable pair (NA). When the two phones of a non-native contrast are assimilated to two native categories, the contrast is perceived as TC and when both are assimilated to a single category, then the contrast is SC. CG refers to the case when both phones are assimilated to one category but one is assimilated better than the other. When one is categorized and one is not, the contrast is UC and when both are not categorized, the contrast is UU. When both phones are unassimilable, then the contrast is predicted to be NA. The discrimination of the six non-native contrasts can be affected by the native phonological system in different ways. To be more specific, L1 phonology should have a positive effect on discrimination of TC and UC contrasts. When a non-native segmental contrast can be categorized as either TC or UC, learners would find it easier to differentiate the non-native sound segments. The effect of L1 phonology may be neutral for NA contrasts, neither positive nor negative. For SC or CG, L1 phonology is predicted to have a negative effect.

Different from both NLM and SLM (discussed below) which focus on the attributes of phonetic categories, PAM is the modal that is phonological in nature. As Best commented (2001: 791), “PAM instead focuses on the functional organization of the native phonological system, specifically on the phonological distinctions between, and phonetic variations within, native phonological equivalence classes.”

2.3.3 Speech learning model

Flege (1987, 1991, 1995, 2004) and colleagues proposed the Speech Learning Model (SLM). Working hypotheses of SLM were proposed by Flege in 1995 with supporting evidence from empirical studies. Similar to the previous two models, SLM posits that listeners' speech perception is attuned to the contrasts in the L1 phonological system. In the acquisition of an L2, contrastive phones may not be perceived as contrastive because the L1 phonology may have prevented the listeners from attending to "the features or properties of L2 sounds that are important phonetically but not phonologically, or both" (Flege 1995). SLM's difference from PAM is that SLM is not specifically built on articulatory phonology. Flege has not been very explicit about what "the features and properties" of L2 sounds are. They may be articulatory gestures or acoustic properties. In his experimental studies, he has focused on the acoustic properties in speech learning. Unlike NLM, SLM focuses on adult speakers, especially bilingual speakers' acquisition of speech sounds in L2. Furthermore, SLM assumes that the construction of new L2 categories is possible. It is proposed that non-native sounds are classified according to their "equivalence" to existing sounds. It is less possible for a new L2 category to be created when the shared similarity is large. The correct perception of a more similar sound is moredifficult. In other words, L2 learners can master a 'new' sound in the target language but not a 'similar' sound.

Flege and his colleagues have conducted experimental studies to test SLM in the perception of L2 speech sounds and the contributions made by acoustic correlates were discussed in these studies (Flege 1987, 1993, 1995). For example, in the studies on the acquisition of English voiceless stops, it was found that learners with the same phonological voiceless stops in their L1 but different VOT settings have great difficulties with the perception and production of the English voiceless stops. Flege suggested that the correct categorization may "be blocked by the continued perceptual linkage of L1 and L2 sounds" (Flege 1995: 258). In a different study with German learners of English, Bohn and Flege (1992) found that German learners can be trained to perceive and produce the 'new' vowel /æ/ in English. Thus, the researchers concluded that, unlike a similar sound, with enough time and exposure to the new phone in L2, a new category in L2 can be created.

In general, SLM is not specifically designed to account for speech perception in a non-native language, and it uses the accuracy and failure in L2 speech perception to explain acquisition of L2 production. On the one hand, it offers a broader view of L2 speech acquisition, which incorporates not only perception and production, but also discusses factors such as Age Of Arrival (AOA) or Age Of Learning (AOL). On the other hand, it lacks the detailed account of why and how L2 perception is different from (or similar to) L1 perception, or why AOA or AOL would have an effect on L2 perception and production.

Although the three models differ in their beliefs about the native perceptual framework and how L2 sounds or sound contrasts are mapped to the L1 system, they hold the same view that "adults' discrimination of non-native speech contrasts is systematically related to their having acquired a native speech system" (Best 2001: 776). All three models have made important contributions to the study of speech learning. They have offered sophisticated proposals for the possible influence of L1 on L2 speech learning from different levels, phonological, phonemic, phonetic and acoustic. Many experimental studies were conducted to verify and evaluate these different claims.

In the discussion of speech learning, Flege (1995) also pointed out the importance of suprasegmentals in L2 acquisition and indicated that not only segmental but also prosodic divergences may lead to foreign accent. While the proposals about speech learning made by these models should apply to the perception of suprasegmentals, none of the models has made explicit predictions about the acquisition of suprasegmentals. Furthermore, the experimental methods have been used mainly with the study of segment perception. The scarcity of studies on suprasegmentals may be attributed to the complicated nature of suprasegmentals. While it is comparatively easier to identify a distinctive phoneme, to define stress is never an easy task.

The present study focuses on one aspect of suprasegmentals, lexical stress, in L2 acquisition. It is not designed to compare or evaluate the three models. Rather, it is an attempt to apply experimental methods used in the studies of L2 segment perception to the study of L2 stress perception. The speech learning models discussed above are not viewed as conflicting models and among which one is the best. Rather, they have their different perspectives and focus, and they are used to guide research design and to inform the discussion of results from different levels.

2.4 Acoustic Features of Mandarin Chinese Tone and English Stress

According to the taxonomy of stress systems mentioned above, in Figure 2.1,, there are four kinds of stress systems in natural languages which belong to either accentual or nonaccentual languages, following Altmann & Vogel (2002). Hirst and Di Cristo differ in the refinements of these distinctions. Stress languages all belong to accentual languages and tone languages are all nonaccentual languages. Pitch accent languages share features of both accentual and nonaccentual languages. While Chinese is a typical tone language, English is a typical accentual language with movable stress assignment according to this taxonomy.

Beckman (1986: 27-44) in her book *Stress and Non-stress Accent* provided four characteristic differences between stress and tone, i.e. speakers' attitude, historical development, distinctive load and alternations and restrictions. It was pointed out that "accent is fundamentally different from tone, suggesting that the fundamental difference lies in tone's functional role in differentiating words as opposed to accent's role in organizing utterances"(Beckman 1986: 44). From the phonetic perspective, different acoustic cues or cue combinations other than pitch are employed to signal stress in a stress accent language (Beckman 1986). For example, in addition to fundamental frequency, English uses duration, intensity and vowel quality for stress distinctions (Beckman 1986, Lehiste 1970). In contrast, in a tone language, pitch is used to contrast an individual lexical item with another that could have occurred in the same place (Jones 1950 in Nguyen 2003, Gandour 1978). In the following sections, the acoustic correlates of Mandarin tone and English stress are discussed in more detail.

2.4.1 Acoustic correlates of English stress

The acoustic correlates of English stress have been explored extensively by researchers. There are four generally recognized correlates of English stress, F0, duration, intensity and formant structure. In perception, the four correlates can be translated as pitch, length, loudness, and quality (Fry 1955, 1958, Lehiste 1970). Lehiste (1970: 153) commented that "the perception of stressedness appears to be based on a number of factors, the most influential of which is fundamental frequency, other phonetic correlates of stress, besides fundamental frequency and intensity, include vowel quality and duration."

In a series of classic studies, Fry (1955, 1958, 1964) used synthesized stimuli to study the effect of F0, duration and intensity, and also formant structure in native speakers' perception of stress position. In the first study, Fry (1955) compared the effects of duration and intensity on English stress perception using five recorded real words. All five words were disyllabic and had two possible stress patterns. Duration and intensity were manipulated at the same time, each with five steps, resulting in a total of 125 test items (5 words × 5 steps of duration manipulation × 5 steps of intensity manipulation). One hundred native speakers judged the stress position of these tokens. It was found that native speakers' stress judgments were influenced by both duration and intensity manipulation, but the change of duration had a stronger effect than intensity It was concluded that duration serves a more marked role in stress perception than intensity. In a later study (Fry 1958), the effect of F0 was studied. F0 was varied on one word, *subject*, in combination with a five-level duration manipulation. One of the two syllables was kept at a reference frequency of 97Hz and the frequency of the other was manipulated to be higher in eight unequally-sized steps, 5, 10, 15, 20, 30, 40, 60, 90 Hz. The manipulation was first realized on the first syllable and then on the second syllable. Forty-one subjects participated in the study and judged the stress position. A consistent effect of F0 was observed at each level of duration manipulation. Higher F0 on the first syllable significantly brought up the percentage of initial stress judgments compared to higher F0 on the second syllable. Fry (1958: 151) also indicated that F0 "tends to produce an all-or-none effect, that is to say the magnitude of the frequency change seems to be relatively unimportant while the fact that a frequency change has taken place is all-important". Whether the first syllable was 90 Hz or 5 Hz higher than the second syllable was of little importance. The fact that the first syllable was higher in F0 could make listeners judge the syllable to be stressed. F0 as a cue to stress outweighed both duration and intensity. These findings were confirmed by several other studies using both synthesized and natural speech (Mol &

Uhlenbeck 1965, Morton & Jassen 1965). All these studies suggest that F0 is the most important cue for English stress, and duration is the second most important. Intensity has the smallest contribution.

Differing from Fry, Bolinger (1965) concluded from his study that pitch prominence was indeed the primary cue of stress, but both duration and intensity were only secondary. He commented that duration is only an "auxiliary and residual cue" and intensity is only "negligible both as a determinative and as a qualitative factor in stress". Lehiste (1970) agreed that the status of F0 is a very strong cue for stress, while the status of intensity is somewhat ambiguous, and is thus, a comparatively weak cue for stress. Lieberman (1960) used noun-verb word pairs in American English, such as OBject-obJECT, as stimuli in his study. Both F0 and envelope amplitude were found to be closely correlated with stressed syllables. "The fundamental frequency seems most relevant" (453), a result which is consistent with Fry's (1958). However, "the envelope amplitude seems more important than duration" (454). Sluijter & van Heuven (1996) examined the intensity as a cue for stress. Intensity differences located above 0.5 kHz was found to be an important correlate for stress. Also duration proved the most reliable correlate of stress. Overall intensity and vowel quality were the poorest cues.

A fourth correlate, formant structure, has also been found to be relevant for English stress perception and production. Fry (1958: 128) pointed out that “certain quality differences in English have particular significance in stress judgments. The substitution of the neutral vowel /ə/ for some other vowel, the reduction of a diphthong to a pure vowel, or the centralization of a vowel were all powerful cues in the judgment of stress.” In Fry’s 1965 study, the relative weights of formant structure and duration were compared by using synthesized stimuli. It was concluded from the study that formant structure served as a much weaker cue for stress perception, and it should be considered the least effective cue among F0, duration and amplitude.

It must be pointed out that all of the previously mentioned studies agreed that stress is not the result of a single mechanism. The perception of stress is, instead, the result of a composition of all these factors and can only be accounted for by referring to all the factors mentioned above (Crystal 1969 in Adams & Munro 1978). Although researchers don’t agree totally with each other on the weight of an individual correlate for English stress, “it has now been fairly well established that the most consistent acoustic correlates of stress are fundamental frequency, amplitude (or intensity), and duration” (Adams & Munro 1978: 128). As McClean and Tiffany (1973: 173) summarized, “in relation to unstressed syllables stressed syllables have increased magnitude of these parameters.”

2.4.2 Acoustic correlates of Mandarin tone

The acoustic correlates for Mandarin Tone are not very different from that for English stress. However, the weights of these correlates are quite different in the two languages. The three main acoustic correlates for Mandarin tones include fundamental frequency (F0), duration and amplitude[2]. Among them, F0 is believed to be the primary acoustic cue for Mandarin tones (Rumjacev 1972 in Coster & Kratochvil 1984; Howie 1976; Wu 1986; Tseng 1990; Moore and Jongman 1997; Fu & Zeng 2000; Jongman et al. in press). Most phonetic analyses of Mandarin tones focus on the measurement and description of F0.

2.4.2.1 Fundamental frequency

Mandarin has a syllable-based tone system (Lin 1996). Pike (1948) divided the syllable based tone system into a register tone system and a contour tone system. In a contour tone language such as Mandarin Chinese, the pitch trajectory is more important than pitch height (Pike 1948). While pitch is a perceptual concept, the acoustic correlate for this perceptual correspondence is expressed in fundamental frequency (Laver 1994).

Phonetic studies of Mandarin Chinese tone contours have found that F0 is the primary acoustic parameter in characterizing Mandarin tones. (Rumjacev 1972 in Coster & Kratochvil 1984, Howie 1976, Lin 1988, Wu 1986, Tseng 1990,

[2] Amplitude and intensity are used interchangeably in the present paper

Moore & Jongman 1997, Fu & Zeng 2000, Jongman et al. in press). Howie (1976) based his description of F0 pattern on the recording of two male native Mandarin speakers' production of 15 syllables of each tone. The F0 values of four tones are listed in Table 2.1.

Table 2.1 F0 values for Mandarin tones (generalized from Howie 1976)

Tone	F0 pattern			
	beginning	F0 drop	F0 rise	ending
Tone 1	150 Hz	-	-	150 Hz
Tone 2	115 Hz	-	35 Hz	150 Hz
Tone 3	113 Hz	40 Hz	40 Hz	113 Hz
Tone 4	157 Hz	52 Hz	-	105 Hz

Figure 2.2 from Moore and Jongman (1997: 1865) gives a more direct representation of the tone contours.

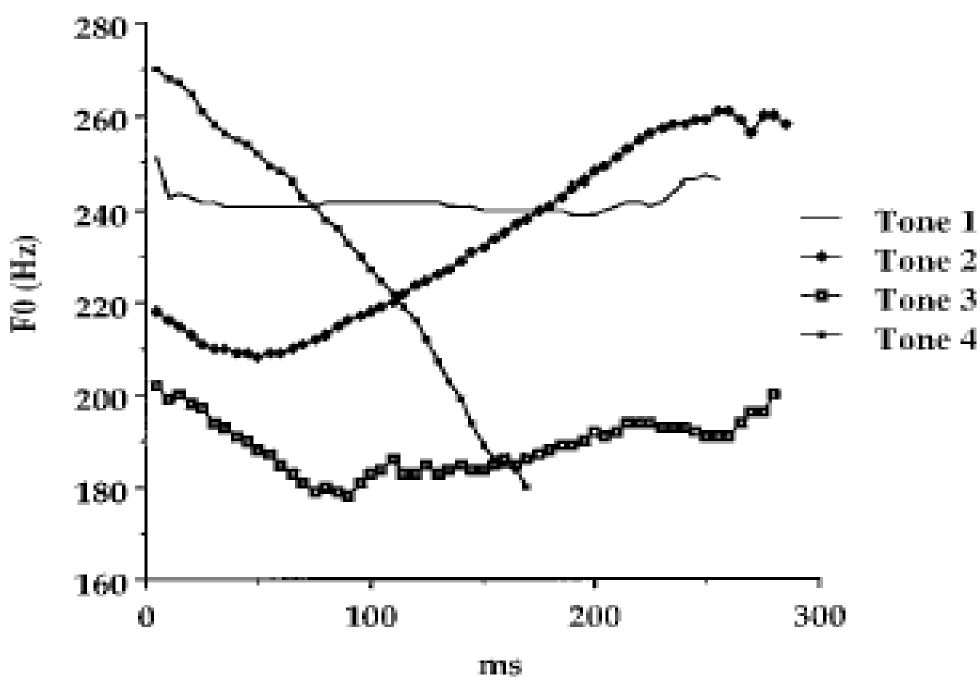


Figure 2.2 F0 contour (Moore & Jongman 1997: 1865)

These F0 characteristics presented above generally agree with Chao's (1948) phonological description of the four tones. Tone 1 is a high level tone which starts high and remains high throughout its duration. Tone 2 is a rising tone which starts at mid pitch and ends slightly higher above the pitch level of Tone 1. Tone 3 starts low and extends lower but finally increases by the offset. Tone 4 begins high and falls to the bottom of the range.

Different studies support the view that F0 is the primary cue to the perception of Mandarin tones. Howie (1976) used synthetic speech in perception tests in three conditions, synthetic speech modeled after natural F0 patterns, synthetic stimuli in which the F0 contours were made to sound monotone, and stimuli synthesized to sound like a whisper. Subjects were found to perform much better at identifying stimuli in which the pitch pattern was maintained. The perception of the monotone and whisper stimuli was just slightly above chance level. Howie (1976, 245) concluded from the study that Mandarin speakers "could apparently make little use of any features other than pitch as cues for the perception of tonal distinctions." Using the recording of one native female speaker in free speech, Coster and Kratochvil (1984) found that "the F0 properties of the syllables' tone-carrying parts alone are …powerful in tone discrimination." With F0 value alone, over 86% of the syllables can be identified correctly in the tone discrimination analysis. Lin (1988), using synthesized speech, reached similar conclusions. F0 was found to provide the highest discrimination percentage. If F0 contour was changed into a constant F0 value, any change in duration or amplitude parameters would not lead listeners to perceive different tones. Gandour (1984) studied the tone dissimilarity judgment by native Chinese speakers. It was found that while both F0 height and F0 contour are important cues for tone perception, Mandarin speakers seem to rely more on F0 contour.

2.4.2.2 Duration

In production, tones are consistently different from each other in the domain of duration, as can be seen clearly from Figure 2.2 above. In other words, from a descriptive perspective, Mandarin tones differ in terms of overall duration (Kratochvil 1971, Howie 1976, Tseng 1990). While Tone 3 is recognized as the longest one, Tone 4 is usually the shortest. In between are Tones 1 and 2 with Tone 1 being significantly shorter than Tone 2. Table 2.2 generalizes the average tone duration in Mandarin as measured by Tseng (1990)

Table 2.2 Tone Duration (taken from Tseng 1990)

Tone	**Tone 1**	**Tone 2**	**Tone 3**	**Tone 4**
Duration (ms)	351.76 ms	376.28	457.04	236.12

In perception, however, the role of duration in the tonal distinction is very limited. Coster and Kratochvil's (1984) study showed duration to be an unsuccessful cue in tone categorization. When the tone discrimination test was solely based on duration value, the classification result was hardly meaningful (only around 40% correctly classified). This is also supported by Lin's (1988) perceptual study of Mandarin. It was estimated from the discrimination test results that the influence of duration in tone perception is only of around 3%. Fu and Zeng (2000) used processed stimuli in their study to investigate the effect of duration in perception test. The stimuli with the naturally varying vowel duration preserved yielded the lowest tone recognition score (35.6%).

On the other hand, some researchers do find duration to be a relevant cue when there is ambiguity in the identification between two tone categories. For example, duration is found to enhance the tonal distinction between Tones 2 and 3 in Mandarin for both native and non-native speakers (Blicher et al. 1990, Tseng 1990, Lin & Repp 1989).

2.4.2.3 Intensity

Studies of Mandarin tones have also investigated the relevance of amplitude as a correlate of tone, in addition to F0 and duration. Chuang et al. (1972 in Jongman et al. in press) found that Tone 4 has the highest overall amplitude while Tone 3 has the lowest. Fu and Zeng (2000: 46) suggested that amplitude contour is highly correlated with F0 contour, as can be demonstrated in Figure 2.3.

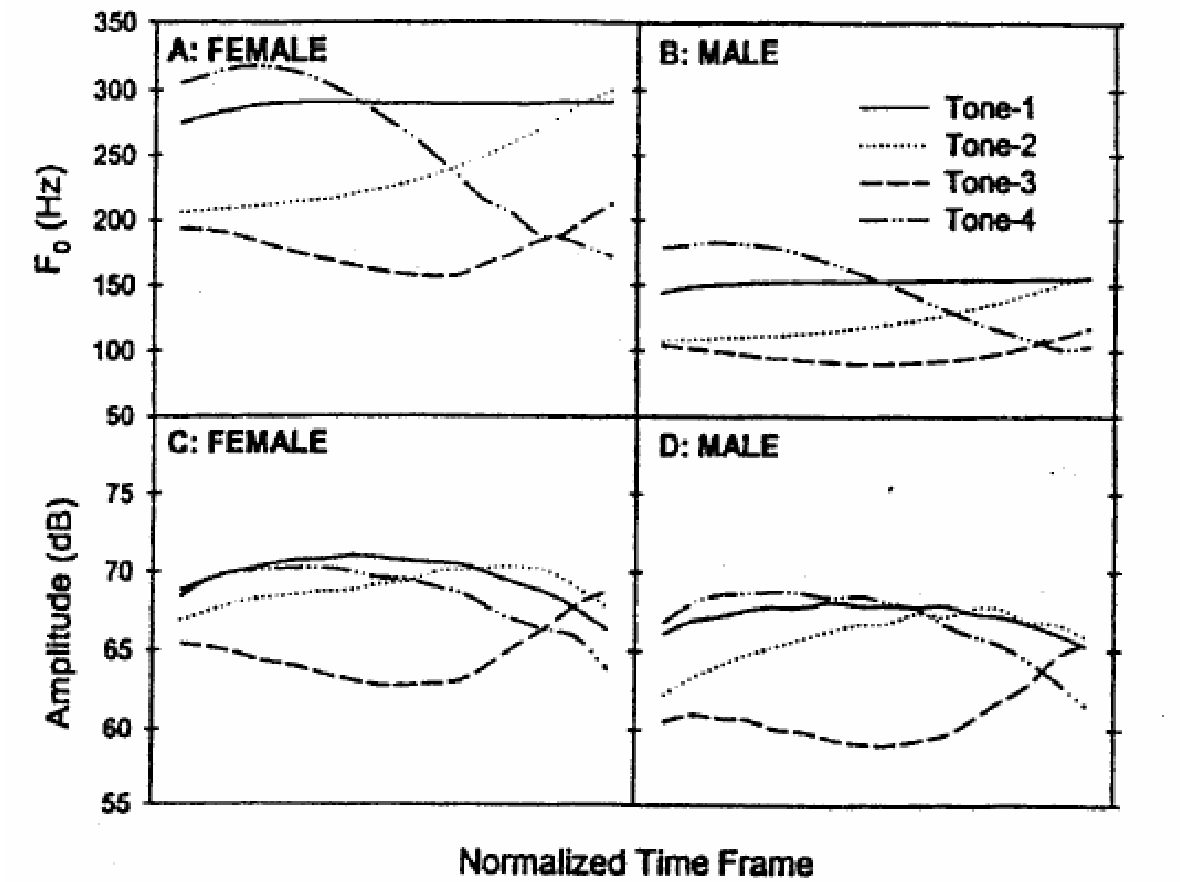


Figure 2.3 Tone Amplitude Contour (Fu & Zeng 2000)

The top two panels are the F0 contour from a female and a male speaker and to compare with these are the amplitude contour from both speakers. As can be seen, tone 1 has a flat amplitude contour with tone 2 rising. Tone 3 exhibits a falling-rising contour and tone 4 is falling. The amplitude contour resembles F0 contour to a great extent.

Many researchers recognize the secondary role played by amplitude in tone perception. In Fu and Zeng's (2000) study, listeners achieved higher perceptual scores (58%) based on amplitude-stimuli than duration-stimuli (35.6%). Whalen and Xu's study (1992) used manipulated stimuli in a perception test of Mandarin tones. F0, formant structure and duration were all controlled. The result showed that listeners can achieve 'reasonable accuracy solely from the amplitude contour' (Whalen U Xu 1992) in tone perception. Similarly, Coster and Kratochvil's (1984) study found that based on overall amplitude alone, tone discrimination was quite successful (over 71% correctly classified). It was suggested that the two dimensions of F0 and amplitude putting together determine the formation of tone and provide the best discrimination result.

In sum, phonetic studies on acoustic characteristics of tones have identified F0 as the most important cue for tonal distinction. Amplitude seems to be a better cue than duration, although its role is only secondary and supporting in nature.

2.4.3 Stress in Mandarin Chinese

Mandarin Chinese is a typical tone language, and it also has stress (Lin 2001). However, the exact nature of stress is still a topic under investigation in Chinese phonology.

In Lin's (2001) discussion of the phonological system of Mandarin Chinese, three types of stress patterns were identified: contrastive stress, normal stress and weak stress (see Lin 2001 for a more detailed review). An example of contrastive stress can be seen in the cases of the Mandarin words *chaoFAN* vs. *CHAOfan*, with stress on the second syllable in the first item and stress on the first syllable in the second item. The two items have exactly the same segmental contents and exactly the same tones on both syllables. The first item means 'to stir fry RICE (but not noodles or vegetables)' while the second item means 'STIR FRIED rice (but not steamed rice)'. An exaggerated pitch range and greater intensity can be observed on the stressed syllable. The second type of stress is normal stress, which means that no exaggerated pitch or intensity is used. In a sequence of syllables having normal stress, the last syllable is the strongest and the first is the second strongest, and all intervening syllables are weaker. For example, for the word *Victoria*, the Mandarin Chinese version is *weiduoliya*, the last syllable *ya* is the strongest, the first syllable *wei* is the second strongest and *duoli* are the weaker ones in this word. The third type of stress is weak stress.

Grammatical particles in Mandarin are usually pronounced with neutral tones, and at the same time bear weak stress.

In addition to the above view of Mandarin stress, Duanmu (2000) argues that Mandarin Chinese has initial stress (based on left-headed feet) on the word level. His proposal is based on the analysis of toneless particles (neutral tone particles), compound or phrase facts, or the historical views.

It is not an intention of this study to evaluate these views of stress in Mandarin Chinese. While it is true that stress phenomena are observed in the Chinese metrical system, it should be pointed that the domain of stress discussed above is usually larger than word. Most Mandarin Chinese words are monosyllabic, which are independent and meaningful morphemes, and longer words can be formed by grouping these monosyllabic words into larger units (Shen 1993). The examples of contrastive stress discussed above are actually units larger than words. The contrastive stress is different from the word stress in English and can thus be argued to be compound stress or phrasal stress, which is beyond the scope of the present study. Furthermore, stress can also be a covariate of tone in Chinese. The third type of stress discussed above is an example of this. Weak stress co-exists with neutral tone. Stress, in this case, does not serve a demarcative function. The second type of stress discussed above shows the closest resemblance to the idea of lexical stress in English being discussed in this study. However, it does not serve demarcative or contrastive functions at the word level that can be consistently used to separate words. While we recognize the existence of stress in Chinese, we believe that what is comparable to lexical stress in English is tone in Mandarin Chinese.

2.5 Studies of Acoustic Correlates in L2 Stress Acquisition

As it can be seen from section 2.4, the correlates of stress in English and tone in Mandarin have been widely studied for native speakers throughout the years. However, the exact correlates and the nature of the interaction between these correlates are "much less… considered for non-native speakers of the language who in their mother tongue may well signal stress in quite a different way from native speakers of English — especially if the first language is syllable-timed (Adams & Munro 1978: 128)". Even less is known about learners whose first language minimize stress contrasts. The following section introduces some research concerning the employments of different acoustic cues in L2 stress perception and production.

Adams and Munro (1978) compared the production and perception of English stress in connected speech by eight native and eight non-native speakers with an L1 Asian language, such as Thai or Chinese, and found that "native and non-native speakers use the same parameters as cues for stress although the exact ways of how these parameters are used in production and perception of stress by the two groups still bear differences" (1978: 152). It was found that fundamental frequency correlated with stress more often in the utterances of the non-native speakers than in those of their native English-speaking counterparts.

More specifically, for non-native speakers in the investigation, who were all speakers of Asian languages, the direction of the perceived pitch movement rather than the relative level of the tone was of greater importance. In Adams and Munro (1978) and Adams (1979), it was also found that non-native speakers differentiated less in duration between stressed and unstressed syllables than did native speakers. Non-native speakers' unstressed syllables were, generally speaking, longer than that of native speakers, although they resembled native speakers in duration for their production of stressed syllables. In other words, duration was used less as a cue for stress in non-native speakers' production. Amplitude, on the other hand, was shown to be a more important cue for non-native speakers than native speakers in the two studies (Adams & Munro 1978, Adams 1979). For non-native speakers who used amplitude as a cue to stress, the difference between the peak level of amplitude for stressed syllables and the peak level of amplitude for unstressed syllables was much higher. For native speakers, such differences were almost negligible, 32.8 and 32.6 dB. It was further suggested by Adams and Munro (1978: 148) that the staccato effect observed in the non-native speakers' production can probably be accounted for by a sudden amplitude fall at the boundaries of the stressed syllable, which is quite different from native speakers, who used amplitude less as a cue for stress.

As far as the correlate of duration is concerned, Bond and Fokes' (1985) finding confirms Adams' study. Nonnative speakers representing several language backgrounds exhibited less syllable duration differentiation than native speakers at the word level. Anderson-Hsieh and Venkatagiri (1994) studied intermediate-level Chinese learners of English and showed that they "failed to differentiate sufficiently in duration between tonic and unstressed syllables" (Anderson-Hsieh & Venkatagiri 1994: 810)

Montero (2007) compared the effects of F0, duration and intensity on stress perception by three language groups, French and Spanish learners of English and native English speakers. Participants heard synthesized disyllabic reiterant nonsense tokens. Three-step manipulations of each acoustic cue were realized on the nonsense token. The duration baseline was 240ms and the second syllable was set to be 1.25, 1.50, and 1.75 times baseline. The intensity baseline was set to be 60dB, and a range of 9dB was used, the lowest intensity of the variation being 51dB and the highest being 57dB. The baseline F0 was set to be 140Hz, and the manipulation of F0 was set to be 150, 160 and 170Hz, with a range of 30Hz. The results of the perception experiments showed that native English speakers performed better than both French and Spanish speakers. F0 was proven to be used by all three groups in stress perception. On the whole, intensity was also shown to be relevant in stress perception. The effect of duration was the smallest among the three acoustic dimensions investigated. The researcher proposed that unlike the prediction made by the Stress Deafness Model, French speakers were not insensitive to F0 and intensity variations although they showed different patterns of sensitivity to F0, duration and intensity from native English speakers and Spanish learners of English.

Ueyama (2000) studied L1 Japanese L2 English speakers and L1 English L2 Japanese speakers. The contrast between lexically accented and unaccented vowels was analyzed. In order to investigate the influence of L1 phonetic habits in L2 speech acquisition, L2 English production and L2 Japanese production were analyzed. English is a stress language and Japanese is a pitch accent language. F0 is an acoustic correlate for both stress and pitch accent. Duration is actively involved in marking stress but not a correlates for Japanese pitch accent. However, duration is an active phonemic feature in Japanese. There is difference between short and long segments. Thus, duration is an active acoustic cue in Japanese, but not at the level of suprasegmentals. In the experiment with L2 Japanese production, participants read sentences with a pair of Japanese words with the pitch accent on different syllables. The measurements of F0 and duration were taken from the target words. Results showed that both groups of participants used F0 to indicate stress or pitch accent differences in L2 speech. As far as the duration cue was concerned, beginning Japanese learners of English were unable to use duration as a systematic cue for stress discrimination. More advanced learners of English, on the other hand, showed native-like patterns in the use of duration cues. The research concluded that beginning L2 learners tend to import the L1 phonetic habit directly in L2 speech learning. L2 prosody gradually develops when learners modify the L1 habit to accommodate L2 speech prosody.

Nyugen (2003) and Nyugen and Ingram (2005) investigated the prosodic transfer of the tonal constraints on Vietnamese acquisition of English stress and rhythm. The acoustic cues used in Vietnamese speakers' perception and production of English stress were analyzed to examine the influence of the learners' tonal background in L2 learning. Vietnamese is a tone language, and the researcher proposed that both F0 and intensity are acoustic correlates for tone in Vietnamese. Duration is non-contrastive for tone perception in Vietnamese. On the other hand, stress correlates in English include not only F0 and intensity, but also duration and vowel reduction. In the experiments conducted by the researchers, three groups of participants were involved: one group of beginning-level Vietnamese learners of English, one group of advanced-level Vietnamese learners of English and a control group of native English speakers. Participants read pairs of sentences with minimal pairs of nouns and verbs such as 'PREsent' and 'preSENT'. The acoustic measurements included fundamental frequency (F0) and intensity values on the mid-point of vowel, as well as duration of the vowel and the consonants plus rhyme. Analysis of the acoustic measurements indicates that native speakers and non-native speakers used different perceptual strategies. While native English speakers employed a combination of pitch and duration cues, L2 learners employed pitch and intensity parameters without reliance on timing parameters. It was suggested that such a strategy was derived from the tonal background. Studies with Vietnamese learners showed that they "transfer L1 prosodic properties into their L2 English at both perception and production levels, especially at the initial stage of language acquisition" (Nguyen 2003: 149). Learners tend to make reference to acoustic cues that are actively involved in both L1 and L2 in L2 perception. At the same time, in L2 speech, they tend to avoid acoustic cues that are inactive in L1 speech.

Chen (1999) studied the acoustic characteristics of English produced by native speakers of Mandarin. Two groups of subjects were involved in the study, one group of Chinese learners of English and one group of native American English speakers. Each participant produced a group of 12 sentences, with emphasis on different words in the sentences. For example, in one sentence type 'I bought a cat there'. The stress can be put on the word, *I, bought, cat,* and *there*, respectively and results in four sentences. The F0, duration and intensity of the syllable nucleus of the stressed word were measured. Three analyses were conducted: one on the absolute measurements of the three acoustic values of the stressed syllable, one on the comparison between the stressed syllable and the rest of the unstressed syllables within the same sentence, and one on the comparison between the stressed syllable and the same syllable in a sentence where it is not stressed. The analysis of the absolute values showed that the two groups were not significantly different from each other in the intensity value. American males produced significantly longer stressed syllable than the Mandarin males in the study. The duration difference between the female participants was not significant. On the other hand, the absolute F0 values of the two groups were significantly different. Chinese learners of English produced stressed syllables with significantly higher F0 than the native English speakers. The analysis of the ratio between the stressed and unstressed syllables within the same sentence revealed a similar pattern. The two groups were similar in the intensity difference they used to differentiate stressed versus unstressed syllables. The duration difference used by Mandarin males was smaller than that used by American males. The F0 difference used by Mandarin females was higher than that used by American females. The analysis of the syllables across sentences showed the same pattern with the second analysis. Thus, the general conclusion drawn from this study was that Chinese learners of English resemble native English speakers in their use of intensity in marking stress. On the other hand, Chinese learners were less similar to the native speakers in their use of F0 and duration cues in signaling stress. They showed considerable difference from native English speakers in their use of the duration cue and the F0 cue. While they tend to underuse duration to mark stress, they obviously overused F0 as a cue for stress.

2.6 The Current Study

2.6.1 Rationale of the current study

The review of literatures in this section has shown that Mandarin Chinese and English belong to two different prosodic systems. Chinese is a typical tonal language (Gandour 1978) and English is a typical stress language (Beckman 1986). Furthermore, Section 2.4 above presented a detailed description of the acoustic dimensions of both English stress and Mandarin tones and discussed the different weight of F0, duration and intensity cues in both languages. The stress acquisition models discussed in 2.2 predict that the tonal background has a positive effect in L2 stress learning, which is not supported by studies with Chinese learners' and other tonal language learners' production and perception of stress (section 2.2). To explain this apparent contradiction, section 2.3 discusses three more general speech learning models. These speech learning models take into consideration the influence of L1 acoustic properties in L2 acquisition, which is a very important aspect that has been ignored by the studies reported in section 2.2. However, little research at the suprasegmental level has been conducted in the frameworks of the speech learning models. In section 2.5, the research contributing to the acoustic details of L2 stress perception was outlined. However, no studies have been conducted to investigate the use of acoustic cues by Chinese learners of English in stress perception. Thus, it is the intention of this study to focus on one group of L2 learners: Chinese learners of English. More specifically, the study intends to investigate the use of acoustic cues by Chinese learners of English in stress perception. The research questions for the current study and the related hypotheses are described in the next section

2.6.2 Research questions and hypotheses

This study is designed to apply methodologies used in L2 perception studies at segmental level and L1 stress perception studies to the investigation of the influence of different acoustic properties in L2 stress perception. In order to understand Chinese learners' problems in English stress perception, this paper focuses on the three major correlates of English stress (Fry 1958), fundamental frequency (F0), duration and amplitude, and aims to find out how Chinese speakers employ these three correlates in the perception of English stress. A perception experiment was designed to answer the following research questions:

1. Do native English speakers use the three correlates, F0, duration and intensity for the perception of English stress? If they do, what is the relative importance of each of the 3 cues in the perception of English stress for them?
2. Do Chinese learners of English use the same three cues? If they do, what is the relative importance of each of the 3 cues in the perception of English stress for them?
3. If both groups use the three correlates for stress perception, is the relative weight of each correlate the same across the two groups? More importantly, is pitch a more important cue for CE than for NE?

Participants' responses were collected and the results were analyzed in relation to the different weight of the three acoustic cues in stress perception and the difference between the Chinese learners of English and native speaker speakers in their reliance on the three cues. The following hypotheses were formulated accordingly.

1. Previous research on stress perception have shown that F0, duration and intensity are all acoustic correlates for native English speakers. It is thus hypothesized that the systematic manipulation of F0, duration and intensity will lead to systematic changes in the performance of stress judgments by native speakers of English.
2. For the Chinese learners of English, the systematic manipulation of F0, duration and intensity will also lead to systematic changes in the performance of CE stress judgments.
3. Based on the classic studies of Fry (1955, 1958, 1964), it is hypothesized that the effect of F0 manipulation is the most obvious one among the three acoustic correlates for native English speakers in perceiving stress, followed by duration, which, in turn, is followed by intensity.
4. For the Chinese learners of English, the order of importance of the three cues proposed for the native English speakers also holds true for the Chinese learners of English.
5. Previous research has shown the predominant influence of F0 in tone

perception in Chinese. Thus, it is hypothesized that the effect of F0 manipulation is more obvious for Chinese learners' stress judgments in the experiment than for native English speakers.

The description of the methods used to construct the perception experiment is presented in the next chapter, chapter 3.

Chapter Three

Methodology

The purpose of this research is to examine the difference between Chinese learners of English and native English speakers in their reliance on the three acoustic cues, F0, duration and intensity, in English lexical stress perception. To offer convincing answers to the above questions, the key component in the design of this perception experiment is the choice of stimuli. On the one hand, the stimuli have to show controlled variations on the three acoustic dimensions. On the other hand, the stimuli should not be susceptible to uncontrollable intervening factors for stress perception. In the present study, nonsense words were designed to allow for maximum control of the test material, and manipulation of F0, duration and intensity were realized on the nonsense words to allow for calibrated variations of the three cues. Three sections are devoted to the discussion of the creation of stimuli. Section 3.1 discusses two important questions: first, why nonsense words were more appropriate than real English words in the present study; second, why synthesized nonsense tokens were chosen over naturally-recorded nonsense tokens as stimuli for the experiment. Section 3.2 describes the construction of the three nonsense words, *latmab*, *nizdit* and *tetsep*. Section 3.3 outlines the manipulation of the three cues, F0, duration and intensity on the nonsense words to create the 375 nonsense tokens.

In Section 3.4, the two groups of participants, one group of native English speakers and one group of Chinese learners of English, who were involved in the experiment, are introduced. The participants listened to the aurally presented stimuli and judged the stress placement on each token. Section 3.5 outlines the detailed procedure of the perception test.

3.1 Motivation for Choosing Synthesized Nonsense Words as Stimuli

In choosing the appropriate stimuli for the experiment, there are two major concerns, whether to use real English words or to use nonsense words, and whether to use naturally recorded stimuli or manipulated stimuli. Section 3.1.1 discusses the advantage of nonsense words over real English words. Section 3.1.2 presents reasons why manipulated tokens rather than naturally recorded tokens were used as the stimuli.

3.1.1 Real words versus nonsense words

Real word pairs differing only in stress placement were widely adopted in previous research on stress perception (Lehiste 1959a, Lieberman 1960, Adams 1979, Sereno & Jongman 1995, Nguyen 2003). Some other studies have used nonsense words in the experiments with stress perception and production (Guion et al. 2004, 2005, Davis & Kelly 1997, Jusczyk et al. 1994, Vitechvich et al. 1997).

Nonsense words were chosen over real word-pairs because this study is a cross-linguistic study which involves listeners from two different language backgrounds. Native and L2 learners may differ in their familiarity with real English words, which may influence their stress judgments in an uncontrolled way. Researchers found that expectancy always plays a role in speech perception (Adams 1979, Adams & Munro 1978, Fry 1955, McEntee 1973 in Adams 1978). Native speakers may perceive stress where they anticipate stress. The kinaesthetic memories associated with a native speaker's own production of the syllable are always going to affect the way he/she perceives a familiar word. For non-native speakers, their expectancy for stress is related to their learning experience and such expectancy may not even be correct. Thus, to avoid the difference in familiarity and expectancy in perceiving stress on a real English word, it is more reliable to use nonsense words which are new to both native and non-native speakers.

The disadvantage of nonsense words is that they may sound 'foreign' to listeners. To reduce such foreignness, nonsense words were made to abide by English word and syllable structures. High frequency combinations of vowels and consonants were chosen to form the nonsense words. Section 3.2 discusses the constructions of these nonsense words in more detail.

3.1.2 Synthesized tokens versus recorded tokens

Section 3.1.1 argues that nonsense tokens are more appropriate for the experiment. The next question is whether we should present listeners with recorded tokens or present them with tokens of manipulated acoustic cues.

The advantage of recorded tokens is that they sound more natural in a perception test than synthesized stimuli. Their disadvantage is that it is difficult to record naturally produced stimuli with systematic variations of the acoustic cues. This study is designed to compare the influence of the three different acoustic cues on stress perception. It is hard to expect even a well-trained phonetician to manipulate different physical dimensions of stress independently in production. A demand like " please produce the word so that the first syllable is stressed in a way that the first syllable is longer than the second one but not higher in F0 or intensity" would be absurd for any native speaker, even a phonetician. Furthermore, in the comparison of different cues, it is more useful to use tokens with conflicting cue settings, for example, higher F0 on one syllable but higher intensity and longer duration on the other syllable. It is almost impossible to elicit the production of such tokens in natural production. This, on the other hand, is not difficult to create with speech synthesis methodology. Previous studies pointed out that "research using speech-synthesis techniques has probably been the most valuable in clarifying the nature of stress as they allow unambiguous reference to the dimensions of a specific sound stimulus and maximum control by the researcher over all the physical variables" (Crystal 1969: 46 in Adams 1979: 79). Using synthesis techniques in this study, each of the three acoustic cues can be manipulated independently and the manipulation can be controlled with predefined parameters. As Fry (1958, 134) commented: "The essence of this method is that the properties of the speech signals may be closely controlled. This is generally not possible in the case of live speech and only partially so in recorded speech, so that the most satisfactory method is to synthesize the required speech sounds in some way that will afford the necessary control over all the variables of the speech."

The disadvantage of using manipulated tokens rather than recorded tokens is that it may sound unnatural. In order to reduce the unnaturalness of such tokens, a strategy used by many previous researchers is to change the acoustic properties on the *recorded* tokens in such a manner as to achieve a combination of maximum naturalness and accurate control of the three parameters (Fry 1955, 1958, Sluijter & van Heuven 1996). This strategy was used for the present study. Naturally produced tokens were used as the basis for cue manipulation. Section 3.3 introduces the synthesis process in detail.

3.2 Creation of Nonsense Words

This section introduces the creation of nonsense words for the experiment. On the one hand, the nonsense words should be designed to be as natural as possible, which means that they should resemble real English words. These words are introduced in Section 3.2.1. On the other hand, the nonsense words constructed should not remind listeners of any real English words to avoid the influence of the real English words. This is realized by a Similarity Test described in Section 3.2.2. Section 3.2.3 describes the recording of the three nonsense words that were chosen.

3.2.1 Design of nonsense words

In the design of the nonsense words, the following factors were considered: word and syllable structure, the vowels to be used, and the consonant vowel combination.

The nonsense words were designed to be of two syllables and there are two reasons for this. First, in English, most real words with an alternation of stress-placement are disyllabic. Secondly, in order to avoid distraction from the focus on stress perception, the nonsense words should not be too long. In addition to word structure, syllable structure may also influence the perception of stress. Typological evidence of stress placement in the metrical system shows that heavy syllables tend to attract stress, although the definition of a heavy syllable may vary from language to language. English is a quantity sensitive language, and syllable structure has an effect on stress placement in real English words. Such influence is also observed for nonsense words. Guion et al. (2003) suggested that syllable structure has influence on main stress placement in nonsense words for English speakers. Heavy syllables were more likely to be stressed than light syllables. If the two syllables of a nonsense word differ in syllable structure, native English participants of the study may judge stress placement based on the difference in syllable weight instead of the acoustic cues on the two syllables. Furthermore, if the light syllable in a nonsense word is manipulated to have higher F0, greater intensity and longer duration, but the phonological structure prefers stress on the heavy syllable, then the stimuli may sound unnatural for a native speaker. In order to avoid the uncontrolled effect of syllable structure on stress perception, the nonsense words were designed to have the same syllable structure on both syllables, CVC.CVC, where C is the a consonant and V is a vowel and the dot is used to indicate syllable boundary.

Next, we consider the choice of vowels to be used in the nonsense words. In a real word pair such as SUBject, subJECT, the vowels in the first syllable and second syllable are different. This is problematic as previous studies found that native listeners may have established certain intrinsic relative intensity for each vowel (Morton & Jassen 1965, Lehiste & Peterson 1959). In perception, a certain "correction factor" can be applied to the speech signal (Lehiste & Peterson 1959) to calculate the "surface value". For example, the measurement in Lehiste and Peterson's study shows that the intrinsic amplitude of /ɪ/ is 78.1 dB, and /aɪ/ is 80.2 dB. However, in a word like INcline with stress on the first syllable, the amplitude difference between /ɪ/ and /aɪ/ is zero. It is suggested that listeners take into consideration the intrinsic amplitude and apply a "correction factor" in perception. When a vowel is produced with amplitude higher than its intrinsic amplitude, it may be perceived as stressed, although the actual amplitude may not be higher than the unstressed vowel in the same word. To avoid the influence of the intrinsic intensity of different vowels, the same vowel was used for the two syllables of a nonsense word. In the present study, three short vowels were chosen, /ɛ/, /ɪ/, and /æ/. Thus, the three nonsense words designed were: CɛC.CɛC, CɪC.CɪC, and CæC.CæC. Long vowels and diphthongs were avoided because they have a greater chance to be stressed (Guion et al. 2003: 422), which means that to un-stress them would make the stimuli less natural. A short vowel, on the other hand, is natural either in the stressed syllable or the unstressed syllable.

In the choice of appropriate consonants, two criteria were used. First and foremost, the nonsense words should abide by English phonotactics. In other words, the stimuli should contain permissible consonant-vowel sequences in English. In addition, to make the nonsense words natural-sounding, the most 'popular' consonant-vowel combination in the English language is preferable. In the present design, 'popularity' is operationalized as a high frequency of a certain consonant-vowel combination. Jusczyk et al. (1994) defined the frequency of a certain CVC combination by two measures: 1. How often a given segment occurs in a position within a word. 2. Biphone frequency, which is the phoneme-to-phoneme co-occurrence probability[1]. Among the high frequency CVC combinations provided by Jusczyk et al. (1994), 24 were chosen for the present study. Among them eight were CɛC syllables, eight were CɪC syllables and eight were CæC syllables. **Table 3.1** lists the 24 syllables chosen to construct nonsense words.

Table 3.1 Twenty-four high-frequency syllables candidates

C ɛ C	tɛt	ɹɛp	sɛp	mɛk	kɛs	sɛs	fɛs	dɛs
C ɪ C	ɹɪz	dɪt	bɪt	mɪp	dɪv	nɪz	wɪz	fɪk
CæC	læt	mæb	mæz	hæs	fæs	dæz	kæk	mæv

The actual stimuli used in the experiment were only a subset of the nonsense words that can be constructed with the syllable candidates found in Table 3.1. The process of how nonsense words were constructed and why only a subset

[1] For a more detailed description of how these two measures are used to calculate the frequency of a combination, please refer to p 633 of Jusczyk et al. (1994).

was chosen are discussed in section 3.2.2.

Summary of section 3.2.1

Section 3.2.1 discusses the design of the nonsense words to be used in the experiment. The nonsense words were designed to be disyllabic, that is, CVC.CVC. The two vowels in the two syllables of a nonsense word were identical in order to reduce the difference in intrinsic pitch of different vowels. Three short vowels were used for the nonsense words, CɛC.CɛC, CɪC.CɪC and CæC.CæC. Twenty-four high-frequency CVC combinations (see **Table 3.1**) were chosen from previous literature as the candidates for the construction of the nonsense words. The concatenation of these syllables to form nonsense word candidates and the choice of nonsense words are discussed in the next section.

3.2.2 Construction of nonsense words

With the eight syllables of each vowel type, many different nonsense words can be constructed by concatenating two syllables with the same vowel, e.g. sɛp.kɛs. However, not all combinations were suitable for the perception test. Empirical studies have shown that if a nonsense word is similar to a real word, the stress pattern of the real English word is predictive of the stress pattern of the nonsense word (Baker & Smith 1976, Guion et al. 2003). Thus, in selecting the most appropriate nonsense words, a word similarity test was used to make sure that the selected nonsense words did not consistently remind participants of any real English words.

The nonsense words were constructed and selected in three steps. Firstly, the twenty-four isolated CVC syllable candidates (see Table 3.1) were recorded and then they were concatenated to make 30 nonsense word candidates, with ten candidates for each vowel type. In the third step, these bisyllabic candidates were played to native English speakers as well as Chinese learners of English to detect possible influence of similar real English words. Three tokens that had the least similarity to real English words were chosen. The detailed steps of the construction and selection of the nonsense words are introduced below.

In the first step, the 24 CVC syllable candidates listed in **Table 3.1** were recorded in a sound-treated room in the Department of Linguistics, University of Victoria. The microphone used in the recording process was an M-Audio LUNA large diaphragm condenser. The recording was stored directly on a PC with a sampling rate of 44kHz. All further recordings were conducted in the same environment if not otherwise indicated. The target syllables were embedded in a carrier sentence "She says CVC twice." The fricative /s/ in 'says' before the target syllable and the stop /t/ in 'twice' after the target syllable were chosen to make the segmentation easier. The sentences were presented in English orthography but the target syllable was transcribed with phonetic symbols. An experienced phonetician, who is a native speaker of English, read the sentences and was requested to stress the target syllable in the sentence. The eight syllables with /ɛ/ were read first, followed by /ɪ/, and then /æ/. There was a two-minute pause between every two lists. After all items were recorded, there was a five-minute rest and the whole process was repeated again. The order of the items was reversed in the second recording to eliminate possible order effect.

In the second step, these recorded syllables were concatenated to make the nonsense candidates. Before concatenation, the recorded syllables were normalized to -3dB using the software Audacity. For each vowel category, ten nonsense words were constructed. Five of the ten words had the form $\sigma_1\sigma_2$ and the other five the form $\sigma_1\sigma_2$. The first and second syllables of each candidate were always different. As there are three vowel types, a total of 30 nonsense word candidates were constructed. Table 3.2 listed all the nonsense word candidates.

Table 3.2 Thirty nonsense word candidates with ten for each vowel type

	vowel ɛ	vowel ɪ	vowel æ		vowel ɛ	vowel ɪ	vowel æ
σ1σ2	tɛt.sɛp	wɪz.dɪt	mæz.læt	σ2σ1	sɛp.tɛt[2]	dɪt.wɪz	læt.mæz
	tɛt.rɛp	rɪz.bɪt	læt.mæb		rɛp.tɛt	bɪt.rɪz	mæb.læt
	sɛs.rɛp	rɪz.mɪp	hæs.kæk		rɛp.sɛs	mɪp.rɪz	kæk.hæs
	sɛs.mɛk	rɪz.dɪv	fæs.læt		mɛk.sɛs	dɪv.rɪz	læt.fæs
	sɛp.mɛk	nɪz.dɪt	dæz.læt		mɛk.sɛp	dɪt.nɪz	læt.dæz

These candidates were then fed to the third step of the similarity test, and the least similar words were chosen for each vowel category. As was discussed at the beginning of this section, previous studies have shown that stress patterns of real words may greatly influence the perception of stress placement on similar nonsense words (Baker & Smith 1976, Guion et al 2003). "(T)he more similar a pair of words appear, the more likely it is that the nonsense word will receive the

[2] 'septet' is a real English word. It is not excluded from the similarity test and most native speakers identified the word as a real word

same stress pattern as its model" (Baker & Smith 1976: 18). If the constructed nonsense word resembles a certain real English word, it is very possible that the participants' perception will be dominated by the stress pattern of the model instead of the acoustic cues imposed on the nonsense word in perception of stress. Although it was observed by Baker and Smith that longer words are more susceptible to the influence of similar words, the effect on disyllable words still exists. Following the methodology of Baker and Smith (1976), a similarity test was designed to eliminate the contaminating effect of real English words. The thirty candidates were played to both native English speakers and Chinese learners of English to see if they reminded them of any real English words. For any given candidate, if no more than two listeners came up with the same English word, then the candidate qualified as a token that did not have a strong similarity to any real English word.

Participants. Eleven native English speakers, seven female and four male, and ten Chinese learners of English, seven female and three male, participated in the word similarity test. These participants were different from those who participated in the perception test.

Materials. The thirty candidates listed in Table 3.2 were presented aurally as a word with two stressed syllables, since it has been proposed by Guion et al (2003) that presenting the tokens in written form may induce an influence from English orthography.

Procedure. One block of thirty tokens was presented in two orders, one the reverse of the other; half of the subjects received the tokens in one order, the other half of the subjects received the words in the reverse order. In each trial, a word was played three times in a row with a 500ms interval between repetitions. After the three repetitions, participants were given 10 seconds to respond orally. There was a ding sound at the end of 10 seconds to remind the participants that the trial was over and then a chime leading to the next trial. Participants were given verbal instructions to listen to the recording and produce any real English words that were similar in *sound*. They were reminded that they shouldn't worry about the type (common words or proper nouns) or the number of words that they came up with. The recordings were played through high-quality head-phones and participants' verbal responses were recorded on a laptop computer. If the researcher was not clear about a listeners' response, the process was paused at the end of each trial to clarify.

After the similarity test, native English participants were invited to judge the naturalness of the nonsense word candidates. The participants ranked the naturalness of each candidate from one to three, one being the least natural and 3 being the most natural. The result of this was used as an additional criterion in the selection of test words. If more than one candidate in each vowel category satisfied the requirement of the similarity test, then the most natural one was chosen.

Results. Generally speaking, most native English speakers were able to come up with one or more real English words when they heard the nonsense candidates.

Chinese learners of English were far less sensitive to the nonsense tokens. Most Chinese learners of English could not think of any real English words when they heard the nonsense tokens. Furthermore, Chinese participants with higher English proficiency didn't demonstrate better performance than participants with lower proficiency. When the results from both groups were pooled together, it could be seen that for each nonsense word candidate, participants produced a variety of different real English words. The results were transcribed by the researcher (see Appendix A. Similarity Test Results). Table 3.3 below provides a brief summary. Only real English words which were produced by more than two participants are listed.

As can be seen from the following table, altogether six candidates satisfied the requirement of the similarity test, *tɛtsɛp, nɪzdɪt, dɪtnɪz, mɪprɪz, lætmæb, mæblæt*. There is an asterisk beside each of these candidates in the following table. As only one word from each vowel category is needed for the perception test, further filtering was needed. The candidates *mɪprɪz* and *mæblæt* were excluded because they can be possibly re-syllabified as *mɪ.prɪz* and *mæ.blæt*. Between the two candidates *nɪzdɪt* and *dɪtnɪz,* the first one has a higher naturalness score (see Appendix A). Finally, one candidate from each vowel category was chosen. They are *tɛt.sɛp, nɪz.dɪt and læt.mæb*.

Table 3.3 Similarity test results for the nonsense word candidates.
Words produced by all participants in the Word Similarity Task. Numbers in parentheses indicate frequency of the word produced by participants. Words beginning in a capital letter are proper nouns.

	ɛ		**ɪ**		**æ**	
1	rɛpkɛs	repeat/ repetition(4)	nɪzdɪt	*	mæzlæt	Mazda(3)
2	sɛpmɛk	septic(3), September(4)	rɪzdɪv	divide/ division(5)	lætmæb	*
3	tɛtsɛp	*	rɪzmɪp	rhythm/ rhythmic(3)	hæskæk	Pascal(3), pass/ passed(4)
4	tɛtrɛp	repeat/ repetition(4), Tetris(4)	rɪzbɪt	bit(4)	fæslæt	fast/faster(8)
5	sɛsmɛk	seismic(4)	wɪzdɪt	wise/ wisdom(4), wizard(3), whiz(3)	dæzlæt	dazzle/ dazzling(4)
6	kɛsrɛp	repetition(3), ketchup(3), kiss/kissed(4)	dɪtnɪz	*	lætmæz	lats/ latissimusd orsi(4)
7	mɛksɛp	except(4), accept(3)	dɪvrɪz	divide/ division(8)	mæblæt	*
8	sɛptɛt	septet(3), September(3)	mɪprɪz	*	kækhæs	cactus(3), cat(6)
9	rɛptɛt	repeat/ repetition(5)	bɪtrɪz	bit(3)	lætfæs	fast/faster(4)
10	mɛksɛs	\`access/ ac\`cess (4)	dɪtwɪz	wizard(3)	lætdæz	late/later(6)

3.2.3 Recording of nonsense words

The three nonsense test words were recorded by the same native English-speaking phonetician in preparation for the manipulation process.

The nonsense words were recorded in two stress patterns, one with stress on the first syllable and the other with stress on the second syllable. The target words were recorded in isolation. Although it has been common to embed a target word in a carrier sentence to elicit correct stress patterns, there are problems with this elicitation method. Researchers have found that the stress of the immediately preceding word in a sentence will influence the placement of stress in the target word (Kelly & Bock 1988, Guion et al. 2003). Preliminary recording in this experiment using carrier sentences revealed another problem. The prosodic content of the sentence interfered with the stress pattern of the target word. In other words, the duration, intensity and F0 information on the target words is not a single result of the stress pattern on the word but a joint result of sentence prosody and word stress. It was also pointed out in previous literature that "(t)hese parameters, obviously, can be more readily investigated in the word in citation form than in the context of the sentence where various features vie for salience"(Stetson 1951 in Adam & Munro 1978). Thus, the three nonsense words were recorded in isolation.

A second problem in recording was that the speaker found it sometimes difficult to produce the desired stress pattern on the nonsense words. This was solved by using a block elicitation method. The nonsense words were embedded in a block of real English words with the desired stress pattern. The purpose of the real words was to create a natural environment for the production of the stress pattern on the nonsense words. Sixteen real English words which resemble the nonsense words in a number of ways were included in the lists. Firstly, only disyllabic real words were chosen. Secondly, the real English words should match nonsense words in terms of the vowel, the syllable structure and the segmental content of the stressed syllable. If the match of the stressed syllable is not possible, an effort was made to match the unstressed syllable. The nonsense words were embedded in a list of real English words to help the speaker to maintain a stress pattern. In two lists, the words included all have stress on the first syllable and in two other lists, the words all have stress on the second syllable. Table 3.4 lists all the words recorded. Bolded are the target nonsense words. The stressed syllables are capitalized.

Table 3.4 Recorded word lists embedded with nonsense words

Stress on the 1st σ	Set 1	REDhead	**TETsep**	MINit	**NIZdit**	LANguage	**LATmab**	LISten
	Set 2	Tenant	**TETsep**	PUBlic	**NIZdit**	LANDlord	**LATmab**	NORmal
Stress on the 2nd σ	Set 1	exPECT	**tetSEP**	adMIT	**nizDIT**	deMAND	**latMAB**	comMIT
	Set 2	acCEPT	**tetSEP**	comMIT	**nizDIT**	eXAM	**latMAB**	reMAIN

The same phonetician recorded the word list. The instruction given for the recording was to stress the capitalized syllable while NOT reducing the vowel in the unstressed syllable. Each word list was recorded twice. Thus, a total of four repetitions for either stress pattern of a nonsense word form were recorded and this ensured the production of at least one clear, accurate and natural exemplar of each word. A native speaker with no knowledge of the purpose of the study was invited to listen to the four tokens of each of the three words and choose the most natural exemplar of the word. Three tokens were chosen, one token of the word tetSEP, one token of the word nizDIT, both with stress on the second syllable, and one token of LATmab, with stress on the first syllable. These three tokens were used for the manipulation process.

In addition to the nonsense words, some real English words were recorded. These real words were used in the perception tests as fillers. Among the recorded real English words, four were homographs on which stress pattern is related with the lexical class. The four real words chosen are listed in Table 3.5

Table 3.5 Four recorded real English words with two stress patterns

	Words	Stress on the 1st syllable	Stress on the 2nd syllable
	implant	*n.* something implanted, esp. implanted tissue or device	*v.* to set in firmly or establish securely
	increase	*n.* the amount or rate by which something is increased	*v.* to become larger or greater
	export	*n.* something exported	*v.* to send or transport (a commodity) abroad, especially for trade or sale
	concrete	*n.* a heavy-duty building material	*v.* make or become solid

In addition to these four homographs, nineteen fixed-stress English words were also recorded. The purpose of these nineteen real English words was also to help elicit the intended stress pattern on the homographs. The following table shows the lists of the real English words (see Table 3.6).

Table 3.6 Recorded word lists embedded with real English words

1st σ	Set 1	IMpact	**IMplant**	INcome	**INcrease**	EXpert	**EXport**	CONtact	**CONcrete**	CONflict
	Set 2	INfluence	**IMplant**	INterest	**INcrease**	EXtra	**EXport**	CONcept	**CONcrete**	CONflict
2nd σ	Set 1	imPOSE	**imPLANT**	inCLUDE	**inCREASE**	exPLAIN	**exPORT**	conCERN	**conCRETE**	conTAIN
	Set 2	imPROVE	**imPLANT**	inTEND	**inCREASE**	exPRESS	**exPORT**	conTROL	**conCRETE**	conCERNED

3.3 Manipulation on Nonsense Words

This section describes the systematic manipulation of the three acoustic cues, F0, duration and intensity, on the three nonsense words to create the test stimuli. According to the previous research on English stress by Fry (1955, 1958), the change of F0, duration and intensity to signal stress is mainly on the vocalic portion of a syllable. Thus, in this study, the manipulations of the three cues were applied to the syllable nuclei of the disyllabic words, henceforth V1 and V2. For each of the three nonsense words, the F0, duration and intensity of the two vowels were first averaged to create the base forms by a normalization script and then the predefined difference between the two syllables in terms of the three cues was realized on the nonsense words by a manipulation script.

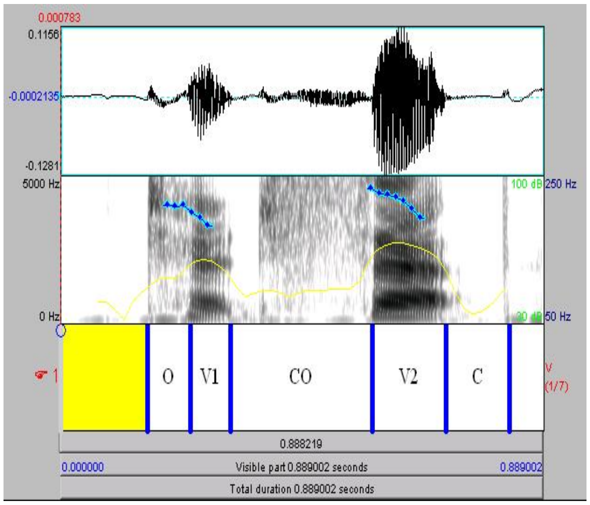
Section 3.3.1 describes the preparation of the base form for manipulation, and Section 3.3.2 discusses the settings of the three parameters for manipulation. Section 3.3.3 outlines the manipulation process of the three nonsense words.

3.3.1 Normalization

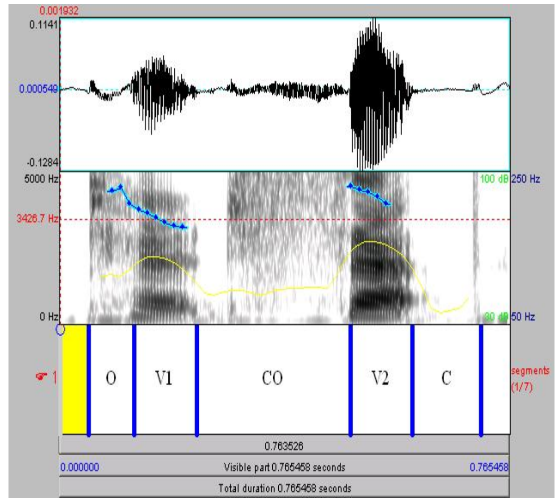
The manipulation of the three cues all depends on the baseline form. The base form of a nonsense token was where its two syllables were normalized to have the same F0 contour, vowel duration, and intensity contour. In order to achieve this, the selected exemplar of the nonsense words was taken, and the average duration, pitch and intensity of the two vowels were calculated and applied to both vowels.

Three Praat scripts were written for this purpose, one for each parameter (see Appendix B, Normalization and Manipulation Scripts). The three Praat scripts were run in the order of Duration Normalization Script, F0 Normalization Script and Intensity Normalization Script. The steps taken in creating the base form are introduced in detail below.

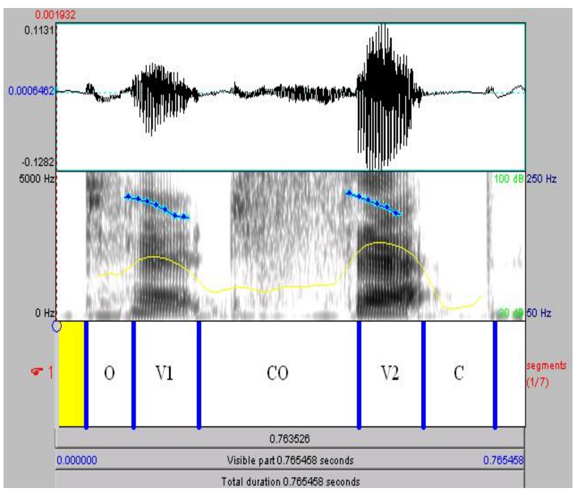
The three nonsense words were first annotated by textgrid files in Praat. Each word was segmented into 5 intervals. The first interval was the onset of the first syllable (O); the second interval was the first vowel (V1); the third interval was the coda of the first syllable together with the onset of the second syllable (CO); the fourth interval was the second vowel (V2) and the fifth interval was the coda of the second syllable (C). The segmentation was based on both the spectrogram and the waveform. The boundary between the onset/coda consonants and the vowels were quite clear because the consonants included in the stimuli were either fricatives or stops, both of which had clear manifestation on the spectrogram and waveform. An example of the word tetSEP before normalization was given in the first graph in Figure 3.1 below. The word had stress on the second syllable with the second vowel having longer duration, higher pitch and higher intensity compared to the first vowel.



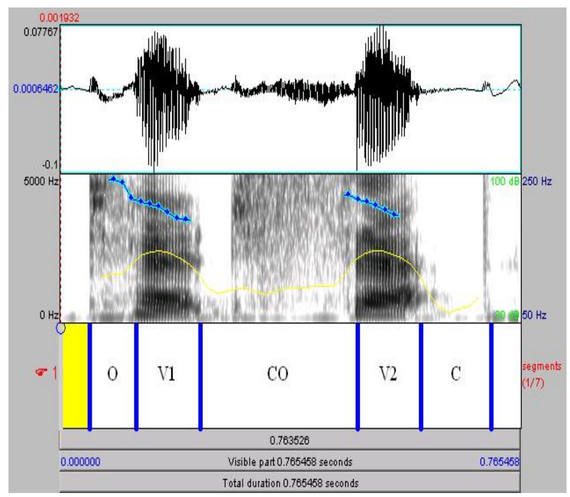
1. The original sound file



2. Duration averaged sound file (V1=V2)



3. Pitch averaged sound file



4. Intensity averaged sound file

Figure 3.1 Normalization of F0 contour, duration and intensity contour of the nonsense words. The blue line on the spectrogram represents the pitch contour of the portion of the sound. The light yellow line on the spectrogram represents the intensity contour of the sound. Below the spectrogram is the annotation tier. The space between the two blue vertical boundaries of V1 indicates the duration of V1, and the same is true for V2.

The Duration Normalization Script was used to measure the duration of the first vowel and the second vowel. The mean of the two vowels was then calculated. The duration of the first or the second vowel was either lengthened or shortened according to the average duration. The other three intervals were kept at their original length. The result of this step is shown by the second graph of Figure 3.1. The V1 interval and V2 interval were of the same length at this stage.

The result of the first step was fed to the next step of pitch manipulation. The new sound file was segmented again into the same five intervals, with the only changing the duration of the two vowels. The first vowel and the second vowel now had the same duration. The Pitch Average Script took down the pitch contour of the five intervals. From the beginning of each interval, a pitch value was taken every 0.01 seconds until the end of the interval. For the manipulation of the pitch of the two vowels, the average pitch value was calculated and then used to replace the original pitch tier. The average was achieved by taking the first pitch value for vowel one and the first value for vowel two and taking the mean of the two. The same was done for each pair of pitch value for V1 and V2 until the end of the vowels. The figure below indicates how this was achieved. Suppose the two vowels were averaged at 0.15ms. A pitch value was taken every 0.01 second and there were a total of 15 pitch points[3]. The dark blue line represents the pitch contour of vowel one (V1), and each circle along the line indicates the pitch value at that time point. Similarly, the light blue line represents the pitch contour of vowel two (V2) and each square indicates the pitch value of V2 at that time point. The dark line between the two pitch contours is a calculated pitch contour. Each triangle was calculated by taking the mean of

[3] In manipulation, the vowels are much longer than 0.15ms and thus a lot more pitch points were sampled.

the corresponding square and circle value at the time point. This dark line was then used to replace the pitch contours of the first and second vowel[4].

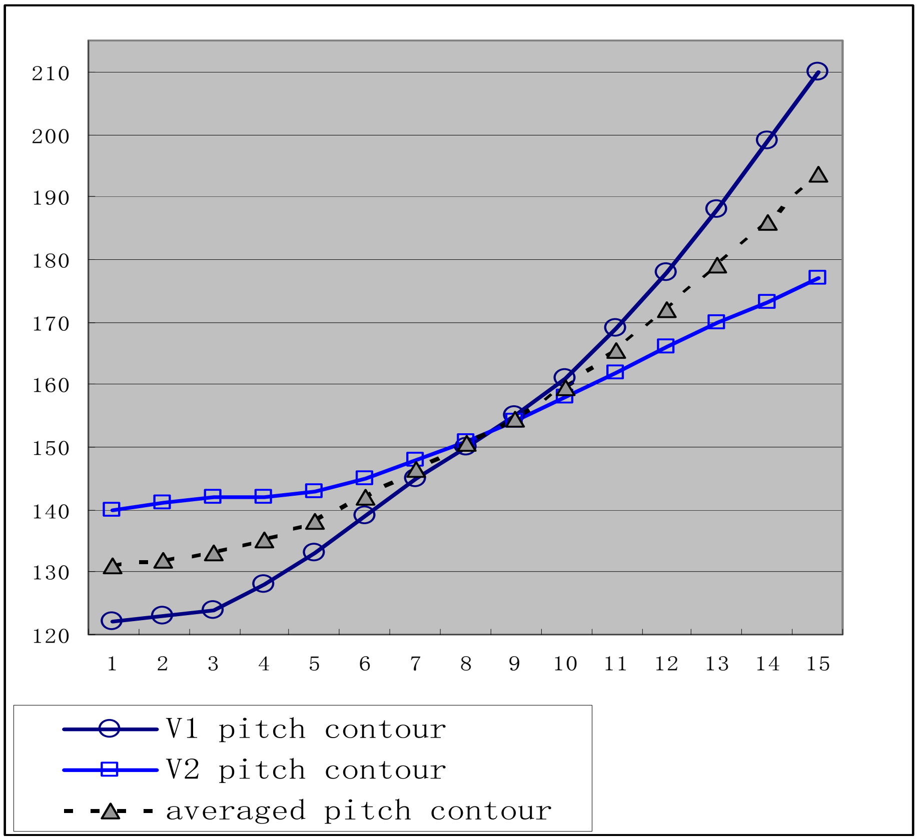


Figure 3.2 Creating the normalized pitch contour by taking the average value of V1 and V2 pitch values at each time point. The X-axis represents 15 time points. The y-axis represents the pitch value (Hz).

Other than the vowels in the nonsense words, the sonorant consonants *l* and *m* in *latmab*, *n* in *nizdit*, and the fricative z in *nizdit* and also the ending part of the aspirated *t* in *tetsep* were also found to carry pitch contours. These pitch contours were preserved in order to keep the tokens as faithful as possible to the original recorded tokens. This is not ideal as there may be disjunctions between the

[4] The pitch replacement is not exactly the same as the calculation. Minor differences were observed between the two vowels in their pitch contours. The difference is no more than 2 Hz apart. This is possibly due to the automatic pitch smoothing function of Praat.

normalized vowel contours and the original neighboring consonant contours. To alleviate the problem, Praat applies an automatic correction function which smooths the pitch contour at the junction.

The third graph in Figure 3.1 shows that the pitch contours for V1 and V2 were the same at this stage.

The last parameter that was normalized was intensity. The manipulation of the intensity contour was similar to that of the pitch contour, but the technical details were different. The result of the last two steps, a token with the same duration and pitch contour of V1 and V2, was fed into this step. The highest intensity level was identified within V1 and V2, respectively. The Intensity Average Script brought the highest intensity of the two vowels to the same level by substituting the original V1 intensity and V2 intensity with the product of V1 intensity multiplying by V2 intensity. Then the intensity of the token was scaled to the average intensity level of the original sound file ((V1+V2)/2). The intensity contour of the whole token can thus be kept unchanged in the process. As can be seen in the last spectrogram in Figure 3.1 , the two vowels were the same not only in duration and pitch contour, but also in intensity contour.

All three nonsense words were normalized this way to have the same F0 contour, intensity contour and same duration on the two syllables. Efforts were made to keep the original shape of the F0 and intensity contour of the vowels, and the two vowels of a token were not as long as the stressed vowel and not as short as the unstressed vowel. We believe it is more important to keep the properties of the vocalic portion of the tokens, even if it means that we had some discrepancies between the vowel contours and the consonant contours. These normalized tokens were used as the base forms in further manipulation. The actual parameters used in manipulation is introduced next.

3.3.2 Parameter settings

To prevent having too many test tokens and making the perception test too long, the manipulation of each parameter was limited to five levels. The range of variation to be covered and the size of each step change were different for the three different parameters, and they are discussed in the order of F0, duration, and intensity below.

Fundamental Frequency. The values to be used for the manipulation of F0 in the present study are based on previous studies involving this parameter.

In Fry's (1958) study with native speakers of English, it was found that a change of 90 Hz in the stimuli produced a rather unnatural effect. In Morton and Jassen's (1965) study, a maximum change of 70 Hz was used in synthesis. Both studies found an "all-or-none" effect of F0. In other words, the magnitude of F0 change does not have a significant influence on stress perception. A one-step change up, which is as little as 5 Hz in Fry's study, is already good enough for syllable to be perceived as stressed. "Increase in the size of the frequency step appears to produce no marked trend in the results." (Fry 1958, 144) Similarly, Morton and Jassen (1965) found that a 58% change in F0 was no more effective than a change of 25% in marking the stress. Thus, in the present study, the difference between the stressed and unstressed vowels was reduced to 50 Hz between the two syllables. At the first level, the first vowel was reduced by 25 Hz, and the second vowel was increased by 25 Hz, so that the difference between the two vowels was -50 Hz. At the second level, V1 was reduced by 12.5 Hz and the second syllable was increased by 12.5 Hz. The difference between V1 and V2 was -25Hz on level two. At the third level, both vowels were kept at the baseline F0. The establishment of the baseline is introduced in section 3.3.2. Level four and five were the mirror image of level two and level one, respectively. At the fourth level, while the first vowel was increased by 12.5 Hz, V2 was reduced by 12.5Hz. And finally at the fifth level, the first vowel was increased by 25Hz and the second vowel was reduced by 25Hz. The differences between the two vowels at level four and level five were 25Hz and 50Hz. In this way, the average F0 of the two vowels at each level was always kept the same, although the difference between the two syllables varied from -50 Hz to 50Hz. In this way, the stimuli did not deviate too much from the original recording and were thus kept as natural as possible.

Table 3.7 below displays the five levels of F0 manipulation. The first column indicates the difference between the two vowels at each level and Columns Two and Three list the values of V1 and V2 at each level.

Table 3.7 Five levels of F0 manipulation

F0 difference b/t V1 and V2 (Hz)	V1 F0	V2 F0
Level 1: -50	baseline - 25	baseline + 25
Level 2: -25	baseline -12.5	baseline + 12.5
Level 3: 0	baseline	baseline
Level 4: 25	baseline + 12.5	baseline-12.5
Level 5: 50	baseline +25	baseline-25

Duration. The range of duration change to be used depends mainly on the ratio between the stressed and unstressed vowel. In a production study, Crystal and House (1987) studied segmental duration in different stress positions and summarized the length of vowels in stressed and unstressed positions as well as positions where stress is unspecified. Their summary is given below in Table 3.8

Table 3.8 Duration of stressed and unstressed vowels in English (taken from Crystal & House 1987, 1576). The duration of vowels in five stress conditions were measured: U = unspecified; -S = unstressed; S2 = secondary stress; S1 = primary stress; +S = primary or secondary stress. The mean durations are listed under the column Mn.

TABLE II. Mean durations (Mn) and standard deviations (s.d.), in milliseconds, for various vowel categories in a variety of stress conditions. *N* — number of tokens. The stress conditions are: U = unspecified; S = unstressed; S_2 = secondary stress; S_1 = primary stress; +S = primary or secondary stress. The number of vocalic types in each category is indicated parenthetically.

	Stress conditions														
	U			-S			S_2			S_1			+S		
Category	*N*	Mn	s.d.	*N*	Mn	s.d.	*N*	Mn	s.d.	*N*	Mn	s.d.	*N*	Mn	s.d.
Vowels (18)	3850	102	59	1633	61	27	229	95	46	1988	137	58	2217	133	58
Long vowels (7)	1328	130	54	237	78	24	78	104	40	1013	144	52	1091	141	52
Short vowels (4)	1443	72	34	842	56	22	101	76	32	500	97	36	601	93	37
Diphthongs (3)	369	176	62	49	113	31	34	143	57	286	190	58	320	185	60

It can be noticed from the cells highlighted above that the duration difference for different kinds of vowels is around 70-80ms, except in the case of the short vowels, which is only around 40ms. However, what we were interested in was not the exact duration difference but the ratio between stressed and unstressed vowels. As can be calculated with the values in the above table, the ratio between stressed and unstressed vowels in general is $137/61 \approx 2.2$. For long vowels, it is $144/78 \approx 1.8$, for short vowels, $97/56 \approx 1.7$ and for diphthongs ≈ 1.7.

Field (2005) used a ratio of 1.5 between the stressed and unstressed syllables in his perception study. In Beristein's (1979: 5) perception study, one syllable is set to the baseline of 100 ms and the other syllable ranges to 70, 100, 120, 140, 160, 200. In other words, the ratios of 1.43, 1, 0.83, 0.71, 0.625 and 0.5 were used in Beristein's study. It was found that the longer the syllable is, the more possible that the syllable is perceived to be stressed.

The review of previous literature led to the decision of a range of 0.5 to 1.5 for duration manipulation in this study. The 5 levels of manipulation were evenly distributed between 0.5 and 1.5. Thus, the ratios between V1/V2 were 0.50, 0.75, 1, 1.25, 1.5. The baseline duration of the vowel was used when the ratio was 1. For example, if the baseline duration was 100 ms, for the level where the ratio of the two vowels is 1, both vowels were 100ms in duration. When the ratio was 1.5, the first vowel was increased to 6/5 times of the baseline, which was 120 ms and the second vowel was reduced to 4/5 times of the baseline, which was 80 ms. The ratio between V1 and V2 was 120/80=1.5. At the same time, the amount added to V1 and deducted from V2 was the same. When the ratio was 1.25, V1 was increase to 10/9 times of the baseline and V2 was reduced to 8/9 times of the baseline. When the ratio was 0.75, on the other hand, V1 was reduced to 6/7 of the baseline and V2 was increased to 8/7 of the baseline. Similarly, when the ratio was set at 0.5, V1 was 2/3 baseline and V2 was 4/3 baseline. Table 3.9 shows the manipulation of vowel duration at five different levels.

Table 3.9 Five levels of duration manipulation

Ratio of V1/V2	V1 duration	V2 duration
level 1: 0.5	baseline * 2/3	baseline * 4/3
level 2: 075	baseline * 6/7	baseline * 8/7
level 3: 1	baseline * 1	baseline * 1
level 4: 1.25	baseline * 10/9	baseline * 8/9
level 5: 1.5	baseline * 6/5	baseline * 4/5

Intensity. Few studies have explored the difference between stressed and unstressed syllables in terms of intensity. Although many studies indicate that intensity is different for stressed and unstressed vowels, the exact difference is hard to define. More recent studies have shown that stress may relate more closely to intensity differences located above 0.5 kHz instead of global intensity (Sluijter & van Heuven 1996). In this study, however, only global intensity was examined.

Three previous studies which manipulated intensity were used as a reference in setting intensity ranges and levels to be used. Fry (1955, 1958) used a maximum 10 dB difference between stressed and unstressed syllables, and five steps were used, ie. +10db, +5db, 0db, -5db and -10db for the first syllable, and the second syllable was kept constant for amplitude. Morton and Jassen (1965) used a maximum 9dB difference and 7 steps, namely, the first syllable was raised by 9, by 6 and by 3 dB, or the second syllable was reduced by 9, 6 or 3dB, and there was the step where the two syllables were kept the same. More recently, Sluijter et al. (1997) used a similar approach. A maximum amplitude difference of 9 dB and 7 steps was employed in their research.

Previous studies seem to suggest that 9-10dB is an appropriate amplitude difference range for stressed and unstressed syllables. Thus, 9dB was chosen to be the range of amplitude variation in the present study. The five levels of intensity difference between the two syllables were set to -9, -4.5, 0, 4.5 and 9dB. At the first level, the first syllable was reduced by 4.5dB and the second syllable was increased by 4.5dB. At the second level, the first syllable was reduced by 2.25dB and the second syllable was increased by the same amount. At the third level, the two syllables had the same intensity, the baseline. At Level Four and Level Five, the first syllable was increased from the baseline, and the second syllable was reduced from the baseline. Table 3.10 below lists the fives levels of intensity manipulation.

Table 3.10 Five levels of intensity manipulation

Intensity Difference b/t V1 and V2	V1 Intensity	V2 Intensity
level 1: -9 dB	baseline – 4.5	baseline + 4.5
level 2: -4.5 dB	baseline – 2.25	baseline + 2.25
level 3: 0 dB	baseline	baseline
level 4: 4.5 dB	baseline +2.25	baseline – 2.25
level 5: 9 dB	baseline + 4.5	baseline – 4.5

Summary of section 3.3.1

In Section 3.3.1, the settings of the three acoustic cues were discussed. Each cue was manipulated in five steps. The range and step size of each cue varied with each parameter. The following table summarizes the settings of the three cues at each of the five levels.

Table 3.11 Five-level settings of the three acoustic cues

Cues	Levels (difference between V1 and V2)				
	1	2	3	4	5
F0: (Hz)	-50	-25	0	25	50
Duration (D): ratio	0.5	0.75	1	1.25	1.5
Intensity (I): dB	-9	-4.5	1	4.5	9

3.3.3 Manipulation of test stimuli

In order to compare the relative weight of the three correlates in stress judgment, the different levels of one parameter were combined with all the levels of the other two parameters. In other words, each level of F0 change was combined with every one of the five levels of duration change and every one of the five levels of intensity change. This produced a total of 125 different tokens for each nonsense word (5 F0 levels × 5 duration levels × 5 intensity levels). Altogether 375 items of three test words were constructed. Figure 3.3 gives a schematic diagram of the construction process.

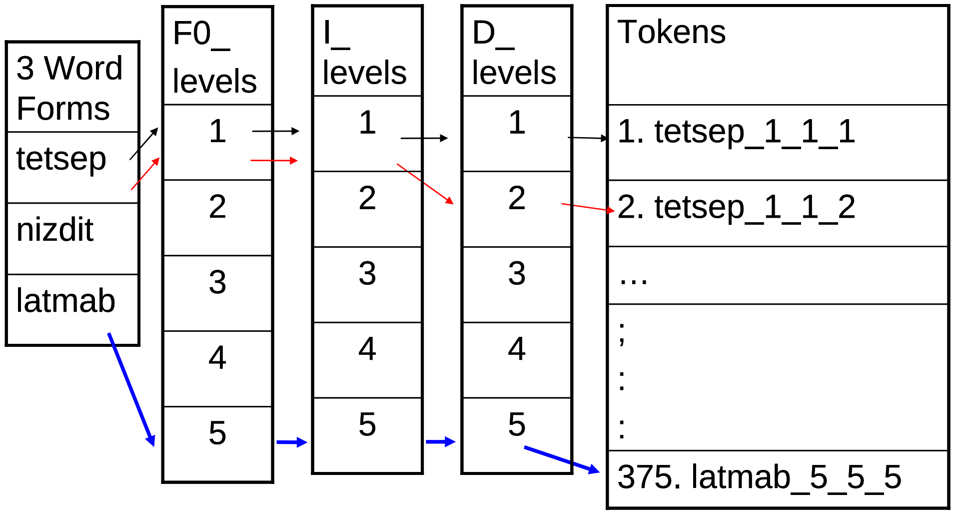
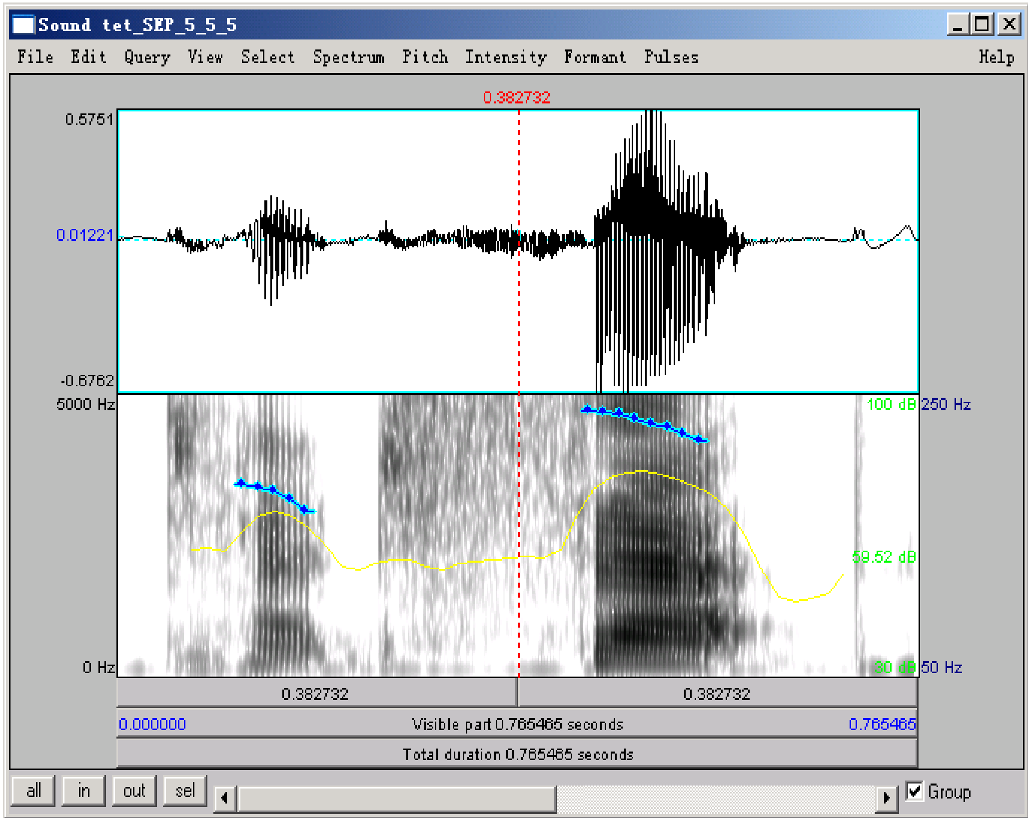


Figure 3.3 Construction of the 375 test tokens

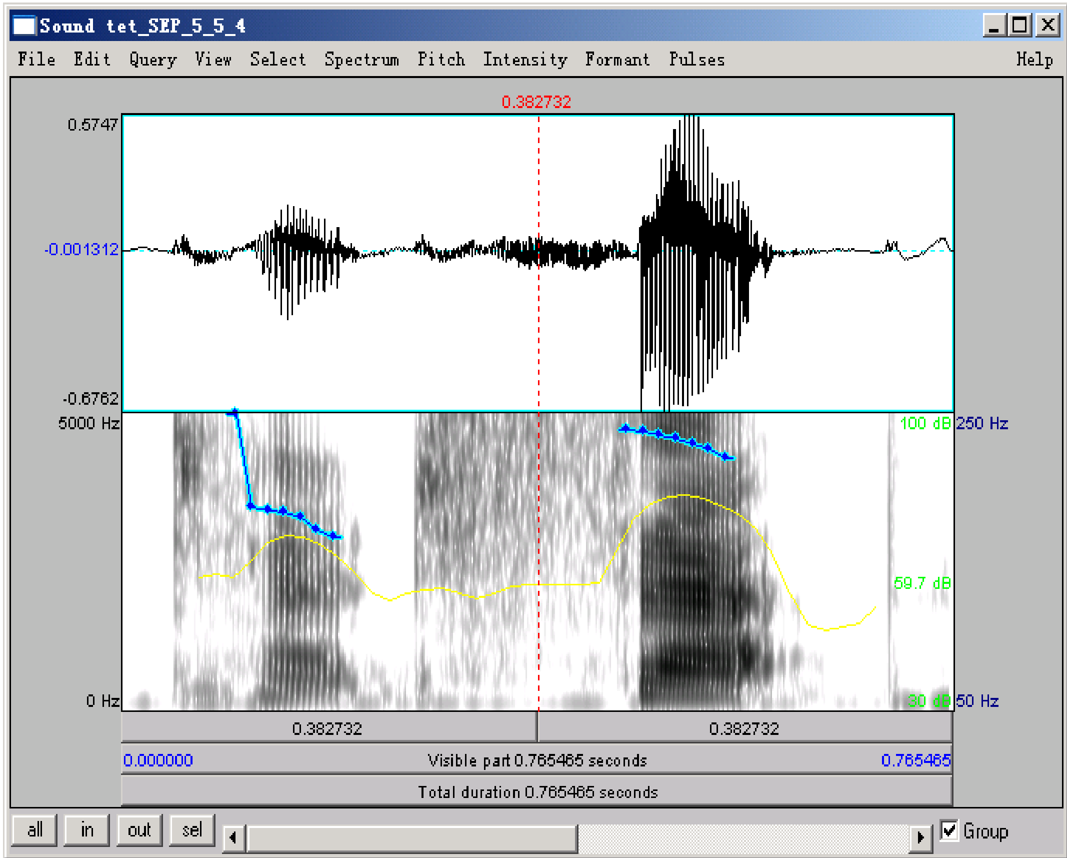
The construction process was realized by a manipulation script (Appendix B Normalization and Manipulation Scripts). This script took the result of the normalization process and changed the three parameters according to the pre-determined values listed in Table 3.11 above. The Manipulation Script was similar to the normalization scripts. It first opened the textgrid file, which included the segmentation of the five intervals of each test word, O, V1, CO, V2 and C. To change the three parameters, three loops were used in the script, one embedded in the other. The pitch loop was the outside loop. The intensity loop was embedded in the pitch loop and the duration loop was the embedded in the intensity loop. There were five steps in each loop to achieve the 5 levels of manipulation of each parameter. The pitch loop changed the F0 contour of V1 and V2 so that the difference between V1 and V2 was -50Hz, -25Hz, 0Hz, 25Hz, 50Hz in each step. At each step of the pitch loop, the intensity changed from step one to step five. At each step of the intensity change, the five levels of duration change were realized one by one. Then the script returned to the intensity loop and proceeded with the second step of intensity change. Again, the five levels of duration change were realized one by one. This process continued until the five levels of intensity loop were finished and the script returned to the pitch loop and proceeded with the second step of the pitch loop. This whole process continued until the last step of the pitch loop was finished.

Take the manipulation of *tetsep* as an example. To produce the first token of *tetsep*, the F0 value was changed first. In order to meet the requirement of an F0 difference of -50Hz, the F0 value on V1 was reduced by 25Hz and the F0 value on V2 was increased by 25Hz. This result was then submitted to the embedded intensity loop. The intensity on V1 was reduced by 4.5dB and the intensity on V2 was increased by 4.5dB. This result was then submitted to the duration loop. V1 was reduced to 6/5 of its original length and V2 was increased to 4/5 of its original length, so that the ratio between V1 and V2 was 0.5. Figure 3.4 shows the spectrogram of this token. On this token, the first syllable had shorter duration, lower F0 and lower intensity than the second syllable.



	V1 and V2 difference		
tetsep	F0	Intensity	Duration
Token 1	- 50Hz	- 9dB	0.5

Figure 3.4 Spectrogram and waveform of the first token of *tetsep*, with the first step of pitch manipulation, first step of intensity manipulation and first step of duration manipulation.

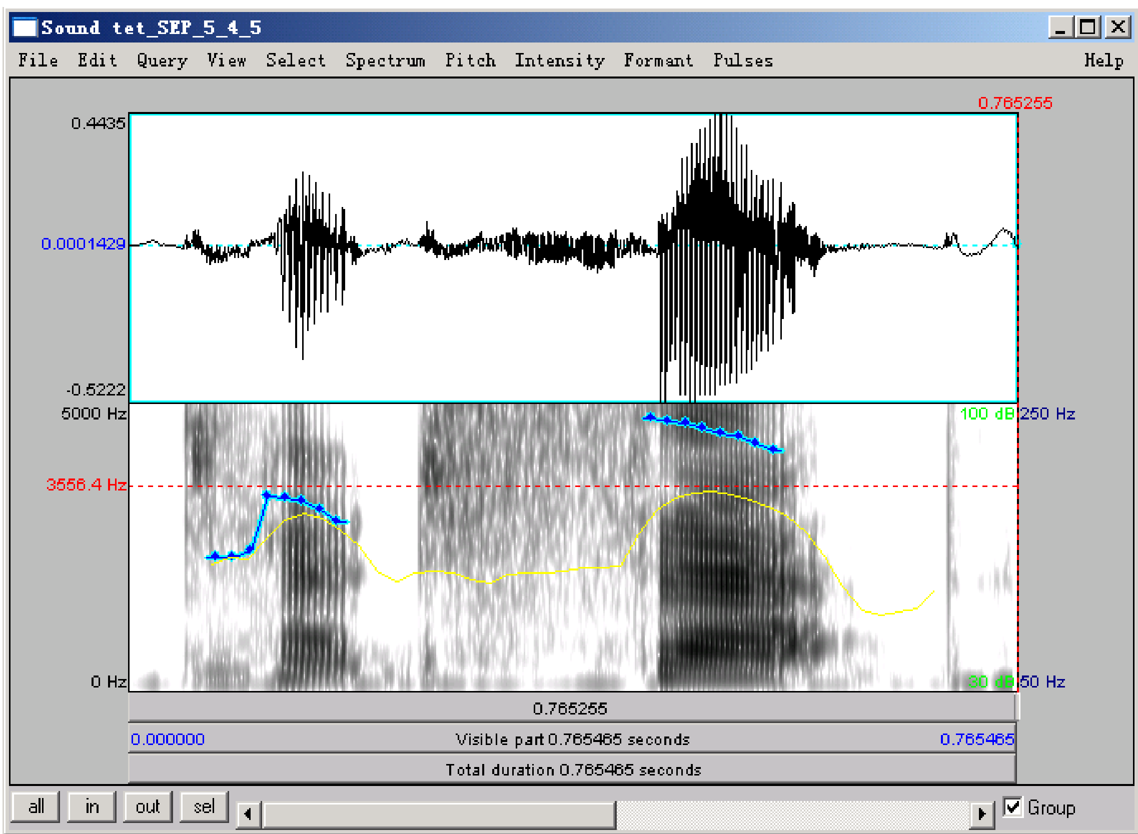


	V1 and V2 difference		
tetsep	F0	Intensity	Duration
Token 2	-50Hz	-9dB	0.75

Figure 3.5 Spectrogram and waveform of the second token of *tetsep*, with the first step of pitch manipulation, first step of intensity manipulation and second step of duration manipulation.

The script then continued to produce the second token of the nonsense word *tetsep*. The difference between this token and the last token lies only in the duration ratio. The script went on to the second step of the duration loop and changed the duration ratio between the two vowels to 0.75. Figure 3.5 shows the second token of *tetsep*.

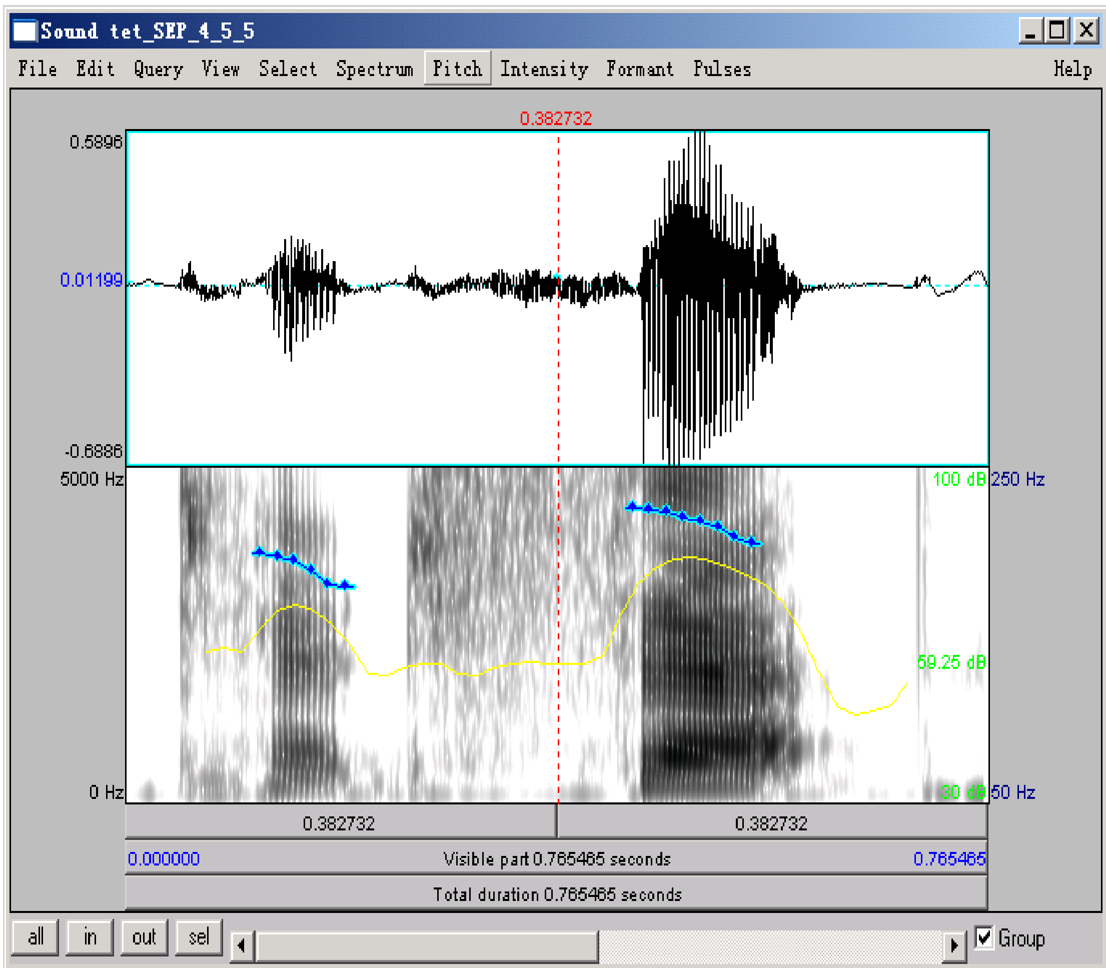
When the duration loop finished all five steps, the intensity loop moved from step one to step two, which was -4.5dB. At this stage (F0= -50 Hz; Intensity= -4.5 dB), duration, again, was changed for five levels. Figure 3.6 below showed the sixth token of *tetsep*, where the F0 difference between V1 and V2 was still -50Hz, and the intensity difference between V1 and V2 was -4.5dB, and the duration of V1 was only half of V2.



	V1 and V2 difference		
tetsep	F0	Intensity	Duration
Token 6	- 50Hz	- 4.5dB	0.5

Figure 3.6 Spectrogram and waveform of the sixth token of *tetsep*, with the first step of pitch manipulation, second step of intensity manipulation and first step of duration manipulation.

The twenty-sixth token is where the pitch manipulation was moved to step two and both intensity and duration were at step one. The following figure shows that the difference between the two vowels in terms of pitch contour was reduced. V1 was also lower in intensity and shorter in duration than V2.



	V1 and V2 difference		
tetsep	F0	Intensity	Duration
Token 26	- 25Hz	- 9dB	0.5

Figure 3.7 Spectrogram and waveform of the twenty- sixth token of *tetsep*, with the second step of pitch manipulation, first step of intensity manipulation and first step of duration manipulation.

The 125 tokens of the word form *tetsep* were created first by running the script for the base form of *tetsep*. In the same fashion, the 125 tokens of the word form of *nizdit* and the 125 tokens of *latmab* were created. These 375 nonsense tokens were used for the present experiment.

3.4 Participants

Two groups of listeners participated in the study. The control group consisted of 38 (17 male and 21 female) native-English-speaking volunteers. The experimental group had 62 Chinese learners of English (19 male and 43 female).

The recruitment of native English speakers was conducted by putting up posters around the Univeristy of Victoria campus and advertising the project in classes. Interested volunteers contacted the researcher via e-mail. Only volunteers who had no experience of Chinese or of any other tone language were accepted for the study. No subject reported hearing problems before. Test appointments were arranged through e-mail and phone calls subsequently. On the day of the appointment, language and music background information was collected from each subject through a questionnaire (Appendix C, Background Questionnaire for native English speakers). The participants were also asked to indicate whether they were familiar or not with the idea of “word stress”. As a reward for their time and effort, each native speaker received a small gift.

Another group of participants were 62 native Chinese speakers who were students of Nanjing University of Science and Technology. These students were from two English classes, one of English majors and the other of non-English majors. In the recruitment process, the researcher contacted five English language instructors in the University by e-mail describing the project and asking for permission to announce the recruitment information in their class.

With permission from two instructors, the researcher made an announcement in one class with students who majored in English and another class with students who majored in engineering. The project was described to the students and it was emphasized that participation in the research was separate from classroom activities and that it was totally voluntary and not related to any grade the students might receive in the courses they were taking. The instructors of the courses or any other school authorities would not have access to the information gathered in the experiment. The project description and consent form were handed out to the students and the researcher encouraged students to ask questions about the project and express any concerns that they might have. A sign-up sheet was used to register interested students, their contact information and their availability. Students were encouraged to take the consent form home where they could review the information about the project in more detail and contact the researcher if they had further questions. Thirty-three students from the English-major class and twenty-nine students from the engineering class volunteered for the project. Two time slots were chosen according to the different schedule of the two classes and the availability of language lab space in the university. Again, it was confirmed that every participant had normal hearing based on self-report. A background questionnaire (Appendix D, Background Questionnaire for Chinese learners of English) was collected to gather information about their use of dialects and knowledge of other foreign languages. The information whether they were familiar with the idea of English stress was also collected in the questionnaire.

The consent form and questionnaire for Chinese learners of English were written in both Chinese and English, and the participants could choose either language to answer the questions on the questionnaire. All participants were requested to sign the consent form and fill out the questionnaire before they started the test.

3.5 The Perception Test

Previous sections introduced the stimuli created for the experiment and the participants involved in the study. This section describes the organization of the stimuli in the perception test and the procedure of the test.

3.5.1 The organization of the tokens

In addition to the 375 constructed nonsense test items, 100 tokens of real English were also included. These 100 tokens of real English words that were recorded were used as foils in the perception test. The responses to these real English tokens were also used for data screening. Among them, there were four real English words with two possible stress patterns, *concrete, export, implant, increase*, each with four repetitions (4 words* 2 stress positions* 4 repetitions =32). The other 68 tokens were 33 real English words with fixed stress, in which each of 32 words had two repetitions and one word had 4 repetitions (32*2 + 1*4 = 68). The 33 real English words were chosen randomly from the lists in which the nonsense words and non-fixed stress words were embedded in the recording. Recall that 16 real words were used in the lists with nonsense words and 19 words were used in the lists with non-fixed stress real English words. Of these 68 tokens, half had stress on the first syllable and half had stress on the second syllable.

The 375 nonsense tokens and 100 real word tokens were divided into 25 blocks, with 19 tokens in each block (19* 25 = 475). In each block, there were 15 nonsense tokens and 4 real word tokens. The 15 nonsense tokens were divided into five sets (5 sets * 3 tokens), each set containing one token of each of the three word forms. The four real word tokens were used as fillers, one word between every set of nonsense tokens. Among the four real English words, two have stress on the first syllable and two on the second syllable.

The exact ordering of the 475 slots is given in Appendix E, Ordering of the Stimuli Slots. Table 3.12 gives the example of one block. *Nt, Nn* and *Nl* refer to the three **n**onsense words, ***t**etsep*, ***n**izdit,* and ***l**atmab*, respectively. RP refers to real English words with two possible stress positions and r refers to real English words with fixed stress position. The number follows RP or r refers to the stress position, 1 for stress on the first syllable and 2 for stress on the second syllable.

Table 3.12 Example of the ordering of stimuli in one block

Trial / Block	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19
1	Nt	Nn	Nl	RP1	Nn	Nl	Nt	r2	Nl	Nt	Nn	r1	Nt	Nn	Nl	RP2	Nn	Nl	Nt

Nt, Nn and Nl refer to the three **N**onsense words, ***t**etsep*, ***n**izdit,* ***l**atmab*, respectively.
RP refers to real English words with two possible stress positions.
r refers to real English words with fixed stress position.
The number following the letter indicates the stress position of this real English word.

The order of presentation for the 475 tokens was randomized for each participant.

Preceding the real test section was a practice section with two blocks. The practice section had exactly the same format as the real test but included tokens not used in the real test.

3.5.2 Procedure

The perception test was a forced choice identification task experiment, conducted on a computer via a C++ program written for this project. Participants listened to the auditory presentation of each token over a headphone connected to an individual station and indicated whether the token had stress on "A. the 1st" syllable or "B. the 2nd" syllable by clicking on the corresponding button on the screen.

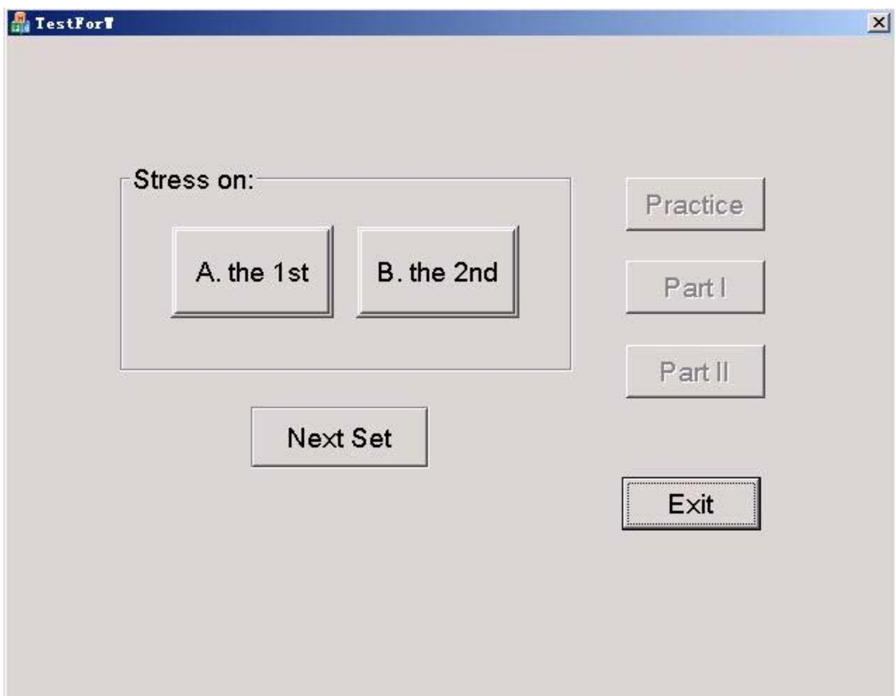


Figure 3.8 Screenshot of the perception test interface

The interval between two tokens was set to 1500 ms. An interval of 3000ms should be long enough for participants to make a judgment. In a similar forced choice test, Davis and Kelly (1997) used 1500ms and Flege (1993) used 1000ms. The next token was played after 3 seconds, no matter if the choice was made or not. Participants were not allowed to make any changes once the choice had been made. Participants were encouraged to make a best guess if they were not sure about the stress position.

The program of the perception test was designed in the way that participants would have to go through the practice section at least once before the real test. In the real test, the 25 blocks were divided into two parts, 15 blocks in Part I and 10 in Part II (see Figure 1). Between the two parts, participants were required to have a break of at least 5 minutes. Each block was introduced with a beep followed by 10s of silence to prepare the participants for the block. Between every block, the participants were required to have a rest of at least 30 seconds during which a recording of an English conversation or monologue was played. This measure was intended to immerse Chinese learners of English in an English speaking environment and to minimize the unnatural effects of nonsense words.

An instruction sheet (Appendix F, Instruction Sheet for Perception Test) was e-mailed to participants two days before the test date, and they were invited to read it if they had time. The same instruction sheet was given to the participants right before the practice section on the test date. Participants were given time to ask questions or express concerns before the test and during the break.

The entire perception test took about 40 minutes with the minimal obligatory break. If a participant chose to take a longer break during the test, the total duration was longer. The Chinese learners of English took the test in a language lab on their campus. Thirty-three English-major students took the test on one day and twenty-nine non-English-major students took the test on a different day. Participants started the test at the same time but proceeded with their own tempo.

Thirty-eight university level native English speakers (NE) took the test individually in the Phonetics Lab oratory of the Department of Linguistics at the University of Victoria, Canada. During the course of the test, the program recorded participant responses in a log file. The choice of stress on the first syllable was recorded as 1, and stress on the second syllable was recorded as 2. If no choice was made during the interval, 0 was used.

Chapter Four

Data Analyses and Results

The primary goal of this chapter is to introduce the statistical analyses of the data collected from the experiment and report on the results of the analyses.

The experiment involved two groups of participants, native English speakers (NE) and Chinese learners of English (CE). They were asked to make stress judgments on 100 real English words and 375 nonsense words with different combinations of cue settings. The stress judgment can be either stress on the first syllable or stress on the second syllable. The aim of this study is to investigate the effects of the three acoustic cues on stress perception by the two different groups. Thus, the data was first screened for validity (see section 4.1). Section 4.2 introduces the methods used in the statistical analyses and presents the general pattern of the data. In section 4.3 and 4.4, the results of two methods of data analysis are reported.

4.1 Data Screening

Thirty-eight NE participated in the study and all completed the tests. Thus, a total of 38 NE data were included for analysis. Two classes of CE participated in the study, one class of English majors and one class of science majors. In the first CE class, there were 33 participants and 31 of them finished the whole test. In the second class, there were 29 CE, and only 28 participants completed the whole experiment. As a result, a total of 57 CE participants' responses were included as valid data. Among the 57 participants who completed the experiments, 17 were male and 40 were female. Their English learning experience ranges from a minimum of 7 years to a maximum of 13 years, with an average of 9.28 years. The age when these participants started to learn English ranges from 7 to 14. Averagely speaking, they started learning English from 11.26 years old. In the Chinese education system, this means that most participants started learning English in the last two years of elementary school.

There were two parts in every participant's responses, responses to real English words and responses to nonsense tokens. The responses to real English words were analyzed and used to screen the collected data. As a forced choice identification task was used in this study, participants can make random choices in the test even if they were unable to perceive stress patterns. This screening step was conducted to make sure that the participants understood the idea of stress that was being tested in the present study. Only participants who showed a genuine understanding of stress in the responses to the real English words in the task were selected for further data analysis.

In the NE group, the highest rate of 100% was achieved by 9 participants, and the lowest rate was only 52%. The average correct rate for NE group was 86.61%, and the standard deviation was 14.54%. A total of 25 NE participants, accounting for 65.79% of the 38 NE participants, achieved a correct rate of over 80%. In the CE group, the highest correct rate was also 100%, achieved by 5 CE participants. The lowest correct rate was 55% and the average was 83.40%, with a standard deviation of 14.52%. Thirty-four out of the 59 CE participants achieved a correct rate of over 80%.

An independent sample t-test was conducted to compare the two groups' performance in the perception of stress on real English tokens. Results showed that the two groups were not significantly different from each other ($t(93)=1.06$, $p > 0.05$).

Previous research has used 80% correct production of any particular structure as the criterion level for acquisition (Andersen 1978, Cancino et al.1975; in Carlisle 1999, Eckman 1991, Carlisle 1998). Similarly, in this study, only subjects who achieved over 80% correct perception were assumed to understand the idea of stress and were thus used in the final data analysis.

4.2 Data Analysis

The targets of data analysis in the present study are the participants' responses to the 375 nonsense word tokens in the perception task. Depending on the combination of the different configurations of the three acoustic cues, a nonsense token can be perceived as having either Initial Stress (IS) or Final Stress (FS). We expect participants to exhibit systematic reaction to the systematically manipulated tokens. For example, when the three cues on a token were manipulated in the same direction, e.g. the first syllable is higher in F0, longer in duration and higher in intensity, then a participant should be more likely to perceive it as having IS. When the first syllable is higher in F0, shorter in duration and has the same intensity as the second syllable, then the participants' judgments may vary according to the importance they assign to each cue or cue combinations in stress perception.

In analyzing participants' responses to different tokens, we anticipate achieving two goals. First, from the participants' reactions to the 375 tokens with different combinations of cue configurations, we could explore the weight of each cue on their stress perception. Second, by comparing the difference in NE and CE participants' reactions in a systematic way, we can delve into the difference in cue reliance between the two groups.

Section 4.2.1 below first introduces how raw data were organized to allow for proper comparisons among the cues and between groups. Section 4.2.2 is intended to describe the general pattern of the data and offer a panoramic view of the two groups and the effects of the three acoustic cues on their stress perception. The results from two more sophisticated statistical methods are introduced below in section 4.3 and 4.4.

4.2.1 Two types of data organization

As indicated above, the dataset for this study is complicated. This section presents the ways in which data were organized for efficient analysis. There are two possible ways to list all the data points. One is subject-entry organization, and the other is token-entry organization.

Subject-entry organization is where each subject represents a line in the dataset. His/her responses to the 375 tokens are listed as different columns. Thus, the raw data include 59 rows (25 NE and 34 CE) and 375 columns (for responses to the 375 tokens). The table below provides a schematic view of data organization.

Table 4.1 Schematic view of subject-entry data organization

Subject ID	Group	Responses to the 375 tokens				
		Token 1: latmab_1_1_1	...	...	...	Token 375: tetsep_5_5_5
1	NE	1				
...	...					
...	...					
...	...					
59	CE					

In addition to the raw data, the responses to the 375 tokens were used to compute Percentage of Initial Stress (ISP). Fifteen categories of ISPs were calculated for each participant, five ISP values for each of the three acoustic cues (see Table 4.2 below). ISP_F0_1 refers to the percentage of IS responses over all tokens with the first level of F0 manipulation, regardless of the duration configuration and intensity configuration on the tokens. This includes a total of 75 tokens, i.e. the tokens with the first level of F0 manipulation in combination with all five levels of duration manipulation, all five levels of intensity manipulation on the three word forms (1 F0 × 5 dur × 5 int × 3 word forms). For each participant, the ISP is calculated as the number of IS responses he or she made over 75 (the total number of) tokens in each category. For example, for a participant, an ISP_Dur_4 of 32% means that he/she judged 24 tokens to be IS among the 75 tokens in the Dur_4 category (with a duration manipulation of level 4, the first syllable is 1.25 longer than the second syllable).

Table 4.2 Fifteen computed Initial Stress Percentage (ISP) values in subject-entry data organization

	F0	Duration	Intensity
Level 1	ISP_F0_1 = the number of one participant's IS response / 75	ISP_Dur_1	ISP_Int_1
Level 2	ISP_F0_2	ISP_Dur_2	ISP_Int_2
Level 3	ISP_F0_3	ISP_Dur_3	ISP_Int_3
Level 4	ISP_F0_4	ISP_Dur_4	ISP_Int_4
Level 5	ISP_F0_5	ISP_Dur_5	ISP_Int_5

Another type of data organization is to take each token as an entry in the dataset. Thus, there is a total 375 lines in the dataset and the responses by the 59 participants were organized into 59 columns. In this data organization, two ISP were calculated for each token, one for CE and one NE. For each token, ISP_CE is the total number of IS over 34, which is the total number of CE participants. ISP_NE is the total number of IS responses over 25, which is the total number of NE participants. Hypothetically, if the token, tetsep_5_1_2, has an ISP_CE of 94% it would mean that 32 out of 34 CEs perceived this token to have IS, and an ISP_NE of 60% would mean that 15 of the 25 NE participants perceived this token to have IS. Table 4.3 provides an overview of this data organization.

Table 4.3 Schematic view of token-entry data organization

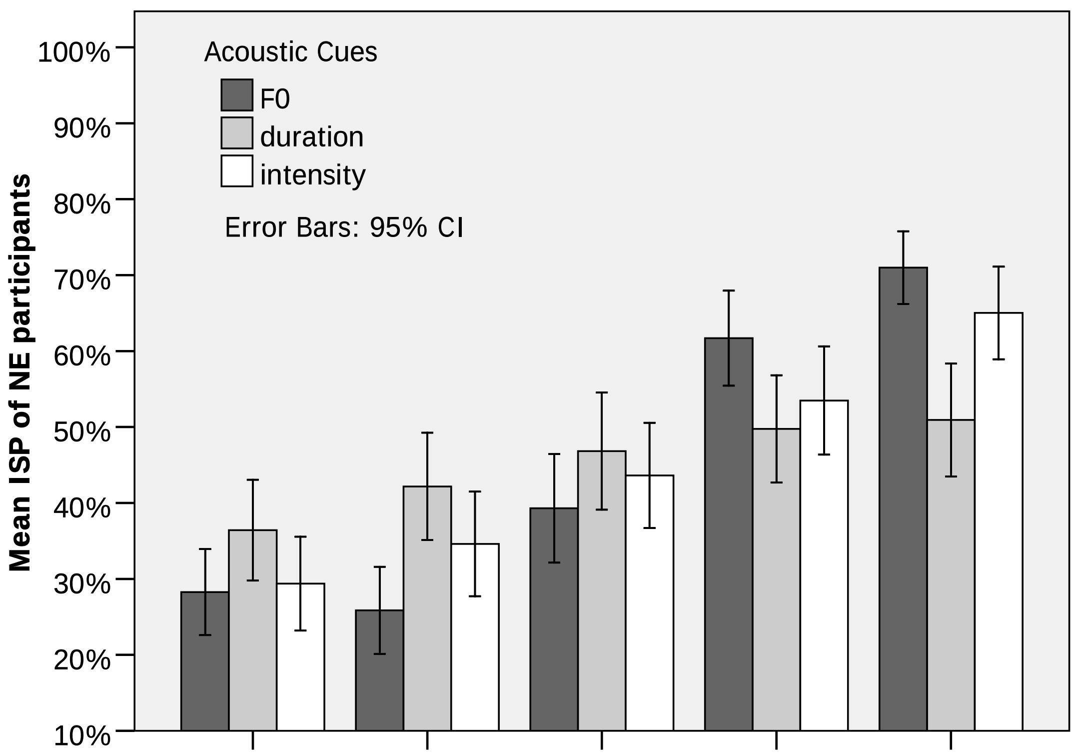
No.	Token list	Responses by the 59 participants				Computed ISP	
		participant 1	...	...	participant 59	ISP_NE	ISP_CE
1	Latmab _1_1_1	1				the number NE participant who judged this token to have IS / 25	the number CE participant who judged this token to have IS / 25
...	...	...	...	...	...	...	...
...	...	...	...	...	...	...	...
375	Tetsep _5_5_5					the number NE participant who judged this token to have IS / 25	the number CE participant who judged this token to have IS / 25

The data were organized in different ways as certain statistical analyses are only appropriate in a specific data organization. The type of data organization used is indicated when the result of this analysis is discussed below.

4.2.2 General results

A total of 22,125 responses from the perception test (59 subjects * 375 tokens) were subjected to analysis. Overall, 44.21% of the responses favored Initial Stress (IS), 54.99% favored Final Stress (FS), and 0.8% of the results were timed-out responses. The percentage of FS responses is larger than that of IS responses. No IS preference was observed in the data.

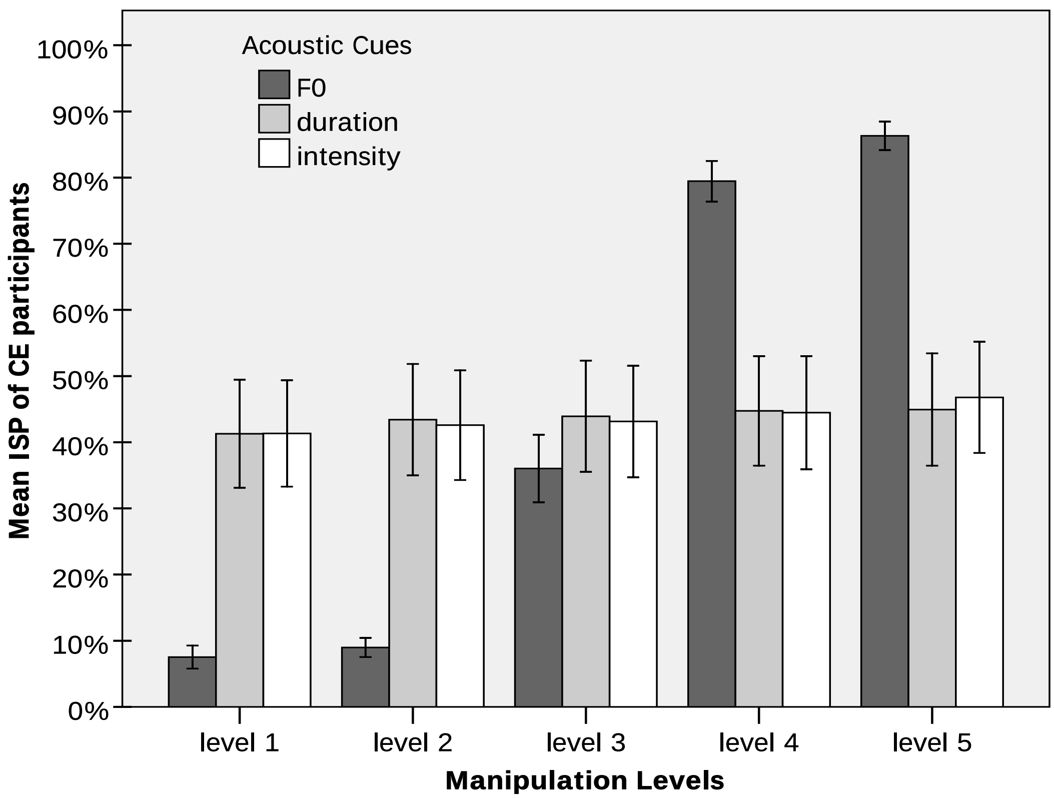
The pattern of responses is summarized in the following two figures, Figure 4.1for NE and Figure 4.2 for CE. The X-axis represents the five different manipulation levels. The Y-axis is the percentage of IS responses (ISP). The three bars represent the three acoustic cues. The percentage of IS responses at each level of F0 manipulation averaged over all duration and intensity levels is indicated by the dark bar. This means that each F0 bar represents the number of IS responses made by 25 NE or 34 CE in response to the 75 tokens at each level (1 level of F0 manipulation × 5 levels of duration × 5 levels of intensity × 3 word forms). Similarly the ISP at each level of duration manipulation is indicated by the grey bar and intensity by the white bar. As level 1 and 2 are where the first syllable has a lower value than the second syllable and level 4 and 5 are where the first syllable has a higher value than the second syllable, we would expect the IS responses to increase as the manipulation level proceeds from level 1 to level 5. The magnitude of the increase along the dimension of an acoustic cue can be taken as a general representation of the main effect of this acoustic cue.



F0	28.27%	25.87%	39.31%	61.71%	70.99%
Duration	36.43%	42.19%	46.83%	49.76%	50.93%
Intensity	29.39%	34.61%	43.63%	53.49%	65.01%

Figure 4.1 Mean Percentage of Initial Stress (ISP) of NE for the 375 tokens as a function of a) F0; b) duration; and c) intensity manipulation. Error bars enclose 95% CI.

For NE participants, the ISPs are different for each of the five levels of the three cue manipulations. As the manipulation changes from level one to level five, all three bars increase. The increase in F0 (the dark bar) is the largest, while the increase in duration is the lowest. Figure 4.1 shows that for NE, when F0 is manipulated from level 1 to level 5, the ISP increased from 28% to 71%. The five-level duration manipulation led to an increase of ISP from 36% to 51%, and the intensity manipulation led to a change from 29% to 65%.



F0	7.53%	8.98%	36.04%	79.45%	86.31%
Duration	41.29%	43.41%	43.92%	44.75%	44.94%
Intensity	41.33%	42.59%	43.14%	44.47%	46.78%

Figure 4.2 Mean Percentage of Initial Stress (ISP) of CE for the 375 tokens as a function of a) F0; b) duration; and c) intensity manipulation. Error bars enclose 95% CI.

For CE participants, on the other hand, the manipulation of F0, the dark bar, shows an increase along the change in the manipulation levels. The grey bar and the white bar stay almost the same at the five different levels. As can be seen from Figure 4.2, for the CE group, when F0 manipulation changed from level one to level five, IS responses increased from 7% to 86%. The change of intensity from level one to five caused an increase of ISP from 41.57% to 46.63%, and similar changes of duration resulted in an increase of ISP from 41.37% to 44.82%.

A four-way independent ANOVA was conducted on the token-entry datasheet. The dependent variable is the ISP value and the four independent variables are Group (two levels), F0 (five levels), Duration (five levels) and Intensity (five levels). Word form was included in the analysis as a covariate. The result of the analysis shows that there is a main effect of F0 ($F(4, 499)=440.32$, $p < 0.001$), Duration ($F(4, 499)= 7.24$, $p < 0.001$) and Intensity ($F(4, 499)=35.61$, $p < 0.001$), but no significant effect of Group ($F(1, 499)=1.62$, $p > 0.05$). Word form as a co-variant has a significant main effect ($F(1, 499)= 298.01$, $p < 0.001$), too. What's more informative here is actually the interaction between group and each of the three cues. Significant effects are observed for the three two-way interactions, Group by F0 ($F(4, 499)= 41.72$, $p < 0.001$), Group by Duration ($F(4, 499)= 2.72$, $p < 0.05$) and Group by Intensity ($F(4, 499)= 20.08$, $p < 0.001$). All three-way and four-way interactions are not significant.

A general review of the data indicates that the manipulation of the three acoustic cues has a significant effect on the ISP of the participants. Furthermore, the two groups are not significantly different from each other, but the way each of the three cues affect the two groups' performance is different.

In order to explore these patterns in more detail, the two methods of statistical analysis proposed above were conducted. The analyses of reliance measures are discussed first in section 4.3, and the analyses of logistic regression are discussed in section 4.4.

4.3 Reliance Measures

In this study, the nature of the data collected is that they are dichotomous data, either 1 (stress on the first syllable) or 2 (stress on the first syllable). Timed-out responses were coded as 'missing' in the dataset. Unlike the case for real English words, there is no right or wrong answers for the stress judgments of these nonsense tokens and thus we can't analyze the correct rate for each token as we did for the real words in data screening. Two ways of analyzing such data were identified in previous research and adopted for the present study. They are the analyses of reliance measures and logistic regression analyses. The use of reliance measures is becoming a major data analysis method in L2 cue-weighting studies. Logistic regression, with its own advantages, is a more recent proposal in analyzing cue-weighting data but has gained much popularity in both L1 and L2 studies. The two ways of statistical analyses are both useful and were adopted in the present study. As the data collected for this study are quite complicated, it is hoped that with results from two types of analyses, the conclusion can be more convincing. The present section, section 4.3, reports the results of reliance measures analyses and section 4.4 reports the results of logistic regression analyses.

Reliance measures refer to "the change in identification rates from one extreme of the stimulus space to the other" (Morrison 2005: 597). This method has been used in previous studies that compare the relative reliance on different acoustic cues in native and non-native speakers' perception of segment contrasts (Bohn 1995, Flege 1993, Flege et al. 1997, Escudero & Boersma 2004). Fry's series of studies (1955, 1958, 1964) on cue-weighting in stress perception by native speakers of English also used similar methods. An important feature of this method is that the raw responses have to be converted to the proportion of one type of response. For example, in the perception of the final consonant in the 'bid vs. bit' contrast, the analysis is conducted on the percentage of the 'bid' response over the total number of responses. In the present study, this can be the Percentage of Initial Stress judgments (ISP). Since the conversion of the Initial Stress Percentage has been introduced in section 4.2.1, this section begins with the introduction of how reliance measures are computed by the two extreme ISPs (section 4.3.1). The results of the analyses of the reliance measures are presented next (section 4.3.2). The detailed analysis of each of the three cues are presented in sections 4.3.3, 4.3.4, and 4.3.5.

4.3.1 The computing of reliance measures

In order to conduct analyses on reliance measures, the reliance scores for F0, duration and intensity have to be calculated. According to previous literature, the reliance score for an acoustic cue is defined by the difference between the values at the two ends of the manipulation continuum. For example, Flege (1993) conducted experiments to find the effect of preceding vowel duration on the contrast between word final /t/ and /d/. The vowel duration was manipulated in 17 steps in the word pair 'beat-bead'. The percentage of /t/ responses given to the stimuli was calculated at each step of the vowel manipulation. The reliance on vowel duration was "assessed by subtracting the percentage of /t/ responses given to the three stimuli with the longest vowel durations (viz. stimuli 15 to 17) from the percentage of /t/ responses given to the three stimuli with the shortest vowel durations (viz. stimuli 1 to 3)" (Flege 1993: 1595). Escudero & Boersma (2004) studied the effects of two acoustic cues, vowel duration and vowel formant (F1) on the perception of /i/ vs. /ɪ/. Vowel duration was manipulated in 7 steps and F1 was manipulated in 7 steps, too. Figure 4.3 shows the manipulation steps of the two acoustic cues.

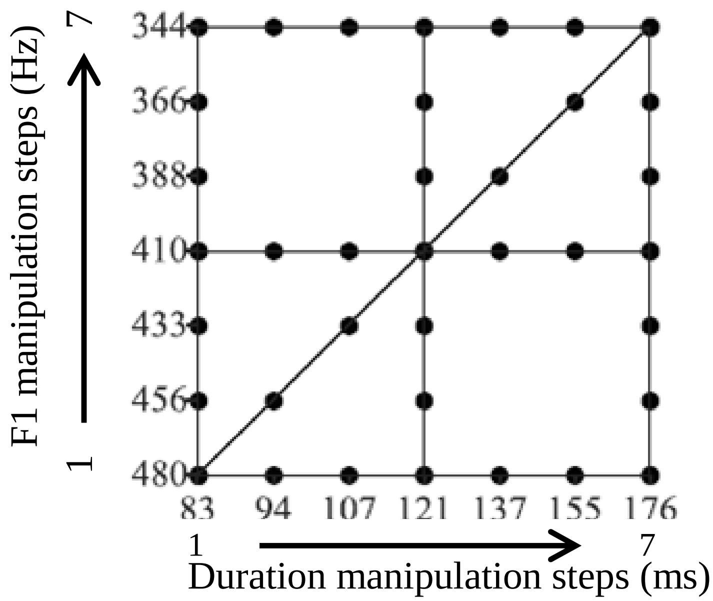


Figure 4.3 The manipulation metrics for Duration and F1 (adapted from Escudero & Boersma 2004)

For each participant, a duration reliance score was computed as the percentage of /i/ responses along the right edge of the stimulus rectangle, at step 7, minus the percentage of /i/ responses along the left edge, at step 1. Similarly, "a spectral reliance was computed as the percentage of /i/ responses along the top edge minus the percentage of /i/ responses along the bottom edge)" (Escudero & Boersma 2004: 558).

Following previous practice, three reliance scores were calculated for each participant in the present study. They were the reliance score for F0 (Reliance_F0), the reliance score for duration (Reliance_D), and the reliance score for intensity (Reliance_I). Using the subject-entry data organization, each reliance score is calculated by taking the ISP at level five minus the ISP at level one (refer to Table 4.2). For example, the reliance score for F0, reliance_F0, is calculated by subtracting ISP_F0_1 from ISP_F0_5. The reliance score for duation, reliance_D, and the reliance score for intensity, reliance_I, were calculated in the same way. Table 4.4 below summarizes the mean reliance scores averaged across participants in the two groups, as well as the standard error and the 95% confidence interval of the mean scores.

Table 4.4 Average Reliance Scores of F0, Duration and Intensity in the Two Groups

group	reliance score	Mean	Std. Error	95% Confidence Interval	
				Lower Bound	Upper Bound
NE	Reliance_F0	42.70%	4.70%	33.30%	52.20%
	Reliance_D	14.50%	1.70%	11.10%	17.90%
	Reliance_I	35.60%	3.10%	29.40%	41.80%
CE	Reliance_F0	78.80%	4.00%	70.70%	86.90%
	Reliance_D	3.60%	1.40%	0.70%	6.50%
	Reliance_I	5.50%	2.70%	0.10%	10.80%

In both groups, the reliance score of F0 is the highest, and the score of duration is the lowest. The F0 score is higher in the CE group than in the NE group, and the duration and intensity score are both higher in the NE group. Figure 4.4 below provides a bar graph of the three reliance scores for the two groups. To evaluate the statistical significance of the differences in these scores, ANOVA tests were conducted.

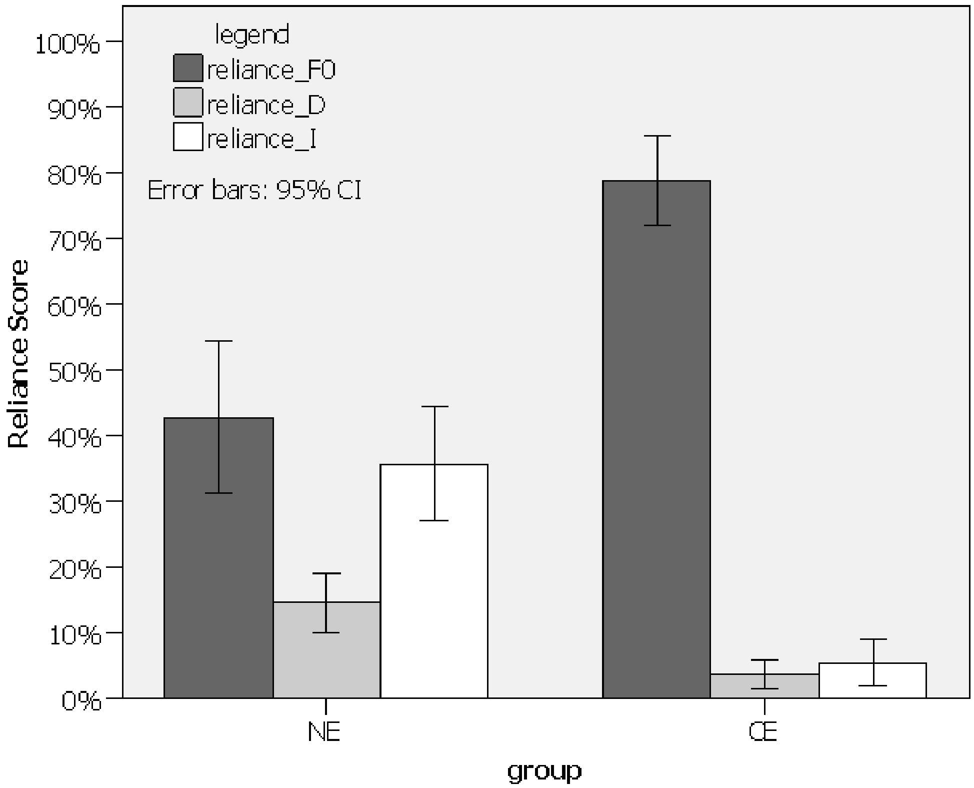


Figure 4.4 Reliance Scores of F0, Duration and Intensity in the NE and CE groups

4.3.2 General results of the analyses of reliance measures

The three reliance scores from the 59 participants were submitted to a two-way mixed-design analysis of variance (ANOVA)[1]. The two factors are Group (two levels), Acoustic Cue (three levels, F0, duration and intensity). Repeated measures were used on acoustic cues, as all the three acoustic cues were tested on the same participants.

The results of the mixed-model ANOVA show that Group doesnot have a significant effect ($F(1, 57) = 0.70$, $p > 0.05$). But effects of different acoustic cues

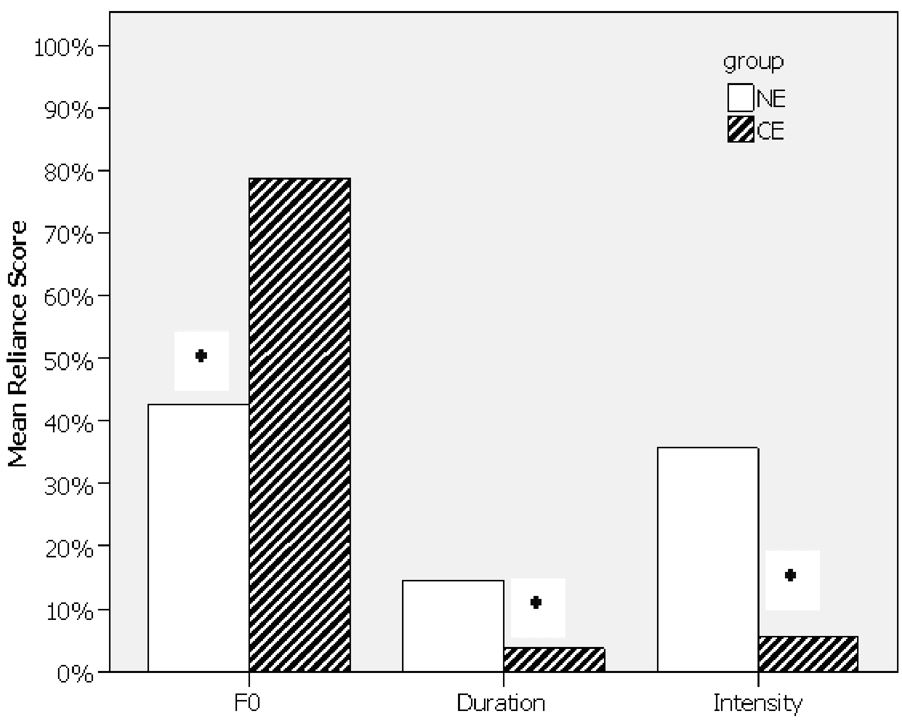
[1] The dataset used in this analysis is the subject-entry dataset.

are significant ($F(1, 57) = 121.85$, $p < 0.001$). Post-hoc (Bonferroni adjustment) test showed that reliance_F0 is significantly larger than reliance_D (mean difference 51.68%, $p < 0.001$). Reliance_F0 is also significantly larger than reliance_I (mean difference 40.21%, $p < 0.001$), and reliance_I is also signficantly larger than reliance_D (mean difference 11.46%, $p < 0.001$). The interaction between the cues and group also has a significant effect on the reliance score ($F(1, 57) = 48.00$, $p < 0.001$), which suggests that the two groups are different in their reliance on the three different cues.

Two groups of tests were conducted to investigate the interaction between the reliance scores and Group. Comparisons are made across the three reliance scores in each group and across the two groups for each acoustic cue.

Two repeated-measure ANOVA tests were conducted on NE and CE, separately. The factor of reliance score has three levels: reliance_F0, reliance_D and reliance_I. For NE, the reliance of the three cues differs significantly ($F(1,24)=9.96$, $p < 0.01$). Post-hoc tests showed that the difference lies between reliance_F0 and reliance_D ($p < 0.001$), and also between reliance_I and reliance_D ($p < 0.001$). NE's reliance on F0 is not significantly different from their reliance on Intensity ($p > 0.05$). The same repeated ANOVA on the CE group showed that CE's reliance on the three cues is significantly different ($F(1,33)=299.79$, $p < 0.01$). Post-hoc tests showed that the difference lies between reliance_F0 and reliance_D ($p < 0.001$), and also between reliance_F0 and reliance_I ($p < 0.001$). CE's reliance on duration is not significantly different from their reliance on intensity ($p > 0.05$).

To compare the difference between the two groups in terms of each cue, three independent *t*-tests were conducted. The two groups are significantly different in all three cue-reliance scores (see Figure 4.5 below, an asterisk indicates significant difference). NE has a significantly lower F0 reliance score than CE ($t(57)= -5.81$, $p < 0.001$). For both reliance_D (t(31.89, equal variance not assumed)= 6.71, $p < 0.001$) and intensity (t(35.59, equal variance not assumed)= 4.51, $p < 0.001$), NE is significantly higher than CE.



	reliance_F0	reliance_D	reliance_I
NE	42.72%	14.51%	35.63%
CE	78.78%	3.65%	5.45%
Difference	-36.06%	10.86%	30.18%

Figure 4.5 Difference between NE and CE reliance scores for the three cues

Summary of the findings of reliance measure analyses

In the analyses above, the reliance scores of F0, duration and intensity were computed for each participant in the two groups. The reliance scores were submitted to mixed-model ANOVA to test the effects of the three cues and group. The reliance scores are significantly different among the three acoustic cues but insignificantly different between the two groups. The interaction between reliance score and Group is significant. Further repeated-measures tests within the two groups showed that acoustic cues also have a significant effect in each individual group. In the NE group, the reliance score on F0 is significantly higher than duration and the reliance score on intensity is also significantly higher than duration. The difference between F0 and intensity reliance score is not significant. In the CE group, the difference is between F0 reliance score and the scores for the other two cues. CE's reliance on F0 is significantly higher than both duration and intensity. The three *t*-tests between the two groups on the reliance score for each cue indicate that the two groups are significantly different from each other in their reliance on every acoustic cue. CE has a significantly higher reliance score on F0, while NE has significantly higher reliance scores on both duration and intensity.

Despite the interesting results summarized above, the reliance scores were only computed by the two extreme ISPs but overlooked the change within the two ends. This has been commented on as a serious disadvantage of the method (Morrison 2005). Thus, it is valuable to explore each acoustic cue and the step change of each acoustic cue in more detail. In the following paragraph, the analyses are broken down into three sections for the three acoustic cues. For each cue, two separate repeated-measures ANOVAs were conducted for the two groups to see if the five-step manipulation has a significant effect on the ISP change of the group. If the five-step manipulation of an acoustic cue showed a significant effect on the ISP change of the group, pairwise comparisons were made between every two steps to pin down the exact location of the difference. In addition, independent *t*-tests were conducted to compare the difference between NE and CE groups at each level of cue manipulation. We begin with the discussion of F0 in section 4.4.3.

4.3.3 Effect of F0

The reliance score for F0 is the highest among three cues in each group, and CE has a higher F0 reliance score than NE. In this section, the effect of the five level F0 manipulation is analyzed.

The effect of F0 manipulation is analyzed in the subject-entry dataset. In the subject-entry dataset, each subject is listed as an entry. For each participant, five ISPs were calculated, each as the number of IS judgments over 75 (the total number of tokens at each level of F0 manipulation in combination with the 5 levels of duration manipulation and intensity manipulation). Table 4.5 shows the mean ISP of the 25 NE participants and the average ISP of the 34 CE participants at the five levels of F0 manipulation, as well as the range, and the standard deviation of the ISPs.

Table 4.5 Mean ISP for the NE group and CE group at each level of F0 manipulation

F0 difference b/t V1 and V2 (Hz)	NE				CE			
	N	Mean	Range	Std. Deviation	N	Mean	Range	Std. Deviation
Level 1: -50	25	28.27%	69.00%	16.21%	34	7.53%	45.00%	9.83%
Level 2: -25	25	25.87%	57.00%	13.12%	34	8.98%	43.00%	10.33%
Level 3: 0	25	39.31%	65.00%	14.33%	34	36.04%	57.00%	13.72%
Level 4: 25	25	61.71%	61.00%	17.28%	34	79.45%	64.00%	16.56%
Level 5: 50	25	70.99%	61.00%	19.47%	34	86.31%	55.00%	14.18%

In the following sections, repeated-measures were used to investigate the main effect of F0 on the two groups. Pairwise comparisons were made to examine the difference between ISP values of every two levels of F0 manipulation, and finally the comparisons were made between the ISP values of the two groups at each level of F0 manipulation.

Main effect of F0 manipulation

Repeated-measures were used on the ISP values of the five levels of F0 manipulation in the two groups to see the effect of F0 manipulation on the ISP change of NE and CE. The result shows that the main effect of F0 manipulation is significant on NE's ISP values ($F(1, 24) = 53.68$, $p < 0.001$, $\eta^2 = .691$). The analysis on CE's ISP values shows that the main effect of F0 manipulation is also significant ($F(1, 33) = 307.33$, $p < 0.001$, $\eta^2 = .903$).

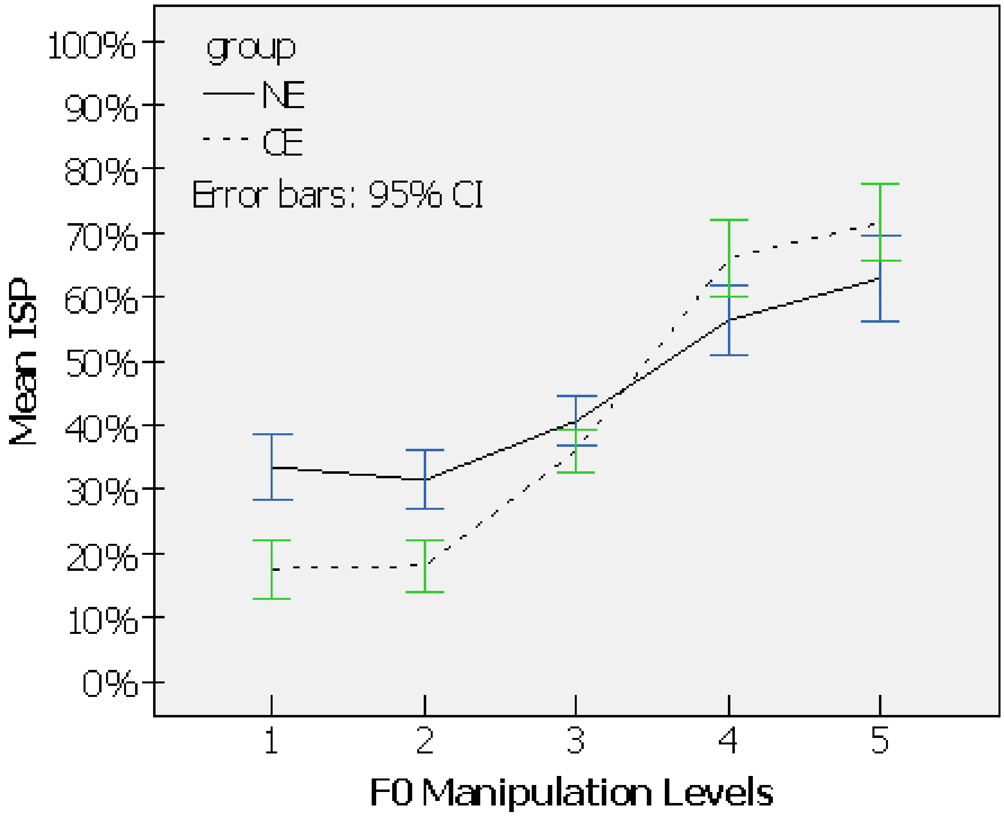


Figure 4.6 ISP of NE and CE group as a function of F0 manipulation

In **Figure 4.6**, two lines were plotted to represent the change in the percentage of IS responses of the two groups as the manipulation of F0 changes from level one to level five. The concrete line is for NE and the dotted line is for CE. It can be seen that the line for NE has a steeper change than the line for CE. Although the two repeated-measures tests indicated that F0 manipulation has a significant effect on both the NE and CE group, it is obvious from this figure that the change of ISP at each step is not equal in the two groups.

Pair-wise comparisons between all different steps.

Pairwise comparisons between every two levels' ISPs were conducted in the two groups, separately. Table 4.6 provides the differences in ISP values for each pair of comparison for NE and CE. An asterisk represents a significant contrast ($p<0.05$) between the pair. In both groups, nine out of the 10 contrasts are significant. The only insignificant contrast in both groups is between level one and level two. Although this is true for both groups, there is a subtle difference between the patterns of the two groups. In the NE group, the ISP value at level one is found to be higher than that at level two. For all other steps change, the NE ISP value increases as the step increases. For CE, level two has a higher ISP than level one, as expected. There is a linear relationship between ISP values, and step changes, and as the step increases, the ISP value increases.

Table 4.6 Pair-wise comparisons between ISP values at each level of F0 manipulation in the NE group and in the CE group.

NE difference b/t ISPs (a.)	level 1	level 2	level 3	level 4	level 5
level 1		-.0024	.110(*)	.334(*)	.427(*)
level 2	0.024		.134(*)	.358(*)	.451(*)
level 3	-.110(*)	-.134(*)		.224(*)	.317(*)
level 4	-.334(*)	-.358(*)	-.224(*)		.093(*)
level 5	-.427(*)	-.451(*)	-.317(*)	-.093(*)	

CE difference b/t ISPs (a.)	level 1	level 2	level 3	level 4	level 5
level 1		.015	.285(*)	.719(*)	.788(*)
level 2	-.015		.271(*)	.705(*)	.773(*)
level 3	-.285(*)	-.271(*)		.434(*)	.503(*)
level 4	-.719(*)	-.705(*)	-.434(*)		.069(*)
level 5	-.788(*)	-.773(*)	-.503(*)	-.069(*)	

* The mean difference is significant at the .05 level.
a Adjustment for multiple comparisons: Bonferroni.

T-test between NE and CE at each step of F0 change.

The previous two paragraphs reported the effects of the step change in F0 on the two groups, separately. In this paragraph, the two groups are compared to each other at each level of the F0 manipulation. Five two-tailed independent *t*-tests were conducted for the five manipulation levels. As can be seen from Figure 4.7, the ISP values of the two groups are significantly different at four out of the five levels of F0 manipulation. Results showed that at level one and two, NE has a significantly higher ISP (for exact statistics, see Figure 4.7) and at level four and five, CE has a significantly higher ISP. The two groups are not significantly different from each other in ISP at level three.

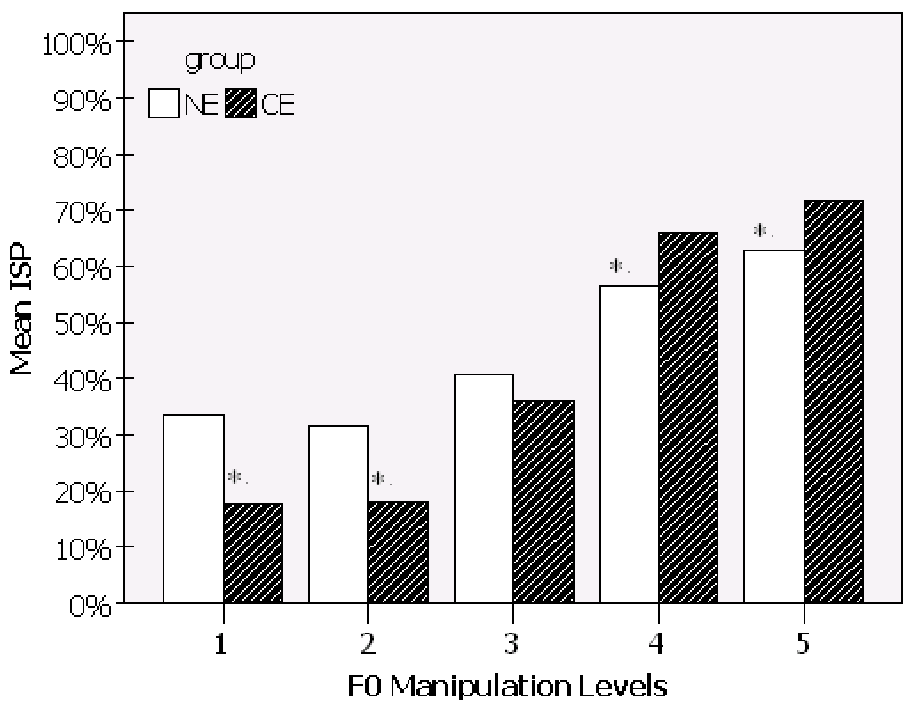


Figure 4.7 **Difference between NE and CE ISP at each level of F0 manipulation**

Table 4.7 ISP Difference between the two groups at each F0 manipulation levels

	ISP_F0_1	ISP_F0_2	ISP_F0_3	ISP_F0_4	ISP_F0_5
NE	28.27%	25.87%	39.31%	61.71%	70.99%
CE	7.53%	8.98%	36.04%	79.45%	86.31%
Difference	20.74%	16.89%	3.27%	-17.74%	-15.32%
***t*-test results**	t (36.78 a.)= 5.68, p < 0.001	t (57)= 5.53, p < 0.001	t (57)= 0.887, p > 0.05	t (57)= -3.99, p < 0.001	t (41.76 a.)= -3.34, p < 0.01

a. df adjusted as equal variance not assumed

Summary of the F0 effect

The exploration of the effect of F0 manipulation on the two groups shows that as the F0 changes from one level to another, the ISP values of the two groups differ, significantly. When we take each step into consideration, it is found that every step change from level two to level five induced significant changes in ISP values in both groups. The change from level one to level two, on the other hand, does not bring about significant changes in either group. As a matter of fact, NE shows a slight decrease in ISP value when F0 is increased from level one to level two. The comparisons of the two groups at each level of F0 manipulation reveal that CE has a significantly lower ISP when the first syllable is lower in F0 (level one and level two) but has a significantly higher ISP when the first syllable is higher in F0 (level three and four). NE shows a slightly higher ISP at level three ,but the difference is not significant.

4.3.4 Effect of duration change

Similar to the analysis of the effects of F0 manipulation in the previous section, the purpose of the present section is to investigate whether the manipulation of duration differences on the two syllables would lead to changes in the NE and CE participants' stress judgments. The analysis is also conducted on the subject-entry dataset. For each subject, five ISPs were calculated, each as the number of IS judgments over 75 (the total number of tokens at each level of duration manipulation in combination with the five levels of F0 manipulation and five levels of intensity manipulation). The average ISP of the 25 NE participants and the average ISP of the 34 CE participants at the five levels of duration manipulation are listed in Table 4.8. The range of ISP values and standard deviation are also included.

Table 4.8 Mean ISP of NE and CE groups at each level of duration manipulation

Duration Ratio of V1/V2	NE				CE			
	N	Mean	Range	Std. D.	N	Mean	Range	Std. D.
level 1: 0.5	25	36.43%	45.33%	11.65%	34	41.29%	45.33%	8.24%
level 2: 0.75	25	42.19%	50.67%	11.59%	34	43.41%	38.67%	8.08%
level 3: 1	25	46.83%	48.00%	11.35%	34	43.92%	34.67%	7.15%
level 4: 1.25	25	49.76%	56.00%	11.36%	34	44.75%	41.33%	7.78%
level 5: 1.5	25	50.93%	49.33%	11.69%	34	44.94%	32.00%	7.46%

Again, two analyses with repeated-measures were used to investigate the main effect of duration manipulation on the two groups, and pairwise comparisons were made between ISP values of different duration levels. The difference between the two groups at each level of duration manipulation is examined by *t*-tests.

Main effect of duration manipulation

The main effect of duration manipulation for each group was investigated by repeated-measures on the five ISPs at different levels of duration manipulation. The results show that the main effect of duration manipulation is significant on the ISP values for NE ($F(1, 24) = 33.12$, $p < 0.001$, $\eta^2 = .580$) and also significant for CE ($F(1, 33) = 4.82$, $p < 0.05$, $\eta^2 = 0.127$).

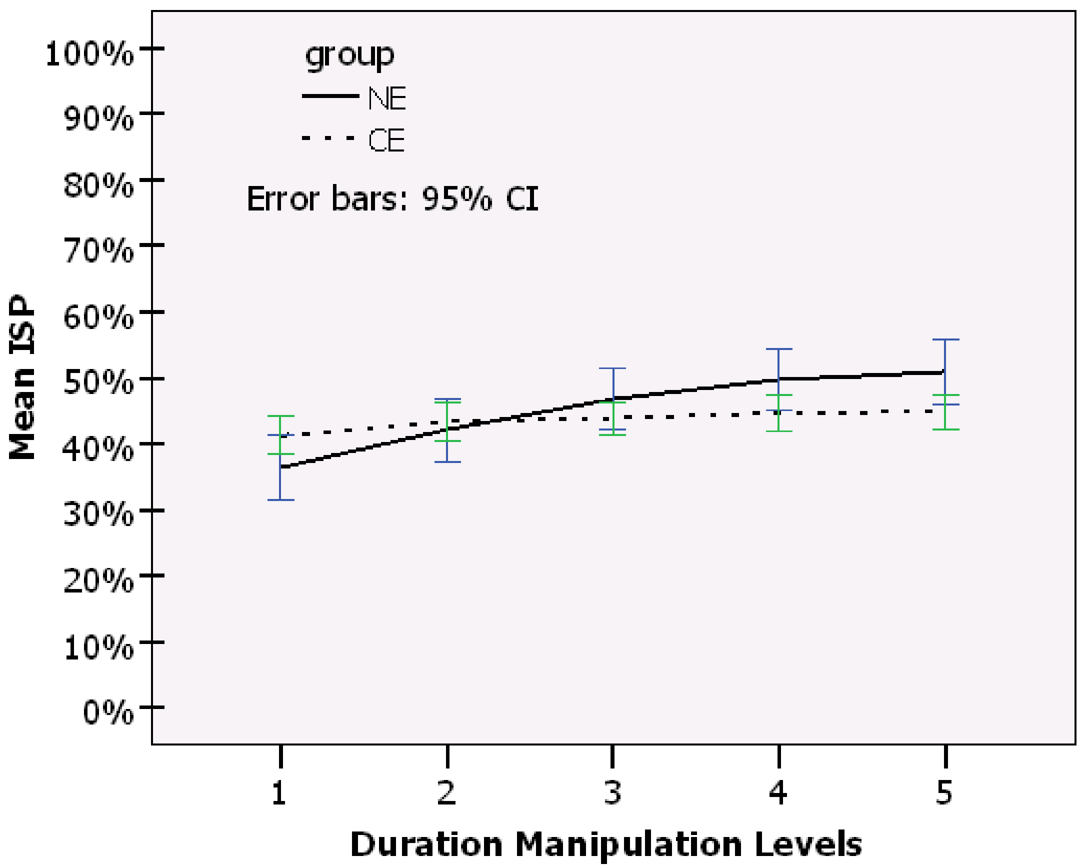


Figure 4.8 **ISP of NE and CE groups as a function of duration manipulation**

Figure 4.8 shows the change of ISP values of NE (concrete line) and CE (dotted line) from level one to level five of duration manipulation. For NE, the ISP increase is quite subtle along the change of duration manipulation levels. On the other hand, the change of ISP values from one level to another is minimal for CE. If duration manipulation has a significant effect in both groups, where does the difference lie? This question is answered in the next section.

Pair-wise comparisons between all different steps.

Pairwise comparisons were made between every pair of ISP values for the different levels of duration manipulation. Tests were conducted to see if adjacent levels differed in ISP values, and if not, where the significant difference lies. Pairwise comparisons were conducted for the two groups, separately.

Table 4.9 provides the contrasts in ISP values for the NE group and the CE group. An asterisk represents a significant difference ($p < 0.05$) between the two levels. In the NE group, level-one ISP is significantly different from ISPs at all four other levels, and so is level two ISP. More specifically, level-one ISP is significantly lower than the ISP of the next adjacent level: level-two ISP is significantly lower than the ISP of the next adjacent level, level three. However, level three ISP is not significantly lower than level-four ISP and again, level-four ISP is not significantly lower than level-five ISP. CE shows a different pattern of contrasts. When we look at the contrasts between levels in the CE group in Table 4.9, only three out of the 10 possible contrasts are significant. The three pairs of contrasts that are significantly different are between level one and level three, level one and level four, and level one and level five. In other words, no two adjacent levels are significantly different from each other in ISP values. Unlike the case of F0, both NE and CE ISPs increase as the duration manipulation increases from level one to level five.

Table 4.9 Pair-wise comparisons between ISP values at each level of duration manipulation in the NE group and in the CE group

NE difference b/t ISPs (a.)	level 1	level 2	level 3	level 4	level 5
level 1		.058(*)	.104(*)	.133(*)	.145(*)
level 2	-.058(*)		.046(*)	.076(*)	.087(*)
level 3	-.104(*)	-.046(*)		.029	.041(*)
level 4	-.133(*)	-.076(*)	-.029		.012
level 5	-.145(*)	-.087(*)	-.041(*)	-.012	

CE difference b/t ISPs (a.)	level 1	level 2	level 3	level 4	level 5
level 1		.021	.026(*)	.035(*)	.036(*)
level 2	-.021		.005	.013	.015
level 3	-.026(*)	-.005		.008	.010
level 4	-.035(*)	-.013	-.008		.002
level 5	-.036(*)	-.015	-.010	-.002	

* The mean difference is significant at the .05 level.
a Adjustment for multiple comparisons: Bonferroni.

T-test between NE and CE at each step of F0 change.

As the two groups exhibit different patterns of contrasts in the pairwise comparisons, it is expected that the comparison between the two groups may reveal further differences. Two-tailed independent *t*-tests were conducted to compare the difference between CE and NE at each of the five levels of duration manipulation. Figure 4.9 shows that two out of the five pairs of comparisons are significant, as indicated by an asterisk in the figure. From the summary of the *t*-tests results in Table 4.10, it can be observed that the NE group has a significantly higher ISP than the CE group at levels four and five (for detailed statistics, see the table below).

For level one, two and three, the two groups are not significantly different from each other. NE has a slightly lower ISP at level one and two but a higher ISP at level three.

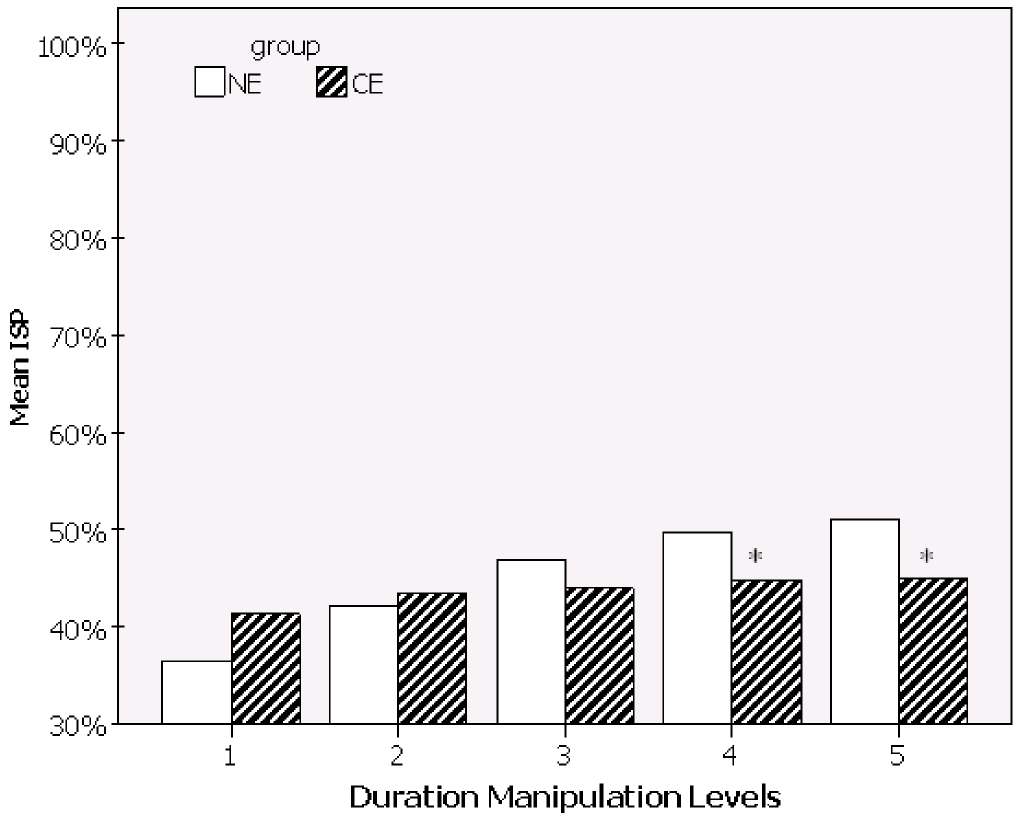


Figure 4.9 Difference between NE and CE ISP at each level of duration manipulation

Table 4.10 ISP Difference between the two groups at each duration manipulation level

	ISP_dur_1	**ISP_dur_2**	**ISP_dur_3**	**ISP_dur_4**	**ISP_dur_5**
NE	36.43%	42.19%	46.83%	49.76%	50.93%
CE	41.29%	43.41%	43.92%	44.75%	44.94%
Difference.	-4.86%	-1.22%	2.91%	5.01%	5.99%
***t*-test results**	t(57)= -1.88, p = 0.65 > 0.05	t(57)= -0.48, p > 0.05	t(57)= 1.205, p > 0.05	t(57)= 2.01, p = 0.049 < 0.05	t(57)= 2.40, p < 0.05

Summary of the duration effect.

The analyses of the effect of duration manipulation involved three types of tests, repeated-measures to test the main effect of duration manipulation, pairwise comparisons between levels of duration manipulation and *t*-tests between the two groups at each level of manipulation. Duration manipulation was shown to have a significant main effect on both NE and CE, although the size of the effect is much larger for the NE group than for the CE group. Comparing ISPs at two different levels, the differences between level three and four, and level four and five are not significant in the NE group. In the CE group, the only significant differences were found between level one and level three, four, five, respectively. None of the adjacent levels were found to be significantly different from each other. The *t*-tests between the two groups indicated that when the first syllable is 1.25 (level four) or 1.5 (level five) times longer than the second syllable, this leads to significantly more IS judgments in the NE group than in the CE group. When the first syllable is shorter than the second syllable (level one and two) or when the two syllables have the same duration, the two groups do not have a significantly different proportion of IS judgments.

4.3.5 Effect of intensity change

The last acoustic cue investigated is intensity. Intensity was also manipulated in five levels in the experiment. Thus, five ISP values were computed for each subject in the subject-entry dataset, ISP_Int_1, ISP_Int_2, ISP_Int_3, ISP_Int_4 and ISP_Int_5. These five ISP values correspond to the proportion of IS judgments among the 75 tokens at each level of intensity manipulation. Table 4.11 summarizes the five subject-averaged ISP values for the 25 NE participants and five subject-averaged ISP values for the 34 CE participants. Also included in the table are the range of the ISP values and the standard deviations.

Table 4.11 Mean ISP for NE and CE groups at each level of intensity manipulation

Intensity difference between V1 and V2 (dB)	NE				CE			
	N	Average ISP	Range	Std. Deviation	N	Average ISP	Range	Std. Deviation
level 1: -9	25	29.39%	60.00%	14.04%	34	41.33%	38.67%	7.75%
level 2: -4.5	25	34.61%	62.67%	13.96%	34	42.59%	40.00%	8.13%
level 3: 0	25	43.63%	54.67%	12.24%	34	43.14%	33.33%	7.54%
level 4: 4.5	25	53.49%	53.33%	12.57%	34	44.47%	40.00%	6.85%
level 5: 9	25	65.01%	57.33%	14.71%	34	46.78%	65.33%	9.84%

To examine the main effect of intensity manipulation on the two groups, two analyses with repeated-measures were conducted. For the difference between the ISP values at different levels, pairwise comparisons were made in the two groups, separately. To compare the two groups' ISP values, *t*-tests were used at each level of intensity manipulation.

Main effect of intensity manipulation

Figure 4.10 shows the ISPs at five levels of intensity manipulation. The concrete line indicates the change of NE ISP values as a function of intensity manipulation and the dotted line indicates the change of CE ISP values. From a glance of the figure, it seems that the manipulation of intensity has brought about considerable change in the NE ISP but very little change in CE ISP.

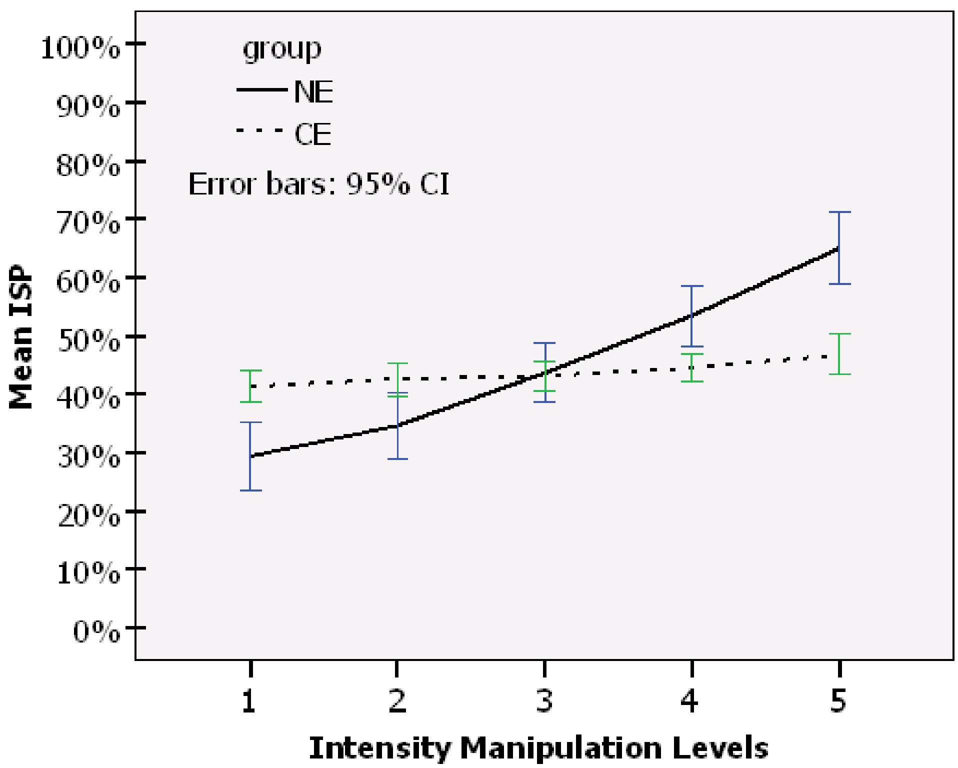


Figure 4.10 **ISP of NE and CE group as a function of intensity manipulation**

Repeated-measures were used on the 5 different levels of ISPs in the two groups. The results shows that the main effect of intensity manipulation is significant not only on ISP values for NE ($F(1, 24) = 57.48$, $p < 0.001$, $\eta^2 = .705$) but also for CE ($F(1, 33) = 6.77$, $p < 0.05$, $\eta^2 = 0.170$).

Pair-wise comparisons between all different steps.

In both groups, pairwise comparisons were made between every two levels to reveal the exact location of the difference.

Table 4.12 below provides the contrasts of different pairs of ISP values for NE and CE. An asterisk represents a significant difference ($p < 0.05$) between the contrast. In the NE group, all contrasts are significantly different. The ISP values at level two of intensity manipulation is higher than level one, level three is higher than level two, level four is higher than level three and level five is higher than level four. On the contrary, in the CE group, all contrasts except for one failed to be significant. The only significant contrast is between level one and level five. From this further examination of the difference between levels, it is shown that although intensity manipulation has a significant effect on both groups, the difference in ISP values exhibits different patterns in the two groups.

Table 4.12 Pair-wise comparisons between ISP values at each level of intensity manipulation in the NE group and in the CE group.

NE difference b/t ISPs (a.)	level 1	level 2	level 3	level 4	level 5
level 1		.052(*)	.142(*)	.241(*)	.356(*)
level 2	-.052(*)		.090(*)	.189(*)	.304(*)
level 3	-.142(*)	-.090(*)		.099(*)	.214(*)
level 4	-.241(*)	-.189(*)	-.099(*)		.115(*)
level 5	-.356(*)	-.304(*)	-.214(*)	-.115(*)	

CE difference b/t ISPs (a.)	level 1	level 2	level 3	level 4	level 5
level 1		.013	.018	.031	.055(*)
level 2	-.013		.005	.019	.042
level 3	-.018	-.005		.013	.036
level 4	-.031	-.019	-.013		.023
level 5	-.055(*)	-.042	-.036	-.023	

* The mean difference is significant at the .05 level.
a Adjustment for multiple comparisons: Bonferroni.

T-test between NE and CE at each step of F0 change.

Two-tailed independent *t*-tests were conducted to compare the difference between NE and CE at each intensity manipulation levels. Figure 4.11 below provides a graphic overview of the difference between the ISP values of the two groups at the five levels of intensity manipulation. The two groups are significantly different from each other in ISP values at all levels except for level 3. Significant differences are indicated by an asterisk in the graph. Statistics of the *t*-tests show that at level one and two, CE has a significantly higher ISP than NE ,and at level four and five CE has a significantly lower ISP than NE (refer to Table 4.13 for the exact statistics).

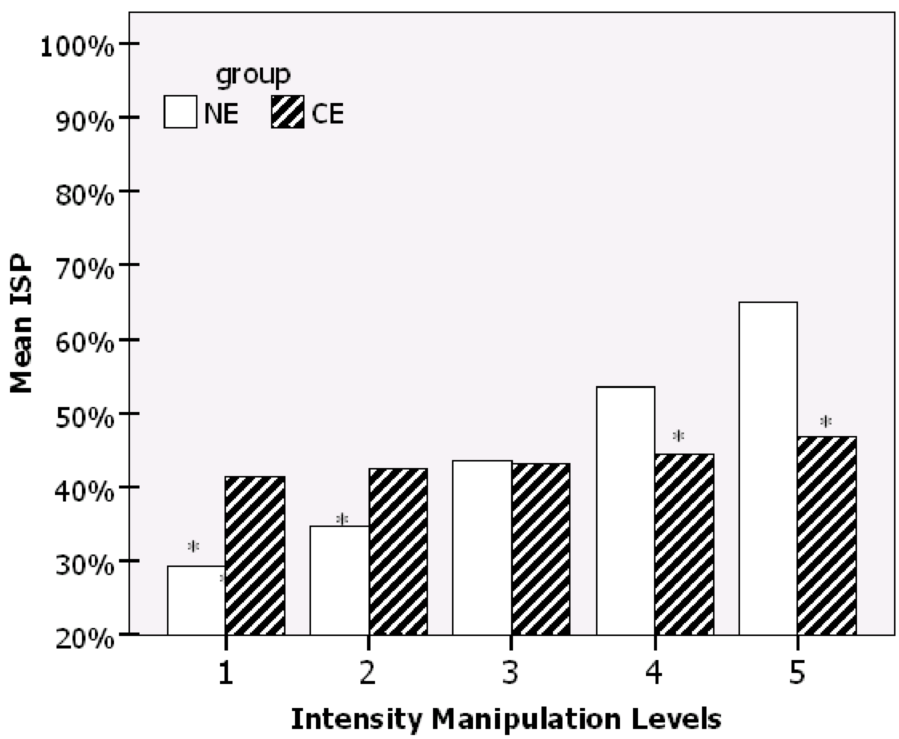


Figure 4.11 **Difference between NE and CE ISP at each level of intensity manipulation**

Table 4.13 ISP Difference between the two groups at each intensity manipulation levels

	ISP_int_1	ISP_int_2	ISP_int_3	ISP_int_4	ISP_int_5
NE	29.39%	34.61%	43.63%	53.49%	65.01%
CE	41.33%	42.59%	43.14%	44.47%	46.78%
Difference	-11.95%	-7.97%	0.49%	9.02%	18.23%
***t*-test results**	t (34.70*a.*)= -3.84, p < 0.001	t (57)= -2.76, p < 0.01	t (57)= 0.19, p > 0.05	t (34.44 *a.*)= 3.25, p < 0.01	t (39.30*a.*)= 5.38, p < 0.001

a. df adjusted as equal variance not assumed

Summary of the intensity effect.

Similar to F0 manipulation and duration manipulation, intensity manipulation was shown to have a significant effect on the ISP values of both groups. In the NE group, the difference between the ISP values at every level of intensity manipulation is different from values at other levels. On the other hand, in the CE group, the only significant difference is between level one intensity ISP and level five ISP. In other words, only a change across four levels can lead to significant change in ISP values for CE. Comparing the two groups, when the first syllable has a lower intensity value (level one and level two), the proportion of IS judgments is significantly smaller in the NE group than in the CE group. When the first syllable is manipulated to have a higher intensity value (level four and five), the proportion of IS judgments is significantly higher in the NE group than in the CE group. The two groups have similar performance when the two syllables have the same level of intensity.

4.3.6 Summary of the analyses of reliance measures

The analyses of the reliance scores and other ISP values in section 4.3 were made to reveal the difference among the weight of the three cues and the difference between the two groups in terms of reliance scores. Further exploration of the ISP values at each level of F0, duration and intensity manipulation were used to examine differences in the effect of each cue manipulation and the effects of their respective step changes. The exact differences between the two groups were investigated by comparisons at each level of each cue manipulation.

The important findings of the statistical analyses in this section include:

- The two groups differ in their reliance scores for the three cues.
  - NE's reliance score for F0 and intensity is significantly higher than that of duration.
  - CE's reliance score for F0 is significantly higher than that of duration and intensity.
  - The two groups differ in all three reliance scores. NE has a higher duration and intensity reliance score while CE has a higher F0 reliance score.
- F0 manipulation has a significant effect on the ISP changes in both NE and CE.
  - NE has a higher ISP value for a higher level of F0 manipulation, except for the contrast between level one and level two. NE has an

insignificantly lower ISP value at level two than level one.

- CE has a same pattern as NE, except for the fact that CE has a higher ISP value at level two than level one, although the difference is insignificant.
- The two groups differ in all levels except for level three. NE has a higher ISP at level one and two and a lower ISP at level four and five.

- Duration manipulation has a significant effect on the ISP changes in both NE and CE.
  - NE has a higher ISP value for a higher level of duration manipulation. The increase is significant from level one to level two and from level two to level three but insignificant from level three to level four and level four to level five.
  - CE also has a higher ISP value for a higher level of duration manipulation. But the increase is not significant for any single step change. The only significant difference is found when duration manipulation increases by two, three or even four steps, i.e. between the ISP at level one and the ISP at level two, level three, and level four.
  - The two groups only differ significantly in level four and five of duration manipulation.
- Intensity manipulation has a significant effect on the ISP changes in both NE and CE.
  - NE has a higher ISP value for a higher level of intensity manipulation. Each step change in intensity manipulation resulted in a significantly

higher ISP value in the NE group.

- CE has a higher ISP value for higher levels of intensity manipulation. Except for the step change from level one to level five, none of the step changes resulted in a significant change in ISP values in the CE group.
- The two groups differ at all levels except for level three. NE has a lower ISP at level one and two and a higher ISP at level four and five.

The analyses of ISP values have provided important insights into the contribution of the three acoustic cues and the difference between the two groups. However, admittedly, all tests were conducted on secondary data, namely, the Initial Stress Percentage data. The ISP values were based on the average of the 25 NE participants and 34 CE participants. In the following analysis, logistic regression was applied to the original dichotomous data to offer robust evidence for our findings.

4.4 Logistic Regression

Morrison (2005) presents a different method for analyzing data of cue-weighting studies, logistics regression. Logistic regression is becoming "a standard and widely used method of analyzing proportional data" (Morrison 2005, 599). In L1 speech perception studies, researchers (Benkí 2001, Nearey 1990, 1997) have used this method in data analysis. One of the advantages of this method is that all collected data contribute to the analysis, and more importantly, logistic regression can be applied directly to dichotomous data collected from each subject without any form of transformation. This point will be elaborated on below.

4.4.1 Introduction to the method of logistic regression

Logistic regression fits a statistical model to response data using an iterative maximum likelihood technique to derive estimates of coefficients for all the cues in an equation. "The coefficients represent the tuning of the response by the stimulus properties" (Morrison 2005: 599). The stimulus-tuned coefficients indicate how fast the response changes from one type of response to the other type of response as the acoustic cues changes from one end to the other. Each coefficient can be exponentiated and results in an odds ratio (Exp(B)). The odds ratio for a predictor represents how much more (or less) likely one type of response is when this predictor changes by one unit (if the predictor is a continuous factor) or when this predictor is present (as in contrary to absent) in the stimulus (if the predictor is categorical). The odds ratios can be used to compare the acoustic cues manipulated in the experiment (F0, duration and intensity in this study) with each other directly and also used to compare with those from other studies (Benkí 2000).

The purpose of this study is to compare the importance of the three cues F0, duration and intensity in the perception of lexical stress. At the same time, we want to compare the importance of each cue in the perception of stress by native speakers of English and Chinese learners of English. Using logistic regression, we can build a model based on the observed responses with F0, duration and intensity to predict the probability of Initial Stress. The coefficient for each predictor represents the tuning effect of this predictor. The larger a coefficient is, the more important the predictor is for the predicting power of the model. In other words, by comparing the coefficients of different predictors, we can judge the weight of each predictor in making predictions. By comparing the coefficient for the same predictor in two models, we can evaluate the weight of this predictor in the two groups.

The model for the present study can be expressed in the following way:

$$P(InitialStress) = \frac{1}{1 + e^{-(b_0 + b_1 * F0 + b_2 * Dur + b_3 * Int + \varepsilon_i)}}$$

In this model, *P(InitialStress)* is the probability of a token being perceived as having IS. *b0* is a constant and ε_i is a residual term. *F0*, *Dur* and *Int* are the three predictors used in the present model. *b1*, *b2* and *b3* are the coefficients for the three predictors, respectively. These coefficients are estimated by fitting the model to the response set through an iterative maximum likelihood method. This means, in order to choose the appropriate coefficient for each predictor, repeated attempts have to be made so that predictions fit the observed data best, i.e. so that the greatest possible number of responses is correctly predicted. This daunting process can be realized easily in the Statistical Package for the Social Sciences (SPSS).

When a model is built, we need to examine three things. First, we need to check if the model is a significant fit of the observed data, expressed by the significance of the chi-squared distribution and how much variance in the response can be explained by the model, expressed by the R^2. Second, we have to evaluate the coefficient for each predictor and see if each predictor makes a significant contribution to the model. This is expressed by the Wald statistic for each coefficient. Third, we have to note down the direction and magnitude of the change in the odds of making an IS judgment resulting from one unit change of each of the predictors. This is termed the odds ratio, i.e. Exp (b), for each predictor.

There are two ways to conduct logistic regression analysis. The first one is a grouped model, and this is to build one model with the responses from all the participants of one group. The second way is to build an individual-based logistic model, which is to build a model for each participant and then to estimate group statistics using all models from one group of participants. The grouped logistic regression can serve as a rough evaluation of different predictors, and the individual-based logistic model can offer more accurate information (Benkí 2000). In the following paragraphs, the group-based logistic regression results are introduced first, followed by the results of individual-based logistic regression analysis.

4.4.2 Group logistic regression results for NE and CE

Two analyses were conducted in SPSS to build two logistic regression models, one for the NE group and one for the CE group. In the analysis of either group, all the responses provided by the group participants were submitted to the analysis. In the NE group, 9319 out of 9375 possible responses (375 tokens * 25 participants) were used and, as in the reliance measures, the timed-out responses were excluded. In the CE group, 12631 out of 12750 possible responses (375 tokens * 34 participants) were used in the construction of the model, and once again, the timed-out responses were excluded. The responses provided by each group were used as the outcome variable, and the three predictors, F0, duration and intensity, were chosen for the construction of the model. All three predictors were entered into the model simultaneously. In the discussion of the group logistic regression results, it was first examined whether a statistically reliable model was built for each group and if each predictor made a unique and significant contribution to the model and finally the coeffecient values were compared across the three predictors in each group and across the two groups.

To assess the performance of the two models built, Table 4.14 cross-tabulates the observed response types with the predicted response types. An overall of 82.9% of the CE responses were predicted correctly and 69.0% of the NE responses were predicted correctly. Although the prediction for FS is better than the IS in both groups, the overall prediction percentages for both groups are larger than 50%.

Table 4.14 Classification table of the logistic regression model for the NE group and the CE group

			Predicted by the model		
	Observed Data		IS	FS	Percentage Correct
NE	IS	4240	2656	1584	62.64
	FS	5079	1301	3778	74.38
			Overall Percentage		69.04
CE	IS	5567	4227	1340	75.93
	FS	7064	826	6238	88.31
			Overall Percentage		82.85

Table 4.15 below shows the detailed statistics for the two different models. Both models constructed are statistically reliable (for NE, χ^2 (3) =1995.710, p < 0.001, for CE, Model χ^2 (3) = 6448.815, p < 0.001). The models constructed with the set of three predictors are a good fit of the data (for NE, R^2 = .258; for CE, R^2 = .536). The larger R^2 value for CE indicates that the CE model is capable of explaining a larger amount of variance found in the data than the NE model.

The Wald statistic for the coefficient (B) for each predictor indicates whether the predictor made a significant contribution to the model. It can be concluded from Table 4.15 that in both models, the contributions made by the three cues are all significant at the level of 0.001.

Table 4.15 Statistics of the NE group logistic regression model and the CE group logistic regression model

	Predictors	**B.(SE)**	**Wald Statistic (Sig.), *df*=1**	**Exp(B)**	**95% CI for Exp(B)**	
					lower	**higher**
NE	F0	.583 (0.018)	1097.504 (***)	1.792	1.731	1.855
	Duration	.186 (0.017)	125.531 (***)	1.204	1.166	1.244
	Intensity	.443 (0.017)	665.265 (***)	1.557	1.505	1.610
	Constant	-3.862 (0.104)	1387.831 (***)	.021		
	Model χ^2 (3) = 1995.710 (***), R^2(Nagelkerke) = .258					
CE	F0	1.358 (0.023)	3618.609 (***)	3.888	3.720	4.064
	Duration	0.063 (0.017)	13.912 (***)	1.065	1.031	1.101
	Intensity	0.092 (0.017)	29.078 (***)	1.096	1.060	1.133
	Constant	-4.967	2048.798	.007		
	Model χ^2 (3) = 6448.815 (***), R^2(Nagelkerke) = .536					

*** $p < 0.001$

The odds ratio (Exp(B)) is similar to the predictor coefficient (B) but offers a more direct indication of the change in odds as a result of *one* unit change of the predictor (Field 2005, 225). When Exp(B) is larger than 1, then the odds of the outcome occurring (IS, in this case) increases as the predictor increases. If Exp(B) is smaller than 1 then the relationship is in the opposite direction. By comparing the odds ratio across the three cues and across the two groups, we can examine the contributions made by the three predictors to the models. In both groups, F0 has the largest odds ratio, intensity the second, and duration the smallest odds ratio. Comparing the contributions made by each predictor across groups, the effect of F0 is much larger in the CE model than in the NE model. On the other hand, the contributions made by duration and intensity are larger in the NE group.

The logistic regression results presented above are consistent with the results of the analyses of reliance measures. The analyses of the reliance scores shows that both F0 and intensity have a significantly higher reliance score than duration in the NE group, and F0 has a significantly higher reliance score than duration and intensity in the CE group. These differences are confirmed in the logistic analysis. Comparing the weight of each cue across the two groups, the analyses of reliance measures show that NE has a significantly lower F0 reliance score and significantly higher duration and intensity reliance score than CE. The results of logistic regression agree with these results. The odds ratio in the logistic regression models shows that the CE group is twice as likely as the NE group in judging a token as having IS if the F0 predictor increases by one unit.

4.4.3 Individual-based logistic regression analysis

Section 4.4.2 above presents the results of logistic regression on group-based responses. In this section, to evaluate the performance of each individual, logistic regressions were conducted on each of the CE and NE participants using SPSS binary logistic regression. Each participant's valid responses (responses that are not timed-out) to all the stimuli (375) were used as the outcome/dependent variable in the analysis, and F0, duration and intensity were used as three predictors. Each predictor was treated as a continuous variable with five levels. As there are 59 participants in total, 59 models were built. For each model, we examined whether the model was a significant fit of the data by the chi-squared statistic and the R2 reported by SPSS, whether each coefficient made a significant contribution to the model (significant level of Wald statistic), and the coefficient for each predictor. The results for NE participants are introduced first, followed by the results of CE participants. Finally, the coefficients for each cue were compared across the two groups.

Results for individual NE participants

Twenty-five models were built for NE participants. The complete statistics are listed in Appendix IV-1. Table 4.16 below summarizes the important statistics for the models and for the predictors. The table is organized according to NE participant ID number (column 1). Column two lists the model χ^2 and the significance level of the model. The Nagelkerke R^2 is listed in the third column. The coefficient for each predictor and the significance level of this predictor as determined by the Wald statistic (Wald statistics are provided in Appendix IV-1) are reported in the order of F0, duration and intensity.

Table 4.16 Important statistics of 25 individual NE logistic regression models

	Model Statistic			Predictor Coefficient Statistic					
	Model χ^2, df =3(sig.)		Nagelkerke R^2	F0 B_F0 (sig.)		Duration B_Dur (sig.)		Intensity B_Int (sig.)	
NE01	121.855	(***)	0.377	0.288	(**)	0.240	(**)	0.942	(***)
NE04	15.607	(**)	0.055	0.036	(NO)	-0.005	(NO)	0.297	(***)
NE05	163.012	(***)	0.477	1.140	(***)	0.296	(**)	0.407	(***)
NE06	267.208	(***)	0.682	0.360	(**)	0.380	(**)	1.893	(***)
NE09	87.264	(***)	0.280	0.412	(***)	0.031	(NO)	0.685	(***)
NE10	95.163	(***)	0.308	0.797	(***)	0.234	(**)	0.234	(**)
NE11	42.026	(***)	0.146	0.472	(***)	0.179	(*)	0.110	(NO)
NE14	102.607	(***)	0.321	0.674	(***)	0.369	(***)	0.453	(***)
NE15	54.674	(***)	0.191	0.616	(***)	0.111	(NO)	0.112	(NO)
NE16	398.133	(***)	0.878	4.255	(***)	0.556	(**)	0.525	(**)
NE17	91.360	(***)	0.298	0.722	(***)	0.235	(**)	0.409	(***)
NE20	172.442	(***)	0.493	1.005	(***)	0.515	(***)	0.738	(***)
NE21	220.410	(***)	0.595	1.380	(***)	0.595	(***)	0.826	(***)
NE22	70.218	(***)	0.249	0.513	(***)	0.118	(NO)	0.601	(***)
NE23	88.328	(***)	0.287	0.691	(***)	0.359	(***)	0.261	(**)
NE24	125.235	(***)	0.379	0.819	(***)	0.204	(*)	0.575	(***)
NE25	372.836	(***)	0.842	3.449	(***)	0.067	(NO)	0.022	(NO)
NE26	75.095	(***)	0.242	0.604	(***)	0.177	(*)	0.345	(***)
NE28	137.324	(***)	0.458	1.145	(***)	0.304	(**)	0.694	(***)
NE30	92.592	(***)	0.294	0.608	(***)	0.377	(***)	0.443	(***)
NE32	101.778	(***)	0.321	0.321	(***)	0.256	(**)	0.786	(***)
NE33	97.575	(***)	0.309	0.489	(***)	-0.101	(NO)	0.703	(***)
NE34	125.366	(***)	0.384	0.803	(***)	0.073	(NO)	0.642	(***)
NE35	33.436	(***)	0.126	0.390	(***)	0.128	(NO)	0.312	(***)
NE37	42.450	(***)	0.145	0.000	(NO)	0.286	(***)	0.432	(***)

Note: The second column presents the Model χ^2 statistic, with the degree of freedom of 3, which determines the significance of the change in –2 Log Likelihood (an estimation of the goodness of fit for the model) when the three predictors are added to the model. The significance level is indicated in parentheses following the χ^2 value. The Nagelkerke R^2 represents how much variance can be predicted in the response for the three predictors. B is the logistic coefficient for each predictor. B_F0, B_Dur and B_Int correspond to b1, b2 and b3 in the equation above, respectively. The significance of the effect of each predictor determined by the Wald statistic is included in parentheses after each B value. For all significance levels, *** indicates that the value is significant at 0.001, ** significant at 0.01 and * at 0.05.

Table 4.16 shows that for all the 25 models, 24 are significant at the level of .001, and each corresponding R^2 indicates a good fit of the model. NE04's model is significant only at the level of 0.01. The R^2 value for NE04's model is only 0.055 and may only account for 5.5% of the variance in the participant's response. For most NE participants (23 out of 25 participants), the predictor of F0 contributes significantly to the model. Similarly, the predictor of intensity contributes significantly to the model built for most NE participants (22 out of 25). In over two thirds of NE participants' models (17 out of 25), duration is also a significant factor. In other words, for most NE participants, all three acoustic cues make significant contributions to their judgments of stress location.

Results for individual CE participants

Thirty-four CE participants' responses were submitted to the logistic regression test and thus 34 models were built. The complete statistics are listed in Appendix IV-2. Table 4.17 summarizes the important statistics for the models and statistics for the predictors.

Table 4.17 Important statistics of 34 individual CE logistic regression models

	Model Statistic			Predictor Coefficient Statistic					
	Model χ^2, df =3(sig.)		Nagelkerke R^2	F0 B_F0 (sig.)		Duration B_Dur (sig.)		Intensity B_Int (sig.)	
CE01	363.169	(***)	0.832	3.242	(***)	0.088		-0.066	
CE02	306.273	(***)	0.744	2.284	(***)	0.030		0.046	
CE03	154.950	(***)	0.454	1.124	(***)	0.282	(**)	0.137	
CE04	209.190	(***)	0.589	1.564	(***)	-0.125		0.019	
CE05	164.829	(***)	0.482	1.220	(***)	-0.003		0.194	(*)
CE06	404.873	(***)	0.884	4.478	(***)	0.268		0.059	
CE07	378.657	(***)	0.857	3.919	(***)	-0.001		-0.070	
CE08	382.919	(***)	0.854	3.750	(***)	0.093		-0.208	
CE15	84.489	(***)	0.322	1.040	(***)	0.074		0.150	
CE17	237.507	(***)	0.631	1.681	(***)	-0.041		0.073	
CE18	77.164	(***)	0.251	0.498	(***)	-0.118		0.539	(***)
CE19	113.156	(***)	0.348	0.897	(***)	-0.052		0.125	
CE21	184.014	(***)	0.525	1.352	(***)	0.108		0.127	
CE22	102.805	(***)	0.330	0.875	(***)	0.083		0.043	
CE23	216.865	(***)	0.596	1.588	(***)	0.223	(*)	0.100	
CE24	46.926	(***)	0.167	0.553	(***)	0.022		0.197	(*)
CE26	345.054	(***)	0.805	2.846	(***)	0.384	(**)	0.259	
CE27	164.066	(***)	0.481	1.197	(***)	0.019		0.076	
CE28	406.383	(***)	0.887	5.117	(***)	0.097		-0.303	
CE29	373.956	(***)	0.843	3.471	(***)	0.045		0.225	
CE30	243.227	(***)	0.649	1.744	(***)	-0.046		-0.030	
CE32	422.418	(***)	0.914	5.047	(***)	0.187		-0.187	
CE36	156.194	(***)	0.467	1.166	(***)	0.024		-0.018	
CE41	299.659	(***)	0.746	2.378	(***)	0.106		0.145	
CE42	185.585	(***)	0.527	1.329	(***)	0.170		0.174	
CE46	361.560	(***)	0.842	3.469	(***)	0.373	(*)	-0.102	
CE55	94.501	(***)	0.303	0.809	(***)	-0.028		-0.086	
CE56	18.716	(***)	0.067	0.297	(***)	0.054		0.148	
CE60	158.566	(***)	0.467	1.157	(***)	0.186		0.144	
CE61	345.054	(***)	0.805	2.846	(***)	0.384	(**)	0.259	
CE62	180.595	(***)	0.514	1.273	(***)	-0.022		0.302	(**)
CE64	139.505	(***)	0.431	1.137	(***)	0.002		0.000	
CE67	209.546	(***)	0.581	1.488	(***)	0.335	(**)	0.109	
CE68	373.288	(***)	0.843	3.495	(***)	0.208		-0.109	

The table above shows that all 34 models of CE participants were statistically reliable (all significant at the level of 0.001), and all corresponding R^2 are comparatively large. For all CE participants, the predictor F0 cue makes a significant contribution (at the level of 0.001). Duration is a significant predictor for six out of the 34 CE participants' responses. Intensity is only significant for four out of the 34 CE participants. In other words, for most CE participants, only F0 contributes significantly to their judgments of stress position.

Comparison of coefficients

The coefficients for the models of individual NE and CE participants were listed in 6.5.2.1 and 6.5.2.2. When we compare the significance of the coefficients for the three predictors in each group, it is found that F0 is a significant predictor for all CE participants, while duration and intensity only serve as significant predictors for very few CE participants. In the NE group, both F0 and intensity are important predictors for the majority of NE participants, and duration contributes to most NE participants' stress judgments. When we compare the significance of each predictor across the two groups, it is obvious that F0 serves as an important and unique predictor for all CE participants and for most NE participants. On the contrary, duration and intensity both make important contributions only to most NE participants but not CE participants.

In order to quantify such differences between the three predictors in each group and between the two groups, the coefficients from all participants are compared with different statistical tests.

Paired-sample *t*-tests were used to compare the three predictors within each group. The three pairs were F0 coefficient with duration coefficient, F0 with intensity, and duration with intensity.

The results showed that the NE group has a significantly higher F0 coefficient than duration coefficient ($t(24) = 3.219$, $p < 0.01$) and also a significantly higher intensity coefficient than duration coefficient ($t(24) = 4.150$, $p < 0.001$). But the difference between F0 coefficients and intensity coefficients is not significant ($t(24) = 1.362$, $p > 0.05$). In the CE group, F0 has significantly higher coefficients than both duration ($t(33)= 8.827$, $p < .001$) and intensity ($t(33) = 8.043$, $p < 0.001$). The difference between duration and intensity coefficients is not significant ($t(33) = 0.727$, $p > 0.05$).

Three one-way ANOVA Welch's[2] tests were conducted between groups, one on the each predictor's coefficients. Welch's F-tests were chosen because the equality of variances in coefficients cannot be assumed (Morrison 2005). The results show that F0 coefficients for the CE individuals are significantly higher than those of the NE individuals ($F(1, 57) = 16.076$, $p < 0.001$). NE participants

[2] Welch statistic can be used to test for the equality of group means and it is preferable to the F statistic when the assumption of equal variances does not hold.

have significantly higher duration ($F(1, 45.393) = 10.948$, $p < 0.01$) and intensity ($F(1, 30.926) = 34.606$, $p < 0.01$) coefficients than CE.

The results of the logistic regression analyses on individual NE and CE participants' responses are consistent with the results of the logistic regression based on group responses and they are also consistent with the results of the analyses of reliance measures. The advantage of the logistic regression on each participant is that it offers a chance to observe the use of the three cues by individual English native speakers and Chinese learners of English. Individual logistic regression analysis shows that 15 out of the 25 NE participants used all three cues simultaneously in deciding the stress position, but none of the 34 CE participants used a combination of all three cues in stress judgments. Furthermore, by looking at the results for individual CE participants, it is found that most participants choose to ignore both duration and intensity cues. When they do make use of any cues other than F0, it is either duration *or* intensity.

4.5 Word Form

If we compare the statistics for the two logistic models constructed for the NE group and the CE group in Table 4.15, it is obvious that the two R^2s are quite different. The R^2 is an estimation of the proportion of the variation in the response that can be explained by the constructed model, and the larger the value, the more the variation can be explained by the model, to a maximum value of 1. NE model has a R^2 value of .258 while CE has a much higher R^2 of .536. If only 25.8% percent of the variation in NE participants' responses can be explained by the three predictors, and 53.6% of the variation in CE participants' responses can be explained by the three predictors, could there be other factors that contribute to the explanatory power of the models, especially in the case of the NE model? As there are three word forms used in the experiment, *latmab*, *nizdit* and *tetsep*, it is possible that the different word forms influence the perception of stress. In the following sections, the influence of word forms on the ISP change in the two groups is discussed and the predictive power of word forms is investigated by logistic regression.

4.5.1 The influence of word form on ISP

In Figure 4.12, the ISP change as a function of F0 (the first row), duration (the second row) and intensity (the third row) are plotted for the two groups, separately. The left column is for the NE group, and the right column is for the CE group. In every graph, the ISP changes are split for the three word forms and plotted with different dashes. The concrete line represents the change in ISP when only the 125 tokens with the word form of *latmab* are included. The dotted line summarizes the ISP changes of the tokens with the word form of *nizdit*, and the dashed line is for tokens with the word form of *tetsep*.

The six graphs in Figure 4.12 show a consistent effect of word form across the manipulation of the three acoustic cues and also across the two groups. All the tokens with the word form of *latmab* have the highest IS percentage. This is true regardless of the acoustic cue under consideration and is true in both groups. On the other hand, the ISP computed from the tokens with the word form of *nizdit* is the lowest, when the ISP is organized along the manipulation levels of F0, duration or intensity. This form of tokens has the lowest percentage of IS in both groups. The tokens with the word form *tetsep* have ISP values between that of *latmab* and *nizdit*, and are closer to the value of *nizdit*. This, again, is true across the three acoustic cues and across the two groups.

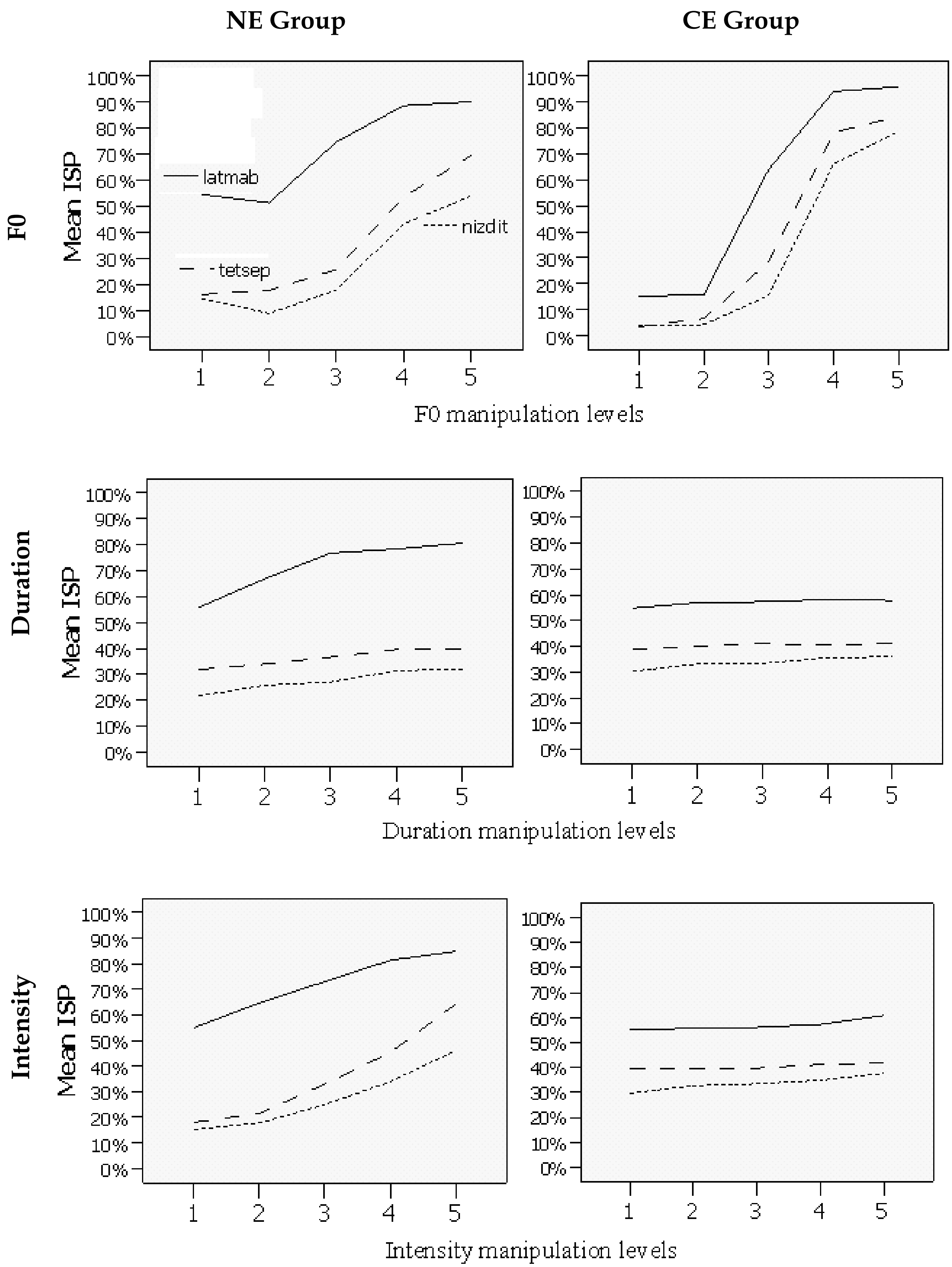


Figure 4.12 ISP change as a function of F0, duration and intensity manipulation for NE and CE, split into three word forms.

Comparing the three graphs down the left column with the right column, it is clear that the difference between the three words is larger in the NE group than in the CE group. In order to investigate the interaction between word form and group, a mixed-model ANOVA was conducted, with the repeated-measures on word-form and Group as the between-group factor. Both word form ($F(1, 248)= 717.183$, $p < 0.001$) and the interaction between word-form and Group ($F(1, 248) = 75.377$, $p <0.001$) are significant. The effect of Group is non-significant ($F(1, 248) = .18$, $p > 0.05$). The interaction is examined in more detail by *t*-tests between every pair of word forms in the two groups and between two groups for each of the word forms.

4.5.2 Pairwise t-test between two word forms

The pairwise comparisons between *nizdit* and *tetsep*, *nizdit* and *latmab*, and between *tetsep* and *latmab* in the NE group indicate that *tetsep* has a significantly higher ISP than *nizdit* ($t(124) = 8.112$, $p < 0.001$). The word form *latmab* also has a significantly higher ISP than *nizdit* ($t(124) = 28.370$, $p < 0.001$) and *latmab* has a significantly higher ISP than *tetsep* ($t(124) = 21.298$, $p < 0.001$). The same trend is also observed in the CE group and all differences are significant at the level of 0.001 (*tetsep* > *nizdit*, $t(124) = 7.485$; *latmab* > *nizdit*, $t(124) = 15.942$; *latmab* > *tetsep*, $t(124) = 15.311$).

4.5.3 Pairwise t-tests between the two groups for each word form

The only difference that could affect the interaction is found to exist in the pairwise comparisons between the NE and the CE group for the word form of *latmab*. The NE group has a significantly higher ISP than the CE group for the tokens with the form of *latmab* ($t(213.790) = 3.792$, $p < 0.001$)). For the other two word forms, the two groups' difference in ISP is not significant (*nizdit*, $t(221.248) = -1.709$, $p > 0.05$, *tetsep*, $t(237.255) = -0.963$, $p > 0.05$).

4.5.4 The Predictive Power of Word Form in Logistic Regression Models

Now that we have identified the influence of word form in the change of ISP, we want to see if including word form in the logistic regression analysis would increase the predictive power of the models. We include word form as a predictor in the construction of the logistic model for the NE group and the CE group, respectively. In the models, word form was included as a categorical predictor with three levels, *nizdit*, *tetsep* and *latmab*, and is deviation coded. The results showed that the addition of word form as a predictor improves the predictive power of the models.

Table 4.18 compares the classification power of the models without word form and with word form for the two groups. An overall higher percentage of responses was predicted correctly by the model with word form. The improvement is more obvious for the NE model, as the overall correct percentage increased from 69% to 77.3%. The CE model only increased by less than 2%, from 82.9% to 84.6%.

Table 4.18 Classification table of the logistic regression models for the NE group and the CE group with and without word form as a predictor

			Predicted by the model without *word form*			Predicted by the model with *word form*		
	Observed Data		IS	FS	Percentage Correct	IS	FS	Percentage Correct
NE	IS	4240	2656	1584	62.64	3074	1166	72.50
	FS	5079	1301	3778	74.38	954	4125	81.22
	Overall Percentage				69.04			77.25
CE	IS	5567	4227	1340	75.93	4708	859	84.57
	FS	7064	826	6238	88.31	1086	5978	84.63
	Overall Percentage				82.85			84.60

Similarly, in Table 4.19, which offers the detailed statistics of the two models, it can be seen that the R^2 for the two models increases over the models without word form as a predictor. The R^2 for the NE model has increased from .258 to .451 and the R^2 for the CE model has increased from .536 to .592. The improvement brought about by word form is more obvious in the NE model than in the CE model.

Table 4.19 Statistics of the NE group logistic regression model and the CE group logistic regression model with word form as a predictor

	Predictors	B.(SE)	Wald Statistic (Sig.), *df*=1	Exp(B)	95% CI for Exp(B)	
					lower	higher
NE	F0	.733 (0.021)	1236.683 (***)	2.080	1.997	2.167
	Duration	.232 (0.019)	154.654 (***)	1.261	1.216	1.308
	Intensity	.554 (0.020)	781.098 (***)	1.740	1.674	1.809
	Word Form		1459.226 (***), *df* = 2			
	nizdit	- 1.040 (0.039)	717.721 (***)	0.354	0.328	0.381
	tetsep	- 0.499 (0.036)	186.704 (***)	0.607	0.565	0.652
	latmab	1.538 (0.041)	1423.157 (***)	4.657	4.300	5.045
	Constant	- 4.807 (.123)	1527.289 (***)	0.008		
	Model χ^2 (5) = 3834.343 (***), **R^2(Nagelkerke) = .451**					
CE	F0	1.503 (0.025)	3520.275 (***)	4.496	4.278	4.724
	Duration	0.069 (0.018)	15.186 (***)	1.072	1.035	1.110
	Intensity	0.099 (0.018)	31.012 (***)	1.104	1.066	1.144
	Word Form		805.833 (***), *df* = 2			
	nizdit	- 0.797 (0.037)	454.358 (***)	0.451	0.415	0.489
	tetsep	- 0.253 (0.036)	50.279 (***)	0.776	0.724	0.832
	latmab	1.050 (0.038)	751.761 (***)	2.859	2.652	3.082
	Constant	-5.486 (0.120)	2104.425 (***)	.004		
	Model χ^2 (3) = 7367.547 (***)， **R^2(Nagelkerke) = .592**					

*** $p < 0.001$

Table 4.19 also shows the odds ratio of each predictor in the models. The odds ratio (Exp (B)) for each level of word form represents the odds of this word form relative to the geometry mean of all the word forms (Benkí 2001). When the odds ratio is greater than one, it is positively correlated with the predictor, and the larger the odds ratio is, the larger the contribution made by the predictor. When the odds ratio is smaller than one, it is negatively correlated with the predictor and the smaller the odds ratio is, the larger the contribution made by the predictor. The results revealed that both *nizdit* and *tetsep* have a negative correlation with the event of initial stress, but *latmab* is positively correlated with IS. In the NE model, the odds ratio for *latmab* is the largest among all predictors and it is even larger than the odds ratio for the F0 predictor in the CE model. In the CE model, the odds ratio for *latmab* is also larger than the other two word forms but not as large as F0. In both groups, the effect of the *nizdit* is smaller than *latmab* and in the opposite direction. The effect of *tetsep* is the smallest among the three word forms and in the same direction as *nizdit*. If we compare the odds ratio of each word form across the two groups, the predictive power of every word form is smaller in the CE group than in the NE group.

4.5.5 Summary of the influence of the word forms on stress perception

Statistical analyses indicated that in both groups, *latmab* led to the highest number of ISP values, and *nizdit* led to the lowest number of ISP values. The two groups differ only in their responses to one word form. NE participants have a significantly higher ISP value than CE participants in response to tokens with the word form of *latmab*. The ISP values of both groups of participants are similar for the word form of *nizdit* and *tetsep*. This is confirmed by the logistic regression analysis of the predictive power of the three word forms. The predictive power of *latmab* in the NE model is larger not only than all predictors in the same model, including F0, duration, intensity and two other word forms, but also larger than the predictive power of F0 in the CE model. The predictive power of the other two word forms is not very different for the two groups.

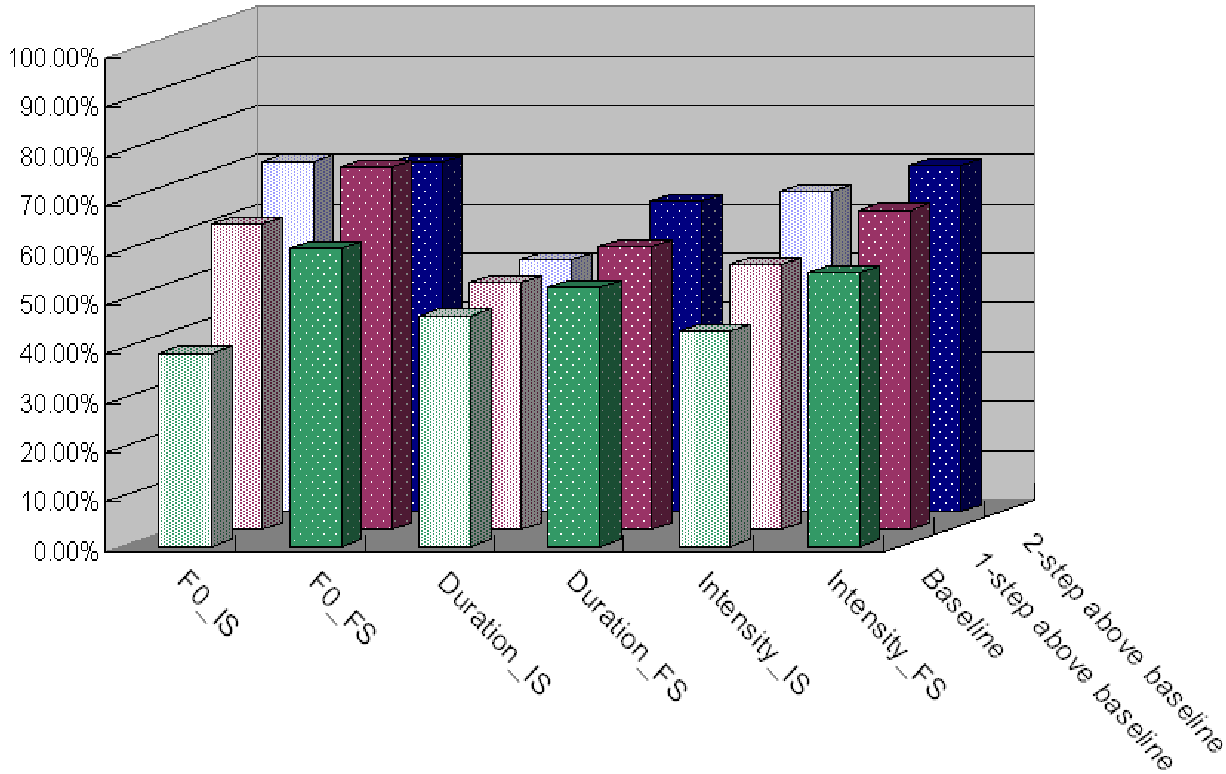
Word form was not planned to be the target of investigation in the present study. We have included three words forms in the present study so that our findings are not specific to one vowel or one type of vowel. Despite the detailed comparisons among the words forms and the difference between the two groups for each word form, the effects of the three word forms are consistent for the three different cues and consistent for the two groups, as can be seen in Figure 4.12 above. Thus, it is concluded that the difference among the three word forms will not change the general findings of this study.

4.6 Syllable Position

Section 4.5 above discusses the possible influence of word form on stress judgments. It was found that although the three word forms showed different effects on IS or FS judgments, no interaction between the three word forms and the three acoustic cues on stress judgments was observed. This section is focused on the interaction of the acoustic cues and syllable position in stress perception by native speakers and Chinese learners of English; that is, to examine if the three acoustic cues would have a stronger effect on stress judgments on the first or on the second syllable. If level three is treated as the baseline level, the question to explore here is whether the effects of F0, duration and intensity are symmetrical for the first syllable and the second syllable. In order to evaluate if the powers of F0, duration and intensity manipulations are the same on the first syllable and on the second syllable, the Percentage of Initial Stress (ISP) judgments when the first syllable is acoustically stronger (level 4 and 5) is compared to the Percentage of Final Stress (FSP) judgments when the second stress is stronger (level 1 and 2 manipulation). Table 4.20 shows how the original data were recoded to compare the difference between the two stress positions.

Table 4.20 Restructuring data to compare the difference between IS and FS

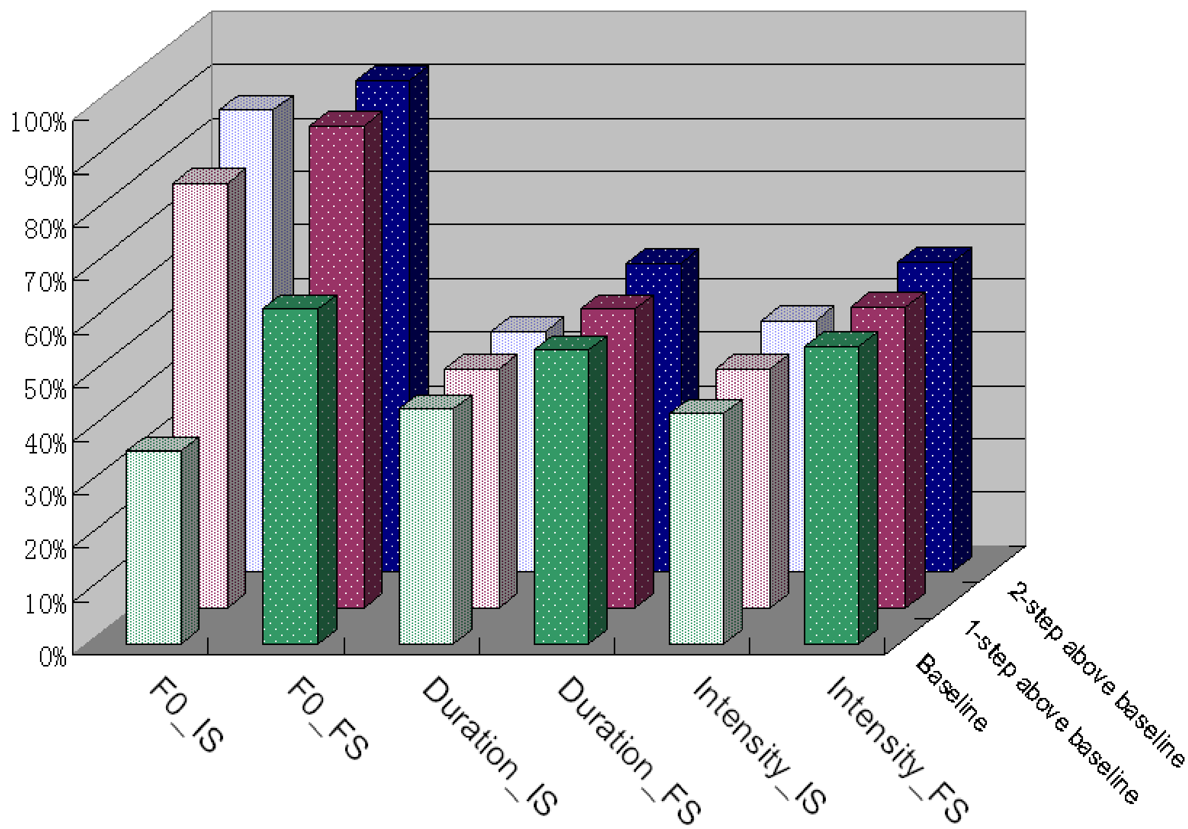
	IS	**FS**
Baseline	percentage of IS judgment at level 3	percentage of FS judgment at level 3
1 step above baseline	percentage of IS judgment at level 2	percentage of FS judgment at level 2
2 steps above baseline	percentage of IS judgment at level 1	percentage of FS judgment at level 1



	F0 _IS	F0 _FS	Duration _IS	Duration _FS	Intensity _IS	Intensity _FS
Baseline	39%	60%	47%	53%	44%	55%
1-step above	62%	73%	50%	57%	54%	65%
2-step above	71%	71%	51%	63%	65%	70%

Figure 4.13 NE ISP and FSP at baseline, 1 step above and 2 steps above baseline manipulations of F0, duration and Intensity

Figure 4.13 here shows the ISP and FSP at the baseline level as well as when the initial syllable or the second syllable was manipulated to be one step or two steps above the baseline from the NE group. Figure 4.14 shows the ISPs and FSPs for the CE group. The results for F0, duration and intensity are listed separately. The set of lighter bars in the figure represents the ISP values when the first syllable is acoustically equal or stronger than the second one, and the set of darker bars represents FSP when the second syllable is acoustically equal or stronger.



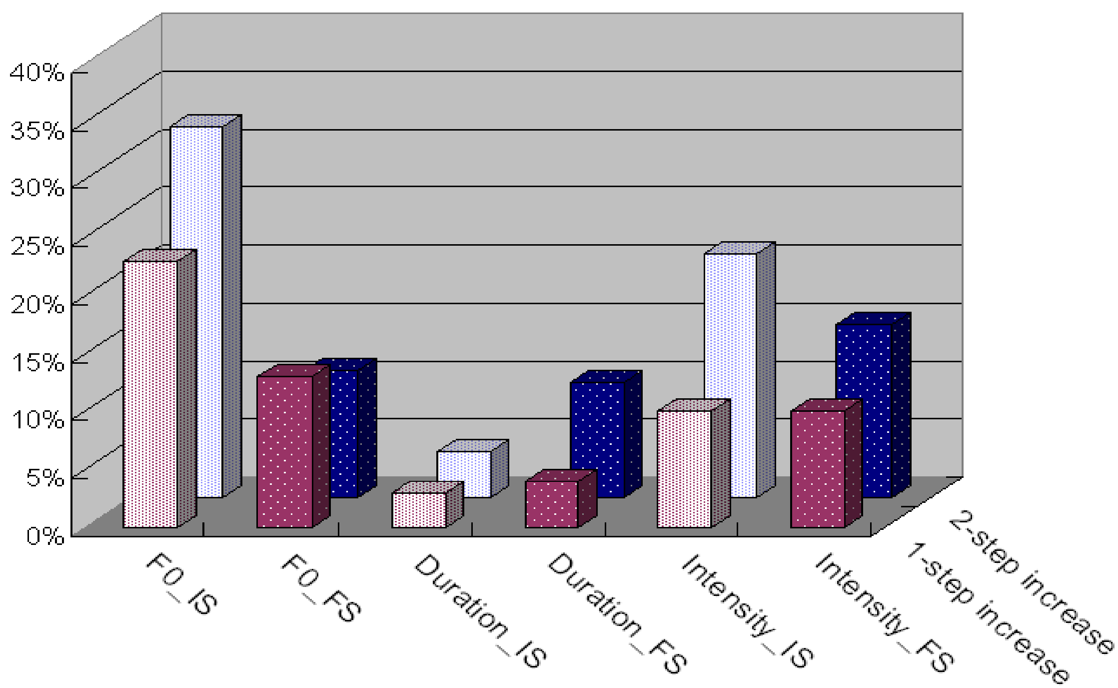
	F0 _IS	F0 _FS	Duration _IS	Duration _FS	Intensity _IS	Intensity _FS
Baseline	36%	63%	44%	55%	43%	56%
1-step above	79%	90%	45%	56%	45%	56%
2-step above	86%	92%	45%	58%	47%	58%

Figure 4.14 CE's ISP and FSP at baseline, 1-step above and 2-step above baseline manipulations of F0, duration and Intensity

At a glance, it seems that for both NE and CE, FSP is almost always higher than ISP at each level of the cue manipulation, with the exception of only one case. F0_IS is equal to F0_FS when the manipulation is 2 steps above baseline. However, the direct comparison of the ISP and FSP at different levels makes the result susceptible to the influence of first or second syllable preference that has been observed for different word forms. From the baseline ISP and FSP in the above figure (the first row of bars), it is clear that listeners tend to judge a token to be final-stressed even if the acoustic cues are neutral.

To reduce the effect of perceptual preference that was not a result of acoustic manipulation, it is preferable to compare the net increase of percentage as a result of the increase in manipulation values. Thus, the baseline percentage was subtracted from percentage at 1 step or 2 steps above baseline. The result shows how many *more IS* judgments have been made as a result of the one step or two-step increase of acoustic manipulations on the first syllable or how many *more FS* judgments have been made as an increase of acoustic manipulations on the second syllable.

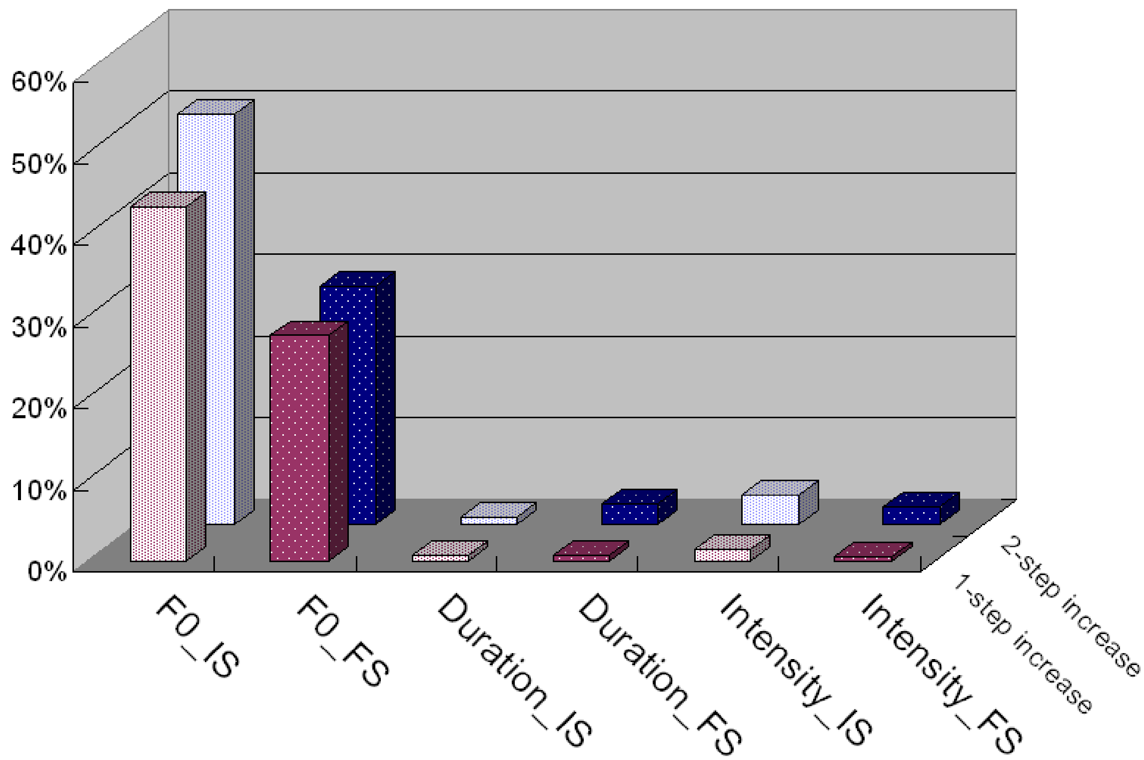
Figure 4.15 presents the net increase from the baseline ISP or FSP as a result of acoustic manipulation of F0, duration and intensity on either the initial syllable or the final syllable for the NE group. Again, the set of lighter bars in this figure represents the increase in ISP when the first syllable is manipulated to be stronger and the set of darker bars represents the increase in FSP when the second syllable is manipulated to be stronger. Different from the figure above, this figure shows that not all three acoustic cues have the same effect when manipulated on the first and second syllable. For F0 and intensity manipulation, the lighter bars are as high as or higher than the corresponding darker bars, which means that making the first syllable one or two levels stronger than the second syllable led to a larger increase in the number of IS judgments. For duration manipulation, on the other hand, the darker bars are higher than the lighter bars, which may suggest that a longer syllable in the second position is more likely to be perceived as a stressed syllable than a longer syllable in the first position.



	F0 _IS	F0 _FS	Duration _IS	Duration _FS	Intensity _IS	Intensity _FS
1-step increase	23%	13%	3%	4%	10%	10%
2-step increase	32%	11%	4%	10%	21%	15%

Figure 4.15 NE net increase in ISP or FSP as a result of 1-step or 2-step increase in F0, duration, and intensity manipulations

Figure 4.16 presents the net increase in the CE group. Similar to the NE group, the increase as a result of F0 manipulation on the first syllable is larger than the increase as a result of F0 manipulation on the second syllable. However, for the CE group, the dark bars for the FSP increase and the light bars for the ISP increase donot show much difference for duration and intensity manipulation.



	F0 _IS	F0 _FS	Duration _IS	Duration _FS	Intensity _IS	Intensity _FS
1-step increase	43%	28%	1%	1%	1%	1%
2-step increase	50%	29%	1%	3%	4%	2%

Figure 4.16 CE net increase in ISP or FSP as a result of 1-step or 2-step increase in F0, duration, and intensity manipulations

4.6.1 Mixed-model analysis of the effects of stress position and group

In order to examine if the above mentioned differences are significant, statistical analyses were conducted. A two-way mixed-model ANOVA was used with the group as the between-subject factor and stress position as the within-subject factor. Each pair of ISP and FSP was tested. The following table summarizes the six tests that were conducted. Group was shown to have a main effect on all six tests, which was no surprising giving the differences between the two groups of participants presented in previous sections. Stress Position was shown to have a main effect in both steps of F0 manipulation but only in the 2-step manipulation of duration and intensity. The group and stress position interaction had a significant effect on the percentage increase as a result of 2-step intensity manipulation.

Table 4.21 Results of two-way mixed-model ANOVA test on the comparison between stress positions and groups

	Stress Position (IS vs. FS)	Group	Group × Stress Position
F0: 1-step increase	* $F(1, 57) = 16.91$, $p < 0.001$	* $F(1, 57) = 33.6$, $p < 0.001$	non-sig. $F(1, 57) = 1.10$, $p = 0.30$
F0: 2-step increase	* $F(1, 57) = 38.29$, $p < 0.001$	* $F(1, 57) = 34.90$, $p < 0.001$	non-sig. $F(1, 57) = 0.001$, $p = 0.98$
Duration: 1-step increase	non-sig. $F(1, 57) = 0.55$, $p = 0.46$	* $F(1, 57) = 13.70$, $p < 0.001$	non-sig. $F(1, 57) = 0.72$, $p = 0.40$
Duration: 2-step increase	* $F(1, 57) = 9.21$, $p < 0.005$	* $F(1, 57) = 22.22$, $p < 0.001$	non-sig. $F(1, 57) = 3.18$, $p = 0.08$
Intensity: 1-step increase	non-sig. $F(1, 57) = 0.33$, $p = 0.57$	* $F(1, 57) = 38.86$, $p < 0.001$	non-sig. $F(1, 57) = 0.01$, $p = 0.4$
Intensity: 2-step increase	* $F(1, 57) = 11.32$, $p < 0.005$	* $F(1, 57) = 56.08$, $p < 0.001$	* $F(1, 57) = 4.67$, $p < 0.05$

4.6.2 Pairwise t-test between the ISP and the FSP in the two groups

As Group was shown to be a significant factor in all six models, further analyses were conducted for the NE group and the CE group separately. For each group of participants, a paired-sample *t*-test was conducted between ISP and FSP. The following table summarizes the results of the two groups.

Table 4.22 Pair-wise comparisons between net ISP and FSP increases in the NE and CE group

	NE (df = 24)	CE (df = 33)
F0: 1-step increase	* t = 2.373, p < 0.001	* t = 3.62, p < 0.005
F0: 2-step increase	* t = 4.833, p < 0.05	* t = 4.33, p < 0.001
Duration: 1-step increase	non-sig. t = -0.94, p = 0.36	non-sig. t = 0.09, p = 0.93
Duration: 2-step increase	* t = **-2.68**, p < 0.05	non-sig. t = **-1.15** , p = 0.26
Intensity: 1-step increase	non-sig. t = 0.32, p = 0.76	non-sig. t = 0.52, p = 0.61
Intensity: 2-step increase	* t = 3.29, p < 0.005	non-sig. t = 1.01, p = 0.32

For the NE group, the F0 manipulation on the first syllable leads to a larger increase in IS judgments than the increase in FS judgments. This is true both when the F0 is only 1 step above and 2 steps above baseline. Intensity shows a similar effect in the NE group, although the 1-step deviation from the baseline hasnot created a significant effect between the ISP and FSP. The effect of duration manipulation is opposite to F0 and intensity. The duration manipulation leads to a smaller increase on the first syllable than on the second syllable, although the difference between the first and second syllable is not significant at 1-step manipulation. The CE group shows a similar general tendency. F0 and intensity manipulation has a stronger effect on the second syllable, but the difference between the first and second syllable is not significant in the case of intensity manipulation. The 2-step duration manipulation leads to a higher increase on the second syllable, which is similar to the NE group. However, the differences between the first and second syllable as a result of duration manipulation at both steps are not significant.

4.6.3 Summary of the effects of syllable position

The general observation from previous sections has indicated that the acoustically stronger syllable is more likely to be perceived as the stressed syllable. This section discusses how this effect interacts with the position of the syllable under consideration. Will the increase of a cue have a same effect when it is manipulated on the first syllable and the second syllable? The data collected and discussed in this section suggest the following two important points.

- A first syllable with higher F0 and intensity is more likely to be perceived as a stressed syllable than a second syllable with higher F0 and intensity.
- A second syllable with longer duration is more likely to be perceived as a stressed syllable than the first syllable with a longer duration.

The two groups, NE and CE, are similar in terms of the general pattern but show considerable difference for all three cues. The positive effect of F0 on the first syllable is stronger in the CE group than in the NE group, as a result of CE's predominant reliance on F0. The difference between the two syllables in terms of duration and intensity manipulation is much less obvious in the CE group than in the NE group.

4.7 Conflicting Cues

In this research, 125 different combinations of cue settings were used (5 F0 ×5 Duration × 5 Intensity). Among these different combinations, some tokens have all three cues manipulated in the same direction, while some other tokens have conflicting cues. For example, on a certain token with conflicting cue settings, F0 may be manipulated to be higher on the second syllable, but duration is longer and intensity is higher on the first syllable. In this case, if F0 is being relied on, then the token can be perceived to have FS, but if duration and intensity are being actively used, then the token can be perceived to have IS.

In this section, listeners' responses to tokens with conflicting cues are analyzed to reveal the effect of cues when they are working against each other on a token. It is shown in previous results that F0 has the strongest influence on both NE and CE perception of stress position, the influence of intensity ranks second, and duration has the weakest influence. Thus, three comparisons are interesting. The first comparison is made on tokens with F0 conflicting both duration and intensity. Thus, the effect of F0 is compared against the composition forces of duration and intensity. The second comparison is made on tokens with F0 and intensity conflicting with each other while duration is controlled at the neutral level. The third comparison is on tokens with duration and intensity conflicting with each other while F0 controlled at the neutral level.

4.7.1 F0 against duration and intensity

In order to compare the F0 against duration and intensity as conflicting cues, 12 tokens were chosen for the analysis. These 12 tokens consisted of four types of manipulation settings and three word forms of each setting. In the following tables, Table 4.23, the first row indicates the level of F0 and the first column indicates the level of duration and intensity manipulation. The cells in white represent the tokens investigated. For example, the cell in the second row indicates tokens with a manipulation setting of F0 at level five and both duration and intensity settings at level one. Three tokens (three word forms) with the same setting contributed to the value of this cell. The value is calculated as the average native speaker ISP over the three tokens. Similarly, three other cells are calculated, F0 at level four and duration and intensity both at level two, F0 at level four and duration and intensity both at level four, and F0 at level five and duration and intensity both at level five. Due to the limited number of tokens for each comparison, no statistical analysis has been conducted. For each cell, if the average ISP value is higher than 50%, then it is taken to indicate that the cue with a higher level of setting (larger number) is the cue in control in this conflicting situation. If the cell value is below 50%, it suggests that the cue(s) with a lower setting is/are in control.

For example, in the first cell of the NE table in Table 4.23, with F0 at level five and both duration and intensity at level one, the cell value is only 45%, below 50%. This cell indicates that duration and intensity together have a stronger influence on stress perception than F0, because if F0 is a stronger cue, then it would lead to more than 50% of participants perceiving IS on the tokens. The observation of the four cells in the NE table shows that there is no consistent winner. The composition forces of duration and intensity have a larger influence in three cells, but F0 alone still has a larger influence when it is manipulated to be level two and both duration and intensity are at level four. The asterisk beside a number indicates that the cue with this setting is stronger than its conflicting cue. For CE participants, on the other hand, F0 always serves as a stronger cue when it is in conflict with both duration and intensity. In the CE table, when F0 is at level four and five, there is a high ISP value, despite the fact that both duration and intensity settings support FS judgments. When F0 is set at level one and two, the ISP values drop to around 10%, even though duration and intensity cues support IS judgments.

Table 4.23 ISP values for the four tokens with conflicting cues, F0 against duration and intensity, in the NE and CE group

NE

F0 D&I	1	2*	3	4	5
1*					45%
2*				49%	
3					
4		44%			
5*	56%				

CE

F0 D&I	1*	2*	3	4*	5*
1					72%
2				77%	
3					
4		8%			
5	11%				

4.7.2 F0 against intensity

As F0 and intensity rank first and second for NE speakers in stress judgments, the power of each of the two cues when they are in conflict is studied in this section. The tokens used for this comparison have their duration manipulation set to level three, the neutral level, and the F0 and intensity on these tokens have the opposite settings. The NE grid in Table 4.24 indicates that although F0 is shown to be a stronger cue in general, intensity is still effective in two of the four cases: when intensity is at level two and F0 is at level four and when intensity is at level five and F0 is at level one. CE participants, again, rely solely on the F0 information with the presence of a conflicting intensity cue.

Section 4.7.1 above and section 4.7.2 here suggest that although F0 is the strongest cue for NE, the interaction of this cue with other conflicting cues may lead to a complex picture of NE judgments. For CE participants, F0 is so strong a cue that it takes over even when it is in conflict with one or even two other cues.

Table 4.24 ISP values for the four tokens with conflicting cues, F0 against intensity, in the NE and CE group

NE

F0 / I	1	2*	3	4	5*
1					55%
2*				45%	
3					
4		36%			
5*	55%				

CE

F0 / I	1*	2*	3	4*	5*
1					83%
2				73%	
3					
4		13%			
5	15%				

4.7.3 Intensity against Duration

The last comparison is made between tokens with conflicting intensity and duration cues while F0 is controlled at the neutral level. Table 4.25 Duration is a stronger cue when it is at level one and two and supports FS judgments, and intensity is a stronger cue when it is at level one and two and supports FS judgments. In other words, when a token has conflicting duration and intensity cues, listeners tend to rely more on the cue on the second syllable in stress judgments. Another finding for CE participants has also emerged when the F0 cue is controlled for. Similar to NE, duration is a stronger cue for CE when it is at level one and two and in conflict with intensity, as the ISP values are low. Intensity is a stronger cue for CE when it is at level one and two and in conflict with duration.

Table 4.25 ISP values for the four tokens with conflicting cues, duration against intensity, in the NE and CE group

NE

I D	1*	2*	3	4	5
1 *					47%
2 *				39%	
3					
4		29%			
5	29%				

CE

I D	1*	2*	3	4	5
1*					29%
2*				29%	
3					
4		37%			
5	41%				

4.7.4 Summary of the effects of conflicting cues

Section 4.7 presents the analyses conducted on tokens with conflicting cues. One general finding is that for the NE participants, there is no cue that is a consistent winner when it is in conflict with other cues. This is true even for the strongest cue, F0. When F0 is in conflict with intensity or both intensity and duration, NE listeners may choose to rely on intensity or duration for stress judgment. In stress perception, NE participants donot simply choose acoustic correlates based on the ranking of these correlates but use a complicated strategy that involves trading relationships among the correlates. CE listeners, on the other hand, always rely on the higher ranking F0 cue whenever it is present. Cues that are in conflict with F0 are ignored. When F0 is controlled for, the contribution of duration and intensity to stress perception in CE group can be revealed.

Chapter Five

Discussion

In chapter four, we have seen that the analyses of the responses to the perception task reveal interesting findings concerning the effects of the three acoustic cues on native English speakers and Chinese learners of English. The analyses have also identified the differences between the two groups in stress perception. This chapter is organized to explore these findings. We will first look at the five hypotheses proposed in chapter three and discuss whether they can be confirmed or not with our findings (section 5.1). Then we will delve into the two groups of participants' performances in the perception task in section 5.2 and 5.3. Section 5.4 discusses the implications for L2 stress acquisition theories based on the comparison of the two groups in this study and also on studies with other learners. In the final section, we attempt to address the unexpected effect of word form and offer an explanation for such an effect.

5.1 Hypotheses Testing

At the end of chapter three, five hypotheses were proposed based on the review of the literature and the design of the present study. Analyses were conducted on the responses by Chinese learners of English and native English speakers in chapter four. The results were presented in relation to the different weightings of the three acoustic cues in stress perception and the difference between the two groups in their reliance on the three cues. Now, we want to use these results to evaluate the five hypotheses proposed.

Hypothesis One

Previous research on stress perception has shown that F0, duration and intensity are all acoustic correlates of stress for native English speakers. It is thus hypothesized that the systematic manipulation of F0, duration and intensity will lead to systematic changes in the performance of NE stress judgments.

The analysis of the Initial Stress Percentage (ISP) values shows that F0, duration and intensity manipulations have a significant effect on the change of ISP in NE participants' responses. As the F0, duration and intensity steps increase, the ISP value increases. The only exception is that when F0 increases from manipulation level one to level two, the ISP value decreases. However, the difference is proven to be insignificant. The group logistic regression analyses show that F0, duration and intensity manipulation steps are all positively correlated with the possibility of Initial Stress (IS) judgments. As F0, duration or intensity increases, the possibility of IS judgments increases in the NE group. Thus, hypothesis one is confirmed.

Hypothesis Two

The null hypothesis for the CE group is that the systematic manipulation of F0, duration and intensity will lead to systematic changes in the performance of CE stress judgments.

The analysis of the ISP values shows that F0, duration and intensity manipulations have a significant effect on the changes of ISP in CE participants' responses. As the F0, duration and intensity steps increase, the ISP value increases. The group logistic regression analysis confirms this and shows that F0, duration and intensity manipulation steps are all positively correlated with the possibility of Initial Stress (IS) judgments. As F0, duration or intensity increases, the possibility of IS judgments increases in the CE group. Hypothesis two is also confirmed.

Hypothesis Three

Based on the classic studies of Fry (1955, 1958, 1964), it is hypothesized that the effect of F0 manipulation is the most obvious one among the three acoustic correlates for native English speakers in perceiving stress, followed by duration, which is, in turn, more important than intensity.

The analysis of the reliance scores shows that F0 has the highest score in the NE group. Intensity has the second highest reliance score. There is no significant difference between these two scores in the NE group. Both scores are significantly higher than the reliance score for duration. In the group logistic regression model for NE, F0 has the highest odds ratio (Exp(B)), followed by intensity, and then duration. Hypothesis three is confirmed in that F0 manipulation has the most significant effect on NE stress judgments. However, the prediction about the order of importance for duration and intensity is defeated.

Hypothesis Four

The null hypothesis for the Chinese learners of English is that they resemble native English speakers in the order of importance of the three cues.

The analysis of the reliance scores shows that F0 has the highest score in the CE group. The CE reliance score for intensity is higher than that of duration, but both are significantly lower than that of F0. The group logistic regression analysis shows that F0 has the highest odds ratio (Exp(B)) in the CE model. The odds ratios for both intensity and duration are very low, although intensity is slightly higher than duration. The individual logistic regression analysis shows that while F0 contributes to the prediction of every individual CE participant's stress judgment, duration and intensity manipulation only contribute to the prediction of stress judgment for very few CE participants. Thus, the individual logistic analysis suggests that CE only relied on the manipulation of F0 but not on duration and intensity in stress perception. Hypothesis four is not confirmed. CE resembles NE in their reliance on F0 as a cue in stress perception, but unlike

NE participants they donot also involve duration and intensity as active cues for stress perception.

Hypothesis Five

Previous studies have shown the predominant influence of F0 in tone perception in Chinese. Thus, it is hypothesized that the effect of F0 manipulation is a more important cue for Chinese learners' stress judgments than for native English speakers.

As far as the F0 cue is concerned, CE has a significantly higher F0 reliance score than NE. This is further supported by the comparison between the two groups at each step of F0 manipulation. When the second syllable is higher in F0, significantly less CEs perceived the stress on the first syllable than the NEs. When the first syllable is higher in F0, significantly more CEs perceived the stress on the first syllable than NEs. Group logistic regression analysis shows that the F0 odds ratio for the CE group model is more than twice as much as that for the NE group model. Hypothesis Five is confirmed.

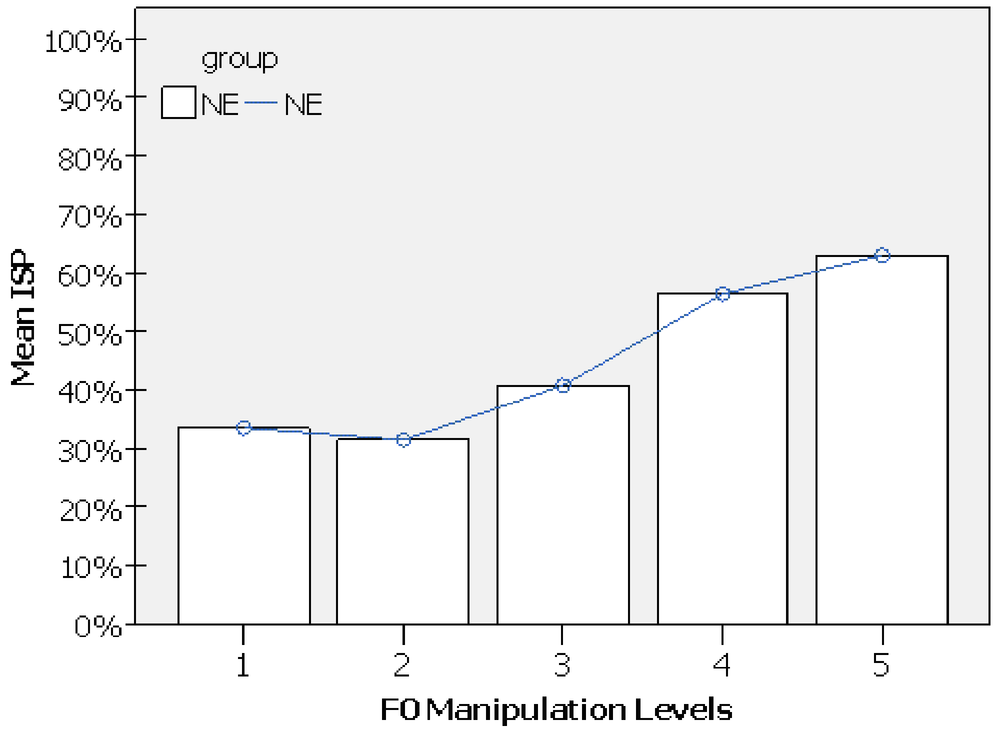
5.2 Native Speakers of English

This section is devoted to the discussion of the native speakers' performance in the perception tasks. We first analyze the effects of F0 for native speakers and then offer explanations for the unexpected ranking of duration and intensity in native speakers' stress perception.

5.2.1 F0 cue

The results of this study have shown that F0 serves as the most important correlate for NE's perception of stress. The reliance score for F0 is the highest compared to duration and intensity in the NE group. The group logistic regression analysis shows that the odds ratio for F0 is the largest among the three cues in the NE model. Furthermore, the individual logistic regression models of NE have added further evidence for the strong effect of F0 in the NE group. In all 25 individual models constructed, F0 makes a unique and positive contribution to the prediction of stress position for 23 individuals. We see a consistent effect of F0 in all three analyses.

This result is not surprising for the NE group. All of the research discussed in the literature review chapter indicate that F0 is the most important cue for NE speakers' stress perception (Fry 1958, Lehiste 1970, Bolinger 1965, Lieberman 1960). In order to reveal the exact nature of the importance of F0 in stress perception, the step change of F0 was studied.



F0 difference b/t V1&V2(Hz)	Level 1: -50	Level 2: -25	Level 3: 0	Level 4: 25	Level 5: 50
Mean	28.27%	25.87%	39.31%	61.71%	70.99%
difference between adjacent levels	-2.40%	13.44% (*)	22.40% (*)	9.28%(*)	

Figure 5.1 Bar graph for the effects of F0 step change for the NE group

There is one surprising finding in the analysis of F0 data in the NE group. It was found that while each step change of F0 manipulation from level two to level five resulted in an increase of ISP values in the NE group, the step change from level one to level two led to a slight drop in ISP value, from 28.27% to 25.87% (see Figure 5.1). Although the direction of this change is not expected, the exact difference between the two levels is insignificant.

One explanation for this finding is the “all-or-none” effect found by Fry (1958) and Morton & Jassen (1965) for F0 manipulation. Both researchers indicated that one step above the baseline can make the syllable be perceived as stressed. Increasing the difference between the two syllables in F0 may not bring a stronger effect. In the present study, level three is the baseline level where the two syllables have the same F0 values. At level two, the first syllable is manipulated to be 25Hz lower than the second syllable. This F0 difference between the two syllables is enough to lead to a significant drop (13.44%) of IS judgments in the NE group. Manipulating the first syllable to be 50Hz lower than the second syllable at level one resulted in insignificant changes in the ISP value (-2.4%). In other words, we can conclude that level one and level two F0 manipulations have equal effects on NE stress perception.

Another interpretation for the drop of the ISP value as F0 manipulation increased from level one to level two is proposed by Xu (2008, presentation feedback) at the 4th Speech Prosody conference in Brazil. He suggested that as F0 is used as a major acoustic correlate not only for lexical stress but also for intonation, the exaggerated F0 difference may mislead native listeners into perceiving the token as carrying question intonation.

Sentence intonation is primarily cued by pitch contour in English. The intonation is added on top of the stress pattern of the word to result in a surface pitch contour that may or may not be different from the original stress-related pitch pattern on the word. Let us consider the four sentences below. There are two target words in the sentences, *window* with Initial Stress (IS) and *exam* with

Final Stress (FS). Statement intonation can be realized by aligning a falling pitch contour with the stressed syllable in the target words. Thus, on the word *window*, this is realized as the high F0 on the first syllable *win* and low F0 on the second syllable *dow*. The surface pitch contour is consistent with the original word pitch contour without sentence intonation. On the word *exam*, the falling contour of statement starts with the stressed syllable and thus the first syllable is still low but the second syllable starts high and ends low. On the other hand, question intonation in English can be realized as a rising contour aligned with the stressed syllable (e.g. Bolinger 1978). On the target word, *window*, this means that the first syllable is lower in F0 and the second syllable is raised to create rising contour. Thus, the surface F0 contour is opposite to the original F0 contour that would be expected for an IS word. On the word *exam*, the contour starts from the stressed syllable, i.e. the second syllable, and creates a rising contour on the second syllable, which is consistent with the FS pattern of the word.

Table 5.1 Realization of statement and question intonation patterns on IS and FS words

	Initial Stress (IS) *window*	Surface F0 pattern	Final Stress (FS) *exam*	Surface F0 pattern
Statement (falling)	This is a **window**.	high-low	This is an **exam**.	low-high(falling)
Question (rising)	This is a **window**?	low-high	This is an **exam**?	low-high(rising)

As can be seen from Table 5.1 above, three out of the four intonation and stress combinations result in a surface low-high F0 contour. They are: FS word with statement intonation, and FS word with question intonation, and IS word with question intonation. In other words, a surface low-high F0 contour is not necessary mapped to a FS word. It can also be mapped to an IS word in the question sentence context. These examples show that, when the first syllable was manipulated to be 50Hz lower than the second syllable in the present experiment, it is possible that some native speakers related the rising F0 to question intonation rather than to FS stress pattern. This explains why there are not fewer IS judgments for manipulation level one than for level two, as had been expected based on the F0 difference between the two levels. Thus, the F0 difference between the two syllables at level one hasnot made a significant contribution to the ISP change in the NE group, i.e. no significant difference between level one and level two ISP values.

Comparing the two interpretations, the second one is more plausible. The evidence lies in the ISP values on level four and five. If we assume the 'all-or-none' effect of F0, then we would expect to see an insignificant change of ISP at both ends of the continuum. However, the ISP value at level five is significantly higher than that at level four. The absolute difference between level three and four is 22.40%, and the difference between level four and five is 9.28%. Despite the fact that the change from step three to step four is greater than the change from step four to step five, the effects of both step changes are significant. We have to arrive at the conclusion that making the first syllable 50Hz higher than the second syllable leads to significantly more IS judgments in the NE participants, and thus reject the al-or-none explanation. This observation, on the other hand, is consistent with the second interpretation. If we consider the second interpretation, the unparalleled performance at the two ends of the continuum is explainable. While it is possible for native speakers to make IS judgment for tokens with a low-high surface F0 contour, they will not reject the IS judgments for tokens with a surface high-low F0 contour. Consider the F0 configuration on level five. These tokens were manipulated to have a first syllable 50Hz higher than the second syllable. This surface pitch contour would only lead to the prediction of IS judgment (refer to Table 5.1 above, the first cell).

5.2.2 The comparison between the effects of intensity and duration

A second point that deserves discussion for the NE is the relative importance of intensity and duration. In this study, NE group was found to have the highest reliance score for F0, which is higher than both intensity and duration. Between the two cues of duration and intensity, the intensity reliance score is found to be significantly higher than duration score. While previous researchers unanimously agree on the importance of F0 over duration and intensity in stress perception, different views are found concerning the relative weightings of duration and intensity.

Sluijter & van Heuven (1996) found that intensity differences located above 0.5 kHz to be a very important correlate of lexical stress. Lieberman (1960) studied the acoustic correlates of stress in the production of twenty-five pairs of verb-noun words in English. Each word was embedded in a carrier sentence and sixteen participants were instructed to read the whole sentence silently except for the target words. The acoustic cues of F0, peak intensity and duration were measured on both the stressed and unstressed syllables in all the test words. Among all words, the stressed syllable has a higher F0 than the unstressed syllable in 90% of the cases. The stressed syllable had a higher peak intensity than the unstressed one in 87% of the cases. But the stressed syllable is only longer than the unstressed one in 66% of the cases. This study suggests two important points. First, intensity difference between the stressed and unstressed syllables is used almost as frequently as F0 difference. Second, intensity is used more often than duration to indicate the difference between the stressed and unstressed syllables.

Although the results of the present study agree with the above literature, Fry's classic study on English stress correlates found duration to be a more robust cue than intensity for stress perception. Fry (1955) compared the weightings of duration and intensity. Duration and intensity were each manipulated at five levels, and it was found that the variation of duration led to a change of 70% in IS judgments, which is much larger than the change brought about by intensity variation, 29%. We can compare these two values with the values achieved in the present study (see Figure 5.2). The manipulation of duration in the present study led to only 15% change of ISP in the NE group and the manipulation of intensity led to a 36% change of ISP in the group. The numbers from Fry's study and the present study are presented side by side below in Figure 5.2 to offer a direct comparison of the difference between the results.

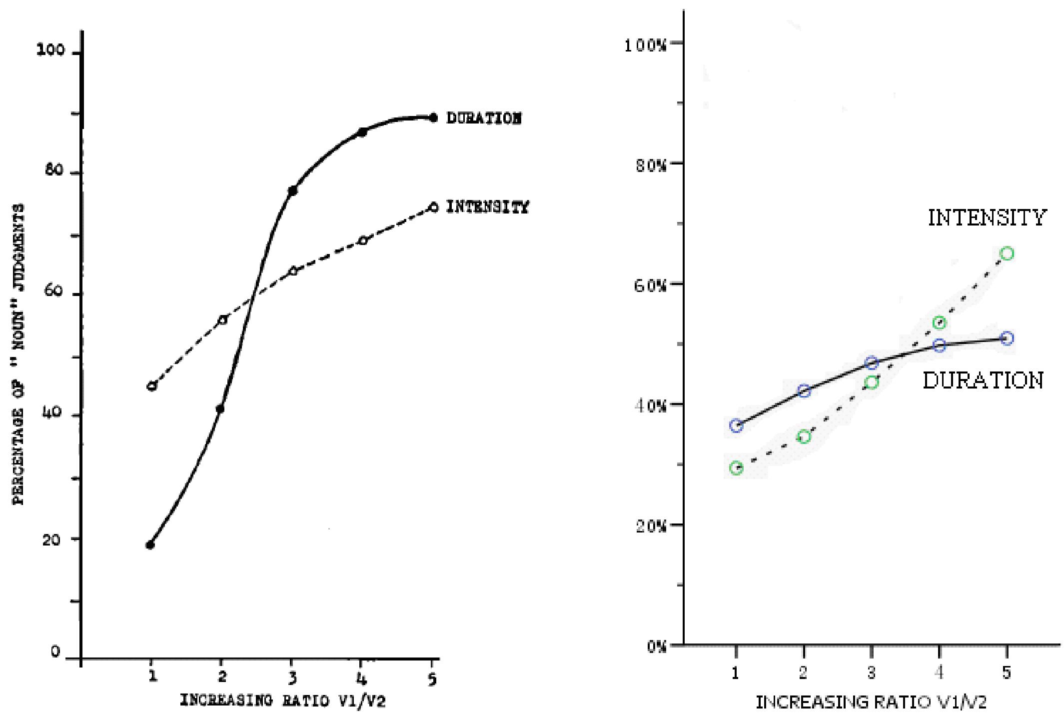


Figure 5.2 Comparison between the results of Fry (1955) and the current study on the effects of duration and intensity

One possible factor that contributes to the different results of the two studies lies in the manipulation process. In Fry's study, five real words were used, *subject, object, contract, digest* and *permit*. For the duration manipulation, no constant range was used for each word. Based on the natural production of these words, the duration range and step size were tailor-made for each word. For example, the word *digest* was pronounced by 12 native speakers with both the verb form and the noun form. The measurements of the 24 tokens (1 word × 2 stress patterns × 12 speakers) indicated that duration ratio between V1 and V2 showed a considerable overlap between the verb and noun form. For the verb form of *digest*, a range of 0.53 to 1.5 was found to exist among speakers. In other words, some speakers produced the verb form of digest with the first syllable shorter than the second one (0.53), but some speakers produced the verb form (final stress) with the first syllable longer than the second syllable (1.5). Similarly, in the production of the noun form of *digest*, the duration ratio range of V1/V2 varied from 0.83 to 2.87, which means that some speakers produced the noun form (initial stress) of digest with a longer first syllable (2.87), but there were also speakers who produced a shorter first syllable (0.83). The whole range of V1/V2 duration ratio is 0.53 to 2.87. For a different word, *contract*, the duration ratio of V1/V2 can be as low as 0.1, which means some speakers produced the verb form with the first syllable one tenth of the duration of the second syllable. For the same word with a noun form, most speakers still produced the word with the first syllable shorter than the second one. The largest ratio between V1/V2 for the noun form is only 1. 06. The range of V1/V2 duration ratio was found to be 0.1 to 1.06. As different words showed different ratio and range of duration differences between the stressed and unstressed syllable, Fry used a different range and step size for each word pair. For example, for the word, *digest*, a range of 0.53 to 2.87

duration ratio was used, and the crossover value of manipulation (equal to level three in the present study) was fixed at 1.25. The range of duration ratio used for *contract* is from 0.1 to 1.06, and the middle level was fixed at 0.5. Thus, the effect of duration that we see in Figure 5.2 above for Fry's study is a result of individualized manipulation which covers the maximized duration difference between the stressed and unstressed syllables observed in the production data. It is not surprising that the manipulation with such an extreme range resulted in an obvious effect in the perception study. Different from the manipulation of duration, a common range and same step size of intensity manipulation of intensity were used for all five word pairs, -10dB, -5dB, 0dB, 5dB and 10dB. The range adopted did not cover the range of intensity difference observed for all production data and it did not reflect the different intensity ratios for the noun form and verb form of a word. One problem with the intensity manipulation in Fry's study is that it failed to take into consideration the intrinsic intensity difference between the stressed and unstressed vowels (Lieberman 1960). In the five real words used by Fry, the first vowel is always different from the second vowel and thus has different intrinsic intensity. The same intensity manipulation may actually be perceived differently. Thus, the different weight of duration and intensity can at least partly be attributed to the different approach to duration and intensity manipulation settings.

Compared to the study by Fry, the same approach is used for both duration and intensity manipulation in the present study. A constant range of duration and intensity were used for all three word forms, and the five levels of manipulation were equally distributed between the extremes. Thus, the duration manipulation is less sensitive to word form differences compared to Fry's study. In addition, the present study used nonsense words with the same vowel in both syllables. This eliminates the intrinsic intensity differences that were not controlled in Fry's study. Although the same intensity manipulation method was used in the two studies, the effect of intensity revealed in the present study should be more accurate. Thus, the different weight of intensity and duration found in Fry's study and in the present study may at least be partly attributed to manipulation differences.

5.2.3 The influence of syllable position

Native English listeners' strategy of using stress correlates in perception is shown to be influenced by syllable position. The influence of a certain acoustic cue for stress perception can be different when the cue is higher on the *first* syllable and when the cue is higher on the *second* syllable. Both F0 and intensity have been shown in this study to result in more ISP increase when they are higher on the first syllable, compared to the FSP increase they produce when they are higher on the second syllable. The effect of duration is influenced by syllable position in the opposite direction. The increase in FSP as a result of longer duration on the second syllable is larger compared to the increase in ISP as a result of longer duration on the first syllable.

Few studies were found to have discussed the effect of syllable position on stress perception. Most studies focused on the difference between IS and FS words in production. McClean & Tiffany's (1973) study was the only one identified to focus on acoustic parameters in relation to syllable position. The discussion was only on production data of native English speakers. The study found that the stressed syllable in a disyllabic IS word was usually produced with a higher pitch than the stressed syllable in an FS word. Furthermore, the unstressed syllable in an FS word was usually produced with a higher pitch than the unstressed syllable in an IS word. In other words, the IS words usually have a comparatively higher pitch on the stressed syllable and a comparatively lower pitch on the unstressed syllable. The FS words, on the other hand, have a comparatively lower pitch on the stressed syllable and comparatively higher pitch on the unstressed syllable. Thus, the pitch ratio between the stressed and unstressed syllables on IS words would be much larger than the ratio for FS words. For duration, the exact opposite was observed. Duration ratios between the stressed and unstressed syllable were almost always higher on FS words than on IS words. Intensity was not found to have different ratios on IS and FS words.

In a later study, Sluijter & van Heuven (1996) found that in the production of the Dutch minimal pair CANnon-kaNON, the duration difference between stressed and unstressed syllables is smaller in CANnon than in kaNON. Despite the possible influence of preboundary lengthening, the authors found that the interaction between syllable position and stress is close to significant level.. Similarly, in an English stress production study (He et al. 2008), the ratio of stressed and unstressed vowel duration was found to be much larger on FS words than IS words. The duration ratio for FS was as large as 2.41 while the duration ratio for IS was only 1.10. In the same study, the pitch ratio of stressed and unstressed vowels was found to be consistently larger on IS words than FS words. Three pitch points were measured on the stressed and unstressed vowels in the same word, the beginning of the vowel, the middle point of the vowel and the final point of the vowel. All three pitch ratios calculated from the IS words (1.17, 1.31, 1.28) were larger than the respective pitch ratios calculated from the FS words (1.15, 1.16, 1.0).

In an unpublished course project, Beaton & Willing found that duration is only effective as a cue for stress perception when it is longer on the second syllable. Only by making the first syllable longer could not lead to IS judgments. Most listeners in the study did not find there to be a stressed syllable.

From the above discussion, it seems that syllable position interacts with acoustic cues in both stress production and perception. Furthermore, the interaction in production corresponds to the interaction in perception. Native speakers use a larger F0 ratio between stressed and unstressed syllables on IS words than on FS words. Their expectation of an IS word is closely related to a falling pitch contour on a disyllabic word. With such an expectation, it is reasonable to assume that native speakers will be more likely to perceive an IS pattern when a falling pitch contour is present. On the other hand, if native speakers use a longer duration ratio for FS words, then they expect a prototypical FS word to show a duration difference. Thus, when a proper duration difference between the stressed and unstressed syllable in an FS word is perceived, it serves as hard evidence for FS judgments.

The relationship between production and perception discussed above is only a hypothesis based on existing evidence. Without a more carefully controlled experiment, it is hard to draw a conclusion on the exact nature of the interaction between syllable position and different acoustic correlates.

5.3 Chinese Learners of English

Despite the findings for native speakers of English in stress perception, it is not the focus of the present study to investigate the native speakers' strategies. Native speakers were used as a baseline against which the performance of Chinese learners could be compared. From the discussion of hypothesis two, four and five above, we identified three interesting findings for Chinese learners of English in stress perception. First we discuss the question of whether Chinese learners are stress 'deaf' or not. Then, we look at Chinese learners' choice of the acoustic cues in stress perception. Finally, Chinese learners' heavy reliance on F0 is compared to the native speakers group.

5.3.1 Stress 'deaf' or not

In this part, we focus on the question of whether Chinese learners of English are 'deaf' or not to English lexical stress. From the discussion of hypothesis one and hypothesis two above, it is clear that both groups show systematic changes in their stress judgments in reaction to the changes in acoustic cues This means that, compared to native speakers of English, Chinese learners are not deaf to stress contrasts. This result is consistent with Altman & Vogel's (2002) Stress Typology Model which predicts that learners with a tonal background would have no problems in stress perception. In their experimental study, Chinese learners' performance was as good as the native speakers.

The results of this study indicate that systematic changes in CEs' stress judgments can be observed for manipulated tokens. Given the appropriate cue, Chinese learners can perceive the difference between stressed and unstressed syllables. When the first syllable is lower in F0, the chance of it being perceived as stressed is significantly lower than when it is higher in F0. If we look at the group logistic regression analysis for NE and CE more closely, it can be found that the regression model for the CE group has more predictive power than NE group model.

Table 5.2 Classification table for the NE group and the CE group with and without word form as a predictor revisited (as a repetition of Table 4.18)

			Predicted by the model* without *word form			***Predicted by the model* with *word form***		
	Observed Data		IS	FS	Percentage Correct	IS	FS	Percentage Correct
NE	IS	4240	2656	1584	62.64	3074	1166	72.50
	FS	5079	1301	3778	74.38	954	4125	81.22
	Overall Percentage				**69.04**			**77.25**
CE	IS	5567	4227	1340	75.93	4708	859	84.57
	FS	7064	826	6238	88.31	1086	5978	84.63
	Overall Percentage				**82.85**			**84.60**

Table 5.2 above shows that the overall percentage of predicted responses in the CE group is 82.85%, which is much larger than the 69.04% percent of the NE group. This difference between the two groups was sustained with the addition of word form as an extra predictor. While the predictive power of the NE model increased to 77.25%, it is lower than the 84.60%. This comparison between the

two groups shows that the CE performance in stress judgments is systematic, and even more systematic than the native speakers.

Another piece of evidence to prove that CE is similar in their ability to perceive stress comes from the perception of real English words in this experiment. One hundred real English words were used in the present study for screening purpose. In the judgments of the stress position in these words, 60% of CE participants (34 out of the 57) achieved a correct rate of over 80%, and 66% of NE participants (25 out of the 38) identified stress correctly 80% of the time or more among the 100 real English words. The independent *t*-test between the two groups shows that they were not significantly different from each other ($t(93)=1.06$, $p > 0.05$). The following boxplot (Figure 5.3) provides a direct overview of the performance of the participants in the two groups. Although NE has a higher median than CE in the correct rate measurements (the line inside the box), there is considerable overlap between the two boxes below. Furthermore, NE performance is more skewed than CE performance.

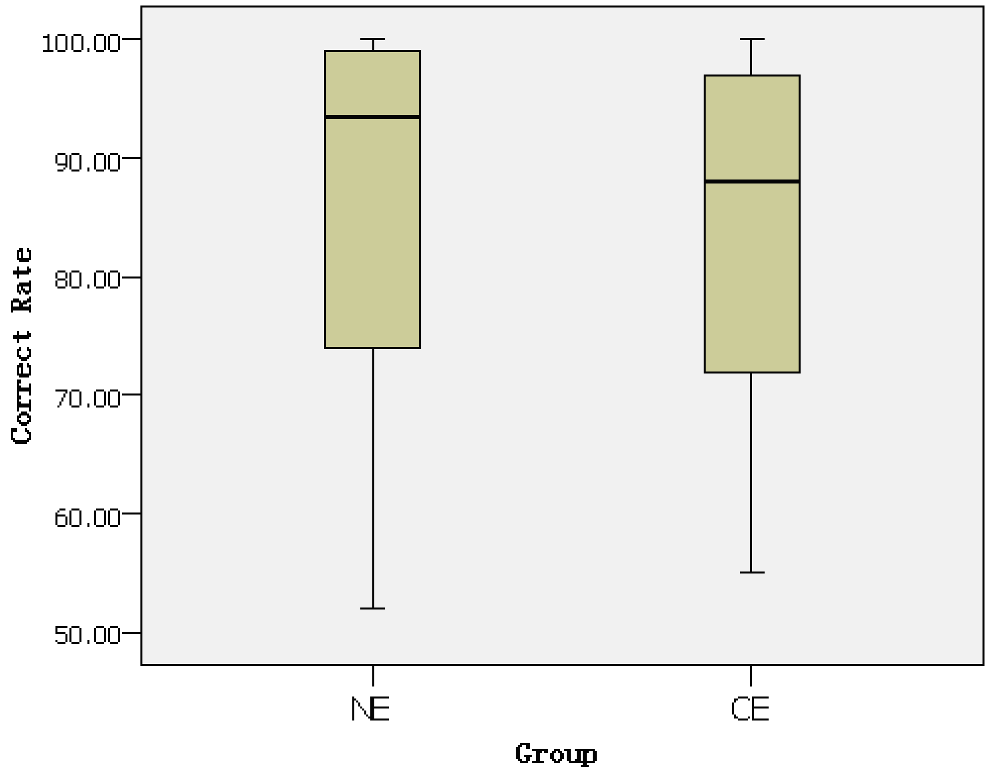


Figure 5.3 Boxplots of correct rates for the two groups in the 100 real English word stress judgments

The good performance by Chinese learners on stress position judgments can be explained by the proposals of the PAM model (Best 1995, 2001). PAM predicts that the native phonological system can affect learners' ability to correctly discriminate sound contrasts in L2 speech learning. In chapter two, we suggested that the effect of L1 phonology may be *neutral* for Non-Assimilable (NA) contrasts, when two sounds in L2 can't be assimilated to any sound category in L1. L1 phonology is predicted to have a *negative* effect for Single Category (SC) sound contrasts and Category Goodness difference (CG) sound contrasts. In the SC and CG cases, as the two sounds are assimilated to one category in L1, it would be difficult for learners to differentiate the two sounds in L2. On the other hand, L1 phonology is argued to have a positive effect on the learning of Two Category assimilation (TC) and Uncategorized-Categorized pair (UC) contrasts. Although the predictions in PAM are described in terms of phoneme contrasts, PAM can also offer valuable insights into the results of suprasegmental studies, such as the present research.

The target of investigation in this research is lexical stress. It is proposed that Chinese learners of English enjoy good performance in lexical stress discrimination as a result of the positive effect of their L1 tonal system. To interpret the results in the framework of PAM, we have to define lexical stress in terms of *contrasts*. There are two ways to define the contrast. The first way is to propose that there is a contrast between Initial Stress (IS) and Final Stress. As only disyllabic tokens are used in the present study, there are only two possible stress patterns, IS or FS. Thus, in the present experiment, participants are actually tested for their ability to discriminate between IS and FS. In accordance with this interpretation, Chinese learners' perception of the IS and FS contrast is argued to be a case of Two Category assimilation. IS is prone to be assimilated to Tone 4 in Mandarin Chinese. As IS can be characterized as having higher F0 on the first syllable and lower F0 on the second syllable, it would exhibit a falling contour of F0 over the domain of the word. Similarly, FS is prone to be assimilated to Tone 2 in Mandarin Chinese. Opposite to IS, FS would have lower F0 on the first syllable and higher F0 on the second syllable and thus exhibit a rising F0 contour. If the assumption is valid, then Chinese learners' proficiency in differentiating T1 and

T4 in their L1 should have a positive effect, making it easy for them to differentiate the two stress patterns. The second way to define the contrast in lexical stress perception is to propose that there is a contrast between the stressed and unstressed syllable(s). The stressed syllable is assimilated to a high level tone, T1 in Mandarin and the unstressed syllable can either be assimilated to a low contour tone, T3, or a neutral tone, or it can be defined as uncategorized. If the unstressed syllable is assimilated to either T3 or neutral tone, then the difference between the stressed and unstressed syllable(s) is a Two Category pair in the PAM model. If the unstressed syllable is uncategorized, then the difference between stressed and unstressed syllable is an Uncategorized-categorized pair. Both types of contrasts are predicted to be easy for L2 learners as the L1 phonology system has a positive effect. Thus, no matter how the unstressed syllable is assimilated, the L1 tone system is predicted to have a positive effect on L2 lexical stress pattern discrimination. This assumption is actually supported by external evidence from production studies introduced in the literature review. Juffs (1990) found that Chinese learners use an inordinately long T1 duration to indicate lexical stress. Chao (1980) indicated that Cantonese speakers associate high and low tones with both stressed and unstressed syllables. Also, a recent study by He, Wang & Wilshire (2008) found that initial stressed disyllabic nonsense English words were produced with high-low pitch patterns, similar to the way they produce a combination of T1 and neutral tone disyllabic nonsense Chinese words. Furthermore, a final stressed disyllabic nonsense English word was produced with a pitch pattern that is similar to the combination of a T3T2 combination.

Within the framework of PAM, it is reasonable to assume that CE learners' good performance in discriminating lexical stress patterns is the result of positive L1 transfer. However, as it is indicated in the review of this model in Chapter Two, PAM focuses on the phonological categorization without a detailed discussion of the phonetic properties contributing to the phonological categorization. The present study suggests that Chinese learners are not 'deaf' to English lexical stress differences and have no problems in differentiating between an IS word and an FS word on the level of phonology. However, this is not to suggest that Chinese learners are native-like in stress perception. As is implied in the discussion of hypothesis three, Chinese learners of English differ from native speakers in their employment of the acoustic cues in stress perception. The sections below discuss their different choice of acoustic cues and the weight of F0 in their perception.

5.3.2 Nonnative employment of acoustic cues

Despite the fact that Chinese learners can perceive stress differences, they showed a clear difference from native speakers in their employment of acoustic cues in stress perception. The nonnative use of acoustic cues is reflected in two ways: first, their reliance on one selected cue rather than on a combination of different acoustic cues in stress perception, and second, the over-sensitivity to F0 in stress perception.

Comparing hypothesis three and hypothesis four, it is found that, different from native English speakers who used a combination of three acoustic cues in stress perception, Chinese learners relied primarily on F0 but largely ignored duration and intensity in stress perception. This difference is most clearly illustrated in the analysis of individual logistic regression models. Among the 25 NE speakers, 15 participants used a combination of all three cues in stress perception. Five speakers used a combination of F0 and intensity cues. One participant used a combination of F0 and duration and one participant used a combination of duration and intensity. Only three participants chose to rely on a single cue in stress perception, two of them relied on F0 and one of them relied solely on intensity. The Chinese learners of English in the study showed a totally different picture in the choice of acoustic cues. None of the CE participants used a combination of all three cues. The majority of them, 24 out of the 34 CE participants, relied on a single acoustic cue, F0. Ten of them used a combination of two cues, and F0 was always one of the two cues. Among the ten CEs, six used a combination of F0 and duration and four used a combination of F0 and intensity. Nobody used a combination of duration and intensity.

Similar to the conclusions of previous research (e.g. McClean & Tiffany 1973), this research has shown that native speakers' stress perception is the result of "complex trading relationships between fundamental frequency, amplitude, and duraiton" (Adams & Munro 1978, 128). The trading relationships between the three cues are clearly seen in the analysis of tokens with conflicting tokens. Despite the fact that F0 ranks first as a cue to stress perception, it is not decisive for native stress perception. When F0 was competing with both intensity and duration, more native listeners judged stress position based on the intensity and duration information in a majority of the cases. Even when F0 was competing with the intensity cue, less than 50% of native listeners relied on it in half of the cases. When duration was competing again intensity, more listeners tend to rely on duration in half of the cases and more listeners tended to rely on intensity in another half of the cases. One of the purposes of the present research was to reveal nonnative speakers' different trading relationships between the three acoustic cues in stress perception. However, in contrast to the complex relationships exhibited by NE in the use of different cues, the relationship exhibited in the stress perception of tokens with conflicting cues by CE can only be described as simple. It can be summarized in one sentence. If F0 differs between the syllables, the syllable with the higher F0 value is perceived as the stressed one. This is true even when F0 is in conflict with both duration and intensity cues. Only when F0 was controlled, the limited influence of duration and intensity can emerge.

Although Chinese learners of English are able to discriminate stress position, their judgments are not built on a complex interaction of three acoustic correlates but are solely dependent on the existence of the F0 difference of the stressed and unstressed syllables. Unlike native speakers, who can switch to other cues when F0 information is absent, Chinese learners' ability to perceive stress patterns is severely compromised when F0 information is minimized.

This finding is supported by a different study (Wang & Yoon 2008) conducted to compare the performance of Mandarin Chinese learners of English (CE) and native English speakers (NE) in the perception of real English words excised from ACCENTED and UNACCENTED conditions. An accented word is a word that is focused in a sentence and an unaccented word is extracted from a sentence where a different word is focused in the sentence. Based on previous literature, the F0 difference between the stressed and unstressed syllables is exaggerated in accented conditions but is minimized in unaccented conditions. In other words, the difference between stressed and unstressed syllables in a word excised from the unaccented condition would be cued primarily by duration and intensity. Two forms of the word permit, noun (IS) or as verb (FS), were embedded in carrier sentences, either accented or unaccented in the sentence. A total of four kinds of tokens were created (2 stress patterns × 2 accent conditions). Six participants' were recorded producing the four kinds of sentences. Their productions were checked, confirming the previous finding that the F0 difference between stressed and unstressed syllables is obvious for accented words and that there is little F0 difference between the stressed and unstressed syllables of a word that is unaccented.

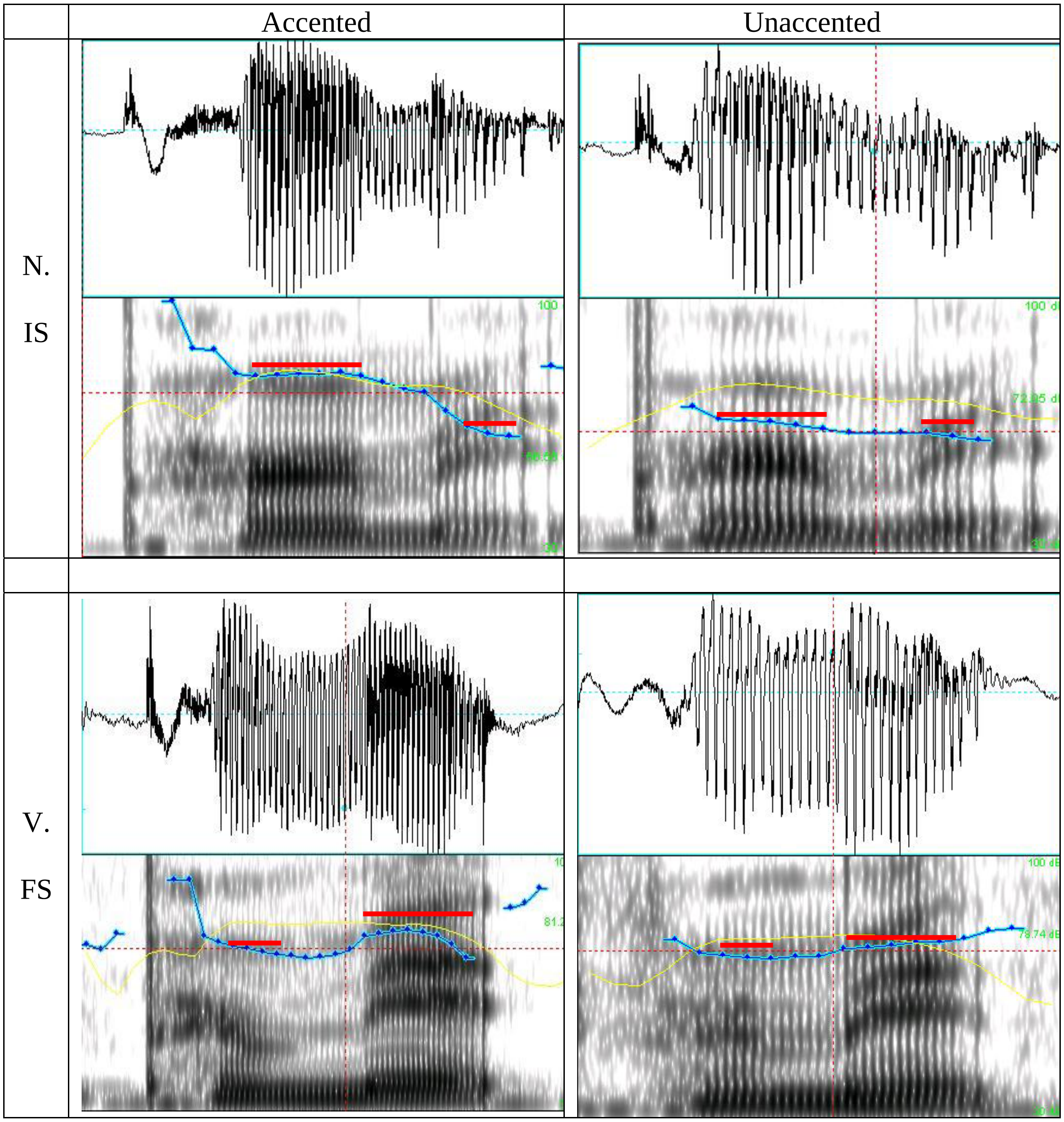


Figure 5.4 Spectrogram and waveform of the word *permit* as a noun and as verb in accented and unaccented condition

Figure 5.4 shows the waveform and spectrogram of the four tokens of *permit* produced by one of the native speakers. Among the four graphs, the top two are tokens with initial stress, i.e. when *permit* was read as a noun and the bottom two are tokens with final stress, i.e. when *permit* was read as a verb. The two tokens in the left column were extracted from an accented condition, and the two tokens on the right were extracted from an unaccented condition. The pitch difference of the two syllables, indicated by the difference between the two red lines, is much larger in the left column than in the right column. F0 is used as a cue for stress to a lesser extent on words in the unaccented condition, i.e. the two graphs in the right column.

The target words were excised from the carrier sentences and used to construct a perception test. Two groups of participants listened to the tokens and decided whether the real English word *permit* has stress on the first syllable or second syllable. CE shows no difference from NE in discriminating the stress pattern of words from an accented context, i.e. tokens similar to those in the left column in Figure 5.4. But their performance is significantly worse than the native speakers for words from the unaccented condition, such as tokens in the right column (Wang & Yoon 2008). The comparable performance between CE and NE in perceiving lexical stress in accented contexts can be ascribed to their sensitivity to F0, whereas their difficulty with unaccented words is due to their insensitivity to duration and intensity. This study with real English words confirms the results of the present study with nonsense words. While native speakers have the choice of three acoustic cues in stress perception, Chinese learners restrict themselves to the difference in F0. In the absence of the F0 cue, Chinese learners' ability to perceive stress differences is severely compromised. Stress perception by native speakers of English doesnot rely on fixed rank of stress correlates but on a dynamic choice of cues depending on the availability and magnitude of the cues in different contexts. Chinese learners' stress perception is built on a strict ranking of acoustic cues with F0 on the top of the list as a dominant cue in all contexts.

In addition to the difference in the choice of acoustic cues, the exact employment of acoustic cues by CE is also different from NE. The discussion of hypothesis five above revealed that although the two groups share the acoustic cue of F0, reliance on the F0 cue by CE is much higher than by NE, as shown by the slope in Figure 5.5.

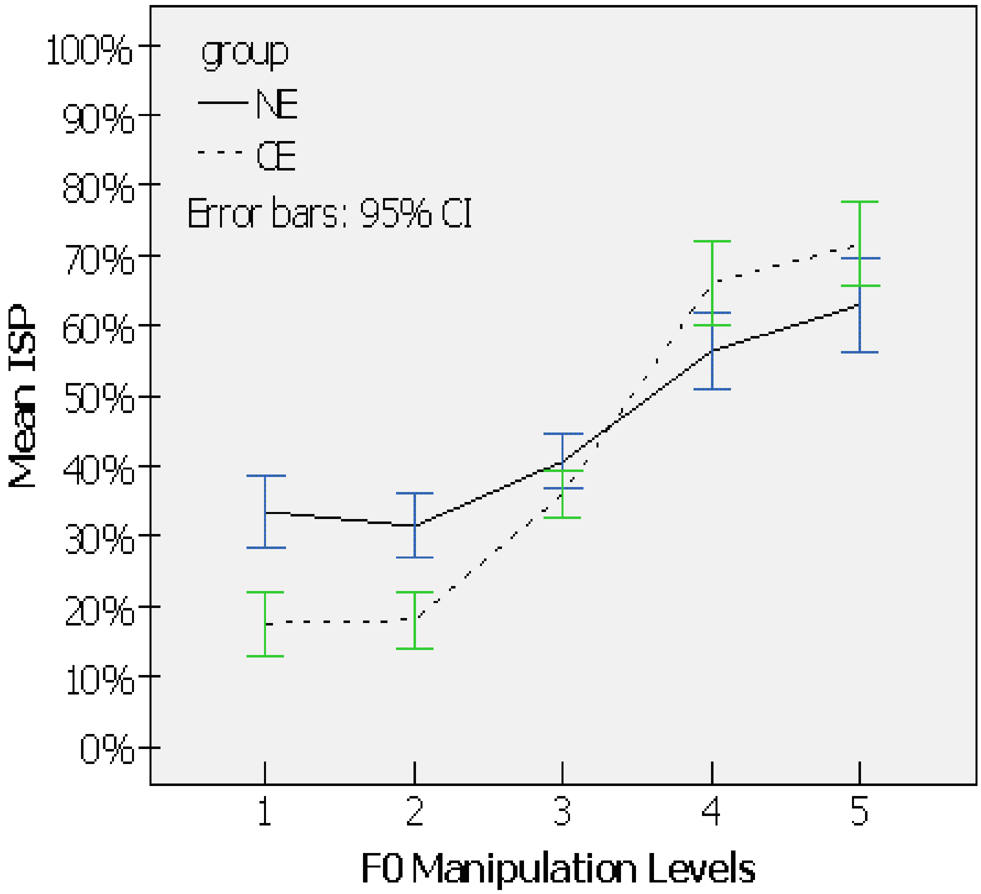


Figure 5.5 ISP of NE and CE groups as a function of F0 manipulation

Along the five manipulation levels of F0, CE shows a significantly lower ISP than NE for level one and level two and a significantly higher value of ISP for level four and five. As a result, the five levels of F0 manipulation lead to a much sharper increase of ISP in the CE group than in the NE group.

The logistic regression analysis arrives at the same conclusion. The F0 odds ratio in the group model of CE is twice as much as the F0 odds ratio in the group model for NE. When the predictor of F0 manipulation increases by one unit, the possibility of IS judgments increases by 3.888. The same increase of F0 manipulation is only predicted to lead to an increase of 1.792 in the possibility of

IS judgments in the NE group. Similarly, the comparison of the coefficients for F0 in the individual logistic regression models shows the same tendency (see Figure 5.6 below).

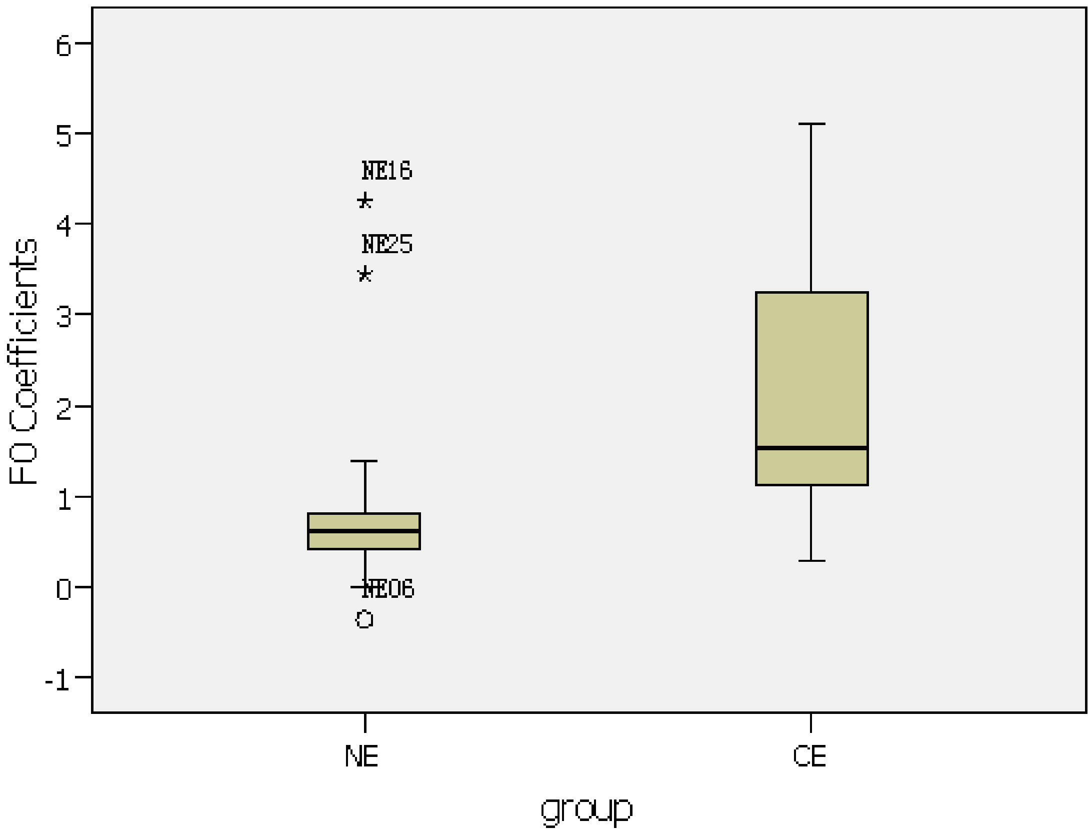


Figure 5.6 Boxplots of F0 coefficients for individual logistic regression models for NE and CE participants

This boxplot above summarizes the coefficients for the CE individuals and NE individuals. There is no overlap between the two boxes, and the median of CE coefficients is much higher than that of the NE coefficients. More importantly, CE is different from NE in the distribution pattern of F0 coeffients. The NE F0 coefficients exhibit a normal distribution, i.e. symmetrical over the median. CE participants' coefficients are heavily skewed.

Such heavy reliance on F0 and insensitivity to duration and intensity cues can be explained by the speech learning models introduced in chapter two. Both the Native Language Magnet model (Kuhl 1993, 2000) and the Speech Learning Model (Flege 1987, 1991, 1995, 2004) posit that sensitivity to an acoustic cue in L2 is largely determined by the learners' experience in L1. If an acoustic cue is actively involved in L1 experience, then the weighting of this cue in L2 speech learning is promoted. On the other hand, those cues that are not actively used in L1 can be systematically overlooked by learners in L2 speech learning. This process of ranking acoustic cues and assigning different weightings to them starts as early as the first year of life. Babies develop the ability to distinguish sharp contrasts found in their native language and gradually lose their sensitivity to contrasts that are not supported in their input (Werker 1989, Werker & Tees 1984). However, researchers emphasize that this loss is not "a sensorineural loss" (Werker & Tees 1984, 1886) but rather a change in speech processing strategies. To make speech processing more effective, the acoustic dimension that is used in native contrasts is not only attended to but also expanded in space, and the acoustic dimension that is not used in native contrasts is unattended and gradually shrunk in space. The result is a so-called "warped" perceptual space with different weightings of acoustic cues that is observed in adult speakers (Kuhl, 2000). Both NLM and SLM predict that such a perceptual space acts as a filter for L2 speech learning and determines what cues will be attended to.

Such views of speech learning have been supported by experimental studies into the perception of segmental contrasts in L2 (e.g. Flege 1995, Iverson et al 2003, Guion & Lee, 2006). There is evidence to believe that this is also true in the area of L2 prosody acquisition.

In an experimental study, Howie (1976) demonstrated that Mandarin speakers "could apparently make little use of any features other than pitch as cues for the perception of tonal distinctions." Using synthesized speech, Lin (1988) also obtained similar results to Howie's. F0 was proved to be the most discriminatory cue. If the F0 contour was changed into a constant F0 value, any change in duration or amplitude parameters would not lead listeners to perceive different tones. On the other hand, various investigations of the acoustic correlates of stress have shown that stress is not the result of a single mechanism (e.g. Fry 1955, 1958, 1964, Lehiste 1970, McClean & Tiffany 1973). Despite the lack of agreement on the relative weightings of duration and intensity, researchers generally recognize stress as the result of a composite of factors such as F0, duration and intensity, and agree that the perception and production of stress can not be attributed to any one single cue. Thus, Chinese learners of English, who have acquired a tone system in L1, have gradually developed a high sensitivity to F0 variation. At the same time, the acoustic cues of duration and intensity are largely unattended to in tonal perception and are thus suppressed in Chinese learners' perceptual space. This perceptual space frames Chinese learners' acquisition of stress contrasts in L2 English acquisition.

5.4 Implications for L2 stress Acquisition Theories

This section discusses the implications of the research and of other studies for the formation of L2 stress acquisition theories.

5.4.1 L1 transfer of phonetic cues

Previous discussions of stress acquisition have mainly attributed the success and the failure of L2 stress acquisition to transfer of L1 phonological units or to L1 phonological settings (e.g. Cheng 1968, Chao 1980, Juffs 1990, Archibald 2000). However, L1 prosodic transfer at the phonological level fails to account for the phonetic details of L2 stress perception and production. It is unclear why there is selective reliance on certain cues and deafness to other cues and also why the reliance on certain cues differs between native and non-native speakers.

This research suggests that, in order to offer more insight into L2 learners' stress acquisition, we have to explore the phonetic cues employed in L1. For example, to study the transfer of tone in stress acquisition, the realization of tone and stress must be taken into consideration. In terms of a phonological point of view, Chinese and English belong to two different prosodic systems, with Chinese as a typical tonal language and English as a typical stress language. In terms of a phonetic point of view, tone and stress differ from each other in their acoustic realizations. As discussed in chapter two, the investigation of tonal perception in Chinese has indicated that the most important acoustic correlate for tone in Chinese is F0 and that the role of duration and intensity in tonal perception is minimal. Given the earlier findings, we argue that what is being transferred to L2 stress perception is not the phonological tonal unit *per se*, but the reliance on the phonetic F0 cue. The acoustic correlate that is actively involved in L1 prosodic structure tends to be borrowed in the acquisition of L2 stress. L2 learners may have a perceptual bias towards the familiar cues in L1. This proposal agrees with the studies of Vietnamese and Japanese learners' choice of acoustic cues in perception and production of English stress (Nguyen 2003, Ueyama 2000). Vietnamese learners can use both F0 and intensity cues in English stress perception and production but failed to use duration, as both F0 and intensity but not duration were actively involved in tonal contrasts in Vietnamese. Japanese learners, who use F0 for pitch accent in L1, are adept at the use of F0 in English stress perception and production. However, they differ from both Vietnamese and Chinese learners in that they can soon learn to use duration as a cue for stress perception and production, much like duration is used in Japanese for segmental contrast.

Evidence can also be found in the acquisition of Chinese tone by English learners. Gandour (1983) found that English speakers rely on pitch height, which is an active acoustic correlate for stress in English, rather than pitch contour in tone perception in learning Chinese. On the other hand, Lee (1996) found that learners with a tonal language background were much better at tone discrimination, even if the tones are different. In other words, it appears that the linguistic function of pitch contour in learners' native tonal language background can help them to perceive tones in an L2 tonal language (Jongman & Wang in press).

In segmental studies, Iverson et al. (2003) found that in perception of /r/ and /l/ contrasts, native American learners relied primarily on F3 as a cue. Japanese learners of English were insensitive to the F3 cue and categorized the stimuli along the dimension of F2 manipulation. It was argued that Japanese listeners' reliance of F2 rather than F3 is a result of the active use of F2 difference between /r/ and /w/ in their native language. Similarly, Flege et al. (1992) found that Spanish learners of English were able to produce /p t k/and/b d g/ contrasts with nativelike closure voicing but failed to use preceding vowel duration as an effective cue for the contrasts. It was suggested by the authors that when learners were trying to establish a new category for L2 sounds, they might only use only use only of features that are phonetically important in the L1.

The investigation of the acoustic details of Chinese learners' stress perception in this study and by learners with other language backgrounds, supported by research into the acquisition of other prosodic and segmental units, suggests that it may be inadequate to remain restricted to the phonological level in the discussion of L1 prosodic transfer in theories of L2 prosodic acquisition. The influence of L1 may not operate as the transfer of a prosodic unit or af a setting rather through the phonetic realizations of L1 and L2 prosodic units.

5.4.2 Universal hierarchy of stress cues

The performance of Chinese learners, Vietnamese learners and Japanese learners in stress acquisition can be readily accounted for by the influence of L1, but it is still hard to explain the performance of French learners of English, reviewed in chapter two, who were also found to rely on F0 in stress perception (Montero 2007). Their reliance on F0 cannot be attributed to the transfer of an L1 acoustic cue in L2 stress acquisition, as French is not a tone language.

One hypothesis concerning the employment of acoustic cues in stress perception, the Functional Load Hypothesis (Beristein 1979, Remijsen 2002) may offer some insights. The Functional Load Hypothesis (FLH) posits that there is an unmarked universal hierarchy of acoustic cues for stress perception, based on the study of English stress correlates by Bolinger and on the generalization for all languages made by Hyman (from Beristein 1979). It was argued that in all languages, the most important cue is the change of F0, followed by duration, and lastly, intensity. In addition, FLH proposes that the extent to which an acoustic cue is used for signaling stress is based on the use of this cue in other aspects of the native phonological system, i.e. the functional load of this cue in the language system. If a language uses duration for phonetic contrasts in vowel length, then the use of this cue for stress perception will be superseded by the other two cues.

The results of a few studies were found to support this view of acoustic cue employment in stress perception. Beristein (1979) studied K'ekchi, a Mayan language which employs duration as a cue for phonemic vowel length, and found that duration was not used as a correlate of stress in the language. Beristein also suggested that in a tone language, F0 variation will not be used most frequently as a cue for stress perception, since it is used for lexical contrast. Instead, it should be "the least important cue" (ibid, 2). Duration will be promoted as the most important cue and intensity the second important cue. Berinstein used a study of Thai by Gandour (1974 in Berinstein 1979: 3) to support his prediction. He concluded from Gandour's study that, in Thai, stressed syllables are less cued by pitch displacement while they are more obviously longer in duration than unstressed syllables. Remijsen (2002) studied the Ma'ya language, which is argued to be "a hybrid word prosodic system, with both lexically contrastive stress and lexical tone" (39). Four acoustic correlates, vowel duration, vowel quality, selective amplitude and F0, were measured on the target words and analyzed as the potential phonetic correlates for tone and stress in the language. It was found that the first three parameters, duration, vowel quality and selective amplitude were closely correlated with variation in stress, but F0 was not. The correct tone classification, on the contrary, was correlated most successfully with F0, while duration, vowel quality and selective amplitude provided a much lower perception score. The author concluded that F0 as a prosodic parameter is used in the language to encode lexical tone and is thus assigned less functional load as a cue elsewhere in the phonological system, such as for lexical stress.

FLH was not designed to explain L2 stress perception. The results for Japanese, Vietnamese and Chinese learners in previous studies as well as in this study clearly violate the predictions of FLH. Vietnamese and Chinese are both tone languages, in which F0 is used for lexical contrasts. Japanese is a pitch accent language and F0 is also used in differentiating lexical items. According to the prediction of FLH, F0 should be avoided by these speakers for signaling lexical stress. However, all these learners used F0 as a primary cue for lexical stress. This contradiction can be explained by the fact that FLH was proposed to explain acoustic parameter settings for stress which co-exist with a tone system in the same language system but are not designed to account for the acquisition of stress in L2. If tone and lexical stress are both present in an L1 phonological system, different acoustic correlates may have to be used for the different units. However, in second language acquisition, speakers first learn to set acoustic parameters for a native prosodic unit, such as tone or pitch accent, and then start to reset these acoustic parameters or learn to assign new acoustic parameters for lexical stress in L2.

Despite the difficulty with FLH in explaining L2 stress acquisition, the universal hierarchy of stress correlates proposed by FLH offers an interesting perspective on L2 stress learning. FLH is based on the proposal of a universally ranked set of stress correlates: F0, duration and intensity. If L1 transfer does not interfere with the choice of correlates, then the universal hierarchy would emerge as the unmarked set of the acoustic cues for stress perception in L2 learning. The universal hierarchy of acoustic cues can be used to account for French learners' reliance on the F0 cue. As F0 is not used in French for lexical contrast, they choose to rely on F0 in perception and production of stress in L2.

Admittedly, without a large scale crosslinguistic study of learners with different L1 backgrounds, nothing conclusive can be drawn. The preliminary proposal is that a set of universal correlates for stress perception work within the possibility of L1 transfer to influence L2 learners' choice of acoustic cues.

5.5 The Influence of Word Form

The present study is intended to focus on the effect of the variation of the three acoustic cues in stress perception. Measurements have been taken in research design to avoid uncontrolled effects of other factors. For example, Guion and colleagues have identified several factors that contribute to the placement of stress on nonsense words by native and non-native speakers of English (Guion 2003, Guion et al. 2004, Guion et al. 2005, Guion et al. 2006). The three factors investigated include syllable structure, the influence of phonologically similar words and the lexical class of the nonsense words (whether a nonsense word is defined to be a noun or a verb). In order to reduce the effect of these three factors, the nonsense words used in this study were designed to have the same syllable structure in both syllables of a nonsense word, and tests were conducted to make sure the these nonsense words do not consistently remind native and non-native speakers of similar real English words. Furthermore, no indication of the lexical class of the nonsense words was made during the experiment. Despite all these measures, one factor other than acoustic correlates emerged as having a systematic influence on stress perception, and this is word form. Participants have exhibited clear preferences for stress placement for different word forms. The word form of *latmab* is preferred with initial stress. The word form of *nizdit* is clearly preferred as having final stress. Participants show a more neutralized preference of stress placement for the word form of *tetsep*, but it tends to be heard as having final stress, too. Despite such preferences for different word forms, the effects of the variation of the three acoustic correlates on the three word forms are consistent. ISP changes as a result of F0, duration and intensity manipulation in the same fashion for the three word forms. Furthermore, the two groups are similar to each other in their preference, which means that the difference among word forms is controlled between the two groups. Thus, we can propose that the stress pattern preference for different word forms does not interfere with the observations made for the effects of the acoustic cues and the comparisons between the two groups.

In an attempt to account for the differences between the word forms, we can refer to the segmental compositions of each word. In the word form *latmab*, the first syllable ends in a voiceless consonant, *t*, and the second syllable ends in a voiced consonant, *b*. In the word form *nizdit*, the opposite is observed. The first syllable ends in a voiced consonant, and the second syllable ends in a voiceless consonant. Acoustic studies have shown that vowel duration can be used as an important cue for voicing of the following consonant. Longer vowel duration is expected in front of a voiced consonant. Thus, if the vowels in the two syllables are of the same duration, the expectancy may make participants perceive the syllable with the voiceless consonant as having a 'longer' duration. Thus, for the word form, *latmab*, the first syllable may be perceived as longer and more stressed, and for the word form, *nizdit*, the opposite may have occurred. The word form, *tetsep*, has voiceless consonants in both syllables and it has proven to be the word form with the most balanced IS and FS judgments.

Chapter Six

Conclusion

In this concluding chapter, we first summarize the whole project. The contributions of the study and its limitations, especially in terms of research design, are discussed next. The last section of this chapter focuses on possible expansions of this study into larger scale research.

6.1 Summary of research

The acquisition of the prosodic unit, lexical stress, is never an easy task for second language learners. Learners from different language backgrounds have been found to have problems in stress perception and production. Most previous studies investigate such problems from a phonological aspect and conclude that whether the L1 has stress or not or what type of stress exists in the L1 are the root for the difficulties in L2 stress acquisition. The present study approaches the problem from a different angle. It compares Chinese learners of English with native English speakers in the employment of the three acoustic correlates for English lexical stress perception. There are two reasons why this is necessary. First, studies of the acquisition of L2 segments have proven that a comparison between the L1 and L2 segment inventory cannot fully predict the difficulties learners will face in learning new segmental contrasts, but rather that phonetic detail must be taken into consideration as well. Speech learning models have been constructed to explain the successes or difficulties in segment learning by looking at the difference between segments at the phonetic, articulatory, and acoustic level. Studies of the L2 acquisition of suprasegments should be extended into the phonetic level as well. Second, stress perception and production in first language development have also been studied from the articulatory and acoustic level. Researchers have not only investigated how speakers indicate stress in production but also what acoustic cues they rely on in stress perception. Experimental studies have been conducted with real words and nonsense words, as well as with isolated words and words in sentences. These methodologies and research designs can be adjusted and applied to L2 stress acquisition.

This project is an attempt to extend such experimental studies to L2 stress acquisition. In this study, two groups of participants listened to nonsense tokens with different configurations of F0, duration and intensity cues and judged the stress position on these tokens. The nonsense tokens were synthesized to reflect a systematic variation. The variation of the three parameters was realized on the two vowels of each test word. There are two major purposes for the experiment, first, to see if Chinese learners of English can make systematic judgments of stress position as a function of the cue variations, and second, to see if Chinese learners employ these cues differently from native speakers. The responses of the two groups of participants were analyzed to see the effects of the three cue manipulations on the change of Initial Stress Percentage. The contribution made by each of the three cues in predicting the respondents' stress judgments was analyzed by logistic regression models. The two groups were compared for their difference in the reliance scores and the difference in the weight of each cue in stress judgments. The results of the statistical analyses were discussed in relation to the five hypotheses raised in chapter three. To summarize the findings, the following two major conclusions can be drawn:

- Compared to the performance of native speakers of English, Chinese learners can also perceive a stress difference .

- Chinese learners differ from native speakers in their employment of acoustic cues in stress perception. The difference lies in two aspects
  - Chinese learners rely on a single acoustic cue, F0, in stress perception, rather than a combination of the three acoustic cues.
  - Compared to the native speakers' reliance on F0, Chinese learners are over-sensitive to the changes of F0 in stress perception.

Three models on speech acquisition, thePerception Assimilation Model (PAM), the Native Language Magnetic model (NLM), and the Speech Learning Models (SLM) were used to explain the findings of this study. According to the PAM model, Chinese learners' ability to perceive stress can be attributed to the assimilation of stressed and unstressed syllables to different tones in L1. Thus, it is not surprising that Chinese learners can perceive the stress position. However, this stress perception ability is flawed. Stressed and unstressed syllables are not tones. For example, the two syllables can carry the same pitch level and only differ in duration, with the stressed syllable being longer than the unstressed one. This means that the Chinese learners' strategy in stress perception is non-native. Their judgments are dependent on the F0 difference of the syllables but not on a combination of F0, duration and intensity. In addition, their reliance on F0 far exceeds the native reliance on F0. According to NLM and SLM, this perceptual bias is a result of L1 speech development. In first language development, speakers restructure acoustic cues in a way that is optimal for their L1 contrasts, segmental or suprasegmental. This restructuring is a distortion of the natural space of acoustic cues. It has a positive effect when the L2 shares the same acoustic cue organization but has a negative effect when the L2 organization is different. The findings of this study offer implications for the construction of L2 stress acquisition theories. It is suggested that phonetic details should be taken into consideration in the discussion of L1 transfer. In the section on future studies, we will introduce the possibility of involving more learners in the same type of experimental study to build a stress acquisition model.

6.2 Contributions and Limitations

One of the important motivations for this research is to improve teaching and learning of stress in second language acquisition. The results of this study suggest that there is no reason to be pessimistic about learning stress, or to be over-optimistic either. To make the teaching and learning of stress more effective, we should shift our focus. Instead of treating stress as an end product, we should teach stress as a process of realizing prominence on a sequence of syllables. The target of stress teaching should be to help learners to reassign weightings to each acoustic cue and to build a perceptual space that resembles that of a native speaker. Studies have found that learners, even adult learners, can be trained to change the weighting of an acoustic cue (Holt & Lotto 2006, Francis & Nusbaum 2002, Guion & Lee 2006). Successful results have been achieved by identification training using both lab speech and a naturalist approach. The key in identification training lies in "varying the input distribution patterns but orienting to a specific dimension, ..." (Guion & Lee 2006: 126) so as to direct learners to attend to one acoustic cue and gradually increase the weighting of this cue.

Another contribution made by this study is that it is the first study using manipulated nonsense tokens in investigating Chinese learners' stress perception. The only study using a similar approach in L2 studies is on French and Spanish learners of English by Montero (2007). Different from previous studies in L1 stress perception, which only focus on two cues at a time, this study combines all three cues, F0, duration and intensity manipulations on a same token. Accurate manipulation and application were made possible by the powerful speech processing and analyzing software Praat. This is a successful attempt to implement and adapt the resources of L1 speech analysis to the field of L2 speech acquisition.

The design used in this research project is quite sophisticated. Although every effort has been made to reduce or control for irrelevant factors, there is still room for improvements.

First of all, as presented and discussed earlier, the segmental composition of the nonsense words used in the perception test seems to impose systematic effects for stress judgment. In order to reduce the unexpected effects of word form in future, the nonsense words should be examined by a preliminary perception test to make sure that around 50% IS or FS will be achieved for tokens with average F0, duration and intensity on the two syllables.

Furthermore, for the manipulation of the acoustic cues in this study, it is found that large F0 differences between the syllables may be perceived as an indication of accent rather than of stress. Thus, it is more desirable to reduce the range of F0 manipulation. For the correlate of intensity, global intensity is manipulated in this study. More recent studies have shown that stress may relate more closely to an intensity difference located above 0.5 kHz instead of global intensity (Sluijter & van Heuven 1996). Future studies should examine intensity differences at higher frequencies ranges.

Another consideration for the perception test is that it is quite long, around 45 minutes on average. Participants may experience fatigue during the test. To reduce the load of the task, the test can be broken down into two or more sessions. As for the form of the test, the forced choice test used in this study is easy to construct and clear for the participants. However, the chance level in a forced choice test is 50%. This is deemed too high and reduces the validity of statistical analyses. Oddity tests or ABX are possible alternatives for future perception tests.

6.3 Future Studies

It is desirable to expand on this research in future studies in two ways, first, to increase the types of participants, and second, to include a production test as a comparison.

The target participants of the present study are Chinese learners of English. Few studies of the same kind have been identified in the area of L2 stress acquisition. To provide insights into the effect of universal settings of stress correlates and the effects of transfer of native acoustic cues from different L1 systems, learners from a wide range of language backgrounds should be involved. Future studies should include learners from an L1 with a similar stress system to English, learners from an L1 with different stress correlates, and learners from an L1 with a mixed system of stress and tone. Another way to expand the population of the experiment is to include learners from different proficiency levels. It is hypothesized that beginning learners and advanced learners will both differ in their use of the acoustic cues in stress perception. Beginning learners may rely more heavily on L1 cues than advanced learners, if the restructuring of perceptual space is indeed possible.

Furthermore, the scope of this study is only limited to the perception of lexical stress. Although many studies have shown that stress perception and production are related, it is expected that the two may exhibit interesting differences in their acoustic realizations in L2 learning. In future studies, the production of lexical stress will be recorded for direct comparison between the reliance on acoustic cues in production vs. perception.

REFERENCES

Abramson, A. S., & Lisker, L. (1970). Discriminability along the voicing continuum: Cross-language tests. *Proceedings of the Sixth International Congress of Phonetic Sciences* (pp.569–573). Prague: Academia.

Adams, C. (1979). *English Speech Rhythm and the Foreign Learner*. The Hague: Mouton.

Adams, C., & Munro, R. R. (1978). In search of the acoustic correlates of stress: Fundamental frequency, amplitude, and duration in the connected utterances of some native and non-native speakers of English. *Phonetica, 35*(3), 125-156.

Altmann, H. & Vogel, I. (2002). L2 Acquisition of Stress: the role of L1. Paper presented at the DGfS Annual Meeting "Multilingualism Today" in Mannheim, Germany, March 2002.

Altmann, H. (2006). The perception and production of second language stress: A cross-linguistic experimental study. Unpublished doctoral dissertation. University of Delaware.

Anderson-Hsieh, J., & Koehler, K. (1988). The effect of foreign accent and speaking rate on native speaker comprehension. *Language Learning, 38*, 561-593.

Anderson-Hsieh, J., & Venkatagiri, H. (1994). Syllable duration and pausing in the speech of chinese esl speakers. *TESOL Quarterly, 28*(4), 804-812.

Anderson-Hsieh, J., Johnson, R., & Koehler, K. (1992). The relationship between native speaker judgments of nonnative pronunciation and deviance in segmentals, prosody and syllable structure. *Language Learning, 42*, 529-555.

Archibald, J. (1992). Transfer of L1 parameter settings: Some empirical evidence from polish metrics. *Canadian Journal of Linguistics, 37*(3), 301-339.

Archibald, J. (1993). Metrical phonology and the acquisition of L2 stress. In F. R. Eckman (Ed.), *Confluence: Linguistics, L2 acquisition and speech pathology* (pp. 37-48): Amsterdam: John Benjamins.

Archibald, J. (1995). A longitudinal study of the acquisition of English stress. *Calgary working papers in linguistics, 17*, 1-10.

Archibald, J. (1997a). The acquisition of second language phrasal stress: A pilot study. In S. J. Hannahs & M. Young-Scholten (Eds.), *Focus on phonological acquisition* (pp. 263-289). Amsterdam/Philadelphia: John Benjamins.

Archibald, J. (1997b). The acquisition of English stress by speakers of nonaccentual languages: Lexical storage versus computation of stress. *Linguistics, 35*, 167-181.

Archibald, J. (2000). Models of l2 phonological acquisition. In B. Swierzbin, F. Morris, M. Anderson, C. Klee & E. Tarone (Eds.), *Social and cognitive factors in second language acquisition: Selected proceedings of the 1999 second language research forum* (pp. 125-157). Cascadilla Press.

Bake, R. G., & Smith, P. T. (1976). A psycholinguistic study of English stress assignment rules. *Language and Speech*, *19*, 9– 27.

Beckman, M. E. (1986). *Stress and non-stress accent*. Dordrecht, Holland: Foris.

Benkí, J. R. (2001). Place of articulation and first formant transition pattern both affect perception of voicing in English. *Journal of Phonetics*, *29*, 1–22.

Berinstein, A. E. (1979). A cross-linguistic study on the perception and production of stress. *University of California Working Papers in Phonetics*.

Bertoncini, J. & Mehler, J. (1981). Syllables as units in infant perception. *Infant Behavior and Development, 4*, 247-260.

Best, C. T. (1994a**)**. Learning to perceive the sound pattern of English. In C. Rovee-Collier & L. P. Lipsitt (Eds.) *Advances in Infancy Research* (pp. 125-157).

Norwood, NJ: Ablex.

Best, C. T. (1994b). The emergence of native-language phonological influences in infants: a perceptual assimilation model,'' In H. C. Nusbaum (Ed.) *The Development of Speech Perception: The Transition from Speech Sounds to Spoken Words* (pp. 247-260), Cambridge, MA: MIT.

Best, C. T., & Strange, W. (1992**)**. Effects of phonological and phonetic factors on cross-language perception on approximants. *Journal of Phonetics*, *20*, 305–330.

Best, C. T., McRoberts, G. W., & Goodell, E. (2001). Discrimination of non-native consonant contrasts varying in perceptual assimilation to the listener's native phonological system. *Journal of the Acoustical Society of America, 109*(2), 775-794.

Best, C.T. (1995). A direct realist view of cross-language speech perception. In W. Strange (Ed.), *Speech Perception and Linguistic Experience. Theoretical and Methodological Issues* (pp. 171-203). Baltimore: York Press.

Blicher, D. L., Diehl, R. and Cohen, L. B. (1990). Effects of syllable duration on the perception of the Mandarin Tone 2/Tone 3 distinction: evidence of auditory enhancement. *Journal of Phonetics 18*, 37-49.

Bohn, O. S. (1995). Cross-language speech perception in adults: First language transfer doesn't tell it all. In W. Strange (Ed.) *Speech perception and linguistic experience: Issues in cross-language research* (pp.279–304). Timonium, MD: York Press.

Bohn, O. S. & Flege, J. E. (1992). The production of new and similar vowels by adult German learners of English. *Studies in Second Language Acquisition*, 14 (2), 131-158.

Bolinger, D. L. (1965). Pitch accent and sentence rhythm. In I. Abe & T. Kanekiyo (Eds.), *Forms of english: Accent, morpheme, order* (pp. 139-180). Tokyo: Hokuou Pulishing Company.

Bolinger, D.L. (1958). A theory of pitch accent. *Word*, *14*, 109-149

Bond, Z. (1999). *Slips of the ear: Errors in the perception of casual conversation*. San Diego, CA: Academic Press.

Bond, Z., & Fokes, J. (1985). Non-native patterns of English syllable timing. *Journal of Phonetics, 13*, 407-420.

Browman, C. P., & Goldstein, L. (1992). Articulatory phonology: An overview. *Phonetica*, *49*, 155–180.

Broselow, E., Hurtig, R. R., & Ringen, C. (1987). The perception of second language prosody. In G. Ioup, & Weinberger, S. (Ed.), *Interlanguage phonology: The acquisition of a second language sound system* (pp. 350-361). Rowley, MA: Newbury House.

Chao, Y. R. (1980). Chinese tone and English stress. In L. R. Waugh & C. H. VanSchooneveld (Eds.), *The Melody of Language* (pp. 41-44). Baltimore, MD: University Park Press.

Chen, Y. (1999). Acoustic characteristics of American English produced by native speakers of Mandarin. Unpublished doctoral dissertation, University of Connecticut.

Cheng, L. (1968). English stresses and Chinese tones in Chinese sentences. *Phonetica, 18*, 77-88.

Chuang, C. K., Hiki, S., Sone, T., & Nimura, T. (1972). The acoustical features and perceptual cues of the four tones of Standard Colloquial Chinese. *Proceedings of the Seventh International Congress on Acoustics* (Adadémial Kiado, Budapest), 297-300.

Coster, D. C., & Kratochvil, P. (1984). Tone and stress discrimination in normal Beijing dialect speech. In B. Hong (Ed.), *New Papers on Chinese Language Use: Contemporary China Papers 18* (pp. 119-132). Canberra: Contemporary China Centre Research School of Pacific Studies, Australian National University.

Crystal, D. (1969) *Prosodic Systems and Intonation in English.* Cambridge: Cambridge University Press.

Crystal, T., & House, A. S. (1987). Segmental durations in connected-speech signals: Syllabic stress. *Journal of the Acoustic Society of America, 83*(4), 1574-1585.

Cutler, A., & Clifton, C. F. (1984). The use of prosodic information in word recognition. In H. Bouma & D. G. Bouwhuis (Eds.), *Attention and Performance x: Control of Language Processes* (pp. 183-196). Hillsdale, NJ: Lawrence Erlbaum.

Dalton, C., & Seidlhofer, B. (1994). *Pronunciation*. Oxford: Oxford University Press.

Davis, S. M., & Kelly, M. H. (1997). Knowledge of the English noun-verb stress difference by native and nonnative speakers. *Journal of Memory and Language*, *36*, 445–460.

Derwing, T. M., Munro, M. J., & Wiebe, G. (1998). Evidence in favor of a broad framework for pronunciation instruction. *Language Learning, 48*, 393-410.

Duanmu S. (2000). *The Phonology of Standard Chinese*. New York: Oxford University Press.

Dupoux, E., S. Peperkamp & N.Sebastián-Gallés (2001). A robust method to study stress deafness'. *Journal of the Acoustical Society of America*, *110*(3) Pt.1, Sep, 1606-1618.

Escudero, P., & Boersma, P. (2004). Bridging the gap between L2 speech perception research and phonological theory. *Studies in Second Language Acquisition*, *26*, 551–585.

Fakuade, G. (1991). Stress and intonation in Junkun and English: Implications for English as a second language. *Language Learning Journal, 3*, 81-83.

Field, J. (2005). Intelligibility and the listener: The role of lexical stress. *TESOL Quarterly, 39*(3), 399-423.

Flege, J. E. (1984). The detection of French accent by American listeners. *Journal of the Acoustic Society of America, 76*(3), 692-707.

Flege, J. E. (1986). The production and perception of foreign language speech sounds. In H. Winitz (Ed.) *Human Communication and its Disorders*, (vol. 2, pp. 224–401). Ablex Norwood, NJ,

Flege, J. E. (1987). The production of "new" and "similar" phones in a foreign language: Evidence for the effect of equivalence classification. *Journal of Phonetics, 15*, 47-65.

Flege, J., Munro, M., and Skelton, L. (1992). Production of the word-final English/t/-/d/contrast by native speakers of English, Mandarin, and Spanish. *Journal of the Acoustical Society of America, 92*, 128-14

Flege, J. E. (1993). Production and perception of a novel, second-language phonetic contrast. *Journal of the Acoustical Society of America*, *93*, 1589–1608.

Flege, J. E., Bohn, O. S., & Jang, S. (1997). Effects of experience on nonnative speakers' production and perception of English vowels. *Journal of Phonetics*, *25*, 437–470.

Flege, J. E. & MacKay, I. R. A. (2004). Perceiving vowels in a second language. *Studies in Second Language Acquisition*, *26*, 1-34.

Flege, J. E. (1995). Second language speech learning theory, findings and problems. In W. Strange (Ed.) *Speech Perception and Linguistic Experience: Issues in Cross Language Research*, 233-277. Baltimore: York Press.

Flege, J. E. (2003). Assessing constraints on second-language segmental production and perception. In N. O. Schiller & A.S. Meyer (Eds.) *Phonetics and Phonology in Language Comprehension and Production: Differences and Similarities* (pp.319-355). Berlin: Mouton de Gruyter.

Flege. J. E. (1991). The interlingual identification of Spanish and English vowels: Orthographic evidence. *The Quarterly Journal of Experimental Psychology*, *43*, 701-731.

Fowler, C. A., Best, C. T., & McRoberts, G. W. (1990). Young infants' perception of liquid coarticulatory influences on following stop consonants. *Percept. Psychophys*, *48*(6), 559–570.

Francis, A. L., & Nusbaum, H. C. (2002). Selective attention and the acquisition of new phonetic categories. *J. Exp. Psychol. Hum. Percept. Perform.*, *28*, 349–366.

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.

Fry, D. B. (1965). The dependence of stress judgments on vowel formant structure *Proceedings of the 5th International Congress on Phonetic Science*, *Münster, 1964* (pp. 306–311). Karger: Basel

Fu, Q., & Zeng, F. (2000). Identification of temporal envelope cues in Chinese tone recognition. *Asia Pacific Journal of Speech, Language and Hearing*, *5*, 45-57.

Gandour, J. (1974). On the representation of tone in Siamese. *UCLA Working Papers in Phonetics, 27*.

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

Gandour, J. (1984). Tone dissimilarity judgments by Chinese listeners. *Journal of Chinese Linguistics*, *12*, 235-261.

Gandour. J. T. (1978). The perception of tone. In V. A. Fromkin (Ed.), *Tone: A Linguistic Survey* (pp. 41-72). New York: Academic Press.

Goh, C. C. M. (1998). The level tone in Singapore English. *English Today, 14*(1), 50-53.

Goto, H. (1971). Auditory perception by normal Japanese adults of the sounds 'L' and 'R'. *Neuropsychologia*, *9*, 317–323.

Guion, S. G. (2005). Knowledge of English word stress patterns in early and late Korean-English bilinguals. *Studies in Second Language Acquisition*, *27*. 503-533

Guion, S. G., Clark, J. J., Harada, T., & Wayland, R. P. (2003). Factors affecting stress placement for English non-words include syllabic structure, lexical class, and stress patterns of phonologically similar words. *Language and Speech*, *46*, 403–427

Guion, S. G., Flege, J. E., Akahane-Yamada, R., & Pruitt, J. C. (2000). An investigation of second language speech perception: The case of Japanese adults' perception of English consonants. *J. Acoust. Soc. Am.*, *107*, 2711–2724.

Guion, S. G., Harada, T., & Clark, J. J. (2004). Early and late Spanish-English bilinguals' acquisition of English word stress patterns. *Bilingualism: Language and Cognition*, *7*, 207–226.

Guion, S.G. & Lee, B. (2006). The role of phonetic processing in second language acquisition. *English Language and Linguistics*, *21*, 123-148.

He, Y. J., Wang, Q., & Wilshire (2008). Production of English lexical stress by inexperienced and experienced learners of English. the annual conference of Canadian Acoustical Association. Vancouver, Canada.

Hogg, R., & McCully, C.B. (1987). *Metrical phonology: A coursebook*. Cambridge, England: Cambridge University Press.

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*(5). 3059-3071.

Howie, J. M. (1976). *Acoustical Studies of Mandarin Vowels and Tones*. Cambridge: Cambridge University Press.

Hyman, L. (1977). On the nature of linguistic stress. In L. Hyman (Ed.), *Studies in*

Stress and Accent, So. Calif. Occasional Papers in Linguistics, *4*, 37-82.

Ioup, G. & Tansomboon, A. (1987). The acquisition of tone: A maturational perspective. In G. Ioup & Weinberger, S. H. (Eds.) *Interlanguage Phonology*, Rowley, MA; Newbury House.

Iverson, P., and Kuhl, P. K. (1996). Influences of phonetic identification and category goodness on American listeners' perception of /r/ and /l/. *J. Acoust. Soc. Am.*, *99*, 1130–1140.

Iverson, P., Kuhl, P. K., Akahane-Yamada, R., Diesch, E., Tohkura, Y., Kettermann, A., & Siebert, C. (2003). A perceptual interference account of acquisition difficulties for non-native phonemes *Cognition*, *87*, B47– B57.

Jassem, W. (1959). The phonology of Polish stress. *Word*, *15*, 252-269.

Johansson, S. (1978). Studies of error gravity: Native reactions to errors produced by Swedish learners of English. Göteborg, Sweden: Acta Universitatis Gothoburgensis.

Jongman, A., Wang, Y., Moore, C. B., & Sereno, J. (in press). Perception and production of Mandarin Chinese tones. In E. Bates, L. H. Tan & O. Tseng (Eds.), *Handbook of Chinese Psycholinguistics*. Cambridge: Cambridge University Press.

Juffs, A. (1990). Tone, syllable structure and interlanguage phonology: Chinese learners' stress errors. *International Review of Applied Linguistics, 28*(2), 99-118.

Jusczyk, P., Cutler, A., & Redanz, N. (1993). Preference for the predominant stress pattern of English words. *Child Development, 64*, 675-687.

Kelly, M. H., & Bock, J. K. (1988). Stress in time. *Journal of Experimental Psychology: Human Perception and Performance*, *14*, 389–403.

Kratochvil, P. (1971). An experiment in the perception of Peking dialect tones. *A Symposium on Chinese Grammar: Scandinavian Institute of Asian Studies*

Monograph Series 6, Scandinavian Institute of Asian Studies, (Lund, Sweden), 7-31.

Kuhl, P. K. (1993). Early linguistic experience and phonetic perception: implications for theories of developmental speech perception. *Journal of Phonetics*, *21*, 125-139.

Kuhl, P. K. (1991). Human adults and human infants show a 'perceptual magnet effect' for the prototypes of speech categories, monkeys do not,'' *Perception and Psychophysics*, *50*(2), 93–107.

Kuhl, P.K. (2000). A new view of language acquisition. *Proceedings of the National Academy of Sciences of the United States of America,* 97 (22), October, 11850-11857.

Laver, John. (1994). *Principles of Phonetics*. Cambridge: Cambridge University Press.

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

Lehiste, I., & Peterson, G. E. (1959). Vowel amplitudes and phonemic stress in American English. *Journal of the Acoustical Society of America, 31*, 428-435.

Lieberman, P. (1960). Some acoustic correlates of word stress in American English. *Journal of the Acoustical Society of America, 32*, 451-454.

Lin, H. (1996). *Mandarin Tonology*. Taipei: Pyramid Press.

Lin, H. (2001). A Grammar of Mandarin Chinese. Munich: LINCOM Europa.

Lin, H-B., & Repp, Bruno H. (1989). Cues to the perception of Taiwanese tones. *Language and Speech*, *32*(1), 25-44.

Lin, M. C. (1988). Putong hua sheng diao de sheng xue texing he zhi jue zhengzhao [Standard Mandarin tone characteristics and percepts]. *Zhongguo Yuwen*, *3*, 182-193.

McClean, M. D., & Tiffany, W. R. (1973). The acoustic parameters of stress in relation to syllable position, speech, loudness, and rate. *Language and Speech, 16*, 283-291.

McEntee, E. (1973). The perception of English syllable stress by native and non-native speakers of English. Paper presented at the 1st International Conference on Foreign Language Testing, Dublin.

Mehler, J., Jusczyk, J.., Lambertz, G., Halsted, N., Bertoncini, J., & Amiel-Tison, C. (1988). A precursor of language acquisition in young infants. *Cognition*, *29*, 143-178.

Mol, H., & Uhlenbeck, E. M. (1965). The linguistic relevance of intensity in speech. *Lingua, 5*, 205-213.

Montero, D. (2007). *The perception of the acoustic correlates of stress: A cross-linguistic study on English, French and Spanish*. Unpublished Masters thesis, University of Washington.

Moore, C. B., & Jongman, A. (1997). Speaker normalization in the perception of mandarin Chinese tones. *Journal of the Acoustic Society of America, 102*, 1864-1877.

Morrison, G. S. (2004). An appropriate metric for cue weighting in L2 speech perception. *Studies in Second Language Acquisition*, 27, 597-606.

Morton, J., & Jassem, W. (1965). Acoustic correlates of stress. *Language and Speech, 8*, 159-181.

Munro, M. J. (1995). Nonsegmental factors in foreign accent: Ratings of filtered speech. *Studies in Second Language Acquisition, 17*, 17-34.

Nearey, T. M. (1990). The segment as a unit of speech perception. *Journal of Phonetics*, *18*, 347–373.

Nearey, T. M. (1997). Speech perception as pattern recognition. *Journal of the*

Acoustical Society of America, *101*, 3241–3254.

Nguyen, T. A. T. & Ingram J. (2005). Vietnamese Acquisition of English Word Stress. *TESOL Quarterly*, *39*(2), 309-319.

Nguyen, T. A. T. (2003). *Prosodic Transfer: The Tonal Constraints on Vietnamese Acquisition of English Stress and Rhythm.* Unpublished dissertation. University of Queensland, Queensland, Australia.

Palmer, J. (1976). Linguistic accuracy and intelligibility. In G. Nickel (Ed.), *Proceedings of the 4th International Congress of Applied Linguistics*, (pp. 505-513). Stuttgart: Hochschul.

Peperkamp, S., & Dupoux, E. (2002). A typological study of stress 'deafness'. In C. Gussenhoven & N. Warner (Eds.), *Laboratory Phonology 7*, (pp. 203-240). Berlin: Mouton de Gruyter.

Pike, Kenneth L. (1948). *Tone Languages*. Ann Arbor: University of Michigan Press.

Polivanov, E. (1974). The subjective nature of the perceptions of language sounds. In E. Polivanov, *Selected Works. Articles on General Linguistics* (compiled by A. Leont’ev), (pp. 223-237). The Hague: Mouton.

Pytlyk, C. (2007). Shared orthography: Do shared written symbols influence the perception of native-nonnative sound contrasts? Unpublished Masters thesis. University of Victoria

Ramus, F., & Mehler, J. (1999). Language identification with suprasegmental cues: A study based on speech resynthesis. *Journal of the Acoustical Society of America, 105*(1), 512-521.

Remijsen, B. (2002). Lexically contrastive stress accent and lexical tone in Ma’ya. In C. Gussehoven & N. Warner (Eds.), *Laboratory Phonology VII*. Berlin/New York: Mouton de Gruyter

Rivera-Castillo, Y., & Pickering, L. (2004). Phonetic correlates of stress and tone in a mixed system. *Journal of Pidgin and Creole Languages, 19*(2), 261-284.

Rumjancev, M. K. (1972). Ton i Intonacija v Sovremennom Kitajskom Jazyke [Tone and intonation in Modern Chinese] (Izdatel'stvo Moskovskogo Universiteta, Moscow). Reviewed by A. V. Lyovin (1978). *Journal of Chinese Linguistics, 6*, 120-168.

Sapir, E. (1921). *Language*. New York: Harcourt Brace Jovanovich.

Sereno, J. A., & Jongman, A. (1995). Acoustic correlates of grammatical class. *Language and Speech*, *38*, 57–76.

Shen, X.S. (1993). Relative duration as a perceptual cue to stress in Mandarin. *Language and Speech*, *36*, 415-433.

Sluijter, A. M. C., van Heuven, V. J., & Pacilly, J. J. A. (1997). Spectral balance as a cue in the perception of linguistic stress. *Journal of the Acoustical Society of America, 101*(1), 503-513.

Teaman, B. D. (1992). Stress in Japanese english: Evidence from native perceptual judgements. *Working papers in educational linguistics, 8*(1), 69-83.

Tseng, Chiu-yu. (1990). *An Acoustic Phonetic Study on Tones in Mandarin Chinese*. Taipei: Institute of History & Philology Academia Sinica.

Ueyama, M., *Prosodic transfer: An acoustic study of L2 English vs. L2 Japanese*. 2000, University of California, Los Angeles.

Vitevitch, M.S., Luce, P.A., Charles-Luce, J., & Kemmerer, D. (1997). Phonotactics and syllable stress: Implications for the processing of spoken nonsense words. *Language and Speech*. *40*, 47– 62.

Vogel, I. (2000). The Acquisition of Prosodic Phonology: Challenges for the L2 Learner. Paper presented at "Structure, Acquisition, and Change of Grammars: Phonological and Syntactic Aspects" in Hamburg, Germany, Oct

2000.

Wang, Q., & Yoon, Tae-Jin (2008). Do Chinese learners of English make use of fine phonetic details as English speakers do when perceiving the English lexical stress? Paper presented at “LabPhon11”, Wellington, New Zealand.

Wang, Y., Spence, M., Jongman, A., & Sereno, J. (1999). Training American listeners to perceive Mandarin tones. *Journal of the Acoustical Society of America*, *106*(6), 3649-3658

Watanabe, K. (1988). Sentence stress perception by Japanese students. *Journal of Phonetics, 16*(2), 181-186.

Werker, J. F., & Tees, R. C. (1984) Phonemic and phonetic factors in adult cross-language speech perception. *Journal of Acoustic Society of America. 75*, 1866–1878.

Werker, J. F., Gilbert, J. H. V., Humphrey, K., *and* Tees, R. C. (1981). Developmental aspects of cross-language speech perception. *Child Development*. *52*, 349–355.

Whalen, D. H., & Xu, Y. (1992). Information for Mandarin tones in the amplitude contour and in brief segments. *Phonetica*, *49*, 25-47.

Wu, Z-J. (1986). *The spectrographic album of mono-syllables of Standard Chinese.* Beijing: Social Science Press.

Appendix A
Similarity Test Results

The following table presents the results of the Nonsense Word Candidates Similarity Test. Words produced by native English participants and Chinese learners of English in the Word Similarity Task. Numbers in parentheses indicate the frequency of the word produced by participants. Words beginning in a capital letter are proper nouns. Natural score is listed in the last column, calculated by adding the scores given by all native English participants. The highest possible naturalness score should be 33.

			Summary	Naturalness Score
1	rɛpkɛs	repkes	repeat/repetition(4), kiss, rap, case, rib, repercussion, Christmas, represent, ribcage, cake	9
2	sɛpmɛk	sepmek	septic(3), September(4), sepia, separate, septicemia, MEC, seven, septor, mechanic, makeup, septum, setback	7
3	tɛtsɛp	tetsep	Tetris, tetrahedral, sepia, sibling, tetanus, septicemia, concept, tete(French), Zepplins(name of a band), September, tempest, tet	10
4	tɛtrɛp	tetrep	repeat/repetition(4), Tetris(4), ten wraps(2), tetanus, tetanus, representative, rip, techo, tete (French), pterodactyl, tab, tempered, temperance	6
5	sɛsmɛk	sesmek	seismic(4), sex, market, Sespool, make, MEC, mechanic, Sess make, succeed, sufficient, solvent, Cessna, six	15
6	kɛsrɛp	kesrep	repetition(3), ketchup(3), kiss/kissed(4), kestrel, kapers, representative, kissed, continuous, concept, canvas, casserole, kiss rip, reputation	7
7	mɛksɛp	meksep	except(4), accept(3), MEC(2), Ikea, sepia, septicemia, septor, supper, mechanoreceptors, concept, mix, sept(French), suscept, make, Macs	12
8	sɛptɛt	septet	septet(3), September(3), accept(2), sextet, septicemia, tit, tetanus, Subtext(Store name), septum, Tetris, septuplets, sept(French), seven, surreptitious	16
9	rɛptɛt	reptet	repeat/repetition(5), Tetris, tetrahedral, representative, head, telephone, recreational, rectangle, wracked it, redhead, reptile,	17

			reputation, repugnant	
10	mɛksɛs	mekses	`access/ac`cess (4), success(2), excess, make, MEC, cist, sespool, incest, ancestor, mechanism, mixes, message, mix us, Mick says, mechanic, modern, mechanism, ex`cess, breakfast, Mac	22
11	nɪzdɪt	nɪzdɪt	misfit, resident, nose, nasal, tit, dot, delightful, shiznit, missed it, needed, nitwit, niece	9
12	rɪzdɪv	rɪzdɪv	divide/division(5), residents, divine, rise/risen(2), rice, raised, ridiculous, rancor, diving, deserve	7
13	rɪzmɪp	rɪzmɪp	rhythm/rhythmic(3), mip, nipper, resin, rice, risen, map, wrist, nip, ridiculous, shiznit, recipe, risky, business, snip, rescind, reap	8
14	rɪzbɪt	rɪzbɪt	bit(4), respite, Ritz Bits, Reese's pieces, rice, dot, did, risen, Ritz, rabbit, risky, tidbit, Crispix	11
15	wɪzdɪt	wɪzdɪt	wise/wisdom(4), wizard(3), whiz(3), ditty, with, wasted, did, with it, whisk, walked, Whistler	17
16	dɪtnɪz	dɪtnɪz	didi, spinster, dot, nose, nasty, ditches, detonize, demise, dimwit, September, tetanus	5
17	dɪvrɪz	dɪvrɪz	divide/division(8), visibility, restitute, differential, resting, respect, decisive, different, driving, derision, defense	11
18	mɪprɪz	mɪprɪz	maples, risen, megaphone, mad, my prison, priss, model, maze, mace, prize	7
19	bɪtrɪz	bɪtrɪz	bit(3), bitten(2), business(2), risen, raised, rice, get rid, BiTTorrent, batteries, busy, betray, repetition, repulsion, repentance	9
20	dɪtwɪz	dɪtwɪz	wizard(3), whiz, didi, dichotomy, with, whilst, dickless, cheesewhiz, dimwit, distress, dimples	11
21	mæzlæt	mæzlæt	Mazda(3), masturbation, batch, batter, mosotop(?), maze, mass, late, tablet, Manhattan, Platz, Maslit, Mazatlan, massive, math	11
22	lætmæb	lætmæb	Latin(2), lab rat(2), latitude, Mod, latrine, map, maybe, Latin, math lab, lactate, madly, later, maybe, Latvia, Latisamus, blanket, mat, blot	8
23	hæskæk	hæskæk	Pascal(3), pass/passed(4), cackle, passion, past, pessimist, passive, pasirin, hassle, casket, cat, catnap, hasn't, pacifist, pacific, pasta	7

24	fæslæt	fæslæt	fast/faster(8), Latvia, caselot, fiz, tablet, fasten, fuss a lot, fat rat	10
25	dæzlæt	dæzlæt	dazzle/dazzling(4), disaster, Latvia, azlin, daze, lot, dazed, tablet, that is a lot, does a lot, dad's lot, Daesin, daschund, days	12
26	lætmæz	lætmæz	lats/latissimusdorsi(4), bar, mitzvah, latitude, Latvia, Lamose, latitude, maze, latte, late, Latin, lactate, lament, litmus, llamas, let	8
27	mæblæt	mæblæt	lab rat(2), maple, latte, tablet, math lab, mad lot, maybe, my blood	12
28	kækhæs	kækhæs	cactus(3), cat(6), hack, hasty, tacphony, cactus, cathouse, cackle, catastrophe, catch, carcass, khakis, capture	12
29	lætfæs	lætfæs	fast/faster(4), latte(2), latitude, latte, latrine, latee facsia(muscle group), breakfast, face, LaPaz, blasphemous, last, laugh, lost, lamb	15
30	lætdæz	lætdæz	late/later(6), latte(2), dazzle(2), adolescent, dance, what does, ladies, Latin, blenders, Latvia, longing, battles, bladder	9

Appendix B
Normalization and Manipulation Scripts

There are three scripts in the normalization script set: the duration normalization script, F0 normalization script, and intensity-normalization script. The manipulation script is a separate script which takes the sound files resulting from the normalization scripts and produces 125 tokens with different combinations of the F0, duration and intensity combinations.

Duration Normalization Script

```
#################################################################################
#for duration manipulation
#Firstly, the averaged duration of V1 and V2 are calculated
#the highest intensity and its corresponding timepoint in V2 is identified
#the shorter vowel among the two is increased to the average duration
#the longer vowel is reduced to the average duration
#################################################################################
###change from the last version
###the blank before the sound and the blank after the sound are both reduced to 0.05s
directory$ = "D:\"
object_name$ = selected$("Sound")
select Sound 'object_name$'
select TextGrid 'object_name$'
before_sound_begin = Get starting point... 1 1
after_sound_end = Get end point... 1 7
sound_begin = Get starting point... 1 2
sound_end = Get end point... 1 6
v1_begin= Get starting point... 1 3
v1_end= Get end point... 1 3
v1_duration= v1_end - v1_begin
v2_begin= Get starting point... 1 5
v2_end= Get end point... 1 5
v2_duration = v2_end - v2_begin
o_duration = v1_begin - sound_begin
co_duration = v2_begin - v1_end
coda_duration = sound_end - v2_end
before_blank = sound_begin - before_sound_begin
after_blank = after_sound_end - sound_end
printline 'v1_duration' 'v2_duration' 'newline$'
printline 'sound_begin' 'sound_end'
printline 'before_sound_blank' 'after_sound_blank'
####################################################################
#create points where I can start and stop manipulating the two vowels
#from the begin of the sound file to 0.001 sec before v1,
#the sound will remain unchanged
#from 0.001 sec after the end of v1 to 0.001 before the begin of v2
#the sound will also remain unchanged
#also, from 0.001 sec after the end of v2 to the end of the sound
#the sound file will remain unchanged
####################################################################
v1_manip_begin=v1_begin-0.001
v1_manip_end=v1_end+0.001
v2_manip_begin=v2_begin-0.001
```

```
v2_manip_end=v2_end+0.001
######################################################################
#calculate shorten and lengthen percentage
#1.calculate the average duration of v1 and v2
#2.the change percentage of v1 will be the ratio between the average and v1
#3.the same is true with v2
######################################################################
average_duration = (v1_duration+v2_duration)/2
v1_percent = average_duration/v1_duration
v2_percent = average_duration/v2_duration
######################################################################
#start manipulating duration
#6 points are needed to manipulate 5 stretches of sounds
#2 vowels need to be changed, the other 3 stretches kept same
######################################################################
a = 0.05/before_blank
b = 0.05/after_blank
printline average_duration 'average_duration' v1_percent 'v1_percent' v2_percent 'v2_percent'
printline after_sound_end 'after_sound_end' 'after_sound_blank' a 'a' b 'b'
Create DurationTier... 'object_name$' 0 'after_sound_end'
Add point... 'before_sound_begin' 'a'
Add point... 'sound_begin'-0.001 'a'
Add point... 'sound_begin' 1/1
Add point... 'v1_manip_begin' 1/1
Add point... 'v1_begin' 'v1_percent'
Add point... 'v1_end' 'v1_percent'
Add point... 'v1_manip_end' 1/1
Add point... 'v2_manip_begin' 1/1
Add point... 'v2_begin' 'v2_percent'
Add point... 'v2_end' 'v2_percent'
Add point... 'v2_manip_end' 1/1
Add point... 'sound_end' 1/1
Add point... 'sound_end'+0.001 'b'
Add point... 'after_sound_end' 'b'
## to get resynthesis
##
##
	select Sound 'object_name$'
	To Manipulation... 0.01 75 600
	plus DurationTier 'object_name$'
	Replace duration tier
	select Manipulation 'object_name$'
	Get resynthesis (PSOLA)
## rename the new sound file as *_dura_changed
	Rename... 'object_name$'_dura_change
	Write to WAV file... 'directory$''object_name$'_dura_change.wav
#	Read from file... 'directory$''object_name$'.TextGrid
```

Pitch Normalization Script

```
#####################################################################################
#for pitch manipulation
#This script should be applied after the Duration Average Script
#Firstly, the pitch contours for V1 and V2 were achieved by
#a pitch range setting of 75Hz-600Hz, and a time step of 0.001
#As the duration for V1 and V2 is the same now, a same number of
#the sampling points for both vowels are in pair
#Thus, the value of every pair of sampling points were averaged and
#the averaged value was used for the creation of the new pitch contour
#Finally the new pitch contour was applied to both vowels
#####################################################################################
# create a directory string and create an object_name string
# the object_name equals to the name of the selected sound file
# in the object list
dir$ = "D:\"
object_name$ = selected$("Sound")
##
##
# select the sound object first in order to
# create a Manipulation
select Sound 'object_name$'
	To Manipulation... 0.001 75 600
sound_name$ = left$(object_name$, 7)
textgrid_name$ = sound_name$ + "_averaged"
Read from file... 'dir$''textgrid_name$'.TextGrid
# define the important duration points in the sound file
# with the help of the textgrid file
# the textgrid file should include info about
# the onset of Syl 1, V1, Coda (syl 1)+ Onset (Syl 2), V2, Coda (Syl 2)
select Sound 'object_name$'
select TextGrid 'textgrid_name$'
sound_begin = Get starting point... 1 2
sound_end = Get end point... 1 6
v1_begin= Get starting point... 1 3
v1_end= Get end point... 1 3
v1_duration= v1_end - v1_begin
v2_begin= Get starting point... 1 5
v2_end= Get end point... 1 5
o_duration = v1_begin - sound_begin
co_duration = v2_begin - v1_end
coda_duration = sound_end - v2_end
#
Create PitchTier... empty 0 sound_end
Rename... average_pitch_contour
#

select Sound 'object_name$'
To Pitch... 0.001 75 600
Rename... pitch
i = 0
echo new info
```

```
printline 'pitch not changed'
printline 'sound_begin' 'sound_end'
printline 'o_duration' 'co_duration' 'coda_duration'
filedelete 'dir$'pitchmani_record.xls
fileappend 'dir$'pitchmani_record.xls "onset contour" 'newline$'
for i from 0 to o_duration/0.001
printline 'i'
	select Pitch pitch
		time_keep_pitch = sound_begin+i*0.001
		select Pitch pitch
		pitch_keep = Get value at time... time_keep_pitch Hertz Linear
		string$ = Get value at time... time_keep_pitch Hertz Linear
		printline 'time_keep_pitch:3' 'pitch_keep:3'
		select PitchTier average_pitch_contour
		a$ = "--undefined--"
		print 'string$'
fileappend 'dir$'pitchmani_record.xls 'i' 'tab$'
		if (string$ <> a$)
			Add point... time_keep_pitch pitch_keep
fileappend 'dir$'pitchmani_record.xls 'time_keep_pitch:3' 'tab$' 'pitch_keep:3' 'tab$' 'newline$'
		else
fileappend 'dir$'pitchmani_record.xls "no point created" 'newline$'
		endif
endfor
fileappend 'dir$'pitchmani_record.xls "coda and onset contour" 'newline$'
for i from 0 to co_duration/0.001
printline 'i'
	select Pitch pitch
		time_keep_pitch=v1_end+i*0.001
		select Pitch pitch
		pitch_keep=Get value at time... time_keep_pitch Hertz Linear
		string$ = Get value at time... time_keep_pitch Hertz Linear
		printline 'time_keep_pitch:3' 'pitch_keep:3'
		select PitchTier average_pitch_contour
		a$ = "--undefined--"
		print 'string$'
fileappend 'dir$'pitchmani_record.xls 'i' 'tab$'
		if (string$ <> a$)
			Add point... time_keep_pitch pitch_keep
fileappend 'dir$'pitchmani_record.xls 'time_keep_pitch:3' 'tab$' 'pitch_keep:3' 'tab$' 'newline$'
		else
fileappend 'dir$'pitchmani_record.xls "no point created" 'newline$'
		endif
endfor
fileappend 'dir$'pitchmani_record.xls "coda contour" 'newline$'
for i from 0 to coda_duration/0.001
printline 'i'
	select Pitch pitch
		time_keep_pitch=v2_end+i*0.001
		select Pitch pitch
		pitch_keep=Get value at time... time_keep_pitch Hertz Linear
		string$ = Get value at time... time_keep_pitch Hertz Linear
		printline 'time_keep_pitch:3' 'pitch_keep:3'
```

```
                select PitchTier average_pitch_contour
                a$ = "--undefined--"
                print 'string$'
fileappend 'dir$'pitchmani_record.xls 'i' 'tab$'
                if (string$ <> a$)
                        Add point... time_keep_pitch pitch_keep
fileappend 'dir$'pitchmani_record.xls 'time_keep_pitch:3' 'tab$' 'pitch_keep:3' 'tab$' 'newline$'
                else
fileappend 'dir$'pitchmani_record.xls "no point created" 'newline$'
                endif
endfor
printline 'pitch changed'
fileappend 'dir$'pitchmani_record.xls "v1 and v2 contour" 'newline$'
        for i from 0 to (v1_end - v1_begin)/0.001
        printline 'i'
        select Pitch pitch
                time1=v1_begin+i*0.001
                select Pitch pitch
                pitch_1 = Get value at time... time1 Hertz Linear
                string1$ = Get value at time... time1 Hertz Linear
                printline 'time1:3' 'pitch_1:3'
                time2= v2_begin + i*0.001
                select Pitch pitch
                pitch_2=Get value at time... time2 Hertz Linear
                string2$ = Get value at time... time2 Hertz Linear
                printline 'time2:3' 'pitch_2:3'
                select PitchTier average_pitch_contour
fileappend 'dir$'pitchmani_record.xls 'i' 'newline$'
                if (string1$ <> a$) and (string2$ <> a$)
                        average_pitch = (pitch_1+pitch_2)/2
                        printline 'average_pitch'
                        Add point... time1 average_pitch
                        Add point... time2 average_pitch
fileappend 'dir$'pitchmani_record.xls 'time1:3' 'tab$' 'average_pitch:3' 'tab$' 'newline$'
fileappend 'dir$'pitchmani_record.xls 'tab$' 'time2:3' 'tab$' 'average_pitch:3' 'tab$' 'newline$'
                else
fileappend 'dir$'pitchmani_record.xls "no point created" 'newline$'
                endif
        endfor
# replace original pitch tier
        plus Manipulation 'object_name$'
        Replace pitch tier
        select Manipulation 'object_name$'
        Get resynthesis (PSOLA)
        Rename... 'object_name$'_pitch_change
        Write to WAV file... 'dir$''object_name$'_pitch_change.wav
```

Intensity Normalization Script

```
####################################################################################
#for intensity manipulation
#ideas partly borrowed from Manual page: Script for listing time-F0-intensity
#the highest intensity and its corresponding timepoint in V1 is identified
#the highest intensity and its corresponding timepoint in V2 is identified
#an intensity tier is created
#two intensity points are added
#the first point is created on the timepoint of V1, with the intensity of V2
#the second point is created on the timepoint of V2, with the intensity of V1
#This tier is MULTIPLIED to the original sound file
#thus the highest intensity in both vowels will be the same, equal V1_intensity*V2_intensity
#Then the new file is scaled down to the level of the average intensity of the original sound
####################################################################################
directory$ = "D:\"
object_name$ = selected$("Sound")
##
##
# select the sound object first in order to
# create a Manipulation
select Sound 'object_name$'
	To Manipulation... 0.01 75 600
sound_name$ = left$(object_name$, 7)
textgrid_name$ = sound_name$ + "_averaged"
Read from file... 'directory$''textgrid_name$'.TextGrid
select Sound 'object_name$'
##
# to get the average intensity of the original file
##
# intensity_scale = Get intensity (dB)
##
# create an intensity analysis tier
##
To Intensity... 75 0.001
Rename... intensity
##
# to identify V1 and V2 time points in the sound
##
select TextGrid 'textgrid_name$'
sound_begin = Get starting point... 1 2
sound_end = Get end point... 1 6
v1_begin= Get starting point... 1 3
v1_end= Get end point... 1 3
v1_duration= v1_end - v1_begin
v2_begin= Get starting point... 1 5
v2_end= Get end point... 1 5
v2_duration = v2_end - v2_begin
o_duration = v1_begin - sound_begin
co_duration = v2_begin - v1_end
coda_duration = sound_end - v2_end
v_duration = v1_begin - v1_end
##
```

```
#create an intensity tier and manipulation intensity
##
Create IntensityTier... int_change 0 sound_end
#
	for i from 0 to o_duration/0.001
		time_keep = sound_begin + i*0.001
		select IntensityTier int_change
		Add point... time_keep 0
	endfor
#
#fileappend 'dir$'\'filename$' "V1 coda and V2 onset intervals intensity contour" 'tab$' "intensity change" 'tab$' "0" 'newline$'
	for i from 0 to co_duration/0.001
		time_keep= v1_end + i*0.001
		select IntensityTier int_change
		Add point... time_keep 0
	endfor
#
#fileappend 'dir$'\'filename$' "V2 coda interval intensity contour" 'tab$' "intensity change" 'tab$' "0" 'newline$'
	for i from 0 to coda_duration/0.001
		time_keep=v2_end+i*0.001
		select IntensityTier int_change
		Add point... time_keep 0
	endfor
	for i from 0 to (v1_end - v1_begin)/0.001
			time1=v1_begin+i*0.001
			time2=v2_begin+i*0.001
	select Intensity intensity
	intensity1 =Get value at time... time1 Cubic
	intensity2 =Get value at time... time2 Cubic
# Cubic is used because this is the standard method of intensity measurement in Praat.
	average_int = (intensity1 + intensity2)/2
	v1int_change=average_int - intensity1
	v2int_change=average_int - intensity2
			select IntensityTier int_change
				Add point... time1 v1int_change
				Add point... time2 v2int_change
	endfor
#fileappend 'dir$'\'filename$' "V1 intensity change" 'tab$' 'v1int_mani' 'tab$' "V2 intensity change" 'tab$' 'v2int_mani' 'newline$'
# replace original intensity tier
select Sound 'object_name$'
plus IntensityTier int_change
Multiply... no
# Scale intensity... 'intensity_scale'
final_name$ = sound_name$ + "_averaged.wav"
Rename... 'final_name$'
Write to WAV file... 'directory$''final_name$'
```

```
##########################################################################
# This script is used to manipulate pitch, intensity and duration of a
# CV1C.CV2C sound file, where V1 and V2 are the same vowel and they share
# the same pitch contour, intensity contour and duration
# The manipulation of the three correlates are realized in 5 steps each.
# Pitch:    V1 p - V2 p = 50 Hz, 25 Hz, 0 Hz, -25 Hz, and -50 Hz
# Intensity: V1 i - V2 i = 9 dB, 4.5 dB, 0 dB, -4.5 dB, -9 dB
# Duration: V1 d / V2 d = 1.5, 1.25, 1, 0.75, 0.5
# Every step is combined with every other step
# Thus, the token number realized from one sound file will be
# 5 pitch manipulation * 5 intensity mani * 5 duration mani = 125 tokens
# There are mainly 4 parts in the scripts.
# The first part is a part of preparation for manipulation and
# communication with outside files, defining of directory and filenames etc.
# The second part is the loop of pitch mani.
# The third part is the loop of intensity mani.
# The fourth part is the loop of duration mani.
# The dura mani loop is embbeded in int mani loop, which is in turn embedded
# in the pit mani loop.
############################################################################
#
##### Define the directory for saving files
#
dir$ = "d:\"
#
##### Define the object_name to be the same as the name of the sound file
##### selected in the praat object list
object_name$ = selected$("Sound")
filename$= object_name$ + ".xls"
filedelete 'dir$'\'filename$'
#
##### Select the sound object first in order to
##### create a Manipulation
select Sound 'object_name$'
Scale intensity... 70
To Manipulation... 0.001 75 600
#
##### Measure and calculate the duration of intervals
sound_name$ = left$(object_name$, 7)
textgrid_name$ = sound_name$ + "_averaged"
Read from file... 'dir$''textgrid_name$'.TextGrid
sound_begin = Get starting point... 1 2
sound_end = Get end point... 1 6
v1_begin= Get starting point... 1 3
v1_end= Get end point... 1 3
v1_duration= v1_end - v1_begin
v2_begin= Get starting point... 1 5
v2_end= Get end point... 1 5
v2_duration = v2_end - v2_begin
o_duration = v1_begin - sound_begin
co_duration = v2_begin - v1_end
coda_duration = sound_end - v2_end
v_duration = v1_begin - v1_end
```

```
token_id = 1
#
##### Create a Pitch object in order to measure the pitch of the original sound file
select Sound 'object_name$'
To Pitch... 0.001 50 300
Rename... pitch
#
##################################################################################
#################
##### Define the two variables used to be added to V1 and V2
##### Step manipulation of pitch for V1 and V2. These 2 values are added to V1 and V2, respectively
##### These two values are only half of step size. This is coz
##### (V1 + v1pit_mani)- (V2 + v2pit_mani) = step size = 50, 25, 0, -25, -50
##### Thus, v1pit_mani - v2pit_mani = step size.
##################################################################################
#################
v1pit_mani= 25
v2pit_mani= -25
#
##### The loop to control the manipulation of pitch. There are 5 steps and thus 'from 1 to 5'.
##### pit_id indicates pitch manipulation ID number.
#
for pit_id from 1 to 5
#fileappend 'dir$'\'filename$' "token id is" 'tab$' 'token_id' 'newline$'
#fileappend 'dir$'\'filename$' 'object_name$' 'tab$' "original sound basic info" 'newline$'
#fileappend 'dir$'\'filename$' "onset duration" 'tab$' 'o_duration' 'tab$' "coda plus onset duration" 'tab$'
'co_duration' 'tab$' "coda duration" 'tab$' 'coda_duration' 'newline$'
#fileappend 'dir$'\'filename$' "V1 duration" 'tab$' 'v1_duration' 'tab$' "V2 duration" 'tab$'
'v2_duration' 'newline$'
#fileappend 'dir$'\'filename$' 'newline$'
#fileappend 'dir$'\'filename$' 'newline$'

#fileappend 'dir$'\'filename$' "pitch manipulation" 'newline$'
#fileappend 'dir$'\'filename$' "pit_id" 'tab$' 'pit_id' 'newline$'
select Sound 'object_name$'
# pause select the original sound file and wait to create pitch tier for manipulation
Create PitchTier... pit_mani 0 'sound_end'
#
##### There are mainly two parts, one part to keep the original pitch in consonantal intervals
##### and the other to manipulate the pitch in vowel intervals, V1 and V2
#fileappend 'dir$'\'filename$' "Pitch not changed" 'newline$'
#fileappend 'dir$'\'filename$' "V1 onset interval pitch contour" 'newline$'
##### Part one of pitch manipulation, to keep the pitch in three consonantal intervals, which
##### is the onset of the onset of V1, the interval between V1 and V2, the coda of V2.
for i from 0 to o_duration/0.001
	select Pitch pitch
		time_keep_pitch = sound_begin + i*0.001
		select Pitch pitch
		pitch_keep = Get value at time... time_keep_pitch Hertz Linear
		string$ = Get value at time... time_keep_pitch Hertz Linear
		select PitchTier pit_mani
		a$ = "--undefined--"
#fileappend 'dir$'\'filename$' 'i' 'tab$'
```

```
            if (string$ <> a$)
                  Add point... time_keep_pitch pitch_keep
#           else
#fileappend 'dir$'\'filename$' 'time_keep_pitch:3' 'tab$' 'pitch_keep:3' 'tab$' 'newline$'
            #fileappend 'dir$'\'filename$' "no pitch point created at this point" 'newline$'
            endif
endfor
##
#fileappend 'dir$'\'filename$' "V1 coda plus V2 onset interval pitch contour" 'newline$'
for i from 0 to co_duration/0.001
      select Pitch pitch
            time_keep_pitch=v1_end + i*0.001
            select Pitch pitch
            pitch_keep = Get value at time... time_keep_pitch Hertz Linear
            string$ = Get value at time... time_keep_pitch Hertz Linear
            select PitchTier pit_mani
            a$ = "--undefined--"
#fileappend 'dir$'\'filename$' 'i' 'tab$'
            if (string$ <> a$)
                  Add point... time_keep_pitch pitch_keep
#fileappend 'dir$'\'filename$' 'time_keep_pitch:3' 'tab$' 'pitch_keep:3' 'tab$' 'newline$'
#           else
#fileappend 'dir$'\'filename$' "no pitch point created at this point" 'newline$'
            endif
endfor
#
#fileappend 'dir$'\'filename$' "V2 coda interval pitch contour" 'newline$'
for i from 0 to coda_duration/0.001
      select Pitch pitch
            time_keep_pitch = v2_end+i*0.001
            select Pitch pitch
            pitch_keep = Get value at time... time_keep_pitch Hertz Linear
            string$ = Get value at time... time_keep_pitch Hertz Linear
            select PitchTier pit_mani
            a$ = "--undefined--"
#fileappend 'dir$'\'filename$' 'i' 'tab$'
            if (string$ <> a$)
                  Add point... time_keep_pitch pitch_keep
#fileappend 'dir$'\'filename$' 'time_keep_pitch:3' 'tab$' 'pitch_keep:3' 'tab$' 'newline$'
#           else
#fileappend 'dir$'\'filename$' "no pitch point created at this point" 'newline$'
            endif
endfor
#
###### Part two of pitch manipulation, to change the pitch on V1 and V2
#fileappend 'dir$'\'filename$' "Pitch changed" 'newline$'
#fileappend 'dir$'\'filename$' "V1 and V2 intervals pitch contour" 'newline$'
      for i from 0 to v1_duration/0.001
#fileappend 'dir$'\'filename$' 'i' 'tab$'
            select Pitch pitch
            time1 = v1_begin+i*0.001
            select Pitch pitch
            pitch_1 = Get value at time... time1 Hertz Linear
```

```
		string1$ = Get value at time... time1 Hertz Linear
		time2 = v2_begin+i*0.001
		select Pitch pitch
		pitch_2 = Get value at time... time2 Hertz Linear
		string2$ = Get value at time... time2 Hertz Linear
		select PitchTier pit_mani
###### if either V1 or V2 pitch points are not identified
		if (string1$ <> a$) and (string2$ <> a$)
			manipulated_pitch1 = pitch_1 + v1pit_mani
			Add point... time1 manipulated_pitch1
			manipulated_pitch2=pitch_2 + v2pit_mani
			Add point... time2 manipulated_pitch2
#fileappend 'dir$'\'filename$' "V1 time point" 'tab$' 'time1' 'tab$' "V1 pitch value" 'tab$' 'pitch_1' 'tab$' "V2 time point" 'tab$' 'time2' 'tab$' "V2 pitch value" 'tab$' 'pitch_2' 'newline$'
#fileappend 'dir$'\'filename$' "V1 pitch change" 'tab$' 'v1pit_mani' 'tab$' 'tab$' 'tab$' 'tab$' "V1 manipulated pitch" 'tab$' 'manipulated_pitch1' 'tab$' "V2 pitch change" 'tab$' 'v2pit_mani' 'tab$' "V2 manipulated pitch" 'tab$' 'manipulated_pitch2' 'newline$'
#		else
```

Appendix C
Background Questionnaire for native English speakers and Information Summary

PARTICIPANT QUESTIONNAIRE

Please fill out the following form before the experiment.

1. Participant Number: **2. Name:**

3. Age: **4. Gender: Male( ) / Female( )**

5. Major: **6. Phone:**

7. Email:

8. My hometown is ____
(city/town, province, country)

9. I can ( ) / can't ( ) speak languages other than English.

Other languages include:

Languages	& Level in speaking and listening (beginning, intermediate, advanced))
1) ____________________________	____________________________
2) ____________________________	____________________________
3) ____________________________	____________________________
4) ____________________________	____________________________

10. Other than English, I also use ______ frequently for everyday life.

11. I have ( ) / don't have ( ) music background.
I have music training in ___________________ for _______ years.

12. I am ( ) / am not ( ) familiar with the idea of WORD STRESS in English.

Summary of NE background information

		Pass 80% Correct Rate	Hometown	Music Background	Type	Music Years	Familiar with Stress
1	NE01	1	Victoria, BC	2	NA	NA	1
2	NE02	0	Portland, Oregon, USA	1	Alto Saxaphone	5	1
3	NE03	0	Constance LakeFirst Nation, Ontario	2	NA	NA	1
4	NE04	1	Victoria, BC	1	Guitar	3	2
5	NE05	1	Yelloknife, NT	1	Piano	5	1
6	NE06	1	Victoria, BC	1	Piano	17	2
7	NE07	0	Victoria, BC	2	NA	NA	2
8	NE08	0	Kincardine, Ont	2	NA	NA	2
9	NE09	1	Summerland, BC	1	Piano Clarinet	6 4	1
10	NE10	1	Victoria, BC	1	Piano Voice Guitar	5 6 7	1
11	NE11	1	Victoria, BC	2	NA	NA	1
12	NE12	0	Ancaster, Ontario	2	NA	NA	1

13	NE13	0	Kitchener-Waterlog, Ontario	1	Percussion	9	2
14	NE14	1	Victoria, BC	1	Violin	16	1
15	NE15	1	Victoria, BC	2	NA	NA	1
16	NE16	1	Cranbrook,BC	1	Piano	8	1
17	NE17	1	Delta, BC	2	NA	NA	1
18	NE18	0	Victoria, BC	2	NA	NA	1
19	NE19	0	Merrit, BC	2	NA	NA	1
20	NE20	1	Victoria, BC	1	Percussion	9	1
21	NE21	1	Vancouver, BC	2	NA	NA	1
22	NE22	1	Victoria, BC	1	Piano	15	2
23	NE23	1	Quito, Ecuador	2	NA	NA	1
24	NE24	1	Victoria, BC	1	Voice	16	1
25	NE25	1	Wetaskiwin, Alberta	2	NA	NA	1

26	NE26	1	Calgary, Alberta	1	flute Saxophone	6 6	2
27	NE27	0	Port Alberni, BC	2	NA	NA	2
28	NE28	1	Victoria, BC	1	piano	17	1
29	NE29	0	Nanaimo, BC	2	NA	NA	1
30	NE30	1	Squamish, BC	2	NA	NA	1
31	NE31	0	Victoria, BC	1	singing	3	1
32	NE32	1	Victoria, BC	1	saxaphone	3	1
33	NE33	1	Brandon, MB	2	NA	NA	1
34	NE34	1	Vancouver, BC	1	piano	10	1
35	NE35	1	Victoria, BC	2	NA	NA	1
36	NE36	0	Mississauga, Ontario	1	Guitar	1	1
37	NE37	1	Victoria, BC	2	NA	NA	1
38	NE38	0	Saanichton, BC	2	NA	NA	1

Note: In the Gender column, 1 indicates male and 2 indicates female. For the column of Foreign Languages, the number following the name of the language indicates the proficiency level of this additional language, 1=beginning, 2=intermediate, 3=advanced; For the column of Mandarin = Primary Language, 1 indicates Yes and 2 indicates No. For the column of Music Background, 1 indicates with music training and 2 indicates no music training. For the column of Familiarity with Stress, 1 indicates Yes and 2 indicates No.

Appendix D
Background Questionnaire for Chinese learners of English and Information Summary

Please fill out the following form before the experiment. You can use Chinese or English. 请您在开始实验前填写以下表格。您可以选用中文或英文作答。

1. Seat Number 座位号**:** **2. Participant Number** 实验者序号**:**

3. Name 中文姓名**:** **4. Age** 年龄**:**

5. Phone 手机/固定电话**:** **6. Gender** 性别**: Male( )/ Female( )**

7. Major 专业**:** **8. Email** 电邮**:**

9. I have studied English for ____ years. 我学习英语已有**____**年了。

10. I come from ___________________(province, city/town).
我的家乡是 **___________________(**省，市/县**)**。

11. I can ( ) / can't ( ) speak foreign languages other than English.
除了英文外，我 会**( )** / 不会**( )** 其他外语。

Other Foreign languages include: 其他语言如下：

Languages **& Level in speaking and listening**

语言 **&**（初级，中级，高级） **(beginning, intermediate, advanced))**

1) _______________________ _______________________

2) _______________________ _______________________

12. Mandarin is( )/isn't( )my primary language for daily communication.
普通话是**( )** / 不是**( )** 我的日常交流用语。

13. I also use ________ dialect for everyday life.
我还经常使用 **________** 方言。

14. I have( )/don't have( ) music background.
我有**( )**/没有**( )**音乐背景。
I have music training in _______ for _________ years.
我受过 **____________** 方面的音乐训练，时间长达 **______**年。

15. I am( )/am not( ) familiar with the idea of word stress in English.
我熟悉**( )** / 不熟悉**( )** 英语单词重音这一概念。

Summary of CE background information

		80%	Major	Age	Gender	English Learning Years	Starting Year (English Learning)	Hometown (City, Province)	Foreign Languages	Mandarin = Primary Language?	Dialect	Music Back-ground?	Type	Music Years	Familiar with Stress?
1	CE01	1	En.	20	2	11	9	Suzhou, Jiangsu	French1	1	Wu	2		NA	1
2	CE02	1	En.	20	1	9	11	Neijing, Sichuan	German1	1	Sichuan	2		NA	1
3	CE03	1	En.	22	2	9	13	Panzhihua, Sichuan	Japanese1; German1	1	NA	2		NA	1
4	CE04	1	En.	20	2	12	8	Qufu, Shandong	French1; Korean1	1	NA	2		NA	1
5	CE05	1	En.	22	2	9	13	Taixing, Jiangsu	Japanese1; French1	1	NA	2		NA	1
6	CE06	1	En.	22	2	9	13	Changzhou, Jiangsu	Japanese1; German1	1	NA	2		NA	1
7	CE07	1	En.	20	2	10	10	Nanjing, Jiangsu	French1	1	NA	2		NA	1
8	CE08	1	En.	21	2	12	9	Dalian, Liaoning	French1; Korean1	1	Dalian	1	Keyboard	3	1
9	CE09	0	En.	21	2	11	10	Xinzhou, Shan1xi	Japanese1	1	Shanxi	2		NA	1
10	CE10	0	En.	20	2	9	11	Jiangyin, Jiangsu	French1	1	Shanghai	2		NA	1
11	CE13	0	En.	22	2	9	13	Jiaxing, Zhejiang	French1; Korean1	1	Zhejiang	2		NA	1
12	CE14	0	En.	22	2	9	13	Yanchen, Jiangsu	Japanese1	1	NA	2		NA	1
13	CE15	1	En.	20	2	9	11	Jinghua, Zhejiang	French1; Japanese1	1	Jinhua	2		NA	1
14	CE16	0	En.	21	2	9	12	Yulin, Shan3xi	German2; Japanese1	1	NA	2		NA	1
15	CE17	1	En.	22	1	9	13	Yanchen, Jiangsu	French1	1	Yanchen	1	Saxophone	1	1
16	CE18	1	En.	21	2	13	8	Suzhou, Jiangsu	Japanese1	1	Suzhou	2		NA	1
17	CE19	1	En.	21	1	12	9	Tianjing	French1	1	NA	2		NA	1
18	CE20	0	En.	22	1	9	13	Baoji, Shan3xi	Japanese1	1	Shan3xi	2		NA	1

19	CE21	1	En.	21	2	10	11	Xuzhou, Jiangsu	Japanese1	1	Xuzhou	2		NA	1
20	CE22	1	En.	23	2	9	14	Wuxi, Jiangsu	Japanese1	1	Wuxi	2		NA	1
21	CE23	1	En.	20	2	9	11	Nantong, Jiangsu	Japanese1	1	Rudong	2		NA	1
22	CE24	1	En.	20	2	9	11	Zhangjiagang, Jiangsu	Japanese1	1	Zhangjiagang	2		NA	1
23	CE25	0	En.	20	2	10	10	Yangzhong, Jiangsu	German1; Japanese1	1	NA	1	Keyboard	2	1
24	CE26	1	En.	22	2	13	9	Baotou, Neimenggu	French1	1	NA	1	Violin	7	1
25	CE27	1	En.	22	2	9	13	Dalian, Liaoning	German1	1	NA	2		NA	1
26	CE28	1	En.	20	2	9	11	Yichun, Jiangxi	Korean1, French1	1	Yichun	1	Piano	6	1
27	CE29	1	En.	21	2	13	8	Beijing	Japanese1	1	NA	2		NA	1
28	CE30	1	En.	21	2	9	12	Panzhihua, Sichuan	German1	1	Sichuan	2		NA	1
29	CE31	0	En.	21	2	9	12	Yangzhou, Jiangsu	Japanese1	1	Yangzhou	2		NA	1
30	CE32	1	En.	21	2	9	12	Deyang, Sichuan	German1	1	Sichuan	2		NA	1
31	CE33	0	En.	21	2	9	12	Wuxi, Jiangsu	German1	1	Wu	2		NA	2
32	CE36	1	Sci.	19	1	8	11	Guangyuan, Sichuan	French1	1	Sichuan	1	Bass	2	1
33	CE37	0	Sci.	20	1	8	12	Henyang, Hunan	NA	1	Hunan	2		NA	2
34	CE39	0	Sci.	19	2	8	11	Fuan, Fujian	NA	1	Fuan	2		NA	2
35	CE40	0	Sci.	20	1	10	10	Shijiazhuang, Hebei	NA	1	NA	2		NA	1
36	CE41	1	Sci.	21	1	7	14	Yizhang, Hunan	NA	1	Yizhang	1	Piano	1	1
37	CE42	1	Sci.	18	2	7	11	Leshan, Sichuan	NA	1	Sichuan	1	Keyboard	5	1
38	CE45	0	Sci.	20	1	8	12	Yixing, Jiangsu	French1	1	Yixing	2		NA	2
39	CE46	1	Sci.	20	2	12	8	Guangzhou, Guangdong	Korean1	2	Cantonese	2		NA	1
40	CE47	0	Sci.	21	1	8	13	Taiyuan,	NA	1	Shan1xi	2		NA	1

								Shan1xi							
41	CE48	0	Sci.	20	1	8	12	Liaoyang, Liaoning	NA	1	NA	2		NA	1
42	CE49	0	Sci.	20	1	8	12	Nantong, Jiangsu	NA	1	NA	2		NA	1
43	CE50	0	Sci.	21	1	9	12	Jiangyan, Jiangsu	NA	2	Jiangyan	2		NA	1
44	CE53	0	Sci.	20	2	8	12	Datong, Shan1xi	NA	1	Datong	1	Piano	4	1
45	CE54	0	Sci.	21	2	8	13	Yingkou, Liaoning	Japanese1	1	Dongbei	2		NA	1
46	CE55	1	Sci.	19	1	8	11	Xiantao, Hubei	NA	1	NA	2		NA	2
47	CE56	1	Sci.	19	2	8	11	Linfen, Shan1xi	NA	1	NA	2		NA	1
48	CE57	0	Sci.	20	2	11	9	Chongqing	NA	2	Sichuan	2		NA	2
49	CE60	1	Sci.	21	2	9	12	Wushan, Heilongjiang		1	Dongbei	2		NA	1
50	CE61	1	Sci.	20	1	7	13	Zhangzhou, Fujian	NA	1	Min	2		NA	1
51	CE62	1	Sci.	20	1	8	12	Taiyuan, Shan1xi	NA	1	NA	2		NA	1
52	CE63	0	Sci.	19	2	8	11	Jingdezhen, Jiangxi	NA	1	Jingdezhen	2		NA	2
53	CE64	1	Sci.	22	2	8	14	Dongtai, Jiangsu	NA	2	Dongtai	2		NA	1
54	CE67	1	Sci.	20	2	10	10	Lishui, Zhejiang	French1; Japanese1	1	Suichang	1	Piano	3	1
55	CE68	1	Sci.	19	2	12	7	Beijing	French1	1	NA	1	Piano	3	1
56	CE69	0	Sci.	20	1	8	12	Dalian, Liaoning	NA	1	Dalian	2		NA	2
57	CE71	0	Sci.	20	2	8	12	Jiangying, Jiangsu	NA	1	Jiangyin	2		NA	1

Note: In the Gender column, 1 indicates male and 2 indicates female. For the column of Foreign Languages, the number following the name of the language indicates the proficiency level of this additional language, 1=beginning, 2=intermediate, 3=advanced; For the column of Mandarin = Primary Language, 1 indicates Yes and 2 indicates No. For the column of Music Background, 1 indicates with music training and 2 indicates no music training. For the column of Familiarity with Stress, 1 indicates Yes and 2 indicates No.

Appendix E
Ordering of stimuli slots

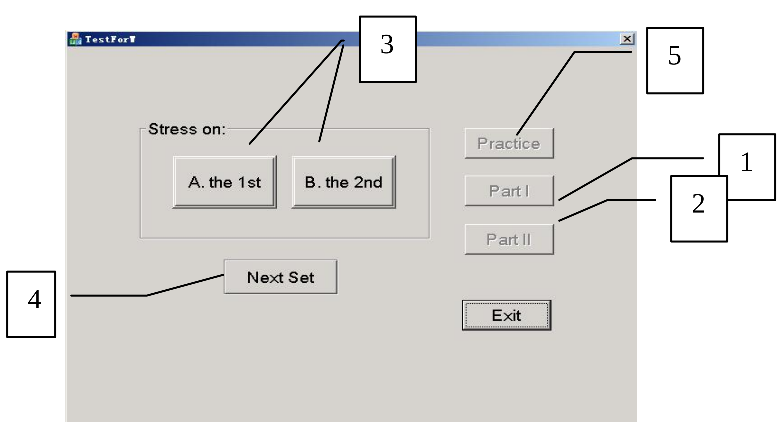
The following table shows the exact ordering of the 475 stimuli slots. Nt, Nn and Nl refer to the three nonsense words, *tet.sep, niz.dit, lat.mab,* respectively. A token of the same word form will never appear in consecutive positions. RP refers to real English words with 2 possible stress patterns. The number following RP refers to the stress position: 1 for stress on the first syllable and 2 for stress on the second syllable. 'r' refers to real English words with either stress on the first syllable or the second, which is indicated by the number following 'r'.

Trial / Block	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19
1	Nt	Nn	Nl	RP 1	Nn	Nl	Nt	r2	Nl	Nt	Nn	r1	Nt	Nn	Nl	RP 2	Nn	Nl	Nt
2	Nl	Nt	Nn	r2	Nt	Nn	Nl	RP 1	Nn	Nl	Nt	r2	Nl	Nt	Nn	r1	Nt	Nn	Nl
3	Nn	Nl	Nt	r2	Nl	Nt	Nn	r1	Nt	Nn	Nl	R P2	Nn	Nl	Nt	r1	Nl	Nt	Nn
4	Nt	Nn	Nl	r2	Nn	Nl	Nt	r1	Nl	Nt	Nn	r2	Nt	Nn	Nl	RP 1	Nn	Nl	Nt
5	Nl	Nt	Nn	RP 2	Nt	Nn	Nl	r1	Nn	Nl	Nt	r2	Nl	Nt	Nn	RP 1	Nt	Nn	Nl
6	Nn	Nl	Nt	r1	Nl	Nt	Nn	RP 2	Nt	Nn	Nl	r1	Nn	Nl	Nt	r2	Nl	Nt	Nn
7	Nt	Nn	Nl	r1	Nn	Nl	Nt	r2	Nl	Nt	Nn	R P1	Nt	Nn	Nl	r2	Nn	Nl	Nt
8	Nl	Nt	Nn	r1	Nt	Nn	Nl	r2	Nn	Nl	Nt	r1	Nl	Nt	Nn	RP 2	Nt	Nn	Nl
9	Nn	Nl	Nt	RP 1	Nl	Nt	Nn	r2	Nt	Nn	Nl	r1	Nn	Nl	Nt	RP 2	Nl	Nt	Nn
10	Nt	Nn	Nl	r2	Nn	Nl	Nt	RP 1	Nl	Nt	Nn	r1	Nt	Nn	Nl	r1	Nn	Nl	Nt
11	Nl	Nt	Nn	r2	Nt	Nn	Nl	r1	Nn	Nl	Nt	R P2	Nl	Nt	Nn	r1	Nt	Nn	Nl
12	Nn	Nl	Nt	r2	Nl	Nt	Nn	r1	Nt	Nn	Nl	r2	Nn	Nl	Nt	RP 1	Nl	Nt	Nn
13	Nt	Nn	Nl	RP 2	Nn	Nl	Nt	r1	Nl	Nt	Nn	r2	Nt	Nn	Nl	RP 1	Nn	Nl	Nt
14	Nl	Nt	Nn	r1	Nt	Nn	Nl	RP 2	Nn	Nl	Nt	r2	Nl	Nt	Nn	r2	Nt	Nn	Nl
15	Nn	Nl	Nt	r1	Nl	Nt	Nn	r2	Nt	Nn	Nl	R P1	Nn	Nl	Nt	r2	Nl	Nt	Nn
16	Nt	Nn	Nl	r1	Nn	Nl	Nt	r2	Nl	Nt	Nn	r1	Nt	Nn	Nl	RP 2	Nn	Nl	Nt
17	Nl	Nt	Nn	RP 1	Nt	Nn	Nl	r2	Nn	Nl	Nt	r1	Nl	Nt	Nn	RP 2	Nt	Nn	Nl
18	Nn	Nl	Nt	r2	Nl	Nt	Nn	RP 1	Nt	Nn	Nl	r1	Nn	Nl	Nt	r1	Nl	Nt	Nn
19	Nt	Nn	Nl	r2	Nn	Nl	Nt	r1	Nl	Nt	Nn	R P2	Nt	Nn	Nl	r1	Nn	Nl	Nt
20	Nl	Nt	Nn	r2	Nt	Nn	Nl	r1	Nn	Nl	Nt	r2	Nl	Nt	Nn	RP 1	Nt	Nn	Nl
21	Nn	Nl	Nt	RP 2	Nl	Nt	Nn	r1	Nt	Nn	Nl	r2	Nn	Nl	Nt	RP 1	Nl	Nt	Nn
22	Nt	Nn	Nl	r1	Nn	Nl	Nt	RP 2	Nl	Nt	Nn	r2	Nt	Nn	Nl	r2	Nn	Nl	Nt
23	Nl	Nt	Nn	r1	Nt	Nn	Nl	r2	Nn	Nl	Nt	R P1	Nl	Nt	Nn	r2	Nt	Nn	Nl
24	Nn	Nl	Nt	r1	Nl	Nt	Nn	r2	Nt	Nn	Nl	r1	Nn	Nl	Nt	RP 2	Nl	Nt	Nn
25	Nt	Nn	Nl	RP 1	Nn	Nl	Nt	r2	Nl	Nt	Nn	r1	Nt	Nn	Nl	RP 2	Nn	Nl	Nt

Appendix F
Instruction sheet for perception test

Instructions for perception test

1. In the perception test, you will listen to some two-syllable English words and English nonsense words. For the same word form (same spelling), stress can be put on the first or the second syllable. For each word, you will have to decide whether the stress is on the 1st syllable or the 2nd syllable. Trust your **EARS**.
2. Some of the words may seem similar, but no two words are exactly the same. They are all **DIFFERENT**. Some of the words may seem more difficult than others. Therefore, if you aren't completely sure, make a **GUESS**. Base your choice on your **BEST** guess, **NOT WILD** guess.
3. There are 25 sets of words, divided into 2 parts, 15 sets in Part I (see 1 in Fig. 1 below) and 10 in Part II (see 2 in Fig. 1). Between the 2 parts, you will have a short break. Please come to me and pick your gift during the break. **DON'T START PART II** before you are instructed to do so.
4. In every set, 19 words will be played to you. After listening to each word carefully, judge where the stress is by clicking on the corresponding button "the 1st" or "the 2nd" (see 3 in Fig. 1) on the screen. Please respond as **FAST** as you can. You only have **3** seconds to respond. It is recommended that you **NOT MOVE** your mouse around during the test. Leave your mouse pointer close to the choices will save you a lot of time! You won't be able to change once you make your choice. So, be cautious! The next word will be played automatically after 3 seconds, no matter if you have made the choice or not.
5. At the end of each set, a window will pop up indicating the end of the current set. After you click "Ok", "Next Set" (see 4 in Fig. 1) will appear **30 seconds** later. It is to say, you **MUST** have a rest of at least 20 seconds between 2 sets. You can click "Next Set" immediately to go to the next set. Or if you need more rest, click it whenever you feel ready. During this rest period, a recording of an English conversation or monologue will be played to you. You **DON'T** have to finish listening to the recording. Be prepared to take longer if you want to listen to the whole recording.
6. Before the real test, there is a practice section, which includes 2 sets of words. Click the "Practice" button (see 5 in Fig. 1) to try this out at least once. You won't be able to start the test without first going through the practice section. During the practice section, please adjust your **VOLUME** to a comfortable level (It will make your task extra hard if the word is either too loud or too low). Click "**Part I**" button when you feel ready to begin the real test.
7. We are about to begin. If you have any questions, please ask them now. If you have questions during the test, ask them in the rest period. Now please click "Test.exe" to begin the test.



Appendix G
Complete statistics of the twenty-five individual NE logistic regression models

The following table presents the full statistics of the logistics regression models constructed for the twenty-five individual NE participants. The first column indicates the B value for each of the three cues and the constant. The Wald statistic is included next to each B value, including the *df* for this Wald statistic and its significance level. The Model chi-squared statistic, with its *df* and the significance level of the model, are listed next. The Cox & Snell R^2 and Nagelkerke R^2 are two statistics to represent how much variance can be predicted in the response with the three predictors.

		B	Wald	df	Sig.	Exp(B)	Model chi-square	df	Sig.
NE01	Pitch	0.288	9.845	1	0.002	1.333	121.855	3	0.000
	Duration	0.240	6.953	1	0.008	1.271	Cox &		Nagelkerke
	Intensity	0.942	78.834	1	0.000	2.566	Snell R^2		R^2
	Constant	-5.107	69.791	1	0.000	0.006	0.277		0.377
NE04	Pitch	0.036	0.229	1	0.632	1.037	15.607	3	0.001
	Duration	-0.005	0.004	1	0.947	0.995	Cox &		Nagelkerke
	Intensity	0.297	14.824	1	0.000	1.346	Snell R^2		R^2
	Constant	-1.341	10.307	1	0.001	0.262	0.041		0.055
NE05	Pitch	1.140	93.234	1	0.000	3.126	163.012	3	0.000
	Duration	0.296	9.475	1	0.002	1.344	Cox &		Nagelkerke
	Intensity	0.407	17.183	1	0.000	1.503	Snell R^2		R^2
	Constant	-5.583	72.348	1	0.000	0.004	0.358		0.477
NE06	Pitch	-0.360	9.198	1	0.002	0.698	267.208	3	0.000
	Duration	0.380	10.248	1	0.001	1.463	Cox &		Nagelkerke
	Intensity	1.893	104.868	1	0.000	6.639	Snell R^2		R^2
	Constant	-6.152	61.198	1	0.000	0.002	0.511		0.682
NE09	Pitch	0.412	22.237	1	0.000	1.509	87.264	3	0.000
	Duration	0.031	0.140	1	0.708	1.032	Cox &		Nagelkerke
	Intensity	0.685	55.809	1	0.000	1.984	Snell R^2		R^2
	Constant	-3.770	52.204	1	0.000	0.023	0.209		0.280
NE10	Pitch	0.797	65.760	1	0.000	2.219	95.163	3	0.000
	Duration	0.234	7.041	1	0.008	1.263	Cox &		Nagelkerke
	Intensity	0.234	7.149	1	0.008	1.264	Snell R^2		R^2
	Constant	-4.446	61.496	1	0.000	0.012	0.226		0.308
NE11	Pitch	0.472	33.029	1	0.000	1.603	42.026	3	0.000
	Duration	0.179	5.057	1	0.025	1.196	Cox &		Nagelkerke
	Intensity	0.110	1.976	1	0.160	1.116	Snell R^2		R^2
	Constant	-2.264	25.076	1	0.000	0.104	0.109		0.146

NE14	Pitch	0.674	52.542	1	0.000	1.962	102.607	3	0.000
	Duration	0.369	17.845	1	0.000	1.446	Cox & Snell R^2		Nagelkerke R^2
	Intensity	0.453	26.258	1	0.000	1.574			
	Constant	-4.523	66.084	1	0.000	0.011	0.240		0.321
NE15	Pitch	0.616	44.328	1	0.000	1.852	54.674	3	0.000
	Duration	0.111	1.696	1	0.193	1.118	Cox & Snell R^2		Nagelkerke R^2
	Intensity	0.112	1.716	1	0.190	1.118			
	Constant	-3.416	43.119	1	0.000	0.033	0.137		0.191
NE16	Pitch	4.255	59.475	1	0.000	70.436	398.133	3	0.000
	Duration	0.556	8.842	1	0.003	1.743	Cox & Snell R^2		Nagelkerke R^2
	Intensity	0.525	8.062	1	0.005	1.691			
	Constant	-17.431	52.225	1	0.000	0.000	0.654		0.878
NE17	Pitch	0.722	54.821	1	0.000	2.058	91.360	3	0.000
	Duration	0.235	7.070	1	0.008	1.265	Cox & Snell R^2		Nagelkerke R^2
	Intensity	0.409	20.277	1	0.000	1.505			
	Constant	-4.894	68.572	1	0.000	0.007	0.217		0.298
NE20	Pitch	1.005	76.022	1	0.000	2.731	172.442	3	0.000
	Duration	0.515	25.960	1	0.000	1.673	Cox & Snell R^2		Nagelkerke R^2
	Intensity	0.738	47.369	1	0.000	2.092			
	Constant	-6.876	87.393	1	0.000	0.001	0.369		0.493
NE21	Pitch	1.380	91.425	1	0.000	3.975	220.410	3	0.000
	Duration	0.595	27.711	1	0.000	1.814	Cox & Snell R^2		Nagelkerke R^2
	Intensity	0.826	45.920	1	0.000	2.285			
	Constant	-8.795	91.643	1	0.000	0.000	0.444		0.595
NE22	Pitch	0.513	27.226	1	0.000	1.670	70.218	3	0.000
	Duration	0.118	1.633	1	0.201	1.125	Cox & Snell R^2		Nagelkerke R^2
	Intensity	0.601	35.599	1	0.000	1.823			
	Constant	-4.983	63.933	1	0.000	0.007	0.171		0.249
NE23	Pitch	0.691	54.755	1	0.000	1.996	88.328	3	0.000
	Duration	0.359	17.021	1	0.000	1.431	Cox & Snell R^2		Nagelkerke R^2
	Intensity	0.261	9.227	1	0.002	1.298			
	Constant	-4.339	61.677	1	0.000	0.013	0.213		0.287
NE24	Pitch	0.819	67.729	1	0.000	2.269	125.235	3	0.000
	Duration	0.204	5.378	1	0.020	1.227	Cox & Snell R^2		Nagelkerke R^2
	Intensity	0.575	37.143	1	0.000	1.778			
	Constant	-4.790	67.622	1	0.000	0.008	0.284		0.379
NE25	Pitch	3.449	66.826	1	0.000	31.466	372.836	3	0.000
	Duration	0.067	0.201	1	0.654	1.070	Cox & Snell R^2		Nagelkerke R^2
	Intensity	0.022	0.022	1	0.881	1.023			
	Constant	-11.291	55.714	1	0.000	0.000	0.630		0.842
NE26	Pitch	0.604	48.115	1	0.000	1.830	75.095	3	0.000
	Duration	0.177	4.702	1	0.030	1.194	Cox & Snell R^2		Nagelkerke R^2
	Intensity	0.345	17.029	1	0.000	1.411			
	Constant	-3.570	50.714	1	0.000	0.028	0.181		0.242
NE28	Pitch	1.145	64.157	1	0.000	3.142	137.324	3	0.000
	Duration	0.304	7.751	1	0.005	1.355	Cox & Snell R^2		Nagelkerke R^2
	Intensity	0.694	33.597	1	0.000	2.002			
	Constant	-4.552	46.297	1	0.000	0.011	0.307		0.458

NE30	Pitch	0.608	44.852	1	0.000	1.837	92.592	3	0.000
	Duration	0.377	18.710	1	0.000	1.457	Cox & Snell R^2		Nagelkerke R^2
	Intensity	0.443	25.281	1	0.000	1.557			
	Constant	-4.647	68.301	1	0.000	0.010	0.219		0.294
NE32	Pitch	0.321	13.263	1	0.000	1.378	101.778	3	0.000
	Duration	0.256	8.610	1	0.003	1.292	Cox & Snell R^2		Nagelkerke R^2
	Intensity	0.786	66.160	1	0.000	2.195			
	Constant	-4.588	65.551	1	0.000	0.010	0.238		0.321
NE33	Pitch	0.489	29.859	1	0.000	1.631	97.575	3	0.000
	Duration	-0.101	1.425	1	0.233	0.904	Cox & Snell R^2		Nagelkerke R^2
	Intensity	0.703	56.676	1	0.000	2.021			
	Constant	-3.626	48.724	1	0.000	0.027	0.230		0.309
NE34	Pitch	0.803	63.982	1	0.000	2.232	125.366	3	0.000
	Duration	0.073	0.673	1	0.412	1.075	Cox & Snell R^2		Nagelkerke R^2
	Intensity	0.642	43.716	1	0.000	1.901			
	Constant	-4.650	64.128	1	0.000	0.010	0.288		0.384
NE35	Pitch	0.390	18.241	1	0.000	1.477	33.436	3	0.000
	Duration	0.128	2.109	1	0.146	1.137	Cox & Snell R^2		Nagelkerke R^2
	Intensity	0.312	12.054	1	0.001	1.367			
	Constant	-1.308	7.822	1	0.005	0.270	0.086		0.126
NE37	Pitch	0.000	0.000	1	0.996	1.000	42.450	3	0.000
	Duration	0.286	12.722	1	0.000	1.331	Cox & Snell R^2		Nagelkerke R^2
	Intensity	0.432	28.029	1	0.000	1.541			
	Constant	-2.510	30.602	1	0.000	0.081	0.108		0.145

Appendix H
Complete statistics of the thirty-four individual CE logistic regression models

The following table presents the full statistics of the logistics regression models constructed for the thirty-four individual CE participants. The first column indicates the B value for each of the three cues and the constant. The Wald statistic is included next to each B value, including the *df* for this Wald statistic and its significance level. The Model chi-square statistic, with its *df* and the significance level of the model, are listed next. The Cox & Snell R^2 and Nagelkerke R^2 are two statistics to represent how much variance can be predicted in the response with the three predictors.

		B	Wald	df	Sig.	Exp(B)	Model chi-square	df	Sig.
CE01	Pitch	3.242	77.276	1	0.000	25.585	363.169	3	0.000
	Duration	0.088	0.351	1	0.554	1.092	Cox & Snell R^2		Nagelkerke R^2
	Intensity	-0.066	0.198	1	0.657	0.936			
	Constant	-10.783	60.885	1	0.000	0.000	0.620		0.832
CE02	Pitch	2.284	97.500	1	0.000	9.818	306.273	3	0.000
	Duration	0.030	0.061	1	0.805	1.031	Cox & Snell R^2		Nagelkerke R^2
	Intensity	0.046	0.137	1	0.711	1.047			
	Constant	-7.035	61.105	1	0.000	0.001	0.558		0.744
CE03	Pitch	1.124	95.982	1	0.000	3.076	154.950	3	0.000
	Duration	0.282	8.861	1	0.003	1.326	Cox & Snell R^2		Nagelkerke R^2
	Intensity	0.137	2.169	1	0.141	1.147			
	Constant	-5.063	66.950	1	0.000	0.006	0.338		0.454
CE04	Pitch	1.564	103.239	1	0.000	4.779	209.190	3	0.000
	Duration	-0.125	1.349	1	0.245	0.883	Cox & Snell R^2		Nagelkerke R^2
	Intensity	0.019	0.033	1	0.856	1.020			
	Constant	-5.356	58.912	1	0.000	0.005	0.433		0.589
CE05	Pitch	1.220	99.650	1	0.000	3.387	164.829	3	0.000
	Duration	-0.003	0.001	1	0.974	0.997	Cox & Snell R^2		Nagelkerke R^2
	Intensity	0.194	4.022	1	0.045	1.214			
	Constant	-4.837	60.395	1	0.000	0.008	0.357		0.482
CE06	Pitch	4.478	51.348	1	0.000	88.098	404.873	3	0.000
	Duration	0.268	2.304	1	0.129	1.307	Cox & Snell R^2		Nagelkerke R^2
	Intensity	0.059	0.119	1	0.731	1.061			
	Constant	-15.510	48.327	1	0.000	0.000	0.660		0.884
CE07	Pitch	3.919	46.830	1	0.000	50.352	378.657	3	0.000

	Duration	-0.001	0.000	1	0.994	0.999	Cox & Snell R^2		Nagelkerke R^2
	Intensity	-0.070	0.211	1	0.646	0.932			
	Constant	-11.684	40.053	1	0.000	0.000	0.643		0.857
CE08	Pitch	3.750	55.265	1	0.000	42.508	382.919	3	0.000
	Duration	0.093	0.369	1	0.544	1.097	Cox & Snell R^2		Nagelkerke R^2
	Intensity	-0.208	1.821	1	0.177	0.812			
	Constant	-11.248	46.057	1	0.000	0.000	0.640		0.854
CE15	Pitch	1.040	54.576	1	0.000	2.828	84.489	3	0.000
	Duration	0.074	0.490	1	0.484	1.076	Cox & Snell R^2		Nagelkerke R^2
	Intensity	0.150	2.012	1	0.156	1.162			
	Constant	-5.799	56.773	1	0.000	0.003	0.202		0.322
CE17	Pitch	1.681	109.410	1	0.000	5.372	237.507	3	0.000
	Duration	-0.041	0.142	1	0.706	0.960	Cox & Snell R^2		Nagelkerke R^2
	Intensity	0.073	0.455	1	0.500	1.076			
	Constant	-5.353	58.723	1	0.000	0.005	0.473		0.631
CE18	Pitch	0.498	32.794	1	0.000	1.646	77.164	3	0.000
	Duration	-0.118	2.017	1	0.156	0.889	Cox & Snell R^2		Nagelkerke R^2
	Intensity	0.539	38.035	1	0.000	1.715			
	Constant	-2.380	25.994	1	0.000	0.093	0.187		0.251
CE19	Pitch	0.897	82.964	1	0.000	2.451	113.156	3	0.000
	Duration	-0.052	0.361	1	0.548	0.950	Cox & Snell R^2		Nagelkerke R^2
	Intensity	0.125	2.112	1	0.146	1.134			
	Constant	-3.033	37.231	1	0.000	0.048	0.261		0.348
CE21	Pitch	1.352	102.967	1	0.000	3.865	184.014	3	0.000
	Duration	0.108	1.177	1	0.278	1.114	Cox & Snell R^2		Nagelkerke R^2
	Intensity	0.127	1.639	1	0.200	1.136			
	Constant	-5.526	66.855	1	0.000	0.004	0.388		0.525
CE22	Pitch	0.875	77.030	1	0.000	2.398	102.805	3	0.000
	Duration	0.083	0.908	1	0.341	1.086	Cox & Snell R^2		Nagelkerke R^2
	Intensity	0.043	0.247	1	0.619	1.044			
	Constant	-3.232	39.417	1	0.000	0.039	0.247		0.330
CE23	Pitch	1.588	104.582	1	0.000	4.894	216.865	3	0.000
	Duration	0.223	4.320	1	0.038	1.250	Cox & Snell R^2		Nagelkerke R^2
	Intensity	0.100	0.897	1	0.344	1.106			
	Constant	-6.649	72.663	1	0.000	0.001	0.439		0.596
CE24	Pitch	0.553	36.756	1	0.000	1.739	46.926	3	0.000
	Duration	0.022	0.065	1	0.798	1.022	Cox & Snell R^2		Nagelkerke R^2
	Intensity	0.197	5.262	1	0.022	1.217			
	Constant	-3.260	40.727	1	0.000	0.038	0.118		0.167
CE26	Pitch	2.846	82.245	1	0.000	17.211	345.054	3	0.000
	Duration	0.384	7.081	1	0.008	1.469	Cox & Snell R^2		Nagelkerke R^2
	Intensity	0.259	3.391	1	0.066	1.295			
	Constant	-10.456	64.849	1	0.000	0.000	0.603		0.805
CE27	Pitch	1.197	102.143	1	0.000	3.310	164.066	3	0.000

	Duration	0.019	0.042	1	0.838	1.019	Cox & Snell R^2		Nagelkerke R^2
	Intensity	0.076	0.645	1	0.422	1.079			
	Constant	-4.057	49.790	1	0.000	0.017	0.360		0.481
CE28	Pitch	5.117	25.998	1	0.000	166.802	406.383	3	0.000
	Duration	0.097	0.348	1	0.555	1.102	Cox & Snell R^2		Nagelkerke R^2
	Intensity	-0.303	3.298	1	0.069	0.739			
	Constant	-14.627	22.820	1	0.000	0.000	0.666		0.887
CE29	Pitch	3.471	66.373	1	0.000	32.154	373.956	3	0.000
	Duration	0.045	0.090	1	0.764	1.046	Cox & Snell R^2		Nagelkerke R^2
	Intensity	0.225	2.184	1	0.139	1.253			
	Constant	-11.861	56.600	1	0.000	0.000	0.631		0.843
CE30	Pitch	1.744	107.300	1	0.000	5.721	243.227	3	0.000
	Duration	-0.046	0.171	1	0.680	0.955	Cox & Snell R^2		Nagelkerke R^2
	Intensity	-0.030	0.074	1	0.786	0.970			
	Constant	-5.227	54.724	1	0.000	0.005	0.486		0.649
CE32	Pitch	5.047	66.592	1	0.000	155.499	422.418	3	0.000
	Duration	0.187	0.728	1	0.394	1.205	Cox & Snell R^2		Nagelkerke R^2
	Intensity	-0.187	0.728	1	0.394	0.830			
	Constant	-17.663	55.398	1	0.000	0.000	0.676		0.914
CE36	Pitch	1.166	99.055	1	0.000	3.209	156.194	3	0.000
	Duration	0.024	0.064	1	0.801	1.024	Cox & Snell R^2		Nagelkerke R^2
	Intensity	-0.018	0.038	1	0.846	0.982			
	Constant	-3.823	46.382	1	0.000	0.022	0.349		0.467
CE41	Pitch	2.378	94.710	1	0.000	10.779	299.659	3	0.000
	Duration	0.106	0.686	1	0.408	1.112	Cox & Snell R^2		Nagelkerke R^2
	Intensity	0.145	1.287	1	0.257	1.156			
	Constant	-9.119	69.484	1	0.000	0.000	0.551		0.746
CE42	Pitch	1.329	104.226	1	0.000	3.777	185.585	3	0.000
	Duration	0.170	2.945	1	0.086	1.185	Cox & Snell R^2		Nagelkerke R^2
	Intensity	0.174	3.100	1	0.078	1.190			
	Constant	-5.526	68.423	1	0.000	0.004	0.393		0.527
CE46	Pitch	3.469	74.870	1	0.000	32.096	361.560	3	0.000
	Duration	0.373	5.207	1	0.022	1.452	Cox & Snell R^2		Nagelkerke R^2
	Intensity	-0.102	0.426	1	0.514	0.903			
	Constant	-12.923	59.886	1	0.000	0.000	0.624		0.842
CE55	Pitch	0.809	73.257	1	0.000	2.245	94.501	3	0.000
	Duration	-0.028	0.110	1	0.740	0.972	Cox & Snell R^2		Nagelkerke R^2
	Intensity	-0.086	1.026	1	0.311	0.918			
	Constant	-2.090	20.378	1	0.000	0.124	0.228		0.303
CE56	Pitch	0.297	14.213	1	0.000	1.346	18.716	3	0.000
	Duration	0.054	0.481	1	0.488	1.055	Cox & Snell R^2		Nagelkerke R^2
	Intensity	0.148	3.624	1	0.057	1.160			
	Constant	-2.011	20.989	1	0.000	0.134	0.049		0.067
CE60	Pitch	1.157	98.868	1	0.000	3.182	158.566	3	0.000

	Duration	0.186	3.829	1	0.050	1.204	Cox & Snell R^2		Nagelkerke R^2
	Intensity	0.144	2.329	1	0.127	1.155			
	Constant	-4.584	58.017	1	0.000	0.010	0.350		0.467
CE61	Pitch	2.846	82.245	1	0.000	17.211	345.054	3	0.000
	Duration	0.384	7.081	1	0.008	1.469	Cox & Snell R^2		Nagelkerke R^2
	Intensity	0.259	3.391	1	0.066	1.295			
	Constant	-10.456	64.849	1	0.000	0.000	0.603		0.805
CE62	Pitch	1.273	104.088	1	0.000	3.572	180.595	3	0.000
	Duration	-0.022	0.051	1	0.821	0.978	Cox & Snell R^2		Nagelkerke R^2
	Intensity	0.302	9.284	1	0.002	1.352			
	Constant	-4.615	56.905	1	0.000	0.010	0.385		0.514
CE64	Pitch	1.137	89.542	1	0.000	3.119	139.505	3	0.000
	Duration	0.002	0.001	1	0.979	1.002	Cox & Snell R^2		Nagelkerke R^2
	Intensity	0.000	0.000	1	1.000	1.000			
	Constant	-4.385	52.813	1	0.000	0.012	0.313		0.431
CE67	Pitch	1.488	105.119	1	0.000	4.429	209.546	3	0.000
	Duration	0.335	9.944	1	0.002	1.398	Cox & Snell R^2		Nagelkerke R^2
	Intensity	0.109	1.090	1	0.297	1.115			
	Constant	-6.417	72.740	1	0.000	0.002	0.432		0.581
CE68	Pitch	3.495	61.096	1	0.000	32.963	373.288	3	0.000
	Duration	0.208	1.912	1	0.167	1.231	Cox & Snell R^2		Nagelkerke R^2
	Intensity	-0.109	0.543	1	0.461	0.896			
	Constant	-10.606	49.657	1	0.000	0.000	0.632		0.843