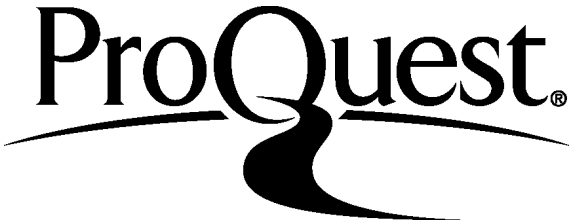
L2 PERCEPTION AND PRODUCTION OF THREE ENGLISH PROSODIC PATTERNS

by

Nadya A. Pincus

A dissertation submitted to the Faculty of the University of Delaware in partial fulfillment of the requirements for the degree of Doctor of Philosophy in Linguistics

Spring 2016



L2 PERCEPTION AND PRODUCTION OF

THREE ENGLISH PROSODIC PATTERNS

by

Nadya A. Pincus

Approved: _____

Benjamin Bruening, Ph.D.
Chair of the Department of Linguistics & Cognitive Science

Approved: _____

George H. Watson, Ph.D.
Dean of the College of Arts and Sciences

Approved: _____

Ann L. Ardis, Ph.D.
Senior Vice Provost for Graduate and Professional Education

I certify that I have read this dissertation and that in my opinion it meets the academic and professional standard required by the University as a dissertation for the degree of Doctor of Philosophy.

Signed: _____

Irene Vogel, Ph.D.
Professor in charge of dissertation

I certify that I have read this dissertation and that in my opinion it meets the academic and professional standard required by the University as a dissertation for the degree of Doctor of Philosophy.

Signed: _____

Jeffrey Heinz, Ph.D.
Member of dissertation committee

I certify that I have read this dissertation and that in my opinion it meets the academic and professional standard required by the University as a dissertation for the degree of Doctor of Philosophy.

Signed: _____

Arild Hestvik, Ph.D.
Member of dissertation committee

I certify that I have read this dissertation and that in my opinion it meets the academic and professional standard required by the University as a dissertation for the degree of Doctor of Philosophy.

Signed: _____

Conxita Lleó, Ph.D.
Member of dissertation committee

ACKNOWLEDGMENTS

First and foremost, I would like to thank my advisor, Dr. Irene Vogel, for her continued patience and support of me throughout what has been a much longer than ideal PhD process. Her persistence in pushing me through to this point is no doubt the reason I have made it this far. I am indebted to her for helping me from the beginning of graduate school through the qualifying exams and qualifying paper, and especially through my many struggles in finishing this dissertation. She aided me in finding a suitable dissertation topic, helped to turn disappointing experimental results into more successful results by fine-tuning methodologies along the way, and just generally supported me beyond what most would deem necessary and reasonable for an advisor to do.

I am also grateful to Dr. Jeffrey Heinz, who encouraged me and pushed me during parts of my graduate studies, and for his patience and valuable comments on my work along the way.

Additionally, I'd like to thank Dr. Arild Hestvik and Dr. Conxita Lleó for their comments on my work and their patience and flexibility.

I am grateful to the ELI, which has both employed me during much of my graduate studies, and supported me through conducting my research there. Also a big thanks to all the students who graciously participated in my experiments. Many thanks also go to Toni McLaughlan, who encouraged many of her students to do my experiments.

I am especially thankful to members of the phonology lab meetings and dissertation group meetings, especially Angeliki Athanasopoulou for extremely valuable discussion about my work along the way.

In addition, I am thankful for the existence of and the members of the stress lab group – this project was so fascinating to me, it was one of the things that sustained me through the PhD, even if it took some time out of my own research.

I couldn't do without my good friends and former classmates, MaryEllen Cathcart, Anne Peng, Gina Chiodo, Yanti, Laura Spinu, Tim McKinnon, Lanny Hidayat, Tim O'Neill, Evan Bradley, Regine Lai, Solveig Bosse, Ozge Ozturk, Alison Tseng, and Veni. Thanks also to all my non-linguist friends and family who supported me in other needed ways.

I am indebted to my parents for their prodding but loving support along the way, always helping me to pursue the best of education and pushing me to choose my own path, and for taking care of my baby daughter during the last month of finishing my revisions. I am also grateful to my brother Jeff for his support and being one of the few *not* to ask me how my dissertation was going, having gone through the process himself.

Finally, I am grateful to my husband, Gavin, for being extremely patient, understanding and emotionally and financially supportive. He was a big part of my getting to this point. This dissertation is dedicated to our daughter, Marilla, for pushing me to finish so that I can spend time with her guilt-free.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

ABSTRACT

This dissertation investigates second language (L2) English learners' perception and production of certain prosodic patterns, with a focus on Chinese learners. A general goal of conducting the experiments laid out in this dissertation was to study in a systematic way how L2 learners perceive and produce certain types of pragmatically ambiguous utterances in English, compared with a set of native speakers. A second goal was to test the ability of L2 learners to improve their performance using a brief linguistically-informed targeted training, involving three components: auditory, visual, and explicit instruction.

The results unsurprisingly showed that L2 learners initially (before training) perform best on prosodic patterns they are likely familiar with. However, remarkably, even after only a ten-minute training, results demonstrated a strong positive influence of training for perception in all areas where there was room for improvement. Training had a moderate effect on production, in many cases aiding in expanding contrasts in duration, F0, and intensity towards native speaker patterns. Results also supported the proposal that the existence of certain lexical-level contrasts (e.g. pitch/tone contrast) in an L1 may make those acoustic properties more accessible to be manipulated in an L2.

The clear evidence from these studies that this type of brief targeted training leads to immediate improvement suggests that this type of training for prosody could be effectively extended to other prosodic patterns, and in other languages, as well.

With further training similar to this, and incorporated into an L2 curriculum, it is expected that much more can even be accomplished.

Chapter 1

INTRODUCTION

1.1 Objectives of this Dissertation

This dissertation investigates second language (L2) English learners' perception and production of certain prosodic patterns. There are many different aspects to prosody, which generally consists of the patterns of stress and intonation in a language, but the current study focuses on the phonetic patterns associated with a few common types of pragmatically ambiguous sentences, and what meanings they are perceived to convey. Part of what is missing in the literature is a careful joint analysis of L2 perception and production of sentential prosodic structures. Hence, a general goal of conducting the experiments laid out in this dissertation was to study in a systematic way how L2 learners perceive and produce certain types of pragmatically ambiguous utterances in English. A second goal was to test the ability of L2 learners to improve their performance using a linguistically-informed targeted training.

L2 learners of English are less likely to have received instruction on the prosodic patterns that are associated with pragmatic ambiguities; thus, a reliance on natural exposure to the language is necessary. Three types of patterns, all of which are commonly used in everyday American English speech, are investigated. The goal was to determine the extent to which high-level learners of English can perceive and produce the slight nuances in prosody that determine the difference in intended meaning for native speakers. Participants from two language backgrounds (Chinese and Arabic) are included, based on degree of distinctness from English in use of

suprasegmental features at the word level; however, this dissertation will primarily concentrate on the performance of Chinese speakers, as this language is prosodically most different from English.

It has previously been shown that there is a relationship between stress type in an L1 and learning stress in an L2 (Altmann, 2006). I thus further extend her findings to show that sensitivity to word-level suprasegmental features in an L1 makes these features more accessible to be used at a higher prosodic level in an L2. It should be noted that there is another possibility: a prediction might have been made that the presence of a lexically-linked contrast in an L1 is more likely to be transferred and difficult to suppress (essentially causing interference) in L2 prosody, as Jun & Oh (2000) suggest. This dissertation seeks to show that in spite of the inevitability of language interference, it is possible to largely overcome it through even just a small amount of linguistically-informed targeted training. Specifically, the brief targeted training introduced here is shown to have a tremendous effect on performance in perception and a moderate effect on production, an expected disparity given the additional mechanisms of articulatory competence involved in production.

1.2 Rationale for the Current Study

Prosody is often said to be acquired first in an L1, at least as far as sensitivity to the patterns of the native language is concerned; however, it has also been shown that children up to the age of 11 or 12 have trouble assigning the appropriate meaning to different prosodic patterns in their language (Atkinson-King 1973, Cruttenden 1985, Vogel & Raimy 2002, and others). While the study of prosody has attracted much more attention in recent years, it remains an area that is not well understood, likely due to the complexities of its interactions with morphology, syntax, semantics,

and pragmatics, in addition to the complexities involved in its acquisition in a second language.

English is used as the target language of investigation of the current study since it has such a rich intonational system, with many prosodic patterns indicating various meanings. By investigating certain common but rarely taught tunes, the current study addresses L2 learners' abilities to extract and apply prosodic knowledge from experiential learning, rather than only from the classroom setting. This study also examines the ability of L2 speakers to be trained on prosody, and how this may be applied to second language instruction; hence not only of interest to linguists, but to the broader audience of L2 learners and teachers throughout the world.

While the acquisition of English as a second language (ESL) has been widely studied, one area that is lacking is the acquisition of its prosody. Non-native speakers' misuse of prosodic patterns can cause great confusion for native speakers. In fact, much research has shown that prosodic errors can have a more detrimental impact than segmental errors on L1 listeners' understanding and judgments of accentedness in L2 speech (Nash 1972, Johansson 1978, Anderson-Hsieh et al. 1992, McNerney & Mendelsohn, 1993, Munro & Derwing 1995, Trofimovich & Baker 2006). This demonstrates how important prosody is in L2 learning.

In general, prosody (or intonation) is one of the least-taught areas in ESL courses, thus it is most likely that L2 learners are not formally taught the patterns tested here (with the possible exception of Contrastive Focus). Hence, L2 learners generally must rely on acquiring these patterns from natural speech. It has remained rather unclear to this point the degree to which L2 learners can understand the various pragmatics and produce the prosodic patterns associated with them. The current study

attempts to shed light on this subject, and focuses on advanced L2 learners of English, since these prosodic patterns require a certain level of experience with the language.

1.3 General Research Questions

There are several general research questions to be addressed in the course of this dissertation. First, against a baseline of English L1 speakers, how well can advanced L2 learners understand certain common patterns of English prosody? Similarly, how well can advanced L2 learners produce these same patterns? Of interest is also how the performance in perception and production compares with respect to these structures. Moreover, with the introduction of a novel linguistically-informed targeted training on prosodic patterns, can rapid improvement in perception and production occur, and how might this differ between perception and production results? Finally, regarding production, can the distinctive use of certain phonetic properties in an L1 aid in the use of these same properties in an L2, with regards to prosodic acquisition?

1.4 Organization of this Dissertation

The rest of the dissertation will be organized as follows: Chapter 2 will discuss background on L2 prosody and the languages under investigation; Chapter 3 discusses the innovative targeted training methodology used in the experiments; Chapter 4 discusses the perception experiment; Chapter 5 presents the production experiment; Chapter 6 presents a discussion of the results and research implications for L2 pedagogy; finally, Chapter 7 concludes the discussion.

Chapter 2

BACKGROUND: L2 PROSODY AND L1 PROSODIC SYSTEMS

This chapter presents background in two parts: literature background on L2 prosody (2.1), specifically regarding L2 perception of prosody, L2 production of prosody, and L2 training studies. The second part presents a brief background on the prosodic systems of the languages under investigation (2.2), starting with the target language, English, followed by the background languages of the L2 speakers: Mandarin Chinese and Arabic.

2.1 L2 Prosody: Perception, Production, and Training Studies

This section provides background on the types of studies that have been done relating to both L2 perception and production of prosody, and with regard to training studies in these areas. As will be reiterated below, few studies exist related to the perception and quantitative acoustic analyses of different pragmatic uses of prosody.

2.1.1 L2 Perception of Prosody

To date, more work has been done on L2 acquisition of lexical prosody (i.e., stress) than on sentential (i.e. utterance-level) prosody. The studies that exist on the perception of sentential prosody tend to test perceptibility of basic tunes relating to specific grammatical issues, like sentence focus type and location, rather than probe pragmatic uses (with the exception of Baker, 2011, as described below). The works that are described here present the main studies that have investigated L2 perception of sentential prosody.

In examining how features like word order, part of speech, syllable type, pitch accent type and boundary tones affect native Mandarin speakers' ability to perceive the location of English pitch accents, Rosenberg, Hirschberg, and Manis (2010) found that native Mandarin speakers were better at identifying pitch accents on two-syllable words than one-syllable words, better on adverbs and determiners than verbs and nouns, and better on words at the end of an utterance than at the beginning. They found that native Mandarin speakers had an easier time perceiving pitch accents that were realized with higher mean and maximum F0s and longer durations. They also found that pitch accents that exhibited a greater difference between the mean F0 on the accented word and the mean F0 for the entire sentence were more easily perceived.

Baker (2011) conducted a perception study on English focus-marking by Korean and Mandarin speakers. She tested non-native speakers' ability to detect location of narrow focus-marking in subject position and two different broad focus markings (verb broad focus and sentence broad focus) and found that Mandarin speakers were less successful at detecting location of pitch accents than Korean speakers and native English speakers in all three focus contexts. She also performed a comprehension study which sought to learn whether the same speakers understood the meaning behind the various focus markings. She discovered that both Mandarin and Korean speakers were less successful than native English speakers at determining whether a sentence had context-appropriate prosody and that their success was a function of their English proficiency level. Non-native speakers were more accurate when presented with matched prosody rather than mismatched prosody; in matched cases, they were better at identification in the subject narrow focus condition and in mismatched cases, they were better at rejections in the verb broad focus condition.

While not an L2 study, Shen (1990) performed an experiment which was aimed at determining whether Chinese speakers (having no knowledge of French) could accurately perceive falling versus rising intonation in French, heard through laryngeal output, and whether they could place them in the correct categories. A comparison with a control group of French speakers showed no significant difference, suggesting that Chinese speakers can be attuned to differentiating intonation patterns in another language, despite the fact that Chinese utilizes different prosodic strategies to distinguish a question versus statement.

Grabe et al. (2003) investigated cross-language effects on the perception of intonation similarities and differences. It is interesting to note, however, that even they state their disinterest in investigating second language acquisition issues. Instead, their goal lies in distinguishing universal from language-specific effects. In their study, Grabe et al. tested listeners of English, Iberian Spanish, and Mandarin on their perception of similarity of seven falling contours and four rising contours that were resynthesized onto an English utterance. They found that all groups separated the falling contours from the rising contours, but that the perceptual space differed among the language groups for the falling contours.

Nguyen et al. (2008) tested Australian English speakers and Vietnamese learners of English on their perception of prominence in compound triplets as compared to noun phrases with narrow focus and broad focus. They found that while the narrow focus pattern was most accurately identified by both native and non-native speakers, native speakers perceived compounds better than broad focus and vice versa for non-native speakers. This result was directly attributed to different strategies and acoustic cues relied upon the speaker groups: specifically, it was proposed (based on

results from a corresponding production experiment) that Vietnamese speakers relied mostly on pitch rather than duration, which resulted in difficulty in perceiving compound patterns, demonstrated to be characterized mainly by timing/duration cues for native speakers.

2.1.2 L2 Production of Prosody

Production studies on L2 prosody tend to be more common than L2 perception studies. Most have been concerned with the comparison of one or two languages to an L1 language and investigate L2 learners' errors in acquiring English. They tend to focus on the production of prosodic patterns from the viewpoint of whether the L2 speakers can accurately produce these patterns, based on accuracy of a variety of phonetic measures, including pitch peak alignment, pitch range, and syllable duration. A few studies do exist on acoustic analyses related to pragmatics and prosody, but they are more rare. The studies outlined below consist of those studies that do involve various kinds of pragmatic contrasts; these tend to be centered around the study of simple declaratives, questions, narrow focus, broad focus, and contrastive focus.

Baker (2011) conducted a production study to determine the extent to which Mandarin and Korean speakers produced appropriate pitch accents for sentences with subject narrow focus and two types of broad focus. Acoustic analyses and a further perception study (involving this production data) by native speakers showed that both native and non-native speakers had context-appropriate prosody much of the time. Non-native speakers differed acoustically from native speakers in that the former had longer utterance durations, higher max F0 and larger F0 ranges on verbs, and higher intensity and lower max F0 on objects. They also produced certain stronger pitch accent cues, as compared with native speakers. It is perhaps less surprising that non-

native speakers performed as well as they did on these experiments, as the types of prosodic structures used are easily found in other languages, albeit not exactly in the same way.

McGory (1997) investigated the production of English word pairs differing in the location of stress in statements and questions and in several focus conditions by Seoul Korean and Mandarin Chinese speakers. All had difficulty producing native English prominence patterns: where native English speakers only produced pitch accents in prominent target words, non-native speakers produced stressed syllables with higher F0 values in both prominent and less prominent words. Moreover, the non-native speakers did not distinguish between statements and questions in their F0 patterns. Results seem to indicate that the differences between non-native and native speakers of English can be attributed to influences of the L1, and an effect of L1 background was found in the different error patterns for the Mandarin and Korean speakers.

Nava & Zubizarreta (2008) investigated placement of Nuclear Stress in Spanish L1 learners of English and English L1 speakers. Various information structure categories were used, including wide-focus, VP-focus, subject focus, anaphoric de-accenting, and compound constructions. They found that there is a negative prosodic transfer in the case of Nuclear Stress for a variety of information structure categories in the speech of L2 learners. They also performed a study to investigate whether the L2 learners were producing speech with rhythmic properties of English or Spanish. They found that as L2 learners progress towards native-like prosodic proficiency at the phrasal level, they begin to adjust their rhythmic/metrical timing.

Lepetit (1989)'s study is based on Martin's (1981, 1982) theory of intonation. The goal in this experiment was to explore the production of four melodic contrasts in French. They used two groups of participants: Canadian Anglophones and Japanese, who had been studying French. Based on Martin's theory, contour errors were marked for the elicited French sentence stimuli for the Canadians and the Japanese. He found that at the phonological level, the data from both the Canadian and Japanese participants proved the existence of a cross-linguistic influence. In addition, at the phonetic level, the Japanese data showed some characteristics of the native intonational background, as the pitch range of the Japanese speakers was rather narrow.

Barlow (1998) investigated the way in which intonational form in the speech of non-native speakers develops as the overall perceived quality of pronunciation becomes more native-like, in addition to the extent to which non-native speakers can attain native-like norms in their intonational production. Twenty-five non-native speakers of English (L1 peninsular Spanish) were divided into four subgroups based on ability (as judged by native speakers) and were compared with eight native speakers. Responses to a series of questions based on pictures were elicited. Normal prominence, different types of contrastive prominence, and listing intonation were tested. Responses were measured with a combination of instrumental analysis and auditory judgment. He found that non-native speakers (NNS) make less use of prosodic features as markers of normal prominence than native speakers (NS), but average use increases slightly with higher proficiency. Pitch-marking is underused by NNSs but increases towards native-like levels in normal prominences, but shows no development in contrastive marking. Loudness is constantly overused on normal

prominences with an overall increased use on contrastive prominences. Durational phenomena appear to be consistently underused, but were not present at significant levels in either NS or NNS speech.

Rasier & Hiligsmann (2007) performed an experiment in which they investigated a hypothesis of whether "unmarked" prosodic patterns (fixed pitch accent location in French) are easier to learn than "marked" prosodic patterns (variable pitch accent location based on pragmatics in Dutch). They tested French L1 language learners of Dutch, and Dutch L1 language learners of French, focusing on pitch accent patterns. They elicited phrases of the form "determiner adjective noun" for Dutch and "determiner noun adjective" for French. They tested these phrases in new, given, and contrastive contexts with L1 and L2 learners of French and Dutch. Results showed that French L1 language learners of Dutch indeed performed worse on producing the correct accentual patterns than Dutch L1 speakers on French, supporting their hypothesis.

2.1.3 L2 Training Studies

There are several L2 training studies aimed at teaching segmental contrasts (Jamieson and Morosan, 1986; Logan et al., 1991; Lively et al., 1994; Bradlow et al., 1997) that show successful short and long-lasting effects. In addition, Wang et al. (1999) were rather successful in training American listeners to perceive Mandarin tones, both in the short-term and long-term, without repeated exposure between those periods. The level of success in their results are comparable to those obtained in segmental training studies.

There are also several L2 studies that have encouraged the use of visual displays of prosody in a computer-based learning approach: Abberton & Fourcin

(1975), Anderson-Hsieh (1992, 1994), de Bot (1981, 1983), de Bot & Mailfert (1982), Leather (1990), Molholt (1988), Pennington & Esling (1996), Spaai & Hermes (1993), and others. Hardison (2004) performed a study aimed at training native English speakers on French prosody of several types, including simple declaratives, questions, and various forms of contrastive focus. The training was given daily for 3 weeks; it was computer-assisted and involved a different set of 30 sentences every day. Feedback consisted of visually providing students with their own contours compared against those of native speakers', along with the audio form of those sentences spoken by native speakers. It does not seem that any explanation was given of the contours. Success was evaluated auditorily by native speaker ratings (according to a 1-7 native-like scale) through a pre-test and post-test in two versions: unfiltered and filtered-out segment information. A significant improvement was found in both cases, although it may be viewed as relatively minor.

Among the literature, Ramirez-Verdugo (2006) uses a multi-sensory approach that appears most similar to the one proposed here; her approach is more intensive, but includes the same stimuli in the training as in the pre-test and post-test. She trained Spanish students in British English intonation over a period of 10 weeks (10 sessions of 50 minutes each) and tested them on their production of controlled conversations before and after the training period. Questions, answers, and statements of various levels of uncertainty were the focus of the study; sparse details are provided regarding the stimuli, though. Native English speakers were used as judges of how native-like the intonation was; in addition, tone, tonality, and tonicity were evaluated through rater annotation. She found significant improvement in all of these areas. While this study seems to be headed in the right direction regarding training in L2 intonation, the

fact that new controlled conversations were not used in the post-test would seem to limit what can be concluded as far as generalization of learned patterns is concerned. In addition, no acoustic measurements were taken, all evaluations being strictly qualitative, which presents us with limited knowledge on the speakers' actual patterns.

2.1.4 Lacking Areas of Research

In sum, it is clear that there remains a paucity of research on L2 perception of prosody, particularly in the area of prosodic understanding (not just identification of pitch accent locations) by L2 learners. In addition, more perception and production studies directed at pragmatic contrasts, especially regarding more complex prosodic contours and corresponding nuanced meanings, are needed to understand the abilities of L2 learners to acquire the more complex patterns that they may not have been taught or exposed to. Trainings tested with L2 perception and production of the same prosodic data do not seem to exist to this point. In fact, very few training studies seem to focus on improving perception. It may also be seen that the training studies testing production do not evaluate data acoustically in a quantitative manner. The current study aims to address those holes in the literature.

2.2 Prosodic Systems of Languages under Investigation

Given the clear role played by the L1 in L2 phonology/prosody (though it may not be the only contributor), some basic properties of the L1 prosody of the speakers investigated in this research are briefly presented below. This section provides an overview of the prosodic systems of the language backgrounds utilized in this dissertation. The results presented later in chapters 4 and 5 mainly emphasize results from Chinese speakers, whose language consists of a very distinct prosodic system

from English. Speakers of Arabic were investigated to a lesser extent to gain a perspective on the effect of prosody from a language with both quantity-sensitive stress and contrastive vowel length.

In the current study, I suggest that experience with L1s that contrast pitch and duration at the segmental or word level can affect not only lexical prosody (e.g. stress) in an L2, as has been shown in Altmann (2006) and Kijak (2009), but also a higher prosodic level: i.e. sentential prosody (or intonation of utterances). As such, as will be seen below, the background information will focus on the properties that will be of most relevance to the research, in particular those that may interact with English sentential prosody: duration, F0, and intensity.

2.2.1 English

English is a stress language and its placement of stress at the lexical level is largely unpredictable. Sluijter & Van Heuven (1996b) provide a variety of conclusions regarding the acoustics of English stress. First, duration, glottal parameters and vowel quality are the most important phonetic cues for English stress. Second, F0 and intensity movements have little involvement in English stress. Stressed syllables tend to have a longer duration and vowels in stressed syllables have a fuller vowel quality than in non-stressed syllables. F0 is used much more at the phrasal level with focused elements. In addition to the tone distinctions, many studies have also discussed duration differences for focused vs. non-focused constituents in English. Cooper et al. (1985), Eady & Cooper (1986), and Eady et al. (1986)'s experiments revealed that various types of focus (including Contrastive Focus, narrow focus, and broad focus) are generally accompanied by an increase in duration on the focused word. Xu & Xu (2005) also indicate that duration increases with narrow

focus. Finally, increases in intensity, through various measures, such as overall intensity (e.g., Fry, 1955) and spectral balance (e.g., Sluijter & van Heuven 1996b) have also been shown to be reliable acoustic correlates of focus.

While it is clear that there are many potential acoustic correlates relating to stress and focus in English, those that are most commonly agreed upon are the fairly consistent use of duration, F0 and intensity to make prosodic distinctions.

2.2.2 Mandarin Chinese

Mandarin Chinese (henceforth Chinese) is a tone language. Tone languages like Chinese are the most distinct from intonation languages, like English, in the sense that pitch is used mainly at the lexical level in tone languages in order to differentiate word meaning. For example, the segmental string /ma/ can be distinguished in meaning in four different ways based on the tone it carries in Chinese: /mā/, containing a high level tone, means 'mother'; /má/, containing a mid-rising tone means 'to bother', /mǎ/, containing a low dipping tone, means 'horse'; /mà/, containing a high falling tone, means 'to scold'. Chinese does have pitch movements at the sentential level, as well; however, its variations seem to be more limited so as not to interfere with the lexical tones. For example, Xu (2004b) indicates that focus only needs to be manifested prosodically when it is not otherwise marked syntactically. Variations in global and local pitch contours may signal certain pragmatic differences (Shen 1990), but they are not very well understood (Peng et al. 2005).

2.2.3 Arabic

Since data from Arabic is also examined (albeit more limited), a brief description of its prosodic system is provided here as well. Arabic has a quantity-sensitive stress system, which means that the location of stress in a word is predictable based on its distribution of heavy and light syllables. Stress in Arabic generally falls on the penultimate or antepenultimate syllable (final stress is restricted to the presence of a super-heavy syllable), and is attracted by the weight of the syllable. For example, [mak'tabha][1]'her desk, office' receives stress on the penultimate syllable because it is the rightmost heavy syllable, whereas in the word [ma'ʕallamak] 'he didn't teach you', stress falls on the antepenultimate syllable because it is the rightmost heavy syllable (besides the final syllable, which is excluded). Arabic also has contrastive vowel (and consonant) length: [kataba] 'he wrote' vs. [ka:taba] 'he corresponded with'. In Jordanian Arabic Stress (a similar variety to the Saudi Arabian version of Arabic spoken by participants in this study), it has been proposed that stress is marked by pitch patterns, measured in terms of fundamental frequency or F0[2] (Al-Ani 1992, Zawaydeh & de Jong 1999, Vogel et al. forthcoming). Based on a large study that separates stress from focus cues in Arabic, Vogel et al. (forthcoming) finds that in addition to stress, corrective focus is also marked by F0, and less so by other cues such as duration and intensity.

---

[1] Stress examples from de Jong & Zawaydeh (1999).

[2] Other cues have been claimed to be useful, as well, such as duration, intensity, and vowel centralization, but F0 seems to be the most agreed upon.

Chapter 3

TARGETED TRAINING METHODOLOGY

A major component of this study centers on the ability of non-native speakers to learn the prosodic patterns in this study and their associated meanings. This chapter will focus on the innovative targeted training that was developed to improve perception and production in L2 learners for these structures. Background on the prosodic structures investigated here (3.1) will first be presented in this chapter as a basis for discussing the targeted trainings that were utilized for each structure. Next, a presentation of the three-pronged training used in this study is given (3.2). The sections that follow will outline how the training was used for each prosodic category: Contrastive Focus category (3.2.1), Compliment category (3.2.2) and Verb Focus Category (3.2.3).

3.1 English Prosodic Structures Used in Study

The three structures used in this dissertation consist of the Contrastive Focus category, the Compliment category, and the Verb Focus category. These will be explained in turn below.

The Contrastive Focus structures used in the present investigation includes sentences such as the following:

1. Noun Focus: “The red *roses* are expensive” (compare with: “The red tulips are cheap”)
2. Adjective Focus: “The *red* roses are expensive” (compare with: “The pink roses are cheap”)

The sentence types included in this category are intended to either have focus on the subject noun or on the adjective modifying the subject noun, as seen in the examples in (1) and (2), respectively. Italicization is present to indicate where focus is located. Contrastive focus is also sometimes termed as 'corrective' focus (Gussenhoven, 2007). This means that the focused element is a direct rejection of an alternative, either spoken by the speaker himself or by the listener. In example 1, the speaker is emphasizing the subject noun 'roses' in order to reject a context-based alternative, such as tulips, balloons, etc. That is, there may be both red roses and red tulips (in addition to flowers of other colors) present, but the speaker wants to make clear that they are only referring to the red flowers that are roses, rather than any other type of flower. Emphasis in a contrastive focus context is typically described by having a high (H*) pitch accent (Pierrehumbert 1980). This means that there is a strong pitch (or F0) peak in the stressed (or accented) syllable of the word that is emphasized (in this case on the first syllable of 'roses'). Similarly, in example 2, the adjective modifier 'red' is emphasized, suggesting that while there may be other colored roses, they are only referring to the ones that are red in color; in this case, 'red' is characterized by a high pitch accent, thereby having a strong pitch peak on 'red'. In this study, these two types of contrastive focus are compared directly with each other in terms of perception and production.

The other two types of structures (the Compliment and Verb Focus categories) involve a comparison of sentences with a nuanced meaning and a more basic meaning. Examples of that comparison in the Compliment category are seen in examples (3) and (4), and examples from the Verb Focus category are seen in (5) and (6). Corresponding potential continuations, to help indicate meaning of the target sentence,

are presented within the parentheses that follow. Italicization is used in (3) and (5) to indicate emphasis; it is left out of (4) and (6) because no special emphasis is intended there.

I use the term "Compliment category" here because both types of sentences (3 and 4) represent a sort of compliment, even though the type in (3) is a more reserved type; the type in (4) indicates a true compliment, without any other implication. I use the term "Verb Focus category" for sentences such as (5) and (6) because the main prosodic difference between these two types of sentences lies in the verb: focus is present on the verb in sentences with nuanced meaning like (5), but absent in sentences with a more basic, neutral meaning like (6).

3. Compliment, nuanced meaning: "Emily has beautiful *hair*" (…but she isn't pretty, otherwise)

4. Compliment, basic meaning: "Emily has beautiful hair" (…and she has a nice face, too)

5. Verb Focus, nuanced meaning: "I *tried* to pay attention in class" (…but I couldn't help daydreaming)

6. Verb Focus, basic meaning: "I tried to pay attention in class" (…and as a result, I did well on the quiz!)

The sentences that have the more nuanced-type meanings here (i.e. 3 and 5) are purported to exhibit falling-rising tone on the word in focus, while the sentences with a basic meaning (i.e. 4 and 6) typically exhibit a falling pattern throughout the sentence. The falling-rising tone's contribution to utterance interpretation has been described in many different ways: 'a statement or answer with reservation ("there's a 'but' about it")' (Halliday, 1967); "focus within a set" (Ladd, 1980); reservation or implied contrast (Bing, 1979); 'selection of a variable from the background' (Gussenhoven, 1983); contrast (Liberman & Sag, 1974); uncertainty (Ward &

Hirschberg, 1985). Essentially, the presence of such a falling-rising tone indicates a degree of uncertainty or reservation regarding the statement.

Therefore, the sentence from the Compliment category in (3), "Emily has beautiful *hair*", with a falling-rising tone on 'hair', could imply that she isn't pretty otherwise, but the speaker wanted to use positive words to be polite. In this type of sentence, the focus always falls on the final content word. It should be noted that intonation/prosody is rarely straightforward, and therefore variations can often occur. As will be seen in the section below that outlines the training for this category, and also later in the L1 production results section, a rising-falling-rising tone is described for this condition. While I have not seen this documented elsewhere, I do believe this is the intonation pattern more commonly used nowadays for reserved compliments. Regarding the basic meaning sentence type in (4), a falling tone on the final content word 'hair' would indicate a true compliment, with no other meaning intended.

In terms of the Verb Focus category, if the main verb in a declarative statement is produced with a falling-rising contour, the speaker is expressing a sense of uncertainty, such that the intent exists, but for some reason, the situation may not be realized as planned. The sentence in (5), "I *tried* to pay attention in class", with a falling-rising contour on 'tried' would imply the failure of that attempt (they couldn't help daydreaming instead), whereas the version in (6), where there is no specialized focus, the verb has a falling intonation, resulting in the meaning that there was some level of completeness to the action (and as a result, they did well on the quiz!).

While I only provided descriptions of pitch patterns for the three categories, it should be noted that other acoustic properties associated with focus also generally

come into play; longer duration and higher intensity is typically seen on focused words than non-focused words.

All three of the patterns described here are commonly used in American English. Non-native speakers may (or may not) have some familiarity with Contrastive Focus, since it is sometimes taught, but they are much less likely to have had exposure to the Compliment and Verb Focus categories, as they are not taught. In the next section, I describe the targeted training aimed at improving the L2 perception and production of all three of these patterns.

3.2 Three-pronged Training

In both the perception and production experiments for L2 speakers, a training is presented between experimental sessions to test whether learning occurs. It was presented through PowerPoint, with the text and images on each slide being narrated (previously recorded) by the author. What is particularly innovative about this training is that it incorporates three components, *auditory*, *visual*, and *verbal description*, into a targeted condensed format (a total of 10 minutes long for all three categories). For the auditory component, participants heard example sentences from each prosodic category, including both versions of Contrastive Focus meaning and both the basic and nuanced meaning for the compliment and verb focus categories. The visual component consisted of pitch contours of the example sentences, where participants were led to concentrate on the relevant parts of the contours. Finally, a verbal description was also provided to teach the various tunes and make the corresponding meanings clear. Situational contexts were also given to help the L2 learners understand where/when this type of prosody would be used. These

components will be demonstrated in more detail below within the discussion of the training on each prosodic pattern.

This three-pronged training allows us to furthermore investigate questions from perception and production. First, can this targeted explicit training, even if brief, be effective in improving L2 perception and production of pragmatic uses of prosody? Second, will the training be more effective on perception than production? This might be expected since the training is brief and production involves additional mechanisms of articulatory competence. Finally, will the training be more successful in certain acoustic properties and/or prosodic categories than others, and does this depend on the L1? Answers to these questions will be addressed in the course of the current study.

3.2.1 Contrastive Focus Category

It should be noted that for this particular pattern, only the auditory and verbal description components were used, as it was deemed that a visual representation of pitch would be less informative here due to the lack of a relevant pitch contour (pitch peak is more relevant here, as described in the previous section). Participants were shown a pair of elephant pictures: one was a pink elephant with glasses; the other was a grey elephant eating leaves. They were told that when given two similar pictures, one must put emphasis on the aspect of the description that is different between the two: i.e. "the *pink* elephant is wearing glasses" or "the *grey* elephant is eating leaves". They were then shown an additional pair of pictures, representing a black dog and a black cat. The purpose of this additional set of pictures was to draw attention to the possibility of placing emphasis on the noun instead of the adjective. They were given examples such as "The black *cat* is sleeping" vs. the "The black *dog* is awake". They were told that placing emphasis on a word in English normally means that the pitch is

made higher and the word is made longer and louder. Participants were given several chances to hear the sentences and they were also given the opportunity to practice repeating each sentence a few times.

3.2.2 Compliment Category

For this category, participants were first presented with an example of neutral intonation, with the basic meaning, indicating a genuine compliment, such as "I liked the special effects", which could be an appropriate answer to a question like "What did you like about the movie?" They were shown a line graph[3] (Figure 1), which they were told visually represents how the pitch of the voice goes up and down throughout a sentence. They were instructed to focus their attention on the circled part and to compare it with what they heard.

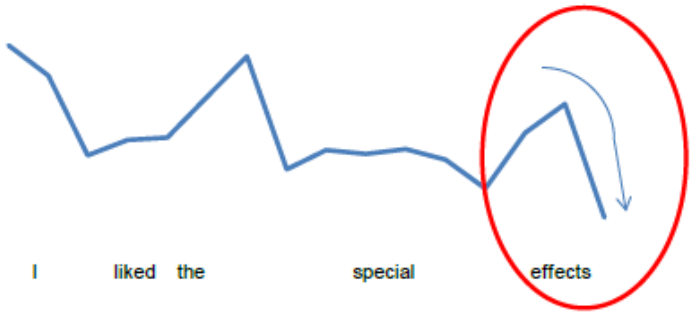


Figure 1: Pitch contour for the sentence "I liked the special effects", where the falling pitch on "effects" indicates a genuine statement.

[3] The pitch contours were created by taking F0 measurements in the speech analysis program Praat at equal points throughout the utterances with the interactive Praat script ProsodyPro (Xu, 2013). The points were then connected to make line graphs in Excel.

To teach the more nuanced meaning, a simple dialogue was presented between two friends where Friend A asks, “So, did you like the movie?” and Friend B responds “I liked the special effects”. In this case, it was pointed out that the final word of the sentence, “effects”, receives a special kind of intonation pattern, a rising-falling pitch, with another slight rise at the end. It was mentioned that this kind of contour should be read as an indirect insult, rather than a true compliment. They were shown a line graph (Figure 2) of the intonation and were instructed only to pay attention to what is circled:

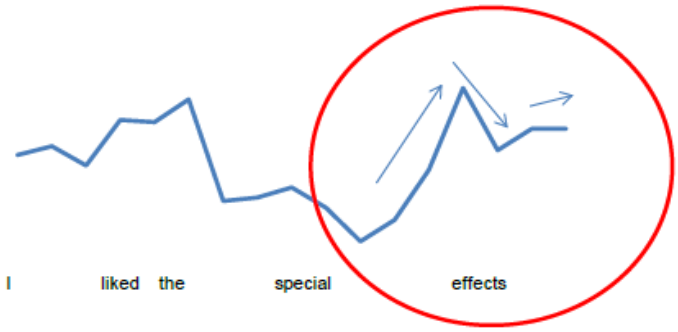


Figure 2: Pitch contour for the sentence “I liked the special effects”, where the rising-falling pitch on “effects” indicates an indirect insult meaning.

They were able to listen to both sentences several times and compare the auditory pitch movement with the visual contour. They were also given the opportunity to practice saying the sentences.

3.2.3 Verb Focus Category

A similar approach was made for the verb focus category: participants were first presented with a neutral sentence “The milk smells okay”, with falling pitch on the verb. It was explained that this kind of intonation pattern suggests a neutral/positive

meaning, in this case that the milk is, in fact, drinkable. They were again given a visual contour (Figure 3) to compare with the auditory stimulus.

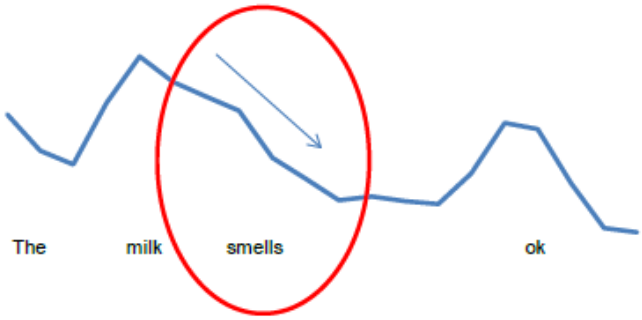


Figure 3: Pitch contour for the sentence "The milk smells ok", where the falling pitch on the verb indicates a neutral/positive meaning.

They were then introduced to another way to say the sentence, where the verb has a special emphasis and the meaning is something like "but we're not sure whether it tastes ok". In this case, the verb "smells" has a falling-rising pitch, as seen in the pitch contour in Figure 4.

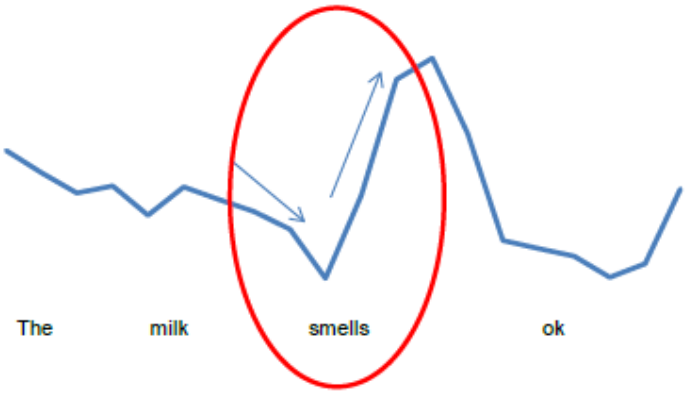


Figure 4: Pitch contour for the sentence “The milk smells ok”, where the falling-rising pitch on the verb indicates a negative/uncertain meaning.

It was also pointed out that a crucial characteristic of this contour/meaning is the final rise in pitch at the end, as can be seen in Figure 5 below. This is an indication of uncertainty.

Figure 5: Pitch contour for the sentence “The milk smells ok”, where the rising pitch at the end indicates uncertainty.

This was followed by a summary of what was taught during the training. A post-test followed the training to measure improvement.

Chapter 4

PERCEPTION STUDY

The perception experiment was designed to test non-native speakers' ability to understand the specific pragmatic meanings associated with different prosodic patterns in English, and to compare this with a set of baseline L1 English speakers. The results concentrate on data from Chinese speakers, but also give a preliminary look at data from Arabic speakers, in order to shed light on whether other language backgrounds would yield similar results, as well as test whether the training could be useful for speakers of other languages besides Chinese.

This chapter will first present the research questions considered (4.1), followed by the methodology used in the experiment (4.2), the accuracy results (4.3), the Signal Detection Theory analysis (4.4), and a discussion (4.5).

4.1 Research Questions

The general question of interest here is how L2 speakers perform in the perception of English prosody. We can explore answers to this question by investigating L2 listeners' understanding of a few varied patterns of prosody. Are there specific types of patterns they do better at? We might expect that they would perform better on previously familiar patterns (that may have been taught) than more unfamiliar patterns that may require experience with more naturalistic data. In the cases where perceptual accuracy is lower, can targeted training in specific types of prosodic patterns noticeably improve it? Moreover, is targeted training able to bring

L2 performance up to the level of L1 speakers? All of these questions will be considered in the course of this experiment.

4.2 Methodology

The basic format of the perception experiment is as follows in Figure 6 below:

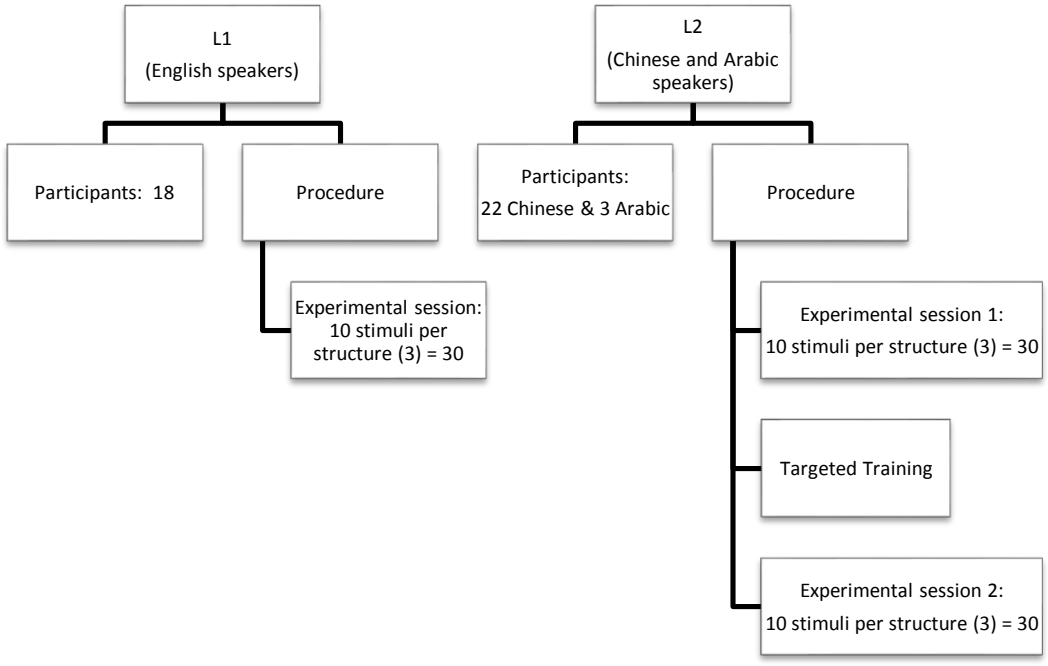


Figure 6: Schematic of Perception Experiment

Figure 6 outlines for L1 and L2 subjects the basic elements of the experiment, including numbers of participants and the procedure followed. This will be explained in more detail in the sections that follow.

4.2.1 Participants

The L1 results presented are from 18 participants who completed the finalized version of the experiment. Participants were between the ages of 18 and 29 (female = 14, mean age = 22). The disproportionate number of females was a result of a similarly off-balance ratio of females to males in undergraduate linguistics courses from which the participants were recruited. However, it was not expected that gender would have any effect on the results. Since the participants were given extra credit for their courses, all students who wished to participate were accepted into the study; however, data from students who had another native language besides English or had a parent with a native language besides English, were excluded. This was done to ensure that all participants' language backgrounds were as similar as possible. In addition, data from students who had previously been diagnosed with any kind of speech, hearing or reading disorder were discarded.

There were 22 Chinese-speaking participants in the experiment. Participants were between the ages of 18 and 39 (female=11; mean age = 26). There were an additional 3 Arabic speakers between the ages of 20 and 29 (all male; mean age = 23). The Chinese speakers' main language was Mandarin and they all came from China; the Arabic speakers came from Saudi Arabia. All participants had a high-level of (but not native-like) English proficiency, and were in levels 5 or 6 (the highest levels) at the English Language Institute, part of the University of Delaware. The duration of presence in the United States varied among the participants, ranging from 2 months to 2 years. Participants were all volunteers and no monetary compensation was awarded; in some cases, extra credit was awarded by the teachers whose class the subjects came from.

4.2.2 Stimuli

The three structures presented in section 3.1 were tested in this perception experiment. Twelve target stimuli were created for each structure, for a total of thirty-six target stimuli. In addition, there were thirty-six filler stimuli dispersed throughout the experiment. Six of the targets and six of the fillers were used solely in the beginning practice session. The stimuli were presented in a fixed pseudo-randomized order. No more than two of the same type of stimuli appeared directly after one another. All stimuli consisted of a target sentence heard auditorily followed by two written continuations that appeared on the screen after the auditory stimulus. Auditory stimuli were short sentences, with at most a simple embedded phrase. The written continuation pairs were designed to be very similar in length to avoid any bias for the time it takes to read them. The author, a native English speaker, recorded all of the stimuli for the experiment. All stimuli can be seen in Appendix A.

There were two distinct recordings of each target stimulus, consisting of different prosody, the details of which are described in section 3.1. Participants received the same two written continuation choices for each auditory version of the stimulus. For the Contrastive Focus category, the auditory stimuli consisted of simple sentences where the subject noun was modified by an adjective, such as "The red shirt is still wet". Either the adjective or the noun received the emphasis, consisting of higher pitch, increased length and loudness. The sentence always began with a determiner (e.g. a/an, the) or a possessive name to ensure that the adjective was not at an utterance boundary. The format for the written continuations in the Contrastive Focus category was a sentence with the subject NP containing either a contrasting adjective or contrasting noun from what was heard in the auditory stimulus' subject NP. For example, the auditory stimulus "The red shirt is still wet", where either "red"

or “shirt” was emphasized, was followed by the written continuations “The green shirt is dry” and “The red pants are dry”. If the adjective was emphasized, the congruent (or expected) choice of continuation would be “The green shirt is dry”, whereas if the noun was emphasized, the congruent choice of continuation would be “the red pants are dry”. For the Compliment and Verb Focus categories, one version of each target auditory stimulus had neutral (falling) prosody, indicating a basic meaning, and the other version had a special kind of prosody, indicating a nuanced meaning. For the Compliment category, an example of an auditory stimulus would be “Emily has beautiful hair”, where the final content word “hair” receives either neutral, falling prosody (basic meaning) or rising-falling prosody (nuanced meaning). A choice of written continuations for this stimulus would be “She isn’t very pretty otherwise, though” or “She has a pretty face, too”. The former continuation (since it indicates a more nuanced meaning) is a congruent response to the auditory stimulus with rising-falling prosody and the latter (indicating a more basic meaning) is a congruent response to the version with falling prosody. An example of an auditory stimulus in the Verb Focus category would be “Bob would like a cheeseburger”, where the verb “like” receives either a falling prosody (basic meaning) or falling-rising prosody (nuanced meaning). The choice of written continuations for this auditory stimulus would be “but he can’t have one” or “and he would like fries, too”, the former continuation a congruent response to the auditory stimulus with falling-rising prosody, and the latter a congruent response to the version with falling prosody. Hence, for both the Compliment and Verb Focus categories, the written continuations were designed to indicate a basic meaning for one (auditory) version of the stimulus and a nuanced meaning for the other version of the stimulus.

In the experiment, participants only heard one version of each stimulus: hence, for example, they heard "The red shirt is still wet", where "red" was emphasized, or they heard the version where "shirt" was emphasized, but did not hear both versions. Thus, half of the participants in this study received one version of the stimuli and the other half received the other version of the stimuli. Which version of stimuli they received depended on their participant number. It was not expected that any differences would result between these sets; it was simply done to ensure that all versions of stimuli were tested, and to make sure that the results found were not due to the order in which each set was presented, but instead were more representative of the general patterns being tested.

The fillers were similar to the target items in that they consisted of short auditory sentences followed by a choice of written continuations; however, the choice was only based on semantic plausibility – prosody did not play a role, as it was always neutral. For example, an auditory filler would be something like "Linda went to the beach" (with neutral prosody) and the written continuation choices would consist of "She went swimming" or "She studied hard", with the more plausible answer always winning (in this case, the former choice).

The set of stimuli used for the L1 participants was also used for the L2 speakers, with an additional set of stimuli for L2 speakers, in order to enable a comparison of their performance before and after the targeted training. L2 participants received 2 sets of stimuli, for a total of 60 target items, split into before and after the training: within each set, there were two versions of auditory stimuli, as described previously for the L1 participants.

4.2.3 Procedure

This experiment has a within-subject design, involving a forced-choice task. The basic task involved first hearing an auditory stimulus, then seeing two written sentence continuations on the computer screen, and choosing which continuation best followed the auditory stimulus. Each continuation had a congruent meaning associated with one prosodic pattern, as previously described in section 4.2.2. The general procedure followed for L1 and L2 participants was very similar. However, in addition, L2 speakers participated in the training discussed in the previous chapter, followed by an added experimental session, as depicted in Figure 6. The experiment with L1 speakers was run with E-prime, a psychology experimental design software, in a sound-attenuated booth in the Phonology and Phonetics Laboratory at the University of Delaware. This experiment was adapted to a javascript format for use with L2 speakers at the English Language Institute computer laboratory. All participants used a Logitech ClearChat USB headset to listen during the experiment.

Before participating in the experiment, participants signed a consent form (Appendix B) and filled out a background questionnaire. They were told that they would be participating in an experiment where the researcher's goal was to learn more about how English speech is perceived. They were also told that they would be part of a group of speakers from their native language, and the data used would not be linked to their names. L1 speakers were told that they were participating to create a baseline comparison for results from non-native speakers.

All participants were told that while there were no right answers, they would be receiving feedback during the experiment which would indicate to them how they were doing. They were informed that this was merely included as a strategy to help them focus more on the task.

A practice session, lasting roughly 2-3 minutes, began the experiment. This was followed by the experimental session (~10-15 minutes). After hearing each sentence stimulus, two written sentences, intended as continuations of the auditorily presented sentence, popped up on the screen. One continuation appeared on the top and the other continuation appeared on the bottom of the screen. Participants had to choose the written continuation that best followed the sentence they heard. They were given unlimited time to respond to each stimulus, but they could not listen to it more than one time. Once they had responded, three seconds passed before the next auditory stimulus was introduced. Participants heard a total of 60 sentences in each experimental session: 30 target sentences (with 10 from each category) and 30 filler sentences. Subjects received feedback intermittently (every 6 sentences), where they were shown their cumulative percentage "correct". Intermittent feedback has been informally observed in previous studies (Vogel et al. 2010) to be effective in improving subjects' focus on the task. It also avoids the training effect of giving feedback following every sentence.

L2 speakers received targeted training (as described in detail in Chapter 3) on the prosodic structures after the experimental session, lasting approximately 10 minutes, along with an additional experimental session to test learning; L1 speakers did not receive the training or additional experimental session, as it was presumed they already had a native understanding of the structures, and their data was being used as a baseline. Thus, in total, the entire experiment lasted roughly 50-60 minutes for L2 speakers, while it lasted 15-20 minutes for L1 speakers.

4.2.4 Analysis

In all categories, mean accuracy was determined based on the percentage of responses matching the congruent written continuation to the auditory stimulus. Accuracy of raw categorical data was compared among speaker groups using binary logistic regression analyses. Specifically, the logistic regression analyses compared the L2 participants' data from Experimental Session 1 and Experimental Session 2 (before/after training comparison) in order to determine whether training affected accuracy. Additional binary logistic regression analyses were performed on the L2 Experimental Session 2 and L1 data to determine whether or not the L2 data after training approximated that of the L1 participants. One-sample t-tests against chance were performed to determine whether each condition differed from chance.

A separate analysis involving signal detection theory was used to investigate the level of sensitivity of the subjects to the stimuli: specifically, whether bias existed in the responses towards choosing a particular type of response.

4.3 Accuracy Results

This section will present the accuracy results, broken down by prosodic category and further into subcategories, of the L1 speakers as a baseline, followed by the results by L2 language background.

4.3.1 Contrastive Focus

The accuracy results of the Contrastive Focus category will be presented for L1 and L2 speakers, with a concentration on the comparison of the L2 Chinese speakers' results to the L1 speakers, as a baseline.

4.3.1.1 L1 Baseline Subjects

Figure 7 below shows the breakdown for L1 subjects of the Contrastive Focus category results into the conditions where the adjective is focused and where the noun is focused.

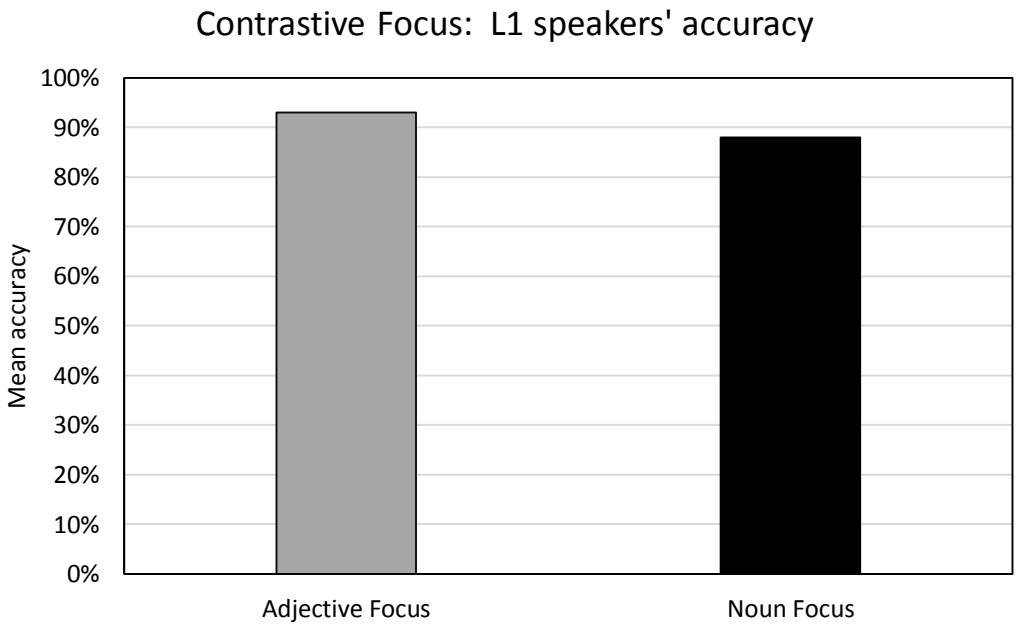


Figure 7: Contrastive Focus Category: L1 speakers' accuracy

As can be seen, the L1 subjects performed very accurately in both conditions, with 93% correct in the Adjective Focus condition and 88% in the Noun Focus condition. These levels of accuracy will be treated as the baseline level that L2 subjects can be compared with.

4.3.1.2 L2 Subjects: Chinese

The results from Chinese speakers are compared to L1 speakers in Figure 8 below.

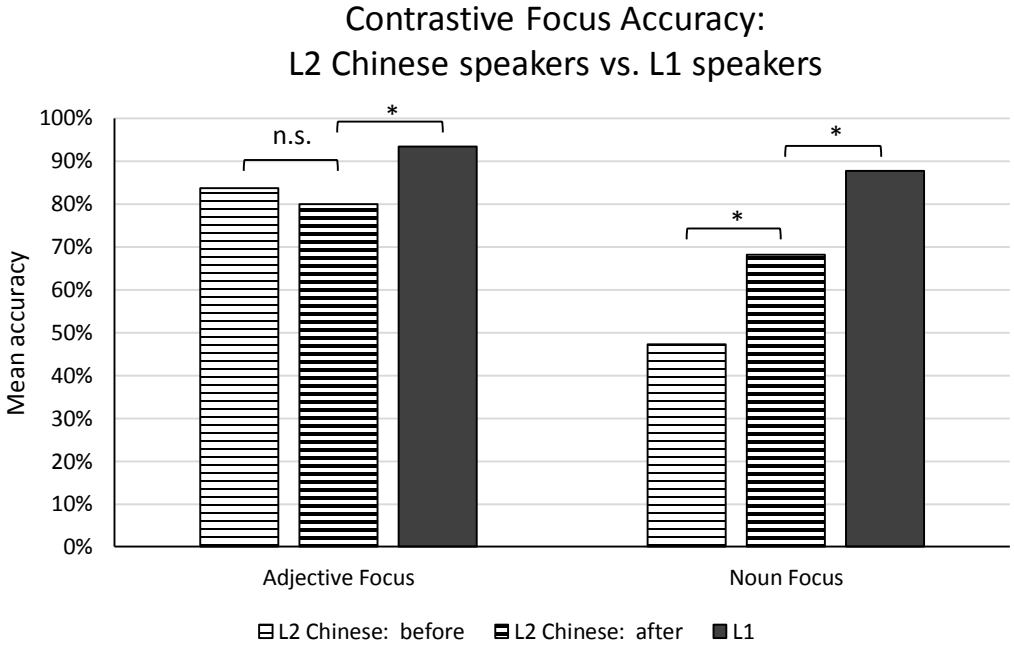


Figure 8: Contrastive Focus Category: L2 Chinese speakers' accuracy compared with L1 speakers. Significance is marked in terms of logistic regression comparisons.

We can see from Figure 8 that Chinese speakers performed well in the Adjective Focus condition even before training (84%), so there was limited room for improvement after training, whereas they performed poorly before training (47%) in the Noun Focus condition, leaving substantial room for improvement after training. Indeed, the graph indicates that Chinese speakers' scores rose to 68% in the Noun

Focus condition following training and stayed roughly the same from before to after training for the Adjective Focus condition, at 80%.

Logistic regression analyses were conducted for each condition (Adjective Focus and Noun Focus) to predict group membership (L1 or L2 Chinese 'after training', and L2 Chinese 'before training' or 'after training'), using accuracy as a predictor. For the Adjective Focus condition, comparing the groups of L1 speakers to L2 Chinese speakers after training, a test of the full model against a constant only model was statistically significant, indicating that the predictor of accuracy reliably distinguished between the L1 and L2 Chinese speakers (chi square = 7.809, p=.005 with df = 1). For the comparison of 'before training' to 'after training' for Chinese speakers, the model was not statistically significant, indicating that the accuracy levels were not different from each other. Thus, while there appears to be a slight decrease in performance for the Adjective Focus condition, this difference is not significant. For the Noun Focus condition, comparing the groups of L1 to Chinese speakers after training, the model was also statistically significant, again indicating that the predictor of accuracy reliably distinguished between the L1 and Chinese speakers (chi square = 11.263, p=.001 with df = 1). For the comparison of 'before training' to 'after training' for Chinese speakers in the Noun Focus condition, the model was also statistically significant, indicating that there was a significant difference in accuracy between before and after training (chi square = 9.936, p=.002 with df = 1).

Hence, Chinese speakers performed similarly to L1 speakers in the condition of Adjective Focus, and the training appeared to cause significant improvement in the Noun Focus condition, although the accuracy rates did not quite reach that of L1 speakers.

4.3.1.3 L2 Subjects: Arabic

Figure 9 below presents the results of the Arabic speakers on the Contrastive Focus category, and compares with those of L1 speakers. It should be noted that since there was limited data available for Arabic speakers, the results will be discussed by speaker and only descriptively in comparison with L1 speakers.

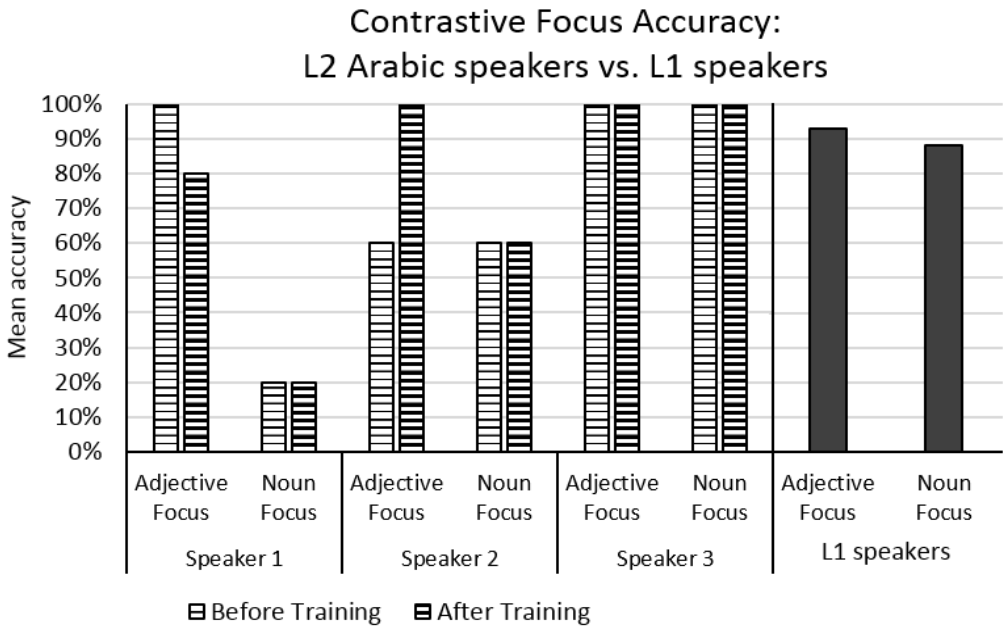


Figure 9: Contrastive Focus Category: L2 Arabic accuracy, by speaker, compared with L1 speakers overall

As can be observed from Figure 9 above, the results are very mixed. Arabic speaker 3 performed perfectly (100%) in this category, both before and after training, in both the Adjective Focus and Noun Focus conditions. Speaker 1 performed perfectly (100%) in the Adjective Focus condition, but poorly (20%) in the Noun Focus condition before training. While in the Adjective Focus condition there was a ceiling effect for

this speaker, the training did not improve this speaker's accuracy in the Noun Focus condition. Speaker 2 performed moderately (60%) in both conditions before training, and improved to 100% in the Adjective Focus condition, but again did not improve in the Noun Focus condition.

The results from the three speakers seem to suggest that the Adjective Focus condition is more easily attained than the Noun Focus condition for Arabic speakers, evident either by an exceptional prior understanding of the Adjective Focus condition or by improvement through training. All three speakers essentially reached the level of L1 speakers in the Adjective Focus condition, while only one speaker performs at (or even exceeds) the mean accuracy level of native speakers in the Noun Focus condition.

4.3.2 Compliment Category

In this section, the accuracy results of the Compliment category will be presented for L1 and L2 speakers, with a concentration on the comparison of the L2 Chinese speakers' results to the L1 speakers, as a baseline.

4.3.2.1 L1 Baseline Subjects

Figure 10 below shows the breakdown for L1 subjects of the Compliment category results into the Basic meaning condition and the Nuanced meaning condition.

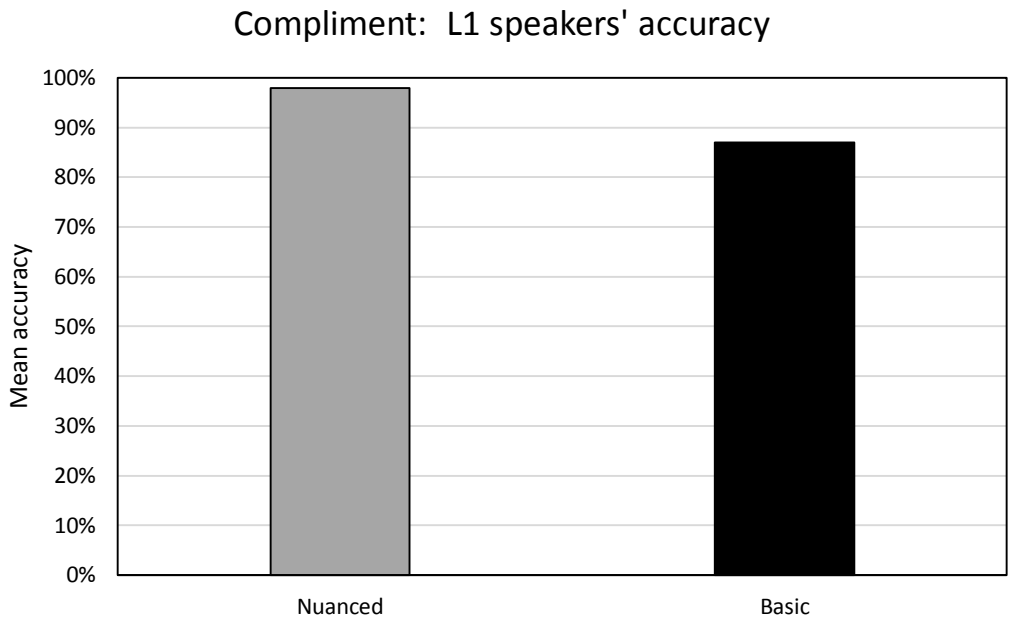


Figure 10: Compliment Category: L1 speakers' accuracy

As can be seen, the L1 subjects performed highly accurately in both conditions, though the accuracy level in the Nuanced meaning condition is still substantially higher (98%) than in the Basic meaning condition (87%). These levels of accuracy will be treated as the baseline level that L2 subjects can be compared with.

4.3.2.2 L2 Subjects: Chinese

Figure 11 below shows the results of L2 Chinese speakers in the Compliment category, broken down into subcategories.

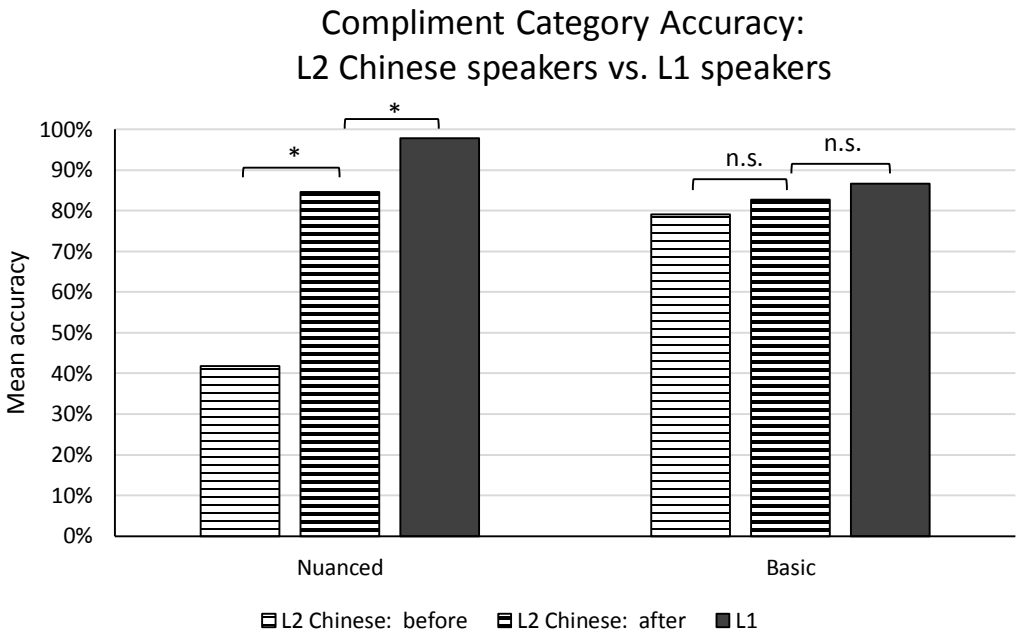


Figure 11: Compliment Category: L2 Chinese speakers' accuracy compared with L1 speakers. Significance is marked in terms of logistic regression comparisons.

As can be seen in Figure 11, participants were fairly accurate before training in the Basic meaning condition (79%), which was expected since they likely had more experience with the Basic meaning condition beforehand. Hence, there was little room for improvement after training, where they performed with 85% accuracy. In the Nuanced meaning condition, there was much room for improvement as participants only performed at 42% before training, and training seemed to help greatly, as accuracy rose to 83% after training.

Logistic regression analyses were conducted for each condition (Basic meaning and Nuanced meaning) to predict group membership (L1 or L2 after training, and L2 before training or L2 after training), again using accuracy as a predictor. For

the Basic meaning condition, comparing the groups of L1 to L2 after training, as well as L2 before and after training, the models were not statistically significant, indicating that there was no difference in accuracy level for L1 speakers as compared with L2 speakers, nor was there a difference after training compared with before training. Thus, while there appears to be a slight increase in performance between before and after training for the Basic meaning condition, this difference is not significant, and there is no difference between the accuracy levels of L1 and Chinese L2 speakers. For the Nuanced meaning condition, comparing the groups of L1 to Chinese L2 speakers after training, the model was statistically significant, indicating that the predictor of accuracy reliably distinguished between the L1 and L2 speakers (chi square = 11.688, p=.001 with df = 1). For the comparison of 'before training' to 'after training' for Chinese speakers in the Nuanced meaning condition, the model was also statistically significant, indicating that there was a significant difference in accuracy between before and after training (chi square = 45.265, p=.000 with df = 1).

Chinese speakers exhibited significant improvement after training occurred in the Nuanced meaning condition, though they did not quite attain the accuracy level of native speakers. For the Basic meaning condition, however, they did attain native speaker level, suggesting that this condition is more easily attained than the former.

4.3.2.3 L2 Subjects: Arabic

The results of the Arabic speakers' data for the Compliment category compared with those of L1 speakers are presented in Figure 12 below. Again, the limited results here will be discussed by speaker and only descriptively in comparison with L1 speakers.

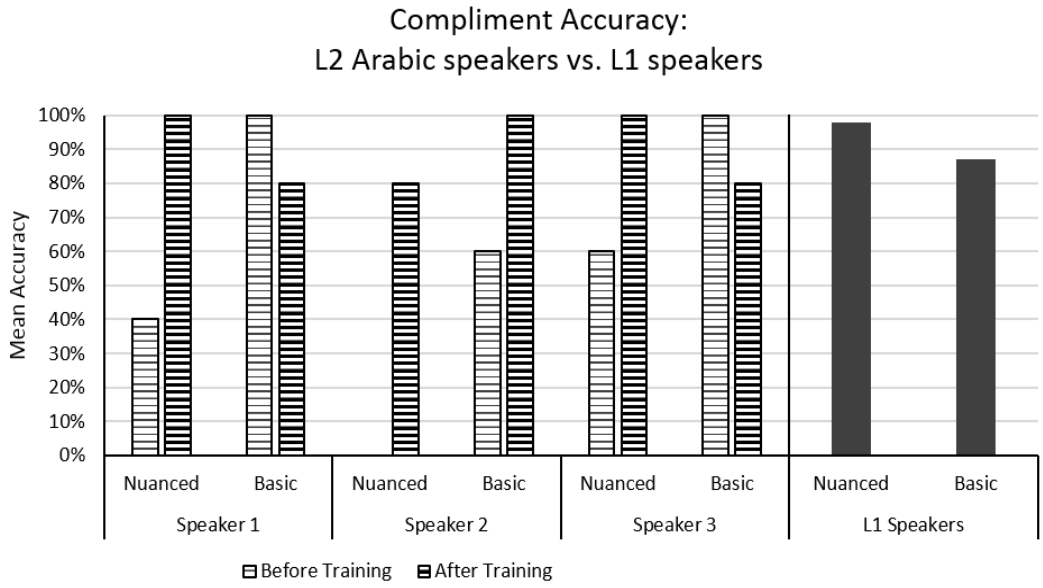


Figure 12: Compliment Category: L2 Arabic accuracy, by speaker, compared with L1 speakers overall

As can be seen in Figure 12 above, before training, the Arabic speakers performed much better in the Basic meaning condition than in the Nuanced meaning condition, with 100% accuracy for speakers 1 and 3 and 60% for speaker 2 in the Basic meaning condition, and 40% for speaker 1, 0% for speaker 2, and 60% for speaker 3 in the Nuanced meaning condition. However, this disparity evened out after the training due to marked improvement in the Nuanced condition: 100% accuracy was achieved for speakers 1 and 3, and 80% was achieved for speaker 2 in the Nuanced meaning condition; speakers 1 and 3 achieved 80%, and speaker 2 achieved 100% in the Basic meaning condition. The two instances where accuracy decreased after training (both from 100% to 80% in the Basic meaning condition) can be attributed to a ceiling effect and is unlikely to be a noteworthy drop.

4.3.3 Verb Focus Category

In this section, the accuracy results of the Verb Focus category will be presented for L1 and L2 speakers, with a concentration on the comparison of the L2 Chinese speakers' results to the L1 speakers, as a baseline.

4.3.3.1 L1 Baseline Subjects

Figure 13 shows the breakdown for L1 subjects of the Verb Focus category results into the Basic meaning condition and the Nuanced meaning condition.

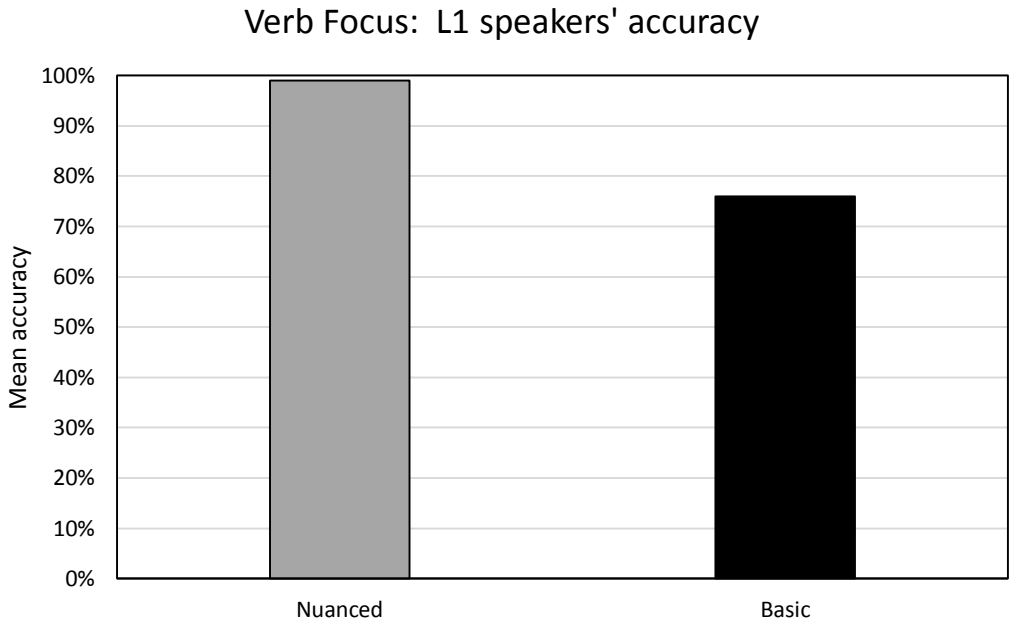


Figure 13: Verb Focus Category: L1 speakers' accuracy

The L1 subjects performed highly in both conditions, although the accuracy level in the Nuanced meaning condition is substantially higher (99%) than in the Basic

meaning condition (76%). These levels of accuracy will be treated as the baseline level that L2 subjects can be compared with.

4.3.3.2 L2 Subjects: Chinese

Figure 14 below presents accuracy results for the L2 Chinese speakers in the Verb Focus category, broken down into subcategories.

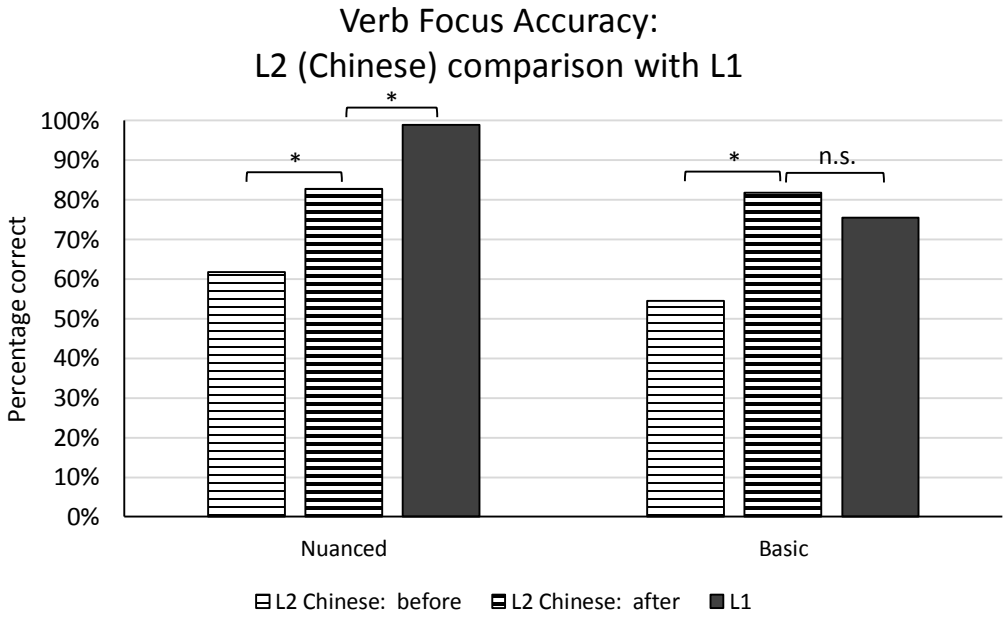


Figure 14: Verb Focus Category: L2 Chinese speakers' accuracy compared with L1 speakers. Significance is marked in terms of logistic regression results.

Before training, the Chinese participants scored only 62% in the Nuanced meaning condition and 55% in the Basic meaning condition. One-sampled t-tests against .5 (chance) revealed that these accuracy levels were not different from chance. We can observe from the graph that substantial improvement occurred in both the

Nuanced and Basic meaning conditions after training, where their accuracy rose to 83% in the Nuanced meaning condition and 82% in the Basic meaning condition.

Similarly to the other prosodic categories, logistic regression analyses were conducted for each condition (Basic meaning and Nuanced meaning) to predict group membership (L1 or L2 Chinese after training, and L2 Chinese before training or after training), again using accuracy as a predictor. For the Nuanced meaning condition, comparing the groups of L1 to Chinese L2 speakers after training, the model was statistically significant, indicating that accuracy level significantly differed between the L1 and L2 Chinese speakers (chi square = 17.804, p=.000 with df = 1). For the comparison of L2 Chinese speakers before and after training in the Nuanced meaning condition, the model was statistically significant, indicating that there was a significant difference in accuracy between before and after training (chi square = 12.229, p=.000 with df = 1). This suggests that the training improved their performance, albeit not quite to the point of native speakers.

For the Basic meaning condition, comparing the groups of L1 to L2 Chinese after training, the model was not statistically significant, indicating that there was no difference in accuracy level for L1 speakers as compared with L2 Chinese speakers. Thus, while the Chinese speakers seemed to outperform native speakers in the Basic meaning condition after the training, this difference was not significant. However, comparing the groups of L2 Chinese before and after training, this model was statistically significant, indicating that there was a significant difference in accuracy between before and after training (chi square = 19.324, p=.000 with df = 1). Hence, the training helped them to attain the level of native speakers in the Basic meaning condition.

4.3.3.3 L2 Subjects: Arabic

The results of the Arabic speakers' data for the Verb Focus category compared with those of L1 speakers are exhibited in Figure 15 below. Again, the limited results here will be discussed by speaker and only descriptively in comparison with L1 speakers.

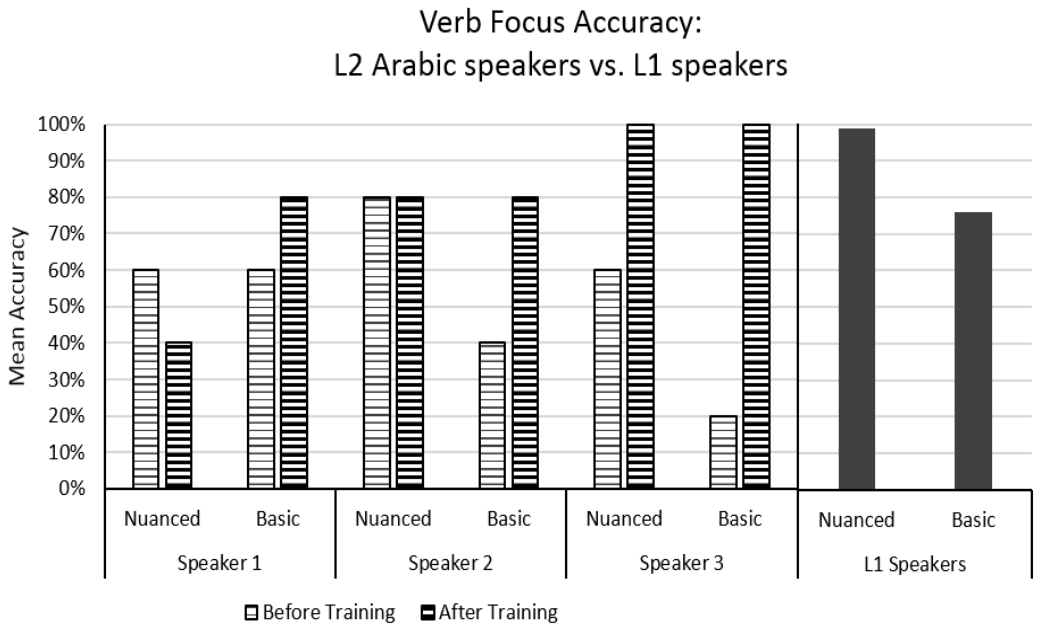


Figure 15: Verb Focus Category: L2 Arabic accuracy, by speaker, compared with L1 speakers overall

This category in general showed relatively poor results for Arabic speakers before training: speaker 1 performed with 60% accuracy in both Nuanced and Basic meaning conditions, speaker 2 performed with 40% accuracy in the Basic meaning condition, and speaker 3 performed with 60% accuracy in the Nuanced meaning condition and 20% in the Basic meaning condition. The one exception to the poor

results before training is the Nuanced meaning condition for speaker 2, who scored 80% there. The results improved for all speakers in the Basic meaning condition after training, especially for speaker 3 where a very substantial improvement occurred. For the Nuanced meaning condition, substantial improvement was seen again for speaker 3, and speaker 2 already performed quite well before training. The lowering of accuracy for speaker 1 was not substantial and likely isn't noteworthy.

Following training, the Arabic speakers all either reached or exceeded the accuracy level of native speakers for the Basic meaning condition. The accuracy levels are also relatively high for the Nuanced meaning condition, with the exception of one speaker. While the data is very limited, these results suggest that the training may indeed help in the attainment of the patterns.

4.3.4 Summary of Accuracy Results

For the L2 Chinese speakers in the Contrastive Focus category, logistic regression analyses showed that there was a significant difference between the before and after training conditions for the Noun Focus condition, but not for the Adjective Focus condition. Comparisons between speaker groups (L1 vs. L2 Chinese after training) show that there were significant differences between the groups in both the Adjective Focus and Noun Focus conditions, suggesting that the L2 Chinese subjects do not quite achieve the accuracy level of native speakers, even though major improvement occurred in the Noun Focus condition. L2 Chinese speakers were already rather proficient in the Adjective Focus condition, thus there was little room for improvement there.

For the Compliment category, logistic regression analyses showed a significant difference between the before and after training conditions for the Nuanced meaning

condition, but not for the Basic meaning condition. Comparisons between speaker groups (L1 vs. L2 Chinese after training) revealed a significant difference between the groups in the Nuanced meaning condition, suggesting that the L2 Chinese subjects did not quite achieve the accuracy level of native speakers, although major improvement occurred here. There was no significant difference between the speaker groups in the Basic meaning condition, suggesting that this condition was mastered for the Chinese L2 speakers.

Finally, for the Verb Focus category, logistic regression analyses revealed a significant difference between the before and after training conditions for both the Nuanced and Basic meaning conditions. Comparisons between speaker groups (L1 vs. L2 Chinese after training) showed a significant difference between the groups in the Nuanced meaning condition, suggesting that the L2 Chinese subjects did not quite achieve the accuracy level of native speakers, although major improvement occurred there. Similar to the Compliment category, since there was no significant difference between the speaker groups (L1 vs. Chinese L2 after training) in the Basic meaning condition, this suggests that the condition was mastered for the Chinese L2 speakers.

For Arabic speakers, while the data is too limited to make strong conclusions about the patterns, the training did seem to help accuracy in most cases where there was room for improvement. It was expected that they would perform better in the Contrastive Focus category than any other, since they were likely to be most familiar with this pattern from prior exposure. While they were ultimately quite successful in the Adjective Focus condition, the results were more mixed with the Noun Focus condition; hence, they were not able to fully master this pattern. In the other two

categories, however, they ultimately performed well in nearly every condition; suggesting that training was very effective for these categories.

4.4 Signal Detection Theory Analysis: d' and C

Signal detection theory was applied here to determine first whether there was sensitivity towards the signal (d') and second whether there was a bias (C) towards a particular meaning response. The d' measure allows separation of the effect of sensitivity to the distinction being measured from the participant's bias for one particular type of response (Stanislaw & Todorov 1999). Both d' and C are implemented here to determine whether some of the higher percentages of accuracy are a result of bias, or whether the subjects were indeed sensitive to the signal in these cases.

A d' score of 0 indicates an inability to distinguish the non-signal from the signal and greater values indicate a greater ability to do so (max: $+\infty$). The d' score is based upon a comparison of the 'hits' and 'false alarms'. In the case of the Compliment category, for example, the number of hits would be the number of times the nuanced auditory stimulus is matched correctly with the nuanced meaning sentence continuation and the number of false alarms would be the number of times the basic auditory stimulus is mismatched with the nuanced meaning sentence continuation.

The C-statistic represents the response bias, essentially the minimum level of certainty needed for the observer to decide a signal was present. Negative numbers indicate a bias for responding in a particular way (usually a 'yes' response, but here it will correspond with the meaning that matches the signal stimulus), whereas positive

numbers indicate a bias of responding in the opposite way (with the meaning that matches the non-signal stimulus).

While bias is not necessarily expected for L1 speakers (i.e. C values very close to zero are expected) in this experiment, there are a couple of factors that could come into play: frequency of observation of each of these types of prosody in real conversation, personal preference for either a more negative or positive response (relevant for the Compliment or Verb Focus categories). Another complication is that in the case of the basic auditory stimuli for the Compliment and Verb Focus categories, the nuanced meaning could sometimes be considered an acceptable alternative to the basic meaning. It should be noted, however, that the basic meaning could never be considered an acceptable continuation to a nuanced auditory stimulus.

For L2 speakers, it was expected that accuracy levels would be lower, since they are likely to have less familiarity with several of the prosodic conditions. Hence, it is especially important to consider bias in the before training condition if higher accuracy levels exist. Bias towards choosing the basic meaning in the Compliment and Verb Focus categories is possible if they are generally more used to hearing basic meaning responses, since they would have less exposure to the nuanced prosody. Similar to the Compliment and Verb Focus categories, one factor that could affect bias in L2 speakers in the Contrastive Focus category is the frequency of observation of Adjective Focus meanings versus Noun Focus meanings.

The sensitivity (d') and bias (C) results will be discussed in the following sections by prosodic category, comparing L1 subjects to L2 Chinese subjects. Arabic L2 subjects' data is not considered here because it is too limited to perform statistical analyses on.

The results will be presented by prosodic category in the sections that follow.

4.4.1 Contrastive Focus Category

Table 1 below shows the results of sensitivity and response bias for L1 speakers and L2 Chinese speakers in the Contrastive Focus category. The focused prosody is produced in the same way whether it is on the adjective or the noun, but it could be argued that Noun Focus is more unmarked than Adjective Focus in this context, given that phrasal stress rules assign focus to nouns in an adjective-noun pair. Hence, we can choose focus on the noun to be the non-signal and focus on the adjective to be the signal.

Table 1: Contrastive Focus Category: SDT results of L2 Chinese speakers compared with L1 speakers

Contrastive Focus Category	L2 Chinese speakers: Before	L2 Chinese speakers: After	L1 Speakers
d'	1.144215	1.552608	2.421074375
C	-0.58404	-0.21036	-0.130217846

L1 speakers showed a high sensitivity to the difference between the adjective stimulus being focused and the noun stimulus being focused, evident by the relatively high d' score. The C statistic is only slightly negative, suggesting that there is a very slight bias towards choosing the adjective-associated meaning. This was not expected, but this may not be noteworthy given that it is only a very slight bias.

Before the training, the Chinese speakers also had a somewhat high d' score, but not as high as the L1 speakers. This value increased slightly after training,

indicating a higher sensitivity level afterwards. However, it does not reach the level of that of the native speakers, so this is clearly still a pattern that is somewhat difficult for them, perhaps surprising given that this prosodic type is more widely known and studied. Similar to L1 speakers, Chinese speakers both before and after training, had a slight bias towards choosing the adjective-associated meaning, as indicated by the negative C value. The bias weakened after training, and thus starts to approximate the level of native speakers.

4.4.2 Compliment Category

Table 2 below encompasses the d' scores and C values for the responses of all of the Chinese speakers and compares against the L1 speakers in the Compliment category. The neutral prosody (Basic meaning condition) for the Compliment category is the non-signal (being more unmarked) and the marked prosody (the Nuanced meaning condition) is the signal.

Table 2: Compliment Category: SDT results of L2 Chinese speakers compared with L1 speakers

Compliment Category	L2 Chinese speakers: Before	L2 Chinese speakers: After	L1 Speakers
d'	0.719765	2.3148186	2.681287016
C	0.5886526	-0.0466377	-0.304210119

L1 speakers had a fairly high d' score, meaning that they are quite sensitive to the difference between the signal (the nuanced auditory stimulus) and the non-signal (the basic auditory stimulus). The C statistic is slightly negative, however, meaning

that there is a slight bias towards choosing the nuanced meaning, which was consistently a more negative meaning. Before the training, the Chinese speakers had a d' score much closer to zero, as compared with after their training and with that of the L1 speakers. This indicates a much lower sensitivity to the signal (nuanced stimulus) before the training as compared with after the training, when they approached the level of L1 speakers. In terms of a bias towards a particular response, the positive value of the C statistic before training indicates there was a bias towards choosing the basic meaning, which was predicted due to less familiarity with the nuanced condition. However, this changed to showing almost no bias either way after the training, becoming much more similar to the L1 speakers.

4.4.3 Verb Focus Category

Table 3 below displays the d' scores and C statistic values for the Verb Focus category responses. Similar to the Compliment category, the neutral prosody (Basic meaning condition) for the Verb Focus category is the non-signal (being more unmarked) and the marked prosody (the Nuanced meaning condition) is the signal.

Table 3: Verb Focus Category: SDT results of L2 Chinese speakers compared with L1 speakers

Verb Focus Category	L2 Chinese speakers: Before	L2 Chinese speakers: After	L1 Speakers
d'	0.558913	2.178342	2.167586586
C	-0.09944	-0.0216	-0.470980302

The Verb Focus category shows a similar pattern to the Compliment category. L1 speakers showed a high sensitivity to the difference between the nuanced stimulus and the basic stimulus, evident by the relatively high d' score. The C statistic is negative, again meaning that there is a slight bias towards choosing the nuanced meaning, which was consistently a more negative meaning. Before the training, the Chinese speakers had a relatively low d' score, but this value increased after training to the level of L1 speakers, indicating a high sensitivity level afterwards. In contrast to the L1 speakers, the Chinese speakers before and after training had very little bias in any direction. The bias is minimized even more after training. Interestingly, the bias does not change in the direction of L1 speakers from before to after training, as it did with the Contrastive Focus and Compliment categories.

4.4.4 Summary of Results on Signal Detection Theory Analysis

The results indicate that before going through the training, the Chinese speakers were not very accurate in the perception of the prosodic categories, especially in comparison with L1 speakers. They showed less sensitivity to the signals and showed a different type of bias from that of the L1 speakers. The accuracy and sensitivity scores rose fairly dramatically after receiving the training, however. The higher accuracy levels that were observed before training in the Adjective Focus condition of the Contrastive Focus category and the Basic meaning condition of the Compliment category were likely due to a bias in choosing the adjective-based meaning responses and the basic meaning responses, respectively. Following training, however, in the Contrastive Focus category, Chinese speakers showed similar bias to the L1 speakers. Thus, in terms of bias, they did seem to attain the pattern, if we consider the ideal bias pattern to be that of the L1 speakers (as opposed to the ideal

pattern being no bias at all). In the Compliment category, the Chinese speakers' bias changed dramatically from going in the opposite direction of bias from L1 speakers before training to going in the direction (albeit only slightly) of the L1 speakers' bias after training. In the Verb Focus category, they showed little bias either way before or after training. This is in contrast with the L1 speakers, who showed a more substantial negative bias (towards the Nuanced meaning responses) in this category.

4.5 Discussion

L2 speakers performed roughly as expected prior to targeted training in this experiment. They performed best in the Contrastive Focus category initially, albeit the high accuracy results were mostly limited to the Adjective Focus condition. In the other two prosodic categories, any initial success lied mainly in the Basic meaning conditions. It was previously unknown what effect targeted training would have on the understanding of these prosodic patterns, but overall we observed a strong positive effect.

Specifically, in the Contrastive Focus category, L1 speakers had a slight bias towards choosing meanings associated with the adjectives being focused. Chinese speakers nearly attained the accuracy level of L1 speakers in the Adjective Focus condition and improved greatly in the Noun Focus condition after training, but a large gap remained between them and the L1 speakers in the latter condition. Similar accuracy results to Chinese speakers are yielded for Arabic speakers, as well; they ultimately performed well on the Adjective Focus condition, and results improved somewhat after training in the Noun Focus condition. The high accuracy levels of the Chinese speakers can in part be attributed to a substantial bias (in excess of that for L1 speakers) towards the adjective-associated meanings, especially before training; the

training seemed to reduce that bias in the direction of L1 speakers, however. Why might listeners prefer an adjective-based meaning instead of a noun-based meaning? This could be due to frequency effects: it may be more common to emphasize adjectives within a phrase than nouns. Alternatively, it could be related to expectations of more natural emphasis on nouns, given it being in a broad focus position phrase-finally. Hence, listeners may overlook audible cues for noun focus in favor of non-prosodic cues for adjective-associated meanings.

In the Compliment category, Chinese speakers did well before training in the Basic meaning condition, but had less success in the Nuanced meaning condition. The SDT analysis presented in Section 4.4 confirmed the strong bias towards choosing the Basic meaning response, which was opposite to that of L1 speakers, who had a slight bias towards choosing the Nuanced meaning response (consistently a negative meaning). Their initial bias toward choosing a Basic meaning response indicates both a strong lack of prior exposure to and lack of prior understanding of this prosodic pattern. This is unsurprising given that the prosodic pattern and meaning of the Nuanced meaning condition is very rarely taught. However, the training vastly improved the sensitivity of Chinese subjects to the stimuli, and changed the direction of bias towards that of the L1 speakers, suggesting that it is possible to successfully teach the meaning of this type of prosody. The preliminary results of Arabic speakers' data support this claim, as they also performed ultimately well following training.

In the Verb Focus category, the accuracy levels greatly increased after training occurred for both Chinese and Arabic speakers, indicating that the training was effective. Chinese speakers exhibited essentially no bias at all either before or after training, while L1 speakers had a moderate bias towards the nuanced meaning.

Hence, training did not seem to affect the bias of Chinese speakers in this category. A question arises as to why L1 speakers exhibited bias towards the nuanced meaning. Perhaps this is due to the type of verbs utilized in this category. Nuanced meanings in this prosodic context normally are used with verbs such as *try* or *want*, verbs that without prosodic context can sometimes be interpreted with an overall negative result as most natural. For example, the more plausible meaning of the sentence "Jim wanted to go fishing" may be that while he had the desire to go fishing, he was not able to for some reason, perhaps because he had another obligation instead. Hence, they may have bypassed prosodic cues in favor of an interpretation of the lexical meaning. Another possibility is that in the absence of obvious prosodic cues (since the prosody is neutral in the Basic meaning condition), they were looking for anything to interpret; perhaps the nuanced meaning continuations were more interesting to them than the basic meaning continuations. L2 Chinese speakers, on the other hand, may not have the same intuitions, since the subtleties of lexical meaning and sentence meanings having differing levels of interest may be limited to the most experienced language users.

It was previously unknown whether training could successfully teach the meaning of lesser known prosodic patterns, but this study with targeted training shows that accuracy indeed improved in most cases where there was room for improvement; moreover, in two out of three prosodic patterns, training shifted the Chinese speakers' bias closer to that of L1 speakers. One question that remains is why L2 speakers ultimately performed with slightly lower accuracy after training in the presumed familiar pattern (Contrastive Focus category) than in the presumed unfamiliar patterns (Compliment category and Verb Focus category). Perhaps having a preconceived

notion actually interfered in the learning process: learning a pattern from scratch may actually yield better comprehension than attempting to correct prior assumptions. In order to obtain maximum effectiveness, more care might need to be taken in creating a training that addresses and targets previously held misunderstandings of these patterns. More data from the language backgrounds studied here as well as other language backgrounds would be necessary to determine whether this training is effective more generally across language backgrounds.

Chapter 5

PRODUCTION STUDY

The production experiment was designed to test non-native speakers' ability to produce prosodic patterns associated with specific pragmatic meanings in English, and to compare against a baseline of L1 English speakers.

The results concentrate on data from Chinese speakers, but in the interest of checking whether other language backgrounds would yield similar results, as well as testing whether the training could be useful for speakers of other languages besides Chinese, the experiment was piloted on some Arabic speakers, whose results will also be presented here.

This chapter will first present the research questions considered, the methodology used in the experiment, followed by the results of the baseline of L1 speakers and L2 Chinese speakers, preliminary results of the L2 Arabic speakers, with a summary/brief discussion of the results. A more extensive discussion related to both experiments will be given in the next chapter.

5.1 Research Questions

The general question of interest here is how L2 speakers perform in the production of English prosody. We can explore answers to this question by investigating L2 speakers' production of a few prosodic patterns. Are there specific types of patterns they do better at? As with the perception study, we might expect that they would perform better on previously familiar patterns (that may have been taught)

than more unfamiliar patterns that may require experience with naturalistic data. In the cases where L2 production patterns are different from L1 speakers, can targeted training on these specific types of prosodic patterns help L2 speakers transform their patterns into L1 patterns? Another question that can be raised regarding L2 learners of English is which acoustic cues they would be most attuned to: specifically, whether certain properties of their L1 (e.g. duration and pitch) influence prosody in the L2 (i.e. English). That is, since Chinese is a language with tone, does this make pitch more accessible to be used for speakers of this language? Similarly, does Arabic as a language with contrastive vowel length make duration more accessible to be used by its speakers?

5.2 Methodology

The format of the production experiment is as follows in Figure 16 below:

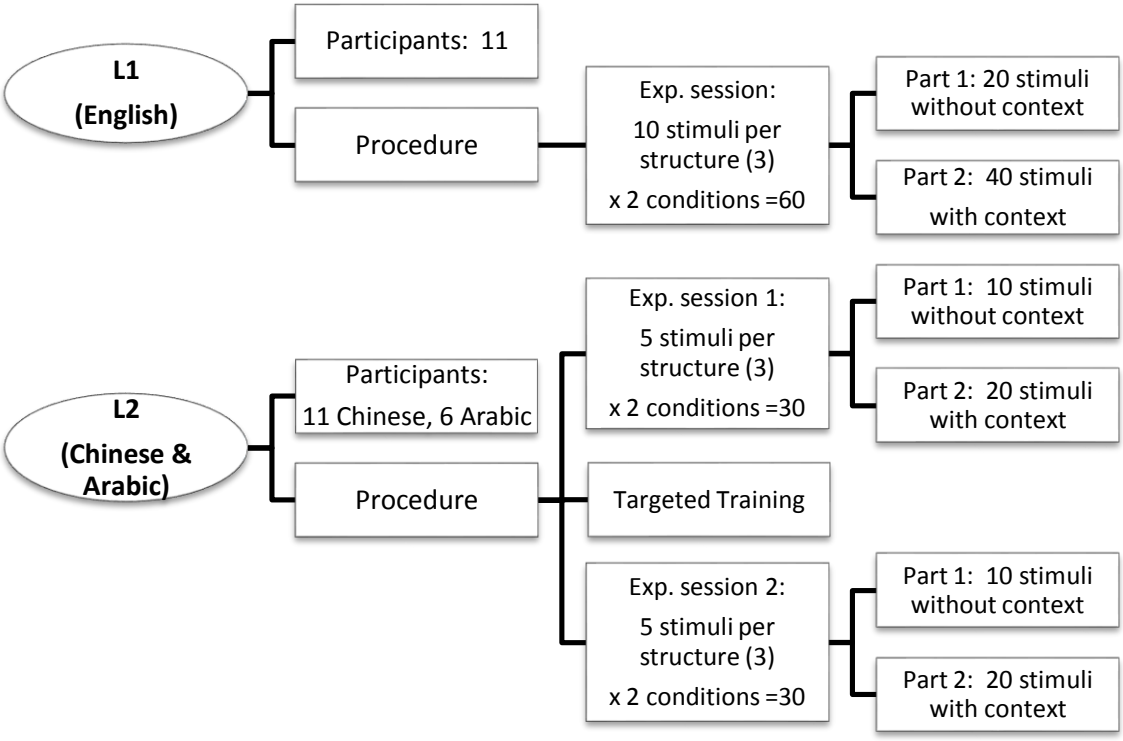


Figure 16: Schematic of Production Experiment[4]

Figure 16 outlines for L1 and L2 subjects the basic elements of the experiment, including numbers of participants and the procedure followed. This will be explained in more detail below.

[4] The numbers of stimuli given in the schematic are the numbers of target stimuli actually measured and presented in the Results section. Additional fillers and conditions originally included but not discussed, are excluded here.

5.2.1 Participants

Sixteen L1 participants were tested in the experiment, of which the data from 11 (female = 10, mean age = 20) will be presented. Five participants were excluded due to having had a speech disorder, being bilingual, or having parents with a different native language. Participants were recruited in the same method as described for the perception experiment.

Eleven Chinese subjects (female = 6, mean age = 21) were recruited in the same method as for the perception experiment from the highest two levels (5 and 6) of English classes at the English Language Institute at the University of Delaware for participation in this study. Additionally, six Arabic subjects were tested (M=4, Mean age = 21). These subjects were distinct from the subjects tested in the perception experiment. The L2 Chinese speakers had all grown up in China and the Arabic speakers grew up in Saudi Arabia. All subjects were between the ages of 18-36 and had been in the United States between 2 months and 2 years. None had any history of speech, hearing, or reading impairments. Additionally, no subject had parents with a different native language than their own.

5.2.2 Stimuli

The same stimuli as in the perception experiment were tested in this production experiment, with an addition of 4 new target sentences to the practice session, as it became clear during pilot studies that more practice was needed for production. For L1 speakers, in Part 1 of the experiment, there were a total of 20 target sentences. The same filler sentences as in the perception experiment were used. Each sentence appeared alone on a slide. They were presented in this way to elicit a neutral prosody, to compare against the other nuanced types of prosody. In Part 2 of the experiment,

40 sentences, all target sentences, were presented in conjunction with a written context. No context was more than 3-4 sentences long, and they only included vocabulary that was deemed[5] to be familiar to the L2 speakers who would be tested (see Appendix A for exact contexts). There were 10 target sentence stimuli for each category, and for one category (Contrastive Focus), those 10 sentences were repeated in order to elicit focus on both the adjective and the noun. Since prosodic patterns are often so difficult to elicit, in each of the sentences involving a context, there was one word that was bolded and italicized to help guide subjects in producing the patterns. An example slide from the Compliment category is as follows:

Context:

My friend Amy tries hard, but she doesn't get good grades in school. My mother asks me if Amy is smart. I look for something positive to say, so I tell her…

"Amy is ***popular***"

An example slide from the Verb Focus category is the following:

Context:

My friend Isabelle always promises to visit, but she never comes. This time she insists that she's actually coming. I don't really believe it, but I tell my family…

"Isabelle ***says*** she's coming to visit"

Finally, 04 below presents slides from the Contrastive Focus category for one example sentence stimulus, with focus on the adjective versus the noun:

---

[5] This was deemed by the author, who has worked extensively with second language learners of the participants' levels.

Table 4: Example contexts for stimuli of Contrastive Focus category

Adjective Focus	Noun Focus
Context: My sister is helping me wash dishes after dinner. I notice that there are still 2 plates left on the table: one is plastic and one is glass. She told me that the plastic plate is dirty, but… "The ***glass*** plate is clean"	**Context:** I'm cleaning off the dinner table. There are two glass dishes on the table: a bowl and a plate. I'm in a rush, so I tell Suzie to only bring me the glass bowl to wash because… "The glass ***plate*** is clean"

No fillers existed in the second part of the experiment, as there was no intended focus in the filler sentences. In fact, fillers were completely removed from the L2 production experiment, as without any context present, it did not seem that subjects would be aware of specific categories that the target stimuli would belong to. All stimuli and their corresponding contexts can be seen in Appendix A.

The practice session at the beginning of the experiment contained 10 items; none of the practice items were used in the experimental sessions.

For the L2 speakers, the format of the stimuli was the same as for the L1 speakers, but they were split into two sessions, before training and after training, as will be described in the procedure section.

5.2.3 Procedure

L1 subjects were tested in the Phonology-Phonetics Laboratory at the University of Delaware in a quiet booth. Due to ease of location for recruiting subjects, L2 subjects were tested at the English Language Institute's Self Access Learning Center, a computer lab, with low volume background noise from other computers and quiet conversation. All were recorded directly to a laptop via an external head-mounted microphone (Logitech ClearChat USB). The recordings were sampled at 44,100 Hz via Praat.

Subjects first signed a consent form (Appendix B) and filled out a background questionnaire, similar to those for the perception experiment. They were instructed that they would be recording sentences in English, and that the study was aimed at investigating the intonation of second language learners of English. L1 subjects were told that they would serve as the baseline group for this study, while L2 subjects were told they would be part of a group of speakers from their language background and that the data used from their recordings would not be linked to their name. All were told that they would be recorded the entire time and that they should say each sentence as naturally as possible.

Recording began, and a PowerPoint presentation led them through the experimental sessions and training (the latter only for L2 speakers). For L1 speakers, who only participated in one experimental session, the experiment lasted roughly 15 minutes. For L2 speakers, receiving a training and additional experimental session, the experiment lasted roughly 45 minutes. At the beginning of the experiment, they were presented with 10 practice slides, the first two of which were sentences without any written contexts, and the latter eight of which were sentences preceded by a context that they read silently beforehand. They were asked to read aloud the sentence

following the context, while pretending to be in that situation. The contexts preceding the target sentences aided in making the intended foci natural. In addition, there was one word in the sentence that was bolded and italicized, and subjects were instructed to put extra emphasis on this word in a way that they felt was appropriate, given the context. These types of stylistic additions have been shown in previous studies (Xu, 1999) to be a successful way to reduce errors. Indeed, in pilot studies where bolding and italicization was not incorporated, the intonation patterns seemed much more varied and often times very different from what was intended. After the practice session, the experimental sessions differed for L1 and L2 subjects.

L1 subjects completed Part 1 of the experiment, which consisted of reading out loud the sentences on 50 slides, 20 of which were targets and 30 of which were fillers. Next, they completed Part 2 of the experiment, consisting of reading out loud the target sentences on the remaining 40 slides[6] after having silently read the associated written contexts. They were instructed that if they stumbled during the pronunciation of the sentence, or if they felt that they did not say the sentence in a way they thought was appropriate, they should repeat the sentence from the beginning. No feedback by the experimenter was given to the subjects, except for situations in which the subjects stumbled within the sentence, and then were subsequently asked to repeat the sentence. Repetitions for most speakers were not often needed. There were two

---

[6] The experiment was initially set up so that 3 different focus conditions could be compared in the Contrastive Focus category: neutral focus (presented without context), adjective stress (presented with context) and noun stress (presented with context). Hence, L1 subjects actually saw 10 additional slides (neutral condition for Contrastive Focus) without context. It was later decided that the neutral focus condition was unnecessary for the current analysis, so that data is not presented here.

orders of stimuli used in the experiment, similar to how the perception experiment was conducted. Half of the participants in this study received one order of the stimuli (A) and the other half received the other order of the stimuli (B) during the experiment. The stimuli were presented in a fixed pseudo-randomized order in both versions A and B. No more than two of the same stimuli condition appeared directly after one another.

For L2 speakers, each experimental session (before training) was nearly identical to that of the experiment for L1 speakers, but contained only half the number of target items that the experiment for L1 speakers had. Following Experimental Session 1, the Targeted Training occurred, after which they participated in Experimental Session 2, in which they received the remaining half of stimuli to see how much they learned from the training. Doubling the number of stimuli, as was done with the perception experiment for L2 speakers, did not seem practical, due to the extension of time to the experiment. Another difference in the experiment for L2 speakers is that the filler sentences were completely removed, as it was deemed that the experiment was too long with the additional sections, and as previously explained, it did not seem that the fillers were actually necessary in this format. Thus, in Experimental Session 1, there were a total of 30 slides, containing 30 target sentences, and in Experimental Session 2, there were an additional 30 slides, containing 30 different target sentences. In both experimental sessions, 10 sentences (5 from the Compliment and Verb Focus categories) were presented without contexts and 20 appeared in contexts[7]: 5 in each of the Compliment and Verb Focus categories, and 5

[7] See previous note. L2 subjects actually saw 5 additional slides (neutral condition for Contrastive Focus) without context.

for each of the adjective and noun Contrastive Focus conditions. The selection of the stimuli for Experimental Session 1 or 2 was random, as it was considered that all stimuli were equally representative of their respective categories. The target sentences throughout this experiment were identical to those for the L1 speakers. Feedback was only provided to L2 subjects if they had a question about the pronunciation of a name used in the sentence or if the experimenter wanted them to repeat the sentence due to speech disfluency. Only questions regarding the meaning of vocabulary were answered, but this was not a common occurrence.

5.2.4 Analysis

The same method of measurement and normalization was applied to L1 speakers and L2 speakers. The Contrastive Focus category was analyzed in a slightly different way from the Compliment and Verb Focus categories due to a different stimuli structure. In all cases, however, F0, duration and intensity (the traditional measurements of prosody) were measured. Inferential statistical analyses are presented for the L1 and L2 Chinese groups, but only descriptive results are provided for the Arabic speaker results, as the statistical power would be considered too low for a meaningful analysis.

5.2.4.1 Contrastive Focus Category

The measurement and analysis methods for duration, F0 and intensity in the Contrastive Focus category will be described in the sections that follow. Results from the Neutral Focus condition will not be presented here, as it is more relevant here to investigate relative properties of the adjective and noun when one or the other is focused.

5.2.4.1.1 Duration

Praat textgrids were created for each target sentence and the target words (adjective and noun in the subject noun phrase) were segmented. The length of the target word (adjective or noun) was compared against the length of the adjective and noun as a unit; the duration of the rest of the sentence was not considered relevant. Segmentation boundaries were marked at positive zero crossings for the target words. A Praat script by Mietta Lennes from SpeCT - The Speech Corpus Toolkit for Praat (2011) was used to extract the durations for the words in each recording. Duration ratios of word:phrase (i.e. adjective: adjective+noun or noun: adjective+noun) were then calculated for each target sentence. Means were taken of the ratios for the Adjective Focus and Noun Focus conditions. Additional normalization between speakers was not considered necessary for this measurement, as the ratios represent a way to determine rate of speech, one common method of normalization.

Two-way repeated measures ANOVAs were performed separately for L1 speakers, L2 Chinese speakers before training, and after training. The independent variables were FocusCondition (Adjective Focus vs. Noun Focus), and MeasuredWord (which word was being measured: adjective or noun); the dependent variable was duration. A three-way mixed ANOVA was not done here, since for the current study, any comparable values between training conditions (between-subject variable) were not relevant; only the patterns within each training condition were of interest.

5.2.4.1.2 F0

A Praat script called ProsodyPro (Xu, 2013) was used to extract F0 and intensity values for the target word in each sentence. ProsodyPro is designed to automate the processing of large amounts of speech data at a high level of accuracy,

and involves F0 trimming and time-normalization algorithms. This is especially relevant for the Compliment and Verb Focus categories, which will be discussed in section 5.2.4.2, but was also used for this prosodic category for the sake of uniformity.

In this category, maximum F0 was determined for the stressed syllables for the adjective and noun in each utterance through the ProsodyPro script. Max F0 was chosen as opposed to mean F0 to avoid averaging means when converting to z-scores. Max F0 values were converted to natural log values to better approximate normal distributions. Per speaker, averages were taken of the max F0 values (transformed into log values) across conditions. Next, z-scores for each max F0 value were calculated for each speaker, so that all speakers could be combined. Since the z-score values are not as informative compared to raw F0 (Hertz) values, a transformation was done to convert the z-scores back into simulated raw values, in order to better visualize the data with meaningful units (Hz). The conversion was completed by using the mean and standard deviation of the values of one speaker, chosen as representative for each group of subjects. Females were chosen to represent the English and Chinese speakers, and a male for Arabic speakers, since females made up the majority of the former two groups, and males made up the majority of the latter group. Two-way repeated measures ANOVAs were performed on the z-transformed data of Max F0, in the same manner as for duration.

5.2.4.1.3 Intensity

Max intensity was used as the intensity measurement in this study, and was only measured on the stressed syllable, since this is the most likely location for a difference in intensity to be clearly exhibited when comparing a word in focus to a

word out of focus. Differently from F0, raw intensity values were not converted into log values, as decibels are already logarithmic.

Normalization of intensity values for this category followed the same process as for max F0, as presented above. Two-way repeated measures ANOVAs were performed on the z-transformed data of Max Intensity, in the same way as for duration and F0 data.

5.2.4.2 Compliment and Verb Focus Categories

The measurement and analysis methods for duration, F0 and intensity in the Compliment and Verb Focus categories will be described in the sections that follow.

5.2.4.2.1 Duration

For the Compliment and Verb Focus categories, Praat textgrids were created for each target sentence and the target word was segmented as well as the entire sentence. For these categories, the duration of the word in focus was measured with respect to the duration of the entire sentence. Segmentation boundaries were marked at positive zero crossings for both the word and sentence. The same Praat script used for the Contrastive Focus category was also used in this category to extract the durations for the word and sentence in each recording. Ratios of word:sentence durations were then calculated for each target sentence. Means were taken of the ratios for the Basic meaning condition separate from the Nuanced meaning condition.

Paired-sample t-tests were run for L1 speakers, as well as L2 Chinese speakers before and after training, to determine if duration differences were significantly different between focus conditions (Nuanced and Basic). Two-way mixed ANOVAs were not performed here, since for the current study, any comparable values between

training conditions (between-subject variable) were not relevant; only the patterns within each training condition were of interest.

5.2.4.2.2 F0

ProsodyPro (Xu, 2013) was used to extract F0 and intensity values for the target word in each sentence. One of the reasons for using the ProsodyPro script was its ability to smooth F0 and specify the number of time intervals desired for analysis, for the purposes of creating F0 contours. Time-normalized F0 values were extracted into a text file for the entirety of sound (.wav) files in each condition (Basic meaning or Nuanced meaning).

Only target words (i.e. the focus-targeted word in the Nuanced auditory sentences and the same word (not focused) in the Basic auditory sentences) were analyzed for each sentence. The sonorant portions of the rimes were analyzed for each syllable of the target word, beginning with the stressed syllable of the word and continuing to the right edge of the word. The maximum number of syllables analyzed in a word was three, because the stressed syllable was never further to the left than the antepenultimate syllable. However, some words only contained one analyzable syllable (the stressed one), others contained two (the stressed one followed by an unstressed one), and still others contained three (the stressed one followed by two unstressed ones). For example, in the sentence from the Compliment category, "Kim is intelligent", the right-most three syllables of the target word "intelligent" were considered, as the stressed syllable of the word was the antepenultimate. In the sentence from the Verb Focus category "Gary appears to be having fun at the party", only the final syllable from the target word "appears" was analyzed, as stress falls on the final syllable, and in the sentence "George is hoping to get the job", both syllables

from the target word "hoping" were analyzed due to stress falling on the first syllable. It was necessary to perform measurements in this manner, as it allowed analysis of a mixed group of stress patterns in target words. Stimuli were first categorized by their number of analyzable syllables, in order to specify to ProsodyPro how many time-normalized segments to calculate over. Target stimuli with one analyzable syllable were given twelve time-normalized segments, those with two analyzable syllables were given six time-normalized segments (six over each syllable), and those with three analyzable syllables were given four time-normalized segments (four over each syllable), so that there were twelve-normalized segments over each target word. It is recognized that the length of each rime differed, which could conceivably have an effect on the shape of the overall contour. However, since the same nuanced target words were compared with their basic counterparts (and across speakers), we can nevertheless be confident that it is consistent across the Basic and Nuanced meaning conditions.

Once time-normalized F0 values were extracted for each target word, they were converted to natural log values. Normalization across speakers then proceeded in the following way: mean F0 values were obtained for each time-normalized point across basic stimuli for each speaker; z-scores were then calculated for each time-normalized point for each nuanced stimulus, based on the mean and standard deviation of the basic stimuli at each time-normalized point; finally, a mean of z-scores was taken along each time-normalized point, resulting in a final average contour for the nuanced pattern. A transformation was later performed to convert the z-scores back into simulated raw values for a better visualization of the contours. The conversion was completed by using the mean and standard deviations of each time-normalized

point for the basic meaning category of the same speaker used for conversion of the data in the Contrastive Focus category. F0 contours were only visually analyzed here, rather than statistically analyzed.

5.2.4.2.3 Intensity

Max intensity was used as the intensity measurement only on the stressed syllable, for the same reasons outlined for the Contrastive Focus category. Normalization of intensity values across speakers for the Compliment and Verb Focus categories consisted of the following: averaging intensity values for the Basic meaning condition, from which a mean and standard deviation for each speaker was determined; creating z-scores for the nuanced condition's values based on the mean and standard deviation of each speaker's basic meaning condition's values; finally, averaging the z-scores for each subject's Nuanced meaning condition. Similar to F0, for intensity, the average z-scores were then transformed back into the relevant units (dB) to form a more accurate visualization of the patterns. One-sample t-tests were performed on the z-scores of the Nuanced condition for L1 speakers, L2 Chinese speakers before training, and after training, to determine if these values were significantly different from zero.

5.3 Results

The results are presented by prosodic pattern, starting with the Contrastive Focus category (expected to be the easiest to master), followed by the Compliment Category and Verb Focus Category (expected to be harder to master). Within each pattern, the L1 results are presented first as a baseline, followed by the results of the L2 Chinese speakers. As was done in Chapter 4, the focus is on Chinese speakers and

the comparison to L1 English speakers. As a preliminary comparison, all of the results from Arabic speakers are presented at the end of the results section; as mentioned previously, more speakers would be needed to draw more solid conclusions.

5.3.1 Contrastive Focus Category

In this category, Adjective Focus is compared with Noun Focus. An example item from this category is "The red shirt is still wet", where either the adjective "red" or the noun "shirt" in the subject noun phrase is focused, depending on which context is given. Measurements were done on the adjective and noun to determine their relative properties. A summary of those measurements is presented in Table 5 below.

Table 5: Measurement summary table for Contrastive Focus category

	Duration	**F0 & Intensity**
Adjective focus Ex. "The ***red*** shirt is still wet"	Ratio: adj. to adj.+noun ***red*** shirt	Max F0 and Max Intensity on stressed syllable of adjective and noun
Noun focus Ex. "The red ***shirt*** is still wet"	Ratio: noun to adj.+noun red ***shirt***	

The results of the L1 and L2 data based on these measurements can be seen in the next sections.

5.3.1.1 Duration

For the Contrastive Focus condition, duration ratios of the target words (adjective or noun) to their corresponding phrases (adjective+noun) are presented here

for two conditions: when adjectives are focused (AF) and when nouns are focused (NF).

5.3.1.1.1 L1 Baseline Subjects

In Figure 17 below, the left two bars show the mean ratios of adjective duration (compared with the phrase), and the right two bars show the ratios of noun duration (compared with the phrase) in the Adjective Focus (AF) and Noun Focus (NF) conditions.

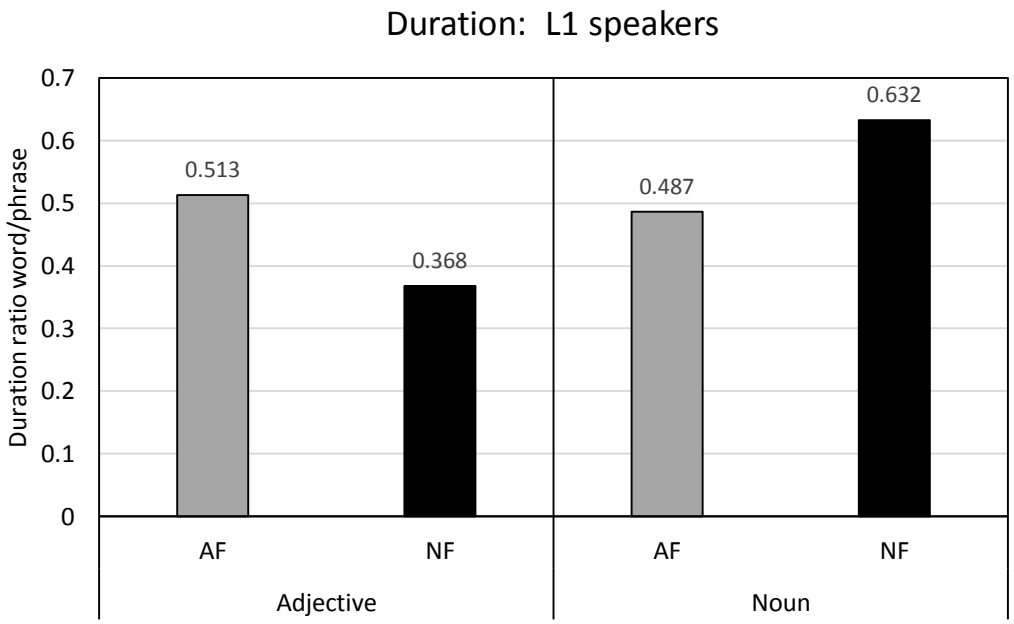


Figure 17: Contrastive Focus Category: duration ratios of adjective and noun to phrase for L1 speakers. AF=Adjective Focus; NF=Noun Focus

As can be seen in the left section of the graph, adjectives appear longer when they are focused (AF) than when they are not focused (NF). Similarly, as can be seen

in the right section of the graph, nouns appear longer when they are focused than when they are not focused. Also noteworthy is that adjectives are similar in length to nouns when in the AF condition (although the adjectives appear slightly longer than the nouns in this case). On the other hand, nouns appear substantially longer than adjectives when in the NF condition. That this is the case for the NF condition but not for the AF condition can be anticipated due to the combination of final lengthening with focus on the nouns in the NF condition: nouns are expected to be much longer than adjectives when the nouns are focused because of the combination of final lengthening with focus, whereas this combination does not exist for the AF condition. When the adjective is focused (AF condition), we expect the adjectives and nouns to be of comparable length (rather than the adjectives being much longer than the nouns) because the noun remains affected by final lengthening.

Two-way repeated measures ANOVAs with independent variables of MeasuredWord (adjective or noun) and FocusCondition (AF or NF) were run to confirm the significance of the observed results. There was a two-way interaction between MeasuredWord and FocusCondition, $F(1, 96) = 251.163$, $p=.000$, confirming the results that adjectives have different lengths from nouns, depending on which is focused. There was a significant main effect of the variable FocusCondition for adjectives and nouns, as well. This confirms that the focus condition made a difference in the duration usage on adjectives and nouns. For the variable MeasuredWord, while it had a significant main effect in the NF condition, $F(1, 96) = 327.933$, $p=.000$, it did not in the AF condition, $F(1, 97) = 3.279$, $p=.073$. Thus, it is confirmed that duration values were significantly different between adjectives and nouns in the NF condition, but not in the AF condition.

Overall, it appears that L1 speakers use duration to distinguish focused words from non-focused words in the Contrastive Focus category.

5.3.1.1.2 L2 Chinese Subjects

In this section, the acoustic property patterns used by L2 Chinese speakers for the Contrastive Focus category will be presented in comparison to the previously described patterns of the baseline L1 speakers, in terms of both descriptive and inferential statistics.

Figure 18 below presents the duration patterns for L2 Chinese speakers before and after training. The values presented here are duration ratios of the target word (adjective or noun) to phrase (adjective+noun) for the Adjective Focus (AF) condition and Noun Focus (NF) condition.

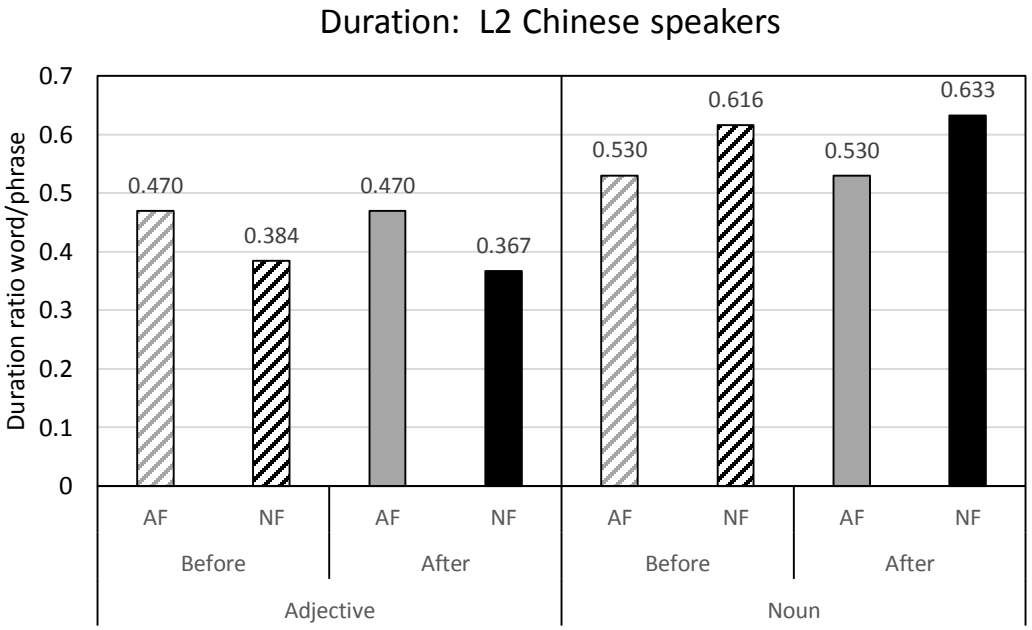


Figure 18: Contrastive Focus Category: Duration ratios of adjective and noun to phrase for L2 Chinese speakers. AF=Adjective Focus; NF=Noun Focus

As can be seen, the Chinese speakers exhibit nearly the same pattern as that of the native speakers, both before and after training: adjectives appear longer in the AF condition than NF condition, nouns appear longer in the NF condition than AF condition, and nouns appear longer than adjectives when nouns are focused (NF). These differences seemed to increase slightly after training, from .08 to .1 for the comparison of AF to NF conditions and .23 to .27 for the comparison of adjectives to nouns in the NF condition, pushing them closer to the distinction made by L1 speakers (.15 and .26, respectively). The only observed pattern difference between L1 and L2 Chinese speakers lies in the AF condition: while for L1 speakers, adjectives were the same length as nouns, for Chinese speakers, nouns appear longer than adjectives (by .06), and this does not diminish after training.

Two-way repeated measures ANOVAs with independent variables of MeasuredWord (adjective or noun) and FocusCondition (AF or NF) were run to compare results of the L2 Chinese speakers before training and after training and confirm the significance of the observed results. Similar to L1 speakers, there were two-way interactions between MeasuredWord (adjective vs. noun) and FocusCondition (AF vs. NF condition): $F(1, 49) = 60.675$, $p=.000$ for L2 Chinese speakers before training, and $F(1, 49) = 108.793$, p=.000 after training. Thus, it appears that like L1 speakers, the L2 Chinese speakers altered the length of adjectives and nouns depending on which was focused.

Moreover, there were significant main effects of the variable MeasuredWord in the NF condition both before training, $F(1, 49) = 95.807$, $p=.000$, and after training, $F(1, 49) = 170.468$, $p=.000$. However, the results of Chinese speakers in the AF condition differed from L1 speakers: L1 subjects did not exhibit a significant main effect of MeasuredWord in this condition, while Chinese speakers did both before training ($F(1, 49) = 6.647$, $p=.013$) and after training ($F(1, 49) = 6.383$, $p=.015$). Nouns for Chinese speakers were significantly longer than adjectives in the NF condition (consistent with the L1 pattern), as well as in the AF condition (departing from the L1 pattern). Training did not change the latter pattern for Chinese speakers.

Nevertheless, it is confirmed that L2 Chinese speakers performed with nearly identical patterns to L1 speakers for the property of duration, with the small exception of the adjective and noun length in the AF condition. In general, they appear to already have been competent with the duration property in the Contrastive Focus category before training.

5.3.1.2 Maximum F0

Maximum F0 values were measured on the stressed syllable of the target adjectives and nouns in the Adjective Focus and Noun Focus conditions.

5.3.1.2.1 L1 Baseline Subjects

Figure 19 below displays the Max F0 values of adjectives and nouns in both the Adjective Focus (AF) condition and Noun Focus (NF) conditions for L1 speakers: the left two bars show the F0 of adjectives and the right two bars show the F0 of nouns in the AF and NF conditions. Error bars represent standard error.

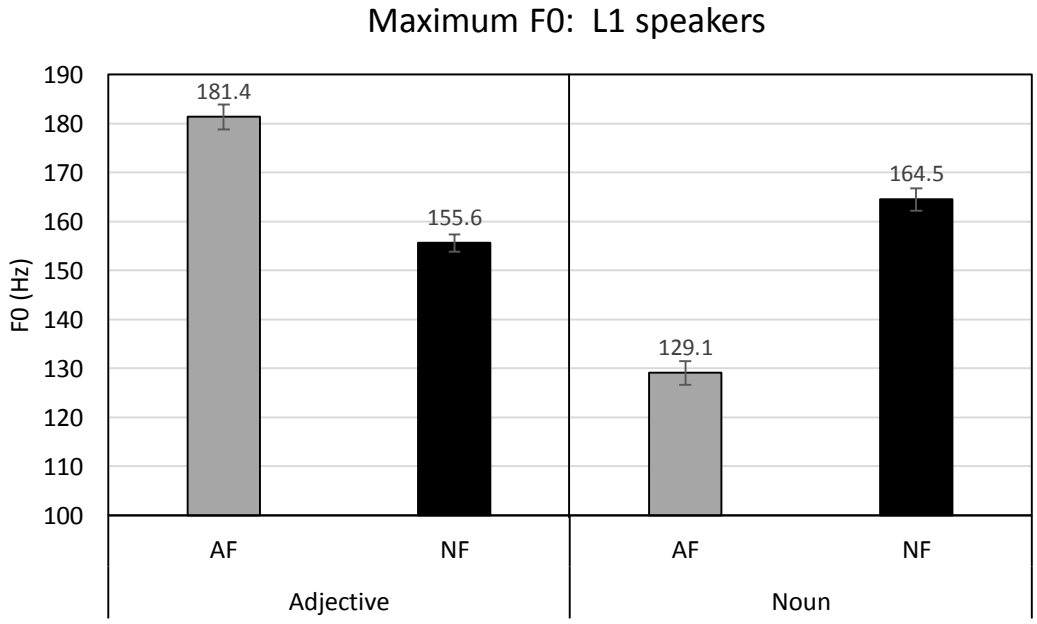


Figure 19: Contrastive Focus Category: Maximum F0 values for L1 speakers. AF=Adjective Focus; NF=Noun Focus.

As can be observed in the figure, adjectives have a considerably higher F0 than nouns when adjectives are focused (AF condition), and nouns appear slightly higher in F0 than adjectives when nouns are focused (NF condition). In addition, focused adjectives have a much higher F0 than non-focused adjectives and focused nouns are much higher in F0 than non-focused nouns. This is generally as expected, since F0 is typically higher on focused words than non-focused words. One somewhat surprising result is that the F0 difference between the adjective and noun in the NF condition is not very large, which suggests that F0 may not be the most important cue in distinguishing focus in the NF condition.

Two-way repeated measures ANOVAs with independent variables of MeasuredWord (adjective or noun) and FocusCondition (AF or NF) were run to confirm the significance of the observed results. There was a statistically significant two-way interaction between MeasuredWord and FocusCondition: $F(1, 96) = 133.503$, $p=.000$ for L1 speakers. Just like with duration, this is as expected, since we would anticipate adjectives to have different F0 from nouns, depending on which is focused.

There were significant main effects of the variable FocusCondition for adjectives, $F(1, 97) = 63.717$, $p = .000$, as well as nouns, $F(1, 96) = 83.919$, $p = .000$. This suggests that the focus condition affected the F0 usage on adjectives and nouns. There were also significant main effects of the variable MeasuredWord in the AF condition, $F(1, 97) = 235.896$, $p = .000$, as well as the NF condition, $F(1, 96) = 6.636$, $p = .012$, suggesting that duration depends on which word is being considered. F0 results were found to be significantly different for adjectives and nouns in both focus

conditions; hence, while the difference in F0 between adjectives and nouns for the NF condition (~9 Hz) appeared small, it was significant.

Thus, it appears that L1 speakers use F0 to distinguish focused words from non-focused words in the Contrastive Focus category.

5.3.1.2.2 L2 Chinese Subjects

Figure 20 below presents the results of the Chinese speakers on Max F0.

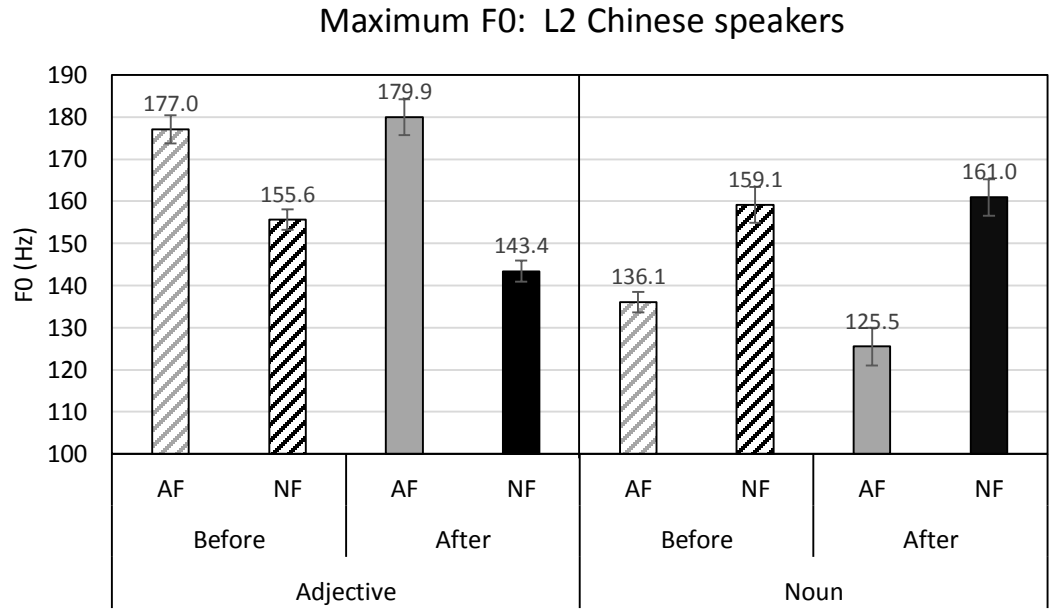


Figure 20: Contrastive Focus Category: Maximum F0 values for L2 Chinese speakers. AF=Adjective Focus; NF=Noun Focus.

As can be seen, the Chinese speakers exhibit a very similar pattern to native speakers, both before and after training: adjectives appear to have higher F0 when they are focused (AF) than not focused (NF), nouns appear to have higher F0 when they are

focused (NF) than not focused (AF), and adjectives have higher F0 than nouns in the AF condition. One detail that appears to differ: L1 speakers had slightly higher F0 for nouns than adjectives in the NF condition, while the Chinese speakers' data before training hardly seem to show any difference (4 Hz at most) between the word types. They do show this pattern after training, however, with a difference of nearly 18 Hz (slightly exceeding the difference made by L1 speakers).

Two-way repeated measures ANOVAs with independent variables of MeasuredWord (adjective or noun) and FocusCondition (AF or NF) were run to compare L2 Chinese speakers before training and after training and confirm the significance of the observed results. There was a statistically significant two-way interaction between MeasuredWord and FocusCondition before training, $F(1, 49) = 70.759$, $p=.000$, and after training $F(1, 49) = 98.816$, $p=.000$. Just like with duration, this is as expected, since we would anticipate adjectives to have different F0 from nouns, depending on which is focused.

There were significant main effects of the variable FocusCondition for adjectives before training, $F(1, 49) = 37.456$, $p=.000$, as well as after training, $F(1, 49) = 59.406$, $p=.000$. There were also significant main effects for nouns before training, $F(1, 49) = 27.238$, $p=.000$, and after training, $F(1, 49) = 45.537$, $p=.000$. This suggests that focus condition affected the F0 usage on adjectives and nouns.

There were also main effects of the variable MeasuredWord in the AF condition before training, $F(1, 49) = 200.927$, $p = .000$, and after, $F(1, 49) = 91.389$, $p = .000$. Thus, F0 results were significantly different for adjectives and nouns in the AF condition. On the other hand, for the NF condition, while there was a significant main effect of MeasuredWord for L1 subjects, there was no main effect for L2

Chinese subjects before training, $F(1, 49) = .671$, $p = .417$; after training, however, there was a main effect, $F(1, 49) = 16.152$, $p = .000$. Hence, we can confirm that while nouns had higher F0 than adjectives in the NF condition for L1 speakers, there was no difference in F0 for Chinese speakers before training. After training, however, the L1 pattern emerged for Chinese speakers, demonstrating a learning effect from training.

Hence, before training, L2 Chinese speakers performed with nearly identical patterns to L1 speakers for the property of F0 in the Contrastive Focus category, with the exception of the difference between adjectives and nouns in the NF condition. In this case, after training, the Chinese speakers' pattern approximated that of L1 speakers, showing that training was effective in the only area with room for improvement.

5.3.1.3 Maximum Intensity

Similar to F0, maximum intensity was measured on the stressed syllable of the target adjectives and nouns in both the Adjective Focus (AF) condition and Noun Focus (NF) conditions.

5.3.1.3.1 L1 Baseline Subjects

In Figure 21 below, the left two bars display the intensity of adjectives and the right two bars the intensity of nouns in the AF and NF conditions. Error bars represent standard error.

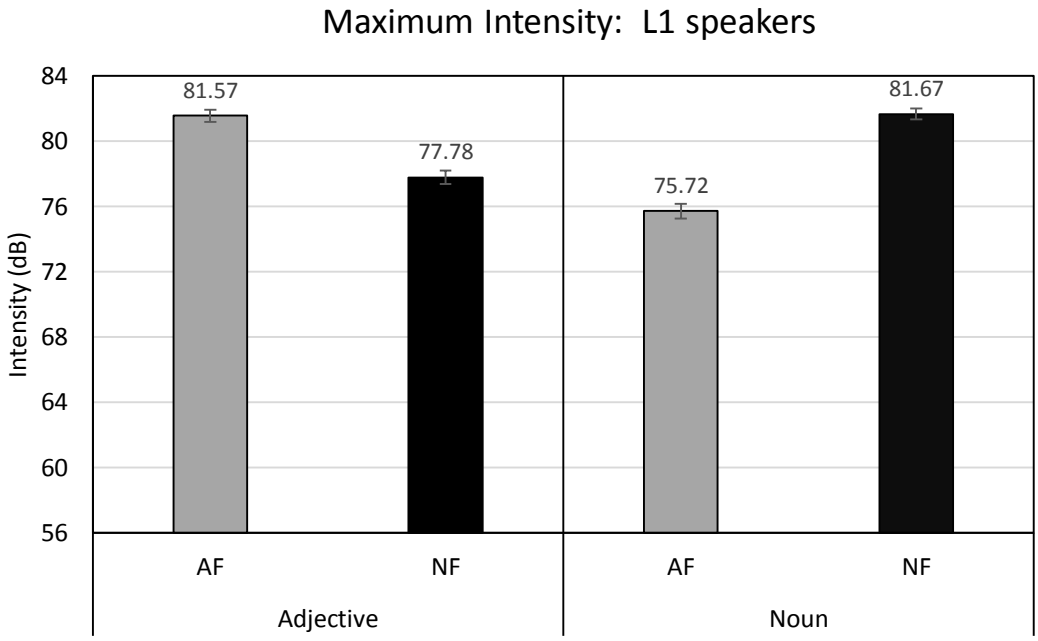


Figure 21: Contrastive Focus Category: Maximum intensity values for L1 speakers. AF=Adjective Focus; NF=Noun Focus.

If we compare Figure 19 (Max F0) with Figure 21, the same patterns for max F0 emerge for max intensity in L1 speakers. Adjectives have higher intensity than nouns when the adjective is focused (AF condition), and the reverse is seen when the noun is focused (NF condition). In addition, adjectives have higher intensity when they are focused than when they are not focused (seen in left section of graph), and nouns have higher intensity when they are focused than when they are not focused (right section of graph); this difference appears slightly greater for nouns than for adjectives.

Two-way repeated measures ANOVAs with independent variables of MeasuredWord (adjective or noun) and FocusCondition (AF or NF) were run to confirm the significance of the observed results. There was a statistically significant

two-way interaction between MeasuredWord and FocusCondition, $F(1, 96) = 243.699$, $p=.000$. As with duration and F0, this was expected, since we would anticipate adjectives to have different intensity from nouns, depending on which is focused.

There were significant main effects of the variable MeasuredWord in the AF condition, $F(1, 97) = 153.132$, $p=.000$, and the NF condition, $F(1, 97) = 55.362$, $p=.000$. Thus, intensity depended on which word was being considered, and this was true for both focus conditions. There were also significant main effects of the variable FocusCondition for adjectives, $F(1, 96) = 64.085$, p=.000, and nouns, $F(1, 96) = 223.451$, p=.000. These results suggest that the focus condition affected the intensity usage on adjectives and nouns.

Overall, it appears that L1 speakers use intensity as a property to distinguish focused words from non-focused words in contrastive focus.

5.3.1.3.2 L2 Chinese Subjects

Maximum intensity values of the stressed syllable in target adjectives and nouns are presented in Figure 22 below.

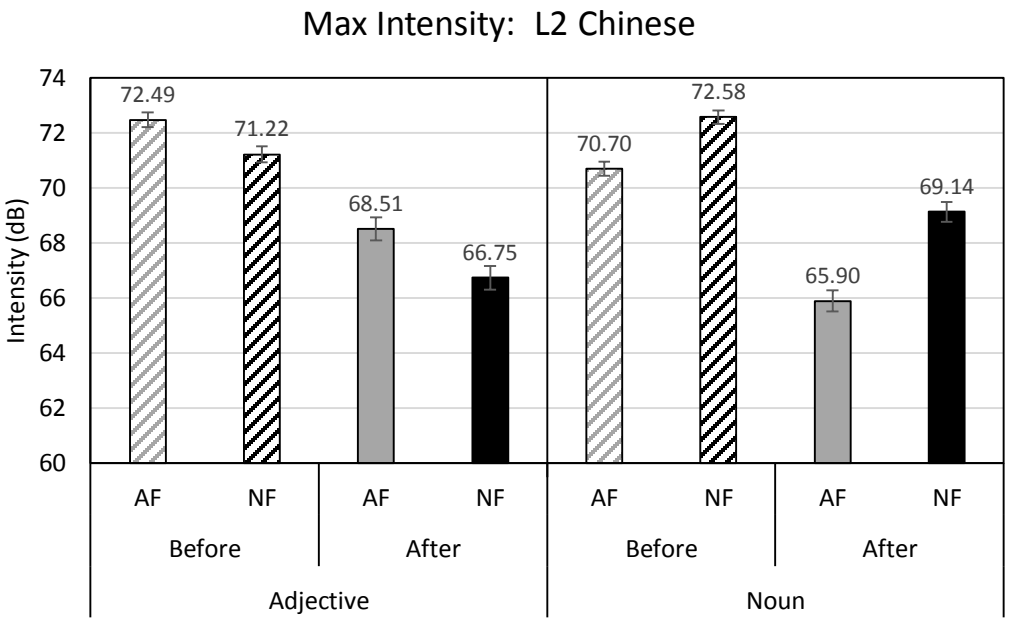


Figure 22: Contrastive Focus Category: Maximum intensity values for L2 Chinese speakers. AF=Adjective Focus; NF=Noun Focus.

It can be noted that the intensity range for Chinese speakers is much narrower than for the L1 speakers. While L1 speakers exhibited a difference in intensity of 4-5 dB between the AF and NF conditions, this difference is much smaller (1-2 dB) in Chinese speakers. However, the difference for Chinese speakers expands after training, to about 2-3 dB. A change by 1 dB is the smallest change the human ear can detect, thus it is very likely that the intensity differences seen here for the Chinese speakers would also be detectable; the differences are simply not as robust as for L1 speakers.

Two-way repeated measures ANOVAs with independent variables of MeasuredWord (adjective or noun) and FocusCondition (AF or NF) were run to compare Chinese speakers' results before and after training, and confirm the

significance of the observed results. Similar to L1 speakers, there was a statistically significant two-way interaction between MeasuredWord and FocusCondition both before training, $F(1, 49) = 50.032$, $p=.000$, and after training, $F(1, 49) = 56.859$, $p=.000$.

There were significant main effects of the variable FocusCondition for adjectives before training, $F(1, 49) = 16.769$, p=.000, and after training, $F(1, 49) = 17.267$, p=.000. In addition, there were significant main effects for nouns before training, $F(1, 49) = 26.251$, p=.000, and after training, $F(1, 49) = 34.787$, p=.000. These results suggest that the focus condition affected the intensity usage on adjectives and nouns (similar to L1 speakers), both before and after training.

There were also significant main effects of the variable MeasuredWord in the AF condition before training, $F(1, 49) = 21.11$, $p=.000$, and after training, $F(1, 49) = 24.515$, $p = .000$. This was also the case for the NF condition before training, $F(1, 49) = 22.082$, $p=.000$, and after training, $F(1, 49) = 22.86$, $p=.000$. Thus, as with L1 speakers, intensity between adjectives and nouns were significantly different, both for the AF and NF conditions.

Hence, it can be confirmed that the Chinese speakers used intensity (even before training) in the same manner as L1 speakers in the category of Contrastive Focus, and the training aided in distinguishing the conditions even more, bringing the level of differentiation closer to that of L1 speakers.

5.3.2 Compliment Category

In this category, a special kind of nuanced focus (Nuanced meaning condition) is compared with broad focus (Basic meaning condition). An example item from this category is “Emily has beautiful hair”, where the final word “hair” can be emphasized

to have a more nuanced meaning (i.e. a backhanded compliment, such as "but nothing else about her is great") or a more basic meaning (i.e. a true compliment). Measurements were performed on the target final word (which was always a content word) to determine the relative properties between the conditions. For duration, ratios of the target final word to the utterance were measured; for F0, contours over the final word were measured; for intensity, the maximum values of the stressed syllable of the target final word were measured. The results of the L1 and L2 data based on these measurements can be seen in the next sections.

5.3.2.1 Duration

Results of duration ratios of the target final word to utterance are presented in this section for the Nuanced meaning condition and the Basic meaning condition.

5.3.2.1.1 L1 Baseline Subjects

Figure 23 below shows the mean ratio of duration of the target final words to their corresponding utterances.

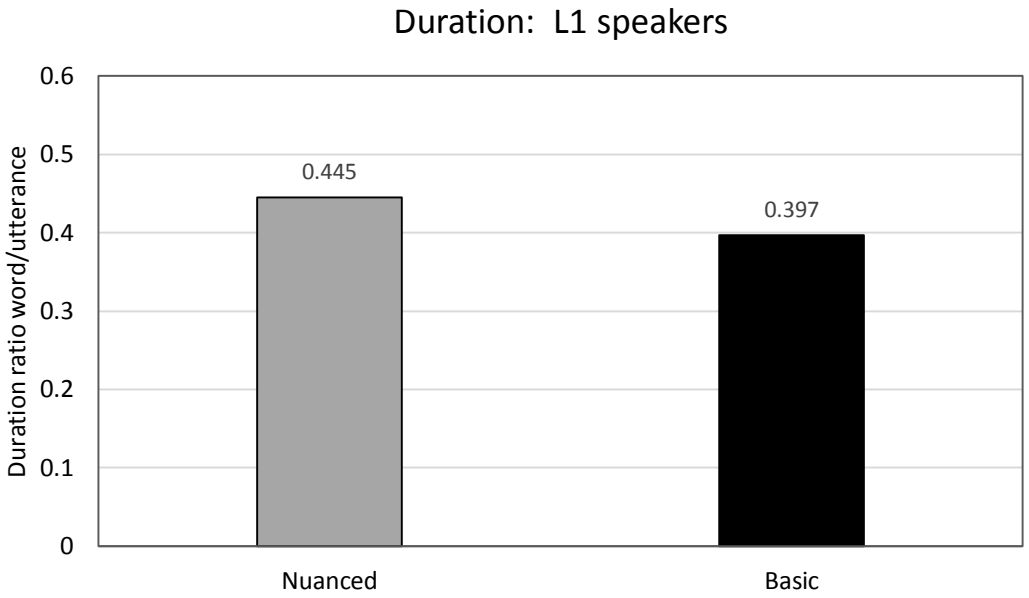


Figure 23: Compliment Category: Ratios of target (final) word to utterance for L1 speakers

As expected, the Nuanced meaning condition, in which the final (content) word is focused in a special way, has a higher ratio of word to utterance duration than the Basic meaning condition. A two-tailed paired-samples t-test was conducted to compare the ratios between the conditions, and a significant difference was found, $t(99)=7.422$, $p = .000$. This suggests that duration is a property used to distinguish these focus patterns in L1 speakers.

5.3.2.1.2 L2 Chinese Subjects

The duration results for Chinese speakers for the Nuanced and Basic meaning conditions, before and after training, are given below in Figure 24.

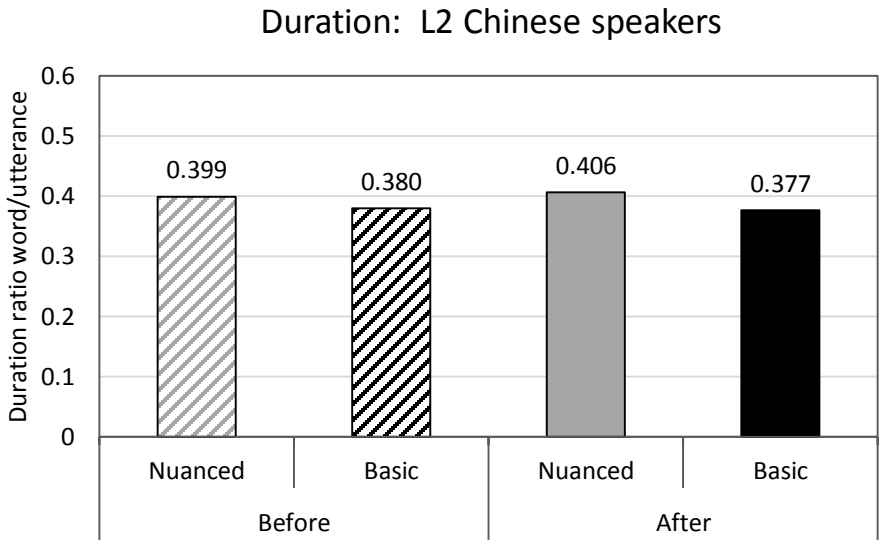


Figure 24: Compliment Category: Ratios of target (final) word to utterance for L2 Chinese speakers

The durational differences appear very minimal before training (.019 difference), but increase slightly following training (.029 difference), coming closer to the difference that L1 speakers exhibit (.048). In fact, both the difference before and after training is significant: $t(48) = 2.518$, $p = .015$ (before training) and $t(50) = 3.24$, $p = .002$ (after training). This indicates that, while small, there were significant differences in duration before and after training, such that Chinese speakers had already attained the pattern of L1 speakers before training was given.

5.3.2.2 F0 Contours

The results sections below on F0 contours will emphasize descriptive results, as it can be problematic to statistically compare contours in a meaningful way.

5.3.2.2.1 L1 Baseline Subjects

Figure 25 below shows the mean F0 contours of the final (content) word of the utterances in the Nuanced and Basic meaning conditions.

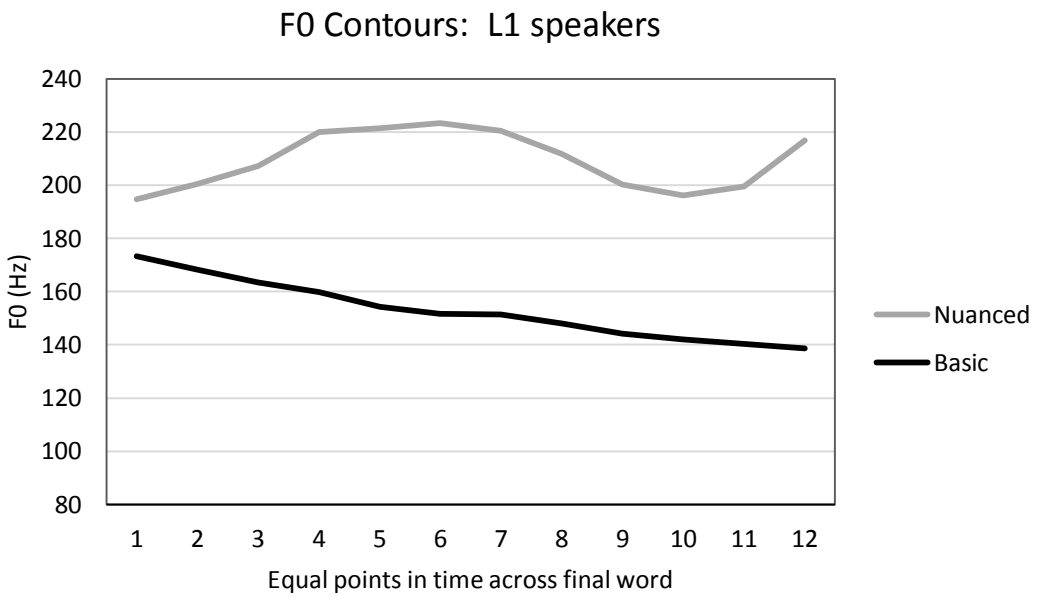


Figure 25: Compliment Category: F0 contours of target (final) word for L1 speakers

It can be seen that the Basic meaning pattern is simply a falling pattern, while the Nuanced meaning pattern involves a rise, followed by a fall, with an additional rise at the end (a boundary tone). Moreover, the F0 values for the Nuanced meaning condition are substantially higher than for the Basic meaning condition (close to 70-80 Hz different at the locations of greatest distinction).

5.3.2.2.2 L2 Chinese Subjects

Figure 26 below illustrates the F0 contour patterns of the Nuanced and Basic Meaning conditions both before and after training for Chinese speakers.

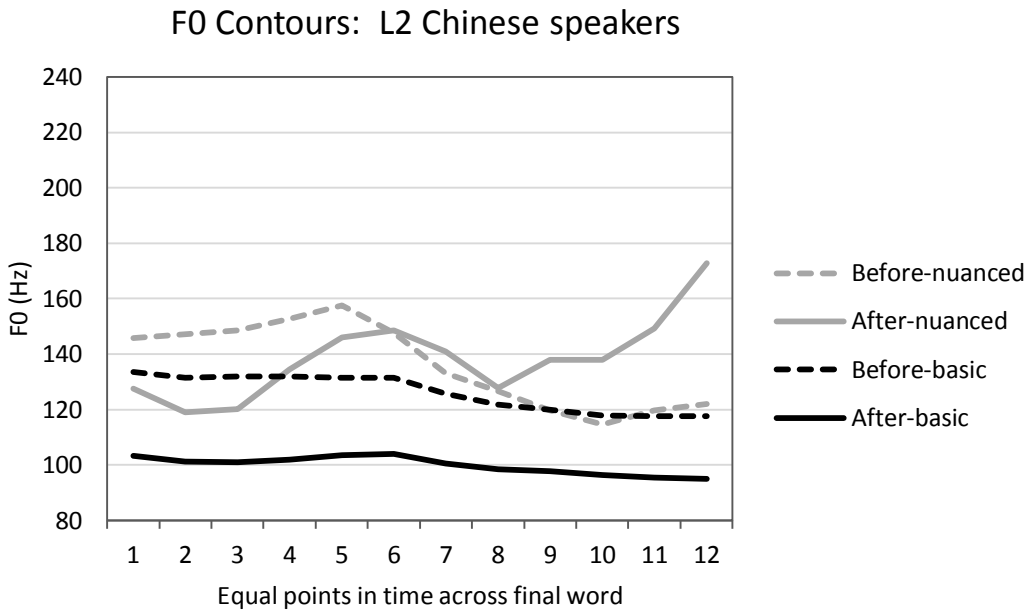


Figure 26: Compliment Category: F0 contours of target (final) word for L2 Chinese speakers

Before training, the initial part of the Nuanced meaning contour (the rise) mimics that of L1 speakers, although it is much less extensive than for L1 speakers. The latter part of the pitch curve illustrates a rather sharp fall, indicating that the Chinese speakers had not mastered the crucial rising-falling-rising pattern utilized by L1 speakers. After the training, however, the Nuanced meaning curve mimics the L1 speakers' fairly accurately, with the exception of a slight fall at the beginning.

The Basic meaning contours of the Chinese speakers before training already approximated the L1 speakers' patterns. After training, the Basic meaning contours decreased greatly in F0, retaining the same falling pattern. The Nuanced and Basic meaning patterns are thus made more distinct after training, very similar to the patterns of L1 speakers. Hence, the training was successful in making the Chinese speakers' patterns closer to L1 speakers'.

5.3.2.3 Maximum Intensity

Max intensity was measured on the stressed syllable of the final (content) word of the utterance.

5.3.2.3.1 L1 Baseline Subjects

Figure 27 below shows the comparison of the maximum values of intensity for the Nuanced and Basic meaning conditions. Error bars represent standard error.

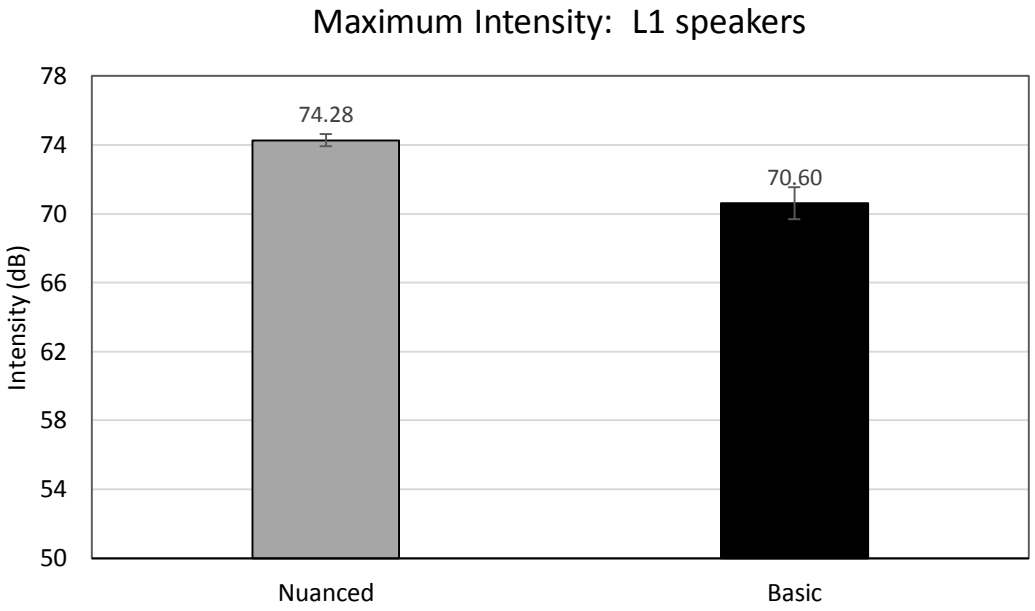


Figure 27: Compliment Category: Maximum intensity of target (final) word for L1 speakers

As expected, the Nuanced meaning category, in which the final (content) word is focused in a special way, has a higher max intensity than the Basic meaning category, with a difference of nearly 4 dB. A one-sample t-test was conducted on the z-scores of the Nuanced category against zero (taken as the mean of the Basic meaning condition), and the Nuanced meaning condition was found to be significantly different from zero, $t(99)=9.947$, $p = .000$. This suggests that intensity is used to distinguish these conditions by L1 speakers.

5.3.2.3.2 L2 Chinese Subjects

Figure 28 below shows the results of max intensity for Chinese speakers.

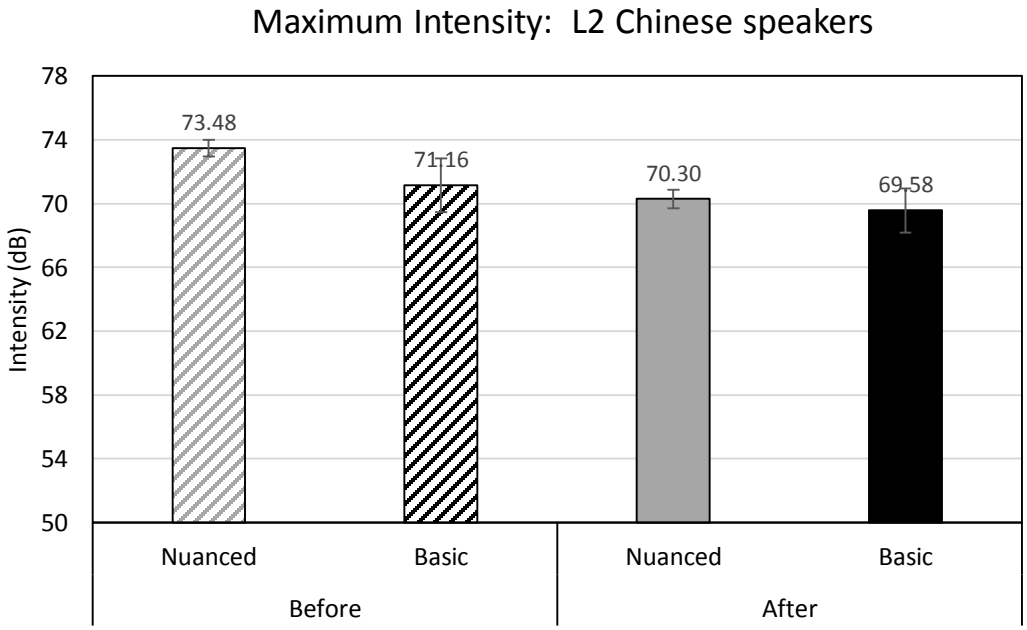


Figure 28: Compliment Category: Maximum intensity of target (final) word for L2 Chinese speakers

It appears that the Chinese speakers distinguished the Nuanced and Basic meaning categories slightly better before training (~2.5 dB difference) than after (<1 dB difference), although in neither training condition is the difference as great as for L1 speakers. Indeed, there is a significant difference before training, $t(48)=4.411$, $p = .000$, but not after training, $t(50)=1.231$, $p = .224$, suggesting that the effect of training goes in the opposite direction from what was expected. This phenomenon will be later commented on in the discussion of this chapter.

5.3.3 Verb Focus Category

Similar to the Compliment category, in this category, a special kind of nuanced focus (Nuanced meaning condition) is compared with broad focus (Basic meaning

condition). An example item from this category is "Isabelle says she's coming to visit", where the main verb "says" can be emphasized to have a more nuanced meaning (e.g. containing an implied meaning of "but we don't believe her") or a more basic meaning (e.g. "We expect that she'll come"). Measurements were performed on the main verb to determine the relative properties between the conditions. For duration, ratios of the main verb to the utterance were measured; for F0, contours over the verb were measured; for intensity, the maximum values of the stressed syllable of the verb were measured. The results of the L1 and L2 data based on these measurements can be seen in the next sections.

5.3.3.1 Duration

Results of duration ratios of the main verb (target word) to utterance are presented in this section for the Nuanced meaning condition and the Basic meaning condition.

5.3.3.1.1 L1 Baseline Subjects

Figure 29 shows the mean ratio of duration of the main verbs (target words) to their corresponding utterances for L1 speakers.

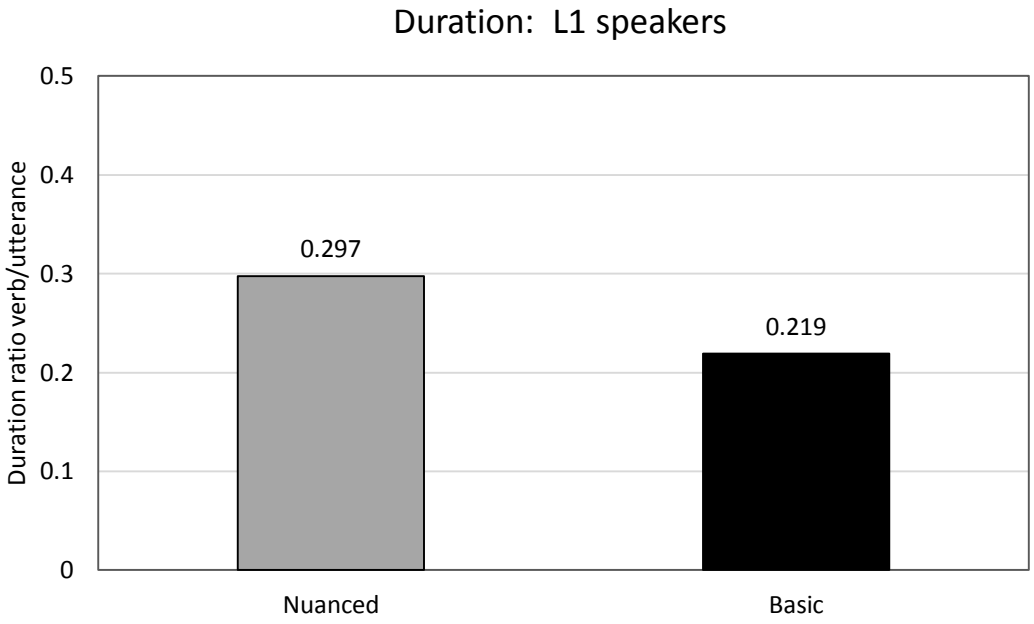


Figure 29: Verb Focus Category: Ratios of duration of verb to utterance for L1 speakers

As expected, the Nuanced meaning condition, in which the verb is focused, has a higher ratio of word to utterance duration than the Basic meaning condition (.078 difference). A two-tailed paired-samples t-test was conducted to compare the duration ratios of the Nuanced meaning condition to that of the Basic meaning condition, and the .078 difference was found to be significant, $t(99)=15.978$, $p = .000$. Hence, L1 speakers use duration to distinguish focus in the Verb Focus category.

5.3.3.1.2 L2 Chinese Subjects

Figure 30 shows the mean ratio of duration of the main verbs (target words) to their corresponding utterances for L2 Chinese speakers.

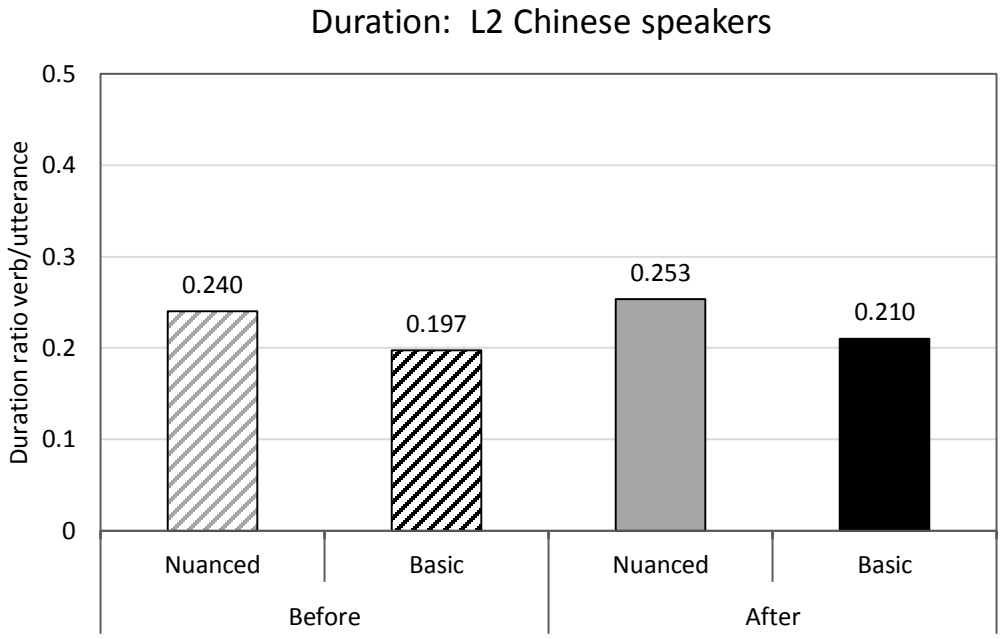


Figure 30: Verb Focus Category: Ratios of duration of verb to utterance for L2 Chinese speakers

As can be observed, the participants show a distinct difference in duration between the Nuanced and Basic meaning conditions both before and after training (with a .043 difference). Indeed, this difference is significant both before training, t (50) = 7.351, p = .000, and after training, t (48) = 7.376, p = .000, suggesting that the Chinese speakers were previously competent with the use of durational differences in this category.

5.3.3.2 F0 Contours

The results sections below on F0 contours will emphasize descriptive results, as it can be problematic to statistically compare contours in a meaningful way. F0 contours are presented on the verb here.

5.3.3.2.1 L1 Baseline Subjects

In terms of F0 contours for the Verb Focus category, there appear to be two patterns in the Nuanced meaning condition for L1 speakers. Figure 31 below shows the mean F0 contours of the verb in the Nuanced and Basic meaning conditions for five speakers. The first pattern shows the canonical pattern for the Nuanced condition: a falling-rising contour.

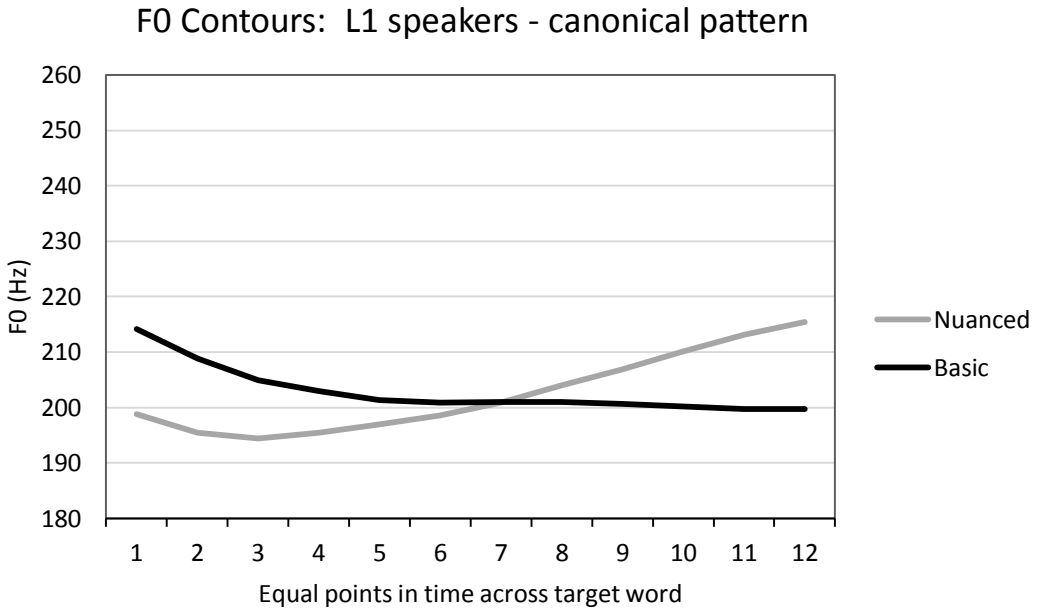


Figure 31: Verb Focus Category: Canonical F0 contours of the verb for five L1 speakers

The Basic meaning contour is simply a falling pattern. The contours cross over each other, suggesting that the shape of the contours matter more than simply higher F0 values for the Nuanced condition.

The remaining six speakers exhibit a pattern distinct from the canonical Nuanced meaning pattern presented above. It appears to be a general rising-falling pattern, as seen in Figure 32.

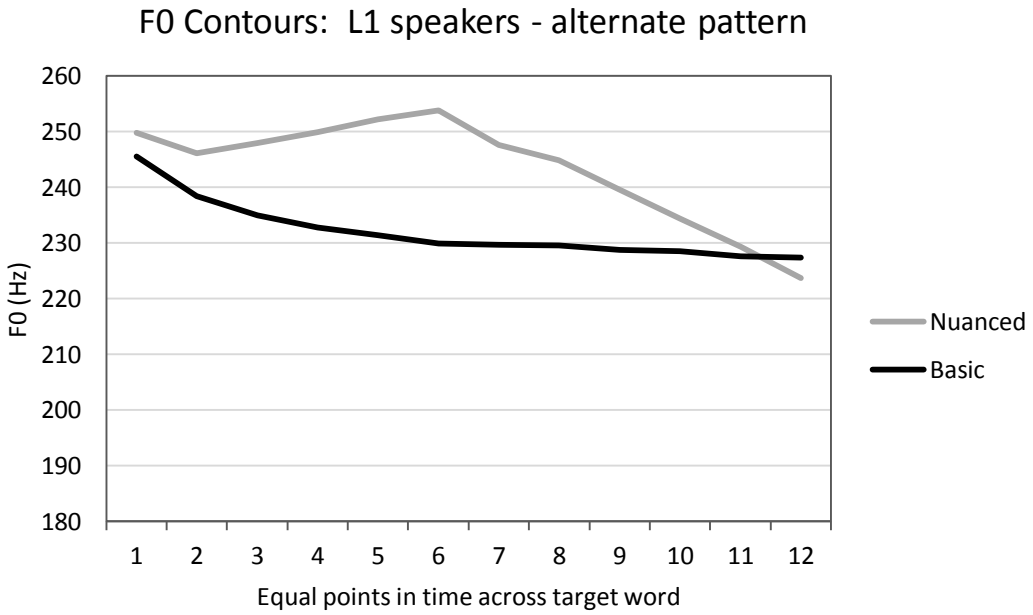


Figure 32: Verb Focus Category: Alternate F0 contours of the verb for six L1 speakers

It is likely that the alternate pattern is representative of some kind of general focus pattern. This will be addressed more in Chapter 6. Only the L1 speaker group that produced the falling-rising pattern will be compared with the L2 subjects' data, since that pattern is what is considered canonical (c.f. Cruttenden, 1985; also, Ward & Hirschberg, 1985), and it is also the pattern utilized in the training; thus, this is the pattern L2 speakers were exposed to in the experiment.

5.3.3.2.2 L2 Chinese Subjects

The F0 contours for the Chinese speakers are shown in Figure 33:

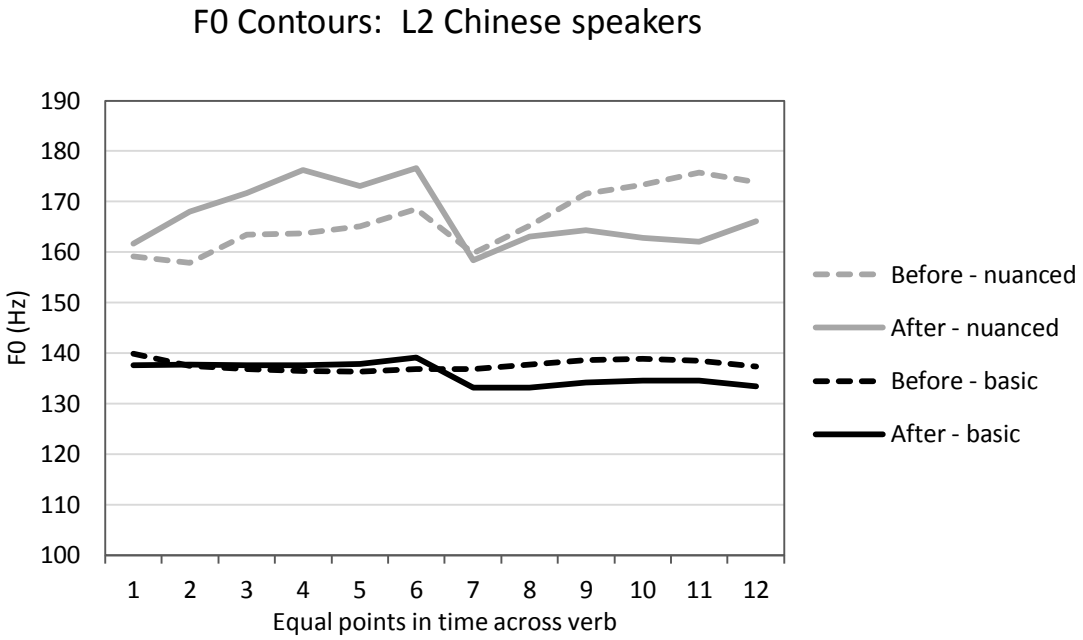


Figure 33: Verb Focus Category: F0 contours of the verb for L2 Chinese speakers

The F0 contours for the Basic meaning condition are not very dissimilar from the L1 speakers, except that they do not have the distinctly clear falling pattern that the L1 speakers exhibited: instead, they are rather flat, both before and after training. The F0 contours of the Nuanced meaning condition before training show a slight rising pattern, with a small dip in the middle of the verb; after training, there is a more substantial rise (than before training), followed by a sharp fall in the middle of the verb, with a fairly flat contour to the end. This is quite different from the canonical pattern used by L1 speakers, who used a falling-rising pattern. Thus, training seemed

to have the effect of making the Chinese speakers' contours more distinct between the conditions, but the canonical F0 pattern used by L1 speakers was not attained.

5.3.3.3 Maximum Intensity

Max intensity was measured on the stressed syllable of the verb, since this was the location focus was likely to be most distinguished in, when comparing the Nuanced and Basic meaning conditions.

5.3.3.3.1 L1 Baseline Subjects

Figure 34 below shows the comparison between the Nuanced and Basic meaning conditions in terms of Max intensity. Error bars represent standard error.

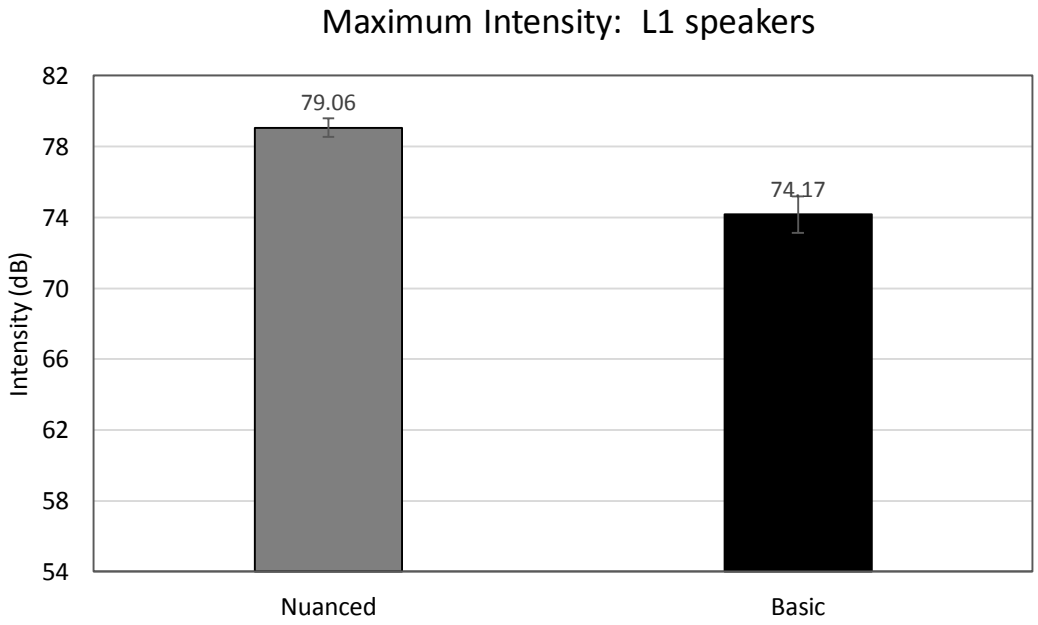


Figure 34: Verb Focus Category: Maximum intensity of the verb for L1 speakers

As expected, in the Nuanced meaning condition, in which the verb is focused, L1 speakers seem to have a higher max intensity than in the Basic meaning condition, with a difference of nearly 5 dB. A one-sample t-test was conducted to compare the z-scores of the Nuanced meaning condition against the mean (taken as z-score of zero) of the Basic meaning condition. The Nuanced meaning condition was found to be significantly different from the mean of zero, $t(99)=8.417$, $p=.000$, thus showing that L1 speakers reliably distinguished the focus conditions through the use of intensity.

5.3.3.3.2 L2 Chinese Subjects

Figure 35 below presents the Max intensity results of the Nuanced and Basic meaning category for Chinese speakers.

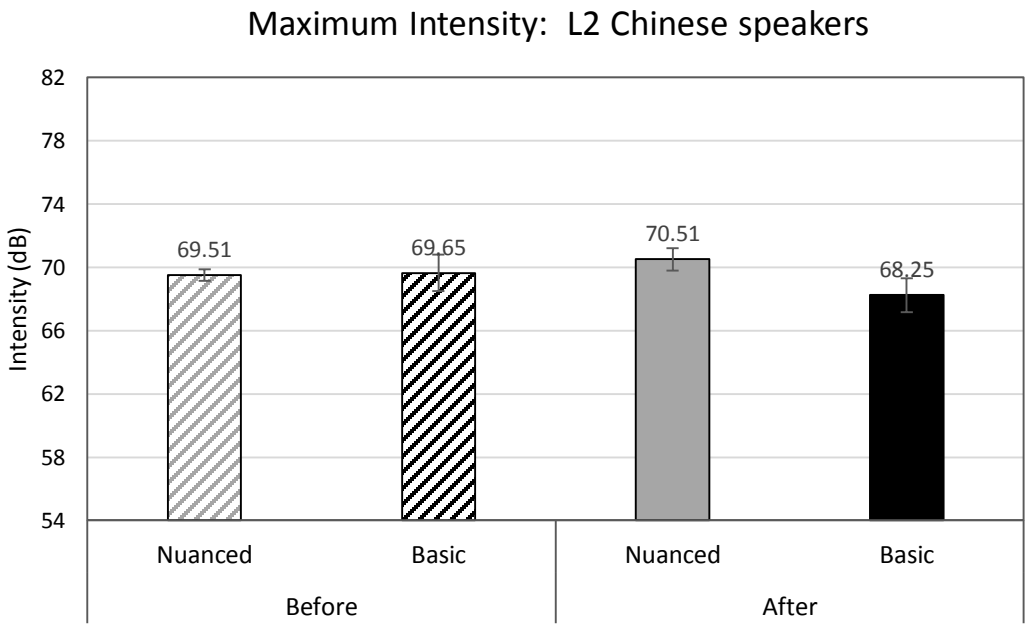


Figure 35: Verb Focus Category: Maximum intensity of the verb for L2 Chinese speakers

For Chinese speakers, while we observe in the graph the lack of differentiation between the Nuanced and Basic meaning conditions before training, statistically speaking, there is a difference between the conditions both before training, $t(50) = 3.517$, $p=.001$, and after training, $t(48)=3.244$, $p=.002$. To make sense of the discrepancy between the graph and t-test results, recall that the values presented in the graph are based on the mean and standard deviation of one speaker. Figure 36 shows the results of the Nuanced meaning condition in terms of z-scores (upon which the statistics are based) calculated from the aggregate of speakers.

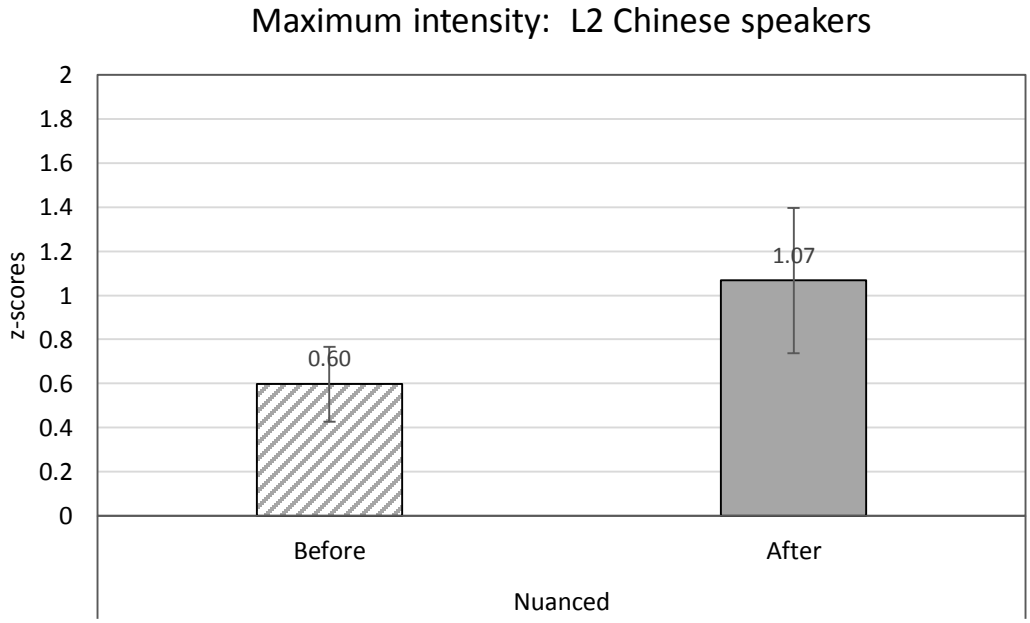


Figure 36: Verb Focus Category: Maximum intensity of Nuanced meaning condition based on z-scores for L2 Chinese speakers

This graph illustrates more accurately that the intensity for the Nuanced meaning condition was indeed higher than the Basic meaning condition (not shown

here, as its mean is taken as zero) both before and after training. Hence, Chinese speakers had mastered the L1 pattern of intensity in the Verb Focus category even before receiving training.

5.3.4 Preliminary Results of L2 Arabic Speakers

This final results section provides preliminary data from Arabic speakers on all three prosodic categories. As indicated above, results here are descriptive rather than inferential, as they are based on fewer speakers than the other groups; hence, results should be considered with more caution here. Nevertheless, with this data, we can hopefully begin to gain some insight into their patterns.

5.3.4.1 Contrastive Focus Category: Duration

Figure 37 below presents the duration patterns for L2 Arabic speakers before and after training. The values presented here are duration ratios of the target word (adjective or noun) to phrase (adjective+noun) for the Adjective Focus (AF) condition and Noun Focus (NF) condition.

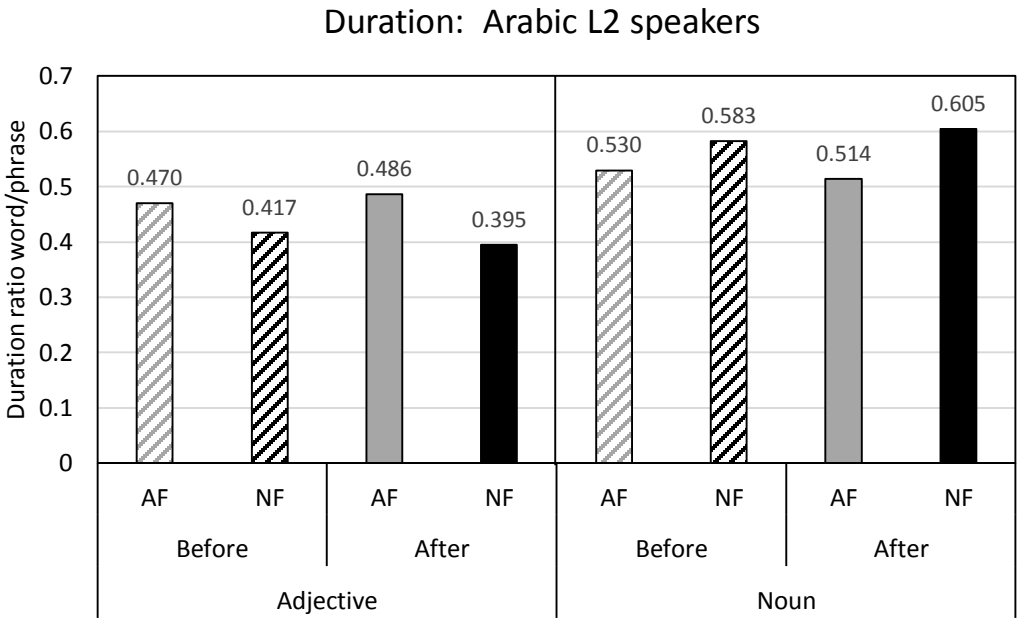


Figure 37: Contrastive Focus Category: duration ratios of adjective and noun to phrase for L2 Arabic speakers. AF=Adjective Focus; NF=Noun Focus.

The Arabic speakers exhibit nearly the same pattern, both before and after training, as that of L1 speakers: adjectives appear longer when focused (AF condition) than not focused (NF condition), nouns appear longer when focused (NF condition) than not focused (AF condition), and nouns appear longer than adjectives in the NF condition. In all three of these cases, the duration ratio difference appears to expand from before training to after training: the ratio differences increase from .053 to .091 in the comparison of AF to NF conditions for both adjectives and nouns, and an increase of .166 to .21 ratio difference is found in the comparison of adjectives to nouns for the NF condition. This suggests that the training may have aided the Arabic speakers in distinguishing the conditions more than they were previously. The only pattern that differs between L1 speakers and Arabic speakers lies in the AF condition:

while for L1 speakers, there was effectively no difference in length between adjectives and nouns, for Arabic speakers, nouns appear slightly longer than adjectives before training. This .06 difference, however, subsides after training to only .028.

Thus, overall, Arabic speakers appear to already have been competent with the duration property in the Contrastive Focus category, but training aided in enlarging distinctions to more closely match the L1 pattern and reducing distinctions that were unlike the L1 pattern.

5.3.4.2 Contrastive Focus Category: Maximum F0

The results of Maximum F0 in target adjectives and nouns in the Adjective Focus and Noun Focus conditions are presented for L2 Arabic speakers in the following graph, Figure 38.

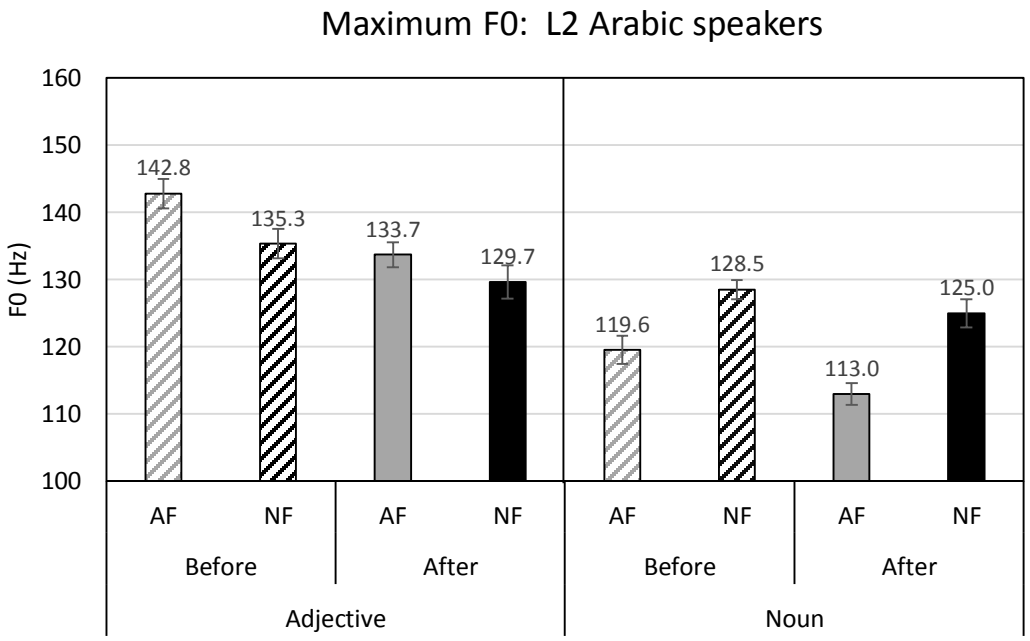


Figure 38: Contrastive Focus Category: Maximum F0 values for L2 Arabic speakers.

For the most part, the pattern of the Arabic speakers seems to resemble that of the L1 speakers: adjectives have higher F0 in the AF condition than the NF condition, nouns have higher F0 in the NF condition than AF condition, and adjectives are higher in F0 than nouns in the AF condition.

However, unlike the L1 pattern, adjectives have higher F0 than nouns in the NF condition. Additionally, the F0 difference between the AF and NF conditions for adjectives is much greater for L1 speakers (~25 Hz difference) than for the Arabic speakers (~7-8 Hz difference), and there does not appear to be an improvement following training. This decreased distinction in F0 for Arabic speakers (as compared with L1 speakers) can at least in part be attributed to the conversion from z-scores being based on a male speaker, rather than a female (as was done for L1 speakers),

since it has been established that females generally use a wider F0 range than males (cf. Chevrie-Muller et al. 1967, Takefuta et al. 1972, Chen 1974, Johns-Lewis 1986). For nouns, the F0 differences between the AF and NF conditions are (slightly) greater, as they also are with L1 speakers (~9 Hz difference for Arabic speakers as compared with ~35 Hz difference for L1 speakers). Here, there appears to be a minor expansion of the F0 difference for Arabic speakers after training (to ~12 Hz difference).

In sum, while Arabic speakers did use F0 (in a similar way to L1 speakers in most cases) to make distinctions in the Contrastive Focus category, the differences appear not quite as great as compared with L1 speakers, and one of the patterns is the reverse of the L1 speakers' pattern.

5.3.4.3 Contrastive Focus Category: Maximum Intensity

Maximum intensity of the stressed syllable in the target adjectives and nouns in different focus conditions is presented in Figure 39.

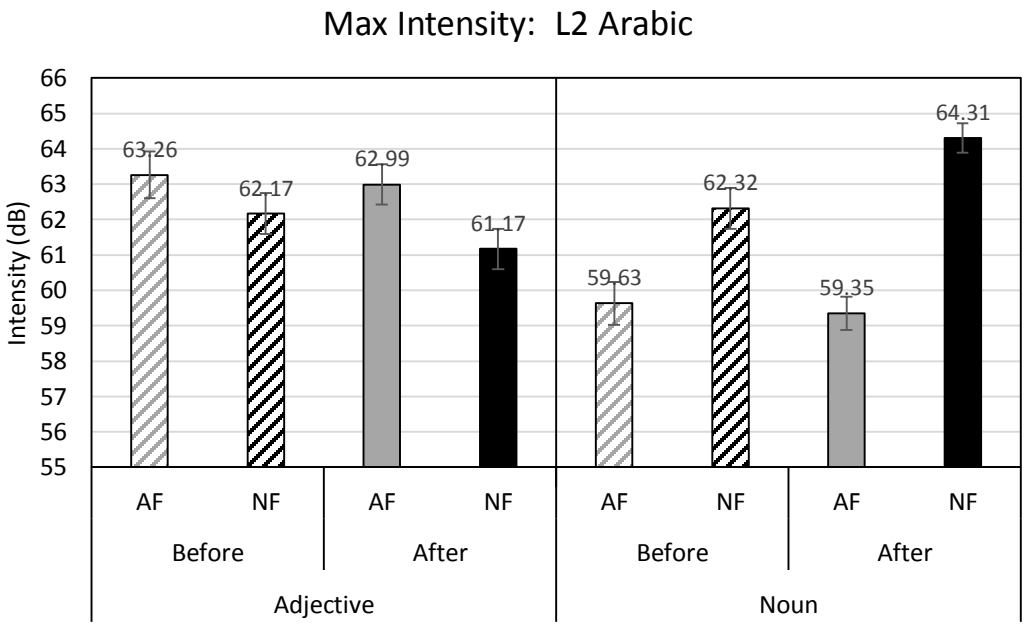


Figure 39: Contrastive Focus Category: Maximum intensity values for L2 Arabic speakers. AF=Adjective Focus; NF=Noun Focus

L2 Arabic speakers show a similar pattern for intensity to L1 speakers, albeit with different ranges being utilized. While the patterns between the speaker groups are similar, we can observe that for L2 Arabic speakers, adjectives are not heavily differentiated with intensity when comparing the AF condition to the NF condition, especially before training occurs (only ~1 dB difference). The nouns appear to be more differentiated between the focus conditions than the adjectives are (similar to L1 speakers), and for both the adjectives and nouns, the intensity differences increase after training, approaching the distinctions of L1 speakers. Ultimately, Arabic speakers used intensity in very similar ways to L1 speakers, and the training seemed to aid in improving the distinctions between conditions.

5.3.4.4 Compliment Category: Duration

The duration results in the Compliment category for Arabic speakers are shown in Figure 40 below, before and after training. The values presented here are duration ratios of the target (final) word to utterance for the Nuanced and Basic meaning conditions.

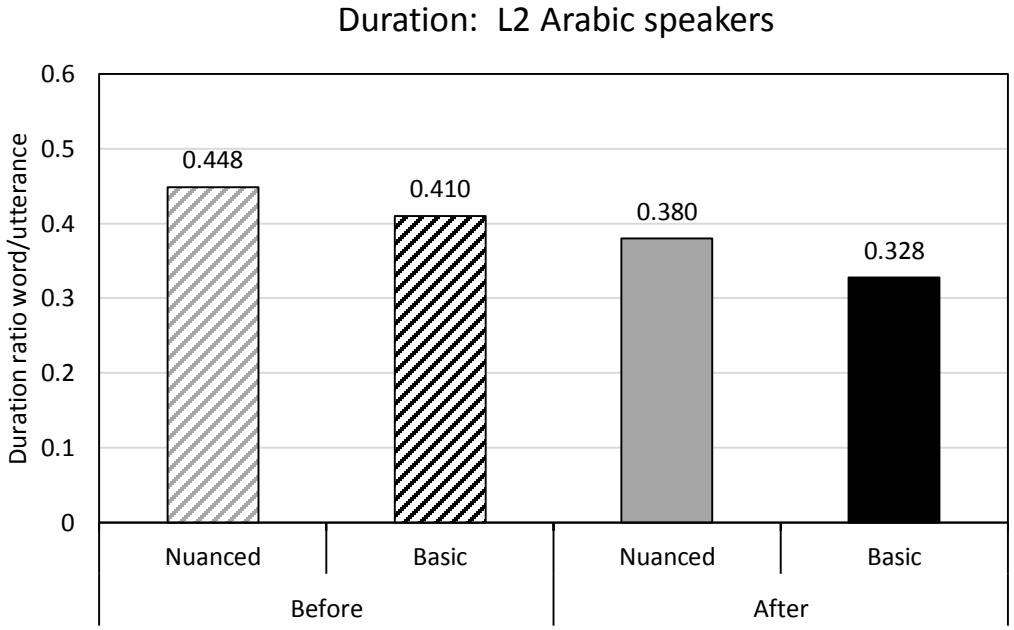


Figure 40: Compliment Category: Ratios of target (final) word to utterance for L2 Arabic speakers

Both before and after training, the differences between the Nuanced and Basic meaning conditions appear very similar to the L1 pattern. However, there does seem to be an expansion of the difference from before training (.038) to after training (.052), suggesting that training aided the Arabic speakers in making a stronger distinction between the focus conditions, to more closely match the L1 pattern.

5.3.4.5 Compliment Category: F0 Contours

The Arabic speakers' F0 contours for the Compliment category are shown below in Figure 41. As stated earlier regarding the Arabic data, note that the range is slightly different, as these values are based upon a male speaker.

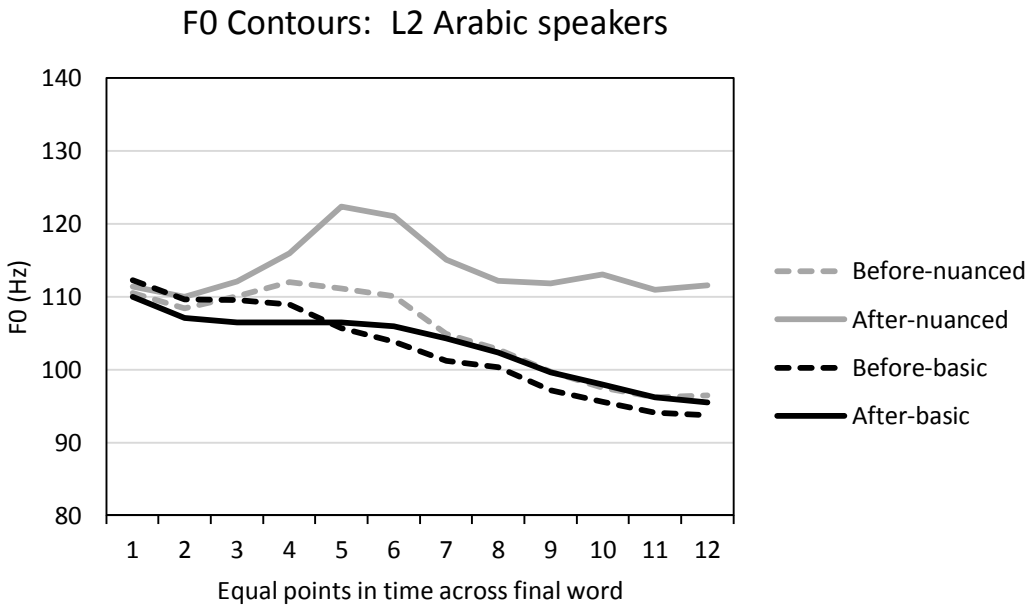


Figure 41: Compliment Category: F0 contours of target (final) word for L2 Arabic speakers

Arabic speakers showed essentially no distinction between the Nuanced and Basic Meaning conditions before training: the Nuanced and Basic meaning F0 contours are both falling, with roughly the same F0 values. Following training, the Nuanced meaning F0 contour did change, exhibiting a very slight rising-falling pattern, with F0 values more distinct from the Basic meaning contour. However, the final boundary rise at the end is crucially lacking. Hence, while training seemed to improve the

Arabic speakers' distinctions of the conditions in terms of F0, they do not produce the same F0 patterns as L1 speakers.

5.3.4.6 Compliment Category: Maximum Intensity

The maximum intensity results in the Compliment category for Arabic speakers are shown in Figure 42 below, before and after training. The values presented here are maximum intensity of the stressed syllable in the target (final) word in the Nuanced and Basic meaning conditions.

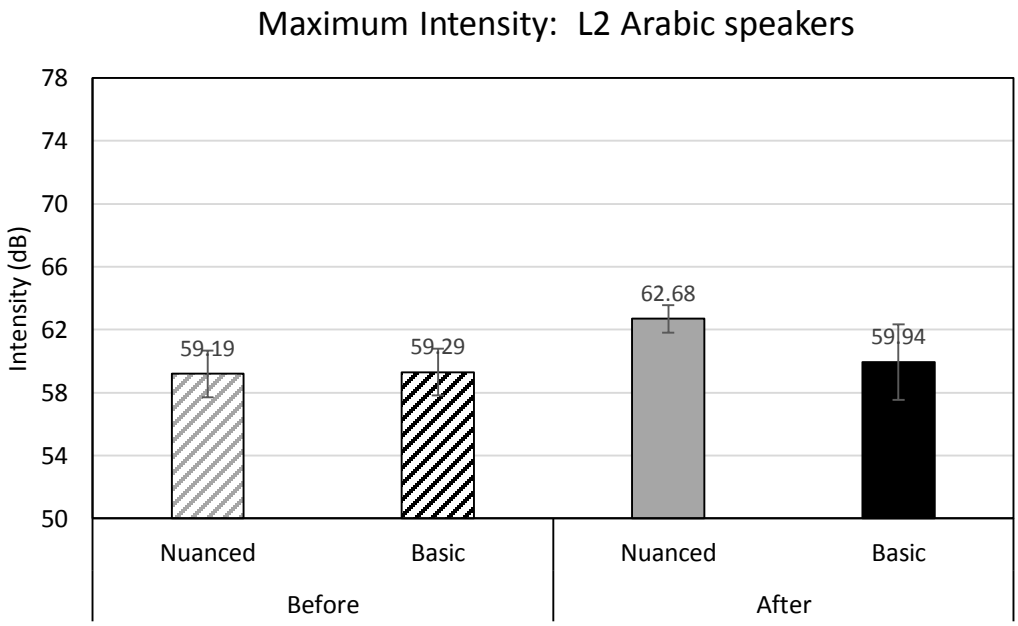


Figure 42: Compliment Category: Maximum intensity of target (final) word for L2 Arabic speakers

There appears to be no distinction for Arabic speakers between the Nuanced and Basic meaning conditions before training, while L1 speakers distinguished the conditions by

~4 dB. However, after training, there is nearly a 3 dB distinction, suggesting that the training helped Arabic speakers attain the L1 pattern.

5.3.4.7 Verb Focus Category: Duration

The duration results for Arabic speakers in the Verb Focus category are presented in Figure 43 below. The duration values are ratios of the target word (verb) to the utterance.

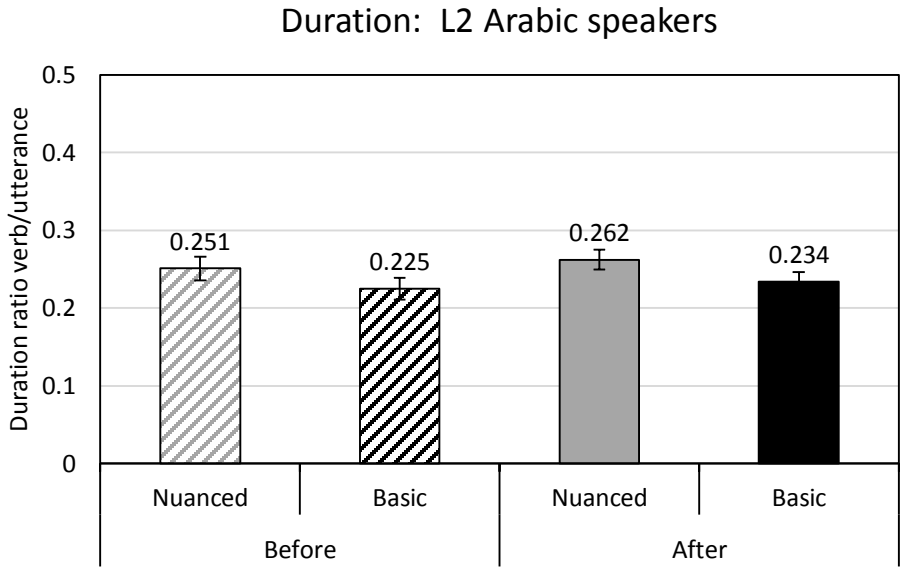


Figure 43: Verb Focus Category: Ratios of duration of verb to utterance for L2 Arabic speakers

Arabic speakers do show a durational difference both before and after training in the same direction as L1 speakers. However, the differences are minimal for Arabic speakers (.026 before training and .028 after training) as compared with L1 speakers

(.078), suggesting that duration may not have played a large role in their strategy for focus differentiation, and the training was not particularly effective in this area.

5.3.4.8 Verb Focus Category: F0 Contours

The Arabic speakers' F0 contours for the Verb Focus category are shown in Figure 44 below. Recall that the range is slightly different from the L1 speakers', as these values are based upon a male speaker, as opposed to a female.

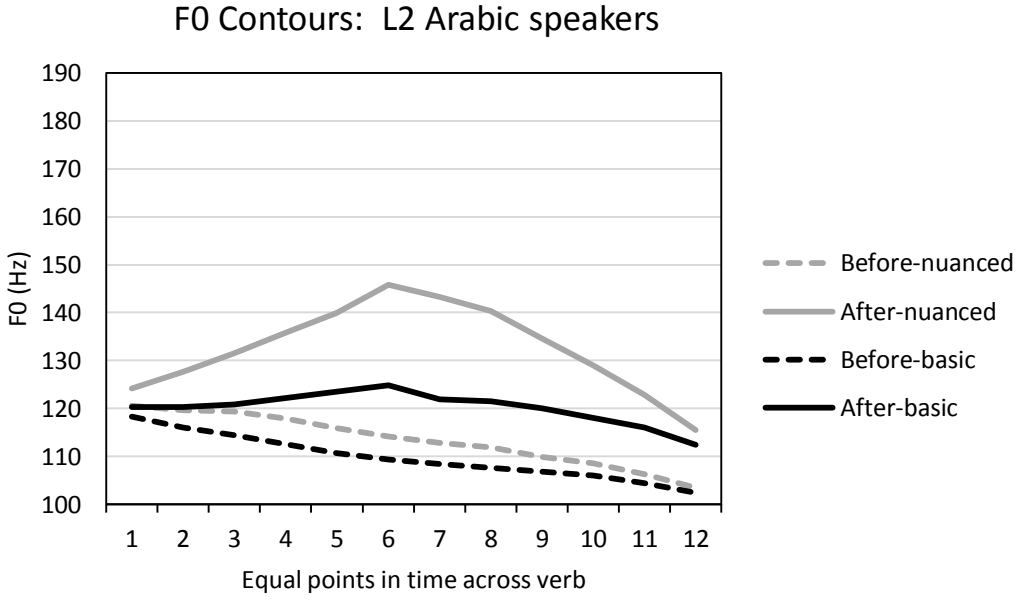


Figure 44: Verb Focus Category: F0 contours of the verb for L2 Arabic speakers

Before training, both the Nuanced and Verb Focus contours were falling and exhibited very similar F0 values throughout. After training, while the Basic meaning pattern retained a general falling contour, the Nuanced meaning contour was a rising-falling pattern. Moreover, after training, the Nuanced meaning contour generally

showed substantially higher F0 than in the Basic meaning condition (20 Hz higher at the location of greatest difference). However, the Nuanced meaning pattern does not bear a resemblance to the L1 canonical pattern. Instead, it resembles the alternate pattern of L1 speakers, even though this was not the pattern used in the training. Thoughts about this phenomenon will be raised in the Discussion chapter.

5.3.4.9 Verb Focus Category: Maximum Intensity

Finally, Figure 45 below presents the maximum intensity patterns for Arabic speakers. The intensity values were measured on the stressed syllable of the target word (verb).

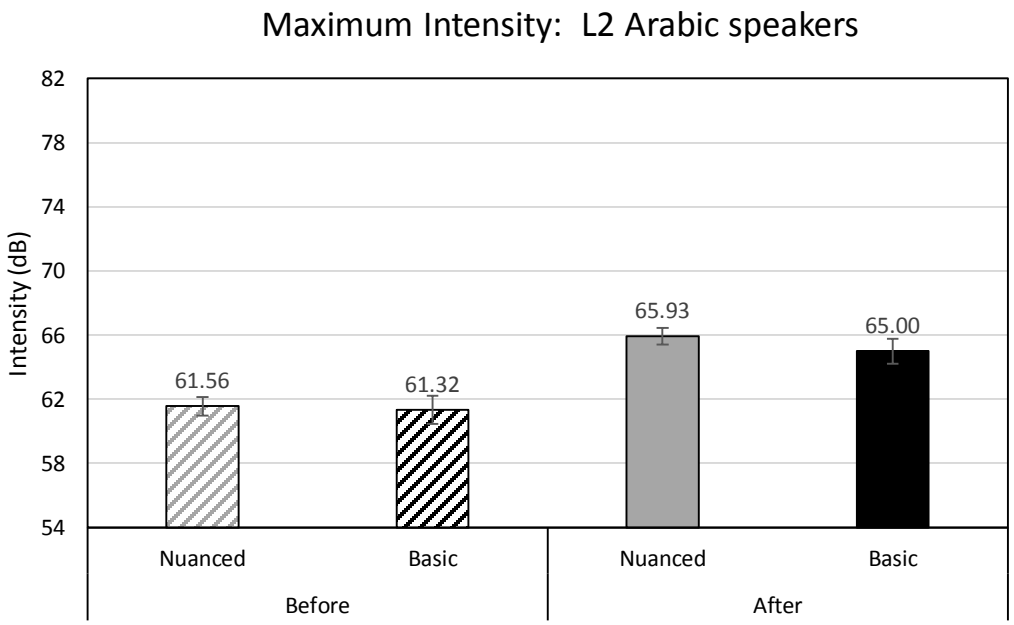


Figure 45: Verb Focus Category: Maximum intensity of the verb for L2 Arabic speakers

For max intensity, there was virtually no difference between the Nuanced and Basic meaning conditions before training. After training, a slight distinction of nearly 1 dB was made, but as previously mentioned, 1 dB is the smallest difference that humans can detect. Hence, Arabic speakers were likely ineffective in producing intensity differences between the conditions in the manner that L1 speakers showed.

5.4 Summary and Discussion of Results

In general, the L2 Chinese speakers exhibited fairly close approximations to the patterns of native English speakers for Contrastive Focus. Duration patterns were nearly identical for L1 speakers and Chinese speakers (with one small exception). F0 patterns were also very similar, and in the one case where the Chinese pattern differed from the L1 pattern, this was rectified through training. Additionally, Chinese speakers used intensity for Contrastive Focus in the same manner as L1 speakers, although with less differentiation between conditions. The training aided in expanding that distinction closer to that of L1 speakers.

In the Compliment category, Chinese speakers exhibited similar durational patterns to L1 speakers, both before and after training. However, before training, the level of differentiation between conditions was much lower for Chinese speakers than for L1 speakers; training increased the level of differentiation, closer to that of L1 speakers. In terms of F0, we observed that Chinese speakers made substantial improvement after training, with the Nuanced meaning F0 contour resembling that of native speakers, and the Nuanced and Basic meaning contours becoming more distinct from one another. For intensity, training had an effect on the Chinese speakers here, but the trend went in the opposite direction, suggesting that the Chinese speakers had not quite mastered the use of intensity in this category.

In the Verb Focus category, Chinese speakers produced duration and intensity differences between the conditions both before and after training, but the training helped in widening those differences, towards the extent of distinction made by L1 speakers. For F0, both before and after training, the Chinese speakers' Nuanced meaning patterns diverged from the canonical contour of L1 speakers. The training did not seem to help here in the intended way; nonetheless, as with duration and intensity, it did seem to have an effect of differentiating the Nuanced from the Basic meaning conditions more so than before training.

Arabic speakers generally demonstrated similar patterns to L1 speakers in Contrastive Focus, particularly with regards to duration and intensity. For these measurements, training had a positive effect of expanding duration and intensity differences between conditions. For F0, most comparisons showed similar patterns, but were not differentiated to the extent of L1 speakers; one comparison showed the reverse of the L1 speakers' pattern; training helped in some cases, but not all. In the Compliment category, Arabic speakers appeared to make clear distinctions with duration, even before training occurred. For F0, they demonstrated mild improvement with the Nuanced meaning contour after training, but lacked the crucial rising boundary tone. As with the Contrastive Focus category, training also had a positive effect on intensity differences in the Compliment category. In the Verb Focus category, Arabic speakers ultimately used the same duration and intensity patterns as L1 speakers, but not nearly to the same extent. While training improved the distinction for intensity, it would remain an undetectable difference to the human ear, thus an essentially ineffective use of the property. For F0, the target patterns of L1

speakers were not quite attained; nevertheless, training did seem to have an effect of differentiating the conditions, just not in the manner intended.

Overall, it would seem that Chinese speakers were rather successful in achieving L1 speakers' patterns, and where they faltered, this was usually rectified with training, with the exception of the F0 pattern in the Verb Focus category and the intensity pattern in the Compliment category. Both Chinese and Arabic speakers were more successful with the Contrastive Focus category than the other prosodic categories, as expected. Arabic speakers appeared to attain the duration patterns of L1 speakers slightly better than F0 and intensity, but some large gaps remained between their patterns and L1 patterns. However, more data from Arabic speakers would be crucial in determining the relationship between the patterns of Arabic speakers and L1 speakers.

Chapter 6

DISCUSSION AND IMPLICATIONS FOR L2 PEDAGOGY

The purpose of the perception and production experiments in this dissertation was primarily to shed light on how L2 speakers (with a concentration on Chinese speakers) understand and produce certain English prosodic patterns, how this differed from L1 patterns, and to what extent a targeted training would improve the performance of L2 speakers. This chapter will discuss the general findings from Chapters 4 and 5 with respect to one another and training (section 6.1), as well as addressing some questions and phenomena regarding L2 production of these patterns (section 6.2**Error! Reference source not found.**). Finally, section 6.3 will present some implications for L2 pedagogy.

6.1 Perception and Production Performance

Phonological perception and production performance can be difficult to compare, given their different formats: perception performance is certainly simpler to evaluate given that results are based on a categorical scale, as opposed to the combination of several continuous variables (e.g. duration, F0, intensity) for production. Nevertheless, general evaluations can be made with respect to one another.

Overall, in both perception and production, the L2 speakers were closer to L1 performance in the Contrastive Focus category than in the Compliment and Verb Focus categories; this was as expected, due to the former being a more familiar

pattern. For the latter two categories, the Basic meaning condition was nearly always perceived and produced more accurately than the Nuanced meaning condition (with the exception of in the Verb Focus category before training, where they were equivalent). Again, this was predicted since they had likely not been exposed to the latter condition, but would be very familiar with the former.

Perception generally precedes production in L2 learning, and production tends to only follow when there is sufficient exposure to the feature in question (Celce-Murcia et al. 1996). Indeed, in most cases in the current study, overall perception results (either before or after training) were closer to L1 patterns than the corresponding production results.

Notwithstanding, before training occurred, it could be argued that some production results were more similar to L1 patterns than the corresponding results for perception. Since L2 participants had a propensity towards choosing the Adjective Focus response in the perception experiment, it is reasonable to suppose that they would not consistently distinguish Adjective Focus from Noun Focus in terms of acoustic properties, or that they might utilize a different strategy for doing so. However, the results showed that L1 and L2 speakers produced clear distinctions between the conditions in a very similar manner. This could partly be due to a different setup between the two experiments: in the perception experiment, participants received one version of each stimulus (e.g. either the Adjective Focus or the Noun Focus version of the stimulus), whereas in the production experiment, they were prompted to record both versions of each sentence (so that proper acoustic comparisons could be made). While the two versions of the same stimulus were never adjacent to each other, simply seeing both versions during the experiment could make

them aware of a necessary distinction to be made. Nevertheless, the fact that L2 speakers used very similar strategies to L1 speakers for making those distinctions suggests that they were previously familiar with how to manifest focus in the Contrastive Focus category. This is one area where the L2 production results were, at least before training, stronger than perception results, with respect to L1 patterns. Several studies have shown that production can precede perception in L2 acquisition (e.g. Sheldon & Strange 1982, Yamada et al. 1994, Strange 1995), and these results lend support to that claim.

6.1.1 Effect of Targeted Training

While some L2 behaviors were already similar to those of the L1 speakers, where there was room for improvement, training boosted perception performance more than production.

In the perception experiment before training, performance for Chinese speakers was highest in the Contrastive Focus category, but with a strong bias towards choosing the Adjective Focus response. After training, the accuracy in the Noun Focus condition rose significantly, and the bias lessened, becoming more similar to the L1 patterns. For the Compliment category, accuracy was relatively high in the Basic meaning condition, but not in the Nuanced meaning condition before training. After training, the accuracy levels rose to similarly high levels for the two conditions, and as a result the bias towards choosing the Basic meaning was essentially eliminated. Even though there remained a significant difference between the accuracy levels of the Chinese speakers and the L1 speakers for the Nuanced meaning, it is evident that training dramatically boosted the results towards the patterns of L1 speakers. In the Verb Focus category before training, Chinese speakers demonstrated equally poor

understanding between the Nuanced and Basic meaning conditions, but accuracy rates again rose dramatically after training to a roughly equal level, matching the L1 pattern for the Basic meaning pattern. While there remained a gap between L1 and L2 speakers for the Nuanced meaning condition after training, the large improvement indicates that the training was highly effective.

In the production experiment, L2 speakers exhibited many of the same patterns, but often with less clear distinction than L1 speakers. For example, both groups may have produced longer final nouns in the Compliment category when they were in the Nuanced meaning condition as opposed to the Basic meaning condition, but the difference between conditions was much larger for L1 than L2 speakers. Thus, typically when there was room for improvement, the effect of training was in expanding those distinctions. While L2 speakers did not often achieve the same degree of distinction as native speakers, the improvement was often substantial. There were only a couple of instances where training seemed to have an effect of reverting to a different pattern from L1 speakers. In these cases, any distinction between conditions beforehand was minimal; thus, it is likely that the L2 speakers did not have a true mastery of the properties there even before training. Regarding the effect of training on F0 contours, it was shown to be highly effective in the Compliment category for Chinese speakers. Not only was there an effect of making the F0 values more distinct between conditions, but the shape of the contours after training became nearly identical to those of L1 speakers. The same effect was not quite achieved for the Verb Focus category. While there was an effect of expansion of the contours for L2 speakers, the shape of the Nuanced meaning contours for L2 speakers remained dissimilar from the L1 speakers'.

Jenkins (2004) points out that the acquisition of discourse intonation likely requires more exposure than for segmental pronunciation. Since the minimal training provided yielded as much success as it did, it is likely that given more extensive training and exposure to the prosodic categories, even more improvement would be observed in the L2 speakers' production.

6.2 Remarks on Production-related Phenomena

In this section, a discussion of the role of L1 properties with respect to the L2 is presented, along with some remarks on the potential emergence of a default pattern for focus.

6.2.1 Role of Properties of L1

One of the research questions raised in this dissertation was regarding the role of acoustic properties in an L1 when acquiring L2 prosody: specifically, whether the presence of a particular contrast in an L1 would make it more accessible to be used in prosody of an L2. Hence, it was proposed that speakers of Chinese may have more success manipulating pitch in an L2 than speakers of languages without a lexical pitch (or tone) contrast. Moreover, it was expected that speakers of Chinese may be more successful with attaining L1 pitch patterns than other properties, such as duration and intensity. Similarly, would speakers of Arabic, a language with contrastive vowel length, have more success in manipulating duration (than other properties) to L1 patterns, and would they be more successful than speakers of languages without a contrastive length contrast?

It was found that Chinese speakers ultimately attained the F0 patterns of L1 speakers for two prosodic categories: Contrastive Focus and Compliment. While their

success on the Contrastive Focus category could be attributed to a previous familiarity with the pattern, this was not the case for the Compliment category, where they showed dramatic improvement from before training to after training. Hence, it is quite possible that a sensitivity to pitch from their L1 contributed to their attainment of these patterns. Unexpectedly, Chinese speakers were also successful at using duration contrasts, suggesting that this is a property that Chinese speakers are also quite sensitive to. A potential explanation is that in Chinese, while tones are typically thought of as being distinguished by F0, there are other cues involved. Certain Chinese tones (tones 2 and 3) exhibit longer duration than others (Dreher & Lee, 1966; Chuang et al., 1972; Howie, 1976; Nordenhake and Svantesson, 1983 and others). Blicher et al. (1990) even showed that durational differences between Tones 2 and 3 in Chinese are perceptually important in that stimuli with ambiguities between Tones 2 and 3 were identified more often as Tone 3 when there was additional lengthening.[8] Thus, this could explain the sensitivity of Chinese speakers to durational usage in English. Chinese speakers were moderately successful at attaining the intensity distinctions used by L1 speakers; since intensity is typically considered a less crucial cue because languages do not use it as a main cue for lexical contrasts

Arabic speakers fared better at attaining duration patterns than F0 and intensity patterns. This observation is not surprising for Arabic speakers, given their contrastive use of duration at the word level. While Arabic speakers showed F0 contrasts between conditions in the Contrastive Focus category, they were substantially less extensive compared with L1 speakers. They also demonstrated some

---

[8] c.f. Jongman et al. (2006) for a more complete discussion of the perception and production of Mandarin tones.

improvement in the F0 patterns of the Compliment category, though the improvement was less dramatic than for Chinese speakers, and the final boundary tone rise was crucially missing. While they wee able to use F0 to a certain degree to make prosodic contrasts, they did not attain the level of L1 speakers. Thus, Arabic speakers may indeed be less sensitive to F0 differences due to its lack of distinctive use at the word level.

While the data here provide support from only two languages, a tentative argument towards the relationship between L1 lexical acoustic contrasts and attainability of like properties in L2 prosody can be established. More data from other L1 and L2 languages would be crucial in determining the extent of this relationship.

6.2.2 Emergence of a Default Focus Pattern?

Regarding the F0 patterns in the Verb Focus category, both groups of L2 speakers were unsuccessful at producing the Nuanced meaning contours, before or after training. Interestingly, only half of the L1 speakers produced the canonical falling-rising contour; the others produced a rising-falling contour. The L2 speakers also appeared to produce something similar to the latter pattern after training, even though it was not the contour taught in the training. For this reason, I propose that some kind of default focus pattern could exist: rising-falling or simply a high pitch accent, such as that used in Contrastive Focus. This would explain why the L2 speakers exhibited a pattern not taught to them during training that is similar to a group of L1 speakers, and why they only produced this pattern clearly following training, after they had been made more aware of the presence of focus.

This default pattern may have also emerged in the Compliment category for Arabic speakers. While the F0 contours for the Verb Focus and Compliment

categories did not appear identical, they did share a prominent peak. This could be interpreted as the same contour, with the timing of the fall shifted based on the target contour of one category (Compliment) containing a sentential boundary tone, and the other category (Verb Focus) not.

What prompted the use of a default focus pattern for Chinese speakers in one category (Verb Focus), but not in another (Compliment), where both categories were likely unfamiliar to them before training? The Verb Focus category could simply be that much harder to interpret (lending support to a default focus), as was previously discussed regarding the perception results. An additional possibility is that it is more difficult to learn to produce a contour that occurs utterance-medially because connection from the target contour to preceding and following contours inside the utterance must also be learned; for a target contour at an utterance boundary, they must only connect it to the rest of the utterance on one end. In fact, results from Jun (2000) lend support to this hypothesis: English high-level learners of Korean were found to be better at producing phrase-final tones that mark a phrase boundary than they were at other surface tones in a phrase. Hence, a default pattern may emerge with a more difficult to interpret prosodic meaning, and/or as a result of being at a phrase-medial, as opposed to phrase-final location.

6.3 Implications for L2 Pedagogy

As has been made clear by this point, intonation (or prosody) is an area that is taught very little to L2 speakers, especially outside of English-speaking countries. An informal survey that I conducted among my own (Chinese) students at the English Language Institute of the University of Delaware, suggests that pronunciation in

general largely goes untaught in China, and when it is taught, no differentiation is delineated between any kind of English dialect.

Yong & Campbell (1995) suggest that although others have claimed the main function of English in China is international communication, that it must be viewed more broadly. Even so, achieving successful international communication requires competent use of oral English, especially with regards to pronunciation. As an ever increasing number of students come to the United States from China to pursue advanced degrees, it is clear that they are searching for more advanced knowledge of the English language. Even back in 1985, Wu showed that English was used to achieve certain sociolinguistic and linguistic effects among Chinese English users. Competence with the types of intonation patterns investigated in this dissertation could certainly apply to these desired effects.

6.3.1 What does Intonation Instruction Consist of in Current L2 Texts?

While there are a variety of pronunciation texts available to second language learners of English, "Clear Speech" (Gilbert, 2012), "Focus on Pronunciation 3" (Lane 2013, and "Well Said" (Grant, 2010) appear to be among the widely used texts intended for more advanced levels (including at the English Language Institute where the studies of this dissertation were carried out). These texts contain a varying amount of instruction on intonation. The following content on intonation is generally found in these texts:

1. Thought groups/pausing/chunking: how to orally divide up sentences into meaningful/useful phrases
2. Using intonation to indicate old/new information
3. Locating a focus word, including various possible types of focus words

4. Yes/no and wh-question intonation; basic declarative intonation
5. List intonation
6. Contrastive Focus
7. Parentheticals

Boundary tones are discussed (with alternate terms) with regards to falling or rising tones, usually related to questions, list intonation, and declaratives. Visual aids regarding intonation, rhythm, and stress usually include selected general contours (often only for boundary tones), dots above stressed words, and variations in typography, including some combination of bolding, italicization, capitalization, and spacing. "Well Said" even introduces pictorial aides (bicycles, cars, and buses) to help with explaining length differences in words. Auditory aids are present usually for exercises (that do not provide answers), rather than introductory explanations.

6.3.2 What is Missing from these Texts?

While each text incorporates valuable instruction on intonation, there are some enhancements that could be made for maximum effectiveness. For example, "Focus on Pronunciation" provides contours to go along with examples of intonation types, but no auditory aids; their exercises introduce audio, but no answers are given in the exercises. In addition, explanation of how to focus words and/or which words to focus is not provided. While "Well Said" includes various approaches (as listed above) to improve prosody, its aim is to more generally improve communication, and therefore lacks enough specific examples and variety of intonation use, to be maximally effective. "Clear Speech" purports to emphasize stress, rhythm and intonation, and while certainly containing more detailed instruction in this area than other texts, an explanation of how to emphasize, besides providing very general contours and audio

of simple rises or falls, is lacking. Moreover, it is in need of a more cohesive structure and sufficient variety of intonation patterns by which to master the subject.

Hence, there are three aspects especially lacking in pronunciation texts regarding intonation: explanation of *how* to focus words, the inclusion of pitch-accented tunes besides boundary tones, and sufficient auditory-visual examples to support the explanations.

6.3.3 Potential Solutions

The three-pronged training introduced in this dissertation was shown to be highly successful in terms of perception, and moderately successful in terms of production on high-level learners of English. In this training, participants received a very limited amount of exposure to the various intonation patterns introduced; however, I propose that introducing an increased amount of exposure perhaps over repeated occasions would have a quick positive effect on production. This type of training could be integrated into a classroom curriculum, used in a tutoring environment, or developed into a self-guided computerized format.

Much of the reason why teaching of intonation is often avoided in classroom settings seems to be due to uncertainty on how best to incorporate it into lessons, since there is so much variation involved. However, similar to learning other aspects of pronunciation (segmental or stress), there are certain fundamental concepts that can be explained in the classroom. The three-pronged approach outlined in Chapter 3, involving auditory, visual, and explicit verbal description of tunes (and their meanings) could be an efficient use of classroom time for teaching intonation.

However, studying pronunciation is, more than any other aspect of learning a language, ideally suited for one-on-one teaching situations, where a student can

receive direct immediate feedback on his/her individual needs. Some institutions offer this kind of supplemental learning. Training teachers and tutors in the three-pronged approach (as well as in the meanings and uses of these tunes) would be of utmost importance in ensuring the adequate instruction and reinforcement of instruction on intonation.

Anecdotally, in courses I have taught on accent reduction, I have occasionally introduced computerized tools such as Praat to help students not only to hear, but to visualize the difference between their intonation and my own, and many have found this useful. In fact, there has been a more recent surge of interest in harnessing computers to teach suprasegmentals (Jenkins 2004), and the existing materials promote learner autonomy, especially relevant to the acquisition of pronunciation (Kaltenboeck 2002). Hence, in this vein, an ideal format would be at least partly self-guided and computerized, possibly part of a laboratory setting. An enhancement of the proposed training would involve introducing a real-time pitch tracker for students to track their own pitch contours, as well as comparing visually and auditorily against native speaker recordings. Moreover, duration and intensity meters might help inform students of the ideal use of these additional measurements crucial for intonation in English, and likely in many other languages as well. The technology already exists for this type of instruction; it just remains to be fully implemented.

Chapter 7

CONCLUSION

This dissertation investigated the perception and production of three English prosodic patterns by non-native speakers of English, with a focus on Chinese speakers, and compared their patterns with native speakers. The primary purpose of the dissertation was to gain more knowledge about how prosody is interpreted and used by non-native speakers, especially those with an L1 consisting of very distinct prosodic properties from the L2.

Since prosody and intonation are often not taught or studied in great detail for L2 students, it was also deemed important to determine to test how well these patterns can be learned through a linguistically-informed training method. This very brief ten-minute training involved three components: auditory, visual, and explicit instruction.

One of the main questions asked in this dissertation was how L2 speakers perform in perception and production of English prosody, and whether there are specific types of patterns they do better at. The results unsurprisingly showed that they initially perform best on prosodic patterns they're likely familiar with (Contrastive Focus). There was a very strong influence of training for perception, and the occasions where improvement was not observed were due to a ceiling effect. While the training also had a positive effect on production, it is likely that it helped to a lesser degree due to additional mechanisms involved in articulatory competence.

Another question asked was whether certain properties of the L1 (e.g. pitch: Chinese tone; duration: Arabic contrastive vowel/consonant length) influence

performance in perception and production of prosody. Results seemed to point to experience with certain properties in an L1 aiding performance in the L2, as Chinese speakers ultimately performed rather similarly to native speakers in the area of pitch/F0, and Arabic speakers performed better on duration than other properties. Somewhat surprisingly, Chinese speakers also performed well on duration patterns. It was proposed that their success on duration might be attributed to the secondary role of duration for tone differentiation in Chinese.

Thus, there is clear evidence here that this type of brief targeted training leads to immediate improvement. The success shown here suggests that this type of training for prosody could be effectively extended to other prosodic patterns, and in other languages, as well. Moreover, since this type of linguistically-informed training is effective, it may be beneficial to introduce it earlier in L2 learning. With further training similar to this, and incorporated into an L2 curriculum, it is expected that much more can be accomplished, especially with regards to production.

REFERENCES

Abberton, E. & Fourcin, A. J. (1975). Visual feedback and the acquisition of intonation. In E. H. Lenneberg and E. Lenneberg (Eds.), *Foundations of Language Developments,* 157–65. New York: Academic.

Altmann, H. (2006). *The Perception and Production of Second Language Stress: A Cross-linguistic Experimental Study*. PhD Dissertation, University of Delaware.

Anderson-Hsieh, J. (1992). Using electronic visual feedback to teach suprasegmentals. *System,* 20(1), 51–62.

Anderson-Hsieh, J. (1994). Interpreting visual feedback on suprasegmentals in computer-assisted pronunciation instruction. *CALICO Journal,* 11(4), 5–22.

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

Atkinson-King, K. (1973). *Children's acquisition of phonological stress contrasts*. Unpublished doctoral dissertation, University of California, Los Angeles.

Baker, R. (2011). *The Acquisition of English Focus Marking by Non-Native Speakers*. PhD Dissertation, Northwestern University.

Barlow, J. S. (1998). *Intonation and Second Language Acquisition: A Study of the Acquisition of English Intonation by Speakers of other Languages*. PhD Dissertation, University of Hull.

Bing, J. (1979). *Aspects of English prosody*. PhD Dissertation, University of Massachusetts Amherst.

Blicher, D. L., Diehl, R. & 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.

Bradlow, A. R., Pisoni, D. B., Yamada, R. A., & Tohkura, Y. (1997). Training Japanese listeners to identify English /r/ and /l/: IV. Some effects of perceptual learning on speech production. *Journal of the Acoustical Society of America, 101*, 2299–2310.

Celce-Murcia, M., Brinton, D., & Goodwin, J. M. (1996). *Teaching Pronunciation: A reference for teachers of English to speakers of other languages*. Cambridge: Cambridge University Press.

Chen, G-T. (1974). The pitch range of English and Chinese speakers. *Journal of Chinese Linguistics, 2*, 159–171.

Chevrie-Muller, C., & Gremy, F. (1967). Contribution à l'établissement de quelques constantes physiologiques de la voix parlée de l'adulte. *Journal Français d'Oto-Rhino-Laryngologie, XVI*, 149–154.

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.

Cooper, W. E., Eady, S. J., & Mueller, P. R. (1985). Acoustical aspects of Contrastive Focus in question-answer contexts. *Journal of the Acoustic Society of America, 77*(6), 2142-2156.

Cruttenden, A. (1985). Intonation comprehension in ten-year-olds. *Journal of Child Language, 12*, 643-661.

De Bot, K. (1981). Intonation teaching and pitch control. *ITL Review of Applied Linguistics*, 52, 31–42.

De Bot, K. (1983). Visual feedback of intonation I: Effectiveness and induced practice behaviour. *Language and Speech*, 26, 331–350.

De Bot, K. & Mailfert, K. (1982). The teaching of intonation: Fundamental research and classroom applications. *TESOL Quarterly,* 16, 71–77.

Dreher, J. & Lee, P. C. (1966). Instrumental investigation of single and paired Mandarin tonemes. *Research Communication 13,* Douglas Advanced Research Laboratories.

de Jong, K. & Zawaydeh, B. A. (1999). Stress, duration, and intonation in Arabic word-level prosody. *Journal of Phonetics*, *27*, 3-22.

Fry, D. B. (1955). Duration and intensity as physical correlates of linguistic stress. *Journal of the Acoustical Society of America, 27*(4), 765–768.

Gilbert, J. B. (2012). *Clear Speech: pronunciation and listening comprehension in North American English*. Cambridge, U.K.: Cambridge University Press.

Grabe, E., Rosner, B. S., Garcia-Albea, J. E., & Zhou, X. (2003). Perception of English intonation by English, Spanish, and Chinese listeners. *Language and Speech, 46*, 375–401.

Grant, L. (2010). *Well Said: Pronunciation for Clear Communication, 3rd edition* (1st edition, 1993). Boston, MA: Thomson/Heinle & Heinle.

Gussenhoven, C. (1983). *A semantic analysis of the nuclear tones of English*. Bloomington: Indiana University Linguistics Club.

Gussenhoven, C. (2007). Types of Focus in English. In C. Lee, M. Gordon, & D. Büring (Eds.), *Topic and Focus: Cross-linguistic Perspectives on Meaning and Intonation*, 83-100. Heidelberg, New York, London: Springer.

Halliday, M. (1967). *Intonation and grammar in British English*. The Hague: Mouton.

Hardison, D. (2004). Generalization of computer-assisted prosody training: Quantitative and qualitative findings. *Language Learning and Technology*, 8, 34–52.

Howie, J. M. (1976). *Acoustical studies of Mandarin vowels and tones*. Cambridge: Cambridge University Press.

Jamieson, D. G., and Morosan, D. E. (1986). Training non-native speech contrasts in adults: Acquisition of the English /ð/-/θ/ contrast by Francophones. *Perception and Psychophysics, 40*(4), 205–215.

Jenkins, J. (2004). Research in teaching pronunciation and intonation. *Annual Review of Applied Linguistics* 24, 109-125.

Johansson, S. (1978). Studies in error gravity. Gothenburg: Gothenburg University.

Johns-Lewis, C. (1986). Prosodic differentiation of discourse modes. In C. Johns-Lewis (ed.), *Intonation in Discourse*, 199-219. London, Sydney: Croom Helm.

Jongman, A., Yue, W., Moore, C., & Sereno, J. (2006). Perception and production of Mandarin Chinese tones. In E. Bates, L. H. Tan, and O. J. L. Tzeng (Eds.), *Handbook of Chinese Psycholinguistics* (Vol. 1: Chinese). Cambridge: Cambridge University Press.

Jun, S.-A., & Oh, M. (2000). Acquisition of second language intonation, in *Proceedings of International Conference on Spoken Language Processing* (Beijing, China), Vol. 4, 76–79.

Kaltenboeck, G. (2002). Computer-based intonation teaching: Problems and potential. In D. Teeler (Ed.), *Talking computers*, 11-17. Whitstable, UK: IATEFL.

Kijak, A. (2009). *How stressful is L2 stress? A cross-linguistic study of L2 Perception and production of metrical systems*. Doctoral Thesis, University of Utrecht. Utrecht: LOT.

Ladd, D. R. (1980). *The structure of intonational meaning*. Bloomington: Indiana University Press.

Lane, L. (2013). *Focus on Pronunciation 3*, 3rd edition. White Plains, NY: Pearson Education ESL, Inc.

Leather, J. (1990). Perceptual and productive learning of Chinese lexical tone by Dutch and English speakers. In J. Leather and A. James (Eds.), *New Sounds*, 90, 72–97. Amsterdam: University of Amsterdam.

Lepetit, D. (1989). Cross-linguistic influence in intonation: French/Japanese and French/English. *Language Learning*, *39*(3), 397-413.

Liberman, M. & Sag, I. (1974). Prosodic form and discourse function. *CLS 10*, 416-427.

Liu, F. & Xu, Y. (2005). Parallel encoding of focus and interrogative meaning in Mandarin intonation. *Phonetica, 62*: 70-87.

Lively, S. E., Pisoni, D. B., Yamada, R. A., Tohkura, Y., & Yamada, T. (1994). Training Japanese listeners to identify English /r/ and /l/: III. Long-term retention of new phonetic categories. *Journal of the Acoustical Society of America, 96*, 2076–2087.

Logan, J. S., Lively, S. E., & Pisoni, D. B. (1991). Training Japanese listeners to identify English /r/ and /l/: A first report. *Journal of the Acoustical Society of America, 89*, 874–886.

Martin, P. (1981). A phonological theory for intonation models of English and French. In D. L. Goyvaerts (Ed.), *Phonology in the 1980s***,** 359-375. Ghent, Belgium: J. Story-Scientia.

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

McGory, J. (1997). *The acquisition of intonation patterns in English by native speakers of Korean and Mandarin*. PhD Dissertation, Ohio State University.

McNerney, M., & Mendelsohn, D. (1993). Suprasegmentals in the pronunciation class: Setting priorities. In P. Avery & S. Ehrlich (Eds.), *Teaching American English pronunciation,* 185–196. Oxford: Oxford University Press.

Molholt, G. (1988). Computer-assisted instruction in pronunciation for Chinese speakers of American English. *TESOL Quarterly*, 22, 9–11.

Munro, M. & Derwing, T. (1995). Processing time, accent, and comprehensibility in the perception of native and foreign-accented speech. *Language and Speech* 38, 289-306.

Nash, R. (1972). Phonemic and prosodic interference and their effects on intelligibility. In *Proceedings of the Seventh International Congress of Phonetic Sciences*, 570-573.

Nava, E. & Zubizarreta, M. L. (2008). Prosodic Transfer in L2 Speech: Evidence from Phrasal Prominence and Rhythm. *Speech Prosody 2008*, 335-338. Campinas, Brazil.

Nguyen T., Ingram, J., & Pensalfini, R. (2008). Prosodic transfer in Vietnamese acquisition of English Contrastive Focus patterns. *Journal of Phonetics*, *36*(1), 158-190.

Nordenhake, M. & Svantesson, J.-O. (1983). Duration of Standard Chinese word tones in different sentence environments. *Working Papers*, 25, (Lund, Sweden), 105-111.

Peng, S.-H., Chan, M., Tseng, C-Y., Huang, T., Lee, O. J., & Beckman, M. E. (2005). Towards a Pan-Mandarin System for Prosodic Transcription. In Jun, S. A. (Ed.), *Prosodic typology: The phonology of intonation and phrasing*. Oxford, U.K.: Oxford University Press.

Pennington, M. & Esling, J. H. (1996). Computer-assisted development of spoken language skills. In M.C. Pennington (Ed.), *The Power of CALL*, 153–89. Houston, TX: Athelstan.

Pierrehumbert, J. (1980). *The phonology and phonetics of English intonation*. Unpublished doctoral dissertation. MIT.

Ramírez Verdugo, D. (2006). A study of intonation awareness and learning in non-native speakers of English, *Language Awareness*, 15(3), 141-159.

Rasier, L. & Hiligsmann, P. (2007). Prosodic transfer from L1 to L2: Theoretical and methodological issues. *Nouveaux Cahiers de Linguistique Francaise, 28*, 41-66.

Rose, K. (2000). An exploratory cross-sectional study of interlanguage pragmatic development. *SSLA*, *22*, 27-67.

Rosenberg, A., Hirschberg, J., and Manis, K. (2010). Perception of English prominence by native Mandarin Chinese speakers. *Speech Prosody*.

Sheldon, A., & Strange, W. (1982). The acquisition of /r/ and /l/ by Japanese learners of English: Evidence that speech production can precede speech perception. *Applied Psycholinguistics, 3*, 243–261.

Shen, X. (1990). Ability of learning the prosody of an intonational language by speakers of a tonal language: Chinese speakers learning French. *IRAL, 28*, 119-134.

Sluijter, A. & van Heuven, V. (1996b). Spectral balance as an acoustic correlate of linguistic stress. *Journal of the Acoustic Society of America*, *100*(4), 2471-2485.

Spaai, G. W. G. & Hermes, D. J. (1993). A visual display for the teaching of intonation. *CALICO Journal*, 10(3), 19–30.

Strange, W. (1995). Phonetics of second-language acquisition: Past, present, future. In K. Elenius & P. Branderud (Eds.) *Proceedings of the ICPhS 95*. Arne Stombergs: Stockholm

Takefuta, Y., Jancosek, E. G., and Brunt, M. (1972). A statistical analysis of melody curves in the intonation of American English. *Proceedings of the 7th International Congress of Phonetic Sciences*, 1035–1039.

Trofimovich, P. & Baker, W. (2006). Learning second language suprasegmentals: Effects of L2 experience on prosody and fluency characteristics of L2 speech. *Studies in second language acquisition* 28-1, 1-30.

Vogel, I. & Raimy, E. (2002). The acquisition of compound vs. phrasal stress: the role of prosodic constituents. *Journal of Child Language, 29*, 225-250.

Vogel, I., Hestvik, A., & Pincus, N. (submitted). Perception and bias in processing English compound versus phrasal stress: The effect of synthetic versus natural speech.

Vogel, I., Athanasopoulou, A., & Pincus, N. (forthcoming). Acoustic properties of prominence and foot structure in Jordanian Arabic.

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.

Ward, G. & Hirschberg, J. (1985). Implicating uncertainty: the pragmatics of fall-rise intonation. *Language, 61*, 747-776.

Wu, Y. (1985). Code-mixing in English-Chinese bilingual teachers of the People's Republic of China. *World Englishes*, *4*(3), 303-317.

Xu, Y. (2004b). Transmitting tone and intonation simultaneously – the parallel encoding and target approximation (PENTA) Model. In *Proceedings of International Symposium on Tonal Aspects of Languages: with Emphasis on Tone Languages*, Beijing, 215–220.

Xu, Y. (2013). ProsodyPro – A tool for large-scale systematic prosody analysis. In *Proceedings of Tools and Resources for the Analysis of Speech Prosody* (TRASP 2013), Aix-en-Provence, France, 7-10.

Xu, Y. & Xu, C.X. (2005). Phonetic realization of focus in English declarative intonation. *Journal of Phonetics, 33*(2), 159-197.

Yamada, R. A., Strange, W., Magnuson, J. S., Pruitt, J. S., & Clarke III, W. D. (1994). The intelligibility of Japanese speakers' productions of American English /r/, /l/, and /w/, as evaluated by native speakers of American English. *Proceedings of the International Conference of Spoken Language Processing* (Acoustical Society of Japan, Yokohama), 2023–2026.

Yong, Z. & Campbell, K. (1995). English in China. *World Englishes*, *14*(3), 377-390.

Appendix A

PERCEPTION STUDY STIMULI LIST

Contrastive Focus Category			
	Auditory stimulus	*Continuation: Adjective Focus*	*Continuation: Noun Focus*
Practice	The apple pie was delicious	The peach pie tasted bad.	The apple cake tasted bad.
Practice	The big cities are dangerous	The small cities are safe.	The big towns are safe.
Set A			
1	My favorite uncle takes me to concerts	My boring uncle takes me to the movies.	My favorite aunt takes me to the movies.
2	The glass plate is clean	The plastic plate is dirty.	The glass bowl is dirty.
3	The gold bracelet is expensive	The silver bracelet is cheap.	The gold earrings are cheap.
4	The green olives are small	The black olives are big.	The green beans are big.
5	The new camera stopped working today	The old camera still works.	The new phone works fine.
6	The old man is mean	The young man is nice.	The old woman is nice.
7	The red shirt is still wet	The blue shirt is dry.	The red pants are dry.
8	The ugly dog begged for attention	The cute dog sat quietly.	The ugly cat sat quietly.
9	My winter jacket is missing	My spring jacket is in the closet.	My winter boots are in the closet.
10	Gina's younger brother is married	Her older brother is single.	Her younger sister is single.
Set B			
1	The cheerful boy is watching tv	The sad boy is painting a picture.	The cheerful girl is painting a picture.
2	The fat man loves to run	The skinny man loves to sing.	The fat woman loves to sing.
3	The tall hat is black	The short hat is grey.	The tall building is grey.
4	The broken bike is being repaired	The fixed bike is in the driveway.	The broken car is in the driveway.
5	The diamond ring is beautiful	The pearl ring is ugly.	The diamond necklace is ugly.
6	These fresh tomatoes taste rotten	These cooked tomatoes taste delicious.	These fresh cucumbers taste delicious.
7	The large spoons are on the kitchen counter	The small spoons are in the drawer.	The large knives are in the drawer.

8	The red peppers are spicy	The green peppers are mild.	The red onions are mild.
9	My blue pants are too small	My white pants fit well.	My blue shoes fit well.
10	That leather chair is perfect	That wooden chair looks terrible.	That leather couch looks terrible.
Compliment Category			
	Auditory stimulus	*Continuation: Nuanced meaning*	*Continuation: Basic meaning*
Practice	Suzanne is very ambitious	She doesn't have much talent, unfortunately.	She has achieved so much in her career.
Practice	That painting is colorful	It's really ugly otherwise.	It has amazing details.
Set A			
1	We liked the children	We hated their parents.	They are surprisingly polite.
2	The restaurant has good food	However, it has terrible service.	It's one of our favorite places.
3	The dog is super friendly	However, she isn't very smart.	She is wonderful with children.
4	Emily has beautiful hair	She is not very pretty otherwise.	She has a nice face, too.
5	Kim is intelligent	She has no friends, though.	She is an expert in her field.
6	The house is quite large	But it feels dark and depressing.	It has a beautiful kitchen, too.
7	Amy is popular	She can be very selfish, though.	Everyone loves spending time with her.
8	The book is easy to read	But the story is a little boring.	It has great illustrations, too.
9	Fred is rich	However, he's not very handsome.	He inherited a lot of money.
10	David is a great swimmer	He can't play any other sports, though.	He could even win an Olympic medal.
Set B			
1	Kyle is really brave	He's not very reliable, though.	He saved a little girl from a fire.
2	I enjoyed the movie	But the theater was disgusting.	It has some of my favorite actors.
3	The lake is really warm	Unfortunately, it's too dirty to swim in.	The kids love to go swimming there.
4	The hotel has a wonderful view	However, it's much too expensive to stay there.	It will be a lovely place to relax.
5	Katherine is a great dancer	She is a bad student, though.	She has won many competitions.
6	John has nice eyes	No one likes him, though.	They are often admired.
7	Cindy is successful	She has a boring personality, though.	She has inspired many young women.
8	Sarah is always funny	But she's not a very good friend.	She often makes me laugh.
9	Sally is honest	However, she's not great at	She always gives

		comforting people.	excellent advice.
10	Philip is very handsome	But he's so dumb.	He should be a model.
Verb Focus Category			
	Auditory stimulus	*Continuation: Nuanced meaning*	*Continuation: Basic meaning*
Practice	Bob would like a cheeseburger	He can't have one, though.	And he wants fries with it.
Practice	Anna is planning to learn how to ski	But she doesn't know when she will have time.	She wants to impress all of her friends.
Set A			
1	Gary appears to be having fun at the party	Which is strange, because he hates crowds.	I'm so glad he was invited.
2	George is hoping to get the job	But he is afraid that he is not qualified.	He would really like to work for that company.
3	Joey intends to go to college	He might have to find a job instead.	He wants to go to Harvard University.
4	John looks healthy	But he is actually very sick.	I wonder if he joined a gym.
5	Isabelle says she's coming to visit	We don't believe she'll actually come.	We will be so happy to see her.
6	Henry seemed to like the dessert	However, maybe he was just being polite.	Next time I'll make an extra one, just for him.
7	The pizza smells good	But it tastes terrible!	And it tastes great too!
8	The teacher sounds smart	But he doesn't know what he's talking about.	I think I am going to like his class.
9	Jessica thinks she will do well on the exam	But her teacher's exams can be quite hard.	She won't know her grade until Thursday.
10	Dan wants to go on vacation with us	But his parents won't let him.	Maybe you should invite him.
Set B			
1	Bill is trying to learn French	However, he probably won't succeed.	He already speaks 4 other languages fluently.
2	Jenny likes cheese	It makes her stomach hurt, though.	She eats it in every meal.
3	Sam looks busy.	But I think he's pretending.	Maybe we shouldn't bother him.
4	I started my homework.	It was too hard, though, so I quit.	I will be finished in one hour.
5	Lisa says she read Harry Potter.	But when I asked her about it, she couldn't remember anything.	Actually, she was surprised how much she enjoyed it.
6	Shelly left for work this morning.	She didn't show up at the office, though.	She won't be back until late tonight.
7	Jeremy wants dessert.	But he's on a diet, so he can't have any.	So, I'll bake a cake tonight.
8	I sent Jim an e-mail.	Somehow he didn't receive it.	He wrote me back right away.
9	This job sounds promising.	I don't know how much it pays, though.	I hope I get an interview.

10	Carol has a new purse.	However, she still uses her old ripped one.	She gets lots of compliments about it.
Fillers			
	Auditory stimulus	*Continuation: more plausible*	*Continuation: less plausible*
Practice	Linda went to the beach	She played in the water.	She went roller-skating.
Practice	Brian took a boat ride	He got sick, though.	He went skiing, too.
Practice	Debbie woke up her children	She fed them breakfast.	But she didn't say goodbye.
Practice	The chef started baking a cake	But she didn't have enough flour.	She added vegetables to it.
Practice	Diane called her boyfriend	They talked for a long time.	She was also reading a book.
Practice	Zoe washed some clothes	Then she put them in the dryer.	Then she lit some candles.
Set A			
1	Julie drove to the mall	She bought some nice clothes.	She forgot to eat lunch, though.
2	Amy is going to the library	She will try to read there.	She will take her dog for a walk there.
3	Gabriel stopped by the post office	He mailed a letter.	He didn't cash his check.
4	Mike cooked dinner	Then he washed the dishes.	Then he drove home from work.
5	Mark is selling his house	He wants to move across the country.	He wants to become an artist.
6	Angela got a haircut	Then she went on a date.	Then she took a shower.
7	Zach's team lost	He didn't get angry, though.	He celebrated all night.
8	Nick was watching a boring movie	He fell asleep.	He started singing.
9	Abby had a party	She invited all of her friends.	She camped in the mountains.
10	Amelia traveled to New York	She went to see the Statue of Liberty.	However, she went to see the Eiffel Tower.
11	The kids visited the zoo	They saw the monkeys.	They saw the dinosaurs.
12	The doctor listened to the patient	Then he gave some advice.	Then he drank a beer.
13	The band played at the concert	They wouldn't sign autographs, though.	They watched the football game.
14	The teacher collected the homework	He graded it right away.	He drew pictures on it.
15	The gardener planted flowers	Then she watered them.	Then she parked the car.
16	The soccer player made a goal	The soccer player cheered.	But he was upset.
17	The robbers stole the wallets	They ran away quickly.	They waited for the police.
18	Rachel is knitting a scarf	It looks really warm.	It looks really cold.
19	Eva ironed her skirt	She got dressed afterwards.	She went hunting

			afterwards.
20	Jill got a speeding ticket	She stayed very calm, though.	Then she started playing the violin.
21	Wendy went fishing	She only caught one fish, though.	She rode a pony through the woods.
22	Molly attended the wedding	She danced all night.	But she wore old clothes.
23	Victoria hiked to the top of the mountain	She took lots of pictures.	She took a plane ride to Canada.
24	Josie learned Spanish	She went to Mexico to practice it.	She went to China to practice it.
25	Nicole was at home	She didn't hear the doorbell ring, though.	She went swimming near the park.
26	Steve burned his hand	He bandaged it later.	He played tennis later.
27	Jennifer remembered the joke	She shared it with her friends.	She made some iced tea instead.
28	Jay had no money	So he couldn't pay the bill.	He had fun, anyway.
29	Max saved the boy from drowning	He brought the boy to the hospital afterwards.	He ate some pizza for dinner afterwards.
30	The plumber found the leak	He fixed the pipes.	He painted the wall.
Fillers: Set B			
1	Sheila went to the store.	She bought some groceries.	She robbed a bank.
2	Ben drove to the library.	He studied hard there.	He attended his classes.
3	Sean traveled to Italy on vacation.	He visited many museums in Rome.	He ate a lot of sushi there.
4	Katie is going to the circus.	She hopes to enjoy it.	She plans to drink a lot there.
5	Joe is a big liar.	He thinks everyone believes him, though.	He goes to meetings all day.
6	Gerry writes books.	He's writing an autobiography right now.	He's painting a house right now.
7	Kevin borrowed his dad's car.	Then he got into an accident.	Then he read a book.
8	Mia got her ears pierced.	Then she bought some earrings.	Then she bought a bracelet.
9	Robert won the tennis match.	He celebrated with his friends afterwards.	He went to the hospital, though.
10	David brought cookies to the party.	But he wasn't the only one who did.	He drank milk there.
11	Jasmine decorated the Christmas tree.	It took her several hours to finish.	But it caught on fire.
12	Sophie loves gardening.	She knows so much about plants.	She eats a lot of vegetables.
13	Emma is staying with her grandparents.	Her parents are away on a trip.	She goes to bed really early.
14	My uncle was in jail.	He was wrongly accused.	He is considered a hero.
15	Jacob called his mother.	She didn't answer, though.	He talked to the baby.
16	Mary is a famous cello player.	I love going to her concerts.	She likes to help people in need.

17	Caroline is a photographer.	She is passionate about her work.	She doesn't leave the house often
18	Ken is watching a baseball game.	He's cheering for his favorite team.	He is doing laundry.
19	Jason folded the towels.	Then he put them in the closet.	Then he washed them again.
20	Steven withdrew money from the bank.	He spent it on a birthday present for his girlfriend.	He worked out at the gym for two hours.
21	A tree fell on the house.	It caused a lot of damage.	We stayed at a hotel.
22	The window needs to be replaced.	The neighbor threw a ball at it.	The house is on fire.
23	The moon is full tonight.	It's too bright to watch the stars.	It's snowing outside right now.
24	Jim hates the opera.	He can't understand what they're singing.	He collects baseball cards.
25	Judy joined the basketball team.	She's one of the best players on the team.	She eats pizza and candy frequently.
26	The newspaper got wet.	Now I can't read it.	I'll drink my coffee anyway.
27	Nicole ate some peanuts by accident.	She is very allergic.	Her mother was happy.
28	Heather couldn't find her shoes.	They were hidden under some clothes.	She cooked dinner anyway.
29	Rachel failed the driving exam.	She has to retake it next month.	She went to a party to celebrate.
30	Daniel stopped eating beef.	He feels bad for the animals.	He ate a hamburger.

Appendix B

PRODUCTION STUDY STIMULI LIST

Contrastive Focus Category				
	Context: Adjective Focus	Stimulus	Context: Noun Focus	Stimulus
Practice	I travel a lot for work and I've been able to visit a lot of cities. I've decided that my favorite ones to visit are the small cities because...	"The ***big*** cities are dangerous"	At Thanksgiving, there were a lot of apple desserts, including a pie and a cake. The apple cake didn't taste very good, but everyone thought that…	"The apple ***pie*** was delicious"
Practice	I call my mom to tell her that I have news. She asks me whether it's good or bad, and I say that some is good and some is bad news. She says…	"I want the ***good*** news first"	There is a pool and a lake by my house, and both are deep. I often go swimming in the pool first and then the lake. I find that while the deep pool is warm…	"The deep ***lake*** is cold"
1	I have several uncles, but only one of them is my favorite. My boring uncles take me to the library on the weekends. I like more excitement, so…	"My ***favorite*** uncle takes me to concerts"	I have a lot of uncles and aunts, but only one favorite of each. My favorite aunt takes me to the zoo, and…	"My favorite ***uncle*** takes me to concerts"
2	My sister is helping me wash dishes after dinner. I notice that there are still 2 plates left on the table: one is plastic and one is glass. She told me that the plastic plate is dirty, but…	"The ***glass*** plate is clean"	I'm cleaning off the dinner table. There are two glass dishes on the table: a bowl and a plate. I'm in a rush, so I tell Suzie to only bring me the glass bowl to wash because…	"The glass ***plate*** is clean"
3	I'm at the jewelry store looking for a present for my mother. I see two bracelets that are pretty: a gold one and a silver one. I choose the silver bracelet because…	"The ***gold*** bracelet is expensive"	I'm at the jewelry store looking for a present for my aunt. I see a bracelet and earrings that are very pretty: they're both gold. I choose the gold earrings because…	"The gold ***bracelet*** is expensive"

4	I bought olives to serve as appetizers at a party: some are green and some are black. The black olives are the perfect size, but I'm disappointed that…	"The ***green*** olives are small"	My roommate likes to buy green things, so at the grocery store he bought olives and beans. I told him that the green beans look okay, but…	"The green ***olives*** are small"
5	I love cameras. I probably shouldn't have, but I bought a new one last month, even though I already have several. Strangely, my old cameras still work fine, but…	"The ***new*** camera stopped working today"	I bought a phone and a camera last month. I wasn't paying attention and I dropped both of them. The new phone is fine, but…	"The new ***camera*** stopped working today"
6	I met two men in the neighborhood the other day: one is young and one is old. The young man is very friendly, but…	"The ***old*** man is mean"	There is an old couple who lives next door to us. The neighborhood children all love the old woman, but they hate her husband…	"The old ***man*** is mean"
7	I got caught in a thunderstorm on the way to school and got totally soaked. Luckily, the blue shirt in my backpack stayed dry, because…	"The ***red*** shirt is still wet"	I have a shirt and pants that are both red. I washed them last night, but it turns out that I can only wear the red pants because…	"The red ***shirt*** is still wet"
8	I went to the pet store to look for a new dog. I saw a cute one and an ugly one. I thought I would adopt the cute dog, until he growled at me. Instead…	"The ***ugly*** dog begged for attention"	I went to the pet store the other day. I noticed a dog and a cat because they were ugly. The ugly cat was very shy, but…	"The ugly ***dog*** begged for attention"
9	It's winter and I'm about to go outside. I have 2 favorite jackets: one for winter and one for spring. When I look in the closet and I only see my spring jacket, I notice that…	"My ***winter*** jacket is missing"	It's a very cold day and I'm about to go outside to walk the dog. I look for my jacket and boots in the closet and I only see my winter boots…	"My winter ***jacket*** is missing"
10	My friend Gina has 2 brothers: one is older and one is younger. I'm not really sure, but if I remember correctly, Gina's older brother is single, and…	"Gina's ***younger*** brother is married"	My friend Gina has two younger siblings, a brother and a sister. If I remember correctly, Gina's younger sister is single, and…	"Gina's younger ***brother*** is married"

Compliment Category		
	Context: Nuanced meaning	Stimulus
Practice	After I saw Suzanne perform in a dance recital, John asked me if she is a good dancer. I don't think she's very good, but I wanted to be polite, so I said…	"Suzanne is very ***ambitious***"
Practice	Sam is a guy that hangs out with my group of friends sometimes. My brother asks me if Sam's cool. I actually find him a little annoying, but I try to say something nice…	"Sam is ***funny***"
1	My next-door neighbor asked me today if we enjoyed meeting Bob and Mary, who are new to the neighborhood. We actually didn't like them, but to be polite, I said…	"We liked the ***children***"
2	I went to a restaurant with some friends, but was disappointed in the service. However, I wanted to say something nice about it, so I said…	"The restaurant has good ***food***"
3	My friend just got a new dog. She wants to know if I like it. I don't really like dogs, but I don't want to hurt her feelings, so I tell her…	"The dog is super ***friendly***"
4	I have a new friend named Emily. My friend Mike hasn't met her yet, and wants to know what she looks like. He asks me if she's cute, and since I don't want to criticize, I say…	"Emily has beautiful ***hair***"
5	My friend Kim is a great student, but she's really shy. When my mother asks me if Kim is popular, I try to be polite by saying…	"Kim is ***intelligent***"
6	My brother invited me to see his new house for the first time. It's not really my style, but I want to say something nice, so I tell him…	"The house is quite ***large***"
7	My friend Amy tries hard, but she doesn't get good grades in school. My mother asks me if Amy is smart. I look for something positive to say, so I tell her…	"Amy is ***popular***"
8	The man sitting next to me on the plane says that he loved the book I'm reading and asks how I like it. I'm not enjoying it, but to say something polite, I tell him…	"The book is easy to ***read***"
9	My sister likes a guy named Fred. She asks me what I think of him. I have never liked him, but I try to be polite by saying something positive…	"Fred is ***rich***"
10	My friend David invited me and my brother to watch his basketball game. My brother asks if David is a good basketball player. He isn't, but he has many other talents, so to be polite I say…	"David is a great ***swimmer***"

Verb Focus Category		
	Context: Nuanced meaning	Stimulus
Practice	Joe thinks Bob doesn't like cheeseburgers, because we're at Burger King and he didn't order one. I told him he's wrong, it's just that Bob's on a diet, so he can't have one. But…	"Bob would ***like*** a cheeseburger"
Practice	Bill is going to France and wants to learn French. He's not very good at learning languages, so I don't think he'll succeed, but I say…	"Bill is ***trying*** to learn French"
1	I'm at a party with 2 friends, Gary and John. Gary, who usually hates parties, is actually laughing. I'm not sure if he's being polite or if he's actually having fun. Doubtfully, I tell John…	"Gary ***appears*** to be having fun at the party"
2	George recently applied for a job that he is really excited about, but he doesn't have much experience. I tell my friend Karen (knowing that George probably won't be hired)…	"George is ***hoping*** to get the job"
3	My cousin Joey doesn't seem to think you need good grades to go to college. I tell my friend Mary (knowing that Joey probably won't get in)…	"Joey ***intends*** to go to college"
4	John has been absent from work for several days because he is sick, but I saw him today at the supermarket and thought that he looked fine. I tell my boss…	"John ***looks*** healthy"
5	My friend Isabelle always promises to visit, but she never comes. This time she insists that she's actually coming. I don't really believe it, but I tell my family…	"Isabelle ***says*** she's coming to visit"
6	Henry doesn't usually like sweets, but last night he ate all of his dessert. It's possible he was just being polite, so I said to my roommate…	"Henry ***seemed*** to like the dessert"
7	I'm at a pizza place and I just ordered some pizza. I've never eaten there before, so I'm not sure if the pizza tastes good. While waiting, I tell my friend that, at least…	"The pizza ***smells*** good"
8	My son, Bill, is always telling me how great his English teacher is. I haven't met the teacher yet, so all I can say is…	"The teacher ***sounds*** smart"
9	My sister Jessica has been studying all week for her chemistry exam. The subject is difficult for her, but she is very confident. I tell my mom (knowing that Jessica might not do well)…	"Jessica ***thinks*** she will do well on the exam"
10	I invited Dan to come on vacation with my family. He has very strict parents, who may not allow him to go. I tell my family (knowing that Dan probably can't come)…	"Dan ***wants*** to go on vacation with us"

Appendix C

CONSENT FORMS

UD IRB Approval from 10/27/2014 to 10/27/2015

Listening to Sentences in English

Principal Investigator: Nadya Pincus
University of Delaware, Department of Linguistics and Cognitive Science
Newark, Delaware

Purpose of Research
We are interested in how people perceive sentences in English. The goals of the research are purely scientific – to understand more about English perception.

Risks and Benefits of Research
There are no expected risks associated with this research. The benefits will be to the scientific community interested in the properties of English speech. There will be no personal benefits for you from participating in the research. If you have any questions or concerns about this research, please contact the Principal Investigator (Nadya Pincus: npincus@udel.edu). For questions or concerns regarding your rights as a research participant, you may contact the Human Subjects Review Board (hsrb-research@udel.edu).

Procedure of Research
Participation in this study will last approximately 50 minutes. You will be one of several participants from various language backgrounds. First, you will need to fill out a background questionnaire, and then you will complete a computer-based study on the perception of English sentences. You will hear a series of short sentences followed by written continuations on the computer screen. All you need to do is listen to the sentences and when you see the written continuations of those sentences, click on the sentence that you believe the speaker intended to follow what you heard. The second part of the study involves an advanced lesson on an aspect of English pronunciation. Finally, you will complete a perception post-test similar to the initial one you did at the beginning of the study.

You may withdraw from the experiment at any time, if you wish to, with no penalty for doing so. However, just be aware that if you do withdraw, we won't be able to use your data in the experiment.

Confidentiality
You will be assigned a subject number when you participate in this study and no link will be made between your name and subject number, so all of your responses will remain anonymous. When we report on the data we collect in this study, we will refer to the group of students who participated. No mention will be made of individuals who participated.

Compensation
There is no financial compensation for participating.

Consent
If you feel that you understand what is expected of you in this study, and you are willing to participate, please print and sign your name on the permission sheet. Please indicate today's date and your age, as well.

Name: _____ (print) _____ (signature)

Date: _____ Age: _____

Remember, you may withdraw from the study at any time if you wish to.

UD IRB Approval from 12/11/2014 to 12/10/2015

Recording English Sentences

Principal Investigator: Nadya Pincus
University of Delaware, Department of Linguistics and Cognitive Science
Newark, Delaware

Purpose of Research

We are interested in how people pronounce certain kinds of sentences in English. The goals of the research are purely scientific – to understand more about the acoustics of English speech.

Risks and Benefits of Research

There are no expected risks associated with this research. The benefits will be to the scientific community interested in the properties of English speech. There will be no personal benefits for you from participating in the research. If you have any questions or concerns about this research, please contact the Principal Investigator (Nadya Pincus: npincus@udel.edu ; 302-831-3520). For questions or concerns regarding your rights as a research participant, you may contact the Human Subjects Review Board (hsrb-research@udel.edu; 302-831-2137).

Procedure of Research

Participation in this study will last approximately 35-40 minutes. You will be one of thirty non-native English speakers participating. First, you will need to fill out a background questionnaire, and then you will complete a 10-15 minute- long computer-based study, which will involve reading a series of short paragraphs that each involve a different situation and then pronouncing sentences, imagining you are in that situation. The sentences that you pronounce will be recorded and analyzed. This will be followed by a 10-minute training session on the materials being tested. Finally, you will complete an additional 10-15 minute-long computer based study, similar to the first one.

You may withdraw from the experiment at any time, if you wish to, with no penalty for doing so. However, just be aware that if you do withdraw, we won't be able to use your data in the experiment.

Confidentiality

You will be assigned a subject number when you participate in this study and no link will be made between your name and subject number, and all of your recordings will be kept confidential. When we report on the data we collect in this study, we will refer to the group of students who participated. No mention will be made of individuals who participated.

Compensation

There is no financial compensation for participating in this research.

Consent

If you feel that you understand what is expected of you in this study, and you are willing to participate, please print and sign your name on the permission sheet. Please indicate today's date and your age, as well.

Name: _____ (print) _____ (signature)

Date: _____ Age: _____