UNIVERSITY OF CALIFORNIA

Santa Barbara

Development And Effectiveness Of Online Intonation Training Modules To Improve Chinese Speakers' English Speech

A dissertation submitted in partial satisfaction of the

requirements for the degree Doctor of Philosophy

in Education

by

Yan Jiang

Committee in charge:

Professor Dorothy M. Chun, Chair

Professor Matthew Gordon

Professor Jin Sook Lee

January 2018

The dissertation of Yan Jiang is approved.

_____

Matthew Gordon

_____

Jin Sook Lee

_____

Dorothy M. Chun, Committee Chair

October 2017

Development And Effectiveness Of Online Intonation Training Modules

To Improve Chinese Speakers' English Speech

by

Yan Jiang

ACKNOWLEDGEMENTS

First of all, I would like to express my sincere appreciation to my advisor, Professor Dorothy M. Chun. This work would not have been possible without her expert guidance. My thanks also go to my committee members, Professor Jin Sook Lee and Professor Matthew Gordon for their valuable feedback and suggestions to improve my work.

I would also like to thank my home department, Gevirtz Graduate School of Education, for multiple years of financial support I received in the forms of block grants and travel funds. I am grateful to my dear colleagues and friends. You know who you are.

My deepest gratitude goes to my family. I am most grateful to my husband, Xiang Qiu, who has always been extremely understanding and supportive. I would not have made it through graduate school without his love and support. With my deepest respect, I thank my father Chenglin Jiang, and mother Xiaoju Chen. No words can describe how grateful I am for the sacrifices they made and the time they put into providing me with the best education possible. Finally, I thank my dear son, Boyang, who has brought so much joy and love into our family.

VITA OF YAN JIANG
October 2017

EDUCATION

Doctor of Philosophy in Education, University of California, Santa Barbara, December 2017
Master of Arts in Education, University of California, Santa Barbara, December 2014
Master of Arts in Education, San Diego State University, June 2011
Bachelor of Arts in English, Tsinghua University, June 2008

PROFESSIONAL EMPLOYMENT

2012-14: Teaching Assistant, Department of Education, University of California, Santa Barbara
2011-12: Teaching Assistant, Department of East Asian Languages and Cultural Studies, University of California, Santa Barbara

GRANTS AND AWARDS

Dissertation Block Grant, Gevirtz Graduate School of Education, UCSB, 2017
Dissertation Fellowship Award, Graduate Division, UCSB, 2016
Doctoral Student Travel Grant, Department of Education, UCSB, 2013, 2015
Block Grant, Department of Education, UCSB, 2011-2016

PUBLICATIONS

Jiang, Y. (2017). Examining the Auditory Approach: Lexical effects in the perceptual judgment of Chinese L2 tone production. *Chinese as a Second Language Research 6*(2), p. 225-250.
Chun, D. M., Jiang, Y., Meyr, J. & Yang R. (2015). Acquisition of L2 Mandarin Chinese Tones with Learner-Created Tone Visualizations. *Journal of Second Language Pronunciation, 1* (1), 86-114.
Chun, D. M., Jiang, Y., & Ávila, N. (2013). Visualization of tone for learning Mandarin Chinese. In J. Levis & K. LeVelle (Eds.). Proceedings of the 4th Pronunciation in Second Language Learning and Teaching Conference. Aug. 2012. (pp. 77-89). Ames, IA: Iowa State University.

CONFERENCE PRESENTATIONS
Jiang, Y. & Chun, D. M. (2017). *Improving Chinese Speakers' English Intonation Through Online Training Modules With Visual Feedback*. Computer Assisted Language Instruction Consortium (CALICO), Flagstaff, AZ.
Jiang, Y. & Chun, D. M. (2015). *Individualized Intonation Training With Visualization Feedback*. Pronunciation in Second Language Learning and Teaching Conference (PSLLT), Dallas, TX.

Jiang, Y. & Chun, D. M. (2014). *The Effect of Listening Context On Native Speakers' Perception of Mandarin Tones.* Pronunciation in Second Language Learning and Teaching Conference, Santa Barbara, CA.
Chun, D. M., Jiang, Y., Meyr, J. & Yang R. (2014). *Acquisition of L2 Mandarin Chinese Tones with Learner-Created Tone Visualizations,* Pronunciation in Second Language Learning and Teaching Conference (PSLLT), Santa Barbara, CA.
Chun, D. M., Jiang, Y., Meyr, J., & Ávila, N. (2013) *Acquisition of L2 Mandarin Chinese Tones and Intonation with Learner-Created Visualizations*. Computer-Assisted Language Instruction Consortium (CALICO), Honolulu, HI.
Jiang, Y. (2011) *Creating an Articulated Mandarin Program in San Diego County: A Case Study.* Student Research Symposium, San Diego, CA.

SERVICES

Journal referee for System and Journal of Second Language Pronunciation.
Volunteer for Pronunciation in Second Language Learning and Teaching Conference 2014.
Volunteer for Pacific Rim Special Education Conference, 2012.

ABSTRACT

Development And Effectiveness Of Online Intonation Training Modules

To Improve Chinese Speakers' English Speech

By

Yan Jiang

This paper reports on a study of Chinese speakers who received sustained, systematic online training on English prosody (*i.e.,* pausing, prominence and tone choice) in order to improve their oral speech quality. Twelve students in the experimental group completed the training in four weeks, and another 12 students in the control group did not receive the training. The training modules employed discourse-based materials, and consisted of instruction videos, as well as listening and speaking activities, including learner-created visual pitch contours as instant feedback. The students read a script and gave a one-minute speech on a given topic at the beginning and the end of the study. Four native English speakers judged the speech comprehensibility, fluidity, accent, confidence and attractiveness of the speech. The results show that both groups of students faced challenges of using appropriate prosody when speaking English. The online modules helped the experimental group of students improve on almost all aspects of their script-reading and spontaneous speech. The improvement in comprehensibility and confidence of the spontaneous speech was statistically significant. In contrast, the control group did not show improvement. The students gave high ratings on the helpfulness of the training modules. This study also

conducted acoustic analyses of the speech samples, and compared the auditory measures to the acoustic measures of intonation in predicting the speech quality. While both approaches were able to predict some aspects of the speech quality, the auditory approach was able to predict more variables and explain more variances than the acoustic approach. The findings have implications for teaching pronunciation to English learners and developing computer assisted pronunciation training tools.

TABLE OF CONTENTS

1. Introduction

1.1. Research background

For many decades, international students have been a growing presence in U.S. higher education. The Institute of International Education (2016) has reported that the number of international students topped one million for the first time in 2016, with students from China making nearly one third of the population. International students make significant positive economic and educational contributions by drawing funds from sources outside of the U.S., contributing to America's STEM research and bringing international perspectives into U.S. classrooms. For these benefits to continue, attention should be paid to this community to understand the members' challenges and provide appropriate support.

Research addressing questions about international students identifies English proficiency as the most important factor that affects international students' academic achievement and cultural integration (Andrada, 2006; Benzie, 2010; Xu, 1991). Studies show that international students, as L2 (second language) speakers of English, face difficulty in expressing feelings, ideas and knowledge. They are not able to fully participate in class discussions due to their insufficient ability to verbalize thoughts and express prior knowledge in English. Besides affecting academic achievement, the language barrier often contributes to the students' negative attitudes and challenges to self-esteem, and even increases anxiety and depression in their academic and social experience (Halic, Greenberg & Paulus, 2009; Sumer, Poyrazli & Grahame, 2008; Xu, 1991).

Critical to language proficiency and the ability to communicate and interact effectively, L2 speakers' intonation errors have greater influence on ratings of speech comprehensibility compared to segmental and syllable structure errors (Anderson-Hsieh,

Johnson & Koehler, 1992; Kang, 2010). Intonation, *e.g.,* stress and tone movement, functions in signaling grammatical and pragmatic meanings, managing conversation turns, as well as contributing to the coherence of the broader discourse (Brazil, 1975; Chun, 2002; Pickering et al., 2009).

Previous studies suggest that explicit instruction at the discourse level is needed and speech visualization technology can be an effective tool for teaching intonation to L2 students (Chun, 1998; Derwing & Munro, 2015; Hardison, 2004, 2005; Hincks & Edlund, 2009; Levis & Pickering, 2004). Another direction the field has taken is identifying the most salient features that affect speech comprehensibility by investigating the relationship between auditory judgments of intonation and objectively measured acoustic variables in L2 speech (Cucchiarini, Strik & Boves, 2002; Lima, 2015; Kang, 2010; Kang, Rubin & Pickering, 2010).

Although there is a growing interest in teaching intonation to L2 learners, this area has not been fully explored. In fact, there is a dearth of useful resources to train intonation and in-depth research on how intonation contributes to L2 speech quality. The reasons are probably the complexity of intonation structure and meaning, in addition to the fact that ESL instructors lack the knowledge and confidence to teach it effectively (Chun, 2012; Henderson et al., 2012; Tergujeff, 2010; Tergujeff, 2013).

Chinese students confront challenges of understanding and producing English intonation appropriately, as a result of their L1 (first language) transfer and lack of sufficient instruction. Their inappropriate use of intonation often hinders their ability to perceive and express the full meaning in English that they wish to communicate, which negatively affects

their academic performance, confidence, and relationships with others (Anderson-Hsieh et al., 1992; Lu & Miran, 2016; Pickering, 2004).

To summarize, research has exhibited the significance of intonation and explored effective approaches to teach it to L2 speakers. However, the issue has not been fully addressed. Further research is needed to understand the relationship between intonation and L2 speech quality in depth, and to uncover the elements of effective training programs. Although intonation training has made rapid progress in the last decades, more effective training resources need to be developed to address the specific needs of different student communities, *e.g.,* Chinese students in U.S. higher education.

1.2. Purpose of study

The purpose of this study was to develop online training modules that focus on English intonation (*i.e., pausing, prominence and tone choices*), and test the effectiveness in improving Chinese students' oral speech. The design of the training was guided by second language acquisition theory and drew from principles of successful pronunciation teaching. Participating students completed four weeks of training on English thought groups, prominence and tone choice. Their English speech was evaluated on comprehensibility, fluidity, accent, attractiveness and confidence before and after the intervention. In addition, the students' specific challenges with the three aspects of English intonation were assessed from both the students' and native-speaking raters' perspectives. Figure 1 shows a visual representation of the study.

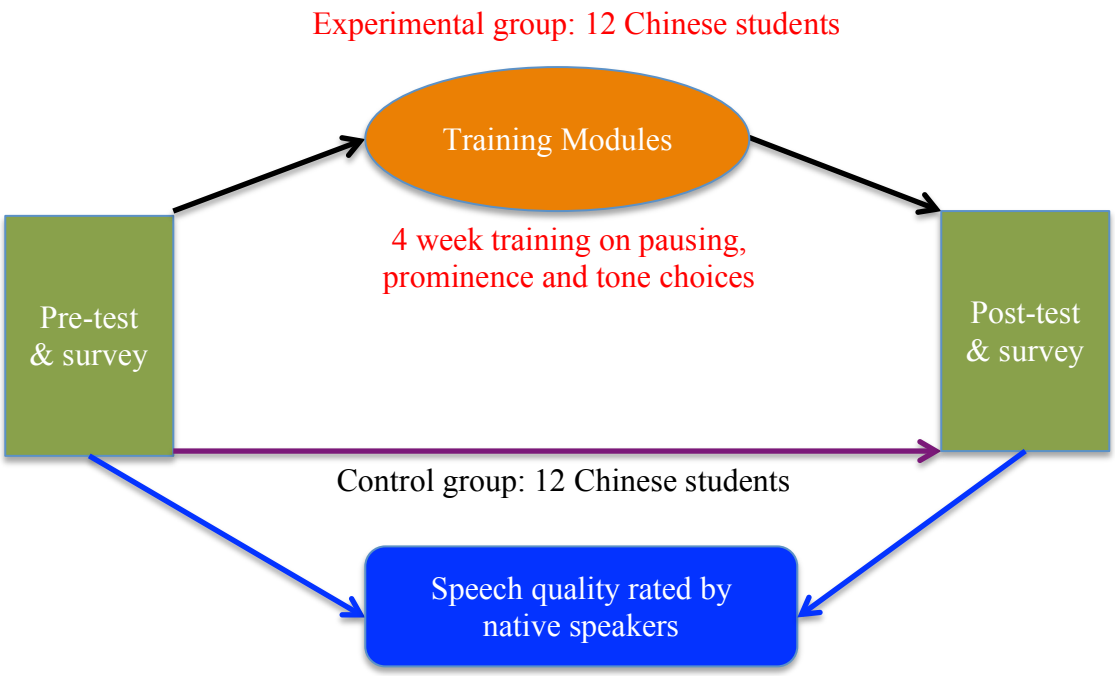


Figure 1. A visual representation of the study.

This study aims to answer the following questions:

1. What are the main challenges with regard to discourse intonation that Chinese speakers confront?

2. What are the effects of the online training modules in improving the comprehensibility, fluidity, accent, speaker's confidence and attractiveness of script-reading and spontaneous speech?

3. What are the students' perceptions of the training modules?

4. What is the relationship between intonation and speech quality?

1.3. Significance of the study

Knowledge attained from this study will contribute to the field of applied linguistics and pronunciation training, especially in development of computer assisted pronunciation training tools and systems for specific groups. Findings from this study will be analyzed

with regard to linguistic theories of the pragmatic functions of English intonation, and will have implications for institutions, teachers, and learners by demonstrating that L2 speakers' problematic intonation and oral speech could be improved using a well designed online training program. This study also has implications for language assessment and instruction in that it identifies the most important factors of intonation that affect L2 speech quality. Last but not least, the findings on the specific challenges of Chinese speakers with intonation would be useful in designing pronunciation training courses and materials for this specific group of students.

1.4. Outline of the dissertation

This dissertation is divided into five main sections. Chapter 1 introduces the research background, purpose and significance of the study. Chapter 2 reviews three topics on past research. It first describes what intonation is and summarizes its meaning and functions. Then it compares the uses of intonation by L1 and L2 speakers, especially those with Chinese as L1 background. And lastly it reviews past research and practices on teaching intonation to ESL learners in classroom and laboratory training. It also discusses their implications for my research project.

Based on past theories and practices, Chapter 3 puts forth four research questions, which survey and address the challenges that Chinese speakers face with discourse intonation in making oral presentations, in addition to revealing the contribution of intonation to speech quality. It proposes that a series of training modules that were discourse-based and aided by visualization technology would enhance the learners' intonation performance and improve their oral presentation quality. This chapter describes the three training modules, which focused on pausing, prominence and tone choice,

respectively. It also discusses the pre- and post- tests and surveys for data collection, as well as the data analysis methods to answer the research questions.

Then Chapter 4 reports on the analysis of the results, which provides responses to each research question. The results showed that the students began with problematic intonation performance and their English speech was less than satisfactory. The experimental group of students benefited from the training modules and improved their presentation quality. In contrast, the control group of students did not make progress. Furthermore, the findings indicated that human raters' impressionistic judgment of students' intonation errors is superior to acoustic measures in predicting learners' presentation quality.

Chapter 5 concludes the dissertation, discussing the research findings and their implications, limitations of the study, and suggestions for future research.

2. Literature Review

2.1. Definition and structure of English intonation

Intonation, as used in this paper, can be broadly defined as the use of suprasegmental phonetic features (*i.e.,* fundamental frequency or F0, intensity, and duration) to convey sentence-level pragmatic meanings in a linguistically structured way (Ladd, 2012, p. 6). A narrower definition of intonation is the use of contrastive pitch movements (*i.e.,* F0 only) over stretches of utterances to "communicate phrasing and discourse meaning" (Levis & Wichmann, 2015, p. 137). Intonation provides additional cues to express the full meaning that speakers wish to communicate. Even though native speakers might not be aware of these functions, they exploit intonation as an integral and meaningful part of language use in listening and speaking.

Before proceeding to the features of English intonation, it is necessary to introduce the structure of spoken utterances for meaning analysis and specify the terms. Since spoken language might be long and messy in structure, in contrast with written language, the unit of intonation is complex and cannot be easily defined by the syntactic structure. Halliday (1967) developed three systems to describe the structure of intonation: *tonality* is the segmentation of the discourse into units of intonation; *tonicity* is the identification of the most prominent word within the unit; *tone* is the specification of the contrastive pitch movement on the tonic syllable. The trio system had been adopted by many linguists, and various terms were used, such as *tone unit / group*, *intonation unit / group, interactional unit*, *phonological / intonational phrase,* and *intonational contour / tune* (Brazil, 1985; Crystal, 1969; Ford & Thompson, 1996; Ladd, 2012; Pierrehumbert & Hirschberg, 1990; Tench, 1996). In the language education field, the corresponding terms that had been

commonly used were *thought group, prominence (or stress* and *accent)* and *tone choice* (Chun, 2002; Gorsuch, Meyers, Pickering & Griffee, 2010).

Due to the pedagogical nature of my project, I follow the concepts used in the language education field to describe the features of intonation, namely *thought group* to denote groups of words that are segmented by *pausing* at the boundaries to form a thought. A thought group may consist of a word, a phrase, a clause or an entire sentence. I use terms *prominence* or *sentence stress* to refer to the most stressed syllable (word) in a thought group. The stressed syllables are usually said longer and with higher pitch and sometimes louder. And I use *tone choice* to describe the pitch movement on the prominent word. In statements and wh- questions, for example, English speakers raise their pitch on the stressed syllable, then fall on the unstressed syllables and drop to low at the end. In yes-no questions, on the other hand, they usually have a rising tone at the end.

2.2. Functions of English intonation

Researchers now agree that the meaning of intonation is determined more by functions or communicative intentions than by phonological rules, although different theories emphasize different aspects of functions (Brazil, 1985; Crystal, 1985; Chun, 2002). The intonation functions were generally interpreted from grammatical, attitudinal, sociolinguistic and interactional perspectives. Furthermore, relatively recent studies had expanded from sentence to the discourse domain (Chun, 2002; Wennertrom, 2001; Wichmann, 2000), after Brazil (1975; 1985) first proposed the full model of how intonation and discourse interact. There are three notions of discourse, firstly defined as “language above the sentence”, secondly defined as “language in use” and thirdly defined as “language

as it constructs or reflects social reality" (Wichmann, 2014, p. 2). The remainder of this chapter summarizes the functions of intonation.

2.2.1. Grammatical functions

The relevance of intonation to grammar has not been agreed by linguists since the beginning. Bolinger (1958) considered that "the encounters between intonation and grammar are casual, not causal" (p. 37). On the other hand, the grammatical function was central for Halliday (1967). One of the major grammatical interpretations was linking intonation with sentence types: statements, wh- questions, commands and exclamations were reported as having falling tones and yes-no questions having rising tones. Nowadays this classification seems oversimplified, and conflicting results have been found (Chun, 2002; Wichmann, 2000). Nevertheless, it still has pedagogical implications as a beginning step in intonation teaching practice.

Tench (1996) described the grammatical function of intonation in distinguishing syntactic patterns, organizing information, grading the pieces of information into major and minor information, and deciding the prominent information in each piece. It is understandable that he classified these functions as grammatical because his analyses derived mostly from sentences or clauses. However, I am inclined to agree with other scholars who subsume similar interpretation under discourse functions, when larger units of speech are employed and contextual factors are considered (Chun, 2002; Wennertrom, 2001; Wichmann, 2000). The discourse function is discussed in subsections.

2.2.2. Attitudinal functions

Using intonation to express attitudes has been considered a primary function in the traditional studies of intonation (Bolinger, 1986; Tench, 1996). Ladd (1990) acknowledged

that “few linguists would disagree with the proposition that intonation affects the interpretation of utterances through the interaction of very general meanings and broad principles of pragmatics” (p. 808), and yet the universal link between attitudes and intonation patterns is difficult to justify. For example, if you ask Russians or Hungarians or Romanians or Greeks how to ask a question like “Did you buy any books?” in comparable contexts, their response would differ consistently from that of speakers of Dutch or German or Italian or French (p. 815). Ladd (1990) suggested that the difference may be described in terms of “rules of grammar” or deeper psychological and functional principles. In other words, the attitudinal and emotional meanings need to be interpreted within the framework of pragmatics or discourse context.

2.2.3. Discourse functions

Brazil (1975) was one of the first linguists who proposed to “let discourse decide” the significance of English intonation features (p.4). According to his framework, each utterance is broken into tone group(s). Two concepts, *tone* and *key* are involved in the analysis of tone group. *Tone* refers to the pitch change of the tonic segment. *Key* is related to the pitch level of the tone group.

There are five tones in Brazil’s system: (1) fall-rise tone, or “referring” tone, marks the matter of the tone group as part of the shared, already negotiated, common ground occupied by the participants; (2) falling tone, or “proclaiming” tone, marks the matter as new; (3) Rising tone is the marked version of the fall-rise tone; (4) rising-falling is the marked version of the falling tone; and (5) low rising tone is neutral.

Speaker’s choice of key has discourse meaning, too. High-key has “contrastive” implication and it serves to “present the matter as if in the context of a closed set of

possibilities". Low-key has "equative" function and it serves to "exploit the possibility of alternative formulations of the same matter from the point of view of both participants" (Brazil, 1975, p. 21).

In addition, key choices have meanings at places of utterance boundaries. At utterance openings, low-key implies what is beginning is related with what has gone immediately before, while high-key marks separateness of the two parts. Utterance final key choice has an interactive implication as if the hearer is to continue. At utterance endings, high-key constrains the hearer to respond in some way, regardless of tone. But ending in low-key inhibits response, regardless of tone. Mid-key tone groups without proclaiming tone expects responses, and other mid-key tone groups set up no special expectations of response (Brazil, 1975, p. 31).

Brazil's theory highlighted the interactive significance of intonation, namely that speakers continuously assess the discourse and choose one intonation pattern over another to achieve discourse coherence. The discourse approach and his interpretations ascribed to speaker's key and tone choices have continuously influenced the subsequent pronunciation and intonation research, training and classroom practice around the world. However, Brazil mostly used "invented (speech) examples", and relied on his own judgments of the speaker's intentions to illustrate the functions of intonation (Couper-Kuhlen & Selting, 1996; Wichmann, 2000). The validity of Brazil's theory and the generalization to authentic data needs to be explored, and the interpretations from the speaker's perspective should be included.

2.2.4. Sociolinguistic functions

Sociolinguistic functions (also termed indexical functions) mean that intonation establishes contrasts indicating the speaker as members of different social groups (*e.g.,* age, gender, socio-economic groups). Loveday (1981) investigated the sociocultural correlations of pitch difference between male and female speakers of English and Japanese. He found very clear gender-based intonational differences in the expression of politeness in both Japanese and English; for Japanese speakers, men adopted a low pitch to emphasize their masculinity and women adopted a high pitch level to signal a stereotypically female role of being very "voluble, decorative and feminine". However, it would be far-fetched to generalize the gender differences among L2 speakers based on Loveday's (1981) study, considering the small sample size, oversimplified materials and the uncontrolled L2 speaker variation, especially that the speakers ages varied and their English proficiency was "impossible to determine" (p. 76).

2.2.5. Interactional functions

Couper-Kuhlen and Selting (1996) proposed an interactive perspective and suggested adopting the following principles to future prosody research:

1. Give priority to the analysis of naturally occurring talk.
2. Treat the data as an integral part of the context in which it occurs.
3. Treat the data as emergent in the real time of ongoing interaction.
4. Ground analytic categories in the data itself.

These principles are largely consistent with the discourse approach, namely that thought group boundaries depend on phonetic cues and need to be subject to repair based on the interactive needs of the participants; intonation function is contextualized; and there is a

necessity of including participants' own handling and interpretation of prosodic cues in context besides that of the researchers'.

To summarize, I view English intonation as an integral part of language use. The important features including *thought group*, *prominence* and *tone choice* play very important functions in conveying grammatical meaning, expressing attitudes and emotions, achieving discourse coherence, reflecting social identity and managing conversation turns. The meaning or functions of intonation are more vital than the phonetic rules for L2 English learners. In other words, it is a crucial skill for L2 learners to perceive the intonation contrasts, interpret the meaning of intonation in contextualized situations, and use appropriate intonation in their speech.

2.3. L1 and L2 speakers' use of intonation

2.3.1. L1 speakers' perception and production of intonation

While section 2.2 discusses the fundamental theories of intonation features and functions in L1 speech, they are mostly based on unnatural speech materials, small units such as clauses and sentences, and the interpretations are largely dependent on the linguists' or researchers' personal perceptual judgments. This sub-chapter adds four relatively recent studies that adopted empirical research methodology to study L1 speakers' perception and production of intonation, such as choosing authentic, discourse level speech; recruiting native speaking participants to perform perceptual and productive tasks; and including acoustic measurements of intonation, such as pitch, volume, speech rate, pause duration into analysis. That being said section 2.2 and section 2.3.1 combine to describe the use of intonation by L1 speakers in various linguistic environments.

Grosz and Hirschberg (1992) examined the correlation between acoustic prosodic features and native speakers' perception of discourse structure by comparing one group of participants who labeled the discourse structure of a sentence-final punctuation removed text while listening to a newscaster reading the news story, to another group of participants who labeled from the text alone. The labelers marked the beginning and ending discourse segments at the global level, and marked parentheticals, direct quotations, indirect reported speech and speaker attributions for reported speech at the local level. The researchers used WAVES speech analysis software to measure pitch range, F0 change, amplitude, intensity, speaking rate and pause duration of each intermediate phrase.

Correlation analysis results showed that the native speakers' markings of local level structures significantly correlated with the prosodic features. It suggested that that the labelers used prosodic variation, in addition to semantic cues to identify parentheticals and quoted phrases. For global structure, the beginning and ending segments appeared to be associated with pitch range and pause.

Furthermore, Grosz and Hirschberg found that prosodic factors could be used to predict both global and local discourse structure as identified by the labelers successfully 86% to 97% of the time. However, the relationship between structure and intonation is complicated. A discourse structure may be signaled by several intonation features that possibly interfere with each other, and the observed values may vary. A more complex model is needed to account for the relationship between intonation and discourse structure.

Ford and Thompson (1996) utilized two 20-minute talks excerpted from face-to-face, multiple party conversations to analyze the relationship among syntactic completion (*i.e.,* a complete clause), intonation completion (*i.e.,* perceptual judgments of the finality of an

utterance) and pragmatic completion (*i.e.,* a complete conversational action) that were marked separately.

Their study provided a clear understanding of the convergence of intonational, pragmatic and syntactic completions, and how interlocutors utilized them in a combined manner to signal and perceive turn completion in natural discourse. Specifically, 71% of the total speaker changes occurred at places where the three completions converged. In addition, intonational completion was nearly always syntactic completions as well (98.8%), although the reverse was not the case (53.6%). In other words, this study showed that intonation played a major role in native speakers' determining of which syntactically complete utterances were being projected as complete pragmatic units.

Besides research on the perception of intonation, a few studies focused on native speakers' production of intonation. Pickering et al. (2009) investigated whether and how speakers mark punch lines prosodically in humorous narratives. Two male and eight female undergraduate students were recorded performing two jokes. Through instrumental measurement of pitch, volume, rate of speech, length of pauses, and voice quality, they found that punch lines were often delivered at a lower pitch and slower speed rate than the preceding text. Although the finding was contradictory to the hypothesis that punch lines would demonstrate higher pitch to mark the saliency, Pickering *et al*. interpreted that punch lines typically occurred at the end of the text, thus the paratone structure that ended with a low pitch took precedence over the rhematic character of the punch line.

Wichmann (2000) conducted systematic analysis on the intonation features of "topic resets" in The Spoken English Corpus, consisting of 53 naturally occurring monologues, both scripted and unscripted in an accent close to RP British English. In addition to the high

pitch start, low pitch end, Wichmann noticed that "topics" did not parallel paragraphs for all types of text. Skilled readers frequently overrode such paragraph divisions in favor of more subtle interpretations of the text. Thus, a number of major tone group boundaries were not at the end of a sentence, but at some other syntactic boundary, usually a clause. Likewise, the signal of finality was not associated exclusively with the end of orthographic sentences. These closures occurred at points where the utterance was at least potentially complete – such as at the end of a main clause or before a final adverbial phrase or clause.

The above four studies jointly revealed the significance and complexity of intonation in L1 speakers' perception and production of natural discourse level speech. L1 speakers used intonation as perceptual cues to understand and even predict discourse structures (Grosz & Hirschberg, 1992), manage conversational turns (Ford & Thompson, 1996) as well as to convey the threads of the information structure of the discourse in their speech (Pickering et al., 2009, Wichmann, 2000). On the other hand, these studies also showed that neither the auditory perceptual nor acoustically measured intonation features matched exactly, although they largely converged with the semantic or syntactic structure in natural discourse.

These empirical studies affirm the importance of using natural discourse level speech in intonation research and practices. Using unnatural or oversimplified material may provide misinformation about intonation to L2 learners, so that the knowledge could not be transferred to authentic communication. These studies also imply that L2 speakers need to learn the perceptual and productive skills of intonation, because the lack of them may cause difficulties to understand and convey the information structure of the discourse and manage

the interactional turns. The following sub-section examines this idea in further depth by reviewing studies on L2 speakers' perception and production of intonation.

2.3.2. L2 speakers' perception and production of intonation

An early study of Luthy (1983) found that L2 speakers made nearly ten times more errors than the L1 speakers in understanding the attitudinal meaning of non-lexical intonation signals in everyday English, *e.g.,* the hesitation signal [əː]. The longer the L2 students had studied English the fewer errors they made. However, residence in the U.S. did not automatically help the learners to understand the signals. This study confirmed the difficulty for L2 learners to interpret the attitudinal meaning of intonation without explicit instruction, though the intonation signals used in this study were de-contextualized sounds without any verbal or visual cues.

Another interesting study of Pickering and Wiltshire (2000) found that international teaching assistants' (hereinafter ITAs) realization of accent was different from L1 speakers. They compared the frequency and amplitude on stressed syllables of three male Indian TAs to that of three native speakers of American English, and found that the native speakers showed consistent increased frequency and amplitude on the accented syllables. In contrast, all three Indian TAs showed a lower frequency and lacked consistent amplitude change on the accented syllables. They attempted to explain that the lack of increase in amplitude of accented syllable may have resulted from transfer from Indian English speakers' L1s, but a conclusion was difficult to make based on such a small sample size.

2.3.3. Chinese speakers' challenges with intonation

Chinese students had been reported having difficulties with discourse intonation in a number of studies. Tyler (1992) observed the misuse of falling tones at utterance-medial

positions by a Chinese TA, which made his lecture extremely difficult for L1 listeners to follow. Anderson-Hsieh et al. (1992) found that L2 speakers' prosodic errors strongly affected their English proficiency test scores. Moreover, their study indicated that the prosodic errors had a greater influence on the pronunciation rating than the segmental and syllable structure errors. Wennerstrom (1998) reported that L2 speakers' master of paratone contributed to Chinese-speaking learners' speech comprehensibility and intelligibility. Those who used intonation at the discourse level in a similar way to native speakers received higher scores.

Pickering (2001) found that the misuse of intonation by Chinese-speaking TAs affected their teaching performances, and even made negative impressions on the L1 students in the class. Compared to L1 speaking TAs' use of rising tone and level tone to "establish common ground" and signal an incorrect student response, the Chinese-speaking TAs used a noticeably higher number of level tones, falling tones and fewer rising tones in similar contexts. Such tonal composition used by Chinese TAs increased the distance between the speaker and the hearer, and may be understood by the students as disinterest and lack of involvement. In addition, a mismatch between syntactic and prosodic cues was also found that may increase the processing load for the hearers.

A later research study by Pickering (2004) continued analyzing the use of pitch variation and pause to create intonational paragraphs by U.S. and Chinese TAs, and found that the Chinese TAs appeared to be unable to exploit key choices appropriately. In particular, the Chinese TAs' pitch ranges (approximately 100-200 hz) were smaller than the U.S. TAs (50-250 hz).

The empirical evidence presented in this chapter suggests that L2 speakers confront challenges with understanding and producing English intonation appropriately. They either neglect or misinterpret the intonational cues or use inappropriate intonation that hinders their speech comprehensibility and intelligibility, proficiency test scores, ITAs' teaching performance, and even negatively impacts their affective relationship with the American students.

2.3.4. The different suprasegmental systems of Mandarin and English

The reason why it is challenging for Chinese speakers to master English intonation is probably the different suprasegmental systems in Mandarin and English. English is often known as a stress-timed language that contains stressed and unstressed syllables, with quasi-uniform durations between consecutive stressed syllables, but Mandarin is syllable-timed where all syllables have quasi-uniform durations (Meng, Tseng, Kondo, Harrison & Viscelgia, 2009; Tseng, Su & Visceglia, 2013).

The primary acoustic feature pitch (F0) plays fundamentally different functions in Mandarin Chinese and English. First of all, Mandarin uses pitch to signal lexical contrast (Duanmu, 2007), while English uses it to convey post-lexical (*i.e.,* discourse or pragmatic) meaning (Ladd, 2012). Second, the F0 contours of lexical tones in Mandarin are syllable bound and align to the end of tone-bearing syllables, regardless of contextual differences (*e.g.,* speech rate and segmental makeups) (Xu, 1998; Xu & Wang, 2001). However, the intonation pitch contour of English is phrase bound, which is contextually governed by speaking rate and syllable duration (Silverman & Pierehunbert, 1990). Table 1 summarizes the prosodic differences in Mandarin and English.

Table 1

Prosodic Differences Between Mandarin And English By Lu and Miran (2016)

	Mandarin	English
	Lexical tone	Intonation
Pitch function	Lexical contrast	Pragmatic contrast
Prosodic domain	Syllable	Accented unit (e.g., prosodic word, phrase)
Tonal alignment	Syllable bound: F0 contours are aligned at the end of the tone-bearing syllable regardless of different speech rates or segmental makeups. (Xu 1998; Xu & Wang 2001)	Metrical structure: The alignment of intonation peaks with their syllables exhibits contextually governed variation. (Silverman & Pierrehumbert 1990)

A recent study by Lu and Miran (2016) examined the contextual effect (*i.e.,* speaking rate) on Chinese speakers' production of L2 English, and found that the speakers' L1 prosody transferred to their production of L2 English. They compared the L* + H tone by Mandarin and English native speakers and found the Mandarin speakers' L* and H tone targets aligned closely with the end of the respective tone-bearing syllable, regardless of speech rate, but the English speakers' production was highly variable and was influenced by speaking rate. In other words, the Mandarin speakers' English speech prosody had similar patterns in Chinese.

Zhang, Nissen and Francis (2008) analyzed the native Chinese speakers' L1 prosody affecting their English lexical stress production. They reported that native Chinese speakers who preserve syllable-timed rhythm used pitch (F0) as the dominant cue to signal English lexical stress, while native English speakers use all four acoustic cues (pitch, intensity, duration and spectral quality).

In addition to the similar findings on Chinese students L1 affecting their English lexical stress production, Mixdorff and Ingram (2009), and Tseng et al., (2013) reported that native Chinese speakers had difficulty reducing vowels in unstressed syllables also due to the syllable-timed rhythm in their L1.

This review chapter makes a point that on one hand, English intonation is a crucial tool for effective communication. On the other hand, it is extremely challenging for L2 learners, especially those L1 with different suprasegmental systems than English, such as Mandarin. Chinese learners of English may preserve their L1 prosody in their L2 productions, often resulting in inaccurate and incomprehensible speech. As discussed above, it has been generally observed that Chinese students frequently face challenges with appropriate placement and realization of stress, vowel reduction, as well as tone choice variations in specific contexts.

What can we do to help the Chinese students learn English intonation? Living in the U.S. does not automatically help all students. Researchers advocate recognizing the role of intonation in communicative competence and proficiency, and raising the pedagogic status of discourse intonation teaching (Chun, 1998; Clennell, 1997). The following sections of Chapter 2 review second language acquisition (SLA) theories, past research and practice of teaching intonation to improve L2 speech comprehensibility, on four aspects, *i.e.,* materials and instruction, visual feedback, and discourse-level input.

2.4. Teaching intonation to improve L2 speech comprehensibility

This project used SLA theories as a framework for creating the online intonation training modules. First, comprehensible input is necessary for SLA. Comprehensible input means one-way input that is both understandable and at the level of just beyond the current linguistic competence of learners. This scaffolding theory is referred to as $i + 1$, (Krashen, 1985; Krashen, 1994), and is similar to Vygotsky's (1962) *zone of proximal development*. Second, comprehensible output is also critical for SLA. Output serves four primary functions in SLA: 1) enhances fluency; 2) creates awareness of language knowledge gaps; 3)

provides opportunities to experiment with language forms and structures; and 4) obtains feedback from others about language use (Swain, 1995). Third, the learner – language content, and learner – instructor interaction facilitates SLA. Scaffolding structures such as modeling, repetition used by more advanced speakers provide support to learners, and enable them to function within their zones of proximal development (Long, 1985; Vygotsky, 1962).

Therefore, the online training modules should be designed to provide interaction that includes negotiation of meaning and functions of intonation, constructed via exchanges of comprehensible input and output. This goal could be achieved through a variety of strategies in the design, such as providing multimedia to facilitate comprehension and accommodate different learning styles, using a variety of genres to provide motivation, and creating activities that allow learners to practice new forms and functions.

The view on pronunciation teaching was pessimistic for several decades, and pronunciation instruction had been de-emphasized since the decline of audiolingulism in the 1980s, not to mention teaching prosody or intonation. A main reason for such negative view was because the traditional goal of getting rid of accent and achieving native-like speech was unrealistic and unnecessary (Derwing & Munro, 2009). As a matter of fact, not all accent problems cause communication problems. For example, Munro and Derwing (1995) found that even though the listeners rated some utterances moderately or heavily accented, they were able to transcribe them perfectly.

However, in the last decade or so, interest in teaching pronunciation has been revived. For example, a new peer-reviewed *Journal of Second Language Pronunciation* has just begun publication in 2015, and is dedicated to research into the acquisition, perception,

production, teaching, assessment, and description of prosodic and segmental pronunciation of second languages in all contexts of learning.

My project sets the goals of intonation training to improve learners' speech comprehensibility and fluency. Even though speech accent was one of the measurements of speech quality in the data analyses, getting rid of accent was not a priority of the training modules. The articles that I review in the following part of this chapter also share similar standards. Comprehensibility is defined as how easy or difficult it is to understand a given speech sample (Derwing & Munro, 2009); and fluency in the narrow sense or fluidity denotes fluid or smooth language use in this project.

Research has shown evidence that prosodic factors contribute to human impressionistic judgment of L2 speech quality. Warren, Elgort and Crabbe (2009) compared naïve native speakers' and experienced teachers' ratings of comprehensibility and nativeness of sentences read by 5 female Chinese learners of English. For the nativeness rating task, most of the segmental information was removed from the speech samples, while keeping prosodic features intact. Factor analysis revealed that prosodic information was important in the two rating tasks.

Kang conducted variance analysis of the 12 suprasegmental variables predicting the accentedness and comprehensibility ratings. She found that prosodic factors accounted for about 35% to 41% of variance in the two ratings, and speakers tended to focus on different suprasegmentals depending on types of rating tasks. Overall pitch range contributed the most to the accentedness ratings, followed by proportion of stressed words, pause duration and articulation, while speaking rate factor mainly accounted for the variance of comprehensibility ratings. The statistical analysis results corresponded to the raters'

comments that they were especially aware of speakers' slow rate of speech, hesitant manner and monotonous intonation.

A later study (Kang, Rubin & Pickering, 2010) included as many as 29 suprasegmental measures of rate, pause, stress and pitch and showed more promising results that suprasegmental features accounted for over 50% of the variance in naïve raters' ratings of the oral proficiency and comprehensibility of NNS (non-native speakers') speech.

Lima (2015) conducted a qualitative analysis on the contribution of suprasegmentals to speech comprehensibility ratings in her dissertation. She employed verbal (think-aloud) protocols and examined the features that trained native listeners focused on when judging the ITAs' speech comprehensibility. Lima did not find clear trends in her data of the training raters' assessments and comments regarding specific suprasegmental and non-suprasegmental scoring. Her study suggests "that even raters with linguistic background are unreliable in how they rate specific pronunciation features in speech" (p. 119).

These findings of different features' contribution to the perceptual ratings have implications for the development of intonation training programs. Researchers should consider the relative weights of individual suprasegmentals and design training exercises focusing on prosodic factors that are directly related to speech comprehensibility, although the reliability of rater's judgment and the relative salience of individual intonation features remain to be further examined.

2.4.1. Materials and instruction

Hurley (1992) found that intonation use was important in conversational management and conversational support. He examined the pragmatic characteristics of intonation through contrasting the intonation realization by L1 and L2 speakers in different languages and cultures. Based on the findings, he suggested including target-culture

videotaped materials of different types of interactions, such as films and TV series, which could help to enhance learner's awareness of using intonation in carrying out interaction. Technique wise, he suggested that L2 learners should interact with native speakers, compare how they differ from native speakers and act again to see if their new knowledge resulted in more native-like performance.

Thompson (2003) confirmed that academic speakers used intonation to organize their speech and help listeners to form a coherent "mental map" of the overall talk after comparing the phonological paragraphing in EAP (English for Academic Purposes) materials to real undergraduate lectures. However, the comparative study also revealed that the phonological paragraphs in the EAP materials were much shorter than in the authentic lectures. Her finding has important pedagogical implications that if teaching materials are unnaturally shorter and correspondingly less complex than the authentic lectures, learners will have unrealistic expectations of how much lecturers chunk the information through phonologically signaling, and may fail to deal with what happens in real lectures.

Gorsuch (2011) proposed providing intensive input that is relevant in topics and communicative functions to improve ITAs' speaking fluency. She trained 28 international teaching assistants by repeatedly and silently reading 500-word basic popular science texts while listening to a recorded model of the text by a native speaker in 20 sessions, 30-minute for each session. The participants conducted two teaching simulation presentations before and after the training. The results showed that the students significantly improved on their speaking fluency after the training with increased percentage of grammatically intact pause groups and decreased percentage of split pause groups. Although the relationship between perception and production is not straightforward, this study suggests that providing the

amount and type of input that the students need would propel their speaking fluency, since the repeated listening and reading activities automatize the students' word-level comprehension processes and allowed them to free up enough cognitive resources to attend to salient intonation features in the audio model.

Gorsuch recommended that the input must be "met" repeated times, and the students need to be able to ask questions about word pronunciation or meaning in the text. Therefore, she did not recommend this approach in an online environment in which the students may listen or read to texts only once. In addition, she emphasized the importance of relevant input that not only has to do with topics but also with communicative functions that the students need in language use.

Relevance of intonation training has been considered in designing exercises, other than input selection. Gorsuch, Meyers, Pickering and Griffee (2012) created audio and video recordings of real classes in their book for ITAs, and developed hands-on excises for teaching functions, *e.g.,* giving instructions, making announcements, defining and explaining terms and asking questions. Some audios from their book were used in my training modules, *e.g.,* interaction between a TA (teaching assistant) and students in a discussion session, because they are not only relevant with regard to topics but also communicative functions for the trainees, Chinese students who are studying or going to study in U.S. universities.

Another good resource for speech models is TED Talks that have been used as rich speech samples for pedagogical use and contextualized teaching of intonation. They serve as authentic resources for more advanced level learners, and are examples of engaging monologic speech (Scotto di Carlo, 2014). McGregor, Zielinski, Meyers and Reed (2016)

suggested starting with meaning making, first asking students what impression they get from a TED speaker, and following with "how does the speaker do this so effectively?" to raise the students' awareness of the discourse intonation features, and motivate the students to imitate the speaker.

Explicit instruction would be effective for improving L2 speakers' pronunciation. Derwing, Munro and Wiebe (1998) compared 3 types of pronunciation instruction: at segmental level, at suprasegmental level and no specific pronunciation instruction to the control group. Their study confirmed the effectiveness of segmental and suprasegmental instructions that significantly improved the L2 learners' performance on reading aloud sentences, but only the suprasegmental group showed significant improvement in comprehensibility and fluency on extemporaneous narratives.

While these researchers suggested including authentic input and revising teaching approaches to improve intonation, empirical study on the role of teacher's feedback on student's intonation production was scarce in the past studies. Two reasons that teachers usually do not give feedback on intonation are because (1) intonation is a subtle and complex system that is difficult to talk about and (2) teachers do not know enough about intonation to be able to explain to the students (Chun, 2012).

Earlier studies on intonation training often involve students getting auditory feedback by monitoring themselves and listening critically to achieve self-correction, even though learners generally lack such phonetic criteria and skills. As a result, tools that can generate real-time and comprehensible feedback on intonation are called for. The following section reviews studies that provide visual feedback on intonation.

2.4.2. Providing visual feedback

Computer technology, especially visualization programs, has advantages of illustrating the subtle intonation changes and displaying instant feedback of students' oral production. With the visual display and real-time pitch contour feedback, teachers are able to explain pitch change in a more apparent manner, and students can imitate model speakers and view how far their pronunciation deviates. Researchers had recognized its capability and continuously applied innovative technology and tested their training effectiveness from using synthesized sounds to natural speech, from sentences to discourse-based texts and integrating multimedia tools. A few relevant studies are discussed as follows:

De Bot and Maifert (1982) are two of the first linguists who adopted electronic visual feedback into language laboratory teaching. They made a short instruction videotape which showed visual graphs of pitch changes of computer generated speech samples, and found that the students not only improved their awareness of intonation, but also developed imitating ability to reproduce the intonation patterns immediately after watching the video. However, the students showed negative reactions to the use of synthesized sounds, insufficient meaningful examples and lack of practice activities, which suggested that a simple tape is not the final solution for learning intonation and much more work needs to be done.

In a comprehensive book about discourse intonation in L2, Chun (2002) suggested exercises for practicing intonation in different discourse types and with different functions. She presented examples of how the four basic function categories, grammatical, attitudinal, discourse and sociolinguistic functions are realized intonationally and how to use pitch visualization technology to teach them to L2 learners. This book may benefit language teachers greatly with its wide range of teaching-related research and pedagogical chapters. It

even has a CD allowing readers to hear some examples used in the book. A shortcoming is that many of the examples are not from natural discourse (Levis, 2004).

Hardison (2004) designed prosody training for English-speaking learners of French. The participants practiced on given sentences with real time pitch contours displayed. They also listened to native speakers' versions and compared their distance from them by seeing the overlaid pitch contours in contrasting colors. This study found that students made significant improvement on sentence prosody, and the prosodic improvement was generalized to the production of novel sentences. The participants expressed positive reactions towards the training that raised their awareness of intonation and increased their confidence in L2 speaking. Similarly, Putro (2013) reported successful application of displaying visual waveforms of pitch change and duration, to teach English sentence stress and intonation to Indonesian college students.

Ramirez Verdugo (2006) replicated previous research on the effects of computer-assisted prosodic training and designed a "multi-sensory" approach that includes auditory, visual and productive speech to train Spanish learners of English. The training materials consisted of 14 short scripted dialogues. *Speech Analyzer* software was used to provide visual feedback of pitch contours. The trainees interpreted and recorded the dialogues, and then compared the pitch contours of their recordings to the native speakers'. Acoustic analyses of the tonality, tonicity, and tone revealed significant improvement for the trainees who achieved intonation variety and an approximate to the native speakers' model after the training. In addition, the comprehensibility of the speech produced by the trainees increased in the post-test as judged by native speakers. In contrast, no progress for the control group

was found in either the acoustic analyses or the speech comprehensibility test. The trainees commented that the program had helped them increase their awareness of English intonation.

2.4.3. Providing discourse-level input

While the earlier teaching and research on intonation were sentence-based, Levis and Pickering (2004) noticed the deficit of this model, because it was inadequate in describing the reality of pitch movement at sentence boundaries in natural speech. So they compared the use of paratones and unit-final tone movement of native English speakers' reading lists of scrambled sentences and then the same sentences as part of discourse-level texts. Their study found that the speakers' pitch change and tonal composition were different in the two situations. Specifically, the speakers created consistent paratone structure and used more rising tones to create discourse cohesion in reading discourse text than in reading sentences.

Levis and Pickering's (2004) study inferred that the traditional sentence-based method for intonation training was insufficient, because it obscured the meaning or use, and provided misleading information about the intonation patterns in a larger discourse context. This finding coincided with Hardison (2005) who examined the use of visualization in discourse-level vs. individual sentence-level situations, and found that only the groups who received visual feedback with discourse-level input, but not those with sentence-level training, produced better transfer of improvement in prosody to natural discourse. These research findings suggest that expanding the context of speech materials to the discourse level would facilitate the effect of visual technology in intonation training, because they helped the learners to mirror authentic pitch patterns in speech.

Tanner and Landon (2009) also recognized the importance of using discourse-level materials and focused on providing students self-directed practice opportunities to train their intonation with minimal teacher involvement. In their design, students marked the pause,

stress and intonation of a passage, checked their judgments by comparing to answer keys and finally repeated after the native speakers' version of the passage. The students made significant improvement on their perception of pausing and stress and controlled production of stress, but not on the sentence-final intonation or overall speech comprehensibility. Furthermore, the students asked for more specific feedback other than answer keys, and they felt that it was difficult to imitate the native speaker without any feedback.

Lima (2015) conducted her dissertation project on pronunciation instruction and developed the first fully online training modules Supra Tutor to teach suprasegmental features, including word stress, rhythm and intonation. Twelve ITAs from different L1 backgrounds completed a four-week online training with Supra Tutor. The training activities consisted of instructional lectures, perception and production exercises, including using *Praat* software to analyze the intonation pattern. Although the online tutor did not achieve consistent effectiveness in improving the ITAs' speech comprehensibility, it did help four students make significant improvement. More importantly, this project showed that the online pronunciation instruction helped solve the lack of training teacher conundrum, and motivated the students to improve their pronunciation skills.

To summarize, pedagogical studies suggest use of relevant (Gorsuch, 2011; Hurley, 1992) and authentic (McGregor et al., 2016; Thompson, 2003) materials, offer explicit instruction on suprasegmentals (Derwing, Munro & Wiebe, 1998), and employ exercises replicating communicative functions (Gorsuch, 2011; Gorsuch et al., 2012). It is also meaningful from a pedagogical perspective to focus teaching on aspects that affect speech comprehensibility and fluency (Kang, 2010, Kang et al., 2010), because teachers and students can only devote limited time to pronunciation in or outside of the classroom.

Research on using visualization technology has shown that seeing real-time pitch contours and waveforms is an effective tool for learning intonation (de Bot & Mailfert, 1982; Hardison, 2004; Putro, 2013; Ramirez Verdugo, 2006), and the effectiveness would be even greater with discourse-level input rather than isolated sentences (Chun, 2002; Levis & Pickering, 2004; Hardison, 2005; Lima, 2015). Tanner and Landon (2009) also reflected that simply using discourse-level materials without providing more specific feedback such as real-time pitch contours to the learners was not enough.

This chapter reviews past teaching-related research on English intonation and concludes that two essential elements should be combined for designing effective training modules: select authentic input materials and design relevant activities based on a discourse approach, and using pitch visualization to provide feedback for perception and production. Any training that only focuses on either one of the two elements would be inadequate.

Generally speaking, the current technology has achieved sufficient advancement to meet the pedagogical requirements of authentic, discourse-level input. It is up to the researcher's and instructor's choice and maybe time and economic constraints to choose appropriate materials and multimedia. As for output, learners need to complete tasks that produce discourse-length speech that mirror authentic pitch patterns. The training activities should contain more than repeating the phonetic features of native speaker models, even with the feedback of pitch contour visualization. It is important for the learners to understand the meaning of intonation in addition to how to realize it in speech. Without the knowledge of the meaning of intonation, interpreting the pitch contour visuals would be difficult and incomplete.

These research findings guided my design of the training modules for this project that utilized natural discourse-length speech materials, contained contextualized meaning-making activities and provided instant, concrete and specific feedback. The description of the training modules is found in the following chapter, in section 3.4.

3. Methodology

3.1. Research design

A quasi-experimental design involving controlled before and after studies was used to answer the research questions presented in Chapter 1. The inclusion of a control group increases the reliability of the evaluation of intervention effects comparing to uncontrolled studies. Participants with similar characteristics were recruited to join either the experimental group or the control group of their own choices. The experimental group received the intervention (the training modules), but not the control group. Speech data and survey responses in both groups before and after the intervention were collected. The gathered data were primarily quantitative and some qualitative data were used to answer the research questions.

Research question 1 (what are the main challenges with regard to intonation that Chinese speakers confront) was addressed from two perspectives. One was from the Chinese speakers' point of view based on the students' reflective ratings, the other one was from the native speakers' side using the raters' judgment on the students' intonation errors. Research question 2 (what are the effects of the online training modules) was addressed by comparing the pre-test ratings to the post-test ratings on the holistic speech quality. To answer Research question 3 (what are the students' perceptions of the training modules), the students' ratings

on the training modules in the post-test survey as well as their responses to the open-ended questions in the training modules were analyzed. Research question 4 (what is the relationship between intonation and speech quality) was addressed by examining the relationship between the raters' ratings on the holistic speech quality and the intonation measurements. Both auditory and acoustic data were used for analyses.

3.1.1. Pilot studies

Two pilot studies were conducted before the main study. The first one was implemented to try out the test and survey questions. Three college students in China completed the pre-test and survey. Three raters in the U.S. rated their speech. The students' and raters' feedback was taken into account and three major changes were made to the materials. The most important revision was to conduct one-on-one interviews, instead of student-directed speaking tasks, as a means to collect spontaneous speech samples. The original task asked the students to describe a picture story and upload the recordings. However, the participants simply added a few transitional words to the given hints (phrases and clauses), which did not reflect their spontaneous speaking skills. So I decided to conduct an interview with each participant and ask him/her about his or her personal college experience. The interview proved to be a better way to collect spontaneous speech that was longer and livelier. In addition, I added brief Chinese translations of the task requirements in the tests and training modules, because two students expressed difficulties with understanding the tasks in English. Lastly, I embedded the audio files directly to Qualtrics surveys instead of storing them on SoundCloud and linking to the surveys. The popular formats of storing audios and videos on SoundCloud or YouTube did not function properly in China.

Another pilot study on the speech rating was employed after the completion of data collection and editing. Three native speakers rated the edited speech samples using revised rubrics to validate the methodology (see Chapter 3.2.2). The following sections describe the participants, intervention, materials and procedures of data collection.

3.2. Participants

3.2.1. Speakers

The target population was Chinese students who were planning or currently studying in U.S. universities. Although the students have been studying English for many years in their home country, the opportunities to speak English are mainly in the classroom with non-native speakers. Accordingly, they often feel that their speaking skill is inadequate for study or social interaction in English-speaking environments (Sawir, 2005).

I recruited participants from two sources. First, I worked with an educational consulting company IvyGate in China and sent my project and contact information to their clients by email. Thirty-four students signed up for the study. Considering that students could quit the study anytime, I wanted to recruit more members to ensure enough participants at the end. So I approached the Chinese Students and Scholars Association at the University of California, Santa Barbara, and sent the same information to the mailing list subscribers. Nine students agreed to join the study.

The students chose to join either the control group or the experimental group by themselves. Thirty students signed up for the experimental group and 13 students enrolled in the control group. However, 11 students decided to switch from the experimental group to the control group after the first week of training because they did not have the time for all the training sessions. Each participant received 120 RMB (approximately $20) incentive

payment via mobile money transfer. For the experimental group of students, the incentive was paid in installments as a reminder to complete each training module.

Twenty-seven students completed the required tasks. After excluding three students' speech samples for the poor audio quality, 24 complete datasets remained to be used. There were equal number of students (12 students in total, 2 male, 10 female) in the control and experimental groups.

The students reported their demographic information and English learning experience in the pre-test survey. The experimental group and the control group of students had similar characteristics in their age, student status, years of living abroad as well as their English speaking proficiency (see Table 2). They were students at the levels of high school (5, 3, the former number represents the number of students in the control group, and the latter means the number of students in the experimental group, the same hereinafter), college preparatory academy (1, 0), undergraduate (4, 3) and graduate school (2, 4) whose age ranged between 16 to 28 years old. The length of living in an English speaking country varied from one month to 10 years for the control group participants, and differed from one year to five years for those in the experimental group. According to their self-reported TOEFL speaking scores in the recent three years (10 students scored 17 – 24 out of a total of 30 for the control group, 9 students scored 19-23 for the experimental group), their speaking proficiency was at fair level using a scale of weak, limited, fair and good (ETS, 2017). Those who didn't report valid standardized test scores evaluated their speaking proficiency (2 intermediate for the control group, 2 intermediate and 1 advanced for the experimental group students) based on a scale of novice, intermediate and advanced levels (ACTFL,

2017). They were either living in the U.S. (6, 8) or China (6, 4) during the period of my study.

Table 2

Participants' Background Information And Number Of Students (In Parenthesis).

	Age group	Gender	Class year	Currently in U.S. or China	Lived abroad[1]	TOEFL speaking score
Control group (12)	< 16 (2) 16-20 (6) 21-25 (2) > 25 (2)	Male (2) Female (10)	High school 5) College prep (1) Undergraduate (4) Graduate school (2)	U.S. (6)	Yes (6) 1 month - 10 years No (6)	17-24 (10) N/A (2, intermediate)
Experimental group (12)	< 16 (1) 16-20 (5) 21-25 (2) > 25 (4)	Male (2) Female (10)	High school (3) Undergraduate (3) Graduate school (4)	U.S. (8) China (4)	Yes (8) 1 - 5 years No (4)	19-23 (9) N/A (2 intermediate, 1 advanced)

In addition, the participants rated their English skills (overall proficiency, speaking skill, pronunciation and prosody) on a scale from 1 to 5 (1 = poor, 5 = excellent). Figure 2 shows that the students' self-ratings fell between 2 and 3.22 (2 = fair, 3 = good). The control group students gave higher ratings to themselves on all four categories compared to the

[1] The majority of the students lived abroad for 1-2 years.

experimental group. The differences between the two groups were relatively big, ranging from 0.22 to 0.53.

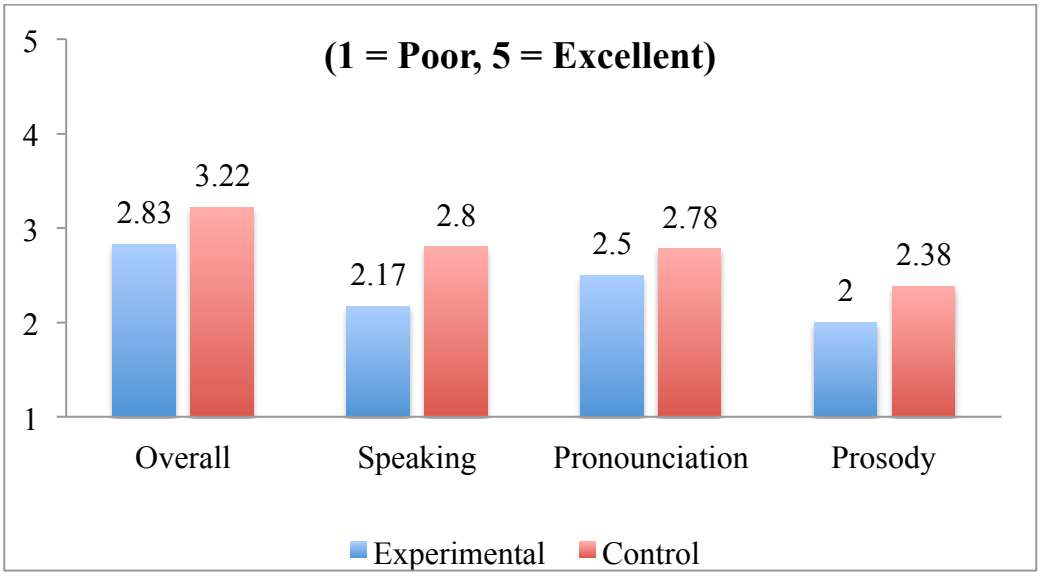


Figure 2. Participants' self-evaluation of their English skills.

They also rated their presentation skills (whether their speech was smooth, natural, whether the speakers were effective, confident and did not fear being misunderstood) using a five-point scale (1 = strongly disagree, 5 = strongly agree). Figure 3 shows that the ratings fell between 3 and 4. While the ratings on smoothness, effectiveness and confidence were about the same, the control group students showed more positive feelings towards the naturalness, and were much less fearful of being misunderstood, with a mean difference of about half point.

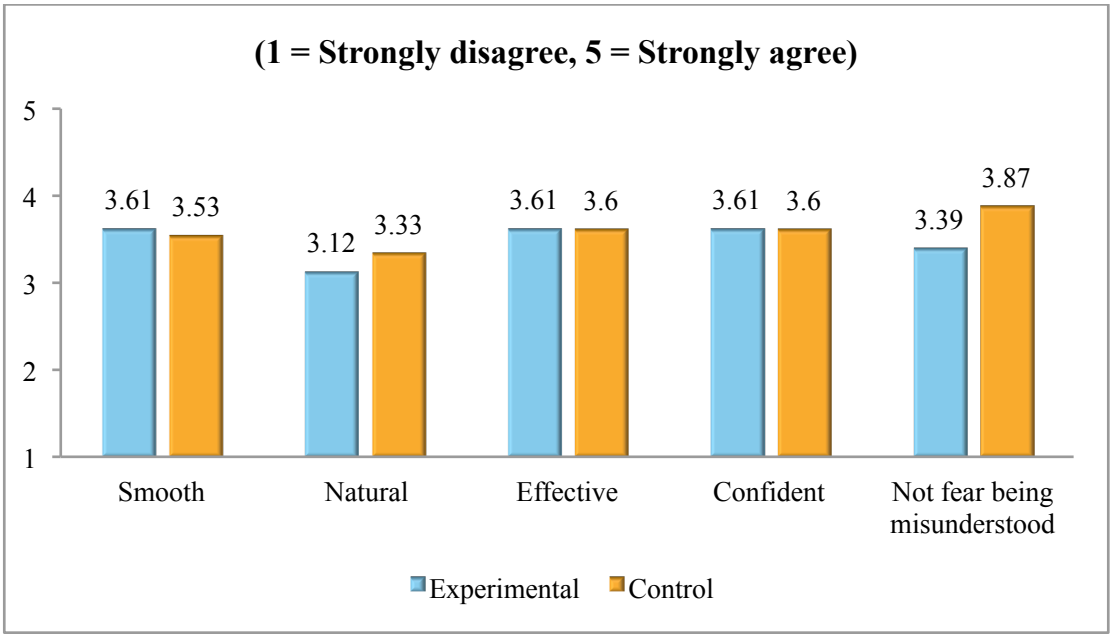


Figure 3. Participants' self-evaluation of their presentation skills.

Based on the students' self evaluations, both groups were not satisfied with their English or presentation skills. Based on their self-ratings, the experimental group of students gave lower ratings to themselves and appeared to be in greater need of improvement.

An analysis of the native speakers' judgment on the participants' pre-test spontaneous speech samples indicated that the experimental group and the control group of students began with about the same proficiency level. The average ratings on the spontaneous speech quality were Comprehensibility: $M_e = 4.23$, $M_c = 4.27$; Fluidity: $M_e = 4.31$, $M_c = 4.44$; Accent: $M_e = 4.67$, $M_c = 4.77$; Attractiveness: $M_e = 4.25$, $M_c = 4.25$; Confidence: $M_e = 4.27$, $M_c = 4.19$. Independent samples tests showed that the differences between the two groups were not statistically significant (all *p*s > .72) (see detailed description in section 4.2.1).

For the script-reading speech, the experimental group of students performed better in reading the script in the pre-test, as they received lower scores on the five categories compared to the control group (Comprehensibility: $M_e = 3.23$, $M_c = 3.78$; Fluidity: $M_e = 3.56$, $M_c = 3.97$; Accent: $M_e = 4.27$, $M_c = 4.97$; Attractiveness: $M_e = 3.73$, $M_c = 4.31$; Confidence: $M_e = 3.06$, $M_c = 3.97$). Independent samples tests showed the difference in speech accent was statistically significant, $F(18) = 0.823$, $p = .027$ (see detailed description in section 4.2.2).

3.2.2. Raters

Four native speakers (3 female, 1 male) were hired to rate the speech samples. They were graduate students majoring in Education (2), Geography and Chemistry at UCSB that could be considered as naïve raters of L2 speech and prosody by Chinese speakers. Two of them have taught English as a second/foreign language, although not for students from China. They described themselves as having heard a Chinese accent in English in casual conversation, but could not distinguish it from other accents in the rater's survey. The raters were selected because they represent the interlocutors the Chinese students interact with in U.S. universities, *e.g.,* classmates, teaching assistants or instructors. Each rater received $150 payment. The raters did not know or meet with the participating students.

3.2.3. Acoustic analysts

Three analysts, including myself worked together on the acoustic transcription. We all had experience in transcribing speech and identifying prosodic elements. I have conducted acoustic analysis for research before. The second analyst graduated with a Bachelor's degree in Linguistics. The third analyst was a professor of Education with expertise in applied linguistics and phonology. The second analyst received an $80 payment.

3.3. The training modules

Three training modules were created on Qualtrics.com, an online survey platform. Qualtrics is versatile for developing learning activities and collecting research data. It offers various question types (*e.g.,* text entry, multiple choice, uploading files, highlighting text) and features (*e.g.,* embedding audio and video files) that can be used to design learning activities. The collected data can be saved and downloaded in multiple formats (*e.g., excel, spss data).* The visualization software program used in the training modules was *Praat* (Boersma & Weenink, 2014), which is free and relatively easy to use for learners without much phonetic knowledge.

The three training modules covered thought group, prominence and tone choice over four sessions. Each training module focused on one topic. Module 3 on tone choice was split into two parts because the tasks required more time to complete. However, the students were allowed to finish the two parts consecutively. The design of the training modules was based on previous research, as discussed in Chapter 2. In the following section, I describe the materials and activities, and also discuss the pedagogical goals of the training modules.

The speech materials were audios taken from *Coursera* lectures, *Ted talks,* and university-level course discussion sessions (Gorsuch et al., 2010). The speech models were authentic and relevant, similar to what students would experience in real occasions (Hurley, 1992; Thompson, 2003; Gorsuch, 2011; Scotto di Carlo, 2014; McGregor et al., 2016; Levis & Pickering, 2004). Each module tried to stick to one theme that was related and appealing to university level students. Module 1 used speech materials that were pertinent to the beginning of a course, such as a faculty introducing a course, students making self-introductions and having conversations with classmates. Module 2 employed speech on

learning English and personal hobbies. Module 3 included speech on job interviewing and work experience.

All three training modules had similar structure consisting of four sessions: watching an instruction video and answering a series of questions; auditory perception activities; reading aloud activities; and producing script-reading and spontaneous speech. The estimated completion time for each module was 40 minutes or more, depending on the participants' pace. The design of the activities was research-driven and aimed to cover the communicative functions of the students (Gorsuch et al, 2012; Derwing et al., 2009).

3.3.1. Instruction videos

Each module began with a 5-7 minutes long lecture of video recorded PowerPoint presentation that covered four parts: the definition of, benefits of, challenges with, and guidelines for using each intonation feature. The content of the lecture was based on past linguistic theories as well as teaching-related research on English intonation, as discussed through Chapter 2. I used textual description, audio sentence examples and visual representations (*e.g.,* pitch contours) to cover the points in the slides. A professor of Education who is also a native speaker of English narrated the presentation. The learners were able to watch the video as many times as they liked. They could download the transcript documents to mark and take notes, and also download the audio files to play and analyze with *Praat* as exemplified in the video. Figure 4 shows a screenshot of the instruction video.

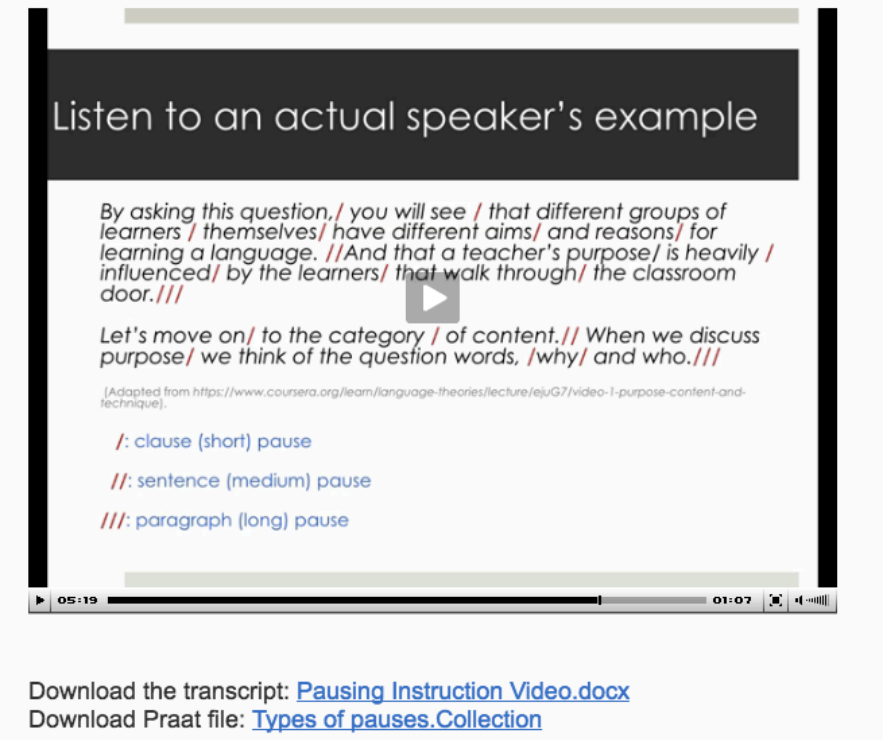
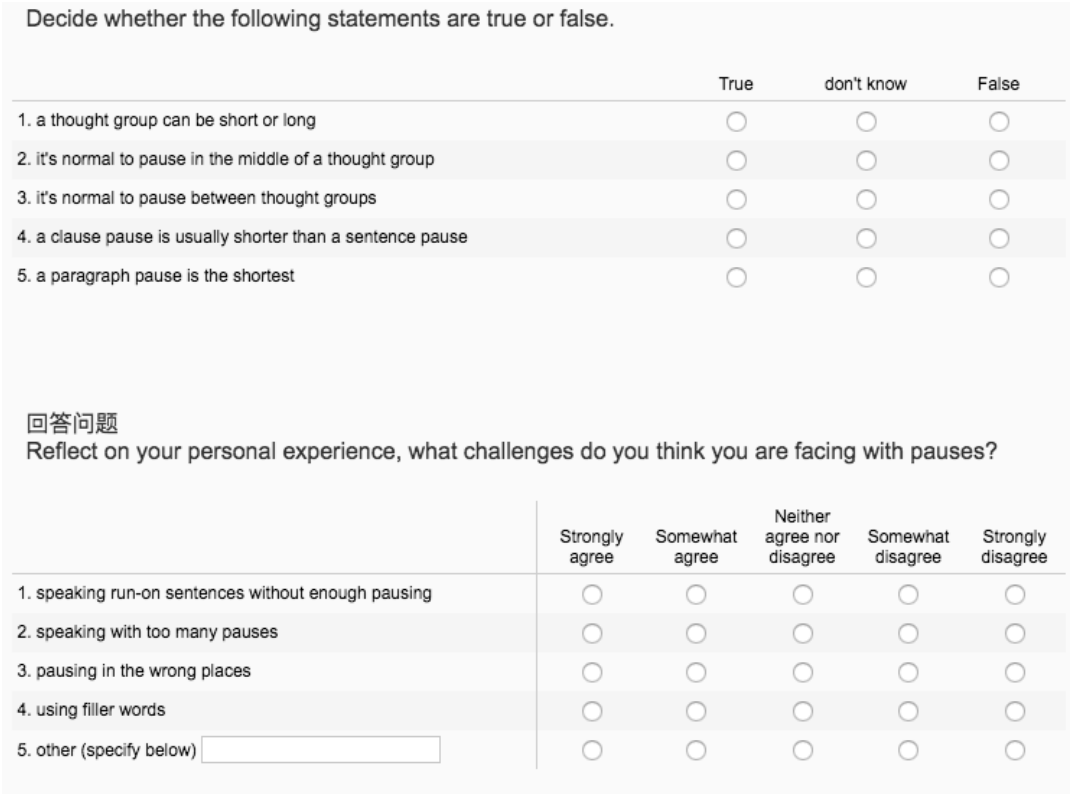


Figure 4. An example of the instructional video.

After watching the instructional video, the learners completed several questions that checked their understanding of the lecture. The questions included: writing down the definition of the terms in their own words; deciding whether the following statements are true or false. They also reflected on their personal experiences and rated the challenges they are facing with the intonation feature on a 5-point Likert scale. See Figure 5 for example.



Decide whether the following statements are true or false.

	True	don't know	False
1. a thought group can be short or long	○	○	○
2. it's normal to pause in the middle of a thought group	○	○	○
3. it's normal to pause between thought groups	○	○	○
4. a clause pause is usually shorter than a sentence pause	○	○	○
5. a paragraph pause is the shortest	○	○	○

回答问题
Reflect on your personal experience, what challenges do you think you are facing with pauses?

	Strongly agree	Somewhat agree	Neither agree nor disagree	Somewhat disagree	Strongly disagree
1. speaking run-on sentences without enough pausing	○	○	○	○	○
2. speaking with too many pauses	○	○	○	○	○
3. pausing in the wrong places	○	○	○	○	○
4. using filler words	○	○	○	○	○
5. other (specify below)	○	○	○	○	○

Figure 5. An example of answering questions based on the instruction video.

Watching the instructional video allowed the learners to activate their past knowledge and gain necessary new knowledge about the intonation features; writing the definitions and completing the true/false statement gave the opportunity to check their understanding of the knowledge; the reflective questions on their individual challenges revealed the learners' individual needs that also helped to answer my research question 1 (What are the main challenges with regard to intonation that Chinese speakers confront?) I also hoped the realization of the challenges would trigger the learners' interest in the following exercises.

3.3.2. Perception exercises

The second session of each training module consisted of two types of perception exercises. In the type I perception exercise, learners listened to given speech samples and

saw intonation marked on the transcript. Then they described their impressions about the intonation and interpret its functions by answering questions (See Figure 6), *e.g.,* what did you observe about the speaker's use of tones, using one or two tones, or a combination of three tones? What tone(s) did the professor use to emphasize new information? What assumptions did the falling tone suggest (choosing from I'm certain, I've finished…)?

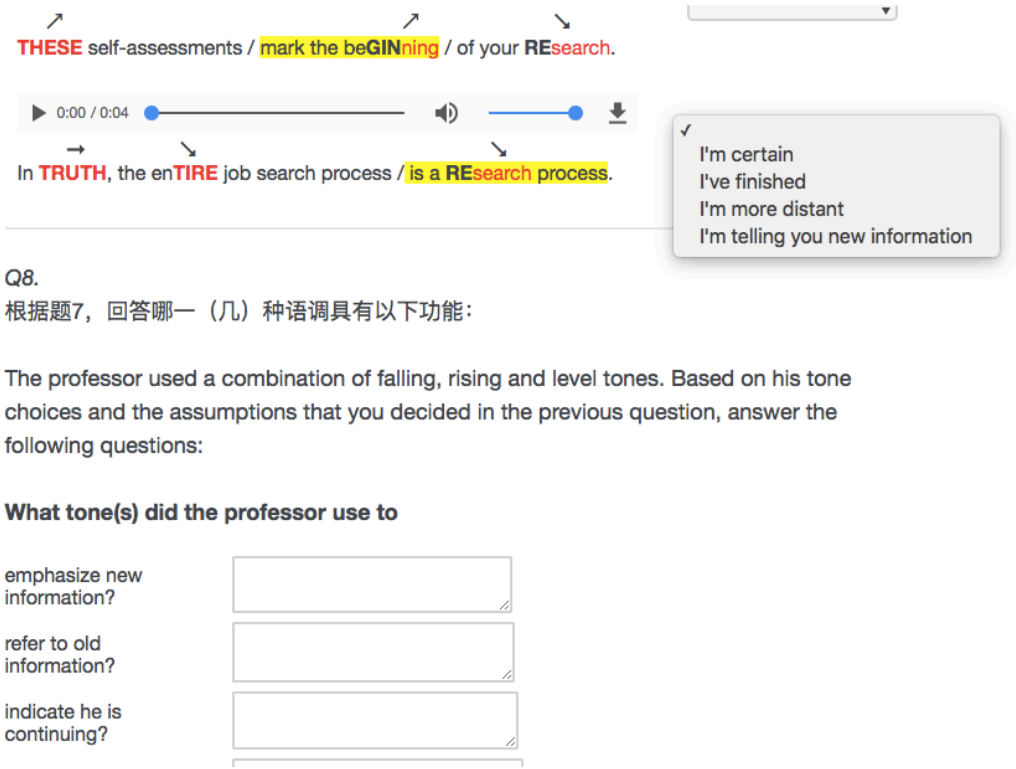


Figure 6. An example of type I perception exercise.

In the type II perception exercises, the learners listened to a new speech, and marked intonation features that they heard on the transcript using symbols. They marked / or // for short and long pauses, highlighted text to mark stressed and unstressed words in different colors, and drew arrow heads to mark rising, falling and level tone choices (see Figure 7 below). They either marked directly on the transcript (Modules 1 and 2), or wrote on a Word document and then uploaded it to Qualtrics surveys for the researcher to review.

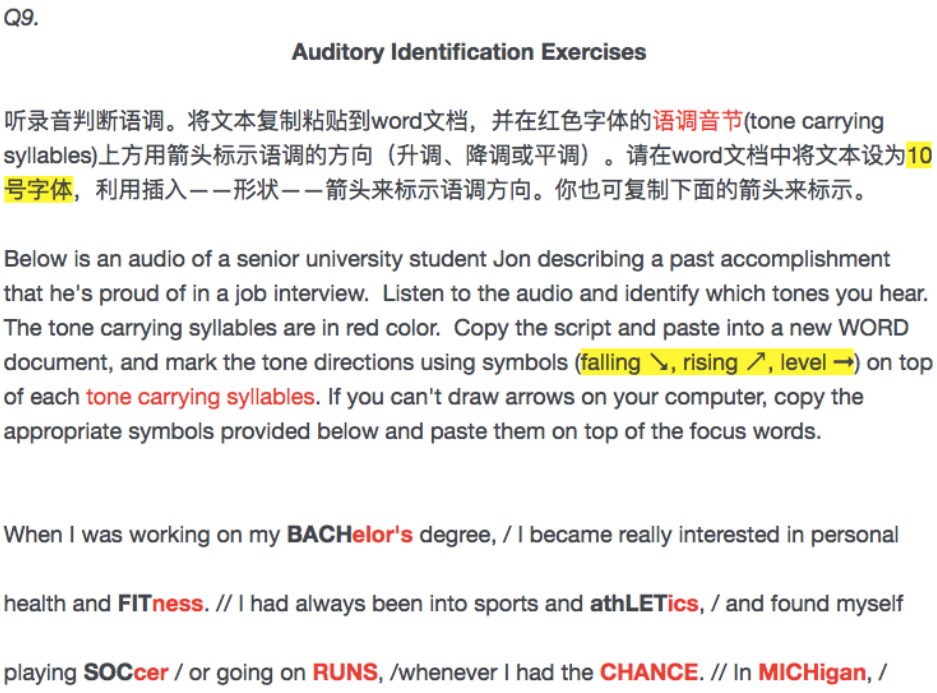


Figure 7. An example of type II perception exercise.

The students conducted activities with visual feedback in the type II perception exercise of Module 3. First, they watched a video that showed how to use *Praat* step-by-step, and then they used the software independently to see the visual graphs. Figure 8 is an example of the visual feedback that students got using *Praat*. The students were able to see the waveforms and pitch contours while listening to the speech, and they could measure the duration of pauses and pitch height of stressed and unstressed words, as well as pitch changes at transitional points. Because pitch contours were considered not very helpful to train pausing or prominence, Modules 1 & 2 did not require the learners to get visualization feedback or answer reflection questions, and yet instruction videos were provided for those

who were especially interested in creating visual representations.

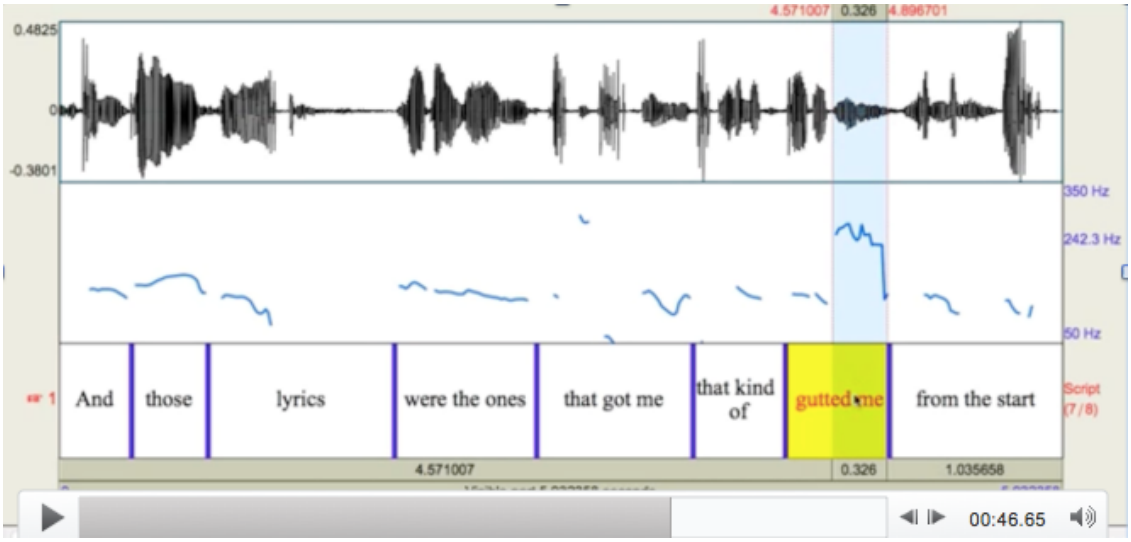


Figure 8. An example of visual feedback with *Praat*.

The students compared the visualized pitch movements to their auditory judgment, and wrote about the results of the comparisons by answering questions, such as, do they match with each other? Do the pitch contours in *Praat* help you to confirm or correct your auditory identifications? Or do they confuse you? Why do you think that *Praat* is/isn't helpful? (See Figure 9).

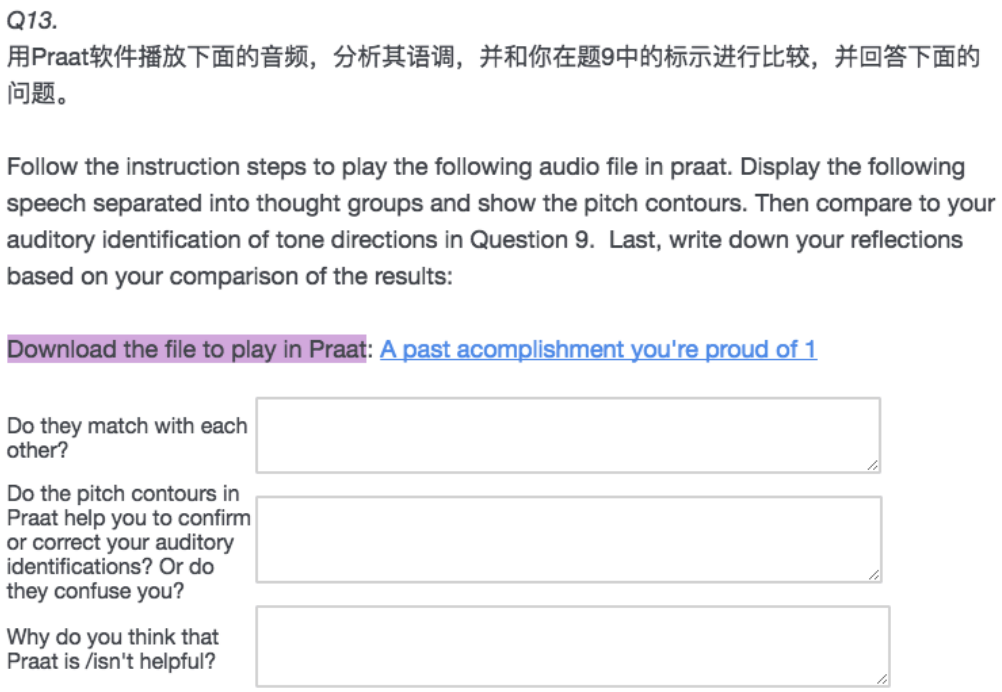


Figure 9. An example of reflecting on the visual feedback.

The type I perception exercise aimed to connect phonetic features of intonation to their meanings and functions. Seeing the marked symbols directed the learners' attention to the native speakers' intonation characteristics, and the impression questions linked to the discourse functions, *e.g.,* achieving discourse coherence, and expressing attitudes. They also acted as an example for the type II exercises in which the learners marked the intonation using symbols by themselves. The mark up activity aimed for the learners to focus on listening to the phonetic signals and deciding the intonation based on their judgment. The visual pitch contours provided feedback that was designed to help check or correct their perceptual judgment. And the reflection questions gave the learners an opportunity to think about and comment on the effects of using visual feedback, which were also valuable to answer Research Question 3.

3.3.3. Production exercises

The third part of each module included two types of oral production exercises: type I was reading aloud a given script with marked intonation (see Figure 10); and type II was marking intonation on a given script, and then recording themselves reading it. The two types of exercises directed the learners' attention to intonation. They were encouraged to practice multiple times and upload the most satisfactory recording. Similar to the perceptual exercises, the learners created pitch contour feedback by playing their recordings in *Praat* and commented on the comparison results and the helpfulness of the feedback in Module 3. In addition, the learners listened to an audio recording of the same paragraph(s) by native speakers, and then compared their uses of intonation and how the different intonation had affected the speech clarity, fluency or attractiveness. They were encouraged to view the pitch contours of the native speakers' in *Praat* for additional help.

Oral Production Exercises

根据标示朗读下文并录音。请注意重读和轻读。

Read the following paragraph with pauses and prominence as marked on the script. The focus words are **CAPITALIZED** and **bold** (大写加粗体). Record yourself with your cellphone or computer. Remember to use a combination of stress and unstress/reduction on appropriate words.

FIRST,/ what is **WORD** stress? // In **ENGLISH**/ we stress just **ONE** syllable in every word. // We say that **ONE** syllable more **LOUDLY**, / we say it **HIGHER**/ and we make it last **LONGER**. // **WHY** is word stress so **IMPORTANT**? //**WORD** stress is like a **MAGIC** key to **CLEAR** English. // As you **TALK** to people,/ **LISTEN** to the **NEWS**, /**WATCH MOVIES**, /you must **LISTEN** for and **PRACTICE WORD** stress. //It will **HELP** you to **UNDERSTAND** what you **HEAR** /and it will let **OTHERS** understand **YOU**.///

Q10.
将你在题8的录音和下文的录音比较，你们的重读 / 轻读处理有何不同？它们是如何影响表达的清晰度的？请回答。

Below is an example of a native speaker reading the paragraph in Question 8. Compare your recording to hers. Did you stress and unstress in the same places as she did? Does the difference of sentence stress affect the speech clarity? Write down your reflection

Figure 10. An example of oral production exercise.

3.3.4. Making speech

The fourth part comprised of one script reading speech and one spontaneous speech. The speech topics were relevant to materials in the previous parts of each module. For the script reading speech, the learners wrote down the text and they could mark the intonation on it if they wanted. Then they recorded themselves reading the script paying attention to their intonation. For the spontaneous speech, the learners only wrote down bullet points for each of the sub-topics to organize the speech, and then they gave the speech without reading the script word for word. After practicing multiple times, they uploaded the most satisfactory speech recordings. I listened to their weekly speech and provided written feedback to each student on whether they have met the training goals and how to improve on their intonation in the following modules.

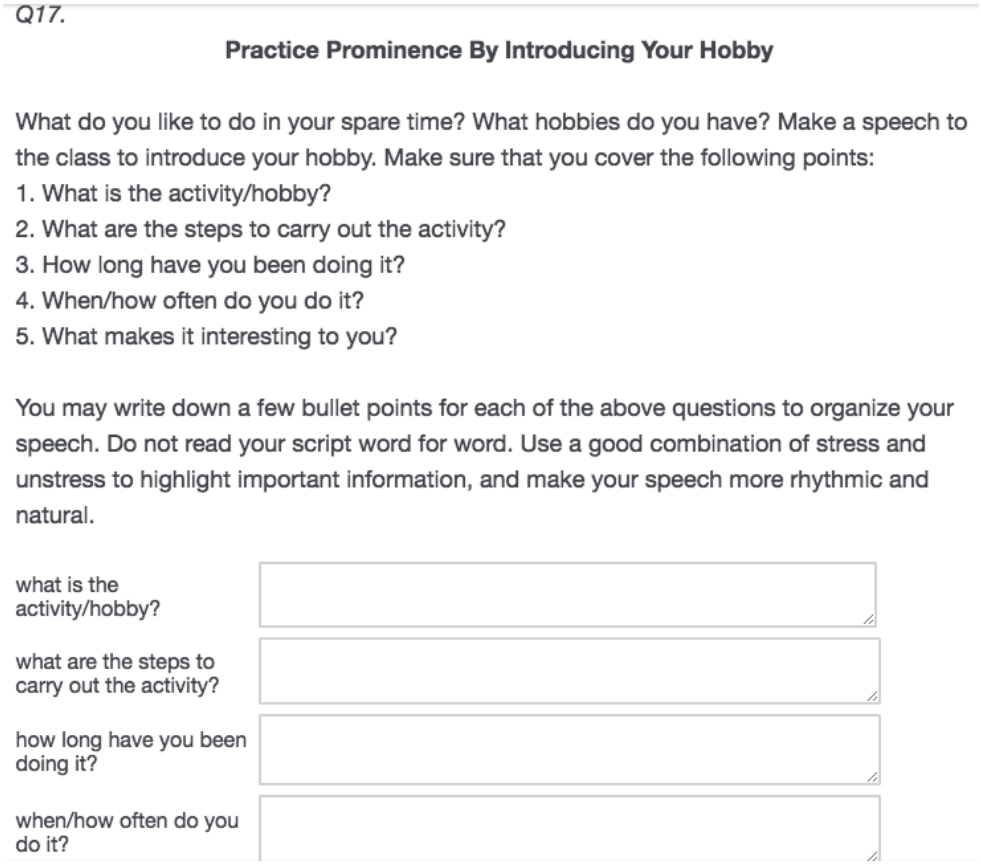


Figure 11. An example of making spontaneous speech.

The speaking tasks paralleled the language use of students at U.S. universities who were often asked to give a presentation on a course project or speak extemporaneously on a given topic. With the repeated practice and receiving instant visual feedback from *Praat* and invidualized written feedback from the researcher, these exercises provided training to improve their intonation and give more comprehensible and fluent oral presentations. Lastly, an open-ended question asked the learners to write down how they felt about the exercises, as well as their comments and feedback at the end of each module.

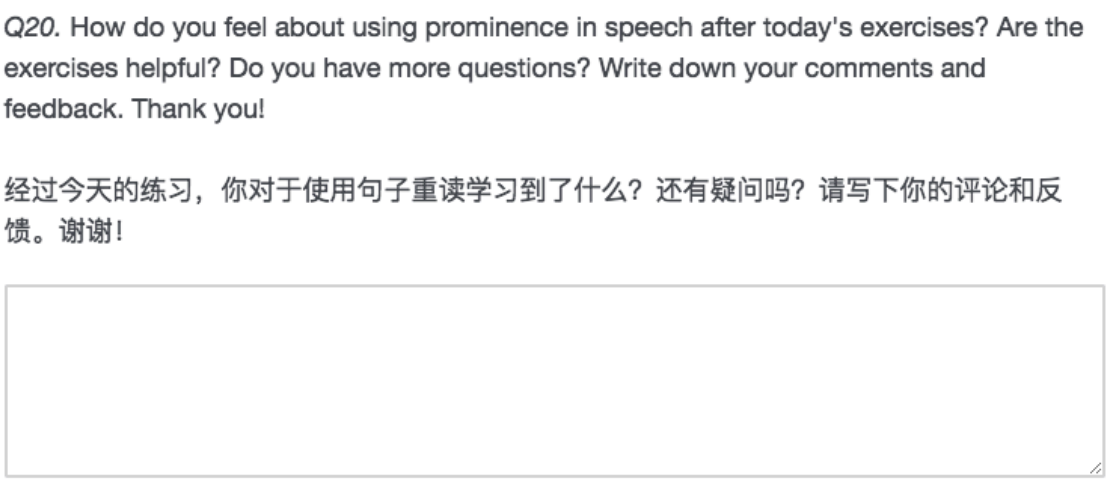


Figure 12. An example of an open-ended question.

3.4. Assessment materials

3.4.1. Pre- and post- tests

The pre- and post- tests measured the participants' perception and oral production of speech before and after the training. The pre-test consisted of 15 questions. Three contextualized speech materials were used. They were excerpts from a male teaching assistant led discussion session, including a monologue by the teaching assistant and a dialogue between him and the students, in addition to an audio recording of a female speaker telling her greatest musical moment taken from an interview.

The perception test asked learners listen to the audio recordings and complete activities on the selected sentences and paragraphs, including marking the pauses in the stream of speech, highlighting stressed words and identifying the tone movements. The terms were explained in the tests in case the participants did not have pre-existent knowledge about intonation.

Two speaking tasks were created to collect a script reading and a spontaneous speech. The students recorded themselves reading a 92-word long news report on an Apple annual event, which was transcribed from an authentic news program (See Appendix 1). The

spontaneous speech samples were collected via one-on-one online conferencing sessions. I interviewed each participant and asked him/her to make a speech. The purpose of interviewing was to make sure that the participants gave real spontaneous speech, rather than writing down the speech and reading it out, which could possibly happen in self-directed online exercises. The interview topics included two warm-up questions (Can you introduce yourself? How would you introduce your school to someone who doesn't know about it?) to establish rapport between the researcher and the participants, and two interview questions about their experience as a student: a) I'd like to know what you do on a day-to-day basis. Can you describe a typical day on campus? b) Where is your favorite place on campus? Why do you like it? The students chose one topic and gave a one-minute speech on it. They were given one minute to prepare before delivering the speech.

The post-tests repeated the whole set of questions in the pre-tests and interviews, with the addition of a new set of similar perception and production tasks. However, the new stimuli data were not used because they were beyond the scope of this study. Appendix 2 describes the interview protocols in the pre- and post- tests.

3.4.2. Pre- and post- surveys

The pre-test survey (see Appendix 3) consisted of 21 questions for all participants, which asked about their demographic characteristics, their study and use of English in daily life, English pronunciation courses that they took, as well as their self evaluation of English skills and presentation skills. The results were reported previously in section 3.2.1.

The post-test survey (see Appendix 4) examined the experimental group of participants' feedback towards the training modules that they had completed. Four sets of questions asked about the improvement on their awareness of English intonation and their spoken English, their evaluation of specific activities (*e.g.,* watching instruction videos), the

use of *Praat* software, their comments and suggestions. The results of the post-survey are reported in Chapter 4.3.

3.4.3. Speech data

The stimuli of this study consisted of 24 students' read and spontaneous speech samples in the pre- and post-tests. While acknowledging the effects of speech perception on its production, this dissertation mainly focused on the production part. Therefore, the participants' responses to the perception exercises were excluded from the analysis. As described in section 3.4.1, the script reading speech consisted of recordings of the students reading aloud a news report, and the spontaneous speech was the students' response to a selected topic in the video conferencing interview with the researcher.

The length of the raw read speech lasted from 33 seconds to 60 seconds, and the uncut spontaneous speech lasted from 32 seconds to 3 minutes 34 seconds. In order to manage the workload for the raters and acoustic analysts, excerpts were edited using *Praat* and *Audacity*. Recordings of the speakers reading the second half of the news report, and the first 30-40 seconds of the spontaneous speech were selected as appropriate stimuli. In order to test whether the selected excerpts were representative of the original speech, four sets of excerpts were compared to the original recordings in terms of speech rate, pause length, pitch range and change, and the differences were minimal. Three native speakers also did a pilot test with the excerpts and identified that they were enough to come up with ratings. As a result, the length of the edited read speech varied from 14 to 29 seconds, and the edited spontaneous speech lasted 24 to 45 seconds.

3.4.4. Rating scales

A total of 94 rating surveys were created for this study. Each survey contained a speech audio and five sets of questions (see Appendix 5). Raters listened to the recordings

and gave ratings using Likert scales. The surveys asked two types of questions: a) the raters' impression of the speech's holistic quality in Question 1; and b) the raters' evaluation of the speakers' speaking rate and specific intonation errors on pausing, prominence and tone choices in Questions 2-5.

Question 1 tested whether the speaker was easy or difficult to understand, spoke smoothly or in a choppy manner, had no accent or a strong foreign accent, was engaging or boring, was confident or hesitant using bipolar scales (*e.g.,* 1 = was easy to understand, 7 = was difficult to understand).

Question 2 measured the speech rate using a bipolar scale (1 = extremely fast, 7 = extremely slow). Question 3 examined whether the speaker spoke run-on sentences without pausing, spoke choppily with constant pausing, paused in the wrong places, and hesitated too much with fillers such as 'eh', 'um'. Question 4 assessed whether the speaker emphasized too many words, didn't emphasize enough words, and emphasized inappropriate words in a sentence. Question 5 evaluated whether the speaker spoke flat/monotonous speech, used too many falling and level tones and sounded distant, used too many rising tones and sounded unconfident, and used pitch changes in a confusing manner (*e.g.,* using falling tones when continuing and rising tones when finishing). Seven-point scales were used in Questions 3-5 (1 = strongly agree, 7 = strongly disagree).

3.5. Procedures

The data were collected in five stages. All stages were conducted online, as it was the most effective approach for this project. After the students were recruited, an email detailing the schedule (by week) and tasks was sent to them. An electronic consent form was also distributed to the students. Written consent was waived by the UCSB human subject

committee. Table 3 summarizes the stages of study implementation and data collection. The remaining part of this section describes the procedure of each stage in details.

Table 3.

Stages of Study Implementation and Data Collection.

Stage	Procedures
Stage 1: Pre-test, interview & survey	∗ All students completed the pre-test; ∗ All students had one-on-one interview with the researcher; ∗ All students completed the pre-training survey.
Stage 2: Training modules	∗ The experimental group of students completed the three training modules. ∗ These students received written feedback on their uploaded speech from the researcher.
Stage 3: Post-test & survey	∗ All students completed the post-test; ∗ All students had one-one-one interview with the researcher; ∗ The experimental group of students completed the post-training survey.
Stage 4: Auditory rating	∗ Four native speakers rated the speech samples for holistic speech quality and specific prosodic features.
Stage 5: Acoustic analysis	∗ Three analysts worked together to transcribe the speech samples. ∗ The acoustic properties were measured.

In the first stage, all the students received the links to access the pre-test and pre-training survey on Qualtrics. They were also instructed to schedule a ten-minute online conference for the interview. The students chose to have video or audio conference with me, via WeChat or QQ on cell phone or computer. This type of synchronous conferencing simulates a formal but comfortable face-to-face environment. The interviews were audio-recorded simultaneously using Voice Memos app and Audacity software in case either method did not function properly. One week was given to complete these tasks, although several students took a few extra days.

In the second stage, the experimental group of students completed the training in four to six weeks. The links to access the training modules were sent to the students in emails, one session at a time. The students used personal computers to view and do the exercises anywhere and anytime they liked. They were asked to complete the training modules 1 and 2 within one week separately. Then one week extension was provided because many students could not finish the tasks within the given time period. Next, two weeks were given for them to complete the two parts of training module 3.

Qualtrics automatically recorded the duration of completing each (part of) the training module, which lasted most often from 45 minutes to two hours. However, the recorded time did not reflect the actual time of learning because the students could pause, close and resume the exercises anytime, and they could be doing other work while keeping the webpages open.

As the course creator and instructor, I monitored the students' progress on Qualtrics. Once they finished a training module, I wrote feedback on their two oral speech samples. I evaluated the speech quality (*i.e.,* comprehensibility, fluidity, accent, attractiveness and

confidence), commented on their use of thought group, prominence and tone choice, and provided suggestions on how to improve them in the following practice. The feedback documents were sent to the students via email. The written feedback intended to enhance the students' learning outcomes, as well as encourage them to continue and finish the remaining training.

In the third stage, all the students completed the post-test and one-on-one interview with me. The experimental group students also took the post-training survey. The procedures were the same as the pre-test in the first stage.

In stage four, I created rating surveys for each speech sample and sent the links to the four raters separately. The links to rate the script-reading speech were put in one spreadsheet, and the links to rate the spontaneous speech were put in another spreadsheet. Within each spreadsheet, the survey links were mixed so that the raters could not tell whether the speech was from the pre-test or the post-test. The speakers' information was also concealed to prevent the raters knowing who made the speech. The raters listened to the speech samples and rated on the holistic quality and specific intonation errors (See section 3.4.4 for the rating questions).

I held an instruction session with the raters to familiarize them with the rating process by doing sample surveys, and clarified their doubts and questions. Research has shown that native speaking listeners' perceptual ratings are highly reliable measures even when they are not trained (Derwing & Munro, 2009). In addition to rating the students' speech, the raters also completed the pre- and post- tests.

In stage five, three analysts worked together transcribing the speech (see section 3.2.3). First, I listened and transcribed the speech samples using Brazil's (1985) framework,

and marked the pauses and prominent syllables. Then the second analyst reviewed my transcripts and made changes where she disagreed. The third analyst addressed the discrepancies between the first and the second analysts.

Based on the transcripts, I generated the linguistic and acoustic indicators with *Praat*: a) number of syllables and words, sample duration, b) number and duration of silent pauses, c) number and duration of filled pauses, d) number and F0 of prominent syllables, e) beginning and ending F0 of the final vowel in a thought group, and f) intensity of stressed vowels. These indicators were then used to compute the suprasegmental variables. Table 4 summarizes the calculation of each acoustic parameter following the methods used by Cucchiarini, Strik and Boves (2002), Kang (2010), Kang, Rubin and Pickering (2010, p. 557-558).

Table 4

Summary of Acoustic Measures

Measures	**Sub-measures**	**Descriptions**
Rate measures	Speech rate	Mean number of syllables produced per minute
	Articulation rate	Mean number of syllables produced over the total amount of time talking, excluding pause time
	Mean length of runs	Average number of syllables produced in utterances between pauses of 0.1 seconds and above
	Phonation time ratio	Percentage of time spent speaking as a Proportion of the total time taken to produce the speech sample
Pause measures	Number of silent pauses	Number of silent pauses per minute
	Mean length of pauses	Total length of pauses of 0.1 second or greater divided by total time of these pauses
	Number of filled pauses	Number of filled pauses (does not include repetitions, restarts or repairs)
	Length of filled pauses	Average of length of filled pauses

Prominence measures	Pace	No. of stressed words per minute
	Space	Proportion of stressed words to the total no. of words
Intonation measures	Pitch range	Calculated by measuring the F0 maxima and minima and producing range of prominent syllables in Hertz
	Mean F0 excursion	Average percentage of the beginning F0 relative to ending F0 of the final vowel in a larger intonation unit
Intensity measure	Mean intensity level	Average intensity of pitch accented vowels relative to the average intensity of the entire utterance (dB).

3.6. Data analysis

Statistical analysis methods were used primarily to answer the research questions. Both auditory and acoustic approaches were used in the quantitative analysis. The qualitative responses in the open questions of surveys and training modules were supplementary data.

Specifically, Research Question 1 (what are the main challenges with regard to intonation that Chinese speakers confront) was answered with descriptive statistics of the average ratings by the speakers and the raters. In response to Research Question 2 (what is the relationship between intonation and speech quality), two sets of multiple regression analysis were performed. The first analysis used the raters' auditory ratings on the holistic speech quality and the ratings on the specific intonation errors as data, to examine the amount of variances of speech quality that could be explained by intonation. The second analysis used the objectively measured acoustic data, and tested how they contributed to the raters' impression of the speakers' speech quality. In other words, these two sets of analyses identified the most salient perceptual and acoustic features of intonation affecting speech quality. Research question 3 (what are the effects of the online training modules) was addressed by running repeated measure t-tests, which measured the students' improvement

on their speech quality in the comparison of pre- and post-tests. The improvement was also compared between the experimental group and the control group to ensure the students' improvement benefitted from doing the training modules. Research Question 4 (what are the students' perceptions of the training modules) was answered with the descriptive statistics of the students' ratings in the post-test survey as well as their responses to the open-ended questions in the training modules.

3.7. Chapter summary

This chapter reports the methodology of the study. It begins with my design of the study and descriptions of the participants, including the students, the raters and the acoustic analysts. Then it describes the three training modules from the pedagogical goals to the activities. Next, it explains the materials covering Qualtrics, *Praat*, the tests and surveys, the speech samples, and the rating scales. After that, it summarizes the procedures of data collection and analysis in five stages. Last, it presents the data analysis methods to answer the research questions.

4. Results

This chapter reports on the data obtained from my project and the statistical analyses of the results that serve to answer the four research questions. It begins with a summary of the challenges that the students confronted with regard to intonation. Then it presents the effects of the training modules on improving the students' script reading and spontaneous speech, as well as their evaluation and feedback on the training. Lastly, it discusses the relationship between intonation and speech quality using both auditory and acoustic data. It also discusses the findings in relation to existing research.

4.1. RQ 1: Challenges with regard to intonation.

Both the Chinese students' self reflection and the native speaking raters' evaluation were used to display the challenges that the learners confronted with regard to intonation. The descriptive statistics of the students' and raters' ratings were compared to uncover the similarity and difference between the two sides' points of view.

The raters' evaluation on the pre-test spontaneous speech was used. Four raters listened to the students' speech samples and rated their specific intonation errors on 11 items using a 7-point scale (1 = strongly agree, 7 = strongly disagree). A total of 528 items were used in the following descriptive statistics (12 students * 11 items * 4 raters = 528).

The 12 experimental group students' reflected on their challenges with intonation using a 5-point scale (1 = strongly agree, 5 = strongly disagree), right after they watched an instruction video explaining one intonation feature in each of the three training modules. Eleven items were selected to match the raters' [2]. A total of 126 items were used in the following descriptive statistics (12 students * 11 items – 6 missing items).

In order to compare the two sides' perspectives, the students' ratings were converted from 5-point scale to 7-point scale (IBM Support, 2017). Figures 13, 14 and 15 display the average ratings on the 11 items judged by the students and the raters on challenges with regard to pausing, prominence and tone choice. On a 7-point scale, any value below four points represents that the students or raters agreed on the challenges. The smaller the value, the larger the challenge.

[2] In addition to the production of DI, the students' also reflected on their challenges with perception of intonation. The ratings on perception were excluded from analysis as they were beyond the scope of this dissertation.

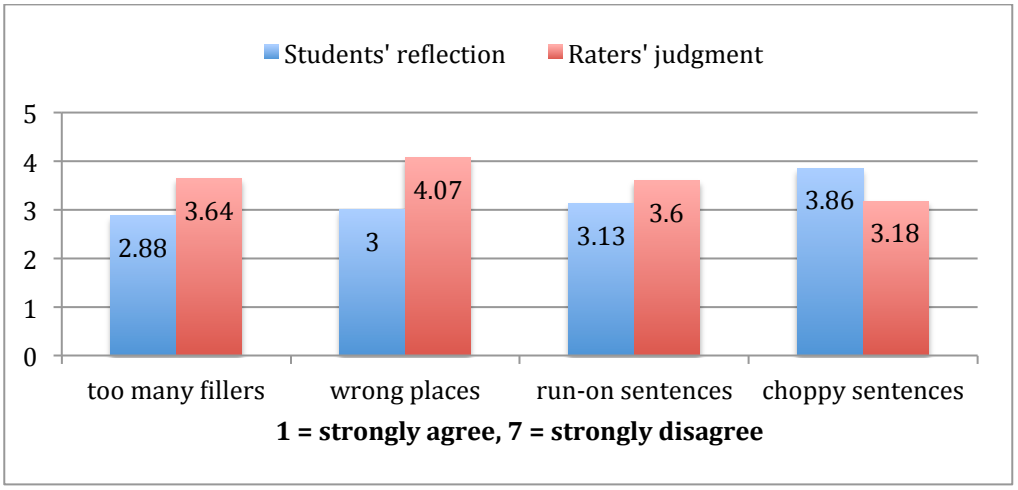


Figure 13. Students' reflection and raters' judgment on the students' challenges with pausing.

Figure 13 shows the students' and the raters' judgment on the students' challenges' with pausing on four items (using too many fillers, pausing at wrong places, speaking run-on sentences and speaking choppy sentences). The students' average ratings ranged between 2.88 to 3.18. In other words, they identified they had difficulty with all the four items with pausing. They reflected that they had bigger challenges with using fillers ($M_S = 2.88$), pausing at wrong places ($M_S = 3$), and speaking run-on sentences ($M_S = 3.13$) than speaking choppy sentences ($M_S = 3.86$).

The native speakers' average ratings ranged between 3.18 and 4.07. Compared to the students' judgment, the raters were more positive on three items, indicated by larger scores, using fillers ($M_R = 3.64$, $M_S = 2.88$), pausing at wrong places ($M_R = 4.07$, $M_S = 3$) and speaking run-on sentences ($M_R = 3.6$, $M_S = 3.13$). In contrast to the students' reflection, the raters identified their biggest challenge was speaking choppy sentences with too many pauses, which was the smallest challenge that the students identified ($M_R = 3.18$, $M_S = 3.86$).

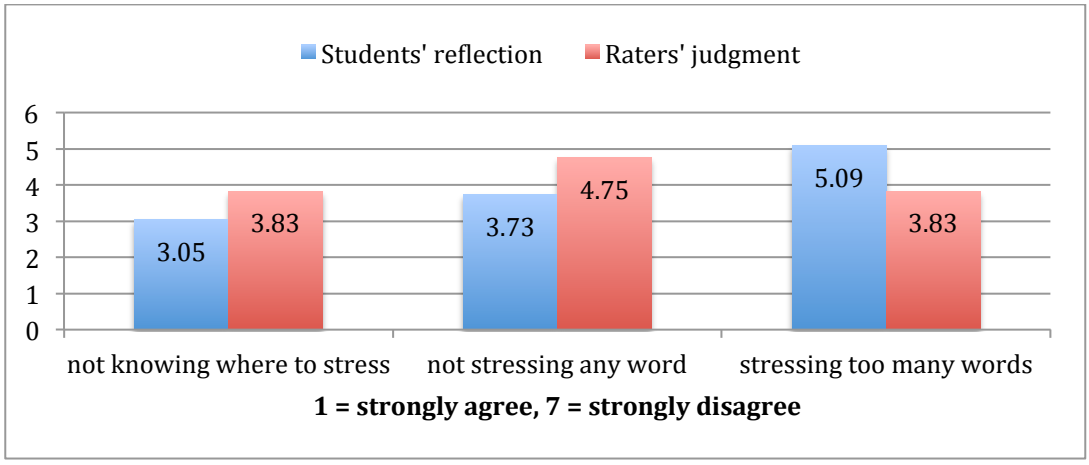


Figure 14. Students' reflection and raters' judgment on the students' challenges with prominence.

Figure 14 shows the students' and the raters' judgments on the students' challenges with prominence on three items. The students' average ratings were between 3.05 and 5.09. They identified the biggest challenge with prominence was not knowing where to stress (M_S = 3.05), they also had some difficulty with not stressing any word (M_S = 3.73). They somewhat disagree that they stressed too many words (M_S = 5.09).

The raters were more positive about the students' intonation production on two items, not knowing where to stress (M_R = 3.83, M_S = 3.05), not stressing any word (M_R = 4.75, M_S = 3.73). In contrast to the students, the raters agreed that stressing too many words (M_R = 3.83, M_S = 5.09) was a challenge for the students.

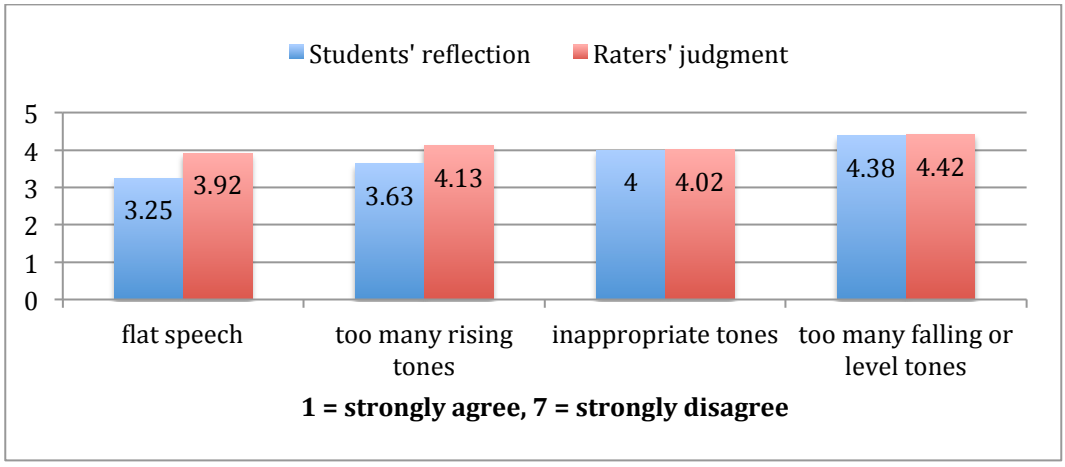


Figure 15. Students' reflection and raters' judgment on the students' challenges with tone choice.

Figure 15 showcases the students' and the raters' judgments on the students' challenges with tone choice on four items. The students' ratings ranged between 3.25 and 4.38. They agreed that their biggest challenge with tone choice was speaking flat speech without pitch variations (M_S = 3.25), they were also challenged by using too many rising tones (M_S = 3.63). They didn't agree with the challenges of using inappropriate tones (M_S = 4) or using too many falling or level tones (M_S = 4.38).

Similar to the other two types of challenges, the raters' scores were higher than the students on four items, ranging from 3.92 to 4.42. They judged that the students had a little difficulty with speaking flat speech (M_R = 3.92). However, they did not agree that the students had difficulties with using too many rising tones (M_R = 4.13), using inappropriate tones (M_R = 4.02) or using too many falling or level tones (M_R = 4.42).

In addition to the above 11 items, the raters judged the speakers' speech rate using a seven-point scale (1 = extremely fast, 7 = extremely slow). They considered the students' speech was slightly slow when speaking spontaneously ($M_{interview}$ = 4.56), but appropriate

when reading the script aloud (M_{script} = 4.02). Because speaking rate was not explicitly trained in the modules, the students did not reflect on this question.

4.2. RQ 2: The effects of the training modules

This question was addressed by running repeated measures t-tests, which examined the students' improvement on their speech quality in the comparison of pre- and post-tests. The improvement was also compared between the experimental group and the control group to ensure the students' progress benefitted from doing the training modules. The following sections report the results of the training effects on the spontaneous speech and script-reading speech separately.

4.2.1. Training effects on spontaneous speech

The data were raters' judgments on the students' speech quality. Raters listened to the pre- and post- spontaneous speech excerpts in random order, and judged the speech comprehensibility, fluidity, accent, attractiveness and confidence using a scale from 1 to 7, with higher values representing a bigger problem. In other words, receiving an average rating larger than 4 suggests the speech was less than satisfactory on that quality measurement.

The pre-test ratings showed that both the experimental group and control group had problematic speech, as the average ratings on the five categories were between 4.19 and 4.77. In addition, the differences between the experimental group and control group were minimal, indicating that both groups of students began with about the same proficiency level (Comprehensibility: M_e = 4.23, M_c = 4.27; Fluidity: M_e = 4.31, M_c = 4.44; Accent: M_e = 4.67, M_c = 4.77; Attractiveness: M_e = 4.25, M_c = 4.25; Confidence: M_e = 4.27, M_c = 4.19, see Figure 16). Independent t-tests showed that none of the differences between the two groups was statistically significant (all ps > .72).

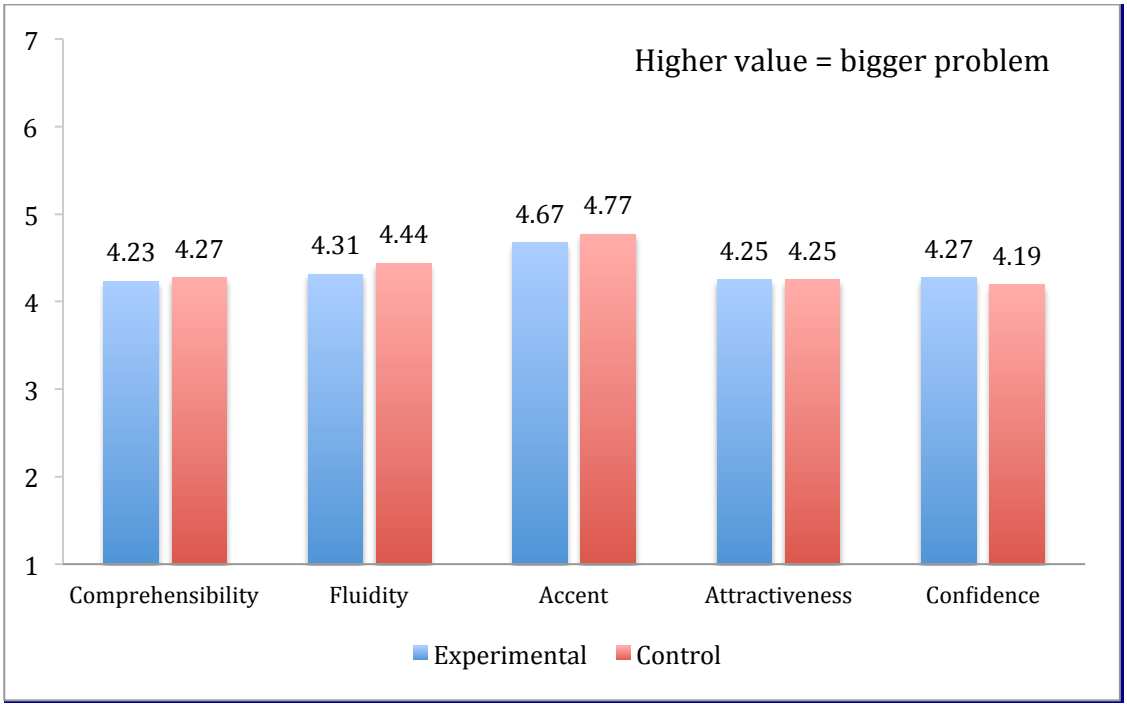


Figure 16. Ratings on the quality of spontaneous speech in the pre-test.

The post-test ratings presented a different picture showing a bigger variation in the five aspects. The average ratings ranged from 3.6 to 4.9. The experimental group received consistently lower ratings than the control group on all five aspects. The average ratings differed from 0.42 to 0.92, showing the experimental group of students' speech was rated more favorably than the control group in the post-test (Comprehensibility: $M_e = 3.6$, $M_c = 4.52$; Fluidity: $M_e = 4$, $M_c = 4.69$; Accent: $M_e = 4.42$, $M_c = 4.9$; Attractiveness: $M_e = 3.75$, $M_c = 4.17$; Confidence: $M_e = 3.65$, $M_c = 4.35$, see Figure 17).

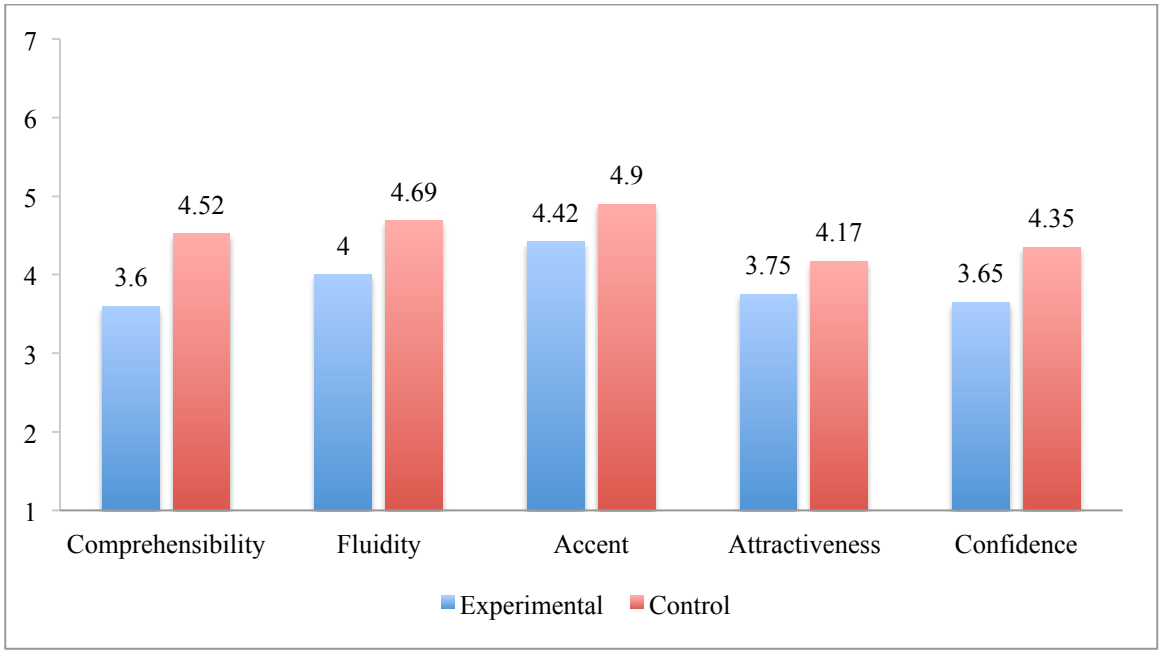


Figure 17. Ratings on the quality of spontaneous speech in the post-test.

The differences of the average ratings of both groups in the pre- and post- tests showed the training probably has helped the experimental group improve their speech quality. To further analyze the data, I ran repeated measures t-tests separately for the two groups on the five aspects of speech quality. Table 5 summarizes the test results.

Table 5

Repeated T-Tests On The Improvement Of Spontaneous Speech Quality.

		M	*SD*	*SEM*	*t*	*df*	*p*
Experimental Group	Comprehensibility	-0.63	0.78	0.23	-2.78	11	0.02
	Fluidity	-0.31	0.64	0.18	-1.69	11	0.12
	Accent	-0.25	0.64	0.18	-1.35	11	0.2
	Attractiveness	-0.5	0.87	0.25	-1.99	11	0.07
	Confidence	-0.63	0.73	0.21	-2.95	11	0.01
Control Group	Comprehensibility	0.25	1.08	0.31	0.8	11	0.44

Fluidity	0.25	1.07	0.31	0.81	11	0.44
Accent	0.13	0.65	0.19	0.66	11	0.52
Attractiveness	-0.08	0.87	0.25	-0.33	11	0.75
Confidence	0.17	0.81	0.23	0.72	11	0.49

For the experimental group, the difference in the ratings on comprehensibility for the pre-test (M = 4.23, SD = 0.84) and the post-test (M = 3.60, SD = 1.02) was statistically significant t (11) = -2.78, p = 0.02. The difference in the ratings on fluidity for pre-test (M = 4.31, SD = 0.92) and post-test (M = 4, SD = .92) was not significant t (11) = -1.69, p = 0.12). The difference in the ratings on accent for pre-test (M = 4.67, SD = .62) and post-test (M = 4.42, SD = .91) was not significant t (11) = -1.35, p = 0.2. The difference in the ratings on attractiveness for the pre-test (M = 4.25, SD = .85) and the post-test (M = 3.75, SD = .86) was not significant t (11) = -1.99, p = 0.07. The difference in the ratings on confidence for the pre-test (M = 4.27, SD = .89) and the post-test (M = 3.65, SD = .79) was significant t (11) = -2.95, p = 0.01.

For the control group, the difference in the ratings on comprehensibility for pre-test (M = 4.27, SD = 1.04) and post-test (M = 4.52, SD = 1.06) was not significant t (11) = 0.8, p = 0.44. The difference in the ratings on fluidity for pre-test (M = 4.44, SD = 0.93) and post-test (M = 4.69, SD = .69) was not significant t (11) = 0.81, p = 0.44. The difference in the ratings on accent for pre-test (M = 4.77, SD = .76) and post-test (M = 4.90, SD = .90) was not significant t (11) = 0.66, p = 0.52. The difference in the ratings on attractiveness for the pre-test (M = 4.25, SD = .71) and the post-test (M = 4.17, SD = .67) was not significant t (11) = -0.33 p = 0.75. The difference in the ratings on confidence for the pre-test (M = 4.19, SD = .72) and the post-test (M = 4.35, SD = .45) was significant t (11) = 0.72, p = 0.49.

In summary, the test results showed that the experimental group of students had made improvements on all five categories of the judged speech quality, which was indicated by receiving lower scores in the post-test evaluation (Comprehensibility: M_{diff} = -0.63; Fluidity: M_{diff} = -0.31; Accent: M_{diff} = -0.25; Attractiveness: M_{diff} = -0.5; Confidence: M_{diff} = -0.63. See Table 5). Among them, the improvements on speech comprehensibility and speakers' confidence were statistically significant. The improvements on speech fluidity, accent and attractiveness were not big enough to reach statistical significance.

In contrast, the control group did not show improvement on speech quality, except the minimal progress on speaker's attractiveness. (Comprehensibility: M_{diff} = 0.25; Fluidity: M_{diff} = 0.25; Accent: M_{diff} = 0.13; Attractiveness: M_{diff} = -0.08; Confidence: M_{diff} = 0.17. See Table 5). The average differences from the post-test to the pre-test were small, and none of them was statistically significant.

4.2.2. Training effects on the script reading speech

The pre-test ratings showed that the experimental group of students performed better in reading the script, as they received lower scores on the five categories compared to the control group (Comprehensibility: M_e = 3.23, M_c = 3.78; Fluidity: M_e = 3.56, M_c = 3.97; Accent: M_e = 4.27, M_c = 4.97; Attractiveness: M_e = 3.73, M_c = 4.31; Confidence: M_e = 3.06, M_c = 3.97, see Figure 18). Independent t-tests showed that the difference in speech accent was statistically significant, F (18) = 2.41, p = .027. Furthermore, the script reading speech quality for both groups was judged adequate on most of the ratings, which were smaller than four points, except the ratings on the accent of both groups, and attractiveness of the control group.

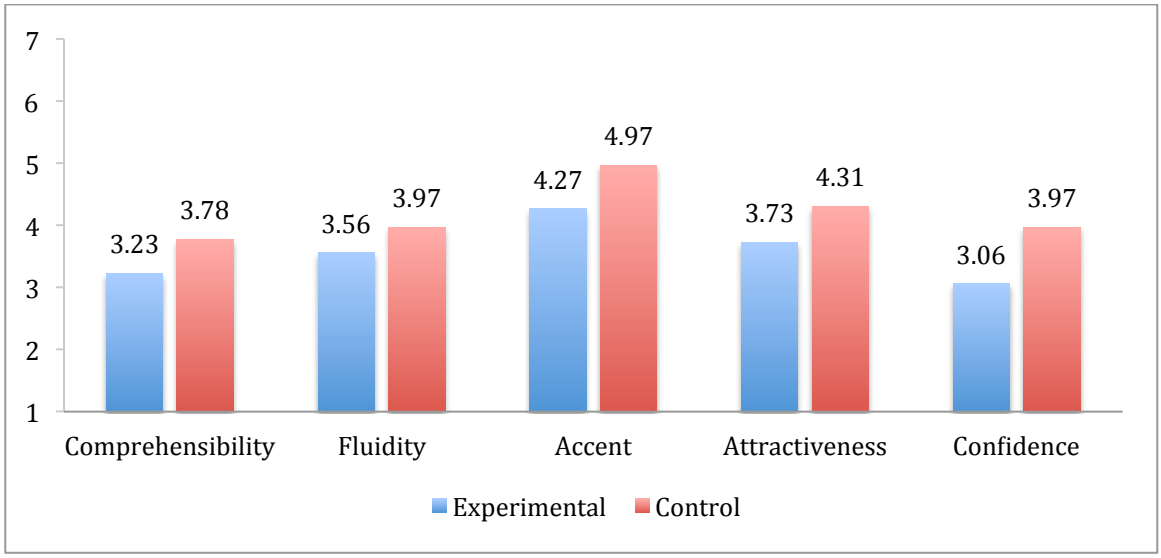


Figure 18. Ratings on the quality of read speech in the pre-test.

The post-test ratings showed similar patterns that the experimental group received lower ratings on the five categories than the control group (Comprehensibility: $M_e = 3.06$, $M_c = 4.16$; Fluidity: $M_e = 3.52$, $M_c = 4.19$; Accent: $M_e = 4.31$, $M_c = 4.9$; Attractiveness: $M_e = 3.38$, $M_c = 4.38$; Confidence: $M_e = 2.96$, $M_c = 4.28$, see Figure 19).

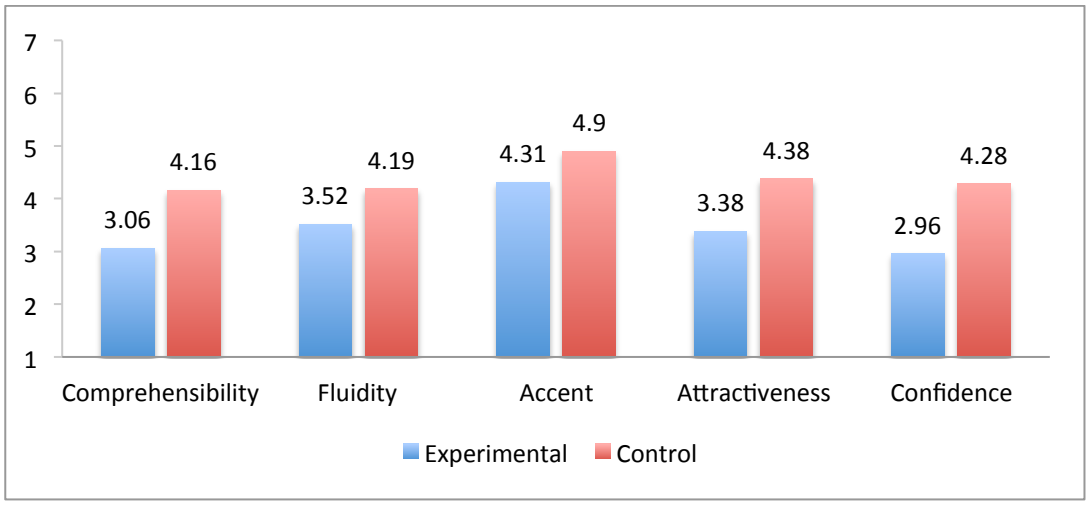


Figure 19. Ratings on the quality of read speech in the post-test.

Repeated measures t-tests were run to examine the students' improvement in their script-reading speech quality (See Table 6). For the experimental group, the students made

some improvement in their speech comprehensibility, fluidity, attractiveness and confidence, although the average ratings only decreased marginally by 0.04 to 0.35. The differences in the ratings were not statistically significant (comprehensibility: t (11) = -0.75, p = 0.47; fluidity: t (11) = -0.16, p = 0.87; accent: t (11) = 0.2, p = 0.84; attractiveness: t (11) = -1.29, p = 0.22; confidence: t (11) = -0.37, p = 0.72).

The control group of students made minimal improvement on their speech accent with mean difference of -0.06, and declined in their speech comprehensibility, fluidity, attractiveness and confidence, as shown by positive mean differences. None of the changes was statistically significant (comprehensibility: t (7) = 1.4, p = 0.2; fluidity: t (7) = 0.58, p = 0.58; accent: t (7) = -0.31, p = 0.76; attractiveness: t (7) = 0.25, p = 0.81; confidence: t (11) = 0.72, p = 0.5).

Table 6

Repeated Measures T-Tests on the Improvement of Read Speech Quality.

		M	*SD*	*SEM*	*t*	*df*	*p*
Experimental Group	Comprehensibility	-0.17	0.77	0.22	-0.75	11	0.47
	Fluidity	-0.04	0.88	0.25	-0.16	11	0.87
	Accent	0.04	0.71	0.21	0.2	11	0.84
	Attractiveness	-0.35	0.95	0.27	-1.29	11	0.22
	Confidence	-0.1	0.97	0.28	-0.37	11	0.72
Control Group	Comprehensibility	0.38	0.76	0.27	1.4	7	0.2
	Fluidity	0.22	1.06	0.38	0.58	7	0.58
	Accent	-0.06	0.56	0.2	-0.31	7	0.76
	Attractiveness	0.06	0.7	0.25	0.25	7	0.81
	Confidence	0.31	1.23	0.43	0.72	7	0.5

To summarize, the training helped the experiment group of students improve a little on their script-reading speech comprehensibility, fluidity, attractiveness and confidence, although none of them reached statistical significance. Furthermore, the control group only made minimal improvement on their speech accent.

4.3. RQ 3: The students' evaluations of the training modules

This question was answered using the descriptive statistics of the students' ratings on the training modules in the post-test survey, along with their responses to the open-ended questions in the training modules.

The 12 students in the experimental group completed a post-survey at the end of the study. They evaluated the training modules' effectiveness on improving their intonation and English presentation using a 7-point scale (1 = strongly disagree, 7 = strongly agree). On average, the students agreed that the training modules had been very helpful in terms of understanding the importance of intonation (M= 6.58) and gaining new knowledge about intonation (M= 6.42). After the training, they would pay more attention to intonation when listening to English (M= 6.33) and speaking English (M= 6.42). They also agreed that the training had helped them increase confidence in making English presentations (M= 5.83), and improve their English speaking skills (M= 5.75) (See Figure 20).

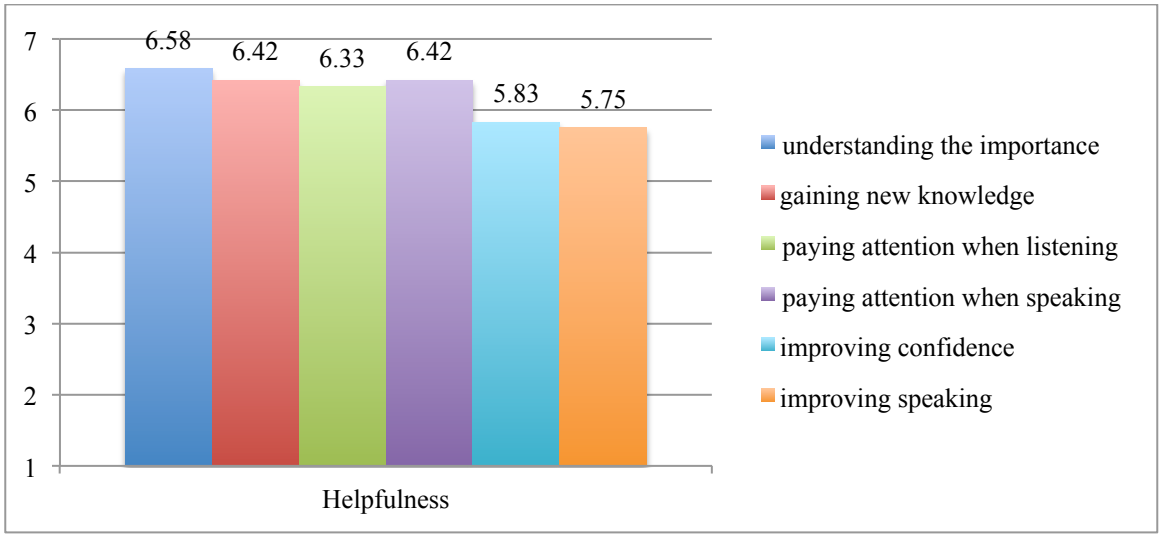


Figure 20. Trainees' evaluation on the helpfulness of the modules.

The students also evaluated how helpful individual activities were in improving their English intonation (See Figure 21). On average, they agreed that all the activities were helpful with the average ratings from 5.92 to 6.58 (1 = strongly disagree, 7 = strongly agree). The most helpful activities were getting written feedback on their spontaneous presentations （*M* = 6.58), and comparing their own recordings to the native speakers' by listening (*M* = 6.42).

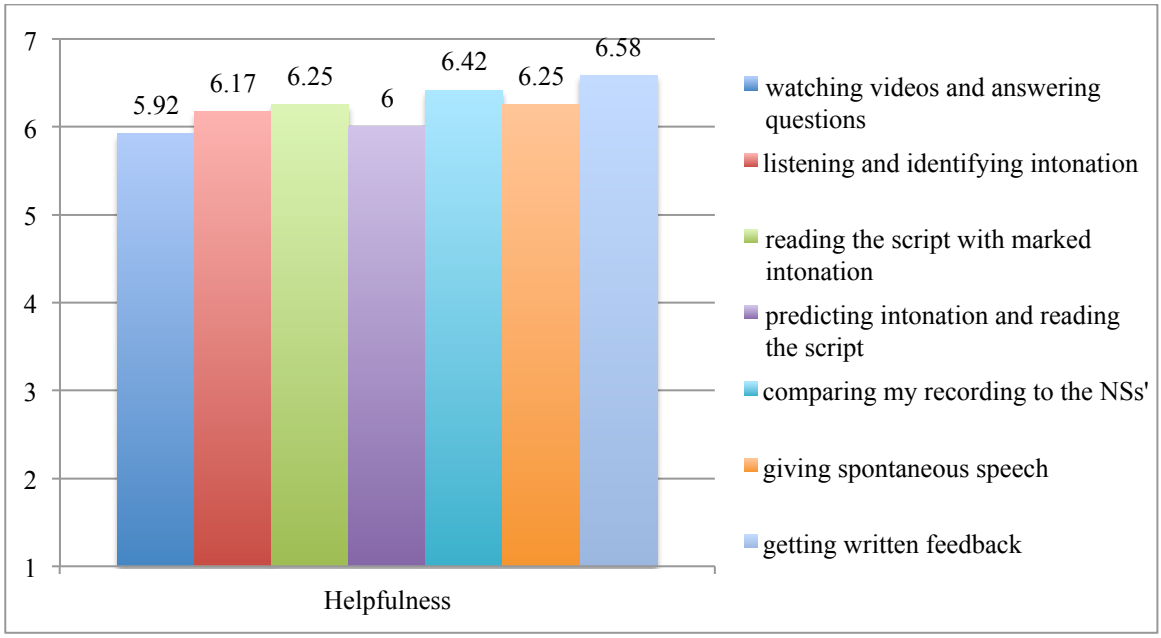


Figure 21. Trainees' evaluation on the helpfulness of individual activities.

In terms of using *Praat* software to visualize their tone choices (See Figure 22), the students agreed that the activities were helpful ($M = 6.17$), comparing their own pitch curves to the NSs' ($M = 6.17$) was more helpful than viewing either the NSs' pitch curves ($M = 6.08$), or their own pitch curves ($M = 6$).

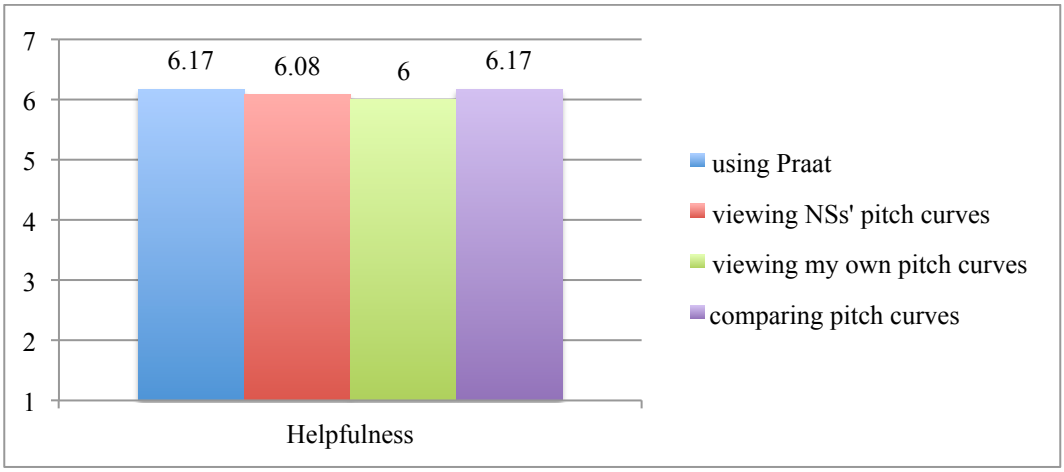


Figure 22. Trainees' evaluation on the helpfulness of using *Praat.*

In the students' responses to the open question, the students offered suggestions for making the training modules more helpful, such as "put more examples", "provide more links to download the speech by NSs", "*Praat* is not very easy to use", and they desired a more automatic and intelligent software that could "label everything (the intonation)", "compare my speech to that of a NS's" and "highlight the difference".

4.4. RQ 4: The relationship between intonation and speech quality

This question was answered using both an auditory approach and an acoustic approach. Two sets of correlation and multiple regression analyses were performed. For both approaches, the dependent variables of the regression analyses were the ratings on the holistic speech quality (comprehensibility, fluidity, accent, attractiveness and speakers' confidence). In the auditory analyses, the independent variables were 12 items on the raters' perceived intonation errors, including speech rate. In the acoustic analyses, the independent variables were 13 calculated acoustic measures of intonation features (See section 3.5). The data consisted of 17 students' script reading speech in the pre-test, and 21 students' interview speech in the pre-test.

4.4.1. The auditory approach

4.4.1.1. Regression analyses of the script reading speech

Table 7 displays the correlations between the speech quality and intonation errors. Out of the 60 pairs of correlations, 30 correlations were statistically significant ($p < .05$). The correlations between speaking rate and speech quality were positive, suggesting that speaking slowly was associated with lower ratings on the speech quality. All the significant correlations between the intonation errors and the speech quality were negative, suggesting that the more errors the raters heard from the speaker, the lower ratings were assigned on the speech quality. The strength of these relationships was moderate to strong.

Speech comprehensibility correlated with speaking choppy sentences (r = -.68), stressing too many words (r =-.54), stressing inappropriate words (r = -.45), using too many falling or level tones (r = -.66) and using confusing tone choices (r = -.58).

Speech fluidity correlated with speaking rate (r = .53), speaking choppy sentences (r = -.88), pausing at wrong places (r = -.60), stressing too many words (r = -.49), stressing inappropriate words (r = -.45), using too many falling or level tones (r = -.56) and using confusing tone choices (r = -.51).

Speech accent correlated with speaking run-on sentences (r = -.48), speaking choppy sentences (r = -.59), stressing too many words (r = -.45), stressing not enough words (r = -.52), and stressing inappropriate words (r = -.47), using too many falling or level tones(r = -.73), using too many rising tones (r = -.56) and using confusing tone choices (r = -.66).

Speech attractiveness correlated with speaking rate (r = .61), speaking choppy sentences (r = -.72), pausing at wrong places (r = -.48), speaking flat speech (r = -.61), using too many falling or level tones (r = -.71), and varying tones in a confusing manner(r = -.47).

Speakers' confidence correlated with speaking rate (r = .72), speaking choppy sentences(r = -.74), pausing at wrong places (r = -.57) and speaking flat speech (r = -.53).

In other words, speaking rate had positive correlations with speech fluidity, attractiveness and speakers' confidence. When the students spoke faster, the raters identified their speech as smoother, more attractive and the speakers' as more confident. However, the correlation between speaking rate and speech comprehensibility (r = .28) and accent (r = 0.36) were small and not significant.

The speakers' pausing errors negatively affected their speech comprehensibility, fluidity, accent, attractiveness and confidence. Specifically, speaking choppy sentences with constant pausing impaired all the five categories. Pausing at wrong places (*e.g.,* in the middle of a thought group) weakened the speech fluidity, attractiveness and speaker's confidence.

The students' prominence errors impaired their speech comprehensibility, fluidity and accent. Specifically, stressing too many words and stressing inappropriate words negatively affected all these three aspects of speech quality. In addition, stressing not enough words was associated with higher ratings in speech accent. However, prominence errors did not affect the ratings on the perception of speech attractiveness and speakers' confidence.

The speakers' tone choice errors had negative impacts on all five categories of the speech quality. Namely, when the speech was flat or monotonous, it was rated less attractive and the speaker less confident. Using too many falling or level tones and confusing pitch changes negatively affected the speech comprehensibility, fluidity, accent, attractiveness and speakers' confidence. Additionally, too many rising tones was associated with more accented speech.

To summarize, the measures of pausing, prominence and tone choice had moderate to strong negative correlations with the five aspects of speech quality. Although correlation doesn't mean causation, the results indicate that reducing the students' intonation errors may help to improve their script-reading speech quality in all five aspects.

Table 7

Correlations between Speech Quality and Auditory Intonation Errors of script reading

		Comprehensibility	Fluidity	Accent	Attractiveness	Confidence
	Speaking rate	0.28	.53*	0.36	.61**	.72**
Pausing	Run-on	-0.44	-0.26	-.48*	-0.29	-0.21
	Choppy	-.68**	-.88**	-.59**	-.72**	-.74**
	Wrong places	-0.40	-.60**	-0.36	-.48*	-.57**
	Fillers	0.00	-0.30	-0.04	-0.29	-0.42
Prominence	Too many	-.54*	-.49*	-.45*	-0.29	-0.10
	Not enough	-0.39	-0.21	-.52*	-0.40	-0.24
	Inappropriate	-.45*	-.45*	-.47*	-0.33	-0.03
Tone choice	Flat speech	-0.32	-0.40	-0.34	-.61**	-.53*
	Falling / level	-.66**	-.56*	-.73**	-.71**	-0.31
	Rising tones	-0.38	-0.26	-.56**	-0.12	-0.05
	Confusing	-.58**	-.51*	-.66**	-.47*	-0.10

** Correlation is significant at the 0.01 level (2-tailed).

* Correlation is significant at the 0.05 level (2-tailed).

Five multiple regression analyses were performed to evaluate how each speech quality measure could be predicted from the linear combination of their speech rate, along with intonation errors of pausing, prominence and tone choices. Stepwise method was used because it minimizes redundancy among independent variables by first including the predictor that had the largest correlation with the dependent variable into the equation, and then incorporating the remaining predictor with the highest partial correlation with the dependent variable while controlling the first predictor. This process was repeated until the R-squared did not increase significantly with the addition of a remaining predictor.

The regression equation of predicting speech comprehensibility comprised of three predictors speaking choppy sentences, using too many falling or level tones and filler words

($F(3, 16) = 18.56$, $p < .001$). The linear combination of these three predictors could explain 78% of the variance in speech comprehensibility ($r^2 = .78$).

The regression equation of predicting speech fluidity included two predictors, speaking choppy sentences and using confusing tones ($F(2, 17) = 52.64$, $p < .001$). Together they could explain 86% of the variance in speech fluidity ($r^2 = .86$).

The regression equation of predicting speech accent consisted of three variables, using too many falling or level tones, speaking choppy sentences and using too many rising tones ($F(3, 16)= 14.74$, $p < .01$), which jointly accounting for 73% of the variance ($r^2 = .73$).

The regression model of speech attractiveness involved four variables, speaking choppy sentences, using too many falling or level tones, speaking flat speech and speech rate ($F(4, 15) = 26.32$, $p < .01$), which explained 88% of the variance ($r^2 = .88$).

The regression model of speakers' confidence contained four variables, speaking choppy sentence, speaking rate, speaking run-on sentences and using too many rising tones ($F(4, 15) = 25.02$, $p < .01$), which could explain 87% of the variance ($r^2 = .87$).

Table 8.

Linear Regression of Auditory Intonation Errors on Script Reading Speech Quality.

Dependent Variable		Unstandardized Coefficients		Standardized Coefficients	*t*	*p*	R^2
		B	Std. Error	Beta			
Comprehensibility	(Constant)	2.66	2.00		1.33	0.20	
	Choppy	-0.67	0.14	-0.70	-4.86	<.01	0.78
	Falling/level	-0.61	0.17	-0.46	-3.59	<.01	
	Filler words	1.05	0.36	0.39	2.87	0.01	

Fluidity	(Constant)	8.65	0.54		16.02	<.01	
	Choppy	-0.78	0.09	-0.80	-8.57	<.01	0.86
	Confusing tone	-0.37	0.11	-0.31	-3.29	<.01	
Accent	(Constant)	8.55	0.64		13.31	<.01	
	Falling/level	-0.45	0.13	-0.49	-3.37	<.01	0.73
	Choppy	-0.22	0.09	-0.35	-2.50	0.02	
	Rising tones	-0.30	0.14	-0.30	-2.14	0.05	
Attractiveness	(Constant)	7.65	0.99		7.71	<.01	
	Choppy	-0.34	0.11	-0.35	-3.12	0.01	
	Falling/level	-0.58	0.14	-0.43	-4.18	<.01	0.88
	Flat speech	-0.30	0.11	-0.27	-2.65	0.02	
	Speech rate	0.30	0.13	0.26	2.41	0.03	
Confidence	(Constant)	0.91	1.19		0.76	0.46	
	Choppy	-0.36	0.13	-0.35	-2.91	0.01	0.87
	Speech rate	0.92	0.16	0.74	5.83	<.01	
	Run-on	-0.48	0.12	-0.45	-3.93	<.01	
	Rising tones	0.53	0.17	0.33	3.14	0.01	

4.4.1.2. Regression analyses of the spontaneous speech

Table 9 displays the correlations between the quality of spontaneous speech and auditory ratings on intonation errors. Out of the 60 pairs of correlations, 41 correlations were statistically significant ($p < .05$). In line with the script reading speech analyses, all the significant correlations between the intonation errors and the speech quality were negative, suggesting that the more errors the raters heard from the speaker, the lower ratings were assigned on the speech quality. The correlations between speaking rate and speech quality

were positive, suggesting that speaking slowly was associated with lower ratings on the speech quality. The strength of these relationships was moderate to strong.

Speech comprehensibility correlated with speaking choppy sentences (r = -.46), pausing at wrong place (r = -.5), stressing too many words (r = -.65), stressing inappropriate words (r = -.56), speaking flat speech (r = - .58), using too many falling or level tones (r = -.64) and using confusing tone choices (r = -.58).

Speech fluidity correlated with speaking rate (r = .59), speaking choppy sentences (r = -.65), pausing at wrong places (r = -.71), stressing too many words (r = -.64), stressing not enough words (r = -.43), stressing inappropriate words (r = -.58), speaking flat speech (r = -.62), using too many falling or level tones (r = -.54) and using confusing tone choices (r = -.48).

Speech accent correlated with speaking choppy sentences (r = -.59), stressing too many words (r = -.48), pausing at wrong places (r = -.5), stressing not enough words (r = -.66), and stressing inappropriate words (r = -.59), speaking flat speech (r = -.54), using too many falling or level tones(r = -.66), and using confusing tone choices (r = -.54).

Speech attractiveness correlated with speaking rate (r = .78), pausing at wrong places (r = -.49), stressing too many words (r = -.48), stressing not enough words (r = -.53), stressing inappropriate words (r = -.5), speaking flat speech (r = -.67), and using too many falling or level tones (r = -.58).

Speakers' confidence correlated with speaking rate (r = .69), speaking run-on sentences (r = -.41), speaking choppy sentences (r = -.72), pausing at wrong places (r = -.77) , using too many fillers (r = -.46), stressing too many words (r = -.69), stressing not enough

words (r = -.51), stressing inappropriate words (r = -.59), speaking flat speech (r = -.77), using too many falling or level tones (r = -.53), and using too many rising tones (r = -.44).

In other words, speaking rate had positive correlations with speech fluidity, attractiveness and speakers' confidence. When the students spoke faster, the raters identified their speech as smoother, more attractive and the speakers' as more confident. However, the correlation between speaking rate and speech comprehensibility (r = .38) and accent (r = .39) was small and not significant.

The students' pausing, prominence and tone choice errors had negative impacts on all five categories of the speech quality. Specifically, pausing at wrong places was associated with more unfavorable ratings in all the five categories, speaking run-on sentences and using too many fillers negatively affected the listeners' judgment of the speaker's confidence; and speaking choppy sentences was associated with more negative ratings in comprehensibility, fluidity, accent, and speaker's confidence. Stressing too many words and stressing inappropriate words negatively affected the ratings on all the five categories, and stressing not enough words was negatively associated with the ratings on speech fluidity, attractiveness and speaker's confidence. When the students spoke flat speech and used too many falling or level tones, all the five categories of their speech quality were negatively affected. Their use of too many rising tones was associated with unfavorable ratings in the judgment of the speaker's confidence, and using confusing tone choices was negatively associated with speech comprehensibility, fluidity and attractiveness.

In accordance with the script reading speech analyses, the measures of pausing, prominence and tone choice had moderate to strong negative correlations with the five aspects of the spontaneous speech quality. Although correlation doesn't mean causation, the

results indicate that reducing the students' intonation errors may help to improve their spontaneous speech quality in all five aspects.

Table 9.

Correlations between Spontaneous Speech Quality and Auditory Intonation Errors.

		Comprehensibility	Fluidity	Accent	Attractiveness	Confidence
	Rate	0.38	.59**	0.39	.78**	.69**
Pausing	Run-on	-0.26	-0.35	-0.21	-0.18	-.41*
	Choppy	-.46*	-.65**	-.48*	-0.38	-.72**
	Wrong place	-.50*	-.71**	-.50*	-.49*	-.77**
	Fillers	-0.22	-0.29	-0.13	-0.03	-.46*
Prominence	Too many	-.65**	-.64**	-.66**	-.48*	-.69**
	Not enough	-0.36	-.43*	-0.32	-.53**	-.51*
	Inappropriate	-.56**	-.58**	-.59**	-.50*	-.59**
Tone choice	Flat speech	-.58**	-.62**	-.54**	-.67**	-.77**
	Falling / level	-.64**	-.54**	-.66**	-.58**	-.53**
	Rising tones	-0.4	-0.4	-0.34	-0.21	-.44*
	Confusing	-.58**	-.48*	-.54**	-0.22	-0.36

** Correlation is significant at the 0.01 level (2-tailed).

* Correlation is significant at the 0.05 level (2-tailed).

Regression analyses were performed to examine how well the auditory intonation errors could predict the script-reading speech quality (See 10). The regression equation of predicting speech comprehensibility comprised of two predictors stressing too many words and using too many falling or level tones ($F(1, 22) = 16.01$, $p < .001$). The linear combination of these three predictors could explain 53% of the variance in speech comprehensibility ($r^2 = .53$).

The regression model of predicting speech fluidity included one predictor, pausing at wrong places ($F(1, 22) = 22.72$, $p < .001$), which could explain 51% of the variance in speech fluidity ($r^2 = .51$).

The regression equation of predicting speech accent consisted of two variables, stressing too many words and using too many falling or level tones ($F(1, 22)= 17.27$ $p < .001$), which jointly accounted for 55% of the variance ($r^2 = .55$).

The regression equation of speech attractiveness involved two variables, speaking rate, and using too many falling or level tones ($F(1, 22) = 33.25$, $p < .001$), which explained 72% of the variance ($r^2 = .72$).

The regression model of speakers' confidence contained two variables, speaking flat speech ad pausing at wrong places ($F(1, 22) = 31.81$, $p < .001$), which could explain 71% of the variance ($r^2 = .71$).

Table 10.

Linear Regression of Auditory Intonation Errors on Spontaneous Speech.

Dependent Variable		Unstandardized Coefficients		Standardized Coefficients	*t*	*p*	R^2
		B	Std. Error	Beta			
	(Constant)	8.48	0.89		9.54	<.01	
Comprehensibility	emphasized too many falling/level tones	-0.55	0.25	-0.42	-2.26	0.04	0.53
		-0.47	0.22	-0.40	-2.14	0.04	
Fluidity	(Constant)	7.29	0.63		11.64	<.01	0.51
	paused at wrong	-0.73	0.15	-0.71	-4.77	<.01	

	places						
Accent	(Constant)	7.92	0.64		12.45	<.01	
	emphasized too many	-0.41	0.18	-0.42	-2.35	0.03	0.55
	falling/level	-0.36	0.16	-0.41	-2.30	0.03	
Attractiveness	(Constant)	2.64	0.93		2.85	0.01	
	Speech rate	0.69	0.13	0.65	5.29	<.01	0.72
	falling/level tones	-0.35	0.12	-0.36	-2.93	0.01	
Confidence	(Constant)	7.50	0.47		16.10	<.01	
	flat speech	-0.41	0.14	-0.47	-2.99	0.01	0.71
	paused at wrong places	-0.41	0.14	-0.46	-2.93	0.01	

To summarize, the regression analyses found that the human raters' judgments of intonation errors were good predictors of their impressions of the quality of the students' speech. At least one or a linear combination of several auditory measurements of intonation errors were able to explain at least 73% of the variance for the script-reading speech, and at least 51% of the variance for the spontaneous speech.

4.4.2. The acoustic approach

4.4.2.1. Acoustic analysis results of the script reading speech

Most of the acoustic items had weak and non-significant correlation with the speech quality, except that pitch range had moderate and significant correlation with speech attractiveness (r = -0.51, p = .037). Two submeasures on filled pauses, the number and mean length of filled pauses, had significant correlation with the speech comprehensibility, fluidity, accent, and attractiveness. However, the raw data showed that there were only two data points on these two items. That is, two students had two filled pause when reading the

script. The number and mean length of filled pauses were excluded a priori from the regression analysis regardless of any potential contributions they make to the model.

Table 11

Correlations between Script-reading Speech Quality and Acoustic Measures.

		comprehensibility	**fluidity**	**accent**	**attractiveness**	**confidence**
Rate	speech rate	-0.15	-0.37	-0.11	-0.20	-0.41
	articulation rate	-0.14	-0.31	-0.09	-0.21	-0.46
	mean length of runs	-0.28	-0.42	-0.28	-0.17	-0.10
	phonation time ratio	-0.12	-0.39	-0.12	-0.13	-0.12
Pause	number of silent pauses	0.24	0.22	-0.10	0.19	0.33
	mean length of pauses	-0.16	0.13	-0.26	0.16	0.45
	number of filled pauses	.66**	.51*	.54*	.50*	0.27
	mean length of filled pauses	.66**	.54*	.53*	.53*	0.35
Prominence	no. of stressed words per minute	-0.10	-0.13	0.02	-0.39	-0.29
	proportion of stressed words to the total no. of words	0.00	0.09	0.05	-0.23	0.00
Intonation	pitch range	-0.18	-0.46	-0.26	-0.51*	-0.29
	terminal contour	-0.10	0.04	-0.22	-0.01	0.06
	intensity	0.01	-0.06	0.16	0.25	0.16

** Correlation is significant at the 0.01 level (2-tailed).

* Correlation is significant at the 0.05 level (2-tailed).

The regression analyses showed that no regression models were generated for speech comprehensibility, fluidity, accent, and speaker's confidence. That is, the acoustic measures of rate, pause, prominence and intonation were not able to predict the variance in the assigned ratings on the scripted-reading speech comprehensibility, fluidity, accent or speaker's confidence. The predictor pitch range was able to predict the speech attractiveness (F (1, 15) = 5.22, p = .037), which could explain 26% of the variance (r^2 = .26) (See Table 12).

Table 12.

Linear Regression of Acoustic Measures on Script Reading Speech Quality.

		Unstandardized Coefficients		**Standardized Coefficients**	**t**	**Sig.**	**R^2**
Dependent Variable		**B**	**Std. Error**	**Beta**			
Comprehensibility	No variables were entered into the regression model						
Fluidity	No variables were entered into the regression model						
Accent	No variables were entered into the regression model						
Attractiveness	(Constant)	4.864	0.535		9.1	0	0.26
	pitch range	-0.011	0.005	-0.508	-2.285	0.037	
Confidence	No variables were entered into the regression model						

4.4.2.2. Acoustic analysis results of the spontaneous speech

None of the acoustic measures correlated with speech comprehensibility or accent. Speech rate (r = -.51) and pitch range (r = -.48) significantly correlated with speech fluidity. Speech rate (r = -.54), mean length of runs (r = -.55), and phonation time ratio (r = -.45) significantly correlated with speech attractiveness. Speech rate (r = -.64), articulation rate (r

= -.55), pitch range (r = -.53), and intensity level (r = .47) significantly correlated with speaker's confidence (See Table 13).

Table 13.

Correlations between Speech Quality and Acoustic Measures of Spontaneous Speech

		comprehensibility	fluidity	accent	attractiveness	confidence
Rate	speech rate	-0.28	-.51*	-0.20	-.54*	-.64**
	articulation rate	-0.27	-0.41	-0.15	-0.33	-.55*
	mean length of runs	-0.23	-0.38	-0.21	-.55*	-0.26
	phonation time ratio	-0.05	-0.21	-0.10	-.45*	-0.21
Pause	number of silent pauses	0.08	0.19	0.08	0.33	-0.01
	mean length of pauses	-0.01	0.08	0.06	0.31	0.30
	number of filled pauses	0.20	0.25	0.07	-0.01	0.21
	mean length of filled pauses	-0.09	0.04	-0.34	-0.04	0.25
Prominence	no. of stressed words per minute	0.04	-0.12	0.02	-0.23	-0.21
	proportion of stressed words to the total no. of words	0.26	0.19	0.25	0.21	0.25
Intonation	pitch range	-0.33	-.48*	-0.26	-0.43	-.53*
	terminal contours	0.31	0.38	0.24	-0.06	0.08
	Intensity level	0.30	0.34	0.17	0.21	.47*

** Correlation is significant at the 0.01 level (2-tailed).

* Correlation is significant at the 0.05 level (2-tailed).

No variables were entered into the regression models for speech comprehensibility or accent. The regression model for speech fluidity consisted of two predictors, speech rate and terminal contours (F (2, 18) = 6.64, p = .007), which could explain 42% of the variance. The regression equation for attractiveness included one predictor, mean length of runs (F (1, 19) = 8.02, p = .011), which could account for 30% of the variance. The regression model for speaker's confidence contained one predictor, speech rate (F (1, 19) = 13.29, p = .002), which could explain 41% of the variance (See Table 14).

Table 14.

Linear Regression of Acoustic Measures on Spontaneous Speech Quality.

Dependent Variable		**Unstandardized Coefficients**		**Standardized Coefficients**	***t***	***p***	***R^2***
		B	**Std. Error**	**Beta**			
Comprehensibility	No variables were entered into the regression model						
Fluidity	(Constant)	6.012	1.227		4.9	0	0.42
	speech rate	-0.019	0.006	-0.527	-2.942	0.009	
	Terminal contours	1.292	0.562	0.411	2.298	0.034	
Accent	No variables were entered into the regression model						
Attractiveness	(Constant)	6.02	0.64		9.413	0	0.3
	mean length of runs	-0.247	0.087	-0.545	-2.832	0.011	
Confidence	(Constant)	7.773	0.984		7.903	0	0.41
	speech rate	-0.021	0.006	-0.642	-3.645	0.002	

In summary, the auditory approach and the acoustic approach adopted different methods to measure the speakers' intonation errors, one with the raters' perceptual judgment, and the other with objective acoustic measurements. The two sets of measurements had divergent relations with and abilities to predict the speech quality. In

short, the auditory measures of intonation errors had strong correlation with the speech quality. Furthermore, the auditory measures generated better fitting regression models, and their linear combinations could explain a large portion of the variance in speech quality, at least 73% of the script reading speech and 51% of the spontaneous speech. In contrast, only a few acoustic measures of intonation errors had significant correlation with the speech quality. Likewise, the acoustic measures either did not form a regression equation, or the model could only account for a small portion of variance in the speech quality.

5. Discussion and Conclusion

This project focused on English intonation and speech comprehensibility of Chinese speakers. It involved first the development of three online modules that instructed and trained the students' perception and production of English intonation in discourse, and then the testing of the modules' effectiveness in improving the students' oral speech. This chapter first discusses the main findings of the four research questions in relation to previous research, then describes the pedagogical implications, limitations and suggestions for future research, and concludes with closing statements.

5.1. Discussion of research findings

RQ 1: Challenges with regard to intonation.

This question identified the main challenges with regard to intonation that Chinese speakers confront. It was answered from both the students' and the raters' points of view by calculating the average ratings. The students acknowledged that they had difficulties with nine out of the 11items of intonation. The finding that the Chinese students' had difficulty with pausing, prominence and tone choice in speaking English was broadly consistent with previous research (Lu & Miran, 2016; Mixdorff & Ingram; 2009; Pickering, 2001;

Pickering, 2004; Tseng et al., 2013; Tyler, 1992; Wennerstrom, 1998). This project found the most pressing problems for the Chinese students were using too many fillers, pausing at wrong places, speaking run-on sentences, speaking choppy sentences, not knowing where to stress and speaking flat speech.

The comparison between the native speakers' judgments and the Chinese students' self-reflection on their speech intonation was lacking in previous studies. This study found that the raters were more lenient on the students' intonation performance and assigned more positive scores on all 11 items with two exceptions, speaking choppy sentences and stressing too many words. Speaking choppy sentences was rated as the most pressing problem of pausing. Contradictorily, the students rated speaking choppy sentences as the least challenging problem. The acoustic measures of the number and length of pauses confirmed the raters' perceptions of choppy sentences, as the students used almost two times as many??? pauses in reading the script compared to the raters ($M_s = 8.78$, $M_r = 3.75$), and the mean length of pause was significantly shorter ($M_s = 0.40$, $M_r = 0.52$, $t(19) = .005$).

RQ 2: the effects of the training modules

This question investigated whether completing the training modules helped the students improve their script reading and spontaneous speech. Repeated measures t-tests were run to compare the impressionistic ratings (*i.e.,* comprehensibility, fluidity, accent, attractiveness and confidence) on the pre-test and the post-test speech samples for both experimental and control groups.

The repeated measures t-tests of the spontaneous speech quality showed that the experimental group of students made progress on all five aspects of speech quality after completing the training modules. Among them, the improvements on speech

comprehensibility and speakers' confidence were statistically significant. Contradictorily, the control group either remained the same or even had slight setbacks in the post-test.

The training modules also had impacts on the students' script reading speech. The experimental group of students received better ratings on their speech comprehensibility, fluidity, attractiveness and speakers' confidence, although none of them reached statistical significance. The limited improvement was probably due to the fact that the students' intonation was rated higher than 4 points on a 7-point scale when reading the script in the pre-test. There was not much room for them to improve during the training. Similarly to the spontaneous speech analysis, the control group either remained the same or had slight setbacks in the post-test script-reading speech ratings.

The effectiveness of the training modules corroborated previous studies that explicit instruction and training on suprasegmentals results in improvement in speech comprehensibility (Derwing et al., 1998; Harrison 2004; Putto, 2013; Ramirez Verdugo, 2006).

The first explanation for the trainees' improvement was perhaps that the training modules provided a large amount of input that were authentic, at discourse-level and relevant in both topics and communicative functions. Listening to the university lecture, discussion sessions, public presentations (*e.g.,* TED talks) provided examples of clear and engaging speech that the learners could imitate. The receipt of massive comprehensible input that the learners needed may propel their oral speaking performance (Gorsuch, 2011; Harrison 2005; Hurley, 1992; Krashen, 1994; McGregor et al., 2016; Scotto di Carlo, 2014; Tanner & Landon, 2009; Thompson, 2003).

The second explanation may be the instruction and activities that the training modules provided. First, the instruction lecture explicitly targeted the intonation features (*e.g.,* definition, benefits, challenges, guidelines). Watching the video and answering questions allowed the learners to activate their pre-existing knowledge and gain new knowledge. The perception and production exercises focused the learners' attention to speech intonation. The perception exercises connected phonetic features of intonation to its meanings and functions in discourse, *e.g.,* achieving coherence, expressing attitudes. The production activities of making speech mirrored the communicative functions that American university students needed, *e.g.,* introducing oneself, describing one's experiences and accomplishments. These activities provided abundant opportunities for the students to practice on their perception and production of intonation, which could lead to better quality oral speech (Chun, 2002; Derwing et al., 1998; Lima, 2015; Long, 1985).

The third explanation was possibly the instant and concrete visual feedback as well as individualized written feedback that the trainees received. Using *Praat* software displaying the pitch contours helped the learners see pitch change in a more apparent manner. In addition, they could compare their own speech to the native speakers' and view how far their intonation deviates. The written feedback gave specific comments on individual students' intonation performance. The two types of feedback raised the awareness of intonation, checked their perception and production, and motivated the students to learn (De Bot & Maifert, 1982; Harrison 2004; Putto, 2013; Ramirez Verdugo, 2006; Swain, 1995; Vygotsky, 1962).

RQ 3: the students' evaluation of the training modules

In order to find out the students' perceptions of the training modules, their responses in the post-test survey and the open-ended questions in the training modules were analyzed.

The students indicated that the training modules were helpful in terms of understanding the importance of and gaining new knowledge about intonation. The training also helped them in paying attention to intonation when listening to and speaking English in future. In addition, the training helped them increase their confidence in making English presentations and improve their speaking skills.

In terms of the helpfulness of individual activities, the students agreed that all seven activities were helpful. The most helpful activities were getting written feedback on their presentations and comparing their own recordings to the native speakers' by listening.

The helpfulness of using *Praat* was rated separately. The students agreed that using *Praat* was helpful. Comparing their own pitch curves to the NSs' was more helpful than simply viewing either of them.

The students also provided suggestions to improve the training modules, such as "put more example", and provide more speech models by NSs. They also desired a more automatic and friendly software that could "label", "compare" and "highlight the difference", which *Praat* was not able to do.

The students' high approval of the training modules as to their helpfulness is not new. Previous studies on intonation training often ask for learners' evaluation of the intervention that they receive, and generally obtain positive feedback on the training's usefulness, *e.g.,* raising their awareness of intonation, increasing their confidence in L2 speaking, and enhancing their imitating ability (De Bot & Maifert, 1982; Hardison, 2004; Lima, 2015; Ramirez Verdugo, 2006).

RQ 4: The relationship between intonation and speech quality

This question intended to find out how Chinese students' intonation contributed to the raters' judgment on their speech quality by running regression analyses. Both auditory and acoustic measurements of intonation were analyzed in order to compare which approach was a better predictor.

Auditory analysis results found close correlations between speech quality and intonation errors. Out of the 60 correlation combinations, 30 for the script-reading speech and 41 for the spontaneous speech were statistically significant. When the students spoke faster, both script-reading and spontaneous speech was rated more favorable in fluidity, attractiveness and confidence. However, speaking rate did not significantly correlate with comprehensibility or accent. The more errors on pausing, prominence and tone choice that were heard, the lower the ratings that were assigned on the speech quality. Specifically, for the script-reading speech, the speakers' pausing and tone choice errors negatively affected all the five aspects of speech quality. The prominence errors mainly impaired their speech comprehensibility, fluidity and accent. For the spontaneous speech, the speakers' pausing, prominence and tone choice errors all negatively affected the five aspects of speech quality.

The auditory measurements of intonation errors could explain substantial portions of variances in all the five aspects of speech quality. Specifically, for the script-reading speech, the linear combination of intonation errors could predict 78% of the variance in comprehensibility, 86% of the variance in fluidity, 73% of the variance in accent, 88% of the variance in attractiveness and 87% of the variance in confidence. For the spontaneous speech, the linear combination of intonation errors could predict 53% of the variance in

comprehensibility, 51% of the variance in fluidity, 55% of the variance in accent, 72% of the variance in attractiveness and 71% of the variance in confidence.

Acoustic analyses showed that only a few items had significant correlation with speech quality, and not all speech quality could be predicted by the acoustic measurements. For script-reading speech, the regression equation could explain 26% of the variance in attractiveness. However, comprehensibility, fluidity, accent and confidence could not be predicted by the acoustic items. For spontaneous speech, the linear combination of acoustic measurements could predict and explain 42% of the variance in fluidity, 30% of the variance in attractiveness and 41% of the variance in confidence. Comprehensibility and accent could not be predicted by the acoustic measurements.

The findings of correlation between auditory judgment of intonation errors and speech quality concurred with other studies that showed L2 speakers' prosodic errors strongly affected the evaluation of their English proficiency, their speech comprehensibility, and even the perception of their personality (Anderson-Hsieh et al., 1992; Pickering, 2001; Pickering, 2004; Wennerstrom, 1998). While previous investigation was generally based on correlation analysis results, this study took a further step running regression analysis, and revealed that the auditory measurements of intonation errors could explain 73% to 88% of the variance in ratings of speech quality (*i.e.,* comprehensibility, fluidity, accent, attractiveness, confidence) for script-reading speech, and 51% to 72% for the spontaneous speech.

In contrast to other studies in the literature (Kang, 2010; Kang et al., 2010), the acoustic measures of intonation mostly did not correlate with the ratings of speech quality,

and the linear combination of acoustic measures could not predict the variance of the comprehensibility or accent of either script-reading or spontaneous speech.

5.2. Pedagogical implications

The development of the online training modules was guided by the second language acquisition research and drew from principles of effective pronunciation teaching in classroom and computer-assisted pronunciation training programs. The training provided explicit instruction to teach selected intonation features, their functions and rules in spoken discourse. Learners accessed large quantities of authentic materials produced by different native models in discourse context, including lectures, discussion sessions and public presentations on topics that were relevant to college students.

The training modules provided immediate and individual feedback to the students, which brought the students' attention to specific individual problems that the students may not otherwise notice and stimulated them to attempt self-improvement. First, learners completed activities using *Praat* software to visualize pitch contours and received concrete and immediate feedback on the subtle subject of intonation. Learners agreed that using *Praat* was helpful to learn intonation, especially comparing their own pitch curves to the NSs'. In addition, they received comments and suggestions from me that highlighted each student's specific problems of their oral presentations in each training module. Learners rated the written feedback as the most helpful training activity.

Based on the analysis results in this study, the online training modules successfully helped the Chinese students improve their spontaneous speech comprehensibility and the speaker's confidence. The training modules were perceived to be helpful by the students in raising their awareness of intonation directing their attention to intonation when listening to

and speaking English. The students also rated the training activities to be of good quality and engaging.

One pedagogical implication of this project is that intonation training should be included and explicitly taught in L2 English pronunciation training. L2 speakers' intonation is often problematic due to lack of instruction and L1 prosody transfer, and they would not understand its use or functions just by living and studying in an English-speaking environment. Their misuse of intonation hinders their speech quality and creates barriers in their interaction with L1 speakers. Explicit instruction on thought group, prominence and tone choice, like the training modules, would raise the learners' awareness, help them understand the rules, direct their attention to intonation when listening and speaking, which ultimately contribute to more comprehensible and fluid English speech.

Another pedagogical implication is that the online training modules' success in training L2 speakers' intonation, either used independently or blended with a pronunciation courses, suggests the viability of online modules. Learners are able to learn without sitting in a classroom at a certain time. They could be students at different levels from different countries (*e.g.,* China and the United States). The online modules offer multiple speech models on relevant topics. Learners have opportunities to do a large amount of exercises. Additionally, the online modules create a low-anxiety learning environment, and promote learner autonomy in that the learners control their pace of study, record the data and track their progress.

The training modules are expandable with new content in addition to thought group, prominence and tone choice. Teachers or researchers may add modules focusing on other suprasegmental features, *e.g.,* paratone, or segmental features, *i.e.,* vowels and consonants,

to further improve learners' pronunciation and speech quality. The new modules could incorporate the same types of instruction video and tasks (perception and production activities). Segmental modules may offer visual feedback such as tongue position and lip movement to enhance the learners' understanding.

The findings also suggest that individual feedback is needed and feasible with the online training modules, as the students rated that the written feedback on their presentations was the most helpful activity. Although the students may not meet with the researcher when completing the training modules independently, they can still submit their presentation speech and hear comments and suggestions virtually from the researcher via email or online conferencing. Hybrid learning of classroom instruction and online training (*e.g.,* flipped classroom) would be more flexible in providing individual feedback. It is important for a teacher to have follow-up discussion on the students' questions. Peer evaluation may also be included by grouping the trainees into pairs and asking them to evaluate each other's oral speech following rating rubrics.

In addition, the study implies that *Praat* software would be a good tool to train L2 speakers' intonation. *Praat* is free and easy to use. Learners could use it to record speech and display pitch contours that are relatively easy to interpret. It offers immediate visual feedback on the speakers' intonation, which keeps the students motivated and brings their attention to specific problems that they may not notice by listening.

Lastly, the findings suggest that linking acoustic properties to auditory ones is difficult and relying on human raters might be superior to relying on acoustic measures in terms of being able to hear and predict speech quality.

5.3. Limitations

The major limitation of this study is the length of the training. Condensing such a complex topic of intonation into three short training modules over four weeks is extremely challenging. The training modules try to be as comprehensive as possible, but some simplifications or eliminations are inevitable. Therefore, the training may be more helpful for learners who have never heard or thought about intonation before, but not so beneficial for more advanced learners. In addition, some students may need extended practice to enhance their learning. Some of them expressed eagerness to learn further after completing the modules, such as hearing more speech examples and doing further activities of visualizing pitch contours.

Another limitation is the lack of longer delayed post-test on the improvement retention and transfer to new speech. The post-test was not given immediately after the training, but was conducted at least one week later after the students finished the last training module. However, this short time interval may be not enough to detect whether the improvement is robust. The test on the improvement transfer to new speech was excluded from the study due to the time constraints of the researcher.

Last but not least, this study involved a limited size of subjects, and did not analyze the perception of English intonation even though the training modules focused on both perception and production of English intonation. As a result, it is unclear whether the perception activities have been effective, and how the students' perception knowledge relates to their production of intonation.

5.4. Suggestions for future research

While the training modules have been successful in improving the Chinese students' speech comprehensibility and confidence, there is much room to improve in order to make

them more helpful for the target students. For example, the raters judged that stressing too many words was a big challenge. The reason may be that the students didn't know how to reduce vowels in unstressed syllables due to the syllable-timed rhythm in their L1 Mandarin. However, the training module on prominence designed for this study had more instruction and exercises on stressing but little on de-stressing. Future research could add extra activities on vowel reduction for Chinese students and examine whether it affects the training effects on the students' speech.

An important finding of this study was the difficulty to predict speech quality from the acoustic properties. While previous studies mainly focused on *F*0 (Kang, 2010), the inclusion of intensity in this study showed that increased intensity would enhance the perception of speaker confidence. The reason why linking acoustic properties to auditory ones is difficult may be that one acoustic property may impact another one in the auditory domain. For example, longer sounds give an impression of being louder even if they are not. Future research may include a more comprehensive list of acoustic properties, *e.g., F*0, intensity and duration, to reveal the relationship between acoustic properties and auditory ones and test if the combination of more acoustic properties could better predict speech quality.

Another aspect of this study is the finding that the students' perception of their challenges with regard to intonation varied from the native speakers' judgment on speaking choppy sentences and stressing too many words. Future research may examine whether the students' attitudes would affect their learning outcomes, *i.e.,* do the students make less progress on the less challenging aspects when self-evaluated? If so, future implementation of

the training modules should include a diagnosis on students' speech by native speakers, and the students need to be aware of the diagnosis results before starting the training.

5.5. Conclusion

It is well acknowledged that intonation is an integral and meaningful part of English language use in communication, and English learners need this crucial skill to perceive the intonation contrasts, interpret its meaning and functions in contextualized situations and use it appropriately in their speech. However, it is often neglected in EFL pronunciation instruction.

The Chinese students in this study had difficulties with using appropriate intonation when speaking English, which negatively impacted their speech comprehensibility, fluidity, accent, confidence and attractiveness, as judged by native speakers. The most pressing problems included using too many fillers, pausing at wrong places, speaking run-on sentences, speaking choppy sentences, not knowing where to stress and speaking flat speech.

The online training modules provided sustained, systematic training on pausing, prominence and tone choice, which employed discourse-based materials, and consisted of instruction videos, listening and speaking activities that connected phonetic features to the meaning and functions of each intonation feature, and also included learner-created pitch contours as instant feedback, in addition to individualized written feedback on the speech.

The online modules helped the experimental group of students improve almost all aspects (*i.e.,* comprehensibility, fluidity, accent, confidence and attractiveness) of their script-reading and spontaneous speech. The improvement in comprehensibility and confidence of the spontaneous speech was statistically significant. In contrast, the control group did not show improvement. The students gave high ratings on the helpfulness of the

training modules. The findings have implications for teaching pronunciation to English learners and developing computer assisted pronunciation training tools.

References

ACTFL. (2017). *Performance Descriptors for Language Learners*. Retrieved from https://www.actfl.org/sites/default/files/pdfs/ACTFLPerformance-Descriptors.pdf

ACTFL. (2012). *ACTFL proficiency guidelines 2012.* Retrieved from https://www.actfl.org/publications/guidelines-and-manuals/actfl-proficiency-guidelines-2012/english/speaking

Anderson-Hsieh, J., Johnson, R. & Koehler, K. (1992). The relationship between native speaker judgments of nonnative pronunciation and deviance in segments, prosody, and syllable structure. *Language Learning, 42*(4), 529-555.

Andrade, M. S., (2006). international students in English-speaking universities. *Journal of Research in International Education 5*(2), 131-154. Doi: 10.1177/1475240906065589

Benzie, H. J., (2010). Graduating as a 'native speaker': International students and English language proficiency in higher education. *Higher Education Research & Development 29* (4), 447-459.

Boersma, P. & Weenink, D. (2014). Praat: Doing phonetics by computer [Computer program]. Version 5.3.77 (retrieved 18 May 2014 from http://www.praat.org/).

Bolinger, D. (1985). *Intonation and its parts.* Stanford: Stanford University Press.

Brazil, D. (1985). *The communicative value of intonation in English*. Bleak House Books and English Language Research.

Brazil, D. (1975). *Discourse intonation I*. Birmingham, England: English Language Research Monographs.

Chun, D. M. (2012). Computer-assisted pronunciation teaching. In *The Encyclopedia of*

Applied Linguistics. (pp. 823-834). Blackwell Publishing.

Chun, D. M. (2002). *Discourse intonation in L2: From theory and research to practice.* Amsterdam, Netherlands: John Benjamins.

Chun, D. M. (1998). Signal analysis software for teaching discourse intonation. *Language Learning and Technology, 2,* 61–77.

Clennell, C. (1997). Raising the pedagogic status of discourse intonation teaching. *Language Teaching Journal, 51*(2), 117–34.

Couper-Kuhlen, E., & Selting, M. (1996). Towards an interactional perspective on prosody and a prosodic perspective on interaction. In Elizabeth Couper-Kuhler & Margret Selting (Eds.), *Prosody in conversation* (pp.11-56). Cambridge, England: Cambridge University Press.

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

de Bot, K., & Mailfert, K. (1982). The teaching of intonation: Fundamental research and classroom applications. *TESOL Quarterly, 16*(1): 71–77.

Derwing, T. M., & Munro, M. J. (2015). *Pronunciation fundamentals: Evidence-based perspectives for L2 teaching and research*. Amsterdam, Philadelphia: John Benjamins.

Derwing, T. M., & Munro, M. J. (2009). Putting accent in its place: Rethinking obstacles to communication. *Language Teaching*, *42*(04), 476-490.

Derwing, T. M., Munro, M. J., & Wiebe, G. (1998). Evidence in favor of a broad framework for pronunciation instruction. *Language Learning, 48*(3), 393–410.

Duanmu S. (2007). The Phonology of Standard Chinese. New York: Oxford University Press.

ETS 2017 Understanding Your TOEFL iBT Scores. Retrieved from https://www.ets.org/toefl/ibt/scores/understand/

Ford, C. E., & Thompson, S. A. (1996). Interactional units in conversation: Syntactic, intonational, and pragmatic resources for the management of turns. In E. Ochs, E. A. Schegloff, & S. A. Thompson (Eds.), *Interaction and grammar* (pp. 134–84). New York, NY: Cambridge University Press.

Gorsuch, G. J. (2011). Improving speaking fluency for international teaching assistants by increasing Input. *Journal of Teaching English As Second Language – Electronic Journal, 14*(4). 1-25.

Gorsuch, G., Meyers, C. M., Pickering, L. & Griffee, D. T. (2010). *English Communication for International Teaching Assistants*. Long Grove, Illinois: Waveland Press, Inc.

Grosz, B., & Hirschberg, J. (1992). Some intonational characteristics of discourse structure. *2nd International Conference on Spoken Language Processing*. 429-432.

Halic, O., Greenberg, K. & Paulus, T. (2009). Language and academic identity: A study of the experiences of nonnative English speaking international students. *International Education, 38*(2). Retrieved from: http://trace.tennessee.edu/internationaleducation/vol38/iss2/5

Halliday, M.A.K. (1967). *A course in spoken English.* London: Oxford University Press.

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

Hardison, D. M. (2005). Contextualized computer-based L2 prosody training: Evaluating the

effects of discourse context and video input. *CALICO Journal, 22,* 175–90.

Henderson, A., Frost, D., Tergujeff, E., Kautzsch, A., Murphy, D., Kirkova-Naskova, A., Waniek-Klimczak, E., Levey, D., Cunningham, U. & Curnick, L. (2012). The English Pronunciation Teaching in Europe Survey: Selected results. *Research in Language, 10*(1), 5-27.

Hincks, R. & Edlund, J. (2009). Promoting increased pitch variation in oral presentations with transient visual feedback. *Language Learning & Technology, 13*(3), 32-50.

Hurley, D. S. (1992). Issues in teaching pragmatics, prosody, and non-verbal communication. *Applied Linguistics, 13*(3), 259–281.

IBM Support (2017). Retrieved from http://www-01.ibm.com/support/docview.wss?uid=swg21482329

Kang, O. (2010). Relative salience of suprasegmentals features on judgments of L2 comprehensibility and accentedness. *System 38*, 301-315.

Kang, O., Rubin, D. & Pickering, L. (2010). Suprasegmental measures of accentedness and judgments of language learner proficiency in oral English. *The Modern Language Journal 94*(4), 554-566.

Krashen, S. (1985). *The Input Hypothesis*. Beverly Hills, CA.: Laredo Publishing Company.

Krashen, S. (1994). The input hypothesis and its rivals. In N. Ellis (Ed.) *Implicit and explicit learning of languages*. (p. 45-77). London: Academic Press.

Ladd, D. R. (1990). Intonation: Emotion vs. grammar. *Language, 66*(4), 806-816.

Levis, J. & Pickering, L. (2004). Teaching intonation in discourse using speech visualization technology. *System 32,* 505-524.

Levis, J. M. & Wichmann, A. (2015). English Intonation - Form and Meaning. In M. Reed & J. M. Levis (Eds.), *The Handbook of English Pronunciation* (pp. 139-155). West Sussex, UK: John Wiley & Sons, Inc.

Lima, E. F. (2015). Development and evaluation of online pronunciation instruction for international teaching assistants' comprehensibility (Unpublished doctoral dissertation). Iowa State University. Ames, Iowa.

Long, M. H. (1985). Input and second language acquisition theory. In S. Gass and C. Madden (Eds.) *Input in second language acquisition* (p. 377-393). Rowley, MA.: Newbury House.

Loveday, L. (1981). Pitch, politeness and sexual role: An exploratory investigation into the pitch correlates of English and Japanese politeness formulae. *Language and Speech, 24*(1), 71-89.

Lu, Y. & Miran, K. (2016). Prosody transfer in second language acquisition: Tonal alignment in the production of English pitch accent by Mandarin native speakers. *Tsing Hua Journal of Chinese Studies, New Series 46* (4), 785-816.

McGregor, A., Zielinski, B., Meyers, C. & Reed, M. (2016). An exploration of teaching intonation using a TED Talk. In J. Levis, H. Le., I. Lucic, E. Simpson, & S. Vo (Eds). Proceedings of the 7th Pronunciation in Second Language Learning and Teaching Conference, ISSN 2380-9566, Dallas, TX, October 2015 (143- 159). Ames, IA: Iowa State University.

Meng, H., Tseng, C. Y., Kondo, M., Harrison, A., & Viscelgia, T. (2009). Studying L2 suprasegmental features in Asian Englishes: A position paper. In Proceedings of the

Annual Conference of the International Speech Communication Association, *Interspeech.* 1715-1718.

Mixdorff, H. & Ingram, J. (2009). Prosodic analysis of foreign- accented English, *Interspeech*, 6-10.

Munro, M. J., & Derwing, T. M. (1995). Foreign accent, comprehensibility, and intelligibility in the speech of second language learners. *Language Learning, 45* (1), 73–97.

Pickering, L. (2004). The structure and function of intonational paragraphs in native and non- native speaker instructional discourse. *English for Specific Purposes, 23*, 19–43.

Pickering, L. (2001). The role of tone choice in improving ITA communication in the classroom. *TESOL Quarterly, 35,* 233–55.

Pickering, L., Corduas, M., Eisterhold, J., Seifried, B., Eggleston, A., & Attardo, S. (2009). Prosodic markers of saliency in humorous narratives. *Discourse Processes*, *46*(6), 517–540.

Pickering, L., & Wiltshire, C. (2000). Pitch accent in Indian-English teaching discourse. *World Englishes, 19*(2), 173–83.

Pierrehumbert, J. & Hirschberg, J. (1990). The meaning of intonational contours in discourse. In *Intentions in communication* (pp. 271–311). Cambridge: MIT Press.

Putro, N. H. P. S. (2013). Adobe audition and stress and intonation awareness. *3rd International Conference on Foreign Language Learning and Teaching, 2*(1), 87-96.

Ramírez Verdugo, D. (2006). A study of intonation awareness and learning in non-native speakers of English. *Language Awareness, 15*(3), 141-159. DOI: 10.2167/ la404.0

Sawir, E. (2005). Language difficulties of international students in Australia: The effects of prior learning experience. *International Education Journal, 6*(5), 567-580.

Silverman, K. E. A. & Pierrehumbert. J. B. (1990). The timing of prenuclear high accents in English, in J. Kingston & M. E. Beckman (Eds.), Papers in laboratory phonology 1: Between the grammar and physics of speech. (72 -106). Cambridge: Cambridge University Press.

Scotto di Carlo, G. (2014). The role of proximity in online popularizations: the case of TED talks. *Discourse Studies, 16*(5), 591-606.

Su, C., Tseng, C., & Visceglia, T. (2013). Underdifferentiation of English lexical stress contrasts by L2 Taiwan speakers. SLaTE.

Sümer, S., Poyrazli, S., & Grahame, K. (2008). Predictors of depression and anxiety among international students. *Journal of Counseling & Development, 86*(4), 429-437.

Swain, M. (1995). Three functions of output in second language learning. In. G. Cook and G. Seidhofer (Eds.) *Principles and practices in applied linguistics: Studies in honor of H. G. Widdowson* (p. 125-144). Oxford: Oxford University Press.

Swerts, M., & Geluykens, R. (1994). Prosody as a marker of information flow in spoken discourse. *Language and Speech*, *37*(1), 21–43.

Tanner, M. W., & Landon, J. M. (2009). The effects of computer-assisted pronunciation readings on ESL learners' use of pausing, stress, intonation, and overall comprehensibility. *Language Learning & Technology, 13*(3), 51–65.

Tench, P. (1996). *The INTONATION systems of English.* London: Cassell.

Tergujeff, E. (2010). Pronunciation teaching materials in Finnish EFL textbooks. In A. Henderson (Ed.), *English pronunciation: Issues and practices (EPIP): Proceedings*

of the First International Conference (pp. 189-205). Chambery, France: University de Savoie.

Tergujeff, E. (2013). *English pronunciation teaching in Finland* (Doctoral dissertation). University of Jyvaskyla, Finland.

The Institute of International Education, (2016). Open Doors Report on International Educational Exchange. Retrieved from https://www.iie.org/Research-and-Insights/Open-Doors.

Thompson, S. E. (2003). Text-structuring metadiscourse, intonation and the signaling of organisation in academic lectures. *Journal of English for Academic Purposes, 2*(1), 5–20.

Tyler, A. (1992). Discourse structure and the perception of incoherence in international teaching assistants' spoken discourse. *TESOL Quarterly,* 26(4), 713-729.

Vygotsky, L. (1962). *Thought and language*. Cambridge, MA.: MIT Press.

Warren, P., Elgort, I. & Crabbe, D. (2009). Comprehensibility and prosody ratings for pronunciation software development. *Language Learning & Technology, 13*(3), 87-102.

Wennerstrom, A. (2001). *The music of everyday speech: Prosody and discourse analysis*. New York: Oxford University Press.

Wennerstrom, A. (1998). Intonation as cohesion in academic discourse. *Studies in Second Language Acquisition, 20,* 1–25.

Wichmann, A. (2000). *Intonation in text and discourse: Beginnings, middles and ends.* England: Longman.

Xu, M. (1991). The impact of English-language proficiency on international graduate students' perceived academic difficulty. *Research in Higher Education, 32*(5), 557-570.

Xu, Y. (1998). Consistency of tone-syllable alignment across different syllable structures and speaking rates. *Phonetica,* 55, 179-203. doi: 10.1159/000028432

Xu, Y. & Wang, Q. (2001). Pitch targets and their realization: Evidence from Mandarin Chinese. *Speech Communication*, 33, 319-337. doi: 10.1016/S0167- 6393(00)00063-7

Yule, G. (1980). Speakers' topics and major paratones. *Lingua*, *52*, 33–47.

Zhang, Y., Nissen, S.L. & Francis, A.L.,(2008). Acoustic characteristics of English lexical stress produced by native Mandarin speakers," *Journal of the Acoustic Society of America, 123*(6), 4498-4513.

Appendices

Appendix 1: Text for script-reading speech

Apple opened its annual event Wednesday with a showcase for Apple Watch touting new apps from Facebook Messenger and GoPro Apple CEO Tim Cook said Apple Watch is helping our users live a better day noting customer satisfaction is at 97% after just a matter of months the first new app Apple signaled that is on the way is Facebook Messenger which allows users to send text audio messages and share their location from the device AirStrip an app that allows for messaging health-care-related issues also got some focus at the event.

Appendix 2: Pre-test interview protocol

Interview protocol

The goal of this interview is to collect more samples of extemporaneous speech for a research project. We will talk about your experience as a college student.

Q1: Can you introduce yourself?

Q2: How would you introduce your university and your major to someone who doesn't know about them?

Q3: Are you taking any English classes? What are they?

Q4: Now I'd like to know you on a day-to-day basis. Can you describe a typical day on campus?

Q5: What based on your experience makes a successful experience for a college student?

Q6: What is your biggest gain in a year? What challenges/problems have you met?

To view online: http://ucsbeducation.qualtrics.com/SE/?SID=SV_5hzGcNS07d5QQjX

Appendix 3: Pre-test survey questions

Q1. What is your name/StudentID?

Q2. What is your age?

Q3. What is your gender?

Q4. What is your current or intended major?

Q5. What is your current Class Year?

Freshman Sophomore Junior Senior Graduate Student

Q6. What is your TOEFL speaking score?

Q7. Have you lived in an English-speaking country/region? If yes, where, and for how long?

Q8. Do you speak English outside of class? With whom and how often?

Q9. Where and how long have you been studying English?

Q10. About how many hours a week do you generally spend studying English?

Q11. Where and how many pronunciation courses have you taken?

Q12. Answer the following questions using the scale.

Multiple times a day Several times a week Several times a month Once in a while Never

How often do you listen to English songs?

How often do you watch English TV shows or movies?

How often do you use English in an online environment (e.g., read or write on English websites)?

Q13. Answer the following questions using the scale.

Extremely Important Very Important Neither Important nor Unimportant Very Unimportant Not at all Important

How important is good English pronunciation to you?

How important is good English intonation to you?

How important is having good English grammar to you?

How important is having a good English vocabulary to you?

Q14. Answer the following questions using the scale.

Poor Fair Good Very Good Excellent

How would your rate your overall English skills?

How would your rate your English speaking skills?

How would your rate your English pronunciation?

How would your rate your English intonation (e.g., stress, rhythm)

Q15. How would you rate your English pronunciation courses?

Never Rarely Sometimes Often All of the Time

My instructors value good English pronunciation

My instructors give explicit instructions on English pronunciation

My instructors provide exercises on English pronunciation

My instructors give feedback on pronunciation errors

I am satisfied with my English pronunciation courses

Q16. What are the things that you like the most about your English pronunciation courses?

Q17. What are the things that you do not like about your English pronunciation courses? How would you suggest improve it?

To view online: http://ucsbeducation.qualtrics.com/jfe/form/SV_dmAzuCudXH3Y7xH

Appendix 4: Post-training survey questions

Q1. This survey will ask for your feedback towards the training that you've completed in the past few months.

The term "intonation" covers pausing, prominence and tone choices in the following questions.

Q2. After taking the training modules, I think that

Strongly agree Agree Somewhat agree Neither agree nor disagree Somewhat disagree Disagree Strongly disagree I don't know

I've realized the importance of English intonation

I've gained new knowledge about English intonation

I will pay more attention to intonation when I listen

I will pay more attention to intonation when I speak

I'm more confident about my spoken English

My spoken English has improved

Q3. In terms of improving my English intonation, I found that

Strongly agree Agree Somewhat agree Neither agree nor disagree Somewhat disagree Disagree Strongly disagree I don't know

watching instruction videos and answering questions were helpful

listening and identifying intonation was helpful

reading the script with marked intonation was helpful

predicting intonation and reading the script was helpful

comparing my recording to the native speakers' by listening was helpful

giving spontaneous speeches was helpful

getting feedback on my speeches was helpful

Q4. In terms of using Praat software to learn tone choices, I found that

Strongly agree Agree Somewhat agree Neither agree nor disagree Somewhat disagree Disagree Strongly disagree I don't know

the activities using Praat were helpful

viewing the native speakers' pitch curves was helpful

viewing my own pitch curves was helpful

comparing my own pitch curves to the native speakers' was helpful

Q5. Do you have any comments about the project or suggestions for making it more helpful to you?

Appendix 5: Rating survey questions

Default Question Block

In this survey, you will listen to the beginning of the speaker's speech (about the first 30 seconds) on this topic:
I'd like to know you on a day to day basis. Can you describe a typical day on campus?

In general, the speaker to whom I just listened

	1	2	3	4	5	6	7	
was easy to understand	○	○	○	○	○	○	○	was difficult to understand
spoke smoothly	○	○	○	○	○	○	○	spoke in a choppy manner
has no accent	○	○	○	○	○	○	○	has strong foreign accent
was engaging	○	○	○	○	○	○	○	was boring
was confident	○	○	○	○	○	○	○	was hesitant

The speaker to whom I just listened spoke at a rate that was

Extremely fast	Moderately fast	Slightly fast	Appropriate speed	Slightly slow	Moderately slow	Extremely slow
○	○	○	○	○	○	○

The speaker to whom I just listened

	Strongly agree	Agree	Somewhat agree	Neither agree nor disagree	Somewhat disagree	Disagree
spoke run-on sentences without pausing	○	○	○	○	○	○

	Strongly agree	Agree	Somewhat agree	Neither agree nor disagree	Somewhat disagree	Disagree
spoke choppily with constant pausing	○	○	○	○	○	○
paused in the wrong places	○	○	○	○	○	○
hesitated too much with fillers such as 'eh', 'um'	○	○	○	○	○	○

The speaker to whom I just listened

	Strongly agree	Agree	Somewhat agree	Neither agree nor disagree	Somewhat disagree	Disagree
emphasized too many words	○	○	○	○	○	○
didn't emphasize enough words	○	○	○	○	○	○
emphasized inappropriate words in a sentence	○	○	○	○	○	○

The speaker to whom I just listened

	Strongly agree	Agree	Somewhat agree	Neither agree nor disagree	Somewhat disagree
spoke flat/monotonous speech	○	○	○	○	○
used too many falling and level tones and sounded distant	○	○	○	○	○
used too many rising tones and sounded unconfident	○	○	○	○	○
used pitch changes in a confusing manner (e.g., using falling tones when continuing and rising tones when finishing)	○	○	○	○	○