

THE EFFECT OF WEB-BASED PRAGMA-PROSODIC INSTRUCTION
AND APTITUDE ON LEARNER PRAGMATIC DEVELOPMENT

By Maria Kostromitina

A Dissertation

Submitted in Partial Fulfillment

of the Requirements for the Degree of

Doctor of Philosophy

in Applied Linguistics

Northern Arizona University

May 2023

Approved:

Okim Kang, Ph.D., Chair

Naoko Taguchi, Ph.D.

Jesse Egbert, Ph.D.

Alyssa Kermad, Ph.D.

ABSTRACT

THE EFFECT OF WEB-BASED PRAGMA-PROSODIC INSTRUCTION AND APTITUDE ON LEARNER PRAGMATIC DEVELOPMENT

MARIA KOSTROMITINA

Despite the importance that learners must place on using prosody appropriately in EIL interaction, pragmatic functions of prosody have been largely disregarded in teaching materials and classroom instruction (Nikolić, 2018). Moreover, with a recent change in classroom paradigms instigated by the global development of educational technologies and reinforced by the recent pandemic, there is an increased need for pragmatic instruction that can be delivered to learners digitally and remotely (Taguchi, 2011b). Although distance learning has played an important role in English teaching and learning long before the pandemic, its impact is even more meaningful today, when English learners all over the world are still experiencing difficulties related to in-person instructional access.

The current mixed-methods quasi-experimental dissertation study investigated the extent to which English learners' interaction skills can improve as a result of contextualized prosody instruction on a web-based platform. Seventy-four English learners from China were randomly assigned to a two-week program that consisted of explicit pragma-prosodic instruction as well as perception- and production-based exercises. Learners' use of prosody in requests before and after instruction was recorded through video-based role plays, analyzed acoustically, and evaluated for appropriateness, accentedness, and comprehensibility by 20 trained raters. In order to provide a complete picture of instructional effectiveness, the study accounted for variation in learners' aptitude for acquiring prosodic properties of English by measuring their working memory, musical sensitivity, and auditory processing skills.

The study results demonstrated significant changes in learners' use of pragma-prosodic features of interaction after the training; however, these changes did not have an effect on listeners' appropriateness perceptions, highlighting the complex nature of this construct that requires further clarification. At the same time, a significant albeit weak relationship was detected among listeners' perceptions of accentedness, comprehensibility, and appropriateness. Finally, the results revealed that learners' musical sensitivity, working memory as well as tone choice and prominence discrimination ability affected the success of the training in improving their use of pragma-prosodic features from the acoustic perspective. These findings provide important guidance for designing digitally mediated contextualized prosody instruction and help to better understand the role of learners' aptitude in its effectiveness.

Keywords: pragmatics; prosody; intonation; web-based instruction; pragma-prosodic appropriateness; individual differences; working memory; musical sensitivity; auditory processing; speech perception; speech production

Maria Kostromitina

© 2023

ACKNOWLEDGMENTS

Although the cover page of this document reads “by Maria Kostromitina”, there is a number of people without whom this dissertation would not have been possible and to whom I am forever grateful. First and foremost, I’d like to thank my advisor, mentor, and teacher, Dr. Okim Kang. There was not a single moment when I did not feel her unwavering support in my career and life decisions. Through the many research opportunities she involved me in, I was able to enhance my research skills and gain invaluable experience managing projects. Thank you for pushing me when things were tough and for making me a part of your PhD family.

My gratitude goes to my dissertation committee whose contributions to this dissertation study and my development as a researcher are hard to overstate. I am thankful to Dr. Naoko Taguchi who guided me in finding and refining my research interests, expanding my methodological arsenal, and designing a methodologically sound pragmatic study. I learned all of the quantitative analysis and programming skills that I used in this dissertation (and other studies) from Dr. Jesse Egbert. His patience in explaining complex statistical concepts and utmost attention to every student question, big or small, were extremely valuable during my learning process. Finally, Dr. Alyssa Kermad has been an inspiration as a scholar and a member of the AL community whose research on the pragma-prosodic interface has informed and enriched my work. Her thorough feedback on the many drafts of this study was inestimable.

Beyond my dissertation committee, I feel extremely lucky to have had a chance to learn from brilliant faculty members at NAU. When I came to Flagstaff to pursue my MA in TESL almost seven years ago, I had the opportunity of a lifetime to take courses with the legends in the field, Dr. Fredricka Stoller, Dr. Randi Reppen, and Dr. Doug Biber. The knowledge I gained from those courses, and more importantly their sheer enthusiasm for the field, has stayed with

me until now. Later on, I was privileged to learn from and work with Dr. Bill Crawford, Dr. Tove Larsson, and Dr. Romy Ghanem. There were two more faculty members who played fundamental roles for my academic career and shaped my identity as a scholar and an educator. Dr. Luke Plonsky showed me the value of collaboration, instilled the importance of high-quality research and open science. To top it off, Luke did it all with an air of positivity that I truly admire. Dr. Vedran Dronjic, with his deep psycholinguistic expertise, was instrumental in developing my dissertation idea and providing guidance during my earlier stages in the program.

I don't think I would have made it through these years without my Ph.D. support system, my amazing, talented friends, Kevin Hirschi, Larissa Goulart, Margaret Wood, Amanda Black, and Vito Miao, who made going through this Ph.D. experience that much easier. Thank you for our after-class gatherings, inside jokes, world cuisine-inspired dinners, and friendship. The support system I have had at NAU extends beyond this group to the rest of my peers, colleagues, and friends who were crucial at all stages of this study, from design to piloting and data collection.

There were several funding sources that made this dissertation study possible: *The International Research Foundation* Doctoral Dissertation Grant, *Duolingo, Language Learning*, the *National Federation of Modern Language Teachers Associations*, as well as the Graduate College at NAU. I am very thankful to these organizations for providing the means for conducting the study and supporting graduate research.

I owe everything to my family, my grandparents, and my parents Sveta and Vova. From age 6, when my grandmother gave me an English-Russian dictionary and fostered my love for languages, to this day, they are my support system (even if they are thousands of miles away) and my home. Thank you for instilling the love for learning and for always encouraging my

endeavors. Спасибо за всё. Я вас очень люблю. To my American family, the Campbells, the Rymans, and the Erforts, – this dissertation is a result of you cheering me on and being my home away from home.

Finally, I cannot even begin to imagine what this journey would have been like without my husband Kyle. He provided the technical support in developing the instructional platform and acted as the sounding board for my ideas. Thank you for always making sure I am fed, for taking care of our home, for our brain-break afternoon walks, for indulging my late-night coffee runs, and for reminding me of the big picture.

TABLE OF CONTENTS

ABSTRACT	ii
ACKNOWLEDGMENTS	v
List of Tables	xii
List of Figures	xiv
DEDICATION	xv
CHAPTER 1: INTRODUCTION.....	1
Current Issues at the Intersection of Second Language Prosody and Pragmatics.....	1
Statement of the Problem	4
Significance of the Study	6
Definition of Terms	8
Outline and Overview of Dissertation Proposal.....	9
CHAPTER 2: LITERATURE REVIEW	10
Defining Intonation vs. Prosody vs. Suprasegmentals	11
Describing Prosody and Its Components	12
Discourse Intonation Framework	12
Discourse Functions of Intonation in English.....	15
Requests in the Speech Act framework.....	18
Pragmatics and Prosody	20
Prosody and Speech Acts	26
Pragma-Prosodic Appropriateness in Speech Acts	31
Pragma-Prosodic Needs of Chinese Learners of English.....	35
Theory and Practice of L2 Prosody Instruction	37
L2 Prosody Acquisition Theories	38
Learners' Ability to Perceive Intonation.....	41
Empirical Research on L2 Prosody Instruction.....	45
Instructional Targets, Type, and Length of Instruction	45
Visual Aids in L2 Prosodic Instruction	49
Research on Pragma-prosodic Instruction	51
Cognitive Individual Differences and L2 Pragma-Prosodic Acquisition.....	54
Working Memory	55
Musical Ability	58

Auditory Processing	60
The Study	63
Research Questions	64
CHAPTER 3: METHODS	65
Research Design	65
Participants	67
Learners	67
Raters	71
Baseline Speakers	73
Instruments and Materials	74
Learner and Rater Background Questionnaires	74
Pragma-Prosodic Instructional Design	75
Pre-Test and Post-Test Materials	99
Aptitude Measures	102
Timeline of Experimental Group Data Collection	109
Elicited Imitation Task	110
Raters' Perception Scales	111
Measured Constructs	111
Scale Functioning	112
Scales Used in the Study	116
Rater Training	118
Control Group Instructional Content and Its Delivery	119
Data Collection Procedures	121
Preparation of Sound Files for Rater Assessment	123
Prosodic Analysis	123
Analysis of Learners' Intonation Perception Ability	126
Study Variables	127
CHAPTER 4: RESULTS	129
Research Question 1: Changes in Learners' Use of Prosody After Training	130
Evidence from Acoustic Analysis	130
Evidence from Perception Surveys	153
Summary of Results for Research Question 1	161

Research Question 2: Role of Learners' Individual Differences in Their Use of Prosody	163
Individual Differences and Use of Prosody Before Training	165
Individual Differences and Use of Prosody After Training	168
Summary of Results for Research Question 2	170
Research Question 3: Ability to Perceive Intonation Cues and Training Effectiveness	170
Summary of Results for Research Question 3	174
CHAPTER 5: DISCUSSION	176
Research Question 1: Learners' Use of Prosody Before and After the Training	178
Research Question 1.1: Acoustic Analysis of Learners' Use of Prosody	179
Research Question 1.2: Perceptual Analysis of Learners' Use of Prosody	186
Research Question 2: Learners' Cognitive Individual Differences and Use of Prosody	190
Research Question 3: Learners' Ability to Perceive Prosodic Cues and Prosodic Use	196
CHAPTER 6: IMPLICATIONS AND DIRECTIONS FOR FUTURE RESEARCH	200
Implications	200
Implications for Second Language Pedagogy and Material Development	200
Implications for L2 Pronunciation and Pragmatics Research	204
Implications for Second Language Acquisition	208
Limitations and Future Research	211
Concluding Remarks	215
REFERENCES	217
Appendix A: Learner Background Questionnaire	248
Appendix B: Rater Background Questionnaire	250
Appendix C: Elicited Imitation Task	252
Appendix D: Instructional Training Lesson Plans	253
Appendix D1: Day 1	253
Appendix D2: Day 2	256
Appendix D3: Day 3	260
Appendix D4: Day 4	263
Appendix D5: Day 5	265
Appendix D6: Day 6	267
Appendix E: Pre- and Post-training DCT Scripts	268
Appendix F: Listener Perception Scales	272

Appendix G: Rater Training Session	274
Appendix H: Working Memory Task Stimuli	281
Appendix I: Final Questionnaire.....	286
Appendix J: Instructional Content and Activities for the Control Group.....	287

List of Tables

Table 1: <i>Content of Pragma-Prosodic Instruction</i>	88
Table 2: <i>Overview of the Pragma-Prosodic Instructional Training</i>	97
Table 3: <i>Summary of the DCT Situations</i>	100
Table 4: <i>Example Completion Timeline for Experimental Group</i>	110
Table 5: <i>Data Collection Steps</i>	122
Table 6: <i>Selected Intonation Properties and Measures</i>	125
Table 7: <i>Operationalizations of the Variables in the Study</i>	127
Table 8: <i>Descriptive Statistics for Baseline Speakers' Use of Tone Choices and Prominence</i>	132
Table 9: <i>Baseline Speakers' Use of Tone Choices and Prominence in Request Head Acts</i>	133
Table 10: <i>Descriptive Statistics for Learners' Use of Tone Choices and Prominence</i>	134
Table 11: <i>Summary of the Binomial Mixed Effects Model for Tone Choices</i>	137
Table 12: <i>Results of Post Hoc Pairwise Tukey HSD Tests for Tone Choices</i>	138
Table 13: <i>Summary of the Binomial Mixed Effects Model for Tone Choices with Delayed Post Test</i>	140
Table 14: <i>Results of Post Hoc Pairwise Tukey HSD Tests with Delayed Post Test for Tone Choices</i>	142
Table 15: <i>Summary of the Mixed Effects Model for Prominence (Space)</i>	143
Table 16: <i>Results of Post Hoc Pairwise Tukey HSD Tests for Prominence (Space)</i>	144
Table 17: <i>Summary of the Mixed Effects Model for Prominence (Appropriate Use)</i>	146
Table 18: <i>Results of Post Hoc Pairwise Tukey HSD Tests for Prominence (Approp. Use)</i>	147
Table 19: <i>Summary of the Mixed Effects Model for Prominence with Delayed Post Test (Space)</i>	148
Table 20: <i>Results of Post Hoc Pairwise Tukey HSD Tests with Delayed Post Test for Prominence (Space)</i>	148
Table 21: <i>Summary of the Mixed Effects Model for Prominence with Delayed Post Test (Appropriate Use)</i>	150
Table 22: <i>Results of Post Hoc Pairwise Tukey HSD Tests with Delayed Post Test for Prominence (Appropriate Use)</i>	152

Table 23: <i>Correlations and 95% CI Between Scores on Accentedness and Comprehensibility Scales</i>	154
Table 24: <i>Fleiss' Kappa Inter-Rater Reliability</i>	155
Table 25: <i>Correlations and 95% CI Between Accentedness, Comprehensibility, and Appropriateness Ratings</i>	156
Table 26: <i>Descriptive Statistics for Perceptual Ratings</i>	156
Table 27: <i>Results of LMEM for Appropriateness Perception</i>	159
Table 28: <i>Results of a Multiple Regression Analysis with Prosodic Features and Appropriateness Ratings</i>	160
Table 29: <i>Descriptive Statistics for Cognitive IDs and Proficiency</i>	164
Table 30: <i>Correlation Matrix of Predictor Variables</i>	164
Table 31: <i>Predictor Variable VIF Values</i>	165
Table 32: <i>Results of a Multiple Regression Analysis with Tone Choices Before Training</i>	166
Table 33: <i>Results of a Multiple Regression Analysis with Prominence Before Training</i>	167
Table 34: <i>Results of a Multiple Regression Analysis with Log-Transformed Tone Choices After Training</i>	169
Table 35: <i>Results of a Multiple Regression Analysis with Prominence After Training</i>	169
Table 36: <i>Descriptive Statistics for Perceptual Predictor Variables</i>	171
Table 37: <i>Perceptual Predictor Variable VIF Values</i>	172
Table 38: <i>Correlation Matrix of Perceptual Predictor Variables</i>	172
Table 39: <i>Results of a Multiple Regression Analysis with Tone Choices After Training</i>	173
Table 40: <i>Results of a Multiple Regression Analysis with Prominence After Training</i>	174

List of Figures

Figure 1: <i>Example of a Tone Unit in Brazil's (1997) Model</i>	13
Figure 2: <i>Design of the Study</i>	66
Figure 3: <i>Interface of the Tone Choice Identification Activity</i>	93
Figure 4: <i>Interface of the Prominence Identification Activity</i>	93
Figure 5: <i>Interface of the Intonation Discrimination Activity</i>	94
Figure 6: <i>Interface of the Controlled Production Activity</i>	94
Figure 7: <i>Sample Parselmouth and Chart.js Output</i>	97
Figure 8: <i>Interface of the Pre- and Post-Tests</i>	102
Figure 9: <i>Sample Trial of the PROMS Melody Subset (Law & Zentner, 2012)</i>	106
Figure 10: <i>Auditory Processing Task Interface</i>	109
Figure 11: <i>PRAAT Analysis of Tone Choices and Prominence</i>	124
Figure 12: <i>Residuals Plot for Tone Choices</i>	135
Figure 13: <i>Residuals Plot for Prominence</i>	136
Figure 14: <i>Residuals Plot for Prominence (Appropriate Use)</i>	137
Figure 15: <i>Tone Choice Production in Pre- and Post-Tests</i>	139
Figure 16: <i>Tone Choice Production in Pre-, Post-, and Delayed Tests</i>	142
Figure 17: <i>Prominence Production in Pre- and Post-Tests (Space)</i>	145
Figure 18: <i>Prominence Production in Pre- and Post-Tests (Appropriate Use)</i>	147
Figure 19: <i>Prominence Production in Pre-, Post-, and Delayed Tests (Space)</i>	149
Figure 20: <i>Prominence Production in Pre-, Post-, and Delayed Tests (Appropriate Use)</i>	152
Figure 21: <i>Residuals Plot for Appropriateness Perceptions</i>	158

DEDICATION

To both of my grandmothers – Валентина и Валентина

CHAPTER 1: INTRODUCTION

Current Issues at the Intersection of Second Language Prosody and Pragmatics

Speech production and perception are complex cognitive processes that are integral to human communication. We start producing speech by conceptualizing a message in our cognition, using grammatical and phonological means to encode it, and articulating it by moving parts of our vocal tract (Levelt, 1989). We then extract acoustic-phonetic information from the speech signal in order to map them onto cognitive representations and process the received message through our speech comprehension system (McQueen & Cutler, 2010). In both of these processes, the phonetic and phonological elements of speech fulfil several crucial functions related to successful communication of lexico-grammatical, attitudinal, discursual, and pragmatic meaning.

Moreover, the way we produce and perceive speech can vary depending on the language we are speaking. While in our first language (L1), the neural processing of speech is automatic and occurs rapidly, in the second language (L2), the processing of phonetic and phonological aspects of speech is more effortful, requiring greater attention and cognitive resources (Golestani & Zatorre, 2009), in part due to the underdeveloped L2 phonological system and consequent difficulties with L2 pronunciation (Munro & Derwing, 1999).

L2 learners' difficulties in acquiring and producing sounds in a target language gave rise to the domain of L2 pronunciation that has been one of the most rapidly growing areas of second language acquisition (SLA). From speech production perspective, studies have examined L2 learners' acquisition of segmental features of speech like long vowels and moraic nasals in Japanese (Iizuka *et al.*, 2020) low and high functional load phonemes (e.g., Suzukida & Saito, 2019) and word pronunciation accuracy (Trofimovich & Gatbonton, 2006), to the learning of

suprasegmentals including lexical tones (e.g., Pelzl *et al.*, 2019), intonation and fluency aspects of pronunciation (Derwing & Rossiter, 2003), to a combination of both segmental and suprasegmental features (e.g., Gordon *et al.*, 2013; Kissling, 2018; Kostromitina & Kang, 2021; Sun *et al.*, 2021). From the point of view of speech production, the acquisition of pronunciation features has been examined in connection with judgements about L2 speech accentedness and comprehensibility (Gordon & Darcy, 2016; Suzukida & Saito, 2019; Kang, 2010; Kang *et al.*, 2020). Since L2 pronunciation is oftentimes acquired in instructional settings, the design features of pronunciation training, such as the length of instruction, inclusion of an audio-visual component, explicit vs. implicit instruction, aspects of feedback, have been investigated in both primary and secondary research (e.g., Lee *et al.*, 2015, 2020; McAndrews, 2019; Saito & Plonsky, 2019). Furthermore, due to the complex cognitive nature of L2 pronunciation acquisition, research has tapped into learners' cognitive individual differences (IDs), furthering our understanding of the impact of learners' aptitude (Saito, 2019), age and motivation (Moyer, 1999; Saito, 2012), and bilingualism (Zembrzuski *et al.*, 2018; Schmid & Moosmüller, 2017) in L2 pronunciation acquisition.

Combined, existing studies come together in an intricate mosaic of L2 pronunciation acquisition research; however, this mosaic is not devoid of blank spots. In current research, suprasegmental learning, and prosodic learning in particular, has been limited to investigations of L2 speakers' intelligibility, comprehensibility, and accentedness (e.g., Gordon & Darcy, 2016; Thomson & Derwing, 2015) as well as speech fluency (Trofimovich & Baker, 2007) and the role of L2 suprasegmental features in their perceptions. While these are important aspects of spoken communication, little is known about the learning of another function of L2 prosody, namely, the acquisition of its pragmatic meaning. This gap in our understanding is remarkable, especially

since it is known that prosody carries pragmatic meaning and plays a fundamental role in spoken English discourse, while also helping to identify a speaker's intention behind an utterance (Brazil, 1997; Kang *et al.*, 2021; Pierrehumbert & Hirschberg, 1990).

Initial existing research has suggested that prosodic cues of L2 English speakers, as part of pragmatic meaning-making in spoken discourse, can be easily misinterpreted by native speakers of a target language. Thus, deviations in language learners' intonation patterns may have negative pragmatic consequences on spoken discourse (Anderson-Hsieh *et al.*, 1992; Derwing & Munro, 1997; Pickering, 2009). Within the speech act framework, Staples (2015) analyzed greetings/introductions and expressions of empathy produced by internationally and U.S. trained nurses discovering that L1 nurses were able to use intonation more effectively to express compassion. In the academic context, Ramírez-Verdugo (2008) and Kang *et al.* (2021) revealed that type of speech act and degree of imposition were associated with different tone choice and pitch accent patterns used by L2 speakers that deviated from L1 use of intonation. Building on Kang *et al.*, Taguchi *et al.* (2021) investigated the longitudinal development of prosodic properties in L2 requests and opinions finding evidence of prosodic development overtime with learners' prominence and pitch range choices approaching a NS baseline.

Perceptually speaking, the use of prosodic cues in L2 speech acts has been associated with speakers' appropriate participation in discourse. L1 speakers' perception of speech acts produced in various languages have indicated that particular intonation patterns and overuse of sentence stress may make an L2 speaker be perceived as impolite (Herrero & Devís, 2020; Kostromitina, in press). It has also been suggested that L1 interlocutors can pick up on L2 speakers' attitudes in producing speech acts with a certain intonation (e.g., Kermad, 2021).

When examining acquisition of a certain linguistic feature, especially that involved in such cognitively demanding processes as L2 speech perception and production, it is essential to consider learners individual differences since they may impact how learners perceive, process, and produce that feature (DeKeyser, 2013; Chun et al., 2015). Research has provided emerging explanations for cognitive processes at play in learners' acquisition of L2 prosody and its pragmatic functions. Learners' personality, aptitude, and affective factors have been examined acquisition of L2 prosody (Kissling, 2014; Suzukida, 2021) and pragmatics (Li, 2017; Takahashi, 2019). With regard to pragmatics, some IDs like learners' L1 and prosodic transfer have received growing attention from researchers (e.g., Gabriel & Kireva, 2014; Ortega-Llebaria & Colantoni, 2014; Rasier & Hiligsmann, 2007).

Statement of the Problem

Despite the clear link between prosody and pragmatics in L2 spoken discourse, there is an apparent disconnect between existing research and instructional practices. For one, the discourse pragmatic functions of prosody have been largely ignored in pronunciation textbooks that miss the communicative focus of intonation (Levis, 1999). Furthermore, while intonation activities are underrepresented overall (5% to 8% of all exercises), discourse intonation exercises for specific speech acts are nearly non-existent (Nikolić, 2018; Trouvain *et al.*, 2016). Not only that, but there is also a lack of clarity about the approaches to teaching prosodic features within pragmatic contexts.

Although primary and secondary studies have shown that even short, explicit prosodic training with an audio-visual component can be effective in improving learners' use of intonation, very few studies have framed such training within specific communicative situations (e.g., McKinnon, 2017), and none have investigated the prosodic instruction within a speech act

framework. In addition, a larger issue related to pragma-prosodic instruction concerns the contents of training. Even though research has indicated that L1 and L2 use of intonation in speech acts is different, it is uncertain whether learners can perceive prosodic cues presented to them in the instruction. There is also a risk that such an approach would ignore the perceptual aspect of L2 pragmatic meaning comprehension (Taguchi, 2019). That is, bringing L2 use of prosody closer to the L1 baseline does not necessarily indicate that learners' production is perceived as more polite or appropriate by L1 interlocutors. Deviations in prosody have been shown to affect listeners' perceptions of L2 speakers' comprehensibility and accentedness (e.g., Kang, 2010; Pickering, 2001); yet, no such connection has been established between prosody and perceived appropriateness.

In addition, while some of the learner IDs have received attention in instructional studies in L2 pragmatics and prosody (e.g., Kissling, 2014; Li, 2017), other relevant learner characteristics that comprise learners' aptitude like working memory (WM) or musical ability have been understudied, especially from the pragma-prosodic point of view. This is surprising, especially if the established significance of these measures in L1 intonation acquisition is considered (Baddeley, 2003; Nathoo *et al.*, 2005). When it comes to L2 intonation, however, the role of these aptitude ID measures requires clarification. In particular, the existing studies of WM and L2 intonation offer conflicting findings concerning the relevance of WM for the acquisition of L2 prosody (e.g., Hu *et al.*, 2013). Even though the role of musical aptitude in L2 intonation overall is consistent with L1 findings, its link with prosodic instruction needs further investigation; some argue that musical ability is unimportant for the effectiveness of such instruction (Schwab & Dellwo, 2019) while others claim the opposite (Komar, 2017). At the same time, with the growing interest in L2 intonation learning, other speech-specific individual

difference factors, like auditory processing, start gaining recognition as effective predictors of the acquisition of L2 prosody (e.g., Saito *et al.*, 2020a; Sun *et al.*, 2021). Similarly, L2 pragmatic research has provided emerging evidence of the role of certain learner aptitude factors in pragmatic acquisition (e.g., Li, 2017 for working memory); however, the evidence is scarce and does not investigate the role of learner factors on the intersection of pragmatic and prosodic acquisition.

Significance of the Study

The proposed study is located at the intersection of the domains of L2 prosody instruction, pragmatics, and learners' cognitive individual differences. Specifically, the study will examine the effectiveness of a computer-based instructional training in improving learners' use of intonation in high-imposition requests via several lines of evidence such as acoustic analysis and perceptual ratings. The study will concentrate particularly on high-imposition requests as these speech acts have been defined as particularly face-threatening for interlocutors (Brown & Levinson, 1987), and thus require speakers' attention to several aspects of their performance (Taguchi, 2006). The production of L2 requests has also been found to differ between L1 and L2 English speakers both on the pragmalinguistic level (e.g., Lundell & Erman, 2012) and on the prosodic level (e.g., Kang *et al.*, 2021; Taguchi, 2021).

The study will first examine the extent to which L2 English learners' use of prosodic features change as a result of pragma-prosodic instruction. This change will be quantified through tone choices, which will be split into three separate categories (rising, falling, level tones) and prominence (stressed words in a tone unit).

Since intonation has been found to differ significantly in L1 and L2 speech act production (Kang *et al.*, 2021; Taguchi *et al.*, 2021), learners' pragmatic production of requests before and

after the training will be prosodically analyzed and compared to baseline L1 and highly intelligible L2 production of the same speech acts. The comparison of learners' pre- and post-training production with the baseline samples will indicate whether the instruction was effective at getting learners' use of tone choices and prominence closer to the target-like models.

In addition to relying on phonological analyses, the study will investigate the effectiveness of the training from the perceptual judgement point of view. The extent to which the training affects learners' perceived pragmatic appropriateness will be determined by means of eliciting perceptual judgements from L1 and L2 English listeners about the appropriateness of learners' productions, which will encompass politeness, formality, directness, and sincerity. On top of that, listeners' perceptions of learners' accentedness and comprehensibility will be measured to moderate the possible effect of this variable on raters' perceptions of appropriateness. The perceptual judgements will provide an additional piece of evidence about the effectiveness of the training reflected in the possible change in listeners' ratings. The two types of analysis, namely, the prosodic analysis and the perceptual analysis of learners' pre- and post-instruction speech samples, will be triangulated to find relationships between the prosodic features that may be associated with more or less appropriate speech act production.

The last immediate issue in need of resolving will consist of examining the extent of the changes in learners' use of prosodic features is explained by cognitive individual differences related to intonation learning aptitude. Three predictors of intonation learning aptitude will be examined: (1) working memory (a cognitive system used to store information over short time periods while manipulating it in some way), (2) musical ability (learners' ability to internalize melodic contours), and (3) auditory processing (ability to encode, remember, and reproduce time and frequency characteristics of sounds). The study will investigate the ability of these variables

to predict changes in learners' use of intonation in their production of requests and listeners' perceptual judgments after the instructional training.

The results of the proposed dissertation study will be enlightening not only to the domains of L2 pronunciation, L2 pragmatics, and SLA, but also to the researchers and practitioners who work at the interface of these domains. Besides validating the designed instructional training and making it available to L2 English instructors, the findings of the study will shed further light on the role of prosody in L2 pragmatics from both the perceptual angle and the contrastive analysis of baseline L1 and L2 use of intonation. It will advance the fields of speech perception and pragmatics by providing a clearer understanding of the connection between the pragmatic meaning of prosody and its perceived appropriateness. The study will also provide a readily available instructional tool that, with some training, can be used by English learners and teachers globally. Finally, in terms of learner individual differences, the study will expand our knowledge of how second language intonation is acquired and what role learners' aptitude plays in this process.

Definition of Terms

The section provides definition of some important terms that will be used throughout the proposal.

Discourse - continuous stretch of language (especially spoken) larger than a sentence

Intonation / prosody / suprasegmentals – functional melodic qualities of speech including pitch movement, pause distribution, and prominence, as contrasted with segmentals, or monophthongs, diphthongs, and triphthongs. Following Pickering (2018), these terms will be used interchangeably throughout the dissertation.

Pragma-prosody – an aspect of the prosodic structure of speech that involves form-function mappings for prosodic features and pragmatic meaning associated with them

Prominence – relative weight that is given to a syllable at the discourse level with regards to pitch, loudness, and length.

Tone unit – an organizational unit of speech similar to a thought group, that is guided by speakers' perception of semantic or syntactic chunks of language.

Tone choice – pitch movement associated with the last prominent syllable in a tone unit, or a stretch of language with prosodic characteristics.

Speech act – one or more utterances that are performed with speech in order to achieve real-world effects.

Appropriateness – the extent to which a listener finds a speaker's production of a request polite, direct, formal, and sincere.

Accentedness – a listener's perception of the degree a speaker's oral production deviates from the L1 pronunciation norms.

Outline and Overview of Dissertation

The rest of this dissertation is organized in the following manner. Chapter 2 of the study situates the study within several theoretical frameworks in the fields of L2 intonation, pragmatics, pronunciation instruction, and intonation development aptitude. The chapter opens with an overview of the existing approaches to describing prosody and zooms in on the discourse intonation model. The chapter also reviews the underlying principles of the speech act theory, focusing on requests. The theoretical section is followed by a detailed review of the existing

research in the intersection of L2 pragmatics and prosody as well as the studies in L2 prosody instruction. Finally, research on the cognitive individual differences that comprise learners' intonation learning aptitude and acquisition of L2 prosody is reviewed in the last section of the chapter. In the end, Chapter 2 builds on the reviewed research to motivate the current research questions. Chapter 3 details the methodological design of the study opening with a description of the participants and instruments in the study and advancing to the data collection procedures and the anticipated analyses. Subsequently, Chapter 4 specifies the hypotheses of the study and presents the results of statistical analyses conducted to address the research questions and test the hypotheses. Chapter 5 then contextualizes the findings for each research question in light of previous research. Chapter 6 closes the dissertation with implications of the findings and suggestions for future investigations.

CHAPTER 2: LITERATURE REVIEW

Since the present study lies at the intersection of prosody and pragmatic research, the first section of the literature review provides a theoretical basis for the study by reviewing the existing approaches to describing intonation, focusing mostly on the discourse intonation framework. From the pragmatic perspective, the section outlines speech act theory with an emphasis on requests, as this speech act is the primary concern of the present study. The chapter then discusses the issues related to learners' acquisition of prosody, particularly in the context of pragmatics, as well as the processing of prosody along with technology-based prosody instruction and the role of learners' individual differences in learners' acquisition of L2 prosody.

Defining Intonation vs. Prosody vs. Suprasegmentals

The terms *intonation*, *prosody*, and *suprasegmentals* have been used interchangeably in the domain of speech research since there is no precise agreement as to what each term represents (Pickering, 2018). Therefore, it is important to define these concepts in order to thoroughly understand the meaning behind them and establish their use for the present study.

Intonation has been traditionally described in relation to the melodic qualities of speech (Crystal, 1985). For example, Chun (2002) described intonation simply as a sentence melody that focuses on pitch variations and modulations. Taking a broader perspective, Pickering (2018) defined this concept as the “grammatical system that includes the use of pitch, pause, and prominence” (p. 3). Analogously, the broad definition of intonation corresponds to that of suprasegmentals, which are functional phonological units of a language that include pitch, stress, volume, and pause patterns, as contrasted with segmentals (Chun, 2002; Clark & Yallop, 1990; Pickering, 2018) or prosody, commonly defined as “the linguistic structure which determines the suprasegmental properties of utterances” (Cutler *et al.*, 1997, p. 142). Prosody is usually characterized by two main dimensions, namely, pitch and prominence. According to Pickering (2018), pitch is the perception of how high a voice is that is computed by the brain from the F0, or the acoustic measure of sound produced by humans (Chun, 2002), while prominence is the relative weight that is given to a syllable with regards to loudness, pitch, and length. Thus, one of the main functions of pitch is to mark prominent syllables in an utterance. It is important to note, however, that prosody is marked by features other than pitch and prominence in general. Kang (2010) identified suprasegmental features related to speech rate, pause, stress, and paratone (intonation sequences used to convey sentence structure information) measures that could be used to describe a speaker’s prosody.

Describing Prosody and Its Components

When it comes to describing English prosody and its components, several approaches have been developed. Some center around describing the overall shape of the pitch patterns (e.g., Halliday, 1967; Sweet, 1980) while others focus on the levels of pitch and tone sequences (e.g., Pierrehumbert & Hirschberg, 1990). More recent theories, however, have taken a more discourse-based perspective viewing intonation as a component of naturally occurring communication (e.g., Brazil, 1997; Couper-Kuhlen & Selting, 1996). The following sub-section hones in on Brazil's (1997) discourse intonation framework as the foundation for the present study.

Discourse Intonation Framework

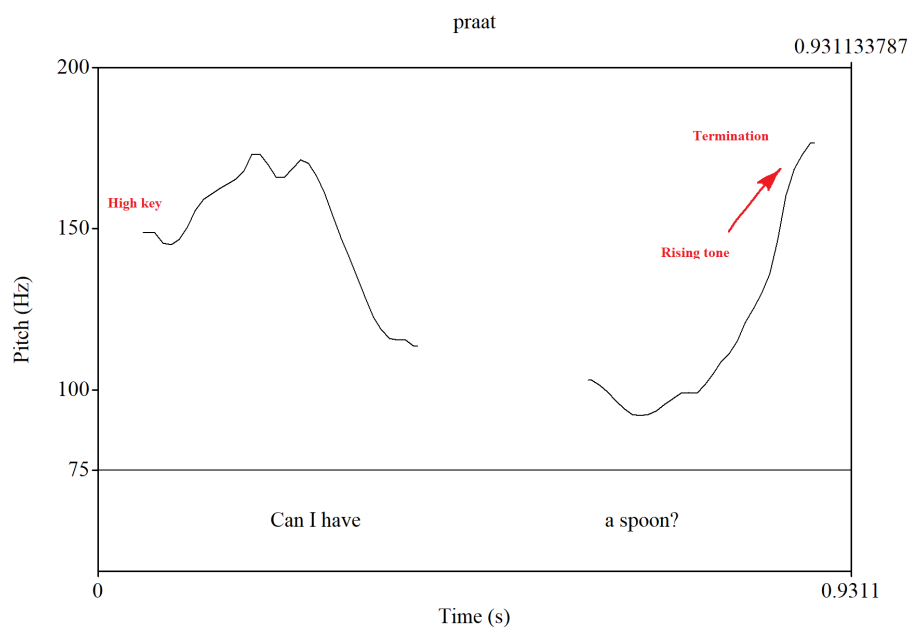
Brazil's (1976, 1980, 1997) discourse intonation framework serves as a basis for the phonological analysis in the current study. This framework was chosen for several theoretical and practical reasons. First, as Pickering (2018) notes, Brazil's model has been successfully adopted in a number of studies that focused on L2 speech production in ESL and EFL contexts (e.g., Hewings, 1995; Koester, 1990; Kang *et al.*, 2021). Additionally, since the framework was developed specifically for classroom discourse (see Brazil *et al.*, 1980), it has been applied in instructional contexts, especially due to its teachability (e.g., Pickering 2001; Levis & Pickering, 2004; Taguchi *et al.*, 2021). Finally, the model was built using naturally occurring L1 English interaction that L2 learners need to be familiarized with at the end of instruction, rather than linguist-generated utterances, as in the case of Pierrehumbert and Hirschberg (1990).

Before taking a closer look at the model itself, it is important to first define *discourse*, as it is an integral component of Brazil's framework. As defined by Crystal (1985), discourse is

simply a “continuous stretch of (especially spoken) language larger than a sentence” (p. 96). However, Crystal also emphasized that this stretch of language usually comprises a *speech event* that has a communicative purpose. In this regard, Brazil (1975, 1987) was concerned with the role intonation played in discourse and what functions it fulfils in communicative situations. He postulated that intonation patterns have different communicative values. For example, a speaker may choose to add prominence to a particular syllable thus amplifying a particular meaning. Thus, since each prosodic configuration has its own meaning in Brazil’s model, intonation gains an *interactional* significance in discourse. In other words, Brazil argues that intonational choices that speakers make are guided by their decisions to add meaning to particular words or groups of words in a specific communicative situation.

Figure 1

Example of a Tone Unit in Brazil’s (1997) Model



A tone unit, or “a stretch of language that carries the systematically opposed features of intonation” (Brazil, 1997, p. 5), is the central concept in the discourse intonation framework. In discourse, the speaker is the one who plans a tone unit to be decoded by the hearer. According to

the model, each tone unit in English has one or two syllables that are more emphasized than others. The first prominent syllable in a tone unit is the onset syllable that precedes the tonic syllable, or the last prominent syllable in a tone unit. While the onset syllable does not contain a pitch movement choice, it is the place where the speaker sets the pitch level (or *key*) of the tone unit. The tonic syllable, in its turn, is where the speaker indicates pitch movement, or tone, for the tone unit. The tone choice associated with the tonic syllable in a tone unit is called *termination*. Figure 1 presents an example utterance with a marked mid-key and falling tone on the termination. The x-axis in the figure plots the duration of the speech file in seconds, and the y-axis plots pitch height in Hertz (Hz). The black line within the plot is the mapped pitch movement in the sound file. Pickering (2018), in explaining Brazil's approach, noted that the creation of tone units is motivated not just by the speaker's understanding of the lexico-grammatical structures in a language, like clauses, items on a list, etc., but also by the perception of components of discourse like ideas pieces of information, or attitudes.

The discourse intonation framework distinguishes between five tone choices in English, each of them having a specific function in a discourse. Rising and fall-rising intonation patterns are also called referring tones and are used when information is shared by both interlocutors. Falling or rise-falling intonation patterns are referred to as proclaiming tones, which are usually used when new information is introduced. As Pickering (2018) explains, a sentence *Do you feel anxious?* pronounced with a falling intonation suggests that the doctor is asking for a new piece of information while rising intonation would indicate a doctor's request to confirm an assumption that was formed based on the other responses provided by a patient (p. 50). Finally, level intonation usually conveys a temporary withdrawal from referring or proclaiming when the speaker withdraws from the interaction. Expanding on Brazil's definition of the functions of the

level tone, Pickering (2018) added that this tone is often used with generic language routines in a conversation. Besides choosing a tone, the speaker is also tasked with selecting the key of the tone unit. Brazil (1997) identified three keys: high, mid, and low. Thus, the high key implies that there is more to follow while the low key indicates a connection between what is being said and some previous information.

The difference between the functions of the tones is connected to the notion of *common ground*, or the shared world views or information of the interlocutors. When the speaker utters a tone unit, they take into consideration the amount of common ground shared between them and the listener in structuring the tone unit prosodically. The listener then applies their common ground knowledge to interpret the intonation of the tone unit they heard. Pickering (2018) also explains that the main principle guiding tone choices in discourse is the *assumptions* the speaker makes about the information that the listener possesses in the interaction. Applying the concept of common ground on to the interpretation of tone choices, the referring tone is used when the listener is already sharing common ground with the speaker. In contrast, when using a proclaiming tone, the speaker hopes to expand the common ground. Apart from the functionality of tone choices in conveying new vs. known information, intonation also bears social significance in discourse. For example, rising tones are usually employed to convey a more intimate relationship between the speaker and the listener.

Discourse Functions of Intonation in English

The previous section touched upon the functions of intonation, as they are reflected in Brazil's (1997) framework. In sum, the discourse intonation framework posits that English speakers produce various intonation patterns to achieve communicative goals by placing prominence onto selected syllables or using rising, falling, or level tones (see also Kreidler, 1989). While Brazil

did not necessarily provide a comprehensive list of the intonation functions in his work, Chun (2002) summarized the prosodic functions reflected in discourse-based models that recognized the multi-functional role of intonation. She identified four major functions of intonation: grammatical, attitudinal, indexical, discourse, and interactional. The two latter functions on this list are of particular interest for the present study since they accentuate the systematic use of prosodic cues not only to regulate turn-taking in a conversation but also to contribute to relationship-building between the interlocutors and to communicate their status differences and social distance (see also Pickering, 2018).

In relation to the functions of intonation, Pickering (2018), in explaining Brazil's model, suggested that the intonation system is essentially *pragmatic* in nature. That is, languages use intonation structures to signal pragmatic messages. The pragmatic functions of intonation have also been previously noted in the works of Sag and Liberman (1975) as well as Brown *et al.* (1980). Sag and Liberman pointed out that intonation cues may be used to make an inquiry, a suggestion, a complaint, or an indirect request. Brown *et al.* (1980), in discussing the illocutionary force of intonation, stated that prosody exists in a close relationship with the illocutionary force of an utterance and the function it fulfils in a speech act. Taking into consideration the pragmatic functions of intonation, Pickering (2018) enhanced Brazil's description of the functions of tone choices. Thus, falling tones are considered reflective of the speaker's knowledge of information of interest in a conversation while rising tones reflect the speaker's attempt to build solidarity with the listener and to include them in the conversation. Level tones, besides their function of withdrawing the speaker from negotiation, are utilized to indicate lack of focus on the hearer or on the communicative message as well as the speaker's neutrality.

The choice of tones in a conversation can also be affected by the dominant/non-dominant relationship between the interlocutors (Kang & Kermad, 2019). In Brazil's framework, dominant speakers usually have a wider variety of tone choices while non-dominant speakers do not have such a range of options. The influence of the power-relationship dynamic on the tonal composition of speech has been observed by Cheng, Greaves, and Warren (2008) who found that while in daily conversations among speakers of equal power the use of falling and rising tones is split in half, professors in supervising positions use more rising tones. Similarly, Kang and Kermad (2019) pointed out that supervisors in business meetings use more rising tones, as well.

Due to the intuitive nature of prosody for L1 speakers, the interactional and pragmatic functions of intonation may present a particular difficulty for L2 learners. Pickering (2018) noted that while native or competent users of a language are often unable to reflect on or explain the underlying conventions of intonation use in a given language, they usually assume that the other speaker/hearer is being as cooperative as possible and is following those norms and conventions. Nevertheless, when those conventions are not followed, it is often interpreted as a sign of a speaker's attitude, bias, or lack of awareness of the pragmatic norms in a given communicative situation. Moreover, L2 learners are oftentimes unaware of the prosodic intricacies in a target language since the L2 phonological system is not internalized in the same way as the L1 system (Jimenez-Vilches, 2017). For example, Gumperz (1982) found that unexpected intonation patterns used by Indian and Pakistani speakers of English who worked at a cafeteria were perceived as uncooperative by L1 British English speakers. However, discussing the conventions of intonation use with the L2 speakers led to improvements in communication between these two groups. The following sections of the literature review will further illustrate the interplay between L2 prosody and pragmatics as well as the role of instruction in this relationship.

Requests in the Speech Act framework

The current study lies at the intersection of intonational and pragmatic research; therefore, apart from reviewing the existing theories of intonation, it is crucial to frame the study within the domain of L2 pragmatics. This section will thus provide an overview of speech act theory and focus particularly on requests, the speech act under investigation.

Speech act theory has been established as one of the most widely used theories in the field of pragmatics (Martínez-Flor & Usó-Juan, 2010; Taguchi & Roever, 2017). According to the theory, language is used not just to say things, but to achieve communicative goals (Austin, 1962; Tanck, 2002). To accomplish these goals, speakers employ a variety of speech acts, or utterances that use language to achieve real-world effects (Taguchi & Roever, 2017, p. 34). Building on Austin's (1962) theory, Searle (1976) set up the following classification of speech acts: (1) assertives, (2) directives, (3) commissives, (4) expressives and (5) declaratives (see Taguchi & Roever, 2017 for an overview). Surely, many other classifications of speech acts exist, but the important distinction they all include is between direct and indirect speech acts. It is based on whether the intention of the speaker coincides with what they actually say or whether they "[communicate] to the hearer more than he actually says by way of relying on their mutually shared background information, both linguistic and nonlinguistic, together with the general powers of rationality and inference on the part of the hearer" (Searle, 1976, p. 60-61). In this regard, a speech act consists of three components: the locution, or the actual uttered words; the illocution, or the intended real-world force of the utterance; and the perlocution, or the real-world effect of the utterance (Austin, 1962; Taguchi & Roever, 2017). From the perspective of Brown and Levinson's (1987) politeness theory, speech acts can be described in terms of the three factors that affect the speakers' level of politeness and their choice of politeness strategies

in performing a speech act; namely, relative power (P), social distance (D), and degree of imposition (R) (see also Taguchi & Roever, 2017). Power refers to the hierarchical societal roles of the interlocutors. For instance, one speaker may have higher power than the other, like in an interaction between a professor and a student, or the interlocutors may have equal power (e.g., both of them are students). Social distance reflects the closeness of the interlocutors' relationships, or how well they know each other (Brown & Levinson, 1987). Thus, close friends have low social distance in an interaction while people who have met each other for the first time (e.g., new classmates) have high social distance. Finally, the ranking of imposition is the 'cost' to the hearer in performing what the speaker intends for them to do, such as the hearer's time, money, effort, etc. (Taguchi & Roever, 2017). Low imposition speech acts, for example, asking a sales assistant for cash back, do not require much effort from the hearer; on the other hand, asking an employer to give one a raise is considered a high imposition speech act since it requires a lot of resources from the hearer.

While many speech acts across all of the categories have received attention in the domains of interlanguage pragmatics, requests have been of particular interest for researchers. As noted by Usó-Juan (2010), these speech acts may occur in a wide range of situations and involve relationships that vary from lower, equal, or higher status interlocutors. The increased attention to requests may also be explained by the fact that these speech acts, especially high imposition ones, are highly face-threatening; thus, if a speaker uses a request inappropriately, they are at risk of being perceived as rude or demanding. Within the group of directives, requests are considered the most face-threatening because of their intrinsic threat to the negative face of the hearer because of the hearer's potential commitment to a future course of action (Brown & Levinson, 1987). In other words, as Trosborg (2011) explains, in requests, a speaker conveys to a hearer

that they want the requestee to perform an act which is for the benefit of the speaker. The hearer's role, in turn, is to either comply with the request or refuse it. Being highly face-threatening, requests may pose a difficulty not only to the L1 speakers of a language but also to the L2 speakers. For a request to be appropriate, L2 speakers need to possess considerable pragmatic expertise to control the politeness of a request. Specifically, it has been argued that performing a more indirect request (by including a softening modifier, for example) is often viewed as more polite since it shows consideration for an interlocutor's face needs.

Because of the face-threatening potential of requests, speakers must pay particular attention to several situational factors when formulating utterances. More specifically, when determining the degree of directness of a request or strategy, the speaker must consider the degree of imposition involved in the speech acts together with the social distance between the interlocutors and their relative power (Taguchi, 2006). In a more formal situation, such as a student making a request to a professor, the speech act would involve a higher degree of imposition and farther distance between the interlocutors. The professor would also be in a position of higher power in this situation. Together, these situational characteristics require the hearer's face to be protected, motivating the greater degree of indirectness in the student's request. This section introduced requests within the Speech Act framework; the following section will bring together the domains of L2 prosody and pragmatics focusing specifically on research at the intersection of speech acts and intonation.

Pragmatics and Prosody

The relationship between intonation patterns and discourse is fundamental to understanding and interpreting the pragmatic meaning in spoken English discourse (Pierrehumbert & Hirschberg, 1990; Brazil, 1997; Wichmann, 2000; Zhuang, 2015). As noted by Kang *et al.* (2021), prosody is

an important element in identifying a speaker's intention behind an utterance, thus contributing to its pragmatic meaning. Several other essential pragmatic functions of intonation identified by researchers include the communication of sociolinguistic information, such as power distance, perceptions of status differences, and expression of politeness (Brown & Levinson, 1987; Brazil, 1997). As a matter of fact, since intonational cues are especially susceptible to misinterpretation by native speakers, deviations in intonation patterns may negatively affect comprehension and provoke pragmatic consequences on the spoken discourse (Anderson-Hsieh *et al.*, 1992; Derwing & Munro, 1997; Pickering, 2009).

Despite the clear link between prosody and pragmatics, the research in this area has been limited with most of the studies investigating indirect meaning and some research focusing on the role of intonation in the expressions of politeness, speech acts, and discourse (Kang & Kermad, 2019). Nevertheless, this section presents a comprehensive overview of the current empirical findings on prosody in pragmatics highlighting the underlying relationships between these two domains as well as outlining the associations between the prosodic features and communicative situations. The section also hones in on the research that explored intonation within speech act theory, as this relationship is central to the current study.

Several recent studies have examined prosodic cues within different pragmatic constructs, one of them being indirect meaning (see Kang & Kermad, 2019 for an overview). In examining the role of vocal cues in listeners' comprehension of sarcasm, Rockwell (2000) collected speech samples of speakers reading sentences in three conditions: non-sarcasm, spontaneous sarcasm, and posed sarcasm (or sarcasm produced following directions). The sound files were then coded for several vocal features; namely, tempo, pitch (low/high, unvaried/varied), intensity (low/high, unvaried/varied), resonance, and articulation. The results

of the study showed that listeners were able to distinguish posed sarcasm from non-sarcasm, but not from spontaneous sarcasm. Analysis of the acoustic features further revealed that sarcasm was conveyed by a slower tempo, lower pitch level, and greater loudness than non-sarcasm. In a similar fashion, Bryant and Fox Tree (2005) measured the mean frequency, pitch range and variability, amplitude variability (changes in loudness), and mean syllable duration in spontaneous ironic and non-ironic speech collected from radio shows to see if these voice qualities were systematically associated with irony. The study participants also rated content-controlled versions of the same ironic and non-ironic utterances on a range of affective dimensions including the degree of sarcasm, anger, inquisitiveness, authority, and new/given information. The listeners revealed in the post-experiment questionnaire that they used the speed of and tone of the speech samples to rate the sarcasm and inquisitiveness of the interactions. The authors concluded that to communicate an ironic message, the speakers provided prosodic cues in contextually dependent ways that were applied differently onto the presented propositions.

Further investigating indirect meaning, Shively *et al.* (2008) examined the interpretation of ironic utterances in Spanish films by asking 55 L2 Spanish learners to assess the ironic tone of film segments. The results suggested that the recognition of irony improved with increases in learners' proficiency and experience with the target language and that participants, especially with a more advanced proficiency, were able to pick up multiple tones of irony (joking, criticizing, etc.). Taguchi *et al.* (2016) investigated the comprehension of indirect refusals, indirect opinions, and irony in 21 L1 and 32 intermediate/advanced L2 Spanish learners through a computer-delivered multimedia listening test. Participants' test responses were analyzed for accuracy and speed. The results showed that L2 Spanish learners attended to prosodic cues, such as intonation and exaggerated speech, in their processing of ironic spoken utterances. Moreover,

verbal interviews with the learners after the test provided evidence for the multimodal processing of irony that included learners using the tone of speakers' utterances in the comprehension of the sarcastic meaning. It is evident from the studies above that interlocutors use prosodic features to comprehend pragmatic meaning, and that this ability increases with the growth in learners' proficiency.

The intersection of prosody and pragmatics has also been investigated from the English-based cross-linguistic discourse perspective re-emphasizing the role of prosody in pragmatic meaning making. Pickering (2009) conducted an acoustic analysis of English as a Lingua Franca (ELF) information gap interactions to examine the role of intonation in this type of discourse. After analyzing the prominence, tone choices, and key choices, Pickering argued that both pitch movement choices and relative pitch height on prominent syllables were used by ELF interlocutors to signal communication breaks and negotiate their resolution regardless of their L1. Applying corpus linguistics methods, Romero-Trillo and Newell (2012) analyzed the use of discourse markers to provide feedback in cross-linguistic conversations. A detailed analysis of the two most frequent pragmatic markers, 'mhm' and 'yeah', in five L1-L2 interviews from the Louvain International Database of Spoken English Interlanguage (LINDSEI), showed that the larger length and higher final pitch of these markers allowed native speakers of English to appear more interactionally involved while the use of markers by non-native speakers gave more of a "transactional flavor" (p. 129).

Continuing his work with LINDSEI, Romero-Trillo (2019) applied corpus methods to analyze the use of feedback pragmatic markers and the associated intonation patterns (tone choices) in intercultural communication. In this study, apart from the interviews from LINDSEI, Romero-Trillo included L1-L1 interactions from the prosodically annotated London-Lund

corpus. The pragmatic markers in the interactions were analyzed for duration, initial, medium, and final pitch, pitch range and intensity. The results demonstrated that L1-L1 interaction, speakers mostly used falling tones, followed by absence of tonicity. Contrastingly, in L1-L2 interaction, native speakers of English, in addition to using a wider variety of feedback markers, used fall-rising and rising tones in interacting with L2 speakers. L2 speakers, in their turn, used rises and fall-rises, similarly to their L1 interactants demonstrating prosodic matching. Thus, additional evidence from corpus-based pragma-prosodic studies (Romero-Trillo, 2019; Romero-Trillo & Newell, 2012) further suggests that prosodic features, especially tone choices and prominence, are used differently by L1 and L2 English speakers, which may lead to differences in interpreting pragmatic meaning.

Some empirical research of L2 intonation and pragmatics (e.g., Kang *et al.*, 2010; Pickering, 2001; Pickering, 2009) has demonstrated that applying tone choices is challenging for learners across proficiency levels. It is often the case that non-native intonation displays differences in pitch range and tone choice patterns compared to native speech, causing misunderstandings between native and non-native interlocutors (Baker, 2010). For example, Hewings (1995) found that L2 English learners from Korea, Greece, and Indonesia overused falling tones in scripted dialogues whereas native speakers chose rising tones to indicate that they were ready for social integration. Hewings suggests that non-native speakers' use of falling tones may be interpreted as an indication of animosity or rudeness by native speakers. In a related study, Pickering (2001) described and analyzed the tone choices made by North American and Chinese International Teaching Assistants (ITAs) to evaluate teaching success in the classroom. The prosody in twelve teaching presentations given by L1 and L2 English TAs was analyzed using instrumental and auditory analysis informed by Brazil's (1997) discourse intonation

framework. Pickering discovered that native English-speaking teaching assistants systematically exploited intonation to increase the accessibility of the lectures and establish rapport with the students. In contrast, ITAs often used more level tones and fewer rising tones that created a monotonous pitch structure unfamiliar to L1 English hearers. Moreover, using sequences of falling and level tones in the presentations regardless of the informational value of the matter muddled the information structure and made ITAs sound ‘unsympathetic’ and ‘uninvolved’ (p. 233). In other words, ITAs did not use tone choices in a pragmatically effective way increasing social distance between them and the students.

Furthermore, studies have demonstrated that suprasegmental features of L2 speech, including tone choices, pitch movement, and pitch range, can explain a substantial part of listeners’ judgement on the speakers’ overall oral performance (e.g., Kang *et al.*, 2010; Pickering, 2001). Not only that, but in relation to pragmatics, Kang *et al.* (2010) emphasized that ineffective prosodic structure of L2 speech may contribute to listeners’ confusion and misinterpretation of speaker intent. Supporting the conclusions of Pickering (2001) and Wennerstrom (1994), Kang *et al.* established that underuse of rising tones in L2 English speech may lead to the speakers coming off as arrogant and overassertive (see also Gumperz, 1982) as well as oriented towards language rather than the listeners.

Turning to the limited research on prosody and (im)politeness, Culpeper (2005) described the interplay between prosody, particularly pitch movement, and speakers’ ways of expressing impoliteness in a TV show ‘The Weakest Link’. In analyzing the intonation of formulaic catchphrases in the show, Culpeper argued that prosody was not just a contextual cue but rather a central aspect in communicating offense in discourse (p. 68). Later on, Culpeper (2011) conducted an analogous qualitative instrumental analysis of two speech segments of judges on

‘Pop Idol’ revealing that polite comments are associated with the use of parallel prosodic structures, a faster speech rate, and marked pitch movement. On the other hand, monotonous speech patterns as well as slower speech rate and excessive pausing signaled boredom and impoliteness, similarly to the findings of Kang *et al.* (2010) and Pickering (2001).

Research at the interface of prosody and pragmatics has demonstrated that prosody is an important factor in pragmatic meaning making as well as in listeners’ pragmatic judgements about their interlocutors. Moreover, current findings in the domain suggest that the use of intonation in pragmatic contexts is not intuitive for L2 learners. Conjointly, these conclusions indicate the need for pragma-prosodic instructional training for L2 learners, which is the focus of this dissertation.

Prosody and Speech Acts

It has been widely recognized both in L1 and L2 research that intonation plays an important role in conveying and understanding pragmatic meaning. Intonation features have pragmatic functions that communicate sociolinguistic information, signal speakers’ feelings or emotions, and establish rapport between interlocutors (Brazil, 1997; Pickering, 2018). Moreover, as Wichmann (2015) and Kang *et al.* (2021) sharply note, suprasegmental intonational features, such as word stress, pitch, pauses, and speech rate, enable speakers to express their intention, politeness, and directness in speech acts that cannot be communicated through grammar or vocabulary. Nevertheless, empirical research on L2 learners’ use of intonation to express the illocutionary force of a speech act has been scarce.

Within the handful of such studies, Hewings (1995) examined the prosodic structure of disagreements detecting the use of falling tones in L2 learners’ speech in contrast to L1 speakers,

who employed more rising or level tones. Hewings explained this distribution of pitch movement by their different functions in discourse: While L2 speakers used falling intonation for mitigation purposes, L1 speakers used them for social integration with the interlocutors.

Pickering *et al.* (2012) investigated pitch concord, or matching of pitch level in adjacent utterances, of (dis)agreements in L1 and L2 English. In the study, pairs of North American English speakers and Chinese learners of English discussed concept cars to choose their favorite one. Pickering used Brazil's (1997) model to find evidence of pitch concord through the acoustic analysis that concentrated on the identification of pitch level choices in adjacent utterances by each speaker in a pair. The study found that in most of the cases, both L1 and L2 speakers displayed pitch concord in agreement sequences. In the cases with no pitch concord, other types of prosodic matching were detected (e.g., matching pitch contours), although they were not systematic. In the disagreements, L1 speakers used significantly lower or higher pitch choices demonstrating discord. However, this was not the case for L2 speakers, leading the author to conclude that an instructional training could be beneficial for improving L2 prosody in disagreements.

In another speech act-focused study, Staples (2015) analyzed expressions of empathy produced by internationally and U.S. trained nurses. Regarding the prosodic composition of the speech acts, speech rate, number and length of silent pauses, pitch range, tone choices, prominence, and paratone were compared between the two groups of nurses. Staples discovered that both groups differed in their tone choices. In particular, L1 English nurses employed significantly more falling tones to express their sympathy towards the patient. L2 English nurses, in contrast, used significantly more level tones that were interpreted as lack of compassion or interest in the patient (p. 142). Additionally, international nurses did not have a clear preference

for a particular tone choice suggesting that they were not as comfortable with the speech act and not completely prepared to express sympathy in a way that was anticipated by U.S. patients.

Prosody in Requests. Within the academic context, Kang *et al.* (2021) examined the effect of study abroad experience and proficiency on the spoken production of high and low imposition requests and opinions made by L2 English learners towards a professor. Using oral DCTs to elicit participants' responses, the researchers were able to assess sentence prominence, pitch range, and intonation in the speech samples. The findings revealed a distinct impact of proficiency and study abroad experience on participants' use of intonation, pitch, and prominence. Specifically, tone choices of higher proficiency learners with study abroad experience resembled NSs' production more closely. The authors also found that the degree of imposition was associated with different prosodic patterns used by the learners. Thus, in high-imposition situations, low proficiency learners used significantly more rising tones. The authors suggested that this tone choice may seem inappropriate in high-stakes situations as the use of rising pitch movement is often associated with the speaker's dominance (see also Cheng *et al.*, 2008). Considering the referring functions of rising tone (Brazil, 1997; Pickering, 2018), its overuse is inappropriate for high-imposition speech acts with a high power-distance relationship (professor vs. student) since the focus of such requests is not likely to be on establishing common ground. Furthermore, lower level L2 learners under-used level intonation compared to high proficiency and L2 speakers implying that they were unable to hold the floor for extended amounts of discourse to sufficiently provide justification for their requests or opinions. In terms of prominence, the study showed that L2 learners, particularly the low proficiency group, overused this suprasegmental feature stressing too many words in a tone unit, thus increasing the risk of miscommunication and reduced intelligibility. This phenomenon was observed in

previous studies that focused on L2 learners' use of prominence (e.g., Hahn, 2004; Wennerstrom, 1994).

Building on Kang *et al.*, Taguchi *et al.* (2021) investigated the longitudinal development of prosodic properties (tone choices, prominence, and pitch range) in L2 learners' requests and opinions. The participants completed speaking tasks that elicited the speech acts of interest three times during an 8-month period (an academic year). The authors found further evidence of development in tone choices with learners employing more level tones and fewer rising tones in requests and opinions by the end of the immersion. Prominence and pitch range were found to change overtime as well, approaching the NS baseline, but at earlier stages of the immersion. The study also reported that learners' proficiency and amount of target language contact significantly affected the changes in tone choices.

In a corpus-based study, Ramírez-Verdugo (2008) compared tone choices and pitch accent in requests and commands produced by L1 English and Spanish L2 English teachers in scripted read-aloud dialogues and naturally occurring classroom discourse. Similar to Kang *et al.* (2021), the results revealed that intonation patterns differed in the two groups. While the L2 English teachers' speech acts were realized mostly with falling tones, L1 teachers employed a wider variety of tones, including low rises, to project themselves as less dominant and orient the students towards collaboration. The author suggested that the observed distinction in tone choices could be explained by differences in student-teacher status relationships and interpersonal discourse management in L1 and L2 contexts. Although classroom discourse is a typical example of asymmetric, hierarchical distribution of power and distance with the teacher controlling the classroom, the power distance was mitigated at a prosodic level by rising intonation in L1 classroom. As Ramírez-Verdugo noted, rising intonation can be applied by

teachers talking to students to give them confidence and reassurance (see also Thompson, 1990). In contrast, L2 teachers in the study did not demonstrate the mastery of intonation and its different speech act functions. That is, L2 instructors' intonation did not correspond to the intended purpose as they realized the requests and commands with mostly falling tone patterns that transmitted the pragmatic meaning of commanding or ordering rather than encouraging.

The overview of the existing research in the interface of L2 prosody and pragmatics above has exemplified several gaps that are yet to be filled. First, while the studies on certain constructs of L2 pragmatics (e.g., feedback markers, attitudinal meanings) and their intonation composition are relatively common, little is known about the link between L2 prosody and requests. Although several studies have described the prosodic patterns and their pragmatic meaning in select speech acts including requests, more research is needed to construct a more comprehensive representation of the functions of prosody in L2 request production.

In terms of the specific prosodic features and their weight in pragmatic meaning making, research seems to highlight tone choices and prominence, that are pervasive in the reviewed studies in their role in discourse. Tone choices, when used inappropriately in discourse overall and in speech acts, in particular, may contribute to misinterpretation of a speaker's intention and attitude (Baker, 2010; Culpeper, 2011; Kang *et al.*, 2010; Pickering, 2001, 2009). Similarly, excessive and misplaced prominence can cause listeners' confusion and ineffectiveness in conveying a pragmatic message (Kang *et al.*, 2010; Staples, 2015). It is due to these reasons that tone choices and prominence were chosen as the focal prosodic features for the developed instructional training.

Pragma-Prosodic Appropriateness in Speech Acts

One avenue of research in the pragma-prosodic composition of speech has engaged in describing speech act production from the pragmatic or prosodic perspective. However, studies have also been concerned with interlocutors' *perception* of speech acts and their appropriateness. The construct of appropriateness has largely been applied in measuring interlocutors' perceptions of learners' pragmatic competence, especially within speech act frameworks. Dewaele (2008) noted that appropriateness is a "slippery" concept that is hard to define (p. 246). Nevertheless, within the meaning-oriented model, Purpura (2004) identified appropriateness as a latent constituent of pragmatic meaning, specifically, its sociolinguistic and sociocultural components, that incorporates the notions of nativeness and politeness, as well as cultural norms, preferences, and expectations. In its common sense, appropriateness has been defined as the product of evaluation that is based on "the nature of the connectedness between co-participants, communicative actions and their linguistic realizations in sociocultural contexts" (Fetzer, 2004).

While empirical pragmatic research has largely accepted this broad characterization of appropriateness, its specific operationalizations have varied. Some have included the issues of grammaticality and discourse organization in describing the components of appropriateness (e.g., Al Masaeed *et al.* 2020; Cunningham, 2017; Li *et al.*, 2019; Taguchi, 2006, 2013), while others focused on the evaluation of learners' awareness of the sociolinguistic variables in the context of speech act production. For example, Hudson *et al.* (1992, 1995) and later Jianda (2006) defined appropriateness as learners' ability to use the correct speech act with a proper degree of formality, directness, and politeness. Taguchi (2011a) applied a similar operationalization of appropriateness in her study where four L1 English speakers with varied cultural backgrounds rated requests and refusals performed by ELF learners on a five-point scale. Critically, while

studies have mostly considered the appropriateness of linguistic forms in relation to these constructs, Taguchi (2006) also highlighted the importance of features other than linguistic expressions that can contribute to the overall appropriateness of a speech act, such as speakers' suprasegmental speech characteristics.

In addition to the theoretically informed approaches to defining appropriateness, studies have attempted to examine rater criteria that L1 and L2 English interlocutors considered when rating the pragmatic production of learners' speech acts. Ishihara (2013) noted that since learners' interlocutors are the real language evaluators, they may operate with specific criteria as to what constitutes pragmatically appropriate production. Thus, Alemi *et al.* (2014), in investigating compliments, elicited several macro criteria that included politeness, interlocutors' characteristics and relationships, tactfulness of production, sincerity, socio-pragmatic and linguistic appropriateness. The criterion of politeness also appeared in parallel studies by Alemi (2012) as well as Tajeddin and Alemi (2014) that investigated refusals and apologies. Later studies from the same series (Alemi & Khanlarzadeh, 2016 and Alemi & Motamedi, 2019) revealed additional appropriateness criteria, that aligned with previously established constructs, such as situationally dependent formality, directness/indirectness, and authenticity.

Pragma-Prosodic Appropriateness of Requests. Looking at requests in particular, Ishihara (2013), supporting the series of studies by Alemi and colleagues, also identified overall directness, politeness, and formality as the assessment criteria that instructors used to interpret learners' production. Other criteria included learners' choice of request strategies and pragmatic tone, or the affect expressed by linguistic and non-linguistic means. Winans (2020) also named politeness as one of the appropriateness criteria in an investigation of syntactic modifiers in requests.

From the point of view of prosody, connections have been drawn between intonation and appropriateness, although studies in this domain are rare. Sydorenko *et al.* (2014) investigated appropriateness criteria that were salient to untrained raters who scored extended request sequences. Raters used a 6-point scale to broadly mark requests appropriate or inappropriate and explained their choice. The results showed that raters attended to overall or expression-specific intonation in assigning appropriateness ratings. Within the sub-constructs of appropriateness, Brazil (1997) and Pickering (2018) related intonation to the perceptions of a speaker's degree of politeness. For example, Pickering (2018) explained that falling intonation in suggestions may be perceived as less polite by interlocutors in higher power (p. 24). Additionally, Herrero and Devís (2020) investigated the native speakers' perception of requests produced by L1 Mandarin learners of Spanish. The study analyzed the intonational patterns in requests that were produced by the L2 Spanish learners with a polite intention but were perceived as impolite by L1 Spanish speakers. The results of a perceptual survey, in which L1 Spanish speakers rated the degree of politeness of the requests produced by the L2 learners, indicated that those requests that contained falling tones as well as tonal emphasis on stressed syllables were perceived as impolite by L1 Spanish speakers. Finally, another construct, *sincerity*, has been associated with pragma-prosodic appropriateness. Brazil (1997), Pickering (2018), and other researchers (e.g., Staples, 2015) related falling intonation contours to speakers' sincerity in discourse while rising intonation used inappropriately is often interpreted as insincere. Thus, Aijmer (1996), in an investigation of apologies in the London-Lund corpus, found that rising or fall-rising tones were employed in conjunction with short formulaic apologies like 'sorry' whereas more serious, weighty apologies included falling intonation. Knowles (2016) found a similar pattern in

apologies and condolences with rising intonation sounding more routine-like and falling intonation being associated with sincerity and remorse.

Generally speaking, there have been some attempts to interpret the pragmatic meaning of prosody in speech acts from the listener perspective. Nevertheless, the involvement of the listener in interpreting the meaning of L2 intonation in speech acts, especially regarding appropriateness, has been somewhat neglected. Much is still unknown about the meaning of prosody in requests, especially as perceived by interlocutors. In addition, studies have been inconsistent in operationalizing and measuring the appropriateness of speech acts. Research is needed to provide a deeper understanding of the meaning behind prosody in speech acts like requests, particularly as it is interpreted by the recipients of requests themselves.

Accentedness and Comprehensibility in Pragma-Prosodic Perception. Apart from the appropriateness construct, two other listener variables, accentedness and comprehensibility, has been linked to pragma-prosodic perception. In their plenary speech, Derwing and Munro (2009) emphasized the need for research in pragmatic factors and L2 accent as it often a “scapegoat” that listeners use to disguise their negative reactions to pragmatic behavior that is unfamiliar to them (p. 488). This issue is strongly related to the problem of linguistic stereotyping where judgements about a speaker are made based solely on their accent (Lambert *et al.*, 1960) and perceived comprehensibility (e.g., Coetzee-Van Rooy & Van Rooy, 2005).

While studies have not investigated the relationship between accentedness, comprehensibility, and L2 speakers’ pragma-prosodic appropriateness directly, some have drawn a parallel between L2 accentedness and comprehensibility and listeners’ pragmatics-related judgements about speakers. In Derwing and Munro (2009), Canadian-born employees took a preference task that presented speech samples according to the degree of L2 accent. The results

revealed accentedness being an important factor in listeners' preferences for potential interlocutors in the workplace with less accented speech being more preferred. Similar results were observed in Kang and Yaw (2021) who found that listeners with negative stereotypes against L2 accents view speakers as less socially attractive. Derwing *et al.* (2004) also found associations between raters' judgements about the "goodness of prosody" and accentedness. Finally, Crowther *et al.* (2018) detected strong correlations between accentedness and intonation appropriateness judgements ($r =$ between .72 and .82).

In a similar vein, the impact of speakers' comprehensibility has been explored with regard to their acceptability for certain occupational contexts (e.g., Kang *et al.*, 2023). Studies have also explored reverse linguistic stereotyping and its reflection in comprehension perceptions. For example, McDonough *et al.* (2022) found links between speakers' ethnicity, comprehensibility ratings, status, and solidarity. A few instructional studies also found that comprehensibility perceptions in speech acts can improve after targeted pragmatic instruction (Derwing *et al.*, 2021) and after study abroad (Iwasaki, 2008).

Within the scope of this study, accentedness and comprehensibility were measured to control for the possible confounding effect of this variable in listeners' perceptions of the pragmatic appropriateness of L2 speakers.

Pragma-Prosodic Needs of Chinese Learners of English

The present study has its focus on one particular group of L2 English learners, namely, EFL learners with Mandarin Chinese background. The choice of this particular population for this dissertation study had several motives. For one, EFL learners from China comprised the largest international student population in the United States in the 2021/2022 academic year (Duffin, 2023). Since the designed training is contextualized specifically in university contexts, it may be

especially useful for this population. Chinese learners of English were also the only group accessible to the researcher during the COVID-19 pandemic. Finally, and most importantly, differences in the two prosodic systems (Mandarin Chinese and English) emphasize the need for targeted pragma-prosodic instruction. It is commonly known that intonation in contrast to English, Mandarin is a tonal language, where speakers' tone choices generally help to distinguish lexical meaning (Xu, 2015). Besides lexical-level functions of intonation, there is some evidence that Mandarin is characterized by a discourse-level intonation system that distinguishes types of sentences and helps to express emotions (Shen, 1990). However, it appears that the lexical function of intonation oftentimes dominates discourse-level prosodic structures as question and statement contours are frequently affected by lexical tones in Mandarin (Shih, 1988; Yuan, 2011). In other words, tones in Mandarin are generally understood in the same way, independently of speakers' intentions (Madella, 2020). Studies have observed that Chinese speakers of English may not be able to identify statements and questions based solely on intonation cues equally well for sentences with different lexical tones (Yuan, 2011). In its turn, as described above, English intonation is independent of lexical meaning and thus is more readily available for conveying discourse pragmatic meaning. Moreover, in cases when intonation is indeed used for discourse purposes in Mandarin, it is usually marked at the *initial* phrase boundary. In English, however, it is the shape of the boundary *at the end* of a phrase that communicates discourse meaning (Shen, 1990; Wennerstrom, 1998). Regarding the use of prominence in Mandarin, the few studies that exist have shown that it is largely rhythm-based and guided by duration rather than pitch height (Wennerstrom, 1998).

It appears that these differences in suprasegmental features in English and Mandarin may affect Chinese L2 English learners' use of prosody in discourse. In fact, recent studies in

pragma-prosodic interface in Chinese learners of English found that the prosodic structure in less proficient learners' speech acts is indeed different from L1 and more proficient L2 learners of English, especially in high-imposition situations talking to interlocutors in higher power (e.g., Kang *et al.*, 2021; Kostromitina, in press; Taguchi *et al.*, 2021). The discrepancies in the prosodic use in Chinese and English requests suggest that Chinese learners of English may experience difficulties in acquiring English intonation caused by the L1 intonation transfer (e.g., Graham & Post, 2018; Lu & Miran, 2016). Instruction that exposes Chinese L2 English learners to pragmatic functions of prosody and that is informed by the form-meaning connections between English intonation contours and their pragmatic meanings may assist in learners' pragma-prosodic acquisition.

Theory and Practice of L2 Prosody Instruction

While prosody is central in managing interaction (Brazil, 1997), prosodic features, especially tone choices are not intuitive for L2 learners (Levis, 1999). To convey pragmatic meaning accurately and appropriately through L2 intonation, learners need to not only know the linguistic systems of a language but also know how prosody is mapped onto those systems (Chun, 1998; Kang & Kermad, 2019). Because of the complex, unintuitive nature of the L2 pragma-prosodic interface, instructional trainings play a particularly prominent role in the acquisition of L2 prosody overall, and in relation to L2 pragmatics. This section reviews the current theories of L2 prosody acquisition, concentrating on the L2 Intonation Learning Theory as the underlying framework for the study. Subsequently, a comprehensive overview of studies in L2 prosody instruction is given, with the pragma-prosodic instructional studies as the chief focus of the section.

L2 Prosody Acquisition Theories

Intonation has been found particularly challenging for second language acquisition (SLA) researchers to model (Graham & Post, 2018; Li & Post, 2014). Furthermore, it has been argued that acquiring L2 intonation is the last stage of SLA due to the semantic, discourse, and pragmatic functions of intonation and its close connection with other aspects of language (Son, 2018). Notwithstanding these difficulties, there have been attempts in SLA research to model the processes behind L2 intonation acquisition.

Some have argued that learners model their L2 intonation patterns based on their L1 intonation. As a child acquires intonation in the first language, intonation becomes automatic and intrinsic; that is, competent L1 speakers become unaware of it and pay attention to other linguistic cues in discourse instead (Snow, 1995). Thus, it is often the case that L2 learners use their L1 intonation as the unconscious resource to apply it in L2 spoken discourse (Anderson-Hsieh, 1992). L2 intonation develops on the basis of L1 intonation, which serves as a mechanism of perceiving and conveying basic speech functions in the target language (e.g., Brown & Levinson, 1987). Furthermore, current research suggests that L2 prosodic features develop differently. While the phonetic aspects of speech generally follow universal developmental paths progressing in the direction of an L2 with an increase in learners' proficiency irrespective of L1 (e.g., Archibald, 1994 for lexical stress; Rasier & Hiligsmann, 2007 for accent distribution), intonational features like pitch range and tone choices often show L1 transfer effects at various proficiency levels (Graham & Post, 2018).

Previous models of L2 phonological acquisition, such as the Speech Learning Model (Flege, 1995) or the Perceptual Assimilation Model (Best, 1994) have largely focused on segmentals ignoring the prosodic aspect of speech. Moreover, as Li & Post (2014) note, these

models are unfit to relate to each other, since they were developed to deal with different L2 speech phenomena. In an effort to unify the theoretical framework for the phonological and phonetic components of L2 intonation development, Mennen (2015) introduced L2 Intonation Learning Theory (LILt) by expanding Ladd's theory of prosodic typology (Ladd, 2001). Some have argued that this theory is currently the only theory that effectively explains the acquisition of L2 intonation (e.g., Son, 2018). The main aim of this working model is to account for the difficulties L2 learners encounter when developing L2 intonation. Thus, the model is composed of four intonation dimensions including the *systemic dimension*, or the inventory of categorical phonological elements, the *realization dimension*, or the implementation of these elements, the *semantic dimension*, or the functionality of these elements, and the *frequency dimension*, reflecting the regularity of use of these elements. Each of these dimensions is rooted in the existing evidence of the occurrence of L2 intonation divergences to predict where the L2 deviation is likely to occur.

One of the underlying assumptions of LILt is the link between prosodic perception and production. Mennen (2015) explains that prosodic meaning may be more accessible to learners through perceptual judgments of intonation categories, and that learners often experience difficulties in the production of L2 intonation due to the intrinsic complexity of intonation dimensions (see Chen, 2020). This perception-production link suggests that the acquisition of L2 prosodic production is contingent on learners' perception-based awareness of the meaning of intonation. In other words, the acquisition of prosody is tied to not only the learners' ability to discriminate the L2 intonation dimensions on a perceptual level, but also to their knowledge of the relevant semantic and pragmatic knowledge of intonation in a target language. For example, He *et al.* (2012) examined the ability of Chinese learners of Dutch to match one of the four

intonational contours to sentences. Hu found that the learners experienced difficulty in choosing the optimal contour for the sentences, seemingly due to the lack of semantic and pragmatic knowledge rather than their inability to perceive acoustic cues. Importantly, this trend was also noticed in learners of other languages. Puga *et al.* (2017) replicated He's study but targeted German learners of English. While Puga hypothesized that German learners would perform better on this task since German is typologically closer to English, certain intonation patterns were still challenging for the learners. Díez (2012) observed a similar trend with Spanish learners of English. This empirical evidence implies universality of difficulties with L2 intonation perception and, therefore, production, independent of learners' L1 background (Grosser *et al.*, 1997).

The LILt model of L2 phonological acquisition served as an underlying learning theory in the development of the instructional training in the study. The semantic dimension in the model was of particular importance for the training as it emphasizes that learners acquire L2 intonation together with its functions in discourse. Moreover, the model links perception and production as the essential steps in L2 intonation acquisition with gains in perception being the condition for improvements in production.

Considering this relationship, the instructional training in the study will include both perception and production activities. Importantly, learners can only move to the production part of the training once they master the perceptual part. To ensure this, a 90% correctness threshold will be established in the perception activities. In other words, participants will be presented with the production exercises only when they score 90% or more on the perception exercises.

Learners' Ability to Perceive Intonation

It has been posited by the existing intonation learning models (e.g., Mennen's (2015) LILt) that acquisition of L2 intonation starts with perception. However, L2 learners' ability to perceive and distinguish intonation cues has been a subject of debate. In particular, when L1 and L2 intonational systems differ significantly, L2 learners may experience difficulties in perceiving meaning communicated through subtle prosodic elements (Pell & Skorup, 2008). Due to this obstacle in learners' perception of prosodic cues, some have argued that learners' ability to perceive speech features needs to be assessed as it is a critical steppingstone leading to speech production (e.g., McGregor & Reed, 2018) and is imperative for the interpretation of pragmatic meaning (Celce-Murcia *et al.*, 2010).

There are several methods of measuring learners' ability to perceive prosodic cues in existing research, most common being discrimination tasks, categorization tasks, and oddity tasks (Nagle & Baese-Berk, 2021). The three tasks are principally similar in that learners are asked to listen to acoustic stimuli and compare them. In the discrimination task, participants listen to two stimuli and decide if they are the same regarding certain prosodic features. For example, if a participant listens to two utterances with rising intonation, they are asked to indicate whether they thought the intonation in the two sentences was the same or different. The categorization task follows the ABX format and involves three auditory stimuli where the target stimulus needs to be assigned to one of the "anchors" (A or B). In this task, a listener hears two sentences with rising (A) and falling (B) intonation. Then, a third sentence with rising intonation (C) is played. The listener is asked to identify which of the first two sentences had the same intonation as the third one. Finally, in the oddity task, listeners are given three stimuli and asked

to decide if they belong to the same category (e.g., all level intonation) or if there is an odd stimulus among the three.

Studies investigating the link between perception and production of L2 phonetic features have shown that this relationship relies largely on learners' individual differences (Nagle & Baese-Berk, 2021). For example, proficiency, length of stay, L1 background, and aptitude have been found to be significantly related to learners' perception of segmentals (Baese-Berk & Samuel, 2016; Jia *et al.*, 2006; Nagle, 2018; Rallo Fabra & Romero, 2012; So & Best, 2010). In relation to prosodic perception, however, research has not been so extensive. The existing findings indicate that L2 learners are usually sensitive to the prosodic cues, although they still experience difficulties in perceiving them. Thus, looking at the individual differences in L2 prosody perception, Kimura *et al.* (2010) examined how well Spanish learners from Japan can perceive the most common intonational cues that are used to distinguish declaratives and interrogatives. The perception tasks in the study included identification tasks where listeners were presented with one sentence at a time and asked to decide if it was a question or a statement. The results showed that learners had trouble perceiving intonational cues not present in their L1. Furthermore, Cho and Dewaele (2021), in an investigation of L2 learners' perception of irony, found that the interpretation of semantic aspects of pitch levels is challenging for Korean learners of English. That is, while L2 learners recognized the different types of cues, they did not perceive them as carrying particular meaning. Analogously, Atoye (2005) found in an examination of perception abilities of Nigerian speakers of English that only 20.7% of them associated intonation cues with correct interpretations of their meaning. These results led Atoye to recommend that instructional trainings should emphasize the teaching of the pragmatic meaning of L2 English intonation on top of mere acoustic analysis. Menjo (2018) investigated

the development of English prosody use in production and perception tasks over time by Japanese learners of English as well as factors that may affect the development of their perception and production of prosodic features. For the perception part in particular, a battery of tasks from the Profiling Elements of Prosody in Speech-Communication (PEPS-C) were employed (Peppé & McCann, 2003). The tasks assessed participants' prosodic abilities in auditory discrimination where no expressive meaning is involved as well as in speech that involved expressive meaning. The perceptual results of the study showed that none of the learners (even those who had spent over 10 years in the target L2 context) met the L1 speakers' level in perceiving pragmatically specific prosodic manipulations supporting the general findings that pragmatic functions of prosody are problematic for L2 speakers of English (Hewings, 1995; Pickering, 2001). Similarly to Atoye (2005), Menjo argued that the perception of the pragmatic features of intonation needs to be explicitly taught to the learners.

While it has been shown that overall L2 learners may face difficulties in perceiving prosody, this pattern varies with regard to certain intonation features. Akatsuka *et al.* (2018) analyzed the effects of explicit teaching on intonation perceptions of L2 English learners from Japan. The learners were offered a five-week perception-based training that focused on falling, rising, and fall-rising intonation. The study revealed that falling tones were the easiest to perceive for the L2 learners, followed by rising and fall-rising. In a follow-up study, Hori *et al.* (2021) compared L1 and L2 English speakers' identification of tone choices before and after a similar training as in Akatsuka *et al.* Although learners made identification errors related to rising intonation, the results showed that their perception of fall-rising and falling tones improved significantly after training, yet, it still did not reach the L1 level. Adding prominence to the analysis of learners' perception of tone choices, Ngo and Setter (2011) investigated the

effectiveness of instruction that focused on these two features reflected in the performance of Vietnamese learners of English in a perception test. The developed training centered around intonation of sentence types (statements, questions, etc.) and discourse information types (new vs. given information). The instruction lasted for 12 weeks and included four stages: planning, acting, observing, and reflecting. As a result, the experimental group improved significantly in their perception of prominence with the results being evident already halfway through the instruction.

While research on L2 perception of intonation is mostly in consensus that learners experience difficulties in identifying prosodic cues in their target language, much is still unknown about the link between the learners' perceptual abilities of intonation features and their production of those features. Additionally, the role of learners' prosodic perceptual ability in the development of the pragma-prosodic form-meaning mappings need to be clarified. In other words, even if learners are able to perceive intonation cues, it is yet to be determined whether they associate the cues with certain meanings.

In order to shed light on these areas of lacking research, the present dissertation study will record and analyze learners' changes in pragma-prosodic perceptual abilities during instruction through the tracking of learners' performance on the perception tasks as well as intonation awareness questionnaires. The study will also analyze the relationship between the learners' pragma-prosodic perceptual abilities and their performance on the request production tasks before and after the instructional training. Moreover, taking into consideration the findings reported in the current section, the developed instructional training will include intonation identification and discrimination tasks that will aim at training learners' perceptual ability. Once

learners reach 90% accuracy in completing the perceptual exercises, they will be able to move on to the production part of the training.

Empirical Research on L2 Prosody Instruction

Recent years have seen a steady increase in research of L2 prosodic instruction, especially since it has been found that deviations in suprasegmental production lower listeners' assessment of L2 speech comprehensibility more than segmental deviations (e.g., Kang *et al.*, 2010). Instructional studies in L2 prosody have shed further light on the proposed acquisition theories discussed above and offered contributions to our understanding of the factors that influence the effectiveness of such instruction. Much as the research in L2 prosodic instruction has been conducted in various contexts with learners from different L1 backgrounds and with different L2 targets, several instructional aspects seem to appear as universal across studies in the domain: (1) the taught suprasegmental feature(s), (2) type of instruction, (3) type of outcome measure, and (4) length of instruction. The next section provides an overview of the current instructional research as well as the factors affecting the effectiveness of prosodic instruction and methodological approaches to teaching L2 prosody.

Instructional Targets, Type, and Length of Instruction

Research on L2 prosody instruction has varied with regard to the instructional foci and duration of instructional trainings. Derwing and Rossiter (2003) provided their learners with 12 weeks of perception- and production-based pronunciation training (20 hours total) that centered on a group of global prosodic features including prominence, intonation, rhythm, amplitude, and speech rate. Before and after the instruction, the participants were asked to produce a short spontaneous monologue describing the events in a cartoon. Their production was then assessed

by judges that provided an overall impression of the speech samples as well as accentedness, comprehensibility, and fluency ratings. The results revealed that learners' fluency improved significantly after the training, and that judges reported a lower number of negative overall impressions of L2 prosody. In another comprehensibility-related study, Saito and Saito (2017) examined the effectiveness of the instruction on word stress, rhythm, and tone choices in developing learners' global comprehensibility. The training lasted over six weeks, and the participants received an explicit instructional treatment that focused on one feature at a time. Each lesson briefly introduced a set of rules about the target intonation features, which was followed by perception- and production-based activities designed within the communicative framework (Celce-Murcia *et al.*, 2010). As the outcome measure, the study included trained and untrained paragraph reading tasks. As a result of the training, learners displayed significant gains in comprehensibility, which were associated with improvements in the use of the prosodic features, particularly intonation ($d = .91$ for intonation in untrained texts). Analogously, Gordon and Darcy (2016) detected significant comprehensibility improvements of English L2 learners' speech, but after a short, three-week explicit instructional training focusing on rhythm, stress, reductions, and linking, that included recognition and discrimination tasks as well as production-based activities that included guided practice and communicative tasks.

Using a sentence-reading and a picture-description task as outcome measures, Zhang and Yuan (2020) examined the effectiveness of an 18-week training in L2 English suprasegmentals that included sentence stress, strong and weak forms, and linking in addition to the suprasegmentals from Saito and Saito (2017). The instruction followed a "presentation-practice-production" structure introduced in Derwing *et al.* (1998). Similarly to the findings in Derwing and Rossiter (2003), the learners that received suprasegmental training in the study made

statistically significant progress in comprehensibility not only at the controlled sentence-reading level, but on the spontaneous level, as well. Importantly, the gains were maintained even in the delayed post-test, 20 days after the end of the course.

While the studies discussed above typically included both perception and production activities, Lee *et al.* (2020) examined the effects of perception vs. production-based methods of pronunciation instruction among Japanese learners of English. In contrast with Derwing and Rossiter (2003), the participants received only two weeks of either perception or production-focused instruction (two sessions of 30 minutes) on English suprasegmental features. In order to measure the learners' progress, the study used controlled and spontaneous production tasks. While controlled tasks were designed to assess the learners' more explicit knowledge of L2 prosodic features, free production tasks measured the unconscious use of such features by the learners. The effectiveness of the type of training varied depending on the time of the post-tests. Thus, whereas production-based training was associated with immediate gains in learners' use of suprasegmentals ($d = .73$), the delayed post-test hinted at the superiority of perception-based instructional methods ($d = 1.27$). However, since the effects of both instructional types were substantial, the authors made a case for implementing both types of instruction in improving learners' prosody.

A considerable body of synthetic and meta-analytic research has also offered evidence for the effectiveness of L2 prosodic instruction. Saito (2012) reviewed 15 studies focusing on the success of pronunciation instruction overall as well as on the factors that affect instructional effectiveness. In particular, Saito examined the relationship between the effectiveness of instruction and the chosen features (segmentals vs. suprasegmentals), type of instruction, and type of the outcome measure (controlled vs. spontaneous). Seven studies in the sample centered

around suprasegmentals including speech rate, rhythm, and prominence (e.g., Derwing *et al.*, 1997; Kennedy & Trofimovich, 2010). All of these studies showed significant improvements in learners' production of suprasegmentals. Additionally, the meta-analysis revealed that instruction that focuses on form is more beneficial for the learners' spontaneous production, and that controlled outcome measures were better indicators of those improvements.

Similarly to Saito (2012), Lee *et al.* (2015), in reviewing 86 studies, found consistent, strong effects of pronunciation instruction on prosodic production for both within- and between-group contrasts (Cohen's $d = .89$ and $.80$ respectively). Some of the highest effects were associated with instruction that focused on intonation ($d = .86$) and prominence ($d = .96$). In parallel with Saito's findings, Lee *et al.* also found that the strongest instructional effects ($d = .96$) were associated with studies that used controlled outcome measures to estimate the learners' gain, possibly due to their similarity to the practice activities in the trainings. More recently, a synthetic study by Saito and Plonsky (2019) investigated the meta-analytic evidence for the effectiveness of perception vs. production-based instruction. Saito and Plonsky (2019) compared articulatory and auditory training and their impact on the development of the perception and production of L2 prosody. After analyzing the results of 77 studies, the authors concluded that prosodic instruction appeared to be more effective when it targeted learners' production of specific suprasegmental features. While this meta-analysis also resulted in stronger effects associated with controlled production outcome measures, Saito and Plonsky supported Saito's (2012) call for the use of both controlled and free tasks in pronunciation instruction to be able to "capture the multifaceted nature of gains resulting from pronunciation teaching" (p. 693).

Length of instruction has also attracted the attention of research synthesists. Thomson and Derwing (2015) reviewed 75 studies on L2 pronunciation instruction overall and found a

spread of instructional time in the collected sample of studies between 30 minutes and 70 hours. Since the aim of the study was to describe the existing research practices in the domain, it did not discuss the relationship between instructional duration and its effectiveness. Adding the missing meta-analytic evidence, McAndrews (2019), in a survey of research investigating the effects of instruction of suprasegmental listening skills, determined that learners displayed significant improvement in their prosodic comprehension skills even after short instructional trainings that lasted a few hours ($d = .94$).

Visual Aids in L2 Prosodic Instruction

There is little doubt that visualization of prosodic aspects has been one of the most frequently used techniques in L2 prosodic instruction, especially in computer-assisted settings. Visual displays of prosodic features, such as pitch, on computer screens have been found to be relatively user-friendly and efficient in giving individualized feedback to the user (Hardison, 2010; Hincks, 2015; Pennington, 1999). Although some have argued that visualized feedback in the form of spectrograms and waveforms is challenging and burdensome for the learners (Levis, 2007; Setter & Jenkins, 2005), the alternative of audiovisual pitch tracing has been found to be interpretable, cognitively less demanding and, therefore, helpful in the acquisition of L2 intonation (Chen, 2020; Chun, 1998; Yuan *et al.*, 2019). Hincks (2015) also argued that displaying the pitch contour and sound form of a learner's utterance next to those of a model utterance while the learner is listening to the recordings of the utterances, may help achieve the goal of matching learner's intonation with the target.

In order to evaluate the effectiveness of visual pitch displays, Hardison (2004) provided L2 French learners with visualized pitch contours for feedback during instruction and found that the participants' L2 prosody improved significantly. In a pre-post experimental study, Ramírez-

Verdugo (2006) included activities using audiovisual displays of pitch contour in a 10-week training in English intonation. The results of the post-test showed that speakers made significant gains in their prominence and tone choices, among other features. Additionally, they were perceived as more intelligible by native speaker raters. Zhao *et al.* (2020) also detected improvement in learners' intonation production after a five-day audiovisual training with feedback, especially for pitch patterns in yes/no questions and prominence. Finally, as a result of a fully online instructional training that included meta-instruction videos as well as instant visual pitch contour feedback, Jiang and Chun (2021) found statistically significant improvements in learners' speech comprehensibility and their speaking confidence.

Although visualization-based techniques have evidently been used with various learning targets, Levis and Pickering (2004) emphasized the importance of providing visual feedback together with the presentation of intonation in connected speech to help the learners understand intonation functions in discourse. After examining the use of paratones (intonational paragraphs) and distribution of tone choices in English learners' reading of out-of-context sentences and discourse-level texts, the researchers argued that providing visualization-based practice with discourse-level intonation features is a must in intonation instruction.

While pragma-prosodic trainings have been rare (see limited overview below), the existing effective practices in intonation instructional design will be adopted in the present dissertation. Beginning from instructional targets and the type of training, research has indicated that tone choices and prominence can improve as a result of explicit instruction that includes perception and production activities. Furthermore, regarding the length of instruction, since even short trainings have been found to be effective, the training designed in this dissertation will not exceed 3 hours 15 minutes. Finally, the training will involve audiovisual pitch contour displays,

especially in the production activities and in feedback to facilitate participants' practice of tone choices and prominence.

Research on Pragma-prosodic Instruction

The studies discussed above looked at the effectiveness of suprasegmental instruction that was usually measured by the improvement in native speakers' comprehensibility, intelligibility, and/or ratings of L2 controlled or free production (see e.g., Derwing & Rossiter, 2003; Jiang & Chun, 2021). However, very few existing studies have examined the pragmatic gains as a result of such instruction. Among these few studies, Ramírez-Verdugo (2006) presented findings from a computer-assisted study on the learning of English intonation (tone choices and prominence) in requests by Spanish learners. The study introduced a multi-sensory approach that incorporated auditory and visual representation of the speech produced by the learners in controlled and spontaneous contextualized production practice. More specifically, participants were asked to listen to dialogues that represented requests and observe the pitch contour displays for the dialogues on the spectrograms made with *Speech Analyzer* software. Additionally, the learners produced their own dialogues and compared the contours that resulted from their dialogues and the model ones. Apart from observing the pitch movement of the models and learners' own production, the training included a focus on the meaning of the intonation features in the dialogues. The learner's production before and after the training was assessed by four L1 English speakers. The 10-week training (one session a week, around 8 hours total) resulted in learners' significant improvements in intonation awareness as well as the use of intonation in speech. That is, the training group improved their use of the tone system and the prominence organization of the utterances while the control group gave equal prominence to every item in an utterance regardless of its importance in the intonation structure. While the training lasted for 10 weeks,

learners' improvements were noticeable throughout its duration. In the final questionnaire, the learners reported that visual displays of intonation helped them become more aware of the functions of prominence as well as tone choices.

McKinnon (2017) investigated the prosody/pragmatic instructional interface in a TBLT-based study that focused on intonation in Spanish declarative and imperative sentences. The study participants received meaning-focused instruction that involved explicit explanations of the functions of the intonation contours in the sentences of interest as well as free production training coupled with feedback on learners' production. The learners' production in pre- and post-tests was then acoustically analyzed using PRAAT. The results showed that while the participants did not use intonation to distinguish between imperatives and declaratives in their production before the training, they modified their pitch range, prominence, and tone choices in the post-test.

In another pragmatics-related study of L2 prosody instruction, Tareva and Polushkina (2020) developed an ESP training aimed at improving the clarity and persuasiveness of engineers in spoken discourse. In developing the training, the researchers conducted discourse analysis of a corpus of engineering discourse in order to select the target components of pragma-prosodic competence (prominence, speech rate, tone choices, pauses) and understand their functions in realizing the communicative purpose in discourse. The instructional method in the study included perception and production activities as well as controlled, guided, and free production following the communicative approach principles (Celce-Murcia *et al.*, 2010). Thus, learners started with listening to recordings of public talks and marking their pause structure, pitch amplitude, as well as prominence and tone choices. For prominence, the learners were taught how to emphasize words they want by using stress and the reasons behind making such an

emphasis (e.g., introducing new information). They then proceeded to recording their own technical presentations after marking the intonation features in them. In addition, learners were provided with opportunities for self-monitoring and reflection on their production. The results of the self-, peer-, and teacher evaluations of the post-instructional presentations showed that the learners became more proficient in using intonation to keep the attention of the audience and signal new and given information.

The current section has highlighted that research in the domain of L2 pragma-prosodic instruction overall has been relatively limited. While some aspects of instruction, like the effectiveness of visual feedback, length of instruction, or explicit vs. implicit instruction, have been studied quite extensively, there is not enough research that could clarify the aspects of intonation instruction framed within pragmatic constructs. The separate fields of L2 pronunciation and pragmatic instruction, as well as the domain at the cross between these two fields, would benefit from a more thorough examination of L2 acquisition of the pragmatic meanings of intonation features as well as the instructional design aspects of pragma-prosodic trainings. Due to the lack of research on pragma-prosodic training, the instructional design in this dissertation will largely rely on the existing findings regarding the instructional features that have been found to be effective in prosodic instruction, i.e., explicit instruction on the pragmatic meaning of intonation cues, the use of controlled perception and production activities leading to free production, etc. At the same time, techniques that were found effective in pragma-prosodic studies as well, (e.g., use of pitch contour displays in Ramírez-Verdugo, 2013) will also be incorporated into the training.

Cognitive Individual Differences and L2 Pragma-Prosodic Acquisition

The acquisition of intonation in a second language is a complex perceptive and productive process that entails a thorough understanding of the prosodic features of a target language and appropriate instructional methods. In this regard, whereas intonation learning in L1 is largely unconscious (e.g., Snow & Balog, 2002), L2 learners often need the knowledge about the form, function, and meaning of intonation in L2 spoken discourse (Ramírez-Verdugo, 2006).

Moreover, the acquisition of L2 intonation can be quite difficult for learners, especially since some prosodic features are language-specific (Yuan *et al.*, 2019). As mentioned earlier, L2 learners often transfer intonation strategies from their first language to the target language. The drawback of this L1 intonation transfer, however, is that learners over-use the limited range of intonation patterns in their target language leading to miscommunication (Ramírez-Verdugo, 2006).

Apart from the effects of learners' L1s on the acquisition of intonation, several aptitude-related individual differences have been found to predict L2 learners' intonation and pragmatic development. Skehan (1989) identified four traits that constitute overall language aptitude: phonemic coding ability, grammatical sensitivity, inductive language-learning ability, and explicit memory. Although these global traits account for the acquisition of a second language overall, not all of them relate directly to the acquisition of the L2 prosodic system. Moreover, while Skehan's model mentions memory, it overlooks the working memory component. To this end, Trofimovich *et al.* (2015) singled out learner factors that reflect pronunciation development aptitude. This cluster of variables included aspects of memory, namely, working memory (WM), musical aptitude, and auditory processing (Hu *et al.*, 2013; Pei *et al.*, 2016; Sun *et al.*, 2021). On the side of pragmatic acquisition, cognitive individual differences like working memory

(Taguchi, 2008) have been shown to play a role in pragmatic production. This section will detail three cognitive individual differences, namely, working memory, musical ability, and auditory processing, as they relate to pragma-prosodic acquisition.

Working Memory

Working memory (WM) has been considered one of the most important predictors not only of L2 learning in general but phonological learning, as well (Baddeley, 2003; Miyake & Friedman, 1998). Baddeley (2003) defined WM through its functions as the temporary *storage* and *processing* of information used for cognitive activities. In describing working memory, Baddeley and Hitch's (1974) multicomponent model has been the most widely used conceptualization of the mechanisms behind WM in SLA research (Juffs & Harrington, 2011). The model posits that WM manipulates the storage of information for better utilization of the cognitive abilities and consists of three subcomponents: the phonological loop, the visuospatial sketchpad, and the attention control system that oversees the information manipulation (Baddeley & Hitch, 1974; Baddeley, 1996). After a series of experiments, Baddeley (2000) introduced another component to the model, the *episodic buffer*. This component is considered to be responsible for the temporary storage of sensory information as well as episodic traces while they are being integrated into long-term memory.

According to the model, the components of WM work simultaneously led by the central executive attention control system that coordinates the rest of the mechanisms. Thus, the phonological loop consists of two subcomponents, a temporary storage and a subvocal rehearsal system. The temporary storage holds verbal information over a few seconds until it fades from the cognition unless it is followed by the subvocalization of information for its retention. The phonological loop is the WM mechanism that is most closely associated with pronunciation

acquisition. Baddeley *et al.* (1998) and later Baddeley (2003) explain that the phonological loop serves as the temporary storage capacity for the phonological forms of newly encountered words. Through learners' recurring exposure to the phonological forms and their repeated subvocalization, the newly built form-meaning "chunks" are then passed into the long-term memory to become part of the phonological system. Simultaneously with the work of the phonological loop, the visuospatial sketchpad integrates spatial, visual, and kinesthetic information into a temporarily stored combined representation (Baddeley, 2003). Finally, the episodic buffer, which heavily depends on the central executive, is responsible for integrating information from different modalities (verbal and non-verbal) into "chunks" that are stored in long-term memory.

In relation to language acquisition, working memory capacity (defined as an individual's ability to hold information in memory temporarily while performing cognitive operations over that information and, often, over other information) has proved to be a robust predictor of a wide range of complex cognitive skills, including language comprehension capacity, particularly in later stages of L1 and L2 development (Caplan & Waters, 2005; Waters & Caplan, 1996). While WM has been examined in L2 contexts as a predictor of grammar and vocabulary learning as well as general L2 proficiency (e.g., Juffs & Harrington, 2011; Martin & Ellis, 2012), links have also been detected between WM and speaking CAF (complexity, accuracy, fluency) measures. For instance, Guara-Tavares (2008) and Ahmadian (2012) found positive medium-sized correlations between WM and L2 learners' speech rate. Rafiei *et al.* (2019) detected a positive relationship between WM and speaking accuracy ($r = .55$). WM was also found to be a predictor of learners' speech imitation abilities (Baker, 2008).

In terms of L2 pronunciation, little research has been done to investigate the relationship between this aspect of speech and WM. In a few existing studies, Rota and Reiterer (2009) found a correlation between WM capacity, measured with forward and backward digit span tasks, and development of L2 pronunciation skills. Papagno *et al.* (1991) also observed that preventing learners from subvocal rehearsal of L2 input reduces the learning of new phonological forms. This finding suggests that WM capacity, particularly, its phonological loop system, facilitates L2 learners' phonological awareness and thereby, the development of their L2 phonological system.

Research has also been sparse regarding WM and L2 pragmatic acquisition. Taguchi (2008) examined the relationship between pragmatic comprehension of conversational implicatures and working memory measured by a reading span test (Osaka *et al.*, 2003) and lexical access measured by a word recognition test (Segalowitz & Freed, 2004). The study found a significant relationship between lexical access and response times in the pragmatic listening task but not with working memory. On the other hand, Ahmadian (2020) reported that WM capacity operationalized through an operation span test was strongly and positively correlated with improvements in production and comprehension of refusals, particularly in the implicit instruction group.

In an investigation of behavioral and neurobiological predictors of L2 learning, Hu *et al.* (2013) used short-term memory measures of digit span and non-word repetition tasks to evaluate the storage and processing capacities of phonological WM and discovered that this individual difference was not able to predict L2 pronunciation aptitude in advanced learners. In contrast, Darcy *et al.* (2015) found that learners with higher phonological WM storage capacity operationalized as a sentence repetition test with last word recall and a paired-associates learning task developed a more native-like phonological processing of L2. In the domain of L1 speech

perception, Jacquemot and Scott (2006) found that better success in prosodic discrimination of two similarly sounding sentences is associated with greater WM capacity. Similar results were observed by Stepanov *et al.* (2020) but with monolingual and bilingual children listening to sentences in an unknown language (French).

Musical Ability

Trofimovich *et al.* (2015) operationalized musical ability as a person's ability to internalize music that they were exposed to and are no longer hearing. A typical example of one's musical ability is the capability to distinguish between two closely related, but not identical, melodic contours after listening to them. Importantly, musical ability is thought to be supported by WM capacity (Christiner & Reiterer, 2013). While some neurological research has suggested that verbal and musical STM mechanisms (sub-systems of WM) are separate (see Caclin & Tillmann, 2018 for a review), Schulze and Koelsch (2012), in a series of neuroimaging experiments, showed that some aspects of verbal processing (e.g., lexical tone in an L1) and music share processing mechanisms supported by WM.

It has been empirically demonstrated in both L1 and L2 speech processing research that suprasegmental features and music share similar acoustic characteristics including fundamental frequency, amplitude, and duration (Patel *et al.*, 2008; Pei *et al.*, 2016). Indeed, Patel (1998) observed that subjects with impairments in musical discrimination and perception often encountered similar challenges in processing of linguistic prosody. In an fMRI study, Nathoo *et al.* (2005) observed that an epileptic patient's right superior temporal gyrus was activated during a word generation task. After an operation on the right temporal lobe, the patient did not experience speech arrest; however, they had loss of intonation and singing abilities. The researchers inferred that suprasegmentals that spread over longer spans of speech, such as

intonation, and music are predominantly processed by the right hemisphere. Supporting the results from Nathoo *et al.*, Schön *et al.* (2004) found evidence of neural overlap in the processing of prosodic and musical pitch by measuring ERPs of musician and non-musician participants who were asked to listen to melodies and sentences and detect pitch violations on the final note or word. Importantly, pitch manipulations in both language and music elicited similar ERPs in both groups of participants emphasizing the role of pitch in musical and linguistic processing. Findings from other perceptual L1 studies also strongly indicate that higher levels of musical aptitude are associated with greater phonological awareness in spoken language (Magne *et al.*, 2016). In other words, those with higher music aptitude are likely to be more sensitive to differences in prosody. The reciprocal relationship between music aptitude and prosody are also manifested in musicians who speak a tonal language having absolute pitch more frequently (Van Hedger & Nusbaum, 2018).

With regard to L2 speech, several studies observed high predictive power of musical aptitude in both receptive and productive L2 phonological ability. For instance, Slevc and Miyake (2006), in measuring L2 phonological ability in Japanese immigrants in the U.S., discovered that musical aptitude could account for 8% of the variance in English productive and 12% of the variance in English receptive phonology. An even stronger predictive power was observed in Sheppard *et al.* (2007) who reported self-assessed productive music aptitude was able to explain 17% of variance in English pronunciation ability. More recently, Pei *et al.* (2016) investigated the effects of musical aptitude and musical training on phonological production in four L2s. They found that the group of learners with musical training performed better in L2 suprasegmental production (tone choices, rhythm, and stress) than their counterparts without musical training. Besides, the study showed that learners with high musical aptitude performed

significantly better in suprasegmental production as opposed to those of low musical aptitude suggesting that musical aptitude may have an effect on L2 phonological production. This observation reiterates the findings of previous research on suprasegmental production with Polish (Zybert & Stępień, 2009) and Chinese (Pei & Ding, 2013) learners of English.

Finally, a small number of instructional studies examined the role of musical aptitude in the effectiveness of phonological instruction. Schwab & Dellwo (2019) measured the effect of musical ability on French and German listeners' performance at discriminating stress contrasts in L2 Spanish before and after a short (3 hours) perceptual training in Spanish. The results were somewhat inconclusive with high music aptitude French listeners showing better stress discrimination before and after training and German listeners not revealing significant effects of musical aptitude on the discrimination of stress contrast. Furthermore, no relationship between musical abilities and the amount of improvement after training was found. Conversely, a study by Komar (2017) revealed significant improvements in the acquisition of prosodic tones of both musically trained and untrained EFL learners after a longer 30-hour intonation instruction. The results also indicated that learners with musical training scored higher during the pre-instruction test than their non-musically educated counterparts.

Auditory Processing

Previous investigations of individual differences have largely overlooked the role of auditory processing in acquisition of L2 intonation. However, this learner characteristic has very recently attracted attention in the field of SLA. According to Sun *et al.* (2021), auditory processing refers to the “ability to precisely and accurately perceive sound characteristics” (p. 3) and is often measured by participants explicitly discriminating or reproducing individual acoustic dimensions of sound, or by examining the size, stability, or accuracy of neural responses to sound.

Emerging research in L1 acquisition has demonstrated that auditory processing is linked with various processing skills, including phonological awareness (e.g., Casini *et al.*, 2018; Won *et al.*, 2016) and, consequently, reading ability (Gibson *et al.*, 2006). In SLA research, auditory processing has been examined with regard to the segmental and suprasegmental aspects of L2 speech perception. Wong and Perrachione (2007) measured the predictive ability of auditory processing in the development of L2 Mandarin tone perception. The L1 English participants without previous exposure to a tonal language received training on identifying three Mandarin tones (falling, rising, and level). The study results showed that the learners who could better identify pitch patterns before training were more successful throughout the training in tonal perception, that is shared with prosodic and musical pitch processing in early L2 learning stages.

Other studies have similarly established the role of auditory processing ability in L2 English prosodic perception. Kachlicka *et al.* (2019), for example, examined the link between auditory processing and perception of L2 English segmentals and found that auditory processing facilitates the earlier stages of L2 speech perception proficiency. In a conceptual replication of Kachlicka *et al.*'s study, Saito *et al.* (2020a) added suprasegmental measures of L2 prosodic proficiency. The additional tasks tapped into the participants' ability to identify and decode prosodic information in sentences. Specifically, they measured the participants' perception of sound duration and amplitude (temporal information processing) and discrimination of pitch and spectral shape (spectral information processing). The results indicated that temporal measures of auditory processing accounted for L2 proficiency.

In a follow-up study, Saito *et al.* (2020b) tested the findings of auditory processing underlying L2 acquisition, in particular, its pronunciation learning component. Thus, the auditory processing profiles of the participants were compared to various dimensions of their L2

pronunciation proficiency. The encoding and reproduction aspects of auditory processing explained differences in L2 pronunciation proficiency, specifically their segmental and prosodic accuracy and speech rate. Based on the results of these studies, Saito *et al.* (2020a, b) argued that more efficient auditory processing, as a component of language aptitude, assists learners' acquisition of suprasegmentals. The association between explicit auditory processing and L2 pronunciation development was further supported in the results of Sun *et al.*'s (2021) longitudinal immersion study. Chiefly, the study showed that participants' L2 ability to remember and reproduce musical sequences measured in the beginning of the training significantly correlated with their prosody perception scores after the immersion ($r = .46$). While the results suggest that auditory processing appears to facilitate extraction and reproduction of pitch contours and sound duration, the causality in this relationship needs further investigation, and the role of other aptitude factors, such as working memory, should not be ignored.

It is evident from the review of studies on the role of phonological working memory, musical aptitude, and auditory processing in L2 pragma-prosodic development that research in this domain is still scarce and somewhat inconclusive. While music aptitude seems to be a strong predictor of L2 intonation learning across studies, the existing findings are uncertain with regard to the role of WM in the acquisition of L2 intonation. Moreover, auditory processing, which appears to be a potentially crucial predictor of L2 prosodic learning, requires more research to clearly determine its significance in intonation acquisition. Finally, very little is known about the interrelations between cognitive individual differences and the pragmatic side of L2 speech acquisition. A more comprehensive study of the aforementioned individual differences and the interplay between them will inform the fields of SLA and L2 prosody acquisition and contribute to the small but rapidly growing body of existing research in the domain.

The Current Study

While research on the importance of prosody in pragmatic production and comprehension is being accumulated rapidly, little is known about the applications of this research in teaching contexts. As noted by Kermad (2021), textbooks still underemphasize the discourse pragmatic function of intonation focusing instead on its grammatical or emotional functions (see also Pickering, 2018). It is possible that speech act-specific pragma-prosodic classroom instruction could supplement or even replace generic textbook instruction; however, research in this domain is scarce. Furthermore, it is not yet clear what aspects of intonation are teachable overall and what instructional methods are effective in teaching prosody. In part, the absence of pragma-prosodic features in textbooks and other teaching materials can be due to the vague nature of pragma-prosodic appropriateness. In other words, it is not clear what exactly is considered a pragmatically appropriate use of prosody in spoken discourse. Although Brazil (1997) and Pickering (2018) outline some of the appropriate and polite applications of prosody in discourse, little is known about the effect of prosodic features on listener appropriateness perceptions. Collecting such perceptions could be informative for developing pragma-prosodic trainings and evaluating their effectiveness.

In discussing the effectiveness of prosodic instructional training, one has to also consider the individual differences of learners and their amenability to prosodic instruction overall. Despite the established significance of individual differences like musical aptitude, auditory processing, and phonological working memory in L1 intonation acquisition (Baddeley, 2003; Nathoo *et al.*, 2005), there are gaps in our understanding of how these differences affect L2 acquisition of speech patterns (e.g., Saito *et al.*, 2021). In particular, the existing studies of phonological working memory and L2 intonation offer conflicting findings concerning its

importance for the acquisition of L2 prosody (e.g., Hu *et al.*, 2013). Even though the role of musical aptitude in L2 intonation overall is consistent with L1 findings, its link with prosodic instruction needs further investigation; some argue that musical ability is not important for the effectiveness of such instruction (Schwab & Dellwo, 2019) while others demonstrate the opposite trend (Komar, 2017). At the same time, with the growing interest in L2 intonation learning, other speech-specific individual difference factors like auditory processing start gaining recognition as effective predictors of the acquisition of L2 prosody (e.g., Sun *et al.*, 2021).

Thus, the goal of the study is three-fold. First, the study develops, implements, and tests through acoustic analyses the effectiveness of a prosodic instructional training specific to the context of high imposition requests. Second, the study sheds light on listeners' perception of pragma-prosodic appropriateness in listening to requests produced by L2 English learners and the role of prominence and tone choices in this perception. Finally, the study further investigates the role of learner characteristics, such as musical ability, auditory processing, and phonological working memory, in predicting the success of the prosodic training. The following research questions guide this dissertation study:

Research Questions

1. To what extent does L2 English learners' use of prosody (tone choices and prominence) in their production of high-imposition request role-plays change as a result of pragma-prosodic training?
 - 1.1. What is the effect of developed pragma-prosodic training as measured by the pragma-prosodic changes in the role-plays before and after the training?
 - 1.2. What is the effect of developed pragma-prosodic training as measured by raters' perceptions of appropriateness of the role-plays?

2. To what extent do learners' cognitive individual differences (phonological working memory, musical ability, and auditory processing) predict the use of prosody in high-imposition request role-plays?
 - 2.1. Do learners' cognitive individual differences predict the prosodic features in L2 learners' production of requests before training?
 - 2.2. Do learners' cognitive individual differences predict learners' changes in the use of prosodic features after training?
3. To what extent does learners' intonation perception ability predict the changes in their use of prosodic features after the instructional training?

CHAPTER 3: METHODS

The study investigates the effectiveness of pragma-prosodic instruction on improving EFL learners' prosodic patterns in requests centered around high-imposition campus situations. Evidence of instructional effectiveness is collected by triangulating acoustic analyses of learners' use of prosody before and after the instruction with raters' perceptual judgements of pragmatic appropriateness. Additionally, the study examines the role of learners' cognitive individual differences in the observed change in prosodic features. The current chapter describes the research design of the study including the participants, instruments, and data collection and analysis procedures.

Research Design

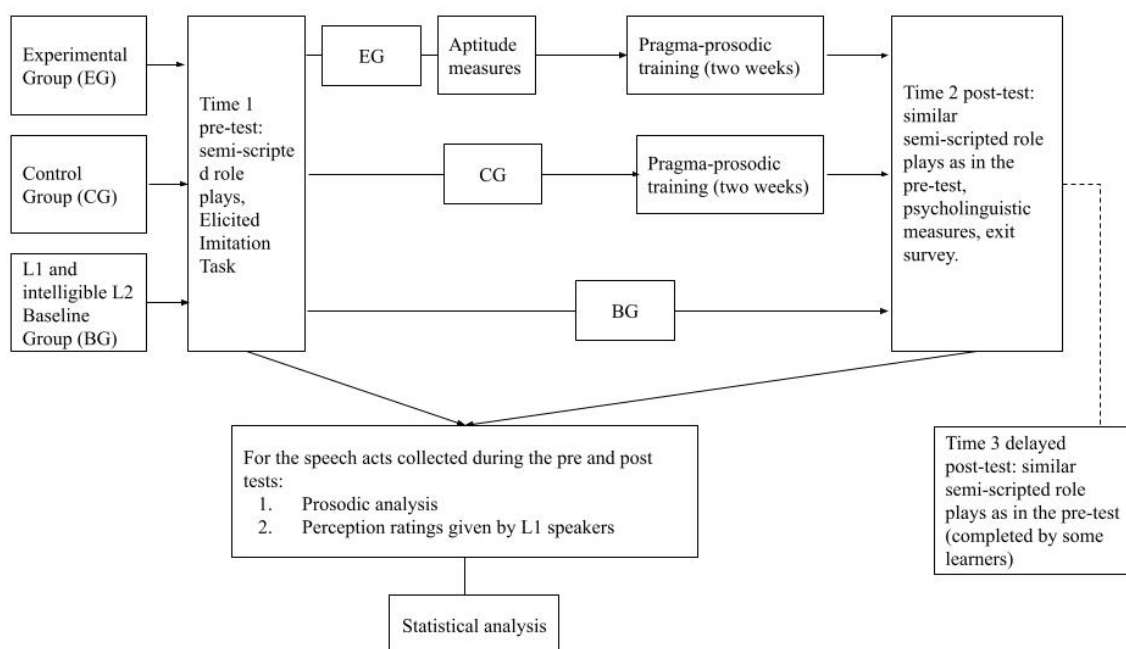
To address the research questions, the study used a quasi-experimental design that includes a pretest, a two-week pragma-prosodic instruction, and a post-test (offered one day after the end of instruction). Two main groups of L2 English learners participated in the study: an experimental

group and a control group. The experimental group received the designed pragma-prosodic training while the control group underwent a sequence of general listening-based activities. The pre-test was given to the two groups one day before the intervention, and the post-test was completed one day after the intervention. Some of the learners in the study from both the experimental and control groups also completed a delayed post-test a week after the instruction was complete. It was impossible to collect delayed post-test production from all of the learners in the experimental and control groups as some of them were recruited as part of a condensed two-week class that ended before they were able to participate in a delayed post-test. During the post-test, the experimental group additionally completed a battery of tasks designed to measure their aptitude for acquiring intonation. On top of that, the study involved a small group of L1 and highly intelligible L2 English speakers who completed the pre- and post-tests to gather the baseline for analyzing learners' production before and after the training.

The second phase of the study design included a prosodic analysis of learners' speech samples collected before and after the instruction and a listener perception survey. The prosodic analysis allowed for an examination of the change in learners' prosodic patterns after the training (within-group comparison) as well as for a comparison of learners' production with the L1 baseline. The perception survey provided insights into the effectiveness of the instruction based on raters' judgments of the pragmatic appropriateness of the learners' production. The survey also elicited raters' perceptions of learners' accentedness and comprehensibility to investigate their relationship with appropriateness perceptions. Figure 2 illustrates the research design of the study.

Figure 2

Design of the Study



Note. EG = Experimental Group, CG = Control Group, BG = Baseline Group

Participants

There were three groups of participants in the study: (1) EFL learners from China, (2) trained L1 and L2 listeners, and (3) baseline speakers. Before participating in the study, all participants signed an informed consent form. At the end of the study, all participants received compensation for their contributions.

Learners

A total of 74 L2 English learners participated in the pragma-prosodic instructional training: 37 learners constitute the experimental group and 37 constituted the control group. To control for L1 background, the participants were Chinese learners of English. Thus, learners in the study were EFL learners recruited from TESOL and Applied Linguistics classes at two different universities in mainland China through researcher's personal contacts. At one university, learners

participated in a study as part of a two-week course taught jointly with Northern Arizona University. At the other university, the study was not attached to a certain course. Learners were assigned to the experimental and control groups randomly. At each of the two universities, there were learners who received experimental treatment and those who participated in control group training. This approach was taken to alleviate potential sampling bias in the case where learners from one university would be assigned to the experimental group, and the other university – to the control group.

While the learners were recruited through convenience sampling, this target group was also chosen for the study for several reasons. First, as evidenced in the literature review above, the pragma-prosodic systems in Mandarin Chinese and English are different, which may cause difficulties in English learners from China in acquiring the pragma-prosodic aspects of communication. Second, Chinese learners of English is the largest group of international students in English-speaking countries (Duffin, 2023) and often find themselves in pragmatic situations like the ones in the developed training. Finally, due to the COVID-19 pandemic and the associated challenges with access to ESL and EFL learners at the time of data collection, this group of learners was the most readily available for online instruction at the researchers home university.

There were 61 female and 10 male learners, while two identified as nonbinary and one preferred not to disclose their gender. The average age of the learners was 23.41 years ($SD = 3.37$). All learners spoke Mandarin Chinese as the first language and reported speaking the following L2s at beginner – intermediate proficiency: Japanese (16), French (4), German (3), Korean (2), Spanish (2), Cantonese (2), and Thai (1). Eight of the learners lived outside of China in English speaking countries for at least two months. In terms of musical training, thirty-nine

learners were able to play a musical instrument and started playing it between the ages of six and 21 years old. Additionally, 33 of the learners had taken music classes at some point throughout their primary, secondary, and high school. The sub-group of learners who took the delayed post-test consisted of 5 males and 34 females; one participant decided not to report their gender. The average age of learners in this sub-group was 24.8 years ($SD = 3.28$).

Learners in the study came specifically from intermediate/high-intermediate proficiency levels. This level was chosen since intermediate learners oftentimes already possess a wider range of vocabulary and grammatical structures needed for appropriate pragmatic performance (Taguchi & Roever, 2017). Thus, it was expected that they would be able to practice the pragmatic-prosodic patterns in their speech act production taking advantage of the lexico-grammatical resources they already possessed. Learners' proficiency was ensured via several means. First, learners were specifically recruited from intermediate-level classes. Second, in the background survey, learners were asked to indicate their scores on standardized tests that they might have taken previously. Not all learners had taken such a test according to survey results; however, those who did fell within the intermediate / high-intermediate benchmark. As such, 37 of the learners had taken the IELTS test with the score of 7.0 being the mode in the dataset (range between 5.5. and 7.5). Four learners took Pearson Test of English (PTE) (scores between 66 and 77) and two learners took TOEIC (scores 805 and 975). Finally, to control for the possible effect of proficiency on the findings, the study employed an Elicited Imitation Task (Yan, 2020) as a measure of learners' proficiency. Details on this task and its benefits for the study are described in the Instruments section below.

Since the study was designed with Chinese learners of English in mind, the recruited sample of participants sufficiently reflected the target population. In its essence, the training

developed in this study targeted English learners from China of various ages and educational backgrounds. With this in mind, the current sample was somewhat constrained by several participant characteristics including their age (early twenties), educational background (university students), and proficiency (intermediate/high-intermediate). These limitations in the learner sample were considered in interpreting the study findings and generalizing them only within the population sample recruited.

Before learners were recruited, an *a-priori* power analysis was conducted using G*Power (Faul *et al.*, 2007) in order to estimate the required sample size for sufficient power ($b = 0.80$). Two main statistical techniques used in the study were linear mixed effects modeling (LMEM) and (multivariate) multiple regression. Although mixed-effects models are powerful techniques, there is yet to be a straightforward methods of correctly rejecting the null hypothesis thus limiting the choice of methods for assessing the sample size needed for building such a model (Kumle *et al.* (2021). One approach to estimating the power, at least for the fixed factors, is to use G*Power for F tests. Since the analysis of fixed effects in LMEM resembles that of an ANOVA, I chose the repeated measures within-factors ANOVA as the statistical test to run power analysis for. The estimated effect size was set at $f = 0.25$, a medium effect size, motivated by previous meta-analytic studies of L2 pronunciation instruction that reported overall medium between-group effect sizes of instructional effectiveness (Saito & Plonsky, 2019). Although for pragmatic instruction, secondary studies (e.g., Plonsky & Zhuang, 2019) reported large between- and within-group effect sizes, a more conservative medium effect size approach was taken to reflect the complex nature of the developed training and account both for prosodic and pragmatic features included in the instruction. The results indicated that with two fixed factors, a sample size of $N = 128$ was needed. Although the number of participants in the study was below this

requirement, each participant provided at least six observations, thus increasing the total number of observations in the study and making LMEM sufficiently powered. For multiple regression analyses with four predictors, the total sample size needed according to G*Power was 36 – a requirement that was met in the study.

Raters

Studies of pragmatic appropriateness and L2 intonation perception have demonstrated substantial variation in the number of trained raters employed to evaluate learner performance. While in pragmatics perception studies, it has been typical to have between two and six trained raters (e.g., Li *et al.*, 2019; Sydorenko *et al.*, 2014; Taguchi, 2011a), studies in L2 speech have generally included larger groups of trained listeners. For example, Isbell (2017) and Crowther *et al.* (2018) used 10 trained raters in evaluating L2 speech samples in their studies whereas Kang (2008) had 29 trained raters judging L2 English speakers' oral performance. Yet, the variable of “goodness of fit” in Derwing *et al.* (2004) was evaluated only by three trained raters.

Considering the general trends in the number of raters in previous research and the feasibility of the dissertation study overall, it was decided to include twenty trained raters in the study from L1 and highly proficient L2 English speakers from a four-year university in the Southwestern United States (e.g., teaching assistants who were MA and PhD students in applied linguistics at the researcher's university). Since the speech acts in the pre- and post-tests included communicative situations that English learners often encounter on campus (e.g., making a request to a professor), the raters had to possess experience teaching in higher education settings to ensure the ecological validity of the pragmatic appropriateness perception survey. Moreover, the rater group consisted of both L1 and highly proficient L2 English speakers as recommended in previous research (Kang *et al.*, 2019). In order to screen the listeners that match the criteria

above, a short survey was conducted at the recruitment stage (see below and Appendix B). The survey included questions about the listeners' linguistic training, teaching experience, and musical training. The raters were recruited after learners had completed the instructional training. This was done in order to ensure raters' retention since all pre- and post-training audio files for the perception surveys had already been collected and so there was no delay in presenting the survey to the raters.

An a-priori power analysis was conducted using G*Power (Faul *et al.*, 2007) before ratings took place. Similarly to the analysis of learner data, LMEM was used as the main analysis technique. For two fixed factors, the required sample size was $N = 128$, which was met since each rater provided between 66 and 68 ratings. According to background survey results, the raters in the study were 14 females and 6 males with an average age of 32.8 ($SD = 8.08$) from seven L1 backgrounds: English (9), Turkish (4), Russian (2), Mandarin Chinese (2), Portuguese (1), Persian/Farsi (1), and Dine (1). Seventeen of the raters spoke several L2s, L3s, and L4s: Spanish (8), German (5), French (3), Italian (2), Japanese (1), Thai (1), Mandarin Chinese (1), Cantonese (1), Korean (1), Kyrgyz (1), Bulgarian (1), Russian (1) and ASL (1), mostly at beginner or intermediate proficiency. All raters had at least one year of teaching experience at the college level, although the reported average was 5 years ($SD = 3.5$). Some of the classes commonly taught included First-Year Composition courses as well as TESOL and Applied Linguistics courses. Seven of the raters were able to play a musical instrument and had started playing it between the ages of six and 45. Thirteen of the raters have taken music classes in elementary, middle and/or high school. Collecting information about musical training was important for the study since research has suggested that those with more musical training may be more perceptive to intonation features of speech.

Before rating the sound files, the raters underwent a short training session to become familiar with the evaluation scales used in the study. Since raters were graduate teaching assistants specializing in the field of applied linguistics and familiar with phonetics and phonology, they already possessed the necessary background knowledge to provide in-depth reports about specific intonation features that motivated their judgements (Gordon & Darcy, 2016). Assessment background was not required since the rating scales used in the study were straightforward and commonly used in L2 phonology research. Nevertheless, a training session with the basic definitions of the intonation features of interest and baseline sound files was provided before rating (see below and Appendix G). It was expected that by relying on their teaching experiences and basic definitions of the assessed constructs, raters would be able to provide instant, impressionistic judgments about the appropriateness, accentedness, and comprehensibility of the speech acts produced by the learners. These constructs were introduced to the raters during the training session.

Baseline Speakers

The last group of participants in the study included three L1 English speakers and three highly proficient L2 English speakers to provide the baseline for the prosodic analysis of the learners' production. The highly proficient L2 speakers were recruited among students in the researcher's graduate program that required high levels of English proficiency for admission and successful completion. The selected participants were three males and three females (M age = 26.17, SD age = 3.76). The L2 English speakers came from Chinese (2) and Spanish (1) background.

Instruments and Materials

This subsection describes the instruments and materials that will be used in the study. The instruments and materials included (1) learner and rater background questionnaires, (2) the pragma-prosodic instructional training materials, (3) the tasks that were given to the learners before and after the instruction, (4) the language aptitude measures used to estimate the learners' working memory, musical ability, and auditory processing, (5) an elicited imitation task to measure learners' proficiency, and (6) rater training materials and the perceptual scale used to measure the rater judgments regarding learners' pre-and post-instructional production, and (6) instruction and delivery for the control group.

Learner and Rater Background Questionnaires

In order to collect the background information on the learners and raters, the study employed two biographical questionnaires (Appendices A and B). The questionnaires recorded basic information about the participants including age, gender, and L1 as well as some group-specific information. The learner questionnaire additionally gathered information about proficiency, the length of English study, and experience living abroad as they have been identified as predictors of L2 English speakers' learning and of pragma-prosodic instructional success (e.g., Kang *et al.*, 2021; Taguchi *et al.*, 2021).

The rater background survey included questions about teaching experience, linguistic knowledge, and accent familiarity. Raters' teaching and language experiences have been found to affect their evaluation of L2 oral production (e.g., Alemi, 2012; Ducasse & Brown, 2009; Johnson & Lim, 2009). Specifically in the field of pragmatics, Walters (2007) and Taguchi (2011a) investigated rater variation showing that factors like language and cultural background

have influenced raters' decisions about the appropriateness of learners' production. In speech perception studies, listeners' accent familiarity has been one of the causes of error variance in rater judgments (Kermad, 2018; Winke *et al.*, 2013).

Additionally, in the domain of prosodic perception, musical training has emerged as a factor connected to listeners' intonation analysis skills. Pastuszek-Lipińska (2004) suggested that listeners' background in musical training may facilitate a higher sensitivity towards auditory discrimination. Dankovicová *et al.* (2007) investigated the relationship between musical training and intonation analysis skills in English. The researchers used a questionnaire to collect information about years of musical training, starting age, grade achieved, instruments played, and listening preferences. In addition, participants completed auditory discrimination tasks (same/different intonation contours) and nuclear tone identification. The study showed that musicians were better in nuclear tone identification. In addition, other studies found musicians to be more accurate in detecting pitch violations in language (e.g., Schön *et al.*, 2004) and to be overall more attuned to the suprasegmental aspects of L2 pronunciation (e.g., Isaacs & Trofimovich, 2010). Since both learners and raters in the study completed perception-based activities, the background questionnaires for both groups included parts about learners' and raters' musical background and training adapted from Dankovicová *et al.* (2007).

Pragma-Prosodic Instructional Design

The main set of materials in the study consisted of a researcher-designed pragma-prosodic instructional training that focused on prominence and tone choices as the two crucial aspects of prosody (Brazil, 1997). The choice of these two features was supported by the existing research in the domain. First, prominence and prosody have shown to be responsive to instructional

trainings (e.g., Lee *et al.*, 2015). Additionally, these features have been found to play a role in interpreting the pragmatic meaning of an utterance (Brazil, 1997; Pickering, 2018).

The training in the study was informed by the existing approaches to teaching discourse-level intonation within a communicative framework for teaching pronunciation (Celce-Murcia *et al.*, 2010; Chun, 2002). As noted by Levis (1999) and Chun (2002), rather than teaching intonation at the phrase- and sentence-levels, prosodic instruction needs to help learners develop awareness of the form-meaning connections between intonational choices and meanings communicated by them. Thus, instructional trainings need to include perception and production activities that teach prosodic features at the *discourse* level. The emphasis of Celce-Murcia *et al.*'s (2010) communicative framework is exactly on such an integration of form and meaning. In the center of the framework is the idea of communicative language teaching, which is applied to pronunciation instruction via five phases: 1. *Description and analysis*, 2. *Listening discrimination*, 3. *Controlled practice*, 4. *Guided practice*, 5. *Communicative practice* (p. 45). The description and analysis phase typically consists of consciousness-raising exercises that are aimed at explaining the meaning of an intonation feature. The listening identification and discrimination phase provides learners with practice in distinguishing the intonation features of interest in speech leading to improvements in production. Following McAndrews's (2019) suggestions, this phase mostly contains categorization activities for they have been more effective in developing prosodic skills. Controlled and guided practice phases are related to the automatization of the intonation skills. Unlike the controlled phase that is related to the learners' acquisition of the target *forms* and accuracy of their use, the guided practice phase has its emphasis on learners' expressing meaning. Finally, the communicative phase lends itself to providing learners with practice of intonation patterns in communicative contexts. Together,

these phases are aimed at helping learners move from the analysis of the target pronunciation features to their perception and finally production, constantly connecting the prosodic forms to their meanings. From this perspective, the framework aligns with Mennen's (2012) L2 Intonation Learning Theory that posits the progression of intonation acquisition from perception to production.

Instruction Delivery Platform. The training in the study was delivered to the experimental group via a custom-built instructional website. The website resembled other Learning Management Systems (LMS) and consisted of six learning modules that were managed by the researcher. While learners completed perception activities independently on the website, production activities were monitored by the researcher, who gave learners individualized feedback on their performance. In addition, the last module of the instruction (Module 6) took place via Zoom, a teleconferencing tool.

The website was built using Python for data processing (e.g., online analysis of learners' production, delivery of speech models and tasks) and Java Script for the learner experience design (i.e., program interface, feedback features, etc.) and hosted on a cloud platform Heroku (Heroku, 2021). The program was linked to two data storage components. The first one, hosted on Heroku, is a SQL database accountable for the participants' responses except for the audio recordings, which were stored on the Amazon S3 container intended for large file storage (Amazon Web Services). The participants were required to access the website via computers or laptops (not cell phones) and complete all training modules in a quiet space wearing headphones with a microphone. The website was available for learners online independent of their physical location. Since data collection took place online, without any physical monitoring, it was not possible to guarantee the learners' lack of distraction. However, all responses in each lesson were

checked by the researcher to ensure that the learners were careful and focused during the training.

In choosing to incorporate online instruction in the study, several points had to be considered. First, as Taguchi (2020) pointed out, the ongoing COVID-19 outbreak has affected the learning processes across the world. Teachers and students have had to adapt to the ever-changing learning conditions, switching between remote and in-person learning. Not only did online instruction delivery ensure the data collection in these unstable conditions, but it also provided teachers with a ready-made course that they could use in their classrooms.

More importantly, however, studies have suggested that digitally mediated remote instruction can be beneficial for L2 learners both in the domains of prosody and pragmatics. Among some of the advantages of using digital spaces for pragmatics instruction, Taguchi and Sykes (2013) named affordances for input, output, and interaction, focus on specific aspects of language, and individualized feedback (Belz, 2007; Sykes, 2008). Taguchi (2020) further explained that the use of structured digital spaces for instruction may provide systematicity and help to make learning objectives clear and create tasks according to those objectives. In addition, such instruction delivery can promote students' learning autonomy in planning, monitoring, and evaluating their progress (see also Cohen, 2005; Kohn, 2009). In prosody instruction, Chun (2019) singled out unlimited input, practice, and repetition as advantages of digitally presented instruction on top of possibilities for visualization of suprasegmental and individualized and instantaneous feedback, parallel to the current discussion in the area of pragmatics.

The website developed for the present study followed the structure of an online curriculum with a live instructor-based component. While prosodic instruction has rarely happened through online platforms, similar websites have been developed in pragmatics research

and have proven to be successful in increasing learners' pragmatic competence (e.g., Cohen & Ishihara, 2005; Cohen & Sykes, 2006; Sykes & Cohen, 2008). For example, Cohen and Sykes (2006) developed an instructional website to explicitly teach 10 speech acts in Spanish. The website consisted of nine instructional modules and included videos that demonstrated the dynamics of interaction, presented considerations for politeness and directness as well as guidelines for using pragmatic strategies for performing speech acts. Reporting on the results of the study, the researchers noted that learners were able to utilize the content of the website in a way that benefitted their own learning and overall positively perceived their experience. Some of the recommendations the authors gave regarding making such learning platforms were varying the instructional activities and methods of content presentation and providing explicit, salient, targeted instruction.

The training developed for this study followed the recommendations in Sykes and Cohen (2008) as well as the overall instructional structure described in previous studies that employed online learner platforms. Each module of the training included video lectures introducing the learning targets as well as prosody perception activities with immediate feedback that the learners completed independently. Learners were allowed to repeat each perception activity an unlimited number of times until they reached the 80% accuracy threshold. The website also automatically recorded learners' attempts in perception tasks as a per-item binary correct/incorrect score on the first attempt. That is, if a learner gave an incorrect response on an item in a perception task, it was recorded as *zero*. In aggregate, these scores showed how many items were recognized correctly on the first try measuring learners' perceptual skills (Research Question 3). As for the production activities, while learners recorded their production through the website, the researcher listened to the recordings and gave feedback about the

appropriateness of learners' use of prosody in requests. Learners were allowed to repeat each production activity taking into consideration the researcher's feedback. They were only able to proceed to a subsequent activity if the researcher found their production appropriate. As mentioned above, to provide learners with communicative practice (Celce-Mutcia *et al.*, 2010), the last module of the training took place over Zoom and consisted of role-plays between the researcher and each of the learners. Thus, while learning in the study was enhanced with the use of technology, there was still an interactive instructor-based component (Blyth & Sykes, 2020).

A pilot was conducted to test parts of the platform with a small group of Chinese L2 English learners. The platform has proven to be successful in presenting the learners with intonation training activities, providing learners with immediate feedback, and recording learners' responses in perception and production exercises (see report on the pilot below). Additionally, the platform was able to integrate prosodic contour images for the sound files and learners' production with the help of *Parselmouth* and *Chart.js*, online pitch contour tools similar to PRAAT (Boersma & Weenink, 2021). Well-designed computer-mediated and web-based training have been found to be effective in scaffolding learners' acquisition of intonation and increasing their L2 speaking confidence (e.g., Jiang & Chun, 2021). In general, the platform facilitated the design of an instructional training by means of training modules that typically include consciousness-raising content and teacher's explanation of the material, and activity modules, where learners actively engage in the exercises (e.g., perception and production activities).

Instructional Design Principles. In agreement with Celce-Murcia *et al.*'s (2010) communicative framework of teaching intonation, the training followed several crucial design principles. The form-meaning mappings of prominence and tone choices were guided by Brazil's

(1997) discourse intonation framework as well as by the existing research in the domain (e.g., Kang *et al.*, 2021; Taguchi *et al.*, 2021). The training offered explicit instruction as it had proven to be more effective in teaching intonation (Lee *et al.*, 2015, 2020; Saito & Plonsky, 2019). In this fashion, the description and analysis phases of the instruction included explanations of the form-meaning mappings of the prosodic features, benefits of using these features, and challenges with incorporating them in speech.

Another principle informing the instruction pertains to the use of visualization techniques, especially to provide feedback (e.g., Jiang & Chun, 2021). Visual aids have been found to be helpful in providing learners with opportunities to compare their production to the model, reflect on the comparison, and thus improve their speech production (Chun, 2002). The instruction incorporated learners' analysis of the prosodic contours of their recordings presented through Parselmouth and Chart.js to allow for self-monitoring and noticing evidence of their progress (Celce-Murcia *et al.*, 2010). The instruction also included individualized teacher feedback on all production activities as it had been suggested that learners prefer such feedback in prosodic trainings (Jiang & Chun, 2021).

Finally, while the prosodic features are central to the training, the teaching of these features was integrated into the communicative contexts of high-imposition requests. The choice of the communicative context for the situations was motivated by previous research on the needs of L2 English learners in EAP settings. In evaluating communicative situations that EAP learners' find themselves in, Ikeda (2017) pointed out that university students, professors, and administrators are "related to each other by their need to negotiate certain matters" (p. 69). For example, a student may be forced to request an extension from an instructor due to unexpected circumstances. Furthermore, in her needs analysis of L2 pragmatic tasks in an EAP context,

Youn (2018) named refusing a professor's request politely and requesting a recommendation letter among the most needed tasks identified by the participants. Informed by the findings of these needs analyses, the instructional situations closely reflected the actual communicative needs of L2 learners in university contexts.

Pragmatic content of training. Since the focus of the instruction in the study is primarily on the prosodic features of requests, the pragmatic content in the training was limited to particular request-making head-act strategies and modifications. The choice of these strategies and modifications was informed by the existing research in L1 English production of requests.

Beginning with Blum-Kulka *et al.* (1989) and their Cross-Cultural Speech Act Realization Project (CCSARP), research has consistently shown that conventional indirectness is the most appropriate and polite way to make requests among L1 English speakers. Even before the seminal CCSARP project, Blum-Kulka (1987) examined the perceptions of politeness and directness in L1 English and Hebrew speakers' interpretation of high- and low-imposition requests. Regarding the L1 English group, she found that Query Preparatory strategies in particular (e.g., *Could you...?*), a sub-type of conventionally indirect strategies, were viewed as the most polite. In fact, they were perceived as more polite than the non-conventional indirect hints (e.g., *It's hot in here* for asking to open the window). L1 English speakers' preference of conventionally indirect strategies can be explained by Grice's (1975) Maxim of Manner, in combination with Brown and Levinson's (1987) notion of face-threat. While the maxim postulates that a speaker must be clear in making their illocutionary point, they also need to minimize the threat to the hearer's face, especially in making such a face-threatening speech act as a request. Blum-Kulka (1987) argues that both conditions are satisfied by conventional indirectness since the speaker appears to be indirect without "burdening the hearer with the

actual cost of true indirectness” (p. 143-144). Other studies of L1 requests (e.g., Barron, 2005; Cenoz & Valencia, 1996; Eslami-Rasekh, 1993; Pinto & Raschio, 2007; Suh, 1999; Tatton, 2008) analogously found the prevalence of Query Preparatory strategies in L1 English request-making.

Since its publication, the CCSARP framework for coding request strategies has been revised by more recent studies. Thus, Taguchi (2006) identified the Mitigated Preparatory (e.g., *I was wondering if...*) as a conventionally indirect strategy that employs a bi-clausal structure. This strategy usually contains a reference to preparatory conditions (e.g., the hearer’s ability or will to comply with the request) in the dependent clause of the speech act. Taguchi’s (2006) study analyzed oral request production of 59 Japanese learners of English in two proficiency groups by rating their performance on a six-point scale that measured overall appropriateness. In addition, the directness levels of the linguistic expressions in requests were identified and qualitatively analyzed. Importantly, learners’ production was compared to an L1 English baseline group. The study showed that 100% of native English speakers used mitigated preparatory strategies to express their requests (e.g., *I’m wondering if...*, *Is there any chance...*) and that this indirect strategy was usually considered more appropriate than any of the direct strategies. Similarly, Taguchi (2012) found that mitigated preparatory strategies appeared in more than 70% of the time in the L1 English speakers’ data in the study that examined the development of Japanese L2 English learners’ production of requests in high- and low-imposition situations. In a follow-up study, Taguchi *et al.* (2015) applied the findings of previous research in choosing mitigated preparatory strategies, namely, *I was wondering if...* and *Would it be OK if...* as the instructional target to examine the effects of explicit instruction in high and low imposition requests on the development of pragmatic competence in L2 English.

In terms of request modifications, research has strongly suggested that grounders (i.e., giving a reason for the request) are the most frequently used mitigation devices in L1 English requests (Barron, 2005; Kostromitina, in press). For example, Barron (2005) investigated the realization of requests by L1 English speakers focusing on the level of directness. A total of 54 participants completed 3 DCTs with varied degrees of power, distance, and impositions. The results of the study revealed that the participants used grounders in over 85% of requests. Analogously, Kostromitina (in press) showed that grounders were the most common mitigating modification (69%) in high-imposition L1 requests elicited with the help of video-based multiple-turn DCTs. Supporting previous research on requestive head-act strategies, the study also found mitigated preparatory to be the most widely used strategy in this type of a speech act (84%).

Based on the results of previous research, it was decided to choose the mitigated preparatory strategy as the focus of instruction in the present study. In particular, to control for the lexico-grammatical forms that the learners were exposed to, two strategies were taught to the experimental group: *I was wondering if...* and *I was hoping that...* These request forms were the most common ones identified in previous studies (e.g., Taguchi, 2006) as well as in one of the pilots conducted in preparation for this study. In addition, the instruction introduced the notion of grounders as the most common mitigation device in requests. The developed instruction for the experimental group (see Appendix D) introduced the learners to this strategy, explained the high-imposition situational contexts in which it can be used, and described what intonation it should be accompanied with.

Prosodic content of training. As explained above, the notions of directness, formality, and politeness have not been directly applied in Brazil's (1997) discourse intonation framework.

Nonetheless, Brazil provided guidelines for interpreting pitch contours and prominence based on the discourse participants' roles and information structuring. Thus, within the framework, proclaiming intonation (i.e., falling tones) is generally used to transmit new information that the interlocutor is not aware of and indicate the end of a turn (see also Hirschberg, 2004). At the same time, referring intonation (i.e., rising contours) does not change the state of speaker-listener convergence as they are used with utterances that contain information that the listener presumably knows. Crucially, Brazil (1997) and Pickering (2018) explained that rising tones are usually used when speakers draw on the information that is already shared between the two of them. While rising intonation is associated with the speaker's desire to have their perspective confirmed, falling tones are linked to a genuine inquiry and request for information. In relation to that, falling tones have been found to indicate sincerity (e.g., Knowles, 2016; Staples, 2015).

Furthermore, one of the central constructs in the discourse intonation framework is dominance, which can be linked to the notion of power distance in pragmatics. Use of rising tones has often been associated with the speaker's dominance. For instance, Cheng *et al.* (2008) found that higher power interlocutors, such as professors in academic supervising positions and supervisors in business meetings, use more rising tones. In discourse, speakers in dominant roles generally have access to all tone choices (rising, falling, level) whereas non-dominant speakers do not have such access. Pickering (2018) stated, for example, that in teaching contexts it is an instructor's prerogative as the person in higher power to use dominant rising or rise-falling intonation. In contrast, in situations where there is an expectation that the speaker will accept a non-dominant role (e.g., a student), such use of intonation would be perceived as rude or even hostile (Brazil, 1997; Pickering, 2018).

Level intonation, on the other hand, helps the speaker shift the focus of the interaction from the state of convergence to the linguistic message itself. It is also generally used when the speaker is planning and organizing the message as they are speaking. Finally, level intonation is at times a signal of hesitancy (Brazil, 1997). When talking about speaker's oblique orientation (or orientation set towards the language item rather than the hearer), Brazil associated level intonation with "immutable knowledge" (p. 138) indicating that the speaker expects that the interlocutors will complete the meaning of the utterance (i.e., "You know what I'm saying, I don't have to continue"). Hirschberg (2004) also suggested that level (or "plateau") contour can be employed for illustrative purposes when the speaker is simply going through the motions of informing their interlocutors. In association to this function, Kang *et al.* (2021) found that L1 English speakers use level contours when providing justification for a request, particularly in high-imposition situations.

Prominence in Brazil's (1997) framework is used to highlight new or contrasting information. Returning to the notion of the speaker's orientation, excessive prominence may cause the listener to switch their attention from the meaning of an utterance to its form—making the speaker sound disengaged and suggesting their oblique orientation. Studies in speech perception have additionally found that overuse of prominence can miscommunicate the speaker's intention in the context of an interaction (e.g., Hahn, 2004; Kang *et al.*, 2018).

If we apply the interpretations above to requests, it becomes clear that L1 English speakers use intonation in this speech act in a deliberate way. Particularly in high imposition requests, the use of rising intonation is oftentimes inappropriate. As Kang *et al.* (2021) explained, when making a request, it is expected that the focus of interaction is on altering the state of convergence between the speaker and hearer rather than on establishing common ground.

The speaker is expected to genuinely ask whether their request can be fulfilled by the listener; thus, falling intonation is more appropriate here. The use of rising intonation may make the speaker sound insincere. When found in requests, rising intonation indicates that the speaker is seeking confirmation that the listener has already agreed to do something or is putting pressure on the speaker to agree to the request (Cheng *et al.*, 2008; Pickering, 2018). Such an assumption, conveyed through rising intonation, may be perceived as extremely inappropriate, especially coming from a non-dominant speaker, as is the case in the target speech act situations in the dissertation study. Turning to level intonation, previous research has suggested that beyond its common function of indicating withdrawal from negotiation (e.g., with a discourse marker *well*) and routine language (Pickering, 2018), level tone is oftentimes used specifically in requestive speech acts when the speaker is giving the reasons for a request (Kostromitina, in press). Since grounders are the most frequently used mitigation device in requests, the use of tone choices is justified by the speakers' need to clarify why they are making the request. Finally, prominence in requests, just like in other speech acts, is related to the clarity of the meaning behind the request and the speaker's direct orientation towards the hearer. By emphasizing only pragmatically meaningful words that are important in a given context, the speakers can effectively convey their request.

The content of the instruction offered in the dissertation study was informed by the interpretations of tone choices and prominence described above. The instruction covered the main content points given in Table 1. It made connections between the mitigated preparatory request strategy, the grounder modification device, and the prosodic features that should accompany them in high-imposition requests. The detailed outline of each instructional session is given in Appendix D.

Table 1

Content of Pragma-Prosodic Instruction

Feature	Meaning
Mitigated preparatory strategy	<i>I was wondering if... / I was hoping if (that)...</i> • Used to make a request to someone in higher power (teacher, employer, advisor, etc.). • Used to make a request that would require a lot of effort/consideration from the listener. • Contains bi-clausal structure. The main clause is followed by the request itself checking the ability or willingness of the listener to comply with the request. ○ ... if you could... ○ ... if you'd be willing to... ○ ... if you'd be able to... ○ ... if I could...
Grounders	• Used to give an explanation or motivation for the request. • Usually go after the request itself (see examples in Appendix D).
Rising tones	• Conveys dominance of a speaker. • Makes the speaker sound insincere and rude in making the request. • Assumes that the speaker knows what you are talking about and that they have agreed to fulfil the request. • Suggests continuity (you have not finished talking). • Oftentimes used in statements and in bi-clausal questions (i.e., mitigated preparatory strategy).
Falling tones	• Does not convey dominance. • Conveys sincerity. • Does not assume that the listener has already agreed to comply with the request. • Genuinely asks if the listener can fulfil the request. • Suggests the end of your turn. • Oftentimes used in yes/no questions (i.e., query preparatory strategy).
Level tones	• Gives a justification for the request (used with grounders). • Used in enumerating reasons for the request.
Prominence (sentence stress)	• Changes the focus of the request from the hearer to the form if excessive prominence (stressing every word or every second word).

- Gets the listener distracted from your request and confused about what you are asking.
 - Helps the listener see the meaning of your request easier if appropriate (1-2 stressed words in a tone unit = thought group).
-

Lessons. The listening materials for the training (i.e., the model audio recordings used in the activities) was adapted from a study conducted by the researcher (Kostromitina, in press). The study used scripted video-based role plays to elicit high-imposition requests (speech acts in focus) from L1 and highly proficient L2 English speakers. Given that another study used to inform this dissertation (Kostromitina & Kang, in preparation) showed that speech elicited with free and scripted role plays was not significantly different in its prosodic composition, the intonational models collected in the pilot were considered authentic and sufficient for the instruction (McGregor *et al.*, 2016). The speech models collected in the studies were used in the listening discrimination phase. In addition, two L1 English speakers recorded models for controlled and guided practice phases of the instruction.

The structure of the exercises in the training was largely adapted from the previously mentioned pronunciation textbooks and studies that are situated within the communicative framework (Celce-Murcia *et al.*, 2010; Chun, 2002; and Jiang & Chun, 2021). The first lesson of the training included description and analysis activities that introduced learners to the meaning of tone choices and prominence in high-imposition requests in relation to their appropriateness (politeness, directness, formality, and sincerity) (partially adopted from Taguchi, 2011a). Apart from introducing the definitions of each intonation feature as well as the speech act of request, the lesson included an example request produced by a L1 English speaker from the pilot study. Learners were asked to listen to the request and observe the pitch contour and prominence plots corresponding to the request that were shown to them through the app. The lesson explained the

essential components of the plots in detail and asked learners to repeat the analysis on a different file. Once learners completed the lesson, the researcher reviewed their answers to make sure that they had a good grasp of the main concepts taught in the training and were ready to proceed to perceptual training (see Appendix D1 for a detailed description of Lesson 1).

Lessons 2 and 3 included perception-based activities involved listening practice that focused on the placement of prominence and the direction of the tone choices in the utterances. For prominence, the learners were asked to (1) listen and mark the words in sentences that are perceived as being stressed and reflect on its appropriateness and its effect in understanding the pragmatic meaning of the utterances and (2) choose the request with more appropriate prominence. The sentences and dialogues that learners listened to represented the target speech act in the study. For tone choices, the exercises invited the learners to (1) listen and identify the overall shape of the pitch movement (rising, falling, or level) and interpret the appropriateness of a speaker's tone choices in a given speech act and (2) choose the request with more appropriate tone choices. Importantly, learners only distinguished between three main tone choices: rising, falling, and level tones. It has been suggested that complex configurations like '*fall-rises*' or '*low rises*' may confuse L2 English learners (Atoye, 2005); thus, it was decided to include only the main tone choice patterns in the instruction. Any utterances that included rise-fall or fall-rise intonation were described as fall or rise. Finally, to ensure that learners were ready to move from one perception activity to another, they only got access to the next exercise when they reached 90% accuracy in a current exercise. In order to control for this, the website monitored the accuracy of each learner as they were completing an exercise and provided immediate feedback asking to repeat the activity if the necessary score was not reached. Before repeating the task, the

learners' incorrect answers were displayed. Appendices D2 and D3 include descriptions of the perception-based activities).

As explained above, the production exercises in the instruction included controlled, guided, and communicative practice. These exercises were included in Lessons 4 and 5. In controlled production activities, learners practiced reading utterances that represented requests incorporating the prominence or tone choice patterns given in a model recording. Importantly, learners were asked to record their reading and analyze its spectrogram and pitch movement to check whether they were able to demonstrate the prominence and tone choices they were provided with. In this exercise, learners were able to see a typed request with marked tone choices and prominence. They practiced producing the same request and, once happy with their prosodic features, the learners recorded their reading. Once the recording was done and processed, they were shown its pitch contour. In a response box, learners reflected on the use of intonation in their recording with the help of the pitch contour. Guided practice exercises encompassed learners reading a request with the target prominence and tone choices based on a model sound file. In this activity, learners were asked to listen to a model request and observe its pitch contour. Then, they recorded their own request in real time and thus compared their production with the target using visualizations. For each production exercise, learners' recordings were submitted to the instructor via the learning platform. Only after the instructor had checked the recordings for tone choices and prominence were the learners allowed to move on to the next lesson. In the case that learners' production did not represent the target use of tone choices and prominence, the instructor asked learners to repeat the exercise until the target prosody was achieved. Appendices D4 and D5 detail the production-based lessons of the training.

Lastly, communicative practice of intonation patterns was provided to the learners through free role plays that imitate request-making with the researchers via Zoom, a video conferencing tool (Zoom Video Communications, 2021). Since one feature of requests was their high power difference and social distance, in each role play, the researcher took on the role of the professor or advisor (person in higher power). The role plays were based on the perception and production exercises that scaffolded learners' use of prominence and tone choices in this communicative activity.

In order to ensure that the proposed sequence of exercises worked for EFL learners, its essential parts were piloted with a group of 14 intermediate/high-intermediate learners. The pilot consisted of four activities that were completed by the learners and constituted the exercise base for the instructional platform. The first three activities tested learners' perceptual abilities and included (1) identification of tone choices in requests, (2) identification of prominent words in requests, and (3) discrimination between more and less appropriate use of tone choices and prominence in pairs of requests (see Figures 3-5). In each perception activity, the platform provided immediate correct/incorrect feedback (e.g., *Your response is incorrect. Try again!*) for each item and recorded the number of attempts it took each participant to reach the 90% accuracy to be able to move on to the next task. In the fourth activity, the learners were asked to record themselves reading a request with appropriate intonation after listening to a provided audio model and analyze the pitch contour of their recording that was displayed immediately after the recording was done (see Figure 6).

Figure 3

Interface of the Tone Choice Identification Activity

≡

Home

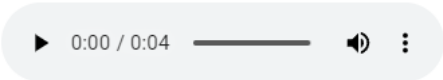
Hello, admin

Logout



Request 1.

Asking for a change in the schedule



What tone choice did you hear at the end of the request?

You've selected:

☐ Rising

☐ Falling

☐ Level

Figure 4

Interface of the Prominence Identification Activity

≡

Home

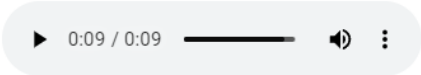
Hello, admin

Logout



Request 11.

Asking for a change in the schedule.



Choose prominent words in the request.

You've selected:HiIwayclasstakespringfall

Hi I was wondering if there is any way I

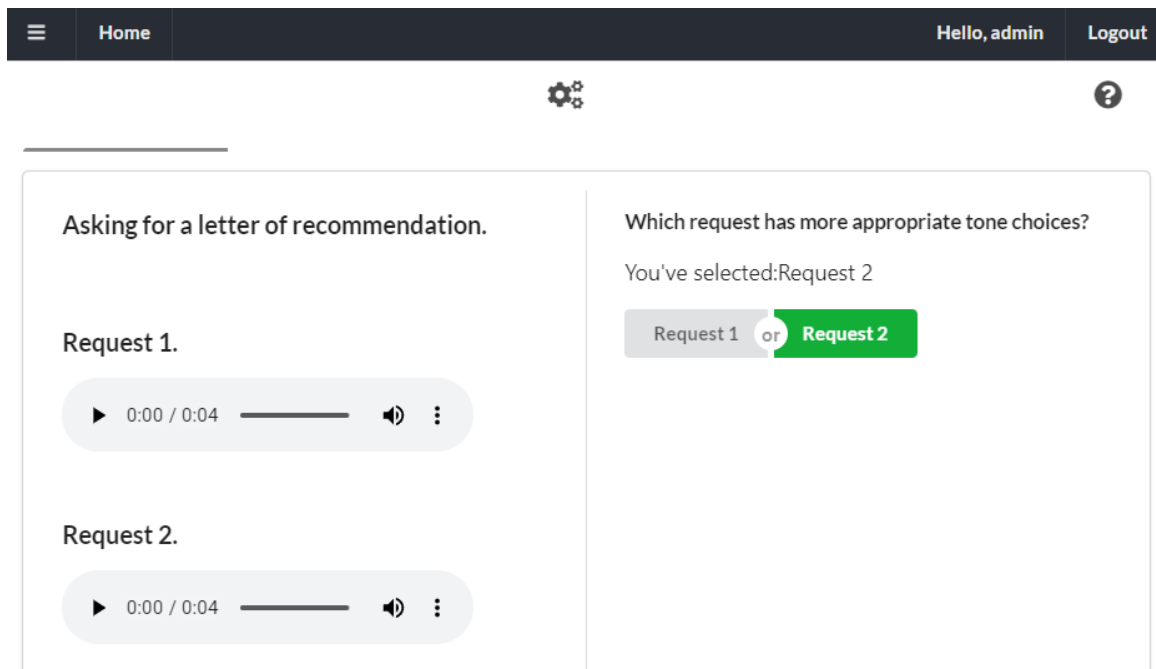
can take a class that I was going to take

originally in the spring semester during my

fall semester

Figure 5

Interface of the Intonation Discrimination Activity



Note. The interface for prominence discrimination was the same as for tone choice discrimination.

Figure 6

Interface of the Controlled Production Activity

Home

Hello, admin

Logout

I was wondering if you could write me a letter of recommendation.

0:03 / 0:03

Pitch Frequency

When you are ready, press "START" to record the request you just heard.

Pitch Frequency

PLAY

START

STOP

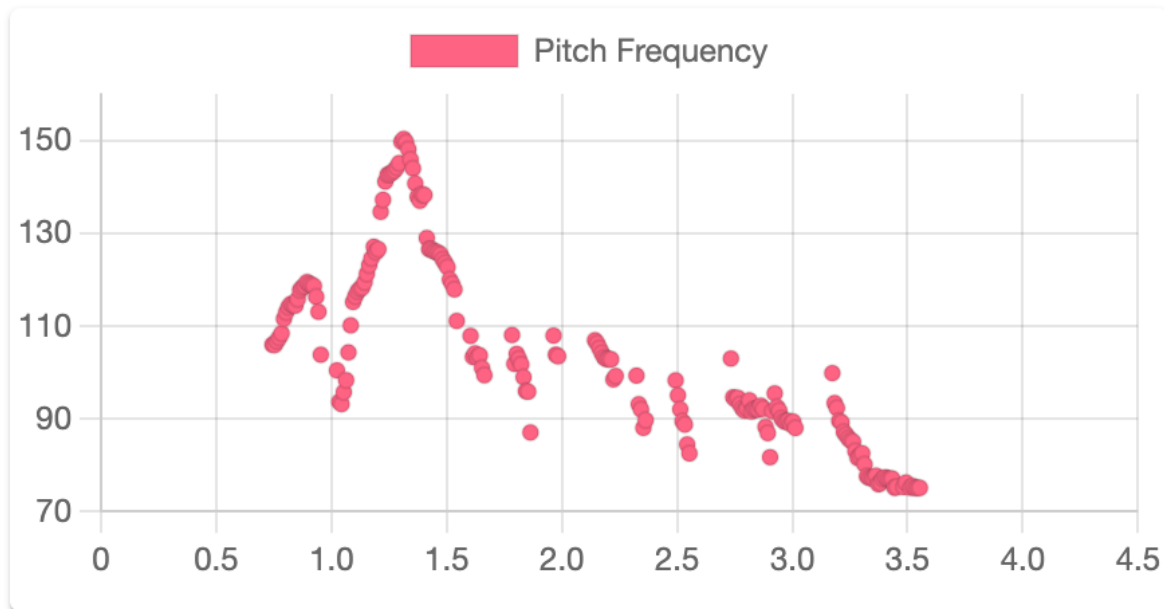
The results of the pilot showed that learners were able to differentiate between the different pitch movements from the first attempt most of the time ($M_{\text{attempt}} = 1.55$, $SD_{\text{attempt}} = .70$ meaning that it took them on average 1.55 attempts to give a correct response). They were also able to identify all of the prominent words in requests, although it usually took them longer compared to tone choices ($M_{\text{attempt}} = 2.96$, $SD_{\text{attempt}} = .26$). In addition, when comparing pairs of requests with the same lexico-grammatical content but different intonation contours, learners were mostly able to identify the requests that were more appropriate after reading a short explanation about the use of intonation features in requests. Specifically, requests with falling tones and pragmatically appropriate prominence were distinguished from the same speech acts pronounced with rising or level intonation and irregular prominence after the first attempt in the majority of cases ($M_{\text{attempt}} = 1.27$, $SD_{\text{attempt}} = .54$ and $M_{\text{attempt}} = 1.24$, $SD_{\text{attempt}} = .51$ for level and

rising intonation respectively). Acoustic analyses of the prominence and tone choices of learners' production in the last activity also showed that in approximately 60% of the cases, learners were able to follow the brief instructions in producing falling pitch contours and stressing only the contextually important words in a request. In the comments following the tasks, the participants noted that pitch contour visualizations were helpful in monitoring their use of intonation in producing requests.

Pitch Contour Visualization. All the exercises that involved learners prosodically analyzing either their own recordings or the model recordings were conducted with the help of Parselmouth and Chart.js, computer programs which had the same functionality as PRAAT but could be used in online settings. Figure 7 is an example of the Chart.js output that includes the pitch contour of a sample utterance. The vertical axis of the spectrogram presents the pitch height of the audio recording measured in Hz; the horizontal axis indicates the length of the audio file. Thus, the soundwave of the example utterance *Hi professor, I was wondering if you could write me a letter of recommendation* illustrates that the tone unit had falling intonation as the pitch on the termination moves down. During the first lesson of the instruction, the learners received training in how to extract, read, and interpret the soundwave and pitch contour of their own recordings and the models. They also used the Chart.js output to compare the prosody in their production with the model. It is important to note, however, that the intonation judgements about the sound files in the instruction were made not just from the pitch contour displays, but also based on the learners' auditory judgements derived from the actual listening of the audios.

Figure 7

Sample Parselmouth and Chart.js Output



Overall Sequence of Lessons. Table 2 provides an overview of the sequence of the lessons which make up the instructional training. As seen in the table, for each feature, the instruction first included perceptual training and then shifted towards production activities, leading ultimately to the final phase, the communicative practice. The complete training took on average 2 hours and 15 minutes and was spread over two weeks with three lessons a week. Based on the findings of the meta-analytic studies that investigated the role of the length of instruction in the success of an instructional training (e.g., McAndrews, 2019), a two-week training was deemed sufficient to see the changes in the learners' use of prominence and tone choices.

Table 2

Overview of the Pragma-Prosodic Instructional Training

Week	Day	Focus of instruction	Phase (Celce-Murcia <i>et al.</i> , 2010)	Duration
Week 1	Day 1	Introduction to prosody and politeness in requests	Definitions, introduction in intonation (<i>description and analysis</i>)	15 minutes
Module 1				
	Day 2	Prominence and tone choices in requests	Perception (<i>description and analysis, listening identification</i>)	20 minutes*
Week 1	Day 3	Prominence and tone choices in requests	Perception (<i>description and analysis, listening discrimination</i>)	20 minutes
Module 2				
	Day 4	Prominence and tone choices in requests	Production (<i>controlled practice, guided practice</i>)	30 minutes
Week 2	Day 5	Prominence and tone choices in requests	Production (<i>controlled practice, guided practice</i>)	30 minutes
Module 3				
Week 2	Day 6	Prominence and tone choices in requests	Production (<i>communicative practice</i>)	20 minutes

* Note. Depending on learners' performance on the tasks, this timing may change. Lessons in Module 2 were longer due to the production nature of the exercises.

The first lesson of the training focused on introducing the learners to the concept of intonation, particularly to its features of prominence and tone choices, as well as the important notions of the speech act theory and politeness theory (e.g., defining requests, power distance, imposition, etc.). The lesson also introduced learners to pitch contour imaging, delivered to students on the instruction website via Parselmouth, a library that adapts PRAAT to Python

(Jadoul, 2020), and Chart.js, an open-source library for data visualization, and its elements (Chart.js Team, 2021). The next module covered the perceptual training of prominence and tone choices in requests. The first lesson in the module (Day 2) included perception-centered exercises for requests. The following two-lesson module had the same structure as Module 1, but its focus was shifted onto controlled and guided production-based exercises. Finally, in order to help students produce different prosodic contours naturally, the last lesson in the instruction (Module 3, Day 6) centered around a role-play session with the researcher, that simulated the communicative situations of high imposition requests (Chun, 2002; Derwing *et al.*, 1998). The last lesson took place on Zoom after learners completed all the exercises in the online training. Each learner met with the researcher individually, and the role-play scenario was given to a learner during the session. Each learner had 2 minutes to read the scenario and plan their responses. The lessons were spaced out with two days between each of them to maintain a balanced pace of the instruction. Crucially, the pre-test happened one day before the instruction, and the post-test took place one day after the instruction (see the Timeline section). A delayed post-test was taken by some of the learners one week after the instruction. Appendix D contains the detailed lesson plans for each day of the training.

Pre-Test and Post-Test Materials

To collect the participants' production of requests before and after the training (see Figure 2), the study used six video-based multiple-turn DCTs. Five of the DCTs were adapted from Kostromitina (in press) and one additional DCT was developed by the researcher. The pre- and post-test consisted of three DCTs that were chosen from a pool to avoid the practice effect from using the same tasks at three different times. These DCTs represented high-imposition (R+) requests. Other situational characteristics in the DCTs included a greater social distance between

the interlocutors and an unequal power relationship. Each DCT consisted of 6 turns: three turns from the interlocutor and three response turns that the learner had to produce. Importantly, to control for the pragmalinguistic aspects of the production, learners watched a video explaining what pragmatically appropriate request forms they should use before they took the pre-test. The video introduced the concept of appropriateness as well as the bi-clausal target pragmatic structure used in the training (*I was wondering if...* / *I was hoping that...*). This step was crucial to ensure that the learners produced lexico-grammatically comparable responses. It was also important that the main strategies were consistent across learners for the subsequent perceptual evaluation of the prosodic composition of learners' requests.

Although it has been widely recognized that DCTs are an effective and practical method of speech act data collection, some have expressed concerns about the authenticity of the pragmatic production elicited with DCTs. Additionally, this method has been criticized for its lack of dynamism that is characteristic of real-life interaction since participants often have a lot of extra time to decide on the response (see Taguchi & Roever, 2017 for a detailed discussion). To alleviate these limitations, Kostromitina (in press) created a multiple-turn video-based DCT task. This instrument allowed for collection of speech samples that could potentially be more similar to natural conversation involving turn-taking, especially because participants were not able to read all the turns before responding as in a written DCT. Table 3 presents a detailed description of the DCT situations, and Appendix E provides the DCT scripts.

Table 3

Summary of the DCT situations

Requests (PDR+)
Asking for a letter of recommendation (student – professor)

Pre-test & Delayed post-test	Requesting for a day off (student – employer)
	Changing class registration (student – advisor)
Post-test	Asking for an extension (student – professor)
	Transferring class credits from another university (student – advisor)
	Rescheduling a shift (student – employer)

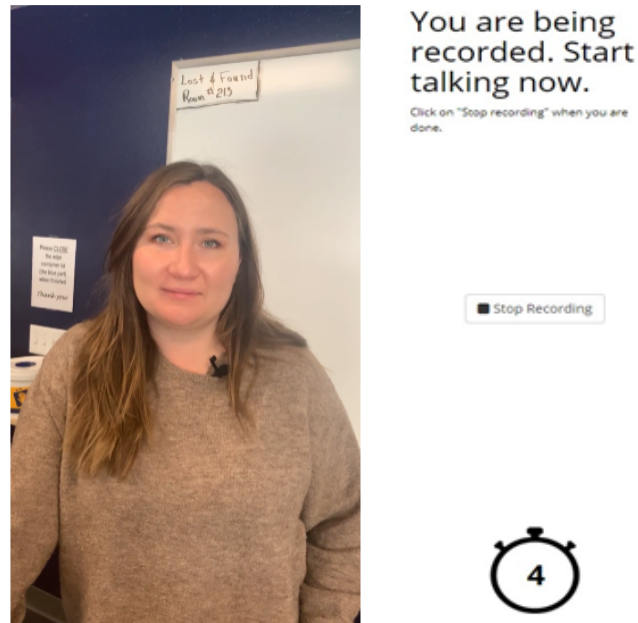
The situational scenarios were partially adopted from previously validated DCTs in studies that investigated the production of requests (Taguchi, 2012). As one can notice from Table 3, the target DCT situations revolved around life on campus where the interlocutors were professors, academic advisors, and on-campus employers (greater social distance and power difference, D+, P+). The high imposition (R+) of the speech acts (e.g., asking for a recommendation) is implied due to the amount of effort needed from an interlocutor to accommodate the request. In all the role plays, the learners performed the student role.

The videos of the interlocutors turns were recorded by four speakers with college education background, two male and two female. Three of them were English NSs, and one was a highly proficient NNS of English who has lived in the United States for the past 5 years and did not possess a noticeable foreign accent. The video-recorded turns were then compiled into DCTs using the online data collection software *Gorilla* (Anwyl-Irvine *et al.*, 2019). This software allowed the participants to watch the videos of each turn and immediately audio-record their responses to the utterances they heard. Figure 8 presents the interface of the software with a screenshot of the situation instructions (top), a screenshot of the presentation of the video turn (bottom left), and the screen that a participant saw to record their response (bottom right). Participants were given 30 seconds to record each turn in a DCT. The timer on the recording screen was used to keep the participants within the given time frame.

Figure 8

Interface of the Pre- and Post-Tests

Situation 1. Asking for a letter of recommendation. Instructions: You want to apply for a scholarship offered by your department for the next semester. The applications are due in a month. To apply for the scholarship, you need a letter of recommendation from one of your professors. You want to ask Professor Smith to write the letter of recommendation for you. What do you say to Professor Smith?



Importantly, as a part of the post-test, learners completed a final questionnaire (Appendix I) to reflect on their progress in pragma-prosodic learning and share their thoughts on the Intonation Trainer.

Aptitude Measures

To investigate the role of the learners' phonological working memory, musical ability, and auditory processing in the effectiveness of the instructional training, these individual differences were measured for each learner at Time 2 (after the training). The following section presents the three tasks that were used for this purpose. It is important to note that each task was previously

validated by studies that investigated the effect of each of the aptitude factors in focus on L2 acquisition (Law & Zentner, 2012; Saito *et al.*, 2020; Shin *et al.*, 2019). The tasks were introduced to the learners sequentially via the *Gorilla* platform after the production post-test (see Section 3.3).

Working Memory (WM). Since WM is thought to include two components, storage and processing (e.g., Baddeley, 2003), the complex tasks that are used to measure it usually impose demands on its storage and processing simultaneously (Colom *et al.*, 2006). In SLA research, two types of tasks have been used to capture the complex nature of WM: non-linguistic and language-based tasks. Most frequently, studies have included digit span tasks to measure WM (e.g., Brunfaut & Revesz, 2015; Kormos & Sáfár, 2008). In this task, participants are usually asked to remember an increasing sequence of numbers presented visually or aurally and reproduce the sequence in order. However, such tasks typically only measure the storage capacity of the WM and not its processing aspect and are thus not suitable for the use in the present study unless used in combination with other non-linguistic measures, such as an automated operation span or a non-word recognition task, for example (e.g., Zalbidea & Sanz, 2020).

In contrast, complex language-based span tasks can provide a comprehensive measure of WM since they require both processing and storage of information (Lesaux *et al.*, 2006; Waters & Caplan, 1996). One of the first language-based span tasks was developed by Daneman and Carpenter (1980) and included sequences of sentences that participants had to read out loud and then recall the last word in each sentence. Thus, the reading part of the task tested the processing component of WM and the recall part made demands on the storage aspect. Improving the procedure in Daneman and Carpenter (1980) to include the processing aspect of WM, Shin *et al.*

(2019) asked their participants to listen to the sentences compiled into 14 trials, judge their semantic acceptability, and remember the last word in each sentence and consequently type all the words they remembered from the sentences.

While reading span tasks as such have been commonly used to measure WM, especially in studies that focused on L2 reading (e.g., Alptekin & Erçetin, 2010; Miyake & Friedman, 1998), adjusting the delivery mode of the task to listening may be more suitable for studies that center around listening and speaking skills, like the present one. This helps to avoid confusion among the participants who need to switch between the different modalities in the experiments and to reduce the possibility of putting additional cognitive load on the participants' WM by having them decode the visually presented stimuli (Sedivy, 2014). Moreover, a listening span task can be expected to share more variance with oral language use (perception and production) and, therefore, be more predictive of performance in the oral/aural modality.

Given the proximity of listening span tasks to oral language use, the focus of this study, the task that measured the participants' phonological working memory (PWM) in the present study was an adaptation of Shin *et al.*'s (2019) reading span task with the stimuli presented aurally (see also Yaw, 2022). The task consisted of 70 stimuli of 10-12 words long grouped into sets of 2-8 sentences. It included two trials of each length with a total of 14 trials that were presented in a set order. In each trial, participants had to listen to a sentence, decide if it is semantically acceptable, and remember the last word in the sentence. After that, in parallel to Shin *et al.* (2019), participants were asked to type as many words they remembered as possible. The stimuli in the task were controlled for lexical difficulty corresponding to the lower intermediate proficiency level and the final word in each sentence did not exceed three syllables. Following Shin *et al.* (2019), the task was administered in English, the target L2 in the study as it

was shown to reveal a stronger relationship between WM and L2 acquisition (Shin, 2020). In terms of scoring, although two indices were recorded for each participant, only one of them was used in the study. That is, both the number of correctly recalled words and response times (RTs) in milliseconds for each semantic judgement were documented. The first index measured participants' PWM storage capacity, and the second one reflected their PWM processing capacity. However, only the first index was included in the analysis. This decision was based on previous research showing that PWM storage capacity in particular has significantly predicted intonation learning (e.g., Ahmadian, 2020; Rota & Reiterer, 2009). Moreover, the two indices were found to be highly correlated during the statistical analyses in the study, a composite PWM variable was created ($r = 0.82$). The complete list of stimuli for this measure can be found in Appendix H.

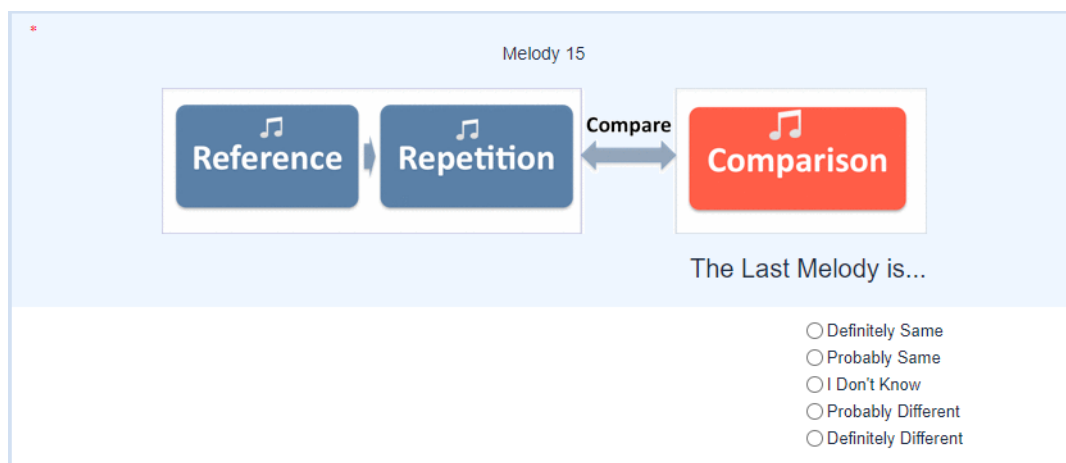
Musical Ability. Typically, musical ability is viewed as a multi-faceted construct that is measured by participants' perceptual sensitivity to timbre, tuning, melody, pitch, rhythm, and tempo (Law & Zentner, 2012). Since each of these acoustic properties has been associated with different properties of speech, studies in L2 pronunciation have employed several musical ability tests to capture the specific aspects of musical ability that were relevant to their research questions. Among the perceptual studies, Delogu *et al.* (2010) used one of the subtests of the Standardized Test of Musical Intelligence (Wing, 1948) to examine the connection between the listeners' melodic ability and their phonological and lexical tone processing of Chinese words. The test included 30 pairs of simple melodies. For each pair, participants had to indicate whether the melodies were similar or different and mark the position of the difference. Isaacs and Trofimovich (2010), in investigating the biasing effects of listeners' musical ability on their L2 pronunciation ratings, used the melody, tempo, and phrasing subtests of the Musical Aptitude

Profile (MAP), a standardized musical ability test (Gordon, 1995, 2001). The three subsets asked the participants to compare the melodic contour of musical statements, indicate whether an established tempo was maintained, and choose a more pleasant rendition of the same musical phrase. Whereas the melody and tempo subsets were indicative of the tonal and rhythmic aspects of musical ability, the phrasing subset targeted the participants' musical sensitivity. In a more recent intonation training study, however, Yuan *et al.* (2018) employed the melody and pitch subsets of the PROMS musical ability test (Law & Zentner, 2012). The authors argued that these two modalities were mostly related to intonation in speech. Each subset in the study consisted of 18 items in which the participants had to compare whether a pair of musical stimuli had the same melody or rhythm.

After careful consideration of the existing tests of musical ability employed in previous empirical studies, the melody and accent subsets of PROMS were chosen for the current study. These two sub-tests matched the prosodic features of interest in the current study with the melody section coinciding with the participants' tonal sensitivity, and the accent subset relating to the learners' ability to perceive prominence. Additionally, with regards to practicality, the PROMS subsets were found advantageous for several reasons. First, each subset has a similar structure; participants are presented with the reference stimulus that is played twice and are then asked to decide if the comparison stimulus matches the reference. Figure 9 provides an example screen with one of the melody trials.

Figure 9

Sample Trial of the PROMS Melody Subset (Law & Zentner, 2012).



Each subset consisted of 18 trials making the completion of the whole battery of musical ability tests take less time in comparison to other tests (e.g., the MAP test takes 3.5 hours). Another benefit of the PROMS test was the continuous nature of scores for each of the trials. Instead of the binary yes/no question, the participants were asked to choose one of the five Likert scale options, which helps to achieve a more fine-grained representation of the learners' musical ability. Finally, the Innsbruck University laboratory offers the stimuli from the test free of charge for the researchers. Thus, after contacting the laboratory, the stimuli were downloaded by the researcher, incorporated into the *Gorilla* interface for the future presentation to the participants. The program recorded the participants' responses that were transformed into scores. In accordance with Law and Zentner's (2012) procedure, a correct response chosen with maximum confidence ("definitely" same or different) were awarded 1 point. Correct responses chosen with less confidence ("probably" same or different) were awarded 0.5 points. Similarly to the PWM measure, the scores on the melody and accent sections were merged into a composite score due to a strong correlations between the two variables ($r = 0.87$).

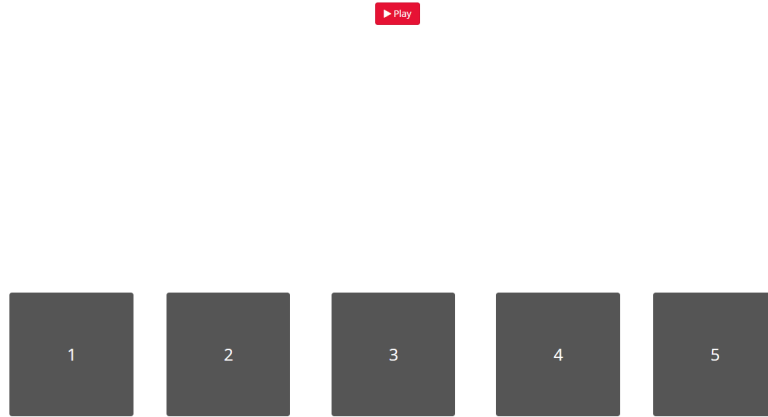
Auditory Processing. The previous chapter illustrates that research on auditory processing and its role in L2 intonation acquisition is emerging and, thus, very limited. While

some behavioral auditory processing measures have been created in the psychological domain (e.g., Moore, 2012), Kazuya Saito and his collaborators recently developed a methodologically similar test for the field of SLA (Saito *et al.*, 2020a, b). The complete test includes four discrimination tasks that measure participants' perceptions of duration, amplitude rise time, pitch, and formant, as well as two reproduction tasks focusing on participants' ability to reproduce a given rhythm and melody. The discrimination tasks consist of 100 stimuli that differ in formant frequencies, pitch, amplitude rise time, and duration. During each trial, participants are presented with three stimuli where either the first or the third sound is different from the other two. They then have to indicate which sound is different. Importantly, the difficulty of the task adapted to the participants' performance; that is, it decreased after every incorrect response. In reproduction tasks, participants are asked to listen, remember, and repeat 10 melodic and 30 rhythmic patterns. The latest validation research on the test has provided evidence that it is able to accurately measure the targeted construct and revealed a medium-to-large amount of variance in L2 production skills, particularly regarding suprasegmentals, that was explained by auditory processing (Saito *et al.*, 2020b; Sun *et al.*, 2021; Tierney *et al.*, 2017).

Based on the results of Sun *et al.* (2021) discussed above, the current study adopted only the musical reproduction task as the measure of auditory processing as this measure was found to be associated with participants' L2 suprasegmental gains. Moreover, learners' perceptual processing were already assessed by the musical ability test. As mentioned above, the task consisted of 10 melodies that were constructed from a set of five 300 ms long tones. Each melody always started with the third tone, and the rest of the tones followed randomly. The task was presented to the participants via Gorilla, similarly to the other aptitude measures, as shown in Figure 10.

Figure 10

Auditory Processing Task Interface



Each melody was played three times with a 1-second interval in between the repetitions. The participants were then asked to reproduce the melody by clicking on the five buttons they saw on the screen (see Figure 10). Each button corresponded to one of the tones, organized in ascending order (higher tones closer to the right side of the screen). Every time it was pressed, a corresponding tone played. Before the main trials, participants were given a training item and a chance to practice pressing each button to acquaint themselves with the five pitches they produce. The learners' scores were calculated based on their accuracy in repeating each melody. In other words, each melody produced by a participant was compared to the model and a total number of matching tones was calculated.

Timeline of Experimental Group Data Collection

Since the production activities in the instruction was monitored by the researcher, it was expected that learners would complete the training at a roughly similar pace. The instruction started for each learner one day after completing the pre-test. After that, learners got a link to the training website as well as a schedule of the days they were required to complete it. Each module

was spaced out with two days in between with the post-test happening on the next day after the last module was finished. Table 4 shows an example timeline for training completion. The activities for each module in the experimental group training are given in Appendix D.

Table 4

Example Completion Timeline for the Experimental Group

Activity	Day of completion
Pre-test & aptitude measures	Monday, April 4, 2023
Module 1	Tuesday, April 5, 2023
Module 2	Thursday, April 7, 2023
Module 3	Saturday, April 9, 2023
Module 4	Tuesday, April 12, 2023
Module 5	Thursday, April 14, 2023
Module 6	Saturday, April 16, 2023
Post-test	Sunday, April 17, 2023

Elicited Imitation Task.

The study used an elicited imitation task (EIT) in order to control for learners' proficiency in measuring the effectiveness of the developed training. This independent measure of general language proficiency has been widely used in academic and research contexts and has demonstrated considerable concurrent and predictive validity (Kostromitina & Plonsky, 2022; Yan *et al.*, 2016). The typical structure of this task includes a listener hearing a sentence and repeating it as accurately as possible either after a pause or after a comprehension question to prevent parroting. The particular task used in the study was adopted from Yan (2020) and included 24 sentences that were assembled in three blocks based on their length as well as syntactic, phonological, and lexical complexity. All items were recorded by an L1 English speaker. The blocks were presented to the participants in the same order; however, the stimuli within each block were randomized with each trial. Before the actual test, listeners were

provided with instructions and a trial item. A list of items used in the EIT is presented in Appendix C.

Raters' Perception Scales

Measured Constructs

Appropriateness. Despite the inconsistencies in defining appropriateness, it appears from the review of the theoretical frameworks and empirical research above that appropriateness is operationalized through several pervasive sub-constructs; namely, context-specific *politeness*, *directness*, *formality*, *sincerity*, *naturalness*, *grammaticality*, *use of pragmatic strategies*, and *discourse organization*. Of particular interest for this study are the first five sub-constructs since the lexico-grammatical variables in learners' production of requests was not part of the instructional focus of the study. These sub-constructs were adapted in the perception scale from the domain of pragmatics. With regards to the independence of the five sub-constructs of interest, the results of the pilot study showed strong significant correlations over $r = 0.8$ between naturalness, sincerity, and politeness ratings. Thus, for the purposes of the dissertation, three listener perception pragmatic-related variables were submerged into three sub-constructs representing appropriateness: politeness, directness, and formality.

The following three constructs were referred to in the listener perception scale in the study:

- *Politeness* – degree of speakers' effort to mitigate the threat to the listener's face in performing a speech act
- *Directness* – degree of face-threat in speakers' use of prosody in performing a speech act
- *Formality* – degree of speakers' observance of the social context of the speech act (power, distance, imposition).

Accentedness. As expounded above, the study applied Derwing and Munro's (2015) definition of accentedness that utilizes the degree of difference between L2 speech and L1 speech as perceived by listeners. This construct was measured via two scales: strength of accent and difference from L1 speech).

Comprehensibility. To operationalize comprehensibility, the study is also guided by Derwing and Munro (2015) who define this construct as the amount of effort needed from a listener to understand a speaker. The construct was represented through two scales: clarity and difficulty of understanding.

Scale Functioning

Performance-based evaluation that involves a rating scale is a common approach to assessing pragmatic competence (e.g., Plonsky & Zhuang, 2019; Taguchi & Li, 2020) and eliciting listeners' impressionistic judgements of speech characteristics (Kang & Kermad, 2017). In speech perception research, studies have validated five-point scales (Kermad, 2018; Isaacs & Thomson, 2013), seven-point scales (Derwing *et al.*, 2008; Southwood & Flege, 1999), and nine-point scales (Munro, 2017) to register listeners' perceptions of accentedness and comprehensibility. In an example study, Derwing *et al.* (2004) collected speech samples from 20 beginner L2 English learners from China and asked 28 untrained judges to rate fluency comprehensibility, accentedness, and "goodness of prosody" in the samples using a 9-point scale. The researchers did not use elaborate descriptors for the scale the reason being that the raters were naïve listeners and, as such, would be overwhelmed with descriptors for each point. Thus, only the ends of each scale were defined.

In their turn, studies in the area of L2 pragmatics have most commonly made use of a 5-point Likert scale to evaluate pragmatic appropriateness. For example, Taguchi (2011a) evaluated the appropriateness of requests and opinions performed by EFL learners by using a 5-point scale with the range of scores from 1 to 5. In contrast to Derwing *et al.* (2004), the study employed only four raters that were provided descriptors for each point on the scale. These descriptors wholistically evaluated appropriateness based on the perceived level of directness, politeness, and formality. Like Taguchi (2011a), Li and Taguchi (2014) used 5-points Likert scales to examine the effects of different levels of L2 proficiency on the development of pragmatically appropriate requests in L2 Chinese study abroad. The holistic scale used in that study covered several criteria within appropriateness, such as realization of the target communicative function, scenario context awareness, and absence of lexico-grammatical errors. While 5-point scales have been predominant in this type of research, some studies have employed six-point scales, adding a zero point on the negative end of the scale for the absence of performance (e.g., Al Masaeed, 2020; Taguchi, 2013).

Although rating scales have been common in both pragmatics and speech perception research, one must consider several issues in choosing the most fitting scale for a study. For one, Barkaoui (2010) emphasized that holistic and analytic scales may have raters attend to different aspects. The results of Barkaoui's study, which involved an analysis of raters' think-aloud protocols collected during evaluation of ESL essays, showed that when using an analytic scale, raters paid more attention to the evaluation criteria themselves, while with the holistic scale raters' focus was more on the target of evaluation. Others have observed variations in rating scale functioning due to the level of learners' performance (Yan, 2014). For example, Grabowski (2013) investigated the construct validity of a speaking role-play test that measured grammatical

and pragmatic knowledge across multiple proficiency levels. The participants' performance on the test was measured with a 5-point analytic rubric that incorporated grammatical knowledge constructs as well as pragmatic knowledge (e.g., sociolinguistic, sociocultural, and psychological appropriateness). Grabowski found that while the use of all points of the scale was effective for the intermediate participant group, there was disordering in some categories, such as sociolinguistic appropriateness, in the advanced group.

Within the speech act framework, Jianda (2006) investigated the functioning of a five-category holistic scale for assessment of L2 English learners' production of apologies and requests in a researcher-made written discourse completion task (DCT). The scale was designed to measure four dimensions: learners' ability to produce the target speech act, use of formulaic expressions, amount of produced speech, and level of appropriateness (formality, directness, and politeness). The study showed that the scale functioned as designed for both speech acts; however, as Li *et al.* (2019) note, the scale might have had too many rating categories for evaluating requests. In their own evaluation of rating scales in pragmatics, Li *et al.* (2019) followed Linacre's (1999, 2002) guidelines to assess a six-point rating scale to see if it functioned differently across raters, speech acts, and L2 proficiency levels. L2 Chinese learners completed computerized oral DCTs that elicited requests, refusals, and compliment responses. Their performance was rated by two L1 Chinese raters using a holistic rating scale that tapped into learners' ability to produce speech acts that were contextually and situationally appropriate as well as grammatical. An important finding that emerged was a suggestion to merge 0 and 1 points on the scale (*No performance* and *Poor performance*) to avoid redundancy as these two categories reflected the same level of pragmatic competence. Additionally, results indicated variance in the scores even among raters with similar backgrounds re-emphasizing the need for

scales with sub-descriptors for specific pragmatic features as well as rater training (see also Youn, 2015).

Studies in speech perception have also investigated the functionality of rating scales, particularly as it pertains to their design. Thus, Isaacs and Thomson (2013) examined the effects of scale length on judgements of L2 comprehensibility, accentedness, and fluency. While no differences in the mean scores on the constructs were detected between the 5- and 9-point scales, Rasch probability plots revealed category overlap, particularly with the 9-point scales suggesting that raters were not able to meaningfully differentiate between the points on the longer scale. A parallel observation was made by Isbell (2017) in his analysis of a 9-point ratings scale for the measurement of comprehensibility and accentedness in Korean learners of English. In the study, ten raters underwent brief pre-training that included an introduction to the speaker tasks used in the data collection, an explanation of the rated concepts, and instructions for using the scale. Raters were also given four samples (two NS and two NNS) to rate as practice. After a short discussion of the practice scores, the listeners completed the rating live. Listeners' post-rating comments uncovered several challenges in the measurement of comprehensibility and accentedness using a 9-point scale. First, analogous to the findings in Isaacs and Thomson (2013), raters reported that the scale was too long, and they could not tell the difference between some of its points (e.g., 2 and 3 out of 9). On top of that, because raters did not have a linguistics background and received only minimal training, they reported difficulty interpreting the measured constructs. This observation was reinforced by a few raters' comments about insufficient training and lack of descriptors for each point of the scale. In a more recent statistically oriented study, Kermad and Bogorevich (2022) explored scale functioning via comparing listener consistency and severity in L2 accentedness and comprehensibility ratings

given on a 9-point Likert scale to the same ratings on transformed 7-point and 5-point scale. The study suggested that the shorter 5-point scale may be a more reliable way to measure listener perceptions leading to more consistent and precise ratings.

The issue of minimally explicated scales (only endpoints are anchored), together with lack of rater training, was also raised in Yan and Ginther (2017), who pointed out that minimal scale explication and training can be attractive as they can provide an understanding of a general listener's perceptions of speech constructs. At the same time, though, the authors noted that inexperienced or naïve listeners may have trouble in the application of scales without enough training on top of being subject to rater bias, discussed above.

Scales Used in the Study

Taking into consideration the existing practices and issues raised in the research on rating scale functioning in the domains of pragmatics and linguistics, the dissertation study made use of six 5-point holistic scales with different degrees of descriptors for each point: two assessing pragmatic appropriateness of learners' production and the other four assessing their accentedness and comprehensibility. The scales used in the dissertation study followed several design principles. First, to alleviate the issues of category overlap, scales consisted of five points ranging from 1 (the lowest score) to 5 (the highest score). This scale length was found to provide reliable ratings and at the same time be short enough to not confuse the raters (Isaacs & Thomson, 2013; Li *et al.*, 2019). Second, the scales provided a holistic judgment of the measured constructs since this type of assessment, in comparison to analytic assessment, had been shown to help raters focus on the target of evaluation itself rather than the rating criteria (e.g., Barkaoui, 2010).

Importantly, the six scales in the study were designed somewhat differently in terms of the amount of description in each of them. The two appropriateness scales included definitions for each point and not just the endpoints. As argued by Li *et al.* (2019) and Youn (2015), providing such definitions may alleviate the discrepancies in the rater scores. One of the two scales measured specifically the appropriateness of lexico-grammatical contents of requests produced by the learners. The other scale assessing appropriateness focused on the prosodic composition of requests. As mentioned before, this step of splitting appropriateness into two subscales (lexico-grammatical and prosodic) in order to reduce the possibility that raters are affected by learners' use of lexico-grammatical features in assessing their use of tone choices and prominence. The two appropriateness scales were the primary focus of the study. In terms of the operationalization of the constructs measured by the scales, the two scales were based on the definition of pragmatic appropriateness derived from previous research (see e.g., Ishihara, 2013; Sydorenko *et al.*, 2014; Taguchi, 2011a and the discussion above) and include the notions of situationally and contextually dependent politeness, directness, and formality, adapting Taguchi's (2011a) descriptors (see Appendix F). The specific definitions of politeness, directness, and formality used in the scale point descriptions were introduced to the raters during training (see the following section). Crucially, only the scale measuring the *prosodic* aspects of appropriateness was included in the analyses since the learners only received instruction on prosodic elements of requests.

As for accentedness and comprehensibility, the constructs were also measured by four five-point scales; however, only endpoints had descriptions on these scales, following Derwing *et al.* (2004). This design was deemed appropriate since the two constructs were measured to control for potential interference effects of listeners' perceptions of accentedness and

comprehensibility on appropriateness perceptions rather than the main focus of the study. As for the operationalization of accentedness and comprehensibility, each of the two constructs were reflected in two sub-constructs using widely accepted definitions in Derwing and Munro (2015) from above. All six scales are presented in Appendix F.

Rater Training

The development of the rater training session, as well as the rating part of the dissertation itself, involved several decisions about the duration of the training and its contents. Regarding the duration and organization of the training, since studies have reported significant improvements in raters' consistency and reduction in the effect of rater background variables on their evaluations after training (e.g., Kang *et al.*, 2019; Tajeddin & Alemi, 2014), raters underwent one extended training session (approximately 30 minutes long). The training was presented in the form of an instructional video and incorporated Construct Introduction, Exposure, and Calibration stages. Appendix G presents a detailed outline of the rater training session. The chosen structure of the training had been applied in both L2 pragmatic assessment (e.g., Chen, 2021) and L2 speech perception studies (Isbell, 2017; Kermad, 2018). The first part introduced raters to the operationalizations of the constructs of appropriateness, accentedness, and comprehensibility together with their measurement scales. It also explained the nature of role plays used in the pre- and post-tests and demonstrated one of the role-plays. The exposure stage provided listeners with example speech files who fall within certain scores on the appropriateness and accentedness scales. Four sound files were presented:

- A highly appropriate request with a prominent accent
- A highly appropriate request with a mild accent

- An inappropriate request with a prominent accent
- An inappropriate request with a mild accent

The Calibration stage was incorporated as the first part of the actual rating surveys. At the beginning of the survey, each rater was asked to assess ten files using the scales that had been introduced to them. Each rater assessed the same audio files, and their ratings were compared between each other. Fleiss Kappa using these ratings was calculated to estimate agreement between the raters. Once the calibration was completed, raters were taken to the next section of the Qualtrics survey with the actual sound files. The Data Collection Procedures section below details the calculations of raters' inter-rater reliability and the composition of the rating survey.

Control Group Instructional Content and Its Delivery

Apart from the experimental group, the study included a control group in order to draw conclusions about the effectiveness of pragma-prosodic instruction. The inclusion of a control group in the study design has been widely recognized in instructional studies. Plonsky and Zhuang (2019) reported that the majority of the 50 studies in their meta-analysis of pragmatic instruction had a control group. Moreover, the effect size of instruction was larger in the between-group designs (meta-analytic Cohen's $d = 1.53$ for studies with a control group and $d = 1.29$ for studies without a control group). In pronunciation instruction, control groups have also been widely used and managed with the help of non-pronunciation specific instruction (NPS) (e.g., Derwing *et al.*, 1998; Gordon & Darcy, 2013; McAndrews, 2021; Satio & Saito, 2017; Zhang & Yuan, 2020). In Gordon and Darcy (2013), for instance, control group participants engaged in listening and speaking practice following Celce-Murcia's (2010) framework but did not receive pronunciation instruction. In a similar manner, Zhang and Yuan (2020) had their

control group participate in their normal English listening, speaking, reading, and writing classes each week of the intervention. In an examination of effectiveness of prosody instruction, McAndrews (2021) provided the control group with product-oriented listening activities organized around TED talks that were related to the themes in an EAP course participants were taking. Participants first completed pre-listening activities, then they listened to a TED talk twice answering comprehension questions. These activities were followed by a discussion of correct answers. Finally, in Saito and Saito (2017)'s analysis of suprasegmental-based instructional effects on comprehensibility and prosody development in Japanese learners of English, control group participants were also given meaning-oriented instruction without a focus on suprasegmentals. The lessons were comparable in terms of duration to those in the experimental group. In addition, while the lessons for the control group differed in terms of the pragma-prosodic focus from those in the experimental training (i.e., perception and production of requests), they were similar in the covered top-level skills (listening and speaking).

It appears that control groups in the studies within the pragma-prosodic domain generally receive instruction that resembles the activities in the control group with the exception of the instruction targets. In other words, while this group typically completes activities related to the overall L2 skill in focus, they do not receive any instruction on the variables of interest for the study. In accordance with this general idea and following McAndrews (2021), the activities chosen for the control group consisted of learners watching short TED talks and answering comprehension questions about each talk. At the end of the instruction, participants in this group were asked to prepare a short monologue based on one of the TED talks and send it to the researcher for feedback. Because learners were recruited among English teaching majors, this type of activity incorporated the skills taught in typical EFL class and involved speech

perception and production practice, similar to the experimental group instruction. There were five TED talks total. Moreover, given the length of each TED talk (approximately 10-15 minutes), and the time it would take participants to answer comprehension questions (10 minutes), the length of each session matched the length of a module in the experimental group. The complete sequence of activities for the control group is given in Appendix J.

All activities for the control group were delivered to participants via the same website as the experimental group. Once participants log in using their name, the website identified them as belonging to the control group and took them to appropriate page. There was a total of six lessons for this group. The first five lessons included a video of one of the TED talks, which learners could watch twice. After that, participants were offered six open-ended questions. The first five questions checked their comprehension of the video and the final question focused on the speaking style of the orator (i.e., *What, in your opinion, was effective in the way the speaker was communicating with the audience? What was not effective?*). The last (sixth) lesson asked participants to record a monologue summarizing their favorite TED talk they watched. The list of the talks was presented in the lesson in case participants wanted to watch it again before recording. The recording was submitted to the researcher via the instruction website. The timeline of the control group completing the activities closely followed the schedule for the experimental group.

Data Collection Procedures

The collection of learner and rater data occurred in three steps outlined in Table 5. First, the experimental and control group learners as well as the baseline speakers (see section 3.2) were given a Gorilla link to complete the pre-test and the EIT (experimental and control group learners only). All participants completed the pre-test on their own, using their computers and

headsets. To check the quality of the recording, the researcher asked the participants to send a sample recording via email prior to data collection. The researcher was able to monitor the completion of the pre-test via Gorilla, which offers live updates on the status of the participants' responses. In terms of the recordings, the platform saved each turn in the DCTs as a separate audio file. The pre-test was prefaced by a short background questionnaire. The instructional training started on the following day and lasted for two weeks. Immediately after the instruction was over, all participant groups were administered a post-test on Gorilla. The experimental group also took the PWM, musical aptitude, and auditory processing tasks online via Gorilla. As noted in the Participants section of this dissertation, a sub-group of participants also completed a delayed post-test one week after the instruction. These were the participants from one of the two universities that took part in the study independently of a certain class and thus were not constrained in time. In the second step, the researcher acoustically and pragmatically analyzed the collected pre- and post-tests with an RA who additionally analyzed 10% of the collected data. Finally, after the recordings were analyzed, they were compiled in Qualtrics surveys and distributed to the raters who were pre-selected based on their responses on a background questionnaire.

Table 5

Data Collection Steps

Step	Data collection procedures
1	Pre-test, aptitude tests, beginning of training
1	End of training, post-test, delayed post-test
2	Acoustic analysis of learners' pre and post-tests
3	Preparation of the survey for the raters
3	Completion of the survey by the raters

Preparation of Sound Files for Rater Assessment

After the learner data collection was complete, the researcher prepared the pragma-prosodic appropriateness survey that was given to twenty trained raters. The raters had to evaluate a total of 538 sound files of approximately 18.1 seconds on average collected during the pre-, post-, and delayed post-tests (this number reflects the final audio file count after certain files were taken out for being incomplete or low quality). The sound files were assigned to the twenty raters randomly so that each file was assessed by at least two raters. Therefore, twenty separate surveys were built. Each survey included 11 core files that were chosen randomly from the whole sample. These files were evaluated by all raters in order to calculate inter-rater reliability of their scoring. In addition, each rater evaluated between 55 and 57 recordings. According to the background data recorded by Qualtrics, complete rating took approximately 96 minutes to complete on average. Raters were able to complete the scoring in two sessions with one break (up to one week) in between. All rating procedures had to be completed within two weeks after the training, as research has shown that rater training effects tend to fall off quickly (e.g., Lumley & McNamara, 1995).

Data Analyses

Prosodic Analysis

The prosodic analyses in this study closely followed the procedures described in Kang *et al.* (2021). The collected speech samples were converted into .wav format and examined using PRAAT (Boersma & Weenink, 2021) with regard to tone (rising, falling, level tones) and

prominence measures (Brazil, 1997; Pickering, 2001). As Kang *et al.* hypothesized, these were some of the discourse intonational features that could be particularly effective indicators of pragmatic competence due to them being especially “unintuitive” for English learners (p. 13). Additionally, the joint results of the pilot studies that prosodically analyzed L1 and L2 English speakers’ production of requests showed that these features were particularly meaningful for the conveyance of pragmatic appropriateness. Thus, falling tones were found more appropriate in request-making; however, level intonation was considered acceptable in request modifications, particularly when giving reasons for a request (e.g., Kang *et al.*, 2021). With regard to prominence, overuse of stressed words in a sentence has been found detrimental to the appropriateness of this speech act. In order to determine the speakers’ tone choices, the speech samples were divided into tone units following Pickering’s (2018) guidelines, and tone choices were identified for the tonic syllable in each tone unit. A tone unit in this analysis was operationalized as a unit of speech that conveys one idea or a piece of information (Pickering, 2018). For prominence analysis, PRAAT spectrograms and pitch contours were obtained, and the percentage of prominent words were calculated out of the total number of words in each utterance. Prominent syllables were those syllables in the spectrogram that were longer, louder, and higher in pitch. In addition, a proportion of appropriately stressed words in a sentence was calculated for each request. Inter-coder reliability between the researcher and a research assistant was calculated for each prosodic feature and is reported in the Results chapter.

For the tonal composition of speech samples, appropriate choice of intonation was measured on the main request strategy for each video-based DCTs produced by each participant. Thus, a falling tone with a mitigated preparatory strategy received one point while other tones received zero points. Regarding prominence, the percentage of prominent words used

appropriately in the context of request making was calculated, meaning that learners stressed content words important for the overall message of the request and not function words distracting the listeners from the intended meaning. After tone and sentence stress measures were analyzed in each speech sample, they were then tabulated in an Excel spreadsheet and compared across the tasks and types of requests to examine the effect of formality and the type of role play on the use of prosodic cues. Two coders performed inter-coder reliability analysis for 10% of the data and reached 95% agreement for tone unit identification, 93% agreement for tone choice identification, and 92% for prominence identification. A complete description of the selected prosodic properties and measures adapted from Kang *et al.* (2021) for the current analysis is given in Table 6.

Table 6

Selected Intonation Properties and Measures (partially adapted from Kang et al., 2021)

Features	Measures
Tone choice	
Rising tones	Tone choices on the termination syllable in the request head act (request itself) used appropriately with the Mitigated Preparatory strategy.
Falling tones	
Level tones	
Prominence	Percentage of stressed words out of the total number of words (Space). Percentage of appropriately stressed prominent words out of the total number of words (Appropriate Use).

Figure 11

PRAAT Analysis of Tone Choices and Prominence

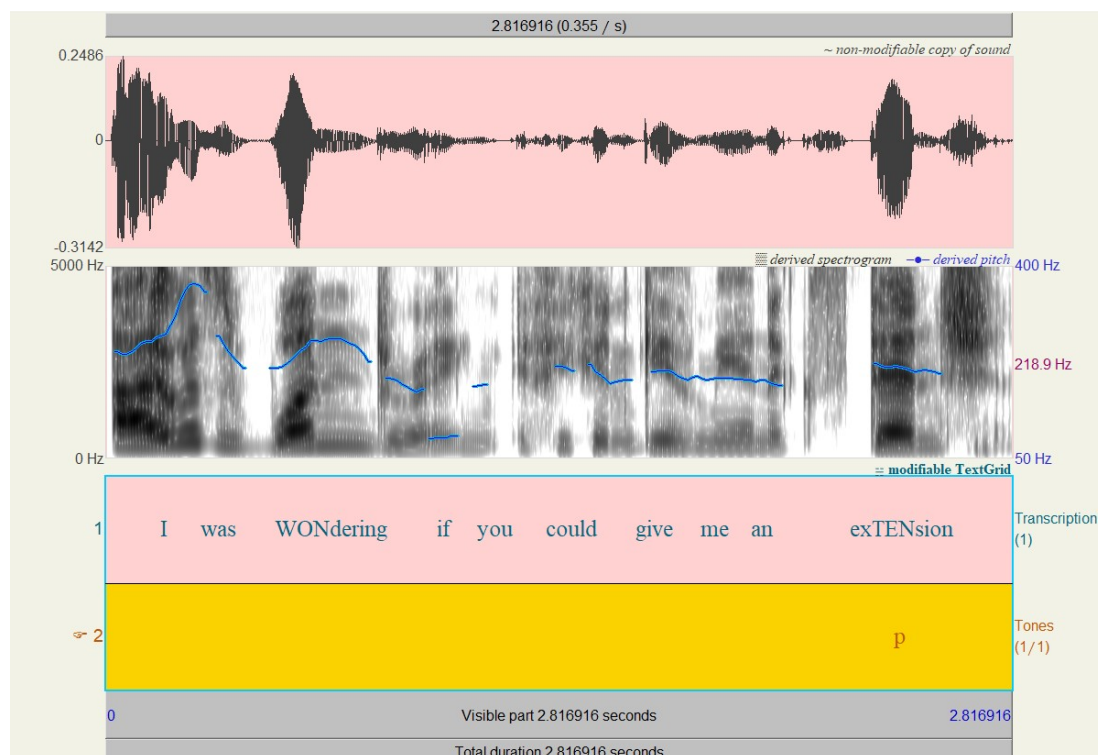


Figure 11 additionally exemplifies the process of tone choice and prominence identification using PRAAT. In the figure, a request *I was wondering if you could give me an extension* is given in its sound form. The prominent syllables in the request are indicated by louder, longer soundwaves that can be seen in the top layer of the PRAAT window. Thus, it appears that the syllables *I*, *WONdering*, and *exTENSion* were prominent in the request. The request in the example is analyzed as one tone unit since it conveys one idea. Therefore, the last stressed syllable in the request is the one that is used to identify the tone choice. Based on the direction of the blue line in the second layer representing the pitch contour, the request received falling intonation.

Analysis of Learners' Intonation Perception Ability

In order to analyze learners' perception ability and how it relates to the changes in their use of intonation after the training as well as its link to cognitive IDs, the correctness of their first

attempt on each item of perception-based tasks was measured. Thus, if a learner answered an item in a task correctly during the first attempt at the task, the website recorded it as 1, and if the answer is incorrect, the website recorded a 0. Then, the sum of learners' correct responses on the first attempt of each perception task was calculated producing four indices: tone choice identification, prominence identification, tone choice discrimination, and prominence discrimination. In identification exercises, learners either indicated the tone choice they heard in a request (rising, falling, or level) or identified all prominent words in a request. In discrimination exercises, learners listened to two requests and chose a more appropriate one from the point of view of tone choices or prominence.

Study Variables

This section presents an overview of the independent (IV) and dependent (DV) variables that were measured to answer the research questions in the study. Table 7 includes the operationalizations of each variable and the instruments that were used to measure them.

Table 7

Operationalizations of the Variables in the Study

RQ	Variable type	Variable name	Operationalization
RQ1	IV	Pragma-prosodic training	Time 1 (before training), Time 2 (one day after training), Time 3 (one week after training – completed by a sub-group of learners)
	IV	Group	Experimental and Control groups
	DV	Changes in the use of prosody	Normed counts (percentages) of appropriately used tone choices and prominence measures in pre- and post-training role plays
		Raters' assessments	Raters' scores on three scales: Pragma-Prosodic Appropriateness, Accentedness, Comprehensibility
RQ2	IV	Working memory	A score on the PWM test

RQ3	IV	Musical ability	A score on the PROMS test
		Auditory processing	A score on the auditory processing test
		English proficiency	A score on the EIT
		Use of prosody before training	Normed counts (percentages) for tone choices and prominence measures in pre-training
		Use of prosody after training	Normed counts (percentages) for tone choices and prominence measures in post-training role plays
	DV	Learners' intonation perception	(1) Number of correct responses during the first attempt in each tone choice identification task item
			(2) Number of correct responses during the first attempt in each prominence identification task item
			(3) Number of correct responses during the first attempt in each tone choice discrimination task item
			(4) Number of correct responses during the first attempt in each prominence discrimination task item
		Changes in the use of prosody	Normed counts (percentages) for tone choices (rising, falling, level) and prominence measures in pre-training and post-training role plays

In addition to the quantitative examinations outlined above, learners' responses from the final questionnaire were coded and thematically analyzed in a quantitative way. The questionnaire primarily focused on learners' perceptions of the effectiveness of the developed training. Although learners' perceptions were not one of the foci of the study, survey responses are summarized in the discussion section of this dissertation and are applied in interpreting the quantitative results of the study.

CHAPTER 4: RESULTS

The following chapter reports on the results of the quantitative analyses conducted to address the research questions in the dissertation study. For each research question, the chapter first presents an overview of the of the conducted analysis together with corresponding assumptions tests, followed by detailed descriptions of the results. The chapter begins with a section on the results for the first research question that examines the changes in learners' use of two selected prosodic features (tone choices and prominence) after the designed training from the acoustic and perceptual perspective. The section is broken down into two parts. The first part describes the results of intonation and prominence analyses and the effect of the training on these prosodic features. The second part approaches the first research question from the perceptual viewpoint, analyzing whether raters' appropriateness perceptions related to learners' use of intonation after the training are different from those before the training. Following the first section is a detailed description of the results addressing the second research question (effect of learners' cognitive IDs in their use of prosody before and after the training). In parallel with the first section in the chapter, the second section is also divided into two parts, one examining the effect of IDs before the training, and the other one looking at this effect after the training. Finally, results for the third research question (learners' ability to perceive intonation cues and training effectiveness) are presented in the third section of the chapter. The section is similarly subdivided into two parts. First, it explores the role of learners' perceptual ability and the observed changes in learners' use of prosody. Second, it details the relationship between learners' perceptual ability and cognitive IDs. Each section in the chapter concludes with a summary of the results for each research question. All data analyses were conducted in *R* (ver. 4.0.3), a free statistical environment (R Core Team, 2021).

Research Question 1: Changes in Learners' Use of Prosody After Training

This section reports on the results of statistical analyses conducted to address the first research question, which examined the effectiveness of the developed instructional training as reflected in learners' use of prosodic features of tone choices and prominence as well as changes in raters' perception of pragma-prosodic appropriateness in learners' requests after the training. In order to address this two-part research question, linear mixed effects modeling technique (LMEMs) was used on the data with the help of the *lme4* package (Bates *et al.*, 2015). Below, the results of LMEMs are presented for each sub-question.

Evidence from Acoustic Analysis

Conducting analyses to examine learners' use of prosodic features required an acoustic analysis of learners' production of requests before and after the training. Therefore, collected audio files were analyzed using an existing pragmatic framework and PRAAT software by the researcher. For each sound file that contained the requestive speech act (i.e., the first turn in each video-based multiple-turn DCT), the main pragmatic strategy was identified to ensure that the learners used the target structure taught in the training. Then, the main tone choice associated with the pragmatic strategy was extracted. If a learner used falling intonation on a Mitigated Preparatory strategy (*I was wondering if...* / *I was hoping that...*), it was counted as appropriate. In cases when rising of level intonation was used, it was counted as inappropriate. For prominence, the total number of stressed words was calculated and then divided by the total number of words in the request for the space measure. In addition, the total number of appropriately stressed words was calculated and divided by the total number of stressed words to find out the pragmatic use of prominence in requests. An appropriately stressed in this case would be a content word that was important for the requestive message. A research assistant (RA) with phonetic training

acoustically analyzed 10% of the data in order to check the reliability of the analysis. A comparison of the researcher's and RA's analyses showed 95% agreement for tone unit identification, 93% agreement for tone choice identification, and 92% for prominence identification. After discussion, agreement reached 100% in all categories.

The first group of analyses investigated the effect of pragma-prosodic training on learners' production of tone choices and prominence. Six LMEMs were fitted on the data with tone choices and prominence (space and appropriate use) as response variables in each model. Before that, however, descriptive statistics were examined and assumptions were checked to ensure that the models could be built with the given data. Tables 8, 9, and 10 present the means (*M*), standard deviations (*SD*), and 95% confidence intervals (95% *CI*) for the production of tone choices and prominence in the baseline, experimental, and control groups before and after the training. The mean values in the table indicate (1) for tone choices, the proportion of tone choices used appropriately by learners in requests and (2) for prominence, the proportion of stressed words out of all the words produced by learners in each request (also known as space), and (3) for prominence, the proportion of appropriately stressed words out of all stressed words (referred to as appropriate use of prominence throughout the study).

Baseline Group Production. The baseline group consisted of six speakers (three L1 English and three highly proficient L2 English) that provided a criterion for the learner data. Each speaker provided six recordings that were elicited via multiple-turn video-based DCTs used in the pre- and post-tests. These data were not included in LMEMs and were examined for reference in order to understand if the learners were approaching target prosodic use.

Table 8 presents an overview of baseline speakers’ use of tone choices in requests. Note that the table includes tone choices extracted from all parts of the requests including not only the main strategy, but also modifications (e.g., attention getters, grounders, etc.). As seen in the table, the overall distribution of tone choices observed in the baseline data closely mirrors previous research that found that L1 and proficient L2 English speakers mostly use falling intonation in requests, followed by rising and level intonation, which is mostly used when providing an explanation for a request (e.g., Kang *et al.*, 2021; Kostromitina, in press). In terms of prominence, both L1 and L2 English speakers in the baseline group maintained similar sentence stress patterns emphasizing approximately 3 words per thought group. On average, L1 English speakers’ prominence (measured by space) was at $M = 0.41$ syllables ($SD = 0.02$, 95% $CI [0.39; 0.43]$). The prominence of L2 English speakers in the baseline group was comparable at $M = 0.43$ syllables ($SD = 0.02$, 95% $CI [0.41; 0.45]$).

Table 8

Descriptive Statistics for Baseline Speakers’ Use of Tone Choices and Prominence

Variable	L1 English			L2 English		
	<i>N</i>	<i>Prop.</i>		<i>N</i>	<i>Prop.</i>	
Falling tones	54	0.63		48	0.55	
Rising tones	21	0.25		26	0.29	
Level tones	10	0.12		14	0.16	
	<i>M</i>	<i>SD</i>	<i>95% CI</i>	<i>M</i>	<i>SD</i>	<i>95% CI</i>
Prominence (space)	0.41	0.02	[0.39; 0.43]	0.43	0.02	[0.41; 0.45]

However, since the developed training focused particularly on tone choices and prominence in the *main head act strategy* rather than the overall speech act composition, additional analyses were conducted. Table 9 presents the results of acoustic analyses specifically of requestive head acts in the baseline data. In this analysis, each head act was treated as one tone unit since it matched Pickering’s (2018) definition of a phrase conveying one idea or thought.

Thus, tone choices in 36 head acts (6 speakers x 6 requests each) were calculated. Out of all the head act strategies, only two (produced by L2 English speakers) received rising intonation. The remaining head acts were all produced with falling tone choices (94%). Importantly, all the head acts in the baseline data contained exclusively the mitigated preparatory request strategy (i.e., *I was wondering if...*). The combination of falling tones and mitigated preparatory strategies that is observed in the baseline data can be interpreted and supported by Brazil's (1997) Discourse Intonation Framework that suggests that falling tone choices are commonly associated with new information and statements that are "world-changing" (*I am telling you something new*) (see also Pickering, 2018). Since requestive head acts typically contain information that is new for the hearer (i.e., the request has not previously been made), falling intonation is appropriate in this context. Moreover, from the grammatical standpoint, the bi-clausal structure of a mitigated preparatory strategy (e.g., an independent clause *I was wondering* and dependent clause *if you could write me...*) resembles a declarative statement rather than a question (cf. *Could you write me a...?*) thus requiring falling intonation in most cases (Levis, 1999). The baseline speakers' prominence in head acts was similar to one reported above and rested at $M = 0.40$ stressed syllables ($SD = 0.02$, $95\%CI [0.39; 0.42]$).

Table 9

Baseline Speakers' Use of Tone Choices and Prominence in Request Head Acts

Variable	<i>N</i>	<i>Prop.</i>	
Tone choices			
<i>Falling</i>	34	0.94	
<i>Rising</i>	2	0.6	
<i>Level</i>	0	0	
	<i>M</i>	<i>SD</i>	<i>95% CI</i>
Prominence (space)	0.40	0.02	[0.39; 0.42]

The results of pragma-prosodic analyses of baseline production, specifically regarding head act strategies, as well as the focus of the developed instruction on head acts, informed the subsequent analyses of learner production. Thus, learners' prominence and tone choice use were measured on head act strategies. If a learner used a mitigated preparatory request strategy with falling intonation, they received 1, if rising or level intonation was used, they received 0. Prominence was operationalized as space (proportion of stressed words over total words).

Experimental and Control Group Production. Table 10 presents a summary of the acoustic analyses for the experimental and control group. Note that the table also includes the descriptive statistics for the sub-group learners in the experimental and control groups who completed the delayed post-test.

Table 10

Descriptive Statistics for Learners' Use of Tone Choices and Prominence

Experimental							Control					
Variable	<i>N</i>	<i>M</i>	<i>SD</i>	95% <i>CI</i>	<i>Kurt</i>	<i>Skew</i>	<i>N</i>	<i>M</i>	<i>SD</i>	95% <i>CI</i>	<i>Kurt</i>	<i>Skew</i>
Tone choices pre-test	37	0.59	0.5	[0.43; 0.75]	-1.96	-0.27	37	0.65	0.48	[0.50; 0.80]	-1.59	-0.66
Tone choices post-test	37	0.88	0.35	[0.77; 0.99]	2.27	-2.05	37	0.68	0.47	[0.53; 0.83]	-0.63	-0.81
Prominence pre-test	37	0.58	0.1	[0.55; 0.61]	0.93	0.35	37	0.56	0.07	[0.54; 0.58]	-1.38	0.34
Prominence post-test	37	0.42	0.07	[0.40; 0.44]	1.27	0.79	37	0.68	0.09	[0.65; 0.71]	-0.63	0.46
With a delayed post-test												
Tone choices pre-test	20	0.53	0.5	[0.31; 0.75]	-2	-0.1	18	0.65	0.48	[0.44; 0.86]	-1.66	-0.62

Tone choices post-test	20	0.86	0.35	[0.71; 1.01]	2.74	-2.16	18	0.57	0.5	[0.35; 0.79]	-1.99	-0.29
Tone choices delayed	20	0.73	0.45	[0.53; 0.93]	-0.92	-1.06	18	0.54	0.5	[0.32; 0.76]	-2	-0.16
Prominence pre-test	20	0.55	0.1	[0.51; 0.59]	1.25	0.72	18	0.56	0.09	[0.52; 0.60]	-0.15	0.46
Prominence post-test	20	0.41	0.08	[0.37; 0.44]	2.09	0.84	18	0.64	0.08	[0.60; 0.68]	0.6	0.81
Prominence delayed	20	0.4	0.06	[0.37; 0.43]	0.49	-0.57	18	0.55	0.07	[0.52; 0.58]	0.02	0.53

Based on the descriptive statistics provided in Table 10, the data largely met the normality assumptions with kurtosis and skewness values being within the -2 and +2 interval with the exception of intonation in the post tests produced by the experimental group. A comparison of the mean values of tone choice use and prominence between the baseline group and the experimental/control group shows that the experimental group indeed started approaching target-like production of prosodic features after the instruction. To visually evaluate the data for normality, linearity, and homoscedasticity, two scatterplots were created. The scatterplots in Figures 12 – 14 do not indicate an obvious curve in the distribution with residuals being located symmetrically above and below the 0 line, although there is a slight deviation from linearity in prominence data in Figure 13. The visual examination of the residual vs. fitted plots, despite some of the deviations from normality in kurtosis and skewness values, suggested that the chosen LMEM analyses could indeed be conducted. The residuals in the plots were distributed relatively equally on both sides of the 0-horizontal line.

Figure 12

Residuals Plot for Tone Choices

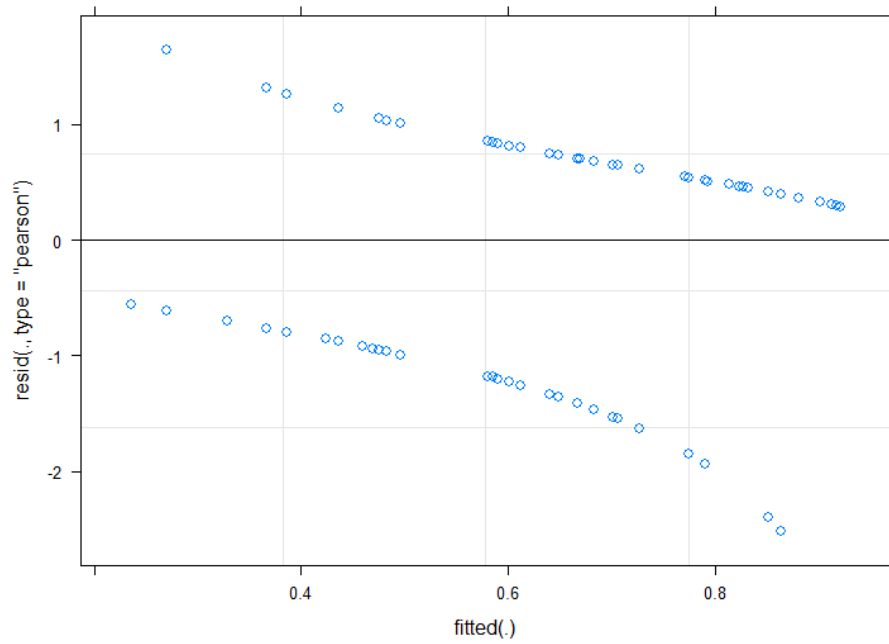


Figure 13

Residuals Plot for Prominence (Space)

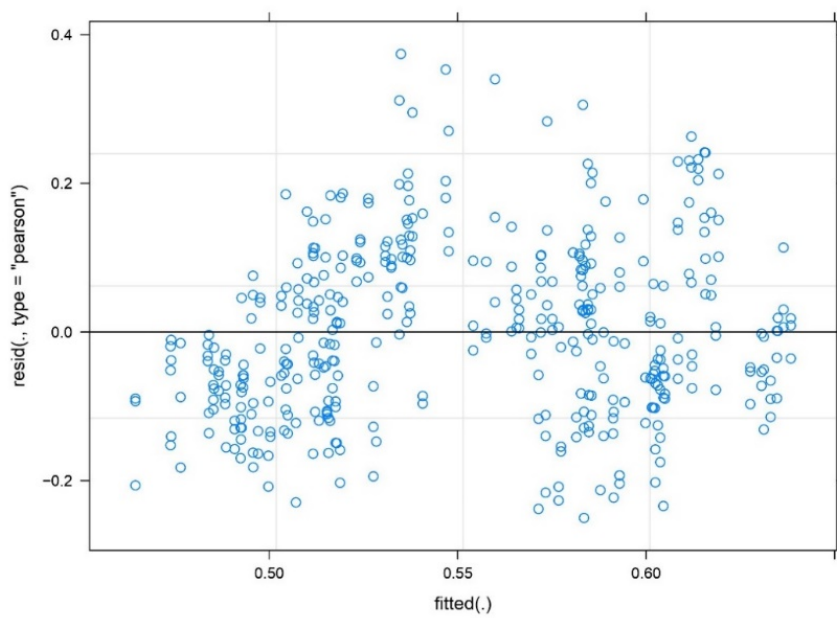
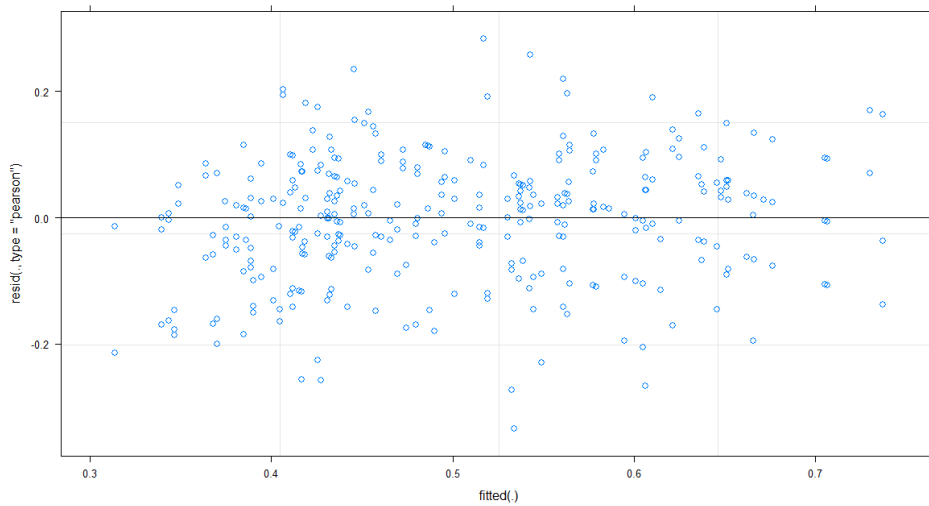


Figure 14

Residuals Plot for Prominence (Appropriate Use)



Changes in Tone Choices. For the first LMEM investigating changes in the use of tone choices after the training, Group (experimental vs. control) and Test (pre-test vs. post-test) were entered as fixed factors. The model also measured an interaction between the two fixed factors. Speakers (i.e., individual speaker IDs) were entered as a random intercept. The response variable was the appropriateness of tone choices used on the main strategy in requests. Since the response variable contained only two values (0 = inappropriate, 1 = appropriate), the binomial (logistic) type of LMEM was chosen for the analysis. The model was fit by maximum likelihood using the adaptive Gauss-Hermite approximation of the likelihood. The results of this model are presented in Table 11 that includes the estimates for each comparison of fixed factors to the intercept as well as standard errors, 95% confidence intervals, and significance values. The table also includes random effect indices, such as standard deviation, variance, inter-class correlations, the number of learners and the number of observations that were collected from the learners in the pre-, post-tests.

The fixed effects in the model explained 9% of the variance in learners' tone choice production ($R^2_m = 0.09$), and the random effect of speakers (i.e., idiosyncrasies related to individual speakers' production) explained an additional 27% ($R^2_c = 0.36$). There was a significant interaction between the fixed effects ($F(1, 352) = 12.76, p < .01$). Thus, post-hoc simple effects were further calculated using Tukey HSD testing.

Table 11

Summary of the Binomial Mixed Effects Model for Tone Choices

Predictors	<i>Estimates</i>	<i>SE</i>	<i>95% CI</i>	<i>p</i>
(intercept)	0.85	0.33	[0.21; 1.50]	<0.01
Experimental	1.32	0.50	[0.34; 2.29]	<0.01
Pre-test	-0.01	0.35	[-0.69; 0.67]	0.97
Experimental : Pre-test	-1.79	0.52	[-2.82; -0.76]	<0.01
Random effect				
σ^2	1.42			
τ_{00}	1.19			
ICC	0.301			
N	74			
Observations	412			

According to the post hoc calculations presented in Table 12, there was a significant difference in tone choice production demonstrated by the experimental group in the pre- and post-test. Specifically, in the post-test, learners in the experimental group used tone choices significantly more appropriately than in the pre-test. In other words, they were more likely to use falling intonation with mitigated preparatory strategies after the training, indicating that the developed training had a medium to strong effect on learners' use of this prosodic feature ($p < 0.01$, Cohen's $d = 0.67$). Furthermore, there was a significant difference in post-test production of tone choices between the experimental and control groups with learners in the experimental group performing significantly better than those in the control group with a small effect ($p <$

0.01, Cohen's $d = 0.48$) as suggested by previous instructional studies in L2 pronunciation (e.g., Saito & Plonsky, 2019) and L2 pragmatics (Plonsky & Zhuang, 2019) as well as by the effect size guidelines for applied linguistics (Plonsky & Oswald, 2014).

Table 12

Results of Post Hoc Pairwise Tukey HSD Tests for Tone Choices

Group contrast	<i>Estimate</i>	<i>SE</i>	<i>df</i>	<i>z</i>	<i>p</i>	<i>d</i>
Control post vs. pre	0.01	0.35	Inf	0.03	0.97	0.29
Experimental post vs. pre	1.80	0.39	Inf	4.66	<0.01	0.67
Pre-test control vs. experimental	0.47	0.43	Inf	1.11	0.27	0.12
Post-test control vs. experimental	-1.32	0.50	Inf	-2.65	<0.01	0.48

Figure 15

Tone Choice Production in Pre- and Post-Tests

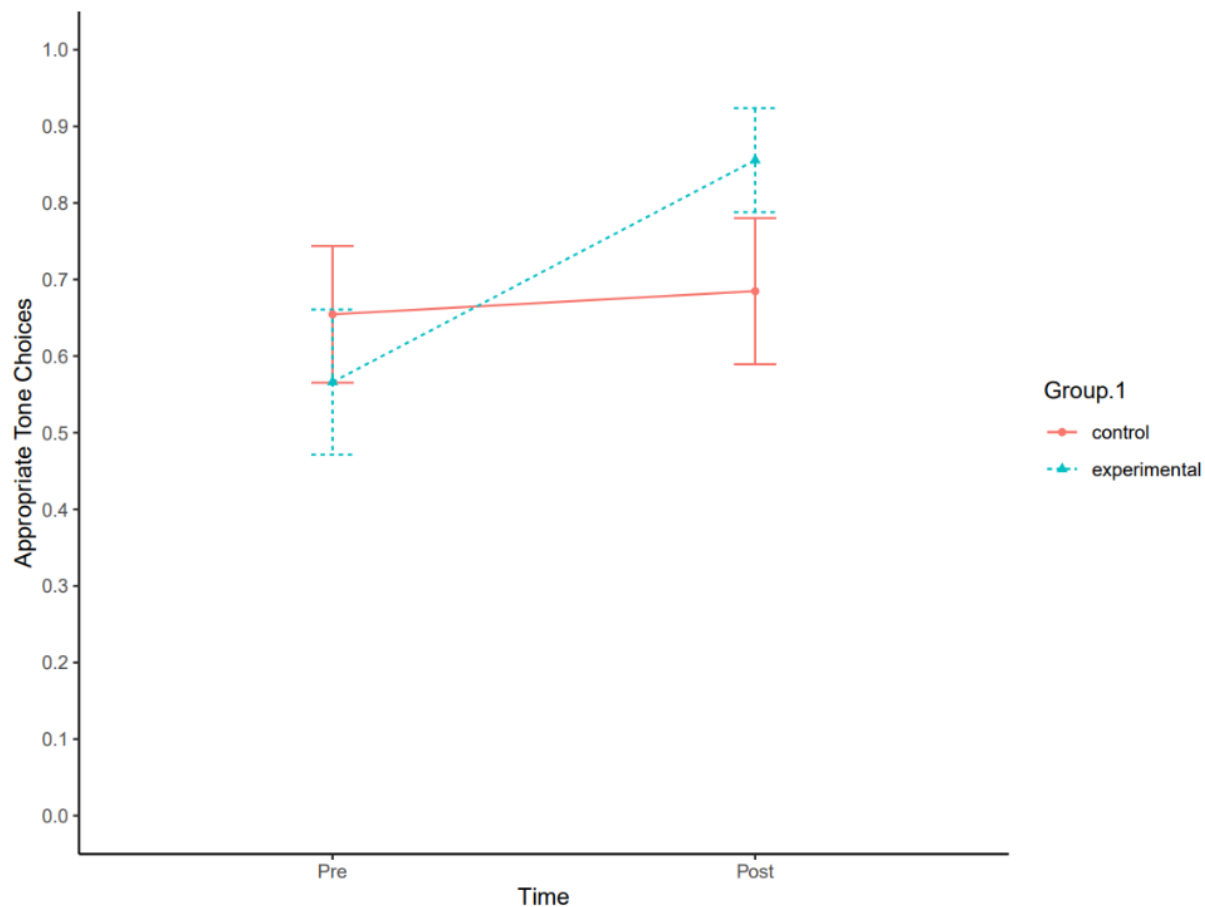


Figure 15 visualizes the results of post-hoc tests in a line graph. In the graph, the x -axis represents the fixed effect of time (before and after the training), and the y -axis reflects the proportion of appropriately produced tone choices on the scale from 0 to 1. The performance of experimental group learners is recorded by the blue dotted line, and the control group – by the red line. Finally, 95% confidence intervals are given around each of the means. Figure 14 further demonstrates the significant improvement in the experimental group learners’ use of tone choices as compared to their pre-test production and to the use of tone choices by learners in the control group.

Changes in Tone Choices Including Delayed Post Test. In addition to the pre-and post-test, some participants in the study (20 in the experimental group, 18 in the control group) also completed a delayed post-test that was offered one week after the end of training. While this sample size was smaller than desired, it was decided to also conduct LMEM analyses with these participants to gain additional evidence for the training effectiveness. Due to the limitations in group sizes, these results should be taken into consideration with caution.

Table 13 summarizes the results of the binomial LMEM with the same fixed and random effects as in the main analysis. That is, Group and Test (this time, pre-, post-, and delayed), as well as an interaction between them, were entered as fixed factors. Speakers were entered as a random intercept. The model accounted for approximately 48% of variance in learners’ production of appropriate tone choices ($R^2_m = 0.11$ for the fixed factors and $R^2_m = 0.37$ for the random factor of speakers). There was a significant interaction detected between the factors of Group and Time ($F[2, 286] = 6.54, p < .01$); thus, post hoc Tukey HSD tests were run.

Table 13

Summary of the Binomial Mixed Effects Model for Tone Choices with Delayed Post Test

Predictors	<i>Estimates</i>	<i>SE</i>	<i>95% CI</i>	<i>p</i>
(intercept)	0.53	0.43	[-0.46; 1.18]	0.21
Experimental	0.72	0.59	[-0.26; 2.07]	0.22
Pre-test	0.24	0.46	[-0.54; 1.23]	0.59
Post-test	-0.16	0.45	[-0.94; 0.81]	0.72
Experimental : Pre-test	-1.33	0.64	[-2.69; -1.20]	0.03
Experimental : Post-test	1.14	0.69	[-0.31; 2.40]	0.09
Random effect				
σ^2				
τ_{00}	1.16			
ICC	0.29			
N	38			
Observations	327			

Table 14 presents the results of Tukey HSD tests for participants in the control and experimental groups who completed delayed post-tests on top of the pre- and post-tests. Similarly to the main analysis, the results indicate several differences in learners' use of tone choices including a significant difference between the pre- and post-test in the experimental group as well as a significant difference in the post-test performance between the experimental and control groups. Of interest, however, is another contrast; namely, the significant difference in the use of appropriate tone choices between the pre-test and delayed post-test in the experimental group. This difference is further illustrated in Figure 16 that presents the production of appropriate tone choices by participants who completed pre-, post-, and delayed post-tests. It is noteworthy that while the appropriate production of tone choices among the learners in the experimental group declines in the delayed post-test as compared to the immediate post-test ($M = 0.86$ $SD = 0.35$ and $M = 0.73$ $SD = 0.45$ respectively), there is still a significant difference between their use of tone choices in the pre- and delayed post-test, albeit with a small effect size

($p = 0.03$, Cohen's $d = 0.42$). These results suggest that the developed training may have a long-lasting positive effect on learners' use of intonation, although it may weaken with time.

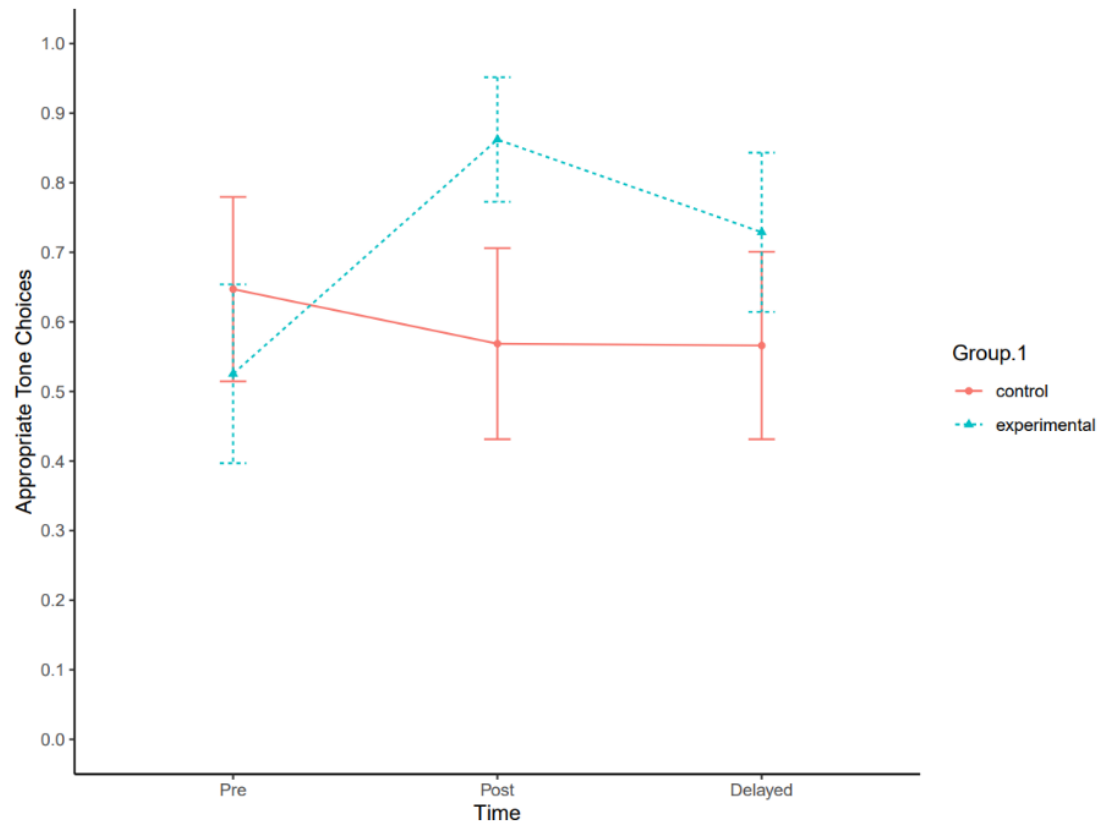
Table 14

Results of Post Hoc Pairwise Tukey HSD Tests with Delayed Post Test for Tone Choices

Group contrast	Estimate	SE	df	z	p	d
Control post vs. pre	-0.41	0.46	Inf	-0.90	0.64	0.16
Control delayed vs. pre	-0.34	0.45	Inf	-0.76	0.73	0.22
Control post vs. delayed	0.07	0.46	Inf	0.15	0.99	0.06
Experimental post vs. pre	2.08	0.52	Inf	4.00	<0.01	0.76
Experimental delayed vs. pre	1.10	0.44	Inf	2.48	0.03	0.42
Experimental post vs. delayed	-0.98	0.52	Inf	-1.87	0.15	0.32
Pre-test control vs. experimental	0.54	0.58	Inf	0.92	0.36	0.24
Post-test control vs. experimental	-1.95	0.66	Inf	-2.96	<0.01	0.67
Delayed control vs. experimental	-0.90	0.60	Inf	-1.52	0.13	0.40

Figure 16

Tone Choice Production in Pre-, Post-, and Delayed Tests



Overall, the findings suggest that from Brazil's (1997) Discourse Intonation Framework standpoint, learners produced requests that included more falling intonation. This means that requests were more appropriately structured prosodically due to the higher alignment of falling intonation with mitigated preparatory strategies.

Changes in Prominence. Besides examining changes in tone choices, this study investigated changes in learners' use of prominence after the training. Thus, two additional LMEMs were fitted on the data with Group and time as the fixed factors as well as an interaction between them, and Speakers were entered as a random intercept. Each LMEM included a response variable that was either space or appropriate use of prominence, as outlined above. The results of the first LMEM in this group are presented in Table 15. The model was able to explain a total of 62% of variance in learners' use of prominence ($R^2_c = 0.62$) with the fixed effects explaining the majority of it ($R^2_m = 0.53$) and the random effect of speakers explaining 9%. A closer examination of the results revealed a significant interaction between the two fixed factors ($F(1, 352) = 32.19, p < .01$); thus, post hoc Tukey HSD tests were used to calculate the simple effects. The results of post hoc tests are given in Table 16.

Table 15

Summary of the Mixed Effects Model for Prominence (Space)

Predictors	<i>Estimates</i>	<i>SE</i>	<i>95% CI</i>	<i>p</i>
(intercept)	0.68	0.01	[0.66; 0.70]	<0.01*
Experimental	-0.26	0.01	[-0.29; -0.23]	<0.01
Pre-test	-0.11	0.01	[-0.14; -0.09]	<0.01
Experimental : Pre-test	0.28	0.02	[0.25; 0.31]	<0.01
Random effect				
σ^2	0.04			
τ_{00}	0.001			
ICC	0.19			
N	74			
Observations	412			

Table 16 shows that there was a significant difference in a few pairwise comparisons. Specifically, the experimental group demonstrated a significant improvement in the appropriate use of prominence after the training with a large effect size ($p < 0.01$, Cohen's $d = 3.11$). For reference, the target prominence established by the baseline speaker group was at $M = 0.40$ stressed syllables. This trend is further illustrated in Figure 17. As seen in the figure, learners in the experimental group (indicated by a the teal boxes) stressed fewer words in requests after the training suggesting that their use of prominence improved.

Table 16

Results of Post Hoc Pairwise Tukey HSD Tests for Prominence (Space)

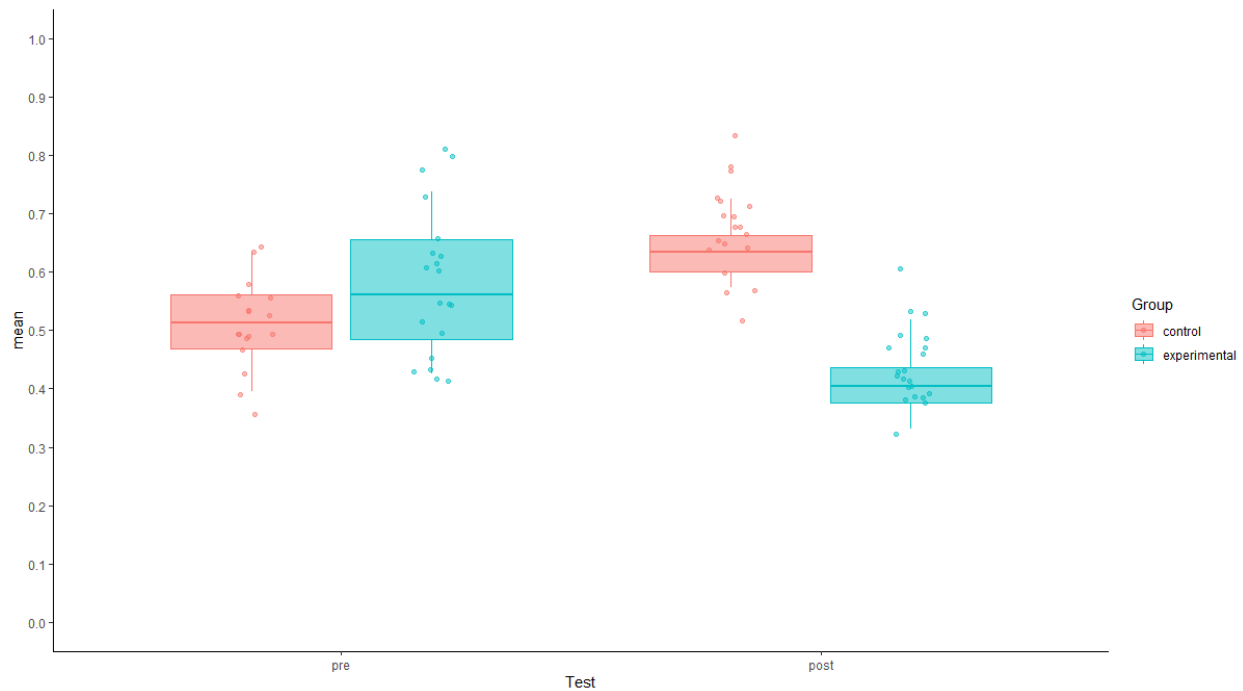
Group contrast	<i>Estimate</i>	<i>SE</i>	<i>df</i>	<i>t</i>	<i>p</i>	<i>d</i>
Control post vs. pre	0.11	0.01	358	10.19	<0.01	1.49
Experimental post vs. pre	-0.17	0.01	343	-15.27	<0.01	1.85
Pre-test control vs. experimental	-0.02	0.01	134	-1.27	0.20	0.23
Post-test control vs. experimental	0.26	0.01	147	18.30	<0.01	3.22

In contrast, the control group seems to have increased the amount of stressed words compared to the total number of words in the post-test as shown in Figure 17, which contains box plots indicating medians, interquartile ranges, and maximum / minimum values for each level of the fixed factors. The dots on the plot are individual learner data points. In the figure, the y-axis reflects the proportion of stressed words in a request, known as space; lower numbers indicate that learners only stress important, content words while lower numbers suggest that learners overuse prominence and stress words that may not be important for communicating the request meaning. In fact, the post-hoc tests indicated a significant difference between the proportion of stressed words in the pre- and post-tests in the control group with a large effect size

($p < 0.01$, Cohen's $d = 1.49$). Finally, there was a significant difference between the use of prominence among the learners in the experimental and control groups after the training with a large effect size ($p < 0.01$, Cohen's $d = 1.85$).

Figure 17

Prominence Production (Space) in Pre- and Post-Tests



A second measure of learners' appropriate use of prominence was a proportion of appropriately stressed words out of the total number of stressed words. Thus, another LMEM was fitted on the data with this response variable while keeping all the other parameters the same as the previous model. The results of the LMEM are summarized in Table 17. Similarly to the analysis of space above, there was a significant interaction between the two fixed effects in the model ($F[1, 352] = 294.57, p < 0.01$). The model explained 53% of variance of learners' appropriate use of prominence before and after the training, with the random effect of speaker explaining only 12% of this variance, indicating that the majority of explained changes in appropriate prominence use was due to the training.

Changes in prominence → appropriateness

Table 17

Summary of the Mixed Effects Model for Prominence (Appropriate Use)

Predictors	<i>Estimates</i>	<i>SE</i>	<i>95% CI</i>	<i>p</i>
(intercept)	0.41	0.01	[0.38; 0.44]	<0.01
Experimental	0.24	0.02	[0.20; 0.28]	<0.01
Pre-test	0.13	0.01	[0.10; 0.15]	<0.01
Experimental : Pre-test	-0.35	0.02	[-0.38; -0.31]	<0.01
Random effect				
σ^2	0.05			
τ_{00}	0.003			
ICC	0.21			
N	74			
Observations	412			

Figure 18 visualizes the changes in appropriate use of sentence stress in learners' requests after the training. In contrast with the previous model, a higher mean value on the y-axis in this figure would indicate a more appropriate use. The figure shows that learners in the experimental group improved their prominence after the training while those in the control group demonstrated a reduced number of appropriately stressed words in the post-test. These findings are further elaborated on in Table 18 that provides the results of Tukey HSD tests to measure the simple effects of group (experimental and control) and time (pre- and post-test). As reported in the table, all four contrasts in the fixed effects were significantly different and with large Cohen's *d* effect sizes. Of importance for the study is the finding that experimental group significantly improved their use of prominence meaning that they more frequently stressed words that were pragmatically salient or important for the request (e.g., content words, words that would help the listener understand the meaning of the request and not appear impolite or too direct). For example, in asking for a letter of recommendation using a bi-clausal mitigated

preparatory strategy *I was wondering if...*, it would be expected that the prominent words would be *wondering*, *letter*, and *recommendation*.

Figure 18

Prominence Production (Appropriate Use) in Pre- and Post-Tests

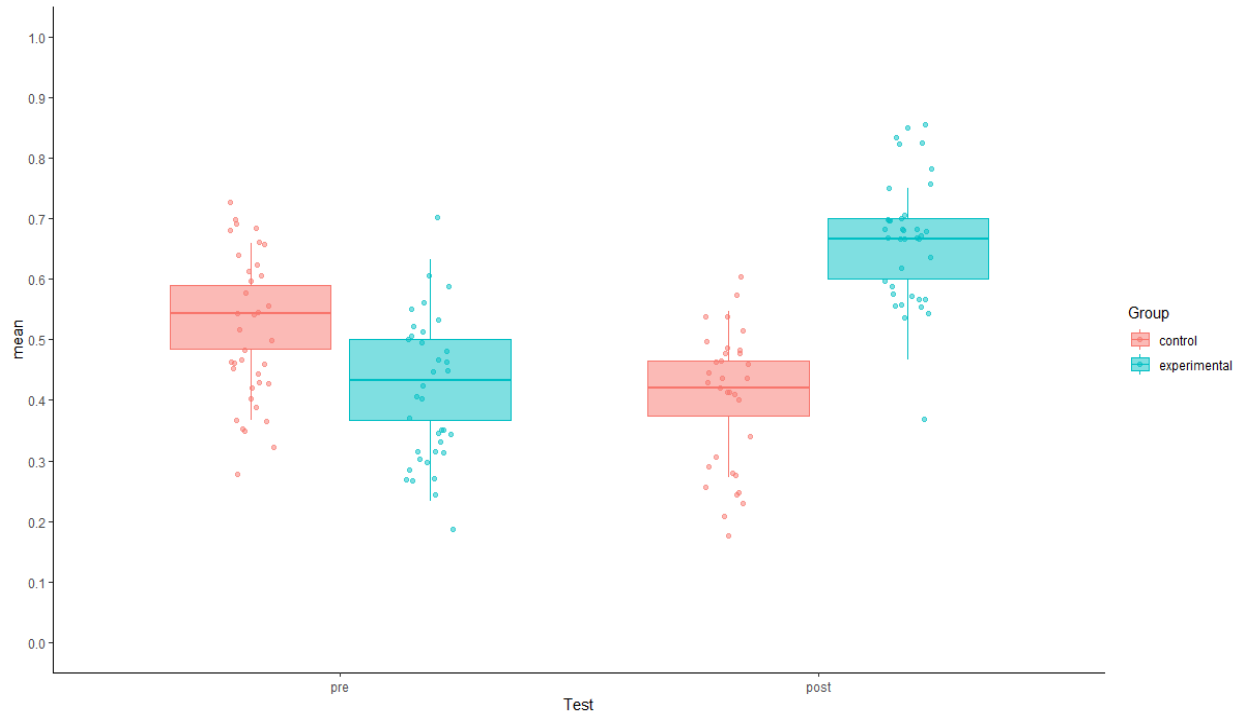


Table 18

Results of Post Hoc Pairwise Tukey HSD Tests for Prominence (Appropriate Use)

Group contrast	<i>Estimate</i>	<i>SE</i>	<i>df</i>	<i>t</i>	<i>p</i>	<i>d</i>
Control post vs. pre	-0.13	0.01	357	-8.68	<0.01	0.92
Experimental post vs. pre	0.22	0.01	342	15.70	<0.01	1.71
Pre-test control vs. experimental	0.11	0.02	130	5.78	<0.01	1.01
Post-test control vs. experimental	-0.24	0.02	143	-12.68	<0.01	2.12

Changes in Prominence Including Delayed Post Test. Parallel to the examination of tone choices, two LMEMs were built with the data from those 38 learners that completed a delayed post-test in addition to the pre- and post-tests. The results of the model are presented in

Tables 19 – 22. The first LMEM included space as the response variable. The fixed and random effects in the model explained around 63% of the variance in learners' use of prominence ($R^2_c = 0.63$) with the fixed effects alone explaining 53% of this variance ($R^2_m = 0.53$) and the random factor of speakers accounting for additional 10%. There was a significant interaction between the fixed effects in the model ($F(2, 287) = 66.21, p < .01$); therefore, simple effects were calculated with the help of Tukey HSD post-hoc tests.

Table 19

Summary of the Mixed Effects Model for Prominence (Space) with Delayed Post Test

Predictors	Estimates	SE	95% CI	p
(intercept)	0.54	0.01	[0.51; 0.57]	<0.01*
Experimental	-0.14	0.02	[-0.14; -0.11]	<0.01
Pre-test	0.01	0.01	[-0.01; 0.04]	0.27
Post-test	0.10	0.01	[0.07; 0.13]	<0.01
Experimental : Pre-test	0.14	0.02	[0.10; 0.18]	<0.01
Experimental : Post-test	-0.09	0.02	[-0.12; -0.05]	<0.01
Random effect				
σ^2	0.04			
τ_{00}	0.001			
ICC	0.23			
N	38			
Observations	328			

Table 20 includes the results of Tukey HSD tests for the current model. The table indicates several significant differences between the fixed factor levels. As such, in the control group, learners produced significantly more prominent words in the post-test compared to the pre- and delayed tests ($p < 0.01, d = 0.94$ and 1.12 respectively). There was also a significant difference in the use of prominence in the experimental group. In particular, learners in this group produced significantly more appropriate prominence in the post-test and delayed post-test compared to the pre-test with a large size of training effect ($p < 0.01$, Cohen's $d = 1.55$ and 1.61

respectively). There was no significant difference between their production in the post and delayed post-tests suggesting that the training had a long-lasting effect (at least one week) on learners' prominence use.

Table 20

Results of Post Hoc Pairwise Tukey HSD Tests with Delayed Post Test for Prominence (Space)

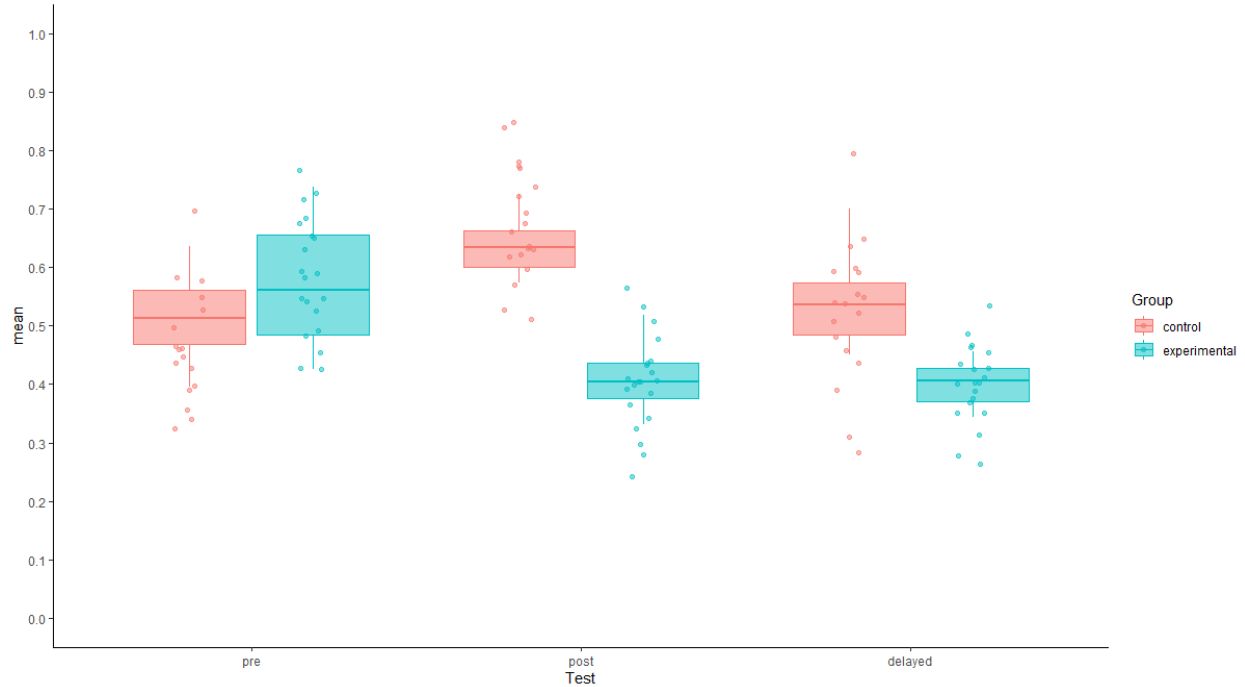
Group contrast	<i>Estimate</i>	<i>SE</i>	<i>df</i>	<i>t</i>	<i>p</i>	<i>d</i>
Control post vs. pre	0.09	0.01	287	5.77	<0.01	0.94
Control delayed vs. pre	-0.01	0.01	287	-1.09	0.51	0.12
Control post vs. delayed	-0.10	0.01	287	-6.84	<0.01	1.12
Experimental post vs. pre	-0.14	0.01	288	-10.53	<0.01	1.55
Experimental delayed vs. pre	-0.15	0.01	287	-11.46	<0.01	1.61
Experimental post vs. delayed	-0.01	0.01	288	-0.87	0.65	0.14
Pre-test control vs. experimental	-0.01	0.02	84.9	0.43	0.67	0.11
Post-test control vs. experimental	0.23	0.02	85.6	12.49	<0.01	2.88
Delayed control vs. experimental	0.14	0.02	85.8	7.81	<0.01	2.30

Finally, there was a significant difference between the experimental and control groups in the post-test and delayed post-test use of prominence. The experimental group used prominence significantly more appropriately with a large effect size in both cases ($p < 0.01$, Cohen's $d = 2.88$ and 2.30 respectively).

Figure 19 further illustrates the trends revealed in post-hoc tests. The overlapping bar plots for the experimental and control groups in the pre-test indicate no significant differences. Contrastingly, the average proportion of stressed words in requests produced by the experimental group in the post-test and delayed post-test is lower compared to the control group.

Figure 19

Prominence Production (Space) in Pre-, Post-, and Delayed Tests



The second LMEM in this group included appropriate use of prominence as a response variable. The model contained a significant interaction of fixed effects ($F[2, 278] = 157.14, p < 0.01$); it also confirmed the trend observed in previous LMEMs with prominence showing that the training explained approximately 48% of variance in learners' appropriateness in sentence stress after the training as indicated by the R^2 value for the fixed factors ($R^2_m = 0.48$). The random effect of speakers explained an additional 8% ($R^2_c = 0.56$). The detailed results of the model are summarized in Table 21 showing that most of the contrasts in the fixed effects are significantly different.

Table 21

Summary of the Mixed Effects Model for Prominence (Appropriate Use) with Delayed Post Test

Predictors	Estimates	SE	95% CI	p
(intercept)	0.53	0.02	[0.50; 0.57]	<0.01
Experimental	0.19	0.02	[0.14; 0.23]	<0.01
Pre-test	0.01	0.02	[-0.03; 0.04]	0.90

Post-test	-0.12	0.02	[-0.16; -0.09]	<0.01
Experimental : Pre-test	-0.29	0.02	[-0.34; -0.24]	<0.01
Experimental : Post-test	0.05	0.02	[0.01; 0.10]	0.03
Random effect				
σ^2	0.04			
τ_{00}	0.002			
ICC	0.15			
N	38			
Observations	328			

Figure 20 illustrates the results obtained from the LMEM analysis. Learners in the experimental group that took the delayed post-test, indicated by the teal color in the figure, showed a continuous improvement in their appropriate use of prominence from the pre-test to the post-test to the delayed post-test. This finding suggests that the effects of the training are long-lasting and are noticeable even a week after. In contrast, control group learners produced significantly less appropriate prominence in the immediate post-test, which could be potentially caused by the cognitive load of the control group training. In the delayed post-test, however, appropriate prominence use increased and reached pre-test levels.

Table 22 provides an additional insights into the observed results via Tukey HSD post hoc tests including estimates and effect sizes for the simple effects in the model. The table confirms the trend suggested in Figure 20 showing that the experimental group consistently demonstrated significant improvements in appropriate tone choice use with a large effect size in the pre – post and pre – delayed comparisons (Cohen’s $d = 1.82$ and 1.98). In contrast, the control group evidently significantly decreased their prominence appropriateness in the immediate post-test (Cohen’s $d = 1.01$); however, it came back to the pre-test levels in the delayed post-test that did not show significant differences compared to the pre-test production ($p = 0.99$, Cohen’s $d = 0.01$).

Figure 20

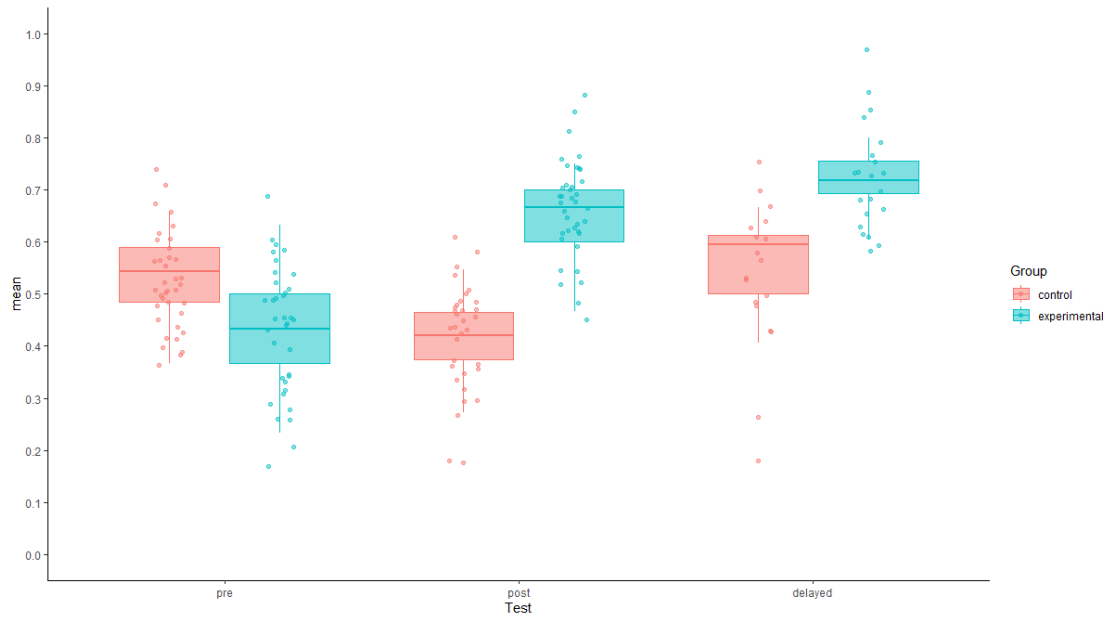
Prominence Production (Appropriate Use) in Pre-, Post-, and Delayed Tests

Table 22

Results of Post Hoc Pairwise Tukey HSD Tests with Delayed Post Test for Prominence (Appropriate Use)

Group contrast	Estimate	SE	df	t	p	d
Control post vs. pre	-0.12	0.02	287	-8.54	<0.01	1.01
Control delayed vs. pre	-0.01	0.01	287	-0.11	0.99	0.01
Control post vs. delayed	0.12	0.02	287	6.56	<0.01	0.77
Experimental post vs. pre	0.22	0.02	288	15.45	<0.01	1.82
Experimental delayed vs. pre	0.29	0.02	287	16.74	<0.01	1.98
Experimental post vs. delayed	0.07	0.02	288	3.99	<0.01	0.47
Pre-test control vs. experimental	0.11	0.02	84.9	6.12	<0.01	1.33
Post-test control vs. experimental	0.24	0.02	85.6	-13.33	<0.01	2.88
Delayed control vs. experimental	0.19	0.02	85.8	-7.94	<0.01	1.71

The findings suggest that learners were able to adjust the amount of prominent words in their requests limiting the stressed words in a thought group only to those that were important for listeners' understanding of intended message in requests. In addition, after receiving the training,

learners are able to stress those words that would be appropriate to highlight for the sake of communicating the pragmatic message successfully to the listener.

Evidence from Perception Surveys

The second part of Research Question 1 (i.e., RQ 1.2) investigated the changes in raters' perception of pragma-prosodic appropriateness in learners' requests. A total of 558 audio recordings of learners' request production were rated by 20 trained raters. In order to collect raters' responses, the audio files were compiled into 20 surveys. Each survey contained 11 "core" audio files that were common for every rater. Scores given to these files were used to calculate inter-rater reliability for the three constructs. The remaining audio files were randomly assigned between the 20 surveys so that each file was evaluated by at least two raters. Thus, each rater evaluated between 55 and 57 sound files in addition to the 11 core sound files.

The raters evaluated learners' request production before and after the training (i.e., pre-, post-, and delayed tests) on three constructs: appropriateness, accentedness, and comprehensibility. Appropriateness was measured via two 5-point rubrics. The first rubric focused on the lexico-grammatical appropriateness of requests, and the second rubric assessed the appropriate use of prosodic features in requests with regards to their directness, politeness, and formality. The two separate rubrics were used to ensure that the raters were not affected by the lexico-grammatical composition of requests when rating their prosodic composition. For the current analyses, only the scores on the second rubric were used because learners only received instruction on the prosodic features of speech acts and not necessarily on their lexico-grammatical composition. Accentedness and comprehensibility in the survey were measured via four 5-point Likert scales. The two accentedness scales assessed the perceived strength of accent (1 = has a strong accent; 5 = has no accent) and difference from L1 English speech (1 = sounds

like a L2 English speaker; 5 = sounds like a native English speaker). Finally, two comprehensibility scales measured the effort (1 = effortful; 5 = effortless) and difficulty in understanding speech (1 = difficult to understand; 5 = easy to understand). Correlations were calculated between the ratings given on the four accentedness and comprehensibility scales to check for the possibility of creating composite variables for the measured constructs. Table 23 presents Pearson correlations and 95% confidence intervals for the four scale ratings. Significant correlations are highlighted in bold font. It is evident from the table that both scales that measured accentedness (*Strength of Accent* and *Degree of Sounding like a L1 Speaker*) as well as the two scales measuring comprehensibility (*Effort* and *Difficulty Understanding*) are strongly related ($r = 0.79$ for accentedness and $r = 0.82$ for comprehensibility). Additionally, the two sets of variables demonstrated strong internal consistency demonstrated through high Cronbach's alpha coefficients (*Effort + Difficulty Understanding*, $\alpha = .90$ [0.89; 0.91]; *Strength of Accent + Degree of Sounding like a L1 Speaker*, $\alpha = .88$ [0.87; 0.89]). Thus, for further analyses, composite variables for accentedness and comprehensibility were created by calculating the mean of the two scores for each sound file rating.

Table 23

Correlations and 95% CI Between Scores on Accentedness and Comprehensibility Scales

	<i>E</i>	<i>DU</i>	<i>SA</i>	<i>NS</i>
<i>E</i>	1	0.82 [0.80; 0.84]	0.37 [0.32; 0.42]	0.32 [0.27; 0.37]
<i>DU</i>		1	0.31 [0.26; 0.36]	0.30 [0.24; 0.35]
<i>SA</i>			1	0.79 [0.76; 0.81]
<i>NS</i>				1

Note. $N = 1055$; *E* = Effort; *DU* = Difficulty Understanding; *SA* = Strength of Accent; *NS* = Degree of Sounding like a L1 Speaker

After composite variables were created, inter-rater reliability was calculated for the three constructs (prosodic appropriateness, accentedness, and comprehensibility) using the core set of

sound files that were assessed by all of the 20 raters. Since the dataset included more than two raters, Fleiss' Kappa inter-rater reliability coefficients (a variation of Cohen's Kappa for more than two raters) were estimated, as summarized in Table 24. Fleiss' Kappa coefficients suggested substantial agreement between the raters in case of appropriateness and accentedness ratings. Comprehensibility scores demonstrated moderate to substantial inter-rater agreement. In the domain of applied linguistics, there is yet to be a guideline for acceptable inter-rater agreement, thus general guidelines provided by Fleiss et al. (1981) had to be applied in interpreting the results. Although the found inter-rater agreement was not perfect (below 0.80), but it was still within the acceptable range to proceed with the data analyses.

Table 24

Fleiss' Kappa Inter-Rater Agreement

Construct	κ	z	p
Appropriateness	0.60	46.4	<0.01
Accentedness	0.61	60.5	<0.01
Comprehensibility	0.56	65	<0.01

Once composite variables were created by averaging the scores on the two subscales in the accentedness and comprehensibility constructs and inter-rater reliability was calculated, Pearson correlations were estimated between Accentedness, Comprehensibility, and Appropriateness. This step was taken to check whether prosodic appropriateness ratings might have been affected by raters' perceptions of accentedness and comprehensibility. Table 25 reports on the correlations and their respective 95% CI showing that there was indeed a significant, albeit weak, relationship between appropriateness and accentedness ($r = 0.27$) and comprehensibility ($r = 0.29$). Significant correlations are indicated in bold font.

Table 25

Correlations and 95% CI Between Accentedness, Comprehensibility, and Appropriateness Ratings

	Accentedness	Comprehensibility	Appropriateness
Accentedness	1	0.36 [0.31; 0.41]	0.27 [0.21; 0.32]
Comprehensibility		1	0.29 [0.24; 0.34]
Appropriateness			1

Note. $N = 1055$

Before running LMEM analyses, descriptive statistics for appropriateness, accentedness, and comprehensibility ratings before and after the instruction were obtained. Table 26 provides the means (M), standard deviations (SD), 95% CI , kurtosis, and skewness values for the scores along the three constructs in pre-, post-, and delayed post-tests for the experimental and control groups. As summarized in the table, kurtosis and skewness values are within the $[-2; 2]$ interval indicating that the data do not deviate from normality in a way that would prevent running the analyses. The table also suggests that the mean scores for appropriateness, accentedness, and comprehensibility in the pre-test performance by the experimental and control groups are comparable. This observation is supported by independent t-tests that were run to ensure that there was no significant difference in the ratings between the two learner groups before the intervention. According to the t-tests, neither of the learner groups was scored significantly higher or lower on the three constructs (Appropriateness pre-test $t(504) = -0.31, p = 0.75$; Accentedness pre-test $t(504) = -0.75, p = 0.45$; Comprehensibility pre-test $t(504) = 0.43, p = 0.67$).

Table 26

Descriptive Statistics for Perceptual Ratings

Experimental	Control

Variable	<i>N</i>	<i>M</i>	<i>SD</i>	95% <i>CI</i>	<i>Kurt</i>	<i>Skew</i>	<i>N</i>	<i>M</i>	<i>SD</i>	95% <i>CI</i>	<i>Kurt</i>	<i>Skew</i>
Approp Pre	231	2.81	1.08	[2.66; 2.94]	-0.49	0.21	276	2.84	1.05	[2.72; 2.97]	-0.71	-0.04
Approp Post	286	2.83	1.15	[2.66; 2.99]	-0.65	0.41	262	2.87	1.06	[2.75; 3.00]	-1.08	-0.21
Approp Delayed	122	2.72	1.10	[2.53; 2.92]	-0.13	0.76	106	2.82	1.05	[2.62; 3.02]	-0.78	-0.14
Accent Pre	231	1.94	1.01	[1.81; 2.07]	-0.05	0.30	276	2.01	1.08	[1.89; 2.14]	-0.67	0.04
Accent Post	286	2.12	1.10	[1.99; 2.25]	-0.46	0.27	262	1.91	0.10	[1.77; 2.04]	-0.35	0.24
Accent Delayed	122	2.35	1.05	[2.17; 2.54]	-0.31	0.13	106	2.04	1.09	[1.83; 2.25]	-0.86	0.39
Comp Pre	231	4.08	0.96	[3.96; 4.21]	0.59	-0.99	276	4.04	1.10	[3.91; 4.16]	0.82	-1.18
Comp Post	286	4.13	1.05	[4.01; 4.25]	1.94	-1.46	262	3.95	1.04	[3.82; 4.07]	0.54	-0.97
Comp Delayed	122	4.45	0.71	[4.32; 4.58]	0.65	-1.15	106	4.15	0.91	[3.98; 4.33]	0.60	-1.10

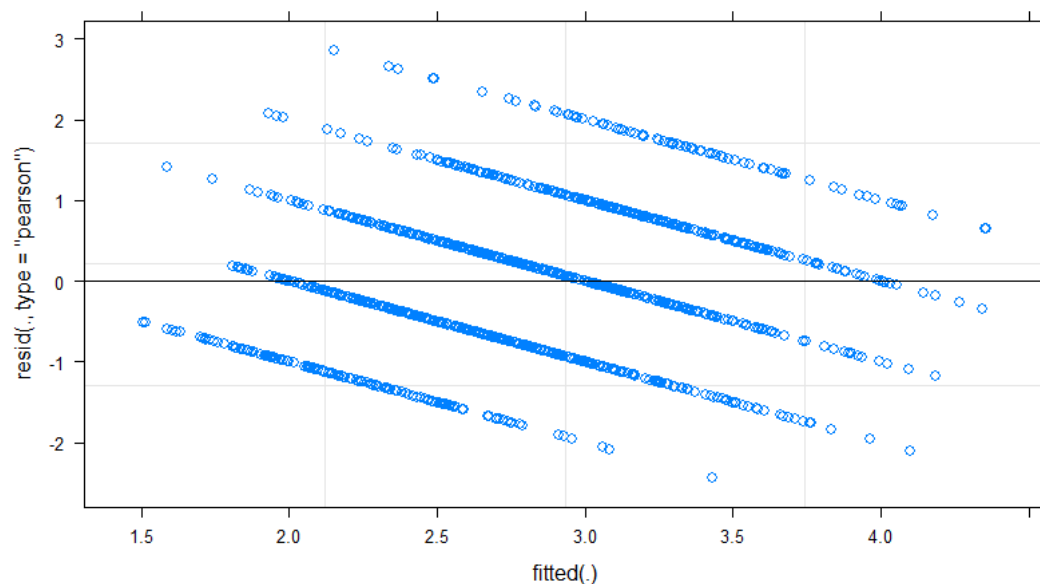
Note. *Approp* = Appropriateness, *Accent* = Accentedness, *Comp* = Comprehensibility

As the next step, a LMEM was built with appropriateness ratings as the response variable. The following variables were entered as fixed factors in the model: Time (pre-, post-, delayed test) and Group (Experimental vs. Control). An interaction was also calculated between these two fixed factors. Accentedness and Comprehensibility were also entered into the model as fixed effects in order to control for these variables. Two random factors were included in the model – Learners and Raters. The model was fitted on the dataset with a total of 7,704 observations (7,698 with missing values excluded). Before interpreting the results of the model, a scatterplot in Figure 20 was created to visually evaluate the residuals for normality, linearity, and homoscedasticity. Additionally, the Shapiro-Wilk normality test was administered for further normality evidence. The test was significant ($W = 0.91, p < 0.01$) suggesting that the data might not have been not normally distributed. However, there was no obvious curve in the distribution,

and the residuals were located symmetrically above and below the 0 line (see Figure 21). The x axis in the plot represents fitted values (i.e. estimated responses) and the y axis represents residuals. *Although* the residuals in the scatterplot were distributed in a diamond shape with more variance in the middle, it is to be expected since the extreme values of a Likert scale are generally used less often than the middle values. The visual examination of the residual vs. fits plot, despite the significant result of the Shapiro-Wilk test, was reassuring, so it was decided to carefully move forward with interpreting the model.

Figure 21

Residuals Plot for Appropriateness Perceptions



The results of LMEM with Appropriateness as the response variable are given in Table 27. The table provides the estimates, standard error values, 95% *CI*, and p values for the fixed factors and interactions. For the random effects, the table provides an unbiased estimated variance of the residual errors (σ^2), as well as variance for each random intercept (τ_{00}), the

intraclass correlation coefficient (ICC) that shows the proportion of the variance explained by the grouping.

Table 27
Results of LMEM for Appropriateness Perception

Predictors	Estimates	SE	95% CI	p
(intercept)	4.24	0.18	[3.89; 4.60]	<0.01
Posttest	-0.07	0.12	[-0.31; 0.17]	0.58
Pretest	0.02	0.12	[-0.22; 0.26]	0.88
Experimental group	0.10	0.14	[-0.17; 0.37]	0.49
Accentedness	0.24	0.03	[0.17; 0.30]	<0.01
Comprehensibility	0.22	0.03	[0.16; 0.29]	<0.01
Posttest : Experimental	0.03	0.18	[-0.32; 0.38]	0.86
Pretest : Experimental	-0.13	0.18	[-0.48; 0.23]	0.48
Random effects				
σ^2	0.19			
τ_{00}	0.22	Learner		
	0.38	Rater		
ICC	0.181			
N	89	Learner		
	20	Rater		
Observations	1282			

The fixed and random factors in the LMEM explained approximately 29% of the variance in appropriateness scores ($R^2_c = 0.29$) while 13% of it was accounted for by fixed effects ($R^2_m = 0.13$). In addition, the explanatory ability of each of the two random effects of learners and raters was calculated showing that while the random effects together accounted for 16% of variance, the majority was attributed to raters (12%) and the remaining 4% of variance was explained by unmeasured learner factors. This finding suggests that in perceptual ratings, and specifically in appropriateness perceptions, individual characteristics of raters (e.g., their L1 background, accent familiarity, and other differences that may not be measured by the instruments in a study) could be at play when raters express their perceptions.

Looking closely at the results in Table 27, one can note that the model did not detect a significant interaction between the Group and Time fixed factors thus permitting to interpret the main effects of the fixed factors. In parallel with the correlations calculated above, the model detected a significant effect of Accentedness and Comprehensibility on raters' appropriateness perceptions. This finding reinforces the earlier observation that raters' perceptions of how accented and comprehensible speech was might have also affected their perceptions of prosodic appropriateness. In other words, those speakers who were rated as more comprehensible and less accented were also more appropriate in their use of prosody in requests (although one should note that the size of the effect in this relationship was rather small as indicated by the correlations in Table 25). In contrast, none of the other fixed factors significantly affected raters' judgements. That is, there was no significant difference in raters' scores before and after the training as well as between the two learner groups.

To further investigate the relationship between prosody and pragmatic appropriateness, a multiple regression was built with tone choices and prominence as predictor variables and appropriateness ratings as the response variable. The goal of this regression was to understand the role these prosodic features play in listeners' appropriateness perceptions. To that end, the variables of interest were entered into a regression model the results of which are reported in Table 28.

Table 28

Results of a Multiple Regression Analysis with Prosodic Features and Appropriateness Ratings

IDs	Est.	SE	Beta (β)	<i>t</i>	<i>p</i>	R^2	Adj.R^2	<i>RI</i>
Intercept	3.43	0.15		22.77	<0.01	0.01	0.01	
TC	-0.07	0.07	-0.03	-1.02	0.31			0.001
PR	-0.58	0.25	-0.07	-2.31	0.02			0.01

Note. *TC* = Tone Choices, *PR* = Prominence

The VIF ($VIF = 1.02$) and Durbin-Watson statistic ($D-W = 1.95, p = 0.44$) values did not indicate the absence of multicollinearity and autocorrelation respectively; thus, the multiple regression results were interpretable. The results indicated that prominence was the only significant predictor of listeners' appropriateness ratings; however, its effect size was minuscule as indicated by the R^2 value of the model as well as the relative importance of the predictor ($RI = 0.01$). Tone choices were not a significant predictor of listeners' appropriateness judgements. Overall, these results indicate that learners' use of these two prosodic features did not sufficiently explain variation in raters' ratings of appropriateness of produced requests.

Summary of Results for Research Question 1

In sum, the LMEM analyses and post-hoc tests conducted to answer the first research question revealed several noteworthy findings. First, regarding tone choices, the analyses suggested that the developed training was moderately effective in helping learners use more appropriate intonation in requests. Specifically, the analyses detected a significant increase in appropriate tone choices in requests produced by learners who received the training, especially in the immediate post-test. While the effect diminished after a week, as indicated by the delayed post-test results, learners still used tone choices significantly more appropriately compared to the pre-test. It is also important to mention that despite the significant differences and moderate effect sizes, the fixed factors explained only 9 – 11% of the changes in the use of intonation. In contrast, the random factor of speakers accounted for between 27% and 37% of variance in tone choice production. This result suggests that there might have been other factors besides the training (i.e., measured and unmeasured individual differences in learners) that have affected

learners' acquisition of tone choices. The control group did not demonstrate any significant changes in tone choices in the post- and delayed post-tests.

With regard to prominence, the analyses also revealed several notable results. In particular, the analyses showed a large effect of the training in improving learners' use of sentence stress in requests. In comparison to tone choices, the training had a much larger effect on prominence as indicated by Cohen's *d* effect sizes as well as *R*-squared values showing that about 53% of changes in the use of prominence could be attributed to the training. Contrastingly to tone choices, the random factor of speakers did not account for a substantial amount of variance in prominence use (9-10%). In addition, the training seemed to have a more lasting impact on prominence since the delayed post-test performance of learners in the experimental group stayed consistently improved even a week after the training. At the same time, learners in the control group did not demonstrate any improvements in the use of sentence stress. In fact, there was a significant worsening of prominence use in the immediate post-test.

Regarding the second part of the research question, the results did not reveal an effect of the training on raters' perceptions of pragma-prosodic appropriateness. However, a few noteworthy observations transpired as a result of this analysis. Although the training did not result in changes in raters' appropriateness perceptions, a significant (but weak) relationship was detected between appropriateness and accentedness and comprehensibility ratings. In addition, tone choices and prominence were not considerable predictors of listeners' appropriateness ratings suggesting that this construct is more fine-grained and may include other pragma-prosodic and linguistic features than those included in the study. These rather unexpected findings are considered in more detail in the discussion chapter of the dissertation.

Research Question 2: Role of Learners' Individual Differences in Their Use of Prosody

The second research question in the study investigated the effect of learners' cognitive IDs on their prosodic development before the training as well as after the training. Specifically, three IDs were considered in the study: musical sensitivity, auditory processing, and phonological working memory. To address the second research question, four multiple regressions were conducted. In the two initial regressions, the effect of learners' cognitive differences was explored with regard to their use of intonation and prominence before the training. The goal of this question was to explore whether learners' IDs had affected learners' development of prosodic features before the training. The other two regressions examined the effect of cognitive IDs on learners' success with the developed training measured via the changes in learners' use of prosody after they completed the training lessons. In all conducted regressions, to control for the potential confounding effect of proficiency, learners' EIT scores were also included in the calculations.

The descriptive statistics for each predictor and response variable in the regression models are summarized in Table 29. The table presents the means (*M*), standard deviations (*SD*), 95% confidence intervals (95% *CI*), as well as kurtosis and skewness values. For auditory processing, musical sensitivity, and phonological working memory, a percent of participants' correct responses (from 0 to 1) was recorded. Most of the variables demonstrated a normal distribution based on the associated kurtosis and skewness calculations except for post-test intonation. Therefore, the variable was log-transformed in further regression analyses. The mean values of cognitive IDs are also of note in the table. The mean values of auditory processing, musical sensitivity, and working memory represent proportions of correct answers in each task. Based on these values, it seems that auditory processing was the most challenging task for the

participants, followed by musical sensitivity where learners were able to recognize on average 45% of the melodies correctly. In the phonological working memory task, learners correctly remembered around 65% of final words in the sentences they heard. Finally, on the EIT task, learners scored on average 56.97 points out of 96 possible points.

Table 29

Descriptive Statistics for Cognitive IDs and Proficiency in the Experimental Group

Variable	<i>N</i>	<i>M</i>	<i>SD</i>	95% <i>CI</i>	<i>Kurt</i>	<i>Skew</i>
AP	37	0.16	0.22	[0.07; 0.23]	0.90	1.53
MS	37	0.45	0.14	[0.40; 0.49]	0.15	-0.78
PWM	37	0.64	0.10	[0.61; 0.67]	-0.09	0.27
EIT	37	56.97	13.23	[52.71; 61.23]	1.86	-0.60
Intonation Pre	37	0.46	0.36	[0.34; 0.58]	-1.26	0.13
Intonation Post	37	0.86	0.24	[0.79; 0.93]	3.43	-1.85
Prominence Pre	37	0.60	0.10	[0.57; 0.63]	-0.81	-0.23
Prominence Post	37	0.41	0.05	[0.40; 0.43]	-0.49	0.42

Note. *AP* = Auditory Processing, *MS* = Musical Sensitivity, *PWM* = Phonological Working Memory, *EIT* = Elicited Imitation Task

Before building regression models, relevant statistical assumptions of multicollinearity and independence of errors were checked. In addition, residual plots were examined to ensure normality, linearity, and equal variance as well as the absence of outliers. First, correlation matrices were created to make sure that no predictor variables were strongly correlated ($r > 0.8$). Table 30 presents the correlation matrix showing that the strongest correlation was between phonological working memory and the EIT score representing learners' proficiency ($r = 0.58$).

Table 30

Correlation Matrix of Predictor Variables

	AP	MS	PWM	EIT
AP	1	0.55 [0.28; 0.74]	0.26 [-0.06; 0.53]	0.27 [-0.05; 0.54]
MS	0.55	1	0.44 [0.14; 0.67]	0.17 [-0.16; 0.46]
PWM	0.26	0.44	1	0.08 [-0.25; 0.39]

EIT	0.27	0.17	0.08	1

Note. *AP* = Auditory Processing, *MS* = Musical Sensitivity, *PWM* = Working Memory, *EIT* = Elicited Imitation Task

In addition, VIF values were calculated for additional lack of multicollinearity evidence. As shown in Table 31, no VIF values were outside of the > 2.5 and < 4 interval indicating that none of the response variables were strongly correlated.

Table 31

Predictor Variable VIF Values

Variable	Tolerance	VIF
AP	0.67	1.49
MS	0.60	1.65
PWM	0.80	1.24
EIT	0.93	1.08

Note. *AP* = Auditory Processing, *MS* = Musical Sensitivity, *PWM* = Phonological Working Memory, *EIT* = Elicited Imitation Task

Individual Differences and Use of Prosody Before Training

The two initial multiple regressions examined the role of cognitive IDs and learners' proficiency in their acquisition of prosody before the training. In these regressions, the IDs of auditory processing, musical sensitivity, working memory, and proficiency (EIT) were entered as predictor variables. The continuous response variables were appropriate tone choices and prominence before training. Before the results of the regression models were interpreted, a few additional assumptions were checked. First, Durbin Watson tests were run with the two response variables to ensure absence of autocorrelation in the residuals. In both tests, the D-W statistic was within the <1 and >3 boundary indicating that observations were independent of each other (D-W = 2.09 and 2.18 for tone choices and prominence respectively). In addition, the residual plots did not suggest considerable deviations from normality, linearity, and homoscedasticity in

the two response variables in the regression. The dots were mostly equally distributed on either side of the diagonal dotted line, there was no funneling or curving of the dots in the scatterplot, meeting the abovementioned assumptions. Finally, the data were checked for outliers. Following Tabachnik and Fidell (2007), the outlier cutoff for four predictor variables was established at the Mahalanobis distance of 18.47. Only one outlier with the Mahalanobis distance of 19.09 was found and winsorized.

Individual Differences and Tone Choices Before the Training. Table 32 presents the results of a multiple regression with learners' use of intonation before training as the response variable. Besides the estimates and standard errors for each predictor, the table includes standardized beta coefficients that estimate how much change in tone choice is explained by every 1-unit change in the predictor variables. In addition, t and p values are included as well as R^2 and adjusted R^2 values that represent how much variance in tone choice use is explained by the predictor variables in the model. Finally, the table includes relative importance (RI) values showing the amount of variance explained by each predictor.

Table 32

Results of a Multiple Regression Analysis with Tone Choices Before Training

IDs	<i>Est.</i>	<i>SE</i>	Beta (β)	<i>t</i>	<i>p</i>	R^2	Adj.R^2	<i>RI</i>
Intercept	0.40	0.48		0.83	0.41	0.09	0.02	
AP	0.44	0.34	0.27	1.31	0.20			0.06
MS	0.04	0.55	0.01	0.07	0.95			0.01
PWM	-0.24	0.68	-0.06	-0.35	0.73			0.01
EIT	0.01	0.01	0.08	0.44	0.66			0.01

Note. *AP* = Auditory Processing, *MS* = Musical Sensitivity, *PWM* = Phonological Working Memory, *EIT* = Elicited Imitation Task

As shown in Table 32, the predictors in the model explained only 9% of variance in tone choices. Moreover, the adjusted R^2 value is even lower at 0.2, which suggests that the selected cognitive individual differences did not sufficiently explain the variance in learners' intonation

use before the training. In other words, learners' auditory processing, musical ability, working memory, and proficiency did not affect their ability to use tone choices on the pre-test.

Individual Differences and Prominence Before Training. The next regression built with the data included the same predictor variables; however, this time, prominence use before the training was entered as the response variable. The results of the regression are given in Table 33. The table shows that, in parallel with the previous regression that examined tone choice use, none of the selected cognitive IDs as well as proficiency significantly affected learners' use of prominence before the training. The model was able to explain only 4% of variance in prominence use (adjusted $R^2=0.04$).

Table 33

Results of a Multiple Regression Analysis with Prominence Before Training

IDs	<i>Est.</i>	<i>SE</i>	Beta (β)	<i>t</i>	<i>p</i>	R^2	Adj. R^2	<i>RI</i>
Intercept	0.54	0.13		4.20	<0.01	0.08	0.04	
AP	-0.01	0.09	-0.01	-0.07	0.95			0.01
MS	-0.18	0.15	-0.27	-1.25	0.22			0.04
PWM	0.13	0.18	0.14	0.73	0.47			0.01
EIT	0.01	0.01	0.15	0.88	0.38			0.02

Note. *AP* = Auditory Processing, *MS* = Musical Sensitivity, *PWM* = Phonological Working Memory, *EIT* = Elicited Imitation Task

Thus, the findings showed that learners' aptitude did not play a significant role in their acquisition of pragma-prosodic features before the training. It is likely that this process mostly involved implicit acquisition where learners were not explicitly exposed to explanations about the use of prosody in specific speech acts. The findings potentially suggest that without explicit training, learners' cognitive IDs do not substantially affect their pragma-prosodic learning.

Individual Differences and Use of Prosody After Training

In order to address the second half of Research Project 2, two additional multiple regression were built. In contrast to the first two regressions, in these analyses, the response variables included learners' use of intonation and prominence *after* the developed training. Before interpreting the results of multiple regressions, Durbin Watson test statistics were calculated and residual plots were visually examined. The D-W statistics did not exceed the acceptable interval of <1 and >3 (D-W = 1.00 and 2.9 for tone choices and prominence respectively) suggesting absence of autocorrelation. The residual plots also did not reveal significant violations of linearity, normality, and homoscedasticity. It is important to acknowledge, however, that the location of the dots and a bend in the regression line in the residuals plot with log-transformed tone choices indicated a possible deviation from the assumptions. However, this bend can be explained by the interval nature of the tone choice variable (log-transformed) since it represents a proportion of appropriately used tone choices in three requests.

Individual Differences and Tone Choices After the Training. Table 34 summarizes the results of the multiple regression model with log-transformed tone choice use after the training as the response variable. Similar to previous analyses, four predictors were included in the model: auditory processing, musical sensitivity, phonological working memory, and proficiency (measured by EIT). The model was able to explain a relatively large proportion (~42%) of variance in learners' application of intonation after the training (adjusted $R^2=0.42$). The majority of this variance was explained by the significant predictor of musical sensitivity as indicated by its relative importance (RI) index (~31%). PWM did not significantly predict tone choice use post-training; however, it explained an additional 10% of variance. Finally, auditory processing accounted for 7% of variation in appropriate intonation use. These results imply that

learners with a more advanced ability to recognize and differentiate between simple musical melodies as well as with a more developed capacity to store auditory input benefitted from the training more readily.

Table 34

Results of a Multiple Regression Analysis with Log-Transformed Tone Choices After Training

IDs	<i>Est.</i>	<i>SE</i>	Beta (β)	<i>t</i>	<i>p</i>	R^2	Adj.R^2	<i>RI</i>
Intercept	-0.43	0.12		-3.50	<0.01	0.48	0.42	
AP	0.03	0.09	0.04	0.27	0.79			0.07
MS	0.50	0.14	0.57	3.51	<0.01*			0.31
PWM	0.23	0.18	0.18	1.31	0.20			0.10
EIT	<-0.01	<0.01	-0.02	-0.20	0.84			<0.01

Note. *AP* = Auditory Processing, *MS* = Musical Sensitivity, *PWM* = Phonological Working Memory, *EIT* = Elicited Imitation Task

Individual Differences and Prominence After the Training. The final multiple regression to address Research Question 2 included learners' use of sentence stress (i.e., prominence) in requests after the training. The results of the model are summarized in Table 35. The model showed that ~33% of variation in learners use of prominence after the training was attributed to the selected predictor variables (adjusted $R^2=0.33$). However, a further examination of results showed that the only significant predictor in the model was learners' phonological working memory that accounted for ~26% of the variance. Furthermore, musical sensitivity, while not a significant predictor, explained ~13% of the variance. The negative beta coefficient associated with phonological working memory suggests that those learners who used prominence more appropriately (i.e., stressed *fewer* words in a request) also had more advanced working memory.

Table 35

Results of a Multiple Regression Analysis with Prominence After Training

IDs	<i>Est.</i>	<i>SE</i>	Beta (β)	<i>t</i>	<i>p</i>	R^2	Adj.R^2	<i>RI</i>

Intercept	0.64	0.06		11.14	<0.01	0.41	0.33	
AP	0.05	0.04	0.23	1.37	0.18			0.02
MS	0.12	0.06	-0.34	-1.95	0.06			0.13
PWM	0.26	0.08	-0.48	-3.20	<0.01*			0.26
EIT	<0.01	<0.01	-0.01	-0.10	0.91			<0.01

Note. *AP* = Auditory Processing, *MS* = Musical Sensitivity, *PWM* = Phonological Working Memory, *EIT* = Elicited Imitation Task

Summary of Results for Research Question 2

The analyses conducted to address the second research question in the study revealed several noteworthy findings regarding the role of learners' cognitive IDs and proficiency in the learning of prosodic features and the effectiveness of the developed instructional training. Specifically, the analyses revealed virtually no effect of auditory processing, musical sensitivity, phonological working memory, or proficiency in learners' applications of intonation in requests produced before the training. However, learners' use of intonation and prominence was explained by their phonological working memory. Those with more advanced PWM skills demonstrated more appropriate use of these prosodic features. Additionally, learners' intonation choices after they completed training lessons were largely accounted for by their musical sensitivity. That is, those learners who could more accurately differentiate between musical melodies also used intonation more appropriately after the training.

Research Question 3: Ability to Perceive Intonation Cues and Training Effectiveness

The third and final research question in the study examined the relationship between learners' ability to perceive prosodic features in perception-based exercises and the observed changes in their use of intonation and prominence. In particular, the analyses conducted to address this question investigated whether learners with more precise perception skills might demonstrate larger improvements in their use of tone choices and sentence stress after the training.

In order to address the research question, two multiple regressions were performed. The predictor variables in the models were learners' scores on four perception-based tasks in the training. Two of these tasks were tone choice and prominence categorization tasks that required learners either to identify the tone choice they heard in a request (rising, falling, or level) or to identify all the prominent words in a request. The other two perception exercises were tone choice and prominence discrimination tasks where learners were asked to listen to two requests and choose one with more appropriate tone choice or prominence. Thus, each predictor represented learners' scores on each of these tasks. Each model included one response variable: proportion of appropriate tone choices or prominence after the training.

Table 36 summarizes the descriptive statistics for the four predictor variables. The mean values for each predictor indicate the percent of times when learners gave a correct response in a perception-based task from the first attempt. In each task, learners completed 10 items. As shown in the table, learners correctly identified prominent words on average in 30% of the items. In contrast, in the tone choice discrimination task, learners correctly discriminated between appropriate and inappropriate intonation in requests in approximately 80% of items.

Table 36

Descriptive Statistics for Perceptual Predictor Variables

Variable	<i>N</i>	<i>M</i>	<i>SD</i>	<i>95% CI</i>
Tone choice categorization	37	0.57	0.08	[0.54; 0.60]
Prominence categorization	37	0.32	0.22	[0.25; 0.39]
Tone choice discrimination	37	0.82	0.20	[0.76; 0.88]
Prominence discrimination	37	0.76	0.19	[0.70; 0.82]

Before regressions were interpreted, the assumptions of linearity, multicollinearity, normality, absence of outliers, equal variance, and autocorrelation were checked. As evidenced

from the VIF values in Table 37, the predictor variables did not demonstrate multicollinearity. None of the variables were correlated at higher than $r = 0.19$, supporting the absence of multicollinearity (see Table 38).

Table 37

Perceptual Predictor Variable VIF Values

Variable	Tolerance	VIF
TC	0.96	1.04
PC	0.93	1.07
TD	0.97	1.03
PD	0.96	1.04

Note. *TC* = Tone choice categorization, *PC* = Prominence categorization, *TD* = Tone choice discrimination, *PD* = Prominence discrimination

Table 38

Correlation Matrix of Perceptual Predictor Variables

	TI	PI	TD	PD
TC	0.19	1	0.02	0.07
PC	1	0.19	0.16	0.10
TD	0.10	0.07	0.15	1
PD	0.16	0.02	1	0.15

Note. *TC* = Tone choice categorization, *PC* = Prominence categorization, *TD* = Tone choice discrimination, *PD* = Prominence discrimination

However, the results of the Durbin Wilson test suggested a risk of autocorrelation in the model with the response variable of tone choices (D-W = 0.68, $p < 0.01$). In order to correct the autocorrelation, the Cochrane-Orcutt estimation was used to transform the data. After the transformation, the autocorrelation was eliminated (D-W = 1.55, $p = 0.28$). The other regression with prominence as the response variable met the autocorrelation assumption without any transformation (D-W = 1.96, $p = 0.82$). Finally, residuals plots for each of the response variables in the multiple regressions were visually examined to evaluate the remaining assumptions. There was no funneling or curve in the location of the residuals. Given that the assumptions for a

multiple regression analysis were mostly met, it was deemed appropriate to interpret the results of the multiple regressions. Importantly, the regression with tone choices as the response variable was re-ran after data transformation discussed above.

The results of the first regression conducted to address Research Question 3 are given in Table 39. Similar to the previous multiple regressions, the table provides estimates and standard errors for each predictor variable as well as standardized beta coefficients, t and p values, R^2 and adjusted R^2 values, and relative importance values. The results of this multiple regression indicate that the model was able to explain ~20 % variance in learners' use of intonation after the training. In particular, tone choice discrimination accuracy was a significant predictor in the model. This result suggests that those learners who were able to identify more accurately a request with appropriate tone choices also demonstrated more appropriate use of intonation in the post-test. Besides tone choice discrimination, none of the other predictors were significant in the model.

Table 39

Results of a Multiple Regression Analysis with Tone Choices After Training

IDs	Est.	SE	Beta (β)	t	p	R^2	Adj.R^2	RI
Intercept	0.17	0.30		0.55	0.59	0.29	0.20	
TC	0.26	0.43	0.09	0.60	0.55			<0.01
PC	-0.26	0.17	-0.24	-1.54	0.13			0.04
TD	0.54	0.18	0.19	2.96	<0.01*			0.20
PD	0.24	0.19	0.45	1.22	0.23			0.04

Note. *TC* = Tone choice categorization, *PC* = Prominence categorization, *TD* = Tone choice discrimination, *PD* = Prominence discrimination

Table 40 contains the results of the second multiple regression. The adjusted R^2 value indicates that the model accounted for 16% of variance in learners' use of prominence after the training. In this model, prominence discrimination accuracy was the only significant predictor, and it explained around 14% of the variance. That is, learners that were more accurate on

prominence discrimination tasks (i.e., choosing a request with more appropriate prominence in a pair of requests) also demonstrated more appropriate prominence use after the training. No other predictors made significant contributions to the model; however, tone choice discrimination accuracy made a substantial contribution to the explanatory power of the model.

Table 40

Results of a Multiple Regression Analysis with Prominence After Training

IDs	<i>Est.</i>	<i>SE</i>	Beta (β)	<i>t</i>	<i>p</i>	<i>R</i>²	Adj.<i>R</i>²	<i>RI</i>
Intercept	0.60	0.07		8.74	<0.01	0.25	0.16	
TC	-0.09	0.10	-0.14	-0.89	0.38			0.02
PC	<0.01	0.03	0.02	0.14	0.89			<0.01
TD	-0.07	0.04	-0.27	-1.72	0.09			0.09
PD	-0.10	0.04	-0.36	-2.30	0.03*			0.14

Note. *TC* = Tone choice categorization, *PC* = Prominence categorization, *TD* = Tone choice discrimination, *PD* = Prominence discrimination

Summary of Results for Research Question 3

To answer the third research question in the study, two multiple regressions were conducted. Each regression estimated the effect of learners' ability to perceive intonation and prominence in categorization and discrimination exercises on their use of these prosodic features after the training. The results of the multiple regressions revealed several noteworthy effects of learners' perception ability on their prosodic use after the training. In particular, learners' ability to discriminate between more and less appropriate intonation in perception tasks significantly predicted their proper use of intonation in requests. In other words, those participants who successfully discriminated between appropriate and inappropriate intonation from the first attempt also used more suitable intonation after the training. In a similar vein, participants who were able to discriminate between more and less appropriate sentence stress in requests also used this prosodic feature more appropriately in post-training requests. However, tone choices and

prominence categorization (i.e., identifying the main tone choice in a request or identifying all of the prominent words in a request) were not significant predictors in the models. These findings suggest that learners' perceptual ability explained to some extent their success in pragmatic prosodic production after the training; however, it appears that the type of perception (discrimination vs. categorization) played a role in this relationship.

CHAPTER 5: DISCUSSION

This dissertation study had three theoretically and practically oriented aims. At the heart of the study was the notion of pragma-prosodic appropriateness as an essential feature of interaction (Brazil, 1997; Kang *et al.*, 2021). Both pragmatics and prosody are aspects of interaction that develop later in the process of L2 acquisition and oftentimes require explicit instruction (e.g., Bardovi-Harlig, 2013; Graham & Post, 2018). Thus, the first aim of the study was to develop and test the effectiveness of a web-based training that would allow learners to practice certain prosodic features of intonation, namely tone choices and prominence, in hopes that these features would be used appropriately in learners' future interactions. Practically speaking, the second aim of the study was to examine the effectiveness of specific elements in the training (i.e., perception-based exercises) in changing learners' use of prosodic patterns in requests. These practical goals of the study were complemented by a third, more theoretically driven aim. This aim was to measure the role of learners' aptitude (measured by phonological working memory, musical sensitivity, and auditory processing) for learning pragma-prosodic features of speech. Although this last aim was motivated mostly by the existing theories of L1 and L2 intonation acquisition, it had practical implications for developing pragma-prosodic trainings similar to one in this study. Currently, few studies have examined instructional methods to teach prosodic features, and even fewer have contextualized these methods in real-life communicative contexts. In addition, there is an apparent lack of a clear understanding of the relationship between learner characteristics and pragma-prosodic acquisition. This dissertation study was designed to inform the current pragma-prosodic instructional practices and to expand the existing body of literature in this domain in three novel ways: (1) a web-based instructional training, (2) measured learners'

aptitude and its role in the training effectiveness, and (3) investigation of the effectiveness of perception-based exercises in the training.

The study followed a quasi-experimental pre-post design. Seventy-four English learners from China were randomly assigned to the instruction (experimental) and control group. Learners' use of prosody in requests before and after instruction was recorded through video-based role plays, analyzed acoustically, and evaluated for appropriateness by 20 trained raters. Learners in the instruction group also completed three cognitive measures that represented their aptitude for acquiring pragma-prosodic speech features: (1) phonological working memory, (2) musical sensitivity, and (3) auditory processing. All learners completed a background survey before the instruction, and the instruction group also completed a reflection about perceived instructional effectiveness at the end of the training. Data were analyzed descriptively as well as through several linear mixed effects (LME) and multiple regression models.

This discussion chapter is organized parallel to the previous chapter. For each research question, the major findings are summarized, explained, and contextualized within relevant existing research. The first research question focused on the effectiveness of the developed training as evidenced by the acoustic analysis and raters' perceptions of learners' prosodic use. Learners' qualitative responses in a final survey are also incorporated to provide additional insights in discussing the results of RQ 1. The second research question asked about the role of learners' aptitude in the observed pragma-prosodic changes. Finally, in the third question, learners' ability to perceive intonational features in perception-based exercises are related to their use of prosody after the training.

Research Question 1: Learners' Use of Prosody Before and After the Training

The first research question in this dissertation study was to examine changes in learners' use of prosody in high-imposition role plays as a result of the developed pragma-prosodic training. This question was examined from two complementary perspectives. First, learners' use of tone choices and prominence before and after the instruction was acoustically analyzed in an objective manner using speech analysis software PRAAT (Boersma & Weenink, 2021). Second, learners' production was evaluated by 20 trained raters, providing a more subjective judgement of the prosodic changes that might have taken place.

The primary, overarching motivation for this question was to provide evidence for the effectiveness of the developed instructional training, especially since instructional materials that focus on prosodic aspects of interaction are largely missing in teaching materials and classroom instruction (Nikolić, 2018). An additional motivation was to add to the research in the domain of L2 *pragma*-prosodic instruction specifically, bringing the fields of L2 pronunciation and pragmatic instruction together. Indeed, at the time of writing this dissertation, there were very few empirical studies in which prosodic instruction was contextualized within pragmatic constructs.

The analysis for this question was based on the hypothesis that learners' use of tone choices and prominence will become more appropriate after the instruction. In other words, learners' production of high imposition requests with the *I was wondering if.. / I was hoping that...* formula would contain fewer rising and more falling tones on the main request strategy, which was the focus of the current study. This hypothesis was supported both by the Discourse Intonation theoretical framework adopted for the study (Brazil, 1997) and by the baseline speakers' use of tone choices. Regarding prominence, the learners were hypothesized to stress

fewer words and place stress more strategically in discourse. These changes would indicate that learners increased their mastery of prosodic features for conveying pragmatic meaning in requests and consequently became more effective in conveying a pragmatic message (Staples, 2015). On the perceptual side, the study hypothesized that speech acts produced with rising intonation on the main strategy and excessive prominence would be rated as less appropriate (i.e., not appropriately direct, polite, and formal). In addition, accentedness and comprehensibility ratings were predicted to be positively associated with the appropriateness ratings. In other words, requests that were rated more appropriate would also be more comprehensible and less accented. The remainder of this section will discuss the findings by the two sub-questions in RQ1.

Research Question 1.1: Acoustic Analysis of Learners' Use of Prosody

Learner data were modeled in several LMEMs. In the first model, tone choices on the main request strategy were entered as a response variable. In the second model, the response variable was space (i.e., the proportion of stressed words over a total number of words in a request). There were two additional LMEMs built with these response variables; these models also included data from the delayed post-test taken by a sub-group of learners. The fixed effects in the models were learners' group (experimental vs. control) and time of data collection (pre-, post, and delayed).

Findings revealed that learners in the experimental group used significantly more appropriate tone choices compared to the control group and also to their own pre-test performance. In addition, this positive trend in learners' tone choices continued in the delayed post-test where learners still demonstrated significantly more appropriate use of tone choices compared to the pre-test, although it was not as high as immediately after the training.

Approximately 36 - 48% of change in learners' use of tone choices was explained by the training (9 - 11%) as well as by individual learner characteristics not captured by the fixed factors (27 - 37%). Post hoc tests and individual effect size calculations showed that the training mostly had a medium-sized effect on learners' ability to use appropriate tone choices in requests.

A similar trend was found in learners' use of prominence before and after the training. After the training, learners in the experimental group stressed significantly fewer function words, focusing instead on content words. In previous research, lack of excessive sentence stress been associated with discourse functions of prosody potentially leading to clarity and effectiveness in differentiating between given / new information in communication (Brazil, 1997; Kang *et al.*, 2010; Pickering, 2018; Staples, 2015; Taguchi *et al.*, 2021). In fact, in contrast with tone choices, this positive trend in learners' use of prominence persisted even a week after the training. The models built to measure changes in learners' use of prominence showed that the developed training accounted for over half of the changes in learners' prominence use (53%). In addition, 9 – 10 % of the variance was accounted for by learner factors extraneous to the current study. Prominence features turned out to be more prone to change after the training, which was supported by large effect sizes observed throughout.

The difference in variance explained by the random factor of speakers in tone choice and prominence use is noteworthy. It appears that while improvements in prominence can be attributed mostly to the developed training, for tone choices, individual factors of speakers played a large role in the observed changes. In other words, changes in learners' use of tone choices seem to have occurred not only due to the instructional training but to a large extent due to other factors in the learner sample. Some of these factors may include cognitive individual differences of learners or other characteristics that were not measured in the study (e.g.,

motivation, age, etc.). This finding implies that the acquisition of tone choices is not a straightforward process and that instructional trainings may need to consider how individuals may respond to such trainings due to their individual differences.

The results in RQ 1.1 supported the study hypothesis and showed that prosodic features can indeed be taught effectively via explicit perception- and production-based instructional training delivered online. The results complement existing empirical research in L2 pronunciation instruction that has demonstrated that explicit instruction in specific prosodic features including tone choices and prominence can be successful in improving learners' use of these features (e.g., Gordon & Darcy, 2016; Saito & Saito, 2017). This finding is also supported by synthetic studies that confirmed the effectiveness of pronunciation instruction centered around suprasegmentals like intonation and prominence (Lee *et al.*, 2015; McAndrews, 2019). However, while most of the existing studies have focused on learners' use of prosodic features in relation to their comprehensibility (Derwing & Rossiter, 2003; Saito & Saito, 2017; Zhang & Yuan, 2020) and intelligibility (Chau *et al.*, 2022; Dalman, 2022; Hu *et al.*, 2022), the results of the acoustic analysis in the current study imply that in addition to comprehensibility and intelligibility, learners' pragmatic appropriateness may improve as a result of targeted prosodic training.

Acoustic analysis results also provide support for the findings of a few previous instructional studies that centered around prosodic training in specific pragmatic contexts. For example, in parallel with the current findings, McKinnon (2017) also found that pragma-prosodic instruction can be helpful in changing learners' pitch range, prominence, and tone choices in declarative and imperative sentences. Similar positive results have also been observed in other

studies at the intersection of pragmatics and prosody (e.g., Tareva & Polushkina, 2020; Ramírez-Verdugo, 2006).

Based on previous primary and synthetic studies of pronunciation instruction, the effectiveness of the developed training may be explained by a few instructional design features. First, several studies have noted that providing both perception- and production-based activities is associated with immediate and delayed gains in suprasegmental use (e.g., Lee *et al.*, 2020). Furthermore, secondary research has suggested that trainings that include controlled outcome measures (Lee *et al.*, 2015) and focus on specific suprasegmental features as opposed to general prosody (Saito & Plonsky, 2019) are associated with larger learners' gains. All of these training features were included in the current study indicating that their use in the training was beneficial for improving learners' prosody and measuring those improvements.

Besides the aforementioned features, another aspect that may have impacted the effectiveness of the training is the use of visual aids including prosodic contours that instantly visualized intonation and prominence in the training production exercises. Several studies have noted that displaying pitch contours of auditory models side-by-side with contours of learner's utterance may help a learner more readily match their intonation with the target (e.g., Hincks, 2015) and provide additional visual feedback opportunities (Harrison, 2004; Jiang & Chun, 2021; Zhao *et al.*, 2020). While these studies investigated the use of pitch contour displays in instruction that centered mostly around comprehensibility, this dissertation study expands the range of instructional contexts demonstrating that this visual element in training could be effective in training learners' prosody for pragmatic appropriateness.

Another instructional feature that might explain the success of the developed training is the hybrid mode of the last activity (free production) that was completed via Zoom. This activity

offered learners an opportunity to apply their pragma-prosodic in a live conversation and hear immediate feedback on their production. During live sessions, learners were also able to clarify any remaining questions with the researcher about appropriate use of tone choices and prosody in requests. Indeed, previous educational research as well as SLA has found that hybrid instruction can be just as effective, if not more effective, than fully asynchronous learning (Dixon et al., 2021; O'Brien et al., 2011).

Instructional features discussed above also appeared in learners' responses to the final questionnaire (Appendix I). After the experimental group completed the training, they were asked to complete a short questionnaire that contained six open-ended questions. The first three questions focused on the platform itself: (1) *Do you think the platform is useful for practicing your intonation skills?* (2) *What features of the training did you find useful?* (3) *What drawbacks did you notice when working with the training?* In addition, three self-reflection questions were posed to the learners: (1) *What do you feel you have accomplished in terms of your intonation in English?* (2) *Do you feel you can apply what you have learned in the training to more natural and spontaneous conversations? Why and how?*, and (3) *What do you feel you would need to practice more now?* Learners' responses to the questions were coded for common themes to gain insights into learners' opinions about the training and its perceived effectiveness.

First, learners' responses to the first three questions were thematically analyzed. This examination showed that the majority of the 37 learners in the instructional group found the rating useful overall (77%), while some of them expressed doubts (16%). Around 6% of the learners did not find the training useful. In terms of specific features of the training, thematic coding showed that the following elements in the training were particularly salient and appreciated by the learners: visual pitch contour displays (45%), real-life application of the

specific constructs taught (36%), videos with explicit instruction (27%), the ability to take multiple attempts in each question (14%), and instructor feedback (14%). A sample response below illustrates the emerged themes:

I think the most important thing is that I started to pay attention to my own pitch and accent each time I pronounced it, and compared to the first test, I can clearly see my progress in the software of PRAAT. <...> I was impressed by the teacher's feedback <...> she told me to pay attention to the stress of a word, which should be the best in 2-3 [prominent words per thought group]. (EG2).

These themes further illustrate that the instructional design decisions made during training development were beneficial for learners' pragma-prosodic acquisition.

Regarding challenges that learners experienced during the training, the most common theme was technical issues related to stable internet connection to access the instructional platform. Several learners also noted that the instruction itself was different from what they were used to, especially in the context of English learning in China. Thus, one learner stated that “Chinese students do more activities or works on paper by answering some questions. But the activities we do used a lot of technologies and we are unfamiliar with the process” (EG14). Another learner noted the following:

Because it was the first time I came into contact with this type of tests, I was a little uncomfortable and dizzy for this platform. In the last few tests, I found that the website included some guidance to correct our intonation errors and so on. For example, the role plays in the [pre-test] and the [post-test]. We can obviously feel our own changes and progress through this comparison. (EG 35).

These reflections suggest that while web-based training like the one developed in this study was a new form of instruction, not previously experienced, the learners were still able to notice their own improvement in the use of prosodic features in the post-test production.

There were also interesting reflections about the three self-assessment questions in the final survey. Over 75% of learners in the instruction group agreed that they were able to apply what they learned in the training to naturally occurring, spontaneous conversations. One learner even mentioned that they have “*already applied it since the first lesson in this module*” (EG29). Among features that learners thought they improved most, one learner noted improvements not only in their prosodic production, but also perception, stating that they are “*able to recognize the tone of a speaker in a sentence, and it is easier to understand the request through the speakers’ intonation*” (EG09). In addition, several learners commented on their improved understanding of the function of level intonation (18%) and ability to make more prosodically appropriate requests overall (20%).

The findings in RQ 1.1 offer a few takeaways. First, it appears that pragma-prosodic features of spoken interaction can indeed be taught via web-based training, at least within the sample of learners in the study and the population it represents (Chinese EFL learners in university contexts). It is also important to note that in this study, the pragma-prosodic training was centered around mitigated preparatory request strategies with a bi-clausal structure (*I was wondering if... / I was hoping that...*). As explained in earlier chapters, this grammatical structure is frequently associated with falling intonation. However, it is possible that with a different request strategy that has the form of a question rather than a statement (e.g., query preparatory, *Could you...?*), the intonation pattern is likely to change. Thus, the results for this research question should only be generalized for this particular strategy.

Nevertheless, providing learners with explicit instruction about chosen features, coupled with ample opportunities for perception- and production-based practice, visualization of prosodic contours, and feedback seem to be effective in helping learners' master suprasegmental features not only in the context of comprehensibility and intelligibility, but also with regards to prosodic features and their pragmatic meaning. Importantly, these findings are supported by objective acoustic analyses of learners' production. While this approach is an important piece of evidence for instructional effectiveness, consideration should also be given to perceptual evaluations of learners' appropriateness. The following section addresses this sub-question (RQ1.2).

Research Question 1.2: Perceptual Analysis of Learners' Use of Prosody

The second part of Research Question 1 examined changes in learners' use of tone choices and prominence from the perceptual perspective. Twenty trained raters listened to learners' requests before and after the training and rated them for pragma-prosodic appropriateness, accentedness, and comprehensibility. Raters' responses were modeled using LMEM analysis. The fixed effects in the model were learners' Group (experimental vs. control), Time (pre-, post-, and delayed post-tests), as well as accentedness and comprehensibility scores. Learners and Raters were entered in the model as random slopes. Finally, pragma-prosodic appropriateness ratings were the response variable in the model.

Findings revealed that there was no difference between raters' perceptions of pragma-prosodic appropriateness before and after the training. In fact, the ratings remained virtually the same for the instruction group in all three tests. Similarly, there was little to no difference in the ratings between the control and experimental group. However, the model revealed a significant (albeit small) positive relationship between raters' appropriateness scores and their evaluations of learners' accentedness and comprehensibility. Learners who were rated more pragma-

prosodically appropriate were also perceived to be more comprehensible and less accented. The model explained around 29% of variance in learners' appropriateness ratings. The fixed effects together accounted for 13% of this variance while the random effects accounted for the rest of explained variance (4 % explained by learners and 12% explained by raters). The large percentage of appropriateness perceptions explained by rater factors is notable especially in relation to the observed lack of differences in appropriateness ratings before and after the instruction. It appears that there may be individual rater factors that affected the ratings. These factors may relate to raters' ability to hear intonation cues as well as their degree of understanding of the prosodic elements of requests.

The lack of instructional effect on raters' appropriateness perceptions was a somewhat unexpected result given that similar previous studies found that suprasegmental and pragma-prosodic training can be effective in improving listeners' impressions of L2 prosody and overall appropriateness (Derwing & Rossiter, 2003; Ramírez-Verdugo, 2006). In this study, results of subjective listener ratings were different in comparison to objective acoustic analyses that revealed significant changes in learners' appropriate use of tone choices and prominence. This perceptual finding may be explained by the complex nature of human judgements. While acoustic analyses offer a straightforward, objective measure of prosody, human judgements, especially in the realm of pragmatics, may include a variety of factors not accounted for by a mere examination of learners' prosody. Previous studies have shown that that prosodic features can predict up to 60% of intelligibility ratings (Kang *et al.*, 2018, 2020) as well as comprehensibility and accentedness (Kang, 2010). However, when it comes to pragma-prosodic appropriateness, there may be other features that can better predict listeners' perceptions of this construct. Some of these may include other prosodic features, such as pause length and

distribution, speech rate, pitch range. Previous research has shown that these features are important predictors of comprehensibility, accentedness, and intelligibility judgements (e.g., Kang, 2010). Studies have also investigated pragmatic appropriateness ratings from the lexico-grammatical perspective (e.g., Sydorenko *et al.*, 2014; Taguchi, 2006). However, aside from this study, there has not been an investigation into the suprasegmental features of speech and how they may affect appropriateness perceptions.

In addition, features beyond suprasegmentals may affect raters' appropriateness perceptions including learners' use of lexico-grammatical modifications or the length of a speech act. At the end of the survey, one of the raters noted that they "*gave lower ratings to speech acts just because the speaker was talking too much. They were given too many explanations, and it was starting to get inappropriate*". This remark suggests that there may be factors beyond prominence, tone choices, and main pragmatic strategies that could influence pragma-prosodic appropriateness perceptions.

Some of the other likely reasons for the apparent lack of change in appropriateness perceptions may be attributed to the design of this study. The study used different communicative situations in pre- and post-tests, although they were tightly controlled for several pragmatic factors. For example, in the pre-test, learners were asked to request a letter of recommendation and an unscheduled day off while in the post-test, they asked for a change in class registration and for an extension. It is thus possible that the communicative situations of requests affected raters' perceptions. Another potential explanation has to do with the raters themselves and their individual characteristics. Tone choices and prominence are not necessarily the most salient prosodic features of speech and may not have been easily accessible to raters solely from listening to requests without PRAAT images. In addition, it is possible that raters'

different degree of familiarity with Chinese English accent interfered with their ratings. Finally, although inter-rater agreement was acceptable based on calculated Fleiss Kappa, it could have been improved by additional rater training as well as scaffolding of the rating procedures (e.g., demonstration of pitch contours for each request) to increase rater agreement. However, because this study aimed to collect appropriateness ratings that resemble real-life perceptions as much as possible, these additional scaffolding was not included.

The lack of training effect on appropriateness perceptions was further explained by the lack of an observed relationships between learners' use of tone choices and prominence and their perceived appropriateness. Additional multiple regression analyses conducted to see if learners' use of tone choices and prominence predicted appropriateness perceptions showed that neither of the two variables was able to explain the variance in ratings substantially. This finding reinforces the idea that prosodic features other than the ones chosen in the study as well as additional linguistic features (and likely a combination of such features) may better explain why listeners may assign a certain appropriateness rating.

In discussing the apparent lack of role of prosodic features of speech in perceived appropriateness, one also needs to consider the scales used to measure the construct of appropriateness itself. While the 5-point scale used in this study was theoretically and empirically motivated by previous research, it is possible that the scale was not fine-grained enough and did not create enough variance to reflect the differences in appropriateness perceptions before and after the training. Using a 100-point scale might have captured more subtle intrinsic differences in raters' perceptions of appropriateness.

Furthermore, findings showed that raters' appropriateness ratings were significantly associated with their perceptions of learners' accentedness and comprehensibility. This finding is

perhaps not surprising. Previous research has consistently demonstrated that listeners judge accented speech to be less acceptable in certain academic and professional contexts (Kang *et al.*, 2023) and less credible than L1-accented speech (Lev-Ari & Keysar, 2010). Speech comprehensibility has also been associated with listeners' perceptions of speaker fluency (Suzuki & Kormos, 2020). Based on the results of this study, it appears that perceived pragma-prosodic appropriateness of speakers' production is also related to accentedness and comprehensibility perceptions. This novel finding contributes to the growing body of speech perception research bringing to light yet another aspect of interaction, namely, its pragma-prosodic aspects.

Taken together, findings in RQ1 indicate that while from the acoustic perspective, learners did improve their prosodic patterns significantly, these changes did not explain listeners' perceptions of appropriateness of learners' request production. More research is needed to tease out the construct of pragma-prosodic appropriateness and what it entails. Furthermore, significant relationships between accentedness, comprehensibility, and appropriateness ratings imply that listeners' perceptions of how pragmatically appropriate a speaker is may be affected by their impressions of the speaker's accent strength and difficulty in understanding their speech. The findings in this research questions present key takeaways for the field of L2 pronunciation and L2 pragmatic training and offer insights into the conceptualization of pragma-prosodic constructs important for learners' interactional success.

Research Question 2: Learners' Cognitive Individual Differences and Use of Prosody

The second research question examined the role of learners' aptitude for acquiring pragma-prosodic features in their use of prosody before and after the training. Similar to RQ1, this question also has two parts. In the first part, the predictive ability of learners' aptitude was investigated in relation to their pragma-prosodic acquisition before the training. In the second

part, the extent to which learners' aptitude predicted their changes in prosody after the training was examined. The research question was mainly motivated by a desire to understand the roots of pragma-prosodic acquisition specifically in L2 learners. Intonation acquisition has been widely investigated in psycholinguistic research showing that there are indeed certain cognitive IDs that can predict the acquisition of this feature in various L1s (e.g., Jacquemot & Scott, 2006; Snow & Balog, 2002). There have also been studies pertaining to L2 learning specifically that have examined the effect of cognitive IDs like working memory, auditory processing, and musical sensitivity in L2 prosodic and pragmatic acquisition (e.g., Li, 2017; Saito *et al.*, 2020 a,b; Schwab & Dellwo, 2019; Taguchi, 2008). However, the role of these IDs in L2 pragma-prosodic development has been inconclusive. Another motivation for this question stemmed from the practical, teaching-oriented perspective. Besides checking the general effectiveness of the developed web-based instruction, it was important to define whom precisely the instruction was likely to benefit (Kissling, 2014). Thus, accounting for variation in learners' aptitude was critical for discussing the effectiveness of the developed instruction.

The three IDs included in the study (phonological working memory [PWM], auditory processing [AP], and musical sensitivity [MS]) were measured via a series of tasks offered to the experimental group at the end of the training, during the post-test. Learners' scores extracted from these measures were included in four multiple regressions as predictors. The response variable in each regression was learners' prominence or tone choice production before or after the instruction. In addition, to control for the potential effect of learners' proficiency, learners' EIT scores were entered into the models. While proficiency was not one of the IDs considered in the study, it was important to account for it as a potential predictor of learners' pragma-prosodic development. Although Research Question 2 consists of two parts, it is best to discuss the results

of these parts together since the main interest in measuring the relationship between cognitive IDs before and after the instruction was to detect differences in predictive power of IDs with regards to learners prosodic production.

Thus, the first part of RQ 2 examined the role of learners' cognitive IDs in their pragma-prosodic acquisition before the instruction. The second part measured the same relationship but after the instruction took place. It was hypothesized that this combination of IDs would predict some variance in learners' pragma-prosodic development. Measuring this relationship provided insights into how L2 pragma-prosodic acquisition usually happens for learners with varied aptitude, and whether targeted instruction may have a varied effect based on learners' cognitive IDs.

In answering the first sub-question, two multiple regressions were run with learner data. One had tone choice production and the other included prominence as the response variable. The predictors in the model were learners' PWM, AP, and MS, as well as their proficiency. The findings of the regression analysis did not indicate any significant effect of learners' IDs in their tone choice and prominence acquisition before the instruction. The models were only able to explain 2 – 4% of the variance in tone choice production as indicated by adjusted R^2 values.

The analyses addressing the second sub-question, however, revealed some significant relationships. Specifically, multiple regression analyses with the same independent variables showed that working memory significantly predicted learners' prominence (26% of variance explained) and musical ability was a significant predictor of learners' tone choice production after the training (31% of variance explained). This finding suggested that in instructional contexts, pragma-prosodic training may be more effective for learners that possess higher PWM capacity and higher musical ability.

These findings are somewhat consistent with other empirical research, especially those that investigated the relationship from the instructional perspective. The findings support other studies that showed significant associations between PWM capacity and L2 pragmatic and phonological learning (e.g., Ahmadian, 2020; Darcy *et al.*, 2015; Rota & Reiterer, 2009). While there is also empirical evidence showing that PWM might not be a significant predictor of phonological learning, this does not seem to be the case in this study. For example, although Hu *et al.* (2013) did not find that phonological PWM can predict L2 pronunciation aptitude, this lack of effect was observed in advanced learners. In the current study, learners were believed to be of low-intermediate / intermediate proficiency as evidenced by their class levels as well as their EIT scores. Therefore, it can be deduced that learners' L2 development was at a stage when PWM still played a significant role in the acquisition of prosodic features anymore. Interestingly, PWM only predicted learners' prominence after the training, and not tone choices. This finding is supported by research in L2 prosodic development in children that detected a meaningful relationship between working memory and prominence (Kushch *et al.*, 2018). It is perhaps not surprising that PWM capacity predicted post-training prominence production in particular. During the training, learners had to complete perception and production-based exercises where they had to listen to relatively long sentences and identify/reproduce prominence patterns in them. To complete such exercises, learners likely had to use their working memory to store a sentence with its prominence pattern in their cognition and then successfully replicate it.

In the same vein, the study supported some of the research that has investigated learners' musical sensitivity and L2 phonological acquisition. In several previous studies (e.g., Pei *et al.*, 2016; Slevc & Miyake, 2006), learners' musical ability accounted for a significant portion of their suprasegmental production. At the same time, however, in Schwab and Dellwo (2019), the

relationship between learners' musical ability and L2 prosodic learning were inconclusive and depended on learners' L1 background. In the present study, participants were L1 Mandarin Chinese speakers. Research has showed that speakers of Mandarin Chinese, which is a tonal language where pitch contributors to word meaning, are better able to imitate and perceptually discriminate musical pitch (Pfordresher & Brown, 2009). It is therefore possible that Chinese learners of English with superior musical sensitivity are able to apply this aptitude feature in prosodic learning.

There was no significant effect of auditory processing on learners' acquisition of prosodic features. This finding was not in line with previous research that revealed a link between AP and L2 suprasegmental acquisition (Wong & Perrachione, 2007) as well as higher prosodic accuracy (Saito *et al.*, 2020b). Connections have also been made between auditory processing and learners' prosodic perception after immersion (Kachlicka *et al.*, 2019; Sun *et al.*, 2021). The observed lack of effect of AP on pragma-prosodic acquisition in this study can possibly be explained by the nature of the tasks used to measure this cognitive ID. Previously, AP has been measured by a range of perception- and production-based tasks (e.g., Saito *et al.*, 2021 a, b; Sun *et al.*, 2021; Zheng *et al.*, 2022). In this study, a musical reproduction task was used as an auditory processing measure. Although this particular measure has been associated with L2 suprasegmental gains (Sun *et al.*, 2021), there have also been more recent reports showing it is only marginally related to L2 pronunciation proficiency. Thus, Zheng *et al.* (2022) argued that production measures like the one in this study may tap into a range of abilities, including short-term memory, and as such, are not a very "pure" measure of auditory processing (p. 492) (see also Trofimovich *et al.*, 2015). The findings in this study offer support to Zhang *et al.*'s claims that the underlying mechanisms of L2 prosodic learning may be perceptual rather than

productive in nature. Similar conclusions have been made by Flege and Bohn (2021) in their revised Speech Learning Model (SLM-r).

While learners' L2 English proficiency variable was not of interest in the current study, it was important to measure this variable in order to control for its possible effect in the instructional effectiveness. Learners' proficiency was not a significant predictor of learners' changes in tone choice and prominence in this study. This finding indicates that for the range of proficiency levels in the study, the training was effective in changing their tone choice and prominence use patterns, at least from the acoustic analysis perspective.

In considering these findings, one has to account for the online mode of data collection. The platform used to collect these cognitive ID data, Gorilla, has been shown to record participants' responses accurately and demonstrated minimized timing issues with presenting tasks (Anwyl-Irvine *et al.*, 2020). However, there are potential limitations to this data collection method related to internet connection issues, differences in devices on which tasks are accessed, as well as the lack of observation opportunities while participants complete the tasks.

Overall, regarding Research Question 2, the study found that learners' phonological working memory and musical ability were indeed significant predictors of their tone choice and prominence production after the training. This finding is of importance for language instructors and curriculum designers as it calls for attention to these cognitive IDs in the development of teaching materials. Ideally, teachers need to account for this variation in learners' working memory and musical ability as they create activities to teach prosodic features of speech. For example, visualization of prosodic contours and segmentation of instruction into smaller chunks may assist those learners with lower PWM and MS capacities. Additionally, PWM and MS measures can be implemented at the beginning of a computer-assisted training to determine what

the rest of the training looks like (e.g., providing additional activities or items on an activity for those with lower PWM and MS). In contrast, auditory processing measured by a melody reproduction task was not a significant predictor of learners' phonological learning. There was also no effect of learners' proficiency on learners' pragma-prosodic learning. Apart from the contributions to our understanding of L2 acquisition of pragma-prosodic features, these results inform instructional design and implementation practices as similar web-based training platforms are developed in the future.

Research Question 3: Learners' Ability to Perceive Prosodic Cues and Prosodic Use

The third and final research question in this study examined the relationship between learners' ability to perceive intonation cues in the perception tasks offered to the learners during the training and the effectiveness of the training itself in changing learners' tone choice and prominence patterns. Primarily, this question had a mostly practical motivation behind it; investigating this question would be informative for the training design as it would show whether including perceptual exercises would lead to learners' production improvements. Yet, further clarifying the perception – production link that has been proposed by existing intonation learning models (e.g., Mennen, 2015) was an additional motivation that guided the research question. While there has been consistent evidence that learners can improve their prosodic perception skills as a result of perceptual training (Hori *et al.*, 2021; Ngo & Setter, 2011), much is still unknown about the effectiveness of perceptual practice on production of pragma-prosodic features.

In tackling this question, two multiple regression analyses were performed. In each of the models, four perception task types from the training were entered as predictors: tone choice identification, prominence identification, tone choice discrimination, and prominence

discrimination. Each of these tasks in the training included ten to twenty items in the training. The response variables in the models were learners' tone choice and prominence use. The results of the first model showed that learners' performance on tone choice discrimination tasks (listening to two requests and selecting one with more appropriate prominence) was a significant predictor of their changes in tone choice production, explaining 20% of the observed variance. The second model indicated that learners' performance on prominence discrimination tasks were significant for changes in prominence production.

The findings in this question are important from several viewpoints. First, the findings imply that to some extent, learners' ability to perceive intonation cues, specifically in discrimination exercises, is related to the changes they demonstrate in post-instructional production. In other words, Chinese learners of English who are more successful in choosing a more pragma-prosodically appropriate request from the first attempt also demonstrate larger gains in their ability to produce tone choices and production more appropriately. Moreover, the findings indicate a match between the perception and production of specific prosodic features. Thus, tone choice discrimination ability seems to predict tone choice use, and prominence discrimination is linked with learners' improvements in prominence use. These results offer support to the perception-production link proposed by speech acquisition models (Mennen, 2015) and extend evidence for this phenomenon in the domain of pragma-prosodic learning. Whereas perceptual abilities have recently been linked with suprasegmental learning (Saito *et al.*, 2020a, b; Zhang *et al.*, 2022), at the time of writing this, research has not offered evidence that the same connection can be made with pragmatically contextualized learning of prosody. The present study provides such evidence and implies that discrimination-based perception exercises can be an effective part of pragma-prosodic instruction. In doing so, the study echoes the

arguments made for the inclusion of perception tasks in speech and pragmatics instruction (Celce-Murcia *et al.*, 2010; McGregor & Reed, 2018).

In addition to the general evidence that prosodic perception is related to prosodic production, the results allow us to tease out the specific perception skill that may be at play in this relationship. As mentioned above, a significant relationship between perception and production was found only in discrimination tasks. Previous research has not necessarily compared the effectiveness of these two tasks in measuring prosodic perception ability. However, some studies have implicitly commented on the potential differences in task difficulty. McAndrews (2019), for example, mentioned that discrimination tasks may be more salient for learners from the beginning while categorization tasks may require some training before they can demonstrate learners' prosodic perception ability. Indeed, the current findings demonstrate that the categorization task was more difficult for learners than the discrimination one. It is therefore possible that learners' categorization ability was not strong enough at the time of the training to reveal any strong connections with their improvements in appropriate prosody use. Moreover, in parallel with the current investigation, other studies in suprasegmental perception have also showed that discrimination tasks are related to learners' use of suprasegmentals (e.g., Hallé *et al.*, 2004).

The key takeaway from the findings in Research Question 3 is that perception-based exercises can indeed help Chinese learners of English change their use of prosodic features. Those learners with superior perceptual ability, especially in discriminating between more or less prosodically appropriate requests, seem to benefit more from perceptual exercises. The findings also shed light on the process of L2 speech acquisition, and pragma-prosodic development in particular. Previous research has employed discrimination tasks for training segmental and

suprasegmental features to improve learners' intelligibility and comprehensibility (Nagle & Bases-Berk, 2021; Peppé & McCann, 2003). The present study goes one step further contextualizing such training in pragmatic scenarios and focusing on specific discourse functions of tone choices and prominence. In the study, learners were exposed not just to prosodic features but also to their meaning in requestive speech acts providing a deeper understanding of the functionality of these features. Finally, the study results offer evidence to the long-standing theory that speech perception practice may lead to successful production of target pronunciation features, specifically in the domain of prosody.

CHAPTER 6: IMPLICATIONS AND DIRECTIONS FOR FUTURE RESEARCH

This closing chapter discusses the main takeaways from this dissertation study for several domains of applied linguistics. First, it highlights the implications of the study followed by its limitations and related opportunities for future research. The chapter ends with some general concluding remarks.

Implications

There are a number of implications of this dissertation study that advance several domains of applied linguistics. First and foremost, the study informs L2 instructional practices, specifically with regards to learners' interactional competence. Moreover, it informs the broad field of second language acquisition as well as research in its specific domains of L2 pronunciation and pragmatics. The subsections below summarize these implications in three groups: (1) implications for L2 pedagogy and material development, (2) implications for L2 pronunciation and pragmatics research and methodology, (3) implications for SLA.

Implications for Second Language Pedagogy and Material Development

The primary motivation for this study was to provide evidence that learners' pragma-prosodic skills can indeed be developed as a result of an instructional training and ultimately lead to more successful participation in global interactions. Therefore, it appears only logical to start the discussion of the study implications with its importance for the pedagogical aspects of L2 learning as well as for future development of instructional materials. Although research has repeatedly demonstrated that English learners' use of prosodic cues, or rather its misuse, can be oftentimes interpreted by the interlocutors as impolite or not appropriate, affecting communication success (e.g., Anderson-Hsieh *et al.*, 1992; Derwing & Munro, 1997; Pickering,

2009; Staples, 2015). Despite that, there still exists an unfortunate disconnect between research findings and their actual applications in instructional practices and materials. In fact, pragmatic functions of prosody have been consistently ignored in pronunciation teaching (Levis, 1999; 2005), in addition to the general lack of attention to suprasegmental and pragmatic features overall (Tatsuki, 2019). What is more, there has been a general scarcity of exercises in textbooks that attend to discourse intonation for specific speech acts (Nikolić, 2018; Trouvain *et al.*, 2016). The instructional training designed in this study meets this pedagogical need in a few meaningful ways.

First, the study adds to a very limited inventory of resources available to teach pragmatic-prosodic aspects of interaction. To some extent, recent studies have addressed this shortage by developing and assessing instruction that centers around either L2 pronunciation or pragmatics as separate aspects of communication (e.g., Bardovi-Harlig & Vellenga, 2012; Derwing & Rossiter, 2003; Lee *et al.*, 2020). While effective, these types of instruction are somewhat decontextualized in that they do not present learners with real-life communicative situations and thus may be difficult to transfer to actual everyday communication. In contrast, unique to this study was contextualization of prosodic features in pragmatic situations that L2 English learners, specifically those studying at an English-speaking university, encounter daily. This approach contributed to the ecological validity of the training in its application in the real world. Besides, the training also incorporated the link that exists between prosody and pragmatics providing learners with insights into how this link can be integrated in communicating in English.

Out of the range of prosodic features of speech, only two were chosen for the study: tone choices and prominence. These features were selected to ensure study feasibility, but more importantly, because they have been found to show the strongest links with pragmatic meaning

in discourse (Brazil, 1997; Kang *et al.*, 2021; Kostromitina & Kermad, 2023; Pickering, 2018; Staples, 2015). The selected features showed significant changes after the training suggesting that they are indeed teachable, at least via the instructional design features included in the study.

Second, the study is novel in its web-based hybrid instructional mode, meeting the current instructional need in language education. With a recent change in classroom paradigms instigated by the global development of educational technologies and reinforced by the recent COVID-19 pandemic, when in-person instruction is not available, there is an increased need for pragmatic instruction that can be delivered to learners digitally and remotely (Taguchi, 2011b). Although distance learning has played a key role in English teaching and learning long before the pandemic, its impact is even more meaningful today, when English learners all over the world are still experiencing difficulties related to in-person instructional access. In China, for example, over 100 million students were taking online classes at the end of 2020, unable to receive in-person instruction (Wang & Zhao, 2020). Development of the digital instruction platform in this study provided inclusive education opportunities for such students. Participants in this study were L2 English learners who were physically located in China. Despite being in a different location from the researcher, they were able to complete the training without any particular difficulties – a clear indication of increased accessibility thanks to the online hybrid instructional mode.

Admittedly, online instruction is not without its drawbacks. Mainly, learners in the study had to overcome unstable internet connection as well as varying performance of the platform on different devices (e.g., laptop vs. tablet). However, those were minor challenges that were easily offset by the convenience and efficacy of the instruction overall. Another potential downside of online instruction is reduced opportunities for teacher-student contact and for peer interaction, as

well. Of course, in-person teaching is an ideal setting for such interaction. However, when such a modality is not available, online instruction presents a viable option for learners' continued growth in their target language. Moreover, as noted by participants, online exercises can find their application as additional at-home practice after in-class instruction. As a matter of fact, almost 30% of the study participants in their final questionnaire indicated that if they were to teach an L2 English Listening and Speaking course, they would include the exercises they completed on the platform as part of homework.

Besides the general usefulness of the platform and its online mode, the current study has many affordances specifically for designing web-based pragma-prosodic training. Instruction in the study included several design features that have been found effective in previous intervention research in L2 pronunciation and pragmatics. Some of these features were *explicit* instruction on pragma-prosodic functions of tone choices and prominence, perception-based exercises, visual aids and feedback in the form of generated prosodic contours. Research concentrating on either prosodic or pragmatic side of interaction separately indicated that including these features in training may lead to larger learner gains in chosen instructional targets (Hincks, 2015; Jiang & Chun, 2021; Lee *et al.*, 2015; Plonsky & Zhuang, 2019; Saito, 2012). The current study expands these findings onto the intersection of pragma-prosodic teaching. The gains in intonation that learners demonstrated in this study (at least from the acoustic perspective), can be attributed to its design features. Although no comparisons were made between the features (i.e., with or without visual aids or implicit vs. explicit instruction), this has implications for teachers and material developers. It is recommended that similarly oriented instruction also includes explicit explanations of target features and their usage, provides visual representations of pitch contours so that learners could observe tone choice and prominence use in audio models as well as their

own production, and offers opportunities to practice listening to target structures before producing them. These recommendations stem not only from the observed acoustic changes in learners' performance, but also from their feedback in the final questionnaire, where multiple participants noted the benefits of the design features above.

Implications for L2 Pronunciation and Pragmatics Research

Besides language educators, curriculum designers, and instructional developers, the study extends takeaways to researchers in the domains of L2 pronunciation and pragmatics. One of such takeaways is a re-evaluation of existing approaches to measuring prosodic appropriateness in discourse. The idea of pragma-prosodic interface that guided this dissertation study is not new and was introduced, among others, by Brazil (1997) in his Discourse Intonation framework. In recent years, a small, but growing, body of research has made use of Brazil's framework in bringing the two domains closer together, drawing connections between intonation features and their pragmatic meaning in discourse (see Pickering, 2018). Brazil's framework, however, is not devoid of some limitations, one of them being that the framework is not necessarily rooted in listeners' perceptions about intonation use. That is, while the framework provides descriptions for why certain intonation contours and prominence patterns occur in English speech, these descriptions stem mostly from subjective interpretations of prosodic features rather than evidence gathered from listeners who actually perceive these features. Being one of such studies, the development and feature selection for this study's instruction was guided by the same framework. At the same time, however, the assessment of pragma-prosodic appropriateness in the study were based not only on Brazil's interpretations, but also on perceptual judgements provided by raters, thus applying the principle of methodological triangulation.

Methodological triangulation has been gaining recognition both in the field of applied linguistics and in domains of speech research (Yaw, 2022). This approach allows a researcher to study linguistic phenomena from two or more complementary perspectives in order to gain more robust evidence for an issue of interest (Egbert & Baker, 2020). In L2 pronunciation studies, global and specific measures have been triangulated to measure learners' comprehensibility development (Darcy & Rocca, 2023), and multiple qualitative tools (e.g., interviews, focus groups, observations) have been amalgamated to profile an understanding of teachers' perspective on L2 pronunciation practices (Burri & Baker, 2021). Similarly, in speech perception, studies have combined measures of linguistic stereotyping with perceptual ratings of L2 accentedness and comprehensibility (e.g., Isaacs & Trofimovich, 2012; Kang & Rubin, 2009). Previous research in L2 pragmatics has also done this to some extent, for example, by measuring learners' pragmatic production after a task-based intervention and supplementing the evidence for instructional effectiveness with perceptual ratings (e.g., Sağdıç, 2022). With regards to prosodic appropriateness, however, this study adds to a small body of research that triangulated both acoustic and perceptual evidence in interpreting and evaluating the construct of prosodic appropriateness in pragmatic situations (e.g., Herrero & Devís, 2020). Acoustic analyses have been widely used in conjunction with Brazil's model to evaluate learners' prosodic use (e.g., Kang *et al.*, 2021; Taguchi *et al.*, 2021). On the other hand, there have been studies that applied perceptual ratings to measuring constructs like accentedness, comprehensibility, and acceptability in speech (e.g., Kang *et al.*, 2023). By triangulating the two methods, this study gained a more wholistic was able to get a complete picture of learners' gains in pragma-prosodic appropriateness after the training. Applying such methodological triangulations is essential for both L2 pronunciation and pragmatics research and especially for instructional studies at the

intersection of the two domains as it can help to clarify the relationship between various constructs of interest.

One of the constructs that was examined in this study was pragma-prosodic appropriateness, or learners' ability to apply prosodic features (i.e., tone choices and prominence) appropriately in request situations. Methodological triangulation in the study showed that while acoustically, learners in the study demonstrated changes in pragma-prosodic use that aligned more with Brazil's framework of intonation use, these changes did not significantly affect raters' perceptions of learners' appropriateness in making the requests. This finding brings to light an important implication for research in L2 pronunciation and pragmatics that has to do with a lack of a clear construct definition for appropriateness. In pragmatics, scholars have suggested their operationalizations of this construct that included factors like *politeness*, *directness*, and *formality* (e.g., Ishihara, 2013; Taguchi, 2011a), although a specific combination of these characteristics has varied from study to study and additionally included the notions of nativeness (Pupura, 2004), *sincerity* (Alemi *et al.*, 2014), and *authenticity* (Alemi & Khanlarzadeh, 2016). In prosodic research, appropriateness has mostly been described within Brazil's Discourse Intonation framework and related to indirect meaning conveyed by speakers. Some studies have also associated prosodic appropriateness with politeness and sincerity (Aijmer, 1995; Herrero & Devís, 2020; Knowles, 2016) while others took a holistic approach without identifying sub-constructs that define appropriateness (Sydorenko *et al.*, 2014).

Looking at the results of the present study, the absence of a clear construct definition of appropriateness in the field and its consequences is evident. The study operationalized pragma-prosodic appropriateness by including the sub-constructs that have been most commonly associated with this construct, namely, *politeness*, *directness*, and *formality*. While these sub-

constructs may be more readily applied on the lexico-grammatical manifestations of appropriateness (e.g., main pragmatic strategies in speech acts), a connection is not as clear between them and prosodic features of speech. This study demonstrated that tone choices and prominence, although being most frequently linked to conveying pragmatic meaning in previous research, may not be as relevant for appropriateness perceptions of requests. In fact, according to some of the comments from the raters, features like the length of a speech act may be more pertinent to this construct. This is further exemplified by the lack of a relationship between tone choices, prominence and perceived appropriateness that emerged in additional regression analyses. These findings reinforce the idea that tone choices and prominence may not be as influential in request-making as they are in other pragmatic elements of spoken discourse like irony and humor that have demonstrated clear links between prosody and listener perceptions (e.g., Shively et al., 2008). The study results give a strong signal that without a transparent definition for the construct of appropriateness, our understanding of how prosody impacts appropriateness perceptions is limited. As a field, we need appropriateness to be consistently operationalized across multiple research contexts and populations. Consequently, when such a definition exists, tools and scales for measuring appropriateness reliably can be developed and tested.

Another implication of the study, especially for the field of L2 speech perception, is the observed overlap in Accentedness, Comprehensibility, and Appropriateness perceptions. Although the association between the three constructs was rather small (though significant), it further brings to light the complexity of appropriateness as a construct and the need to tease out its components. From the viewpoint of accentedness and comprehensibility, this association offers yet another avenue for investigations into the role of these constructs in listener

perceptions besides some of the established connections with acceptability and credibility (e.g., Lev-Ari & Keysar, 2010; Kang *et al.*, 2023). For perception research, this result is another reminder that measuring potential confounding variables is crucial in order to fully understand an investigated relationship.

Some of the additional implications of the study for its research domains have to do with the platform itself. The usefulness of audio recording functionality and pitch contour displays for language instruction has already been discussed above; however, these tools can also be of help for non-instructional studies that want to examine prosodic production or the link between perception and production in various communicative contexts. Moreover, video-based multiple-turn role-plays that were used in this study to record learners' pre- and post-training production could be a viable, more authentic instrument that could help pragmatic research move beyond DCTs as the main method of data collection and provide richer data that would allow for additional analyses of pragmatic production.

Implications for Second Language Acquisition

From the point of view of second language acquisition, the study adds to our understanding of several crucial questions that exist within the field. One of these questions is the link between speech perception and production. As the debate about increased perception leading to improved production is still ongoing for various speech constructs (Bases-Berk, 2021; Lee *et al.*, 2020), this study contributes evidence to the perspective that supports the existence of such a link. Indeed, the study results indicated that learners who were able to perceive tone choices and prominence features more readily in discrimination exercises (i.e., choosing between two options) also used those features more appropriately in requests after the instruction, according to the acoustic analysis. Besides carrying theoretical implications, this finding is valuable for

instructed SLA. Future instructional studies may take advantage of the evidence presented here to develop perception-based trainings targeting the pragma-prosodic features investigated here or other features. In a Listening and Speaking course, learners' performance on perception-based activities can inform course design and guide adjustments needed to ensure that all learners benefit from instruction.

In addition to the perception-production link, the findings regarding learners' aptitude and pragma-prosodic acquisition extend implications to the SLA domain of individual differences. In the current study, learners with higher musical sensitivity were also more likely to produce requests with appropriate tone choices, and those with higher phonological working memory used prominence more appropriately after the training. In L1 acquisition research, studies on cognitive IDs have showed their clear effect on prosodic acquisition (e.g., Nathoo *et al.*, 2005). This relationship has not been fully clarified in SLA research, although there seems to be a growing consensus that aptitude to acquiring suprasegmentals does play a role in L2 learners' success with prosodic instruction (e.g., Saito *et al.*, 2020; Sun *et al.*, 2021). Thus, present study adds to this small, but growing existing evidence connecting specifically learners' phonological working memory and musical sensitivity to their pragma-prosodic gains. For instructed SLA, these findings imply that some learners with higher musical sensitivity and working memory may have an easier time acquiring prosody. If this is the case, as the study shows, suprasegmental instruction needs to provide additional practice opportunities for those whose MS and PWM are not as strong. It may even be beneficial to measure these abilities in learners (using simple tasks like the ones in this study) to identify their potential scaffolding needs. For research, the observed effect of certain cognitive IDs on learners' pragma-prosodic performance after the training underscores a point that was made in the previous section – in

order to fully measure and understand the pragma-prosodic or other SLA-related learning process, one needs to measure not only the variables of interest but also account for possible confounds that may affect the measurement and, therefore, the observed result. Of course, both for teachers and researchers, evaluating learners' cognitive IDs and their potential or observed effects on language acquisition is challenging. However, if we want L2 learners to develop their target language skills more efficiently, such measurements may need to take precedence.

Auditory processing was not one of the significant predictors of learners' pragma-prosodic acquisition in this study. As mentioned in the discussion section, AP is a relatively new construct that still requires refinement (e.g., Trofimovich *et al.*, 2015; Zheng *et al.*, 2022). Despite its lack of effect, this is not a call for excluding AP as a possible predictor of learners' suprasegmental acquisition. Instead, the findings here are an indication that the field is still trying to spell out how auditory processing is best measured and how it may be involved in L2 pronunciation development. A word needs to also be given about learners' proficiency as another non-significant predictor in this study that was measured for control purposes. Although proficiency was not an ID of interest in this dissertation, the observed lack of its effect in participants' pragma-prosodic development implies that learners may benefit from training no matter their language level. At the same time, since the learner group was relatively homogeneous and consisted mostly of intermediate-level learners, this conclusion cannot be fully drawn until a full range of proficiencies are tested and compared.

A final takeaway that deserves to be highlighted in concert with the results on cognitive IDs is the issue of measurement mode. As pointed out above, due to the pandemic, all data collection took place online. While the implications for the web-based method of instruction have already been considered, it needs to be acknowledged that measuring learners' cognitive IDs online

might have affected the results of this study. The impossibility of participant monitoring during data collection as well as potential internet connectivity issues and delays might have introduced external confounds that are typically reduced during in-person measurement. Crucially, however, agreement between these results and previous studies that took place physically do demonstrate that online measurement is a viable option for researchers who might not have access to participants in person.

Limitations and Future Research

There are a number of limitations in this dissertation study that define potential avenues for future research. Some of these limitations are related to study design, participant recruitment, and data collection, while others have to do with the results that emerged.

First, the study design and participant recruitment were somewhat limited by the pandemic and other sociopolitical factors. Originally, the study was planned to be conducted with L1 Russian learners of English. This group of learners is not as frequently investigated, so the study was intended to bring diversity to the fields of L2 pronunciation and pragmatics research. However, due to the aforementioned factors, this group of learners was not accessible and thus replaced by L1 Mandarin Chinese learners of English who were undergraduate students. It would be beneficial if future research explored pragma-prosodic learning in populations other than the one in this study, not just from the point of view of participants' L1, but also their age and educational background. Another pandemic-related limitation was that all data collection had to take place online. Certainly, since web-based instruction is one of the foci in the study, it is self-explanatory that some of the data collection had to happen online. However, in an ideal setting, individual difference measures would have taken place in physical settings with the researcher present. Recently developed online experimental software like *Gorilla* (Anwyl-Irvine

et al., 2020) offer a way out for cognitive measures when in-person access to participants is not possible; however, future research into learners' aptitude might look into comparing the quality of data collected with such platforms vs. in-person data.

Situated in the realm of instructional research, this study has some limitations related to its design. Like with any instruction, the study included certain target features that were taught to the learners – tone choices and prominence in requests. Of course, this approach is necessary to ensure study feasibility both for the researcher and for the participants. Nonetheless, choosing specific features to investigate automatically limits the scope of a research project. In this case, although tone choice and prominence patterns showed changes after the developed training, we do not know whether the same training design would work for other prosodic aspects of pragmatic production. In order to gain a full understanding of the effectiveness of web-based instruction like the one in this study, other prosodic features (e.g., speech rate, pause length and distribution, rhythm) need to be investigated in future studies. Moreover, tone choices in the study were operationalized in a way that was somewhat limited only to the head act strategy of requests rather than their full composition. The findings showed that there may be far more to pragma-prosodic appropriateness than just main strategy tone choice production. Future research could expand the present investigation and examine tone choice and prominence production (as well as production of other prosodic features) in components of requests other than head acts. In addition, it would be beneficial to center instruction around speech acts other than requests. Future research could, for example, examine refusals, apologies, invitations, opinions, or other speech acts that are frequently encountered by L2 learners in various contexts. Finally, studies could compare the effectiveness of different methods used here (e.g., perception vs. production-based exercises, explicit vs. implicit instruction, visualization of prosodic contours vs. audio

only, hybrid vs. in person vs. fully asynchronous) to find those of greatest benefit for the learners. In general, by expanding the number of pragma-prosodic instructional studies, applied linguists will create a robust evidence base for the effectiveness of web-based pragma-prosodic trainings across contexts. Instructional intervention studies may be one of the most challenging research designs in applied linguistics; however, this should not discourage scholars from conducting such studies as they bridge research and actual teaching. Both the fields of L2 pronunciation and pragmatics, and especially the emerging intersection domain, would benefit from more instructional research that can not only inform teaching practices but also increase our understanding of language learning processes.

Conceptually, one of the limitations of this study that transpired during data analysis was the vague nature of perceived appropriateness and a related lack of validated instruments for its measurement. There is certainly a need for further research in refining what appropriateness means for listeners not only in terms of prosodic features, but also regarding other aspects of discourse that were not covered by the current study. In order to identify these aspects, researchers could engage in bottom-up explorations of appropriateness perceptions complemented by interviews with listener groups. When the construct is sufficiently defined and the role of prosodic features in it is transparent, the field can engage in development and validation of appropriateness measures (i.e., scales, rubrics). Regarding scales, it would be useful if future research made comparisons between acoustic features and appropriateness using more fine-grained scales to be able to detect the subtleties of this construct and its components. Another listener perception-related opportunity for future research that was not covered by the current study is the inclusion of Global Englishes both in the learner and the rater samples. In this study, all speakers were Chinese speakers of English. The raters were L1 English and highly

proficient L2 English speakers with linguistic background. In order to fully understand the role of prosody in pragmatic meaning-making and appropriateness in global contexts, future research could investigate the extent to which speakers of Global Englishes (e.g., Spanish English, Korean English, Singaporean English, etc.) differ in their interpretations of appropriateness of speech acts produced by speakers from shared and not shared L1 backgrounds.

With regards to learners' individual differences, shortcomings of this study outline several future research directions. First, the study focused specifically on cognitive IDs that constitute learners' aptitude for pragma-prosodic learning. While the results of the study are informative as for the effect of those IDs on learners' use of prosodic features, future studies could expand the investigation onto other IDs not directly related to aptitude. Especially meaningful for instructional research would be individual differences like motivation, learning style, willingness to communicate, affective factors, etc. Examining the effects of these individual differences would have direct applications for instructional design in L2 pragma-prosodic acquisition. Additionally, large random effects of raters in perceptual analyses indicate that there is much we do not yet know about the influence of individual differences of participants on observed results. In this study, while some of the learners IDs were measured, raters' IDs were not. In future, research needs to make a deep dive into those differences (e.g., accent familiarity, to help the field understand how they may affect provided listener perceptions and given perceptual ratings.

The second limitation has to do with measures used for the cognitive IDs in the study. Although each individual difference was measured with one task validated by previous research, there exists a multitude of measures that could have been used instead, especially given the complex nature of the examined IDs. In the time to come, the field of SLA could benefit from

studies comparing various cognitive ID measures to find those that could be used consistently across research contexts.

The results of this study are limited in their generalizability for various populations and learning contexts. As mentioned before, the learner population in the study consisted of intermediate-level L1 Chinese learners who were undergraduate university students (English majors). Besides, because learner recruitment was hindered by the pandemic, it was difficult to recruit the number of participants that was required by a-priory power analyses. The smaller than expected sample size in the study resulted in reduced statistical power for some of the analyses. In the future, generalizability could be increased if studies included larger groups of learners from various backgrounds (linguistic and educational), ages, and proficiency levels.

Aforementioned limitations in study design and generalizability offer opportunities and highlight the need for synthetic and replication research. As the existing body of research at the interface of pragmatics and prosody is rather small, replication research can help to validate the developed trainings, instruments, as well as expand the generalizability of the existing results onto other contexts and populations. Over time, once there is a sufficient number of studies investigating a certain construct within the domain, synthesists and meta-analysts can aggregate individual study findings not only for pragma-prosodic instructional effectiveness, but also for appropriateness perceptions and for cognitive ID effects.

Concluding Remarks

The overarching goal that guided this research project was to develop an innovative approach to teaching pragmatically contextualized English prosody to Chinese learners of English through a web-based digital instructional platform. To date, digital spaces have only provided instruction

on pragmatic or prosodic aspects of interaction, ignoring their interconnectedness. The study has shown that pragma-prosodic features of speech can indeed be successfully changed as a result of online training that includes explicit instruction, perception and production practice opportunities, pitch contour visualizations, paired with instructor feedback. The study also demonstrates the benefits of triangulation in assessing instructional outcomes showing that objective acoustic analyses of learner production may not reflect listener perceptions of pragma-prosodic appropriateness. Consequently, the construct of appropriateness needs to be refined and the role of prosodic features in this construct needs to be further investigated. Finally, it shows that certain individual differences in learners' cognition, like phonological working memory and musical ability, as well as their ability to perceive prosodic cues in perception exercises facilitate pragma-prosodic learning. In short, the study supplies instructors with a new, accessible way of changing learners' use of tone choices and prominence in context and opens new, important avenues for research at the intersection of pragmatics and prosody. Research such as this brings current trends in applied linguistics scholarship to bear on pedagogical practices and facilitates the development of instructional methods that meet the needs of language learners, instructors, and curriculum developers in the technologically oriented, globalized world.

REFERENCES

- Aijmer, K. (1996). *Conversational Routines in English: Convention and Creativities*. Longman.
- Ahmadian, M. J. (2012). The relationship between working memory capacity and L2 oral performance under task-based careful online planning condition. *TESOL Quarterly*, 46(1), 165–175. <https://doi.org/10.1002/tesq.8>
- Ahmadian, M. J. (2020). Explicit and implicit instruction of refusal strategies: Does working memory capacity play a role?. *Language Teaching Research*, 24(2), 163-188. <https://doi.org/10.1177/1362168818783215>
- Akatsuka, M., Hori, T., & Toyama, N. (2018). Effects of instructing English intonation pronunciation to Japanese learners. *Language Education & Technology*, 55, 151–170.
- Alemi, M. (2012). *Patterns in interlanguage pragmatic rating: Effects of rater training, intercultural Proficiency, and self-assessment*. [Unpublished doctoral dissertation]. Allameh Tabataba'i University, Tehran, Iran.
- Alemi, M., Eslami, Z. R., & Rezanejad, A. (2014). Rating EFL learners' interlanguage pragmatic competence by non-native English-speaking teachers. *Procedia-Social and Behavioral Sciences*, 98, 171–174. <https://doi.org/10.1016/j.sbspro.2014.03.403>
- Alemi, M., & Khanlarzadeh, N. (2016). Pragmatic Assessment of Request Speech Act of Iranian EFL Learners by Non-Native English-Speaking Teachers. *Iranian Journal of Language Teaching Research*, 4(2), 19–34.
- Alemi, M., & Motamedi, M. (2019). Pragmatic criteria in the holistic and analytic rating of the disagreement speech act of Iranian EFL learners by non-native English-speaking teachers. *Journal of Teaching Language Skills*, 38(1), 1–36.
- Al-Gahtani, S., & Roever, C. (2012). Role- playing L2 requests: Head acts and sequential organization. *Applied Linguistics*, 33(1), 42–65. <https://doi.org/10.1093/applin/amr031>
- Al-Gahtani, S., & Roever, C. (2018). Proficiency and preference organization in second language refusals. *Journal of Pragmatics*, 129, 140–153. <https://doi.org/10.1016/j.pragma.2018.01.014>
- Al-Issa, A. (2003). Sociocultural transfer in L2 speech behaviors: Evidence and motivating factors. *International Journal of Intercultural Relations*, 27(5), 581–601. [https://doi.org/10.1016/S0147-1767\(03\)00055-5](https://doi.org/10.1016/S0147-1767(03)00055-5)
- Al Masaeed, K., Taguchi, N., & Tamimi, M. (2020). Proficiency effects on l2 Arabic refusals. Appropriateness, linguistic strategies and multidialectal practices. *Applied Pragmatics*, 2(1), 26–53. <https://doi.org/10.1075/ap.19007.mas>

- Alptekin, C., & Erçetin, G. (2011). Effects of working memory capacity and content familiarity on literal and inferential comprehension in L2 reading. *TESOL Quarterly*, 45(2), 235–266. <https://doi.org/10.5054/tq.2011.247705>
- Anderson-Hsieh, J., Johnson, R., & Koehler, K. (1992). The relationship between native speaker judgments of nonnative pronunciation and deviance in segmentals, prosody, and syllable structure. *Language Learning*, 42(4), 529–555. <https://doi.org/10.1111/j.1467-1770.1992.tb01043.x>
- Anwyl-Irvine, A.L., Massoné J., Flitton, A., Kirkham, N.Z., Evershed, J.K. (2020). Gorilla in our midst: an online behavioural experiment builder. *Behavior Research Methods*, 52, 388–207. <https://doi.org/10.3758/s13428-019-01237-x>
- Archibald, J. (1994). A formal model of learning L2 prosodic phonology. *Second Language Research*, 10(3), 215–240. <https://doi.org/10.1177/026765839401000303>
- Atoye, R. O. (2005). Non-native perception and interpretation of English intonation. *Nordic Journal of African Studies*, 14(1), 26–42. <https://doi.org/10.53228/njas.v14i1.279>
- Austin, J. L. (1962). *How to Do Things with Words*. Oxford University Press.
- Bachman, L. F., & Palmer, A. S. (1982). The construct validation of some components of communicative proficiency. *TESOL Quarterly*, 16(4), 449–465. <https://doi.org/10.2307/3586464>
- Baddeley, A. (1996). The fractionation of working memory. *Proceedings of the National Academy of Sciences*, 93(24), 13468–13472. <https://doi.org/10.1073/pnas.93.24.13468>
- Baddeley, A. (2000). The episodic buffer: A new component of working memory?. *Trends in Cognitive Sciences*, 4(11), 417–423. [https://doi.org/10.1016/S1364-6613\(00\)01538-2](https://doi.org/10.1016/S1364-6613(00)01538-2)
- Baddeley, A. (2003). Working memory and language: An overview. *Journal of Communication Disorders*, 36(3), 189–208. [https://doi.org/10.1016/S0021-9924\(03\)00019-4](https://doi.org/10.1016/S0021-9924(03)00019-4)
- Baddeley, A., Gathercole, S., & Papagno, C. (1998). The phonological loop as a language learning device. *Psychological Review*, 105(1), 158.
- Baddeley, A. D., & Hitch, G. (1974). Working memory. In K.W. Spence & J.T. Spence (Eds.), *The psychology of learning and motivation* (pp. 47–89). Academic Press.
- Baese-Berk, M. M. (2019). Interactions between speech perception and production during learning of novel phonemic categories. *Attention, Perception, & Psychophysics*, 81, 981–1005. <https://doi.org/10.3758/s13414-019-01725-4>
- Baese-Berk, M. M., & Samuel, A. G. (2016). Listeners beware: Speech production may be bad for learning speech sounds. *Journal of Memory and Language*, 89, 23–36. <https://doi.org/10.1016/j.jml.2015.10.008>

- Baker, R.E. (2010, May 10–14). *Non-native perception and interpretation of English intonation* [Conference presentation], Speech Prosody 2010, Chicago, IL, USA. https://www.isca-speech.org/archive_v0/sp2010/papers/sp10_171.pdf
- Baker, W., Bowles, M., Foote, R., Perpiñán, S., & Bhatt, R. (2008). Social, experiential and psychological factors affecting L2 dialect acquisition. In M. Bowles, R. Foote, S. Perpiñán, & R. Bhatt (Eds.), *Selected Proceedings of the 2007 Second Language Research Forum* (pp. 187–198). Cascadilla Proceedings Project.
- Bardovi-Harlig, K. (2013). Developing L2 pragmatics. *Language Learning*, 63, 68–86. <https://doi.org/10.1111/j.1467-9922.2012.00738.x>
- Bardovi-Harlig, K., & Vellenga, H. E. (2012). The effect of instruction on conventional expressions in L2 pragmatics. *System*, 40(1), 77–89. <https://doi.org/10.1016/j.system.2012.01.004>
- Barkaoui, K. (2010). Variability in ESL essay rating processes: The role of the rating scale and rater experience. *Language Assessment Quarterly*, 7(1), 54–74. <https://doi.org/10.1080/15434300903464418>
- Barron, A. (2005). Variational pragmatics in the foreign language classroom. *System*, 33(3), 519–536. <https://doi.org/10.1016/j.system.2005.06.009>
- Bates, D., Kliegl, R., Vasishth, S., & Baayen, H. (2015). Parsimonious mixed models. *arXiv preprint arXiv:1506.04967*.
- Beebe, L. M., Takahashi, T., & Uliss-Weltz, R. (1990). Pragmatic transfer in ESL refusals. In R. Scarcella, E. S. Andersen, & S.D. Krashen (Eds.), *Developing communicative competence in a second language* (pp. 55–73). Newbury House.
- Bella, S. (2012). Pragmatic development in a foreign language: A study of Greek FL requests. *Journal of Pragmatics*, 44(13), 1917–1947. <https://doi.org/10.1016/j.pragma.2012.08.014>
- Belz, J. (2007). The role of computer mediation in the instruction and development of L2 pragmatic competence. *Annual Review of Applied Linguistics*, 27, 45–75. <https://doi.org/10.1017/S0267190508070037>
- Best, C. T. (1994). The emergence of native-language phonological influences in infants: A perceptual assimilation model. In J. C. Goodman & H. C. Nusbaum (Eds.), *The development of speech perception: The transition from speech sounds to spoken words*, (pp. 233–277). MIT Press.
- Blum-Kulka, S., House, J., & Kasper, G. (1989). Investigating cross-cultural pragmatics: An introductory overview. In S. Blum-Kulka, J. House, & G. Kasper (Eds.), *Cross-cultural pragmatics: Requests and apologies*, (pp. 1–34). Ablex Publishing Corporation.

- Blum–Kulka, S., & Olshtain, E. (1986). Too many words: Length of utterance and pragmatic failure. *Studies in Second Language Acquisition*, 8(2), 165–179. <https://doi.org/10.1017/S0272263100006069>
- Blum–Kulka, S., & Levenston, E. A. (1987). Lexical–grammatical pragmatic indicators. *Studies in Second Language Acquisition*, 9(2), 155–170. <https://doi.org/10.1017/S0272263100000450>
- Boersma, P., & Weenink, D. (2021). *PRAAT*, v. 6.0. 2.1. Retrievable online at <http://www.praat.org>.
- Bolinger, D. L. (1958). A theory of pitch accent in English. *Word*, 14(2–3), 109–149.
- Brazil, D. (1976). The teacher's use of intonation. *Educational Review*, 28(3), 180–189. <https://doi.org/10.1080/0013191760280303>
- Brazil, D. (1980). *Discourse intonation and language teaching*. Longman.
- Brazil, D. (1987). *Discussing discourse*. English Language Research.
- Brazil, D. (1997). *The communicative value of intonation in English*. Cambridge University Press.
- Brown, G., Currie, K., & Kenworthy, J. (1980). *Questions of Intonation*. Croom Helm.
- Brown, P., & Levinson, S. C. (1987). *Politeness: Some universals in language usage (Vol. 4)*. Cambridge University Press.
- Brunfaut, T., & Revesz, A. (2015). The role of task and listener characteristics in second language listening. *TESOL Quarterly*, 49(1), 141–168. <https://doi.org/10.1002/tesq.168>
- Bryant, G. A., & Fox Tree, J. E. (2005). Is there an ironic tone of voice?. *Language and Speech*, 48(3), 257–277. <https://doi.org/10.1177/00238309050480030101>
- Burri, M., & Baker, A. (2021). ‘I Feel... Slightly out of Touch’: a Longitudinal Study of Teachers Learning to Teach English Pronunciation over a Six–Year Period. *Applied linguistics*, 42(4), 791–809. <https://doi.org/10.1093/applin/amab009>
- Byon, A. S. (2004). Sociopragmatic analysis of Korean requests: Pedagogical settings. *Journal of Pragmatics*, 36(9), 1673–1704. <https://doi.org/10.1016/j.pragma.2004.05.003>
- Caclin, A., & Tillmann, B. (2018). Musical and verbal short–term memory: Insights from neurodevelopmental and neurological disorders. *Annals of the New York Academy of Sciences*, 1423(1), 155–165. doi:10.1111/nyas.13733
- Campillo, P. S., Safont–Jordà, M. P., & Codina–Espurz, V. (2009). Refusal strategies: A proposal from a sociopragmatic approach. *RAEL: Revista Electrónica de Lingüística Aplicada*, 8, 139–150.

- Caplan, D., & Waters, G. (2005). The relationship between age, processing speed, working memory capacity, and language comprehension. *Memory*, 13(3–4), 403–413. <https://doi.org/10.1080/09658210344000459>
- Casini, L., PechGeorgel, C., & Ziegler, J. C. (2018). It's about time: revisiting temporal processing deficits in dyslexia. *Developmental Science*, 21(2), e12530. <https://doi.org/10.1111/desc.12530>
- Celce–Murcia, M., Brinton, D. M., & Goodwin, J. M. (2010). *Teaching pronunciation hardback with audio CDs (2): A course book and reference guide*. Cambridge University Press.
- Cenoz, J., & Valencia, J. F. (1996). Cross–Cultural Communication and Interlanguage Pragmatics: American vs. European Requests. In L. F. Bouton (Ed.), *Pragmatics and language learning*, (pp. 47–53). University of Illinois.
- Chau, T., Huensch, A., Hoang, Y. K., & Chau, H. T. (2022). The Effects of L2 Pronunciation Instruction on EFL Learners' Intelligibility and Fluency in Spontaneous Speech. *TESL–EJ*, 25(4), n4.
- Chen, H. J. (1995, November). *Metapragmatic Judgment on Refusals: Its Reliability and Consistency* [Paper presentation]. Annual Meeting of the American Council on the Teaching of Foreign Languages, Anaheim, USA.
- Chen, H. J. (1996). *Cross–Cultural Comparison of English and Chinese Metapragmatics in Refusal*. [Unpublished doctoral dissertation]. Indiana University, IN, USA.
- Chen, Y. (2020). *The effect of explicit instruction and auditory/audio–visual training on Chinese EFL learner's perception of intonation*. [Unpublished doctoral dissertation]. Newcastle University, the UK.
- Chen, S. (2021). *Development and Validation of a Web–Based L2 Pragmatic Speaking Test in the University Context* [Unpublished doctoral dissertation]. Northern Arizona University, AZ, USA.
- Chen, S. C., & Chen, S. H. E. (2007). Effects of social status on American and Taiwanese EFL learners' production of English refusals. *Language. Text. Society*, 1(1).
- Cheng, W., Greaves, C., & Warren, M. (2008). *A corpus–driven study of discourse intonation: the Hong Kong corpus of spoken English (prosodic) (Vol. 32)*. John Benjamins Publishing.
- Cho, C. M., & Dewaele, J. M. (2021). A crosslinguistic study of the perception of emotional intonation: influence of the pitch modulations. *Studies in Second Language Acquisition*, 43(4), 870–895. <https://doi.org/10.1017/S0272263120000674>
- Christiner, M., & Reiterer, S. M. (2013). Song and speech: examining the link between singing talent and speech imitation ability. *Frontiers in Psychology*, 4. <https://doi.org/10.3389/fpsyg.2013.00874>

- Chun, D. M. (1998). Signal analysis software for teaching discourse intonation. *Language Learning & Technology*, 2(1), 74–93. <http://dx.doi.org/10.125/25033>
- Chun, D. M. (2002). *Discourse intonation in L2*. John Benjamins.
- Chun, D. M. (2019). Current and future directions in TELL. *Journal of Educational Technology & Society*, 22(2), 14–25.
- Chun, D. M., Cho, G. Y., & Kim, H. (2015). L2 proficiency and its correlates with L1 speech perception: Evidence from Korean learners of English. *Applied Psycholinguistics*, 36(2), 395–417. <https://doi.org/10.1017/S0142716413000296>
- Clark, J., & Yallop, C. (1990). *An introduction to phonetics and Phonology*. Basil Blackwell.
- Coetzee–Van Rooy, S., & Van Rooy, B. (2005). South African English: labels, comprehensibility and status. *World Englishes*, 24(1), 1–19. <http://doi.org/10.1111/j.0883-2919.2005.00384.x>
- Cohen, A. D. (2005). Strategies for learning and performing L2 speech acts. *Intercultural Pragmatics*, 2, 275–301. <https://doi.org/10.1515/iprg.2005.2.3.275>
- Cohen, A. D. & Ishihara, N. (2005). *A web-based approach to strategic learning of speech acts*. Center for Advanced Research on Language Acquisition (CARLA), University of Minnesota.
- Cohen, A. D. & Sykes, J. M. (2006, June 17). *The development and evaluation of a self-access website for learning Spanish speech acts*. [Conference presentation]. The Annual Joint AAALACLA/CAAL Conference, Montreal, Canada.
- Colom, R., Shih, P. C., Flores–Mendoza, C., & Quiroga, M. Á. (2006). The real relationship between short-term memory and working memory. *Memory*, 14(7), 804–813. <https://doi.org/10.1080/09658210600680020>
- Council of Europe. (2001). *Common European framework of reference for languages: Learning, teaching, assessment*. Cambridge University Press.
- Couper–Kuhlen, E., & Selting, M. (1996). Towards an interactional perspective on prosody and a prosodic perspective on. In M. Selting (Ed.), *Prosody in conversation* (pp. 11–56). Cambridge University Press.
- Crowther, D., Trofimovich, P., Saito, K., & Isaacs, T. (2018). Linguistic dimensions of L2 accentedness and comprehensibility vary across speaking tasks. *Studies in Second Language Acquisition*, 40(2), 443–457. <https://doi.org/10.1017/S027226311700016X>
- Crystal, D. (1985). *A Dictionary of Linguistics and Phonetics (2nd edition)*. Wiley Blackwell.

- Culpeper, J. (2005). Impoliteness and entertainment in the television quiz show: The Weakest Link. *Journal of Politeness Research. Language, Behaviour, Culture*, 1(1), 35–72. <https://doi.org/10.1515/jplr.2005.1.1.35>
- Culpeper, J. (2011). “It’s not what you said, it’s how you said it!” Prosody and impoliteness. In Linguistic Politeness Research Group (Eds.), *Discursive approaches to politeness* (pp. 57–84). De Gruyter Mouton.
- Cunningham, D. J. (2017). Second language pragmatic appropriateness in telecollaboration: The influence of discourse management and grammaticality. *System*, 64, 46–57. <https://doi.org/10.1016/j.system.2016.12.006>
- Cutler, A., Dahan, D., & van Donselaar, W. (1997). Prosody in the comprehension of spoken language: A literature review. *Language and Speech*, 40, 141–201. <https://doi.org/10.1177/002383099704000203>
- Dalman, M. (2022). *Online-based intelligibility instruction for second language (L2) learners* (Publication No. 29207245)[Doctoral dissertation, Northern Arizona University]. ProQuest Dissertations and Theses.
- Daneman, M., & Carpenter, P. A. (1980). Individual differences in working memory and reading. *Journal of Verbal Learning and Verbal Behavior*, 19(4), 450–466. [https://doi.org/10.1016/S0022-5371\(80\)90312-6](https://doi.org/10.1016/S0022-5371(80)90312-6)
- Dankovicová, J., House, J., Crooks, A., & Jones, K. (2007). The relationship between musical skills, music training, and intonation analysis skills. *Language and Speech*, 50(2), 177–225. <https://doi.org/10.1177/00238309070500020201>
- Darcy, I., Park, H., & Yang, C. L. (2015). Individual differences in L2 acquisition of English phonology: The relation between cognitive abilities and phonological processing. *Learning and Individual Differences*, 40, 63–72. <https://doi.org/10.1016/j.lindif.2015.04.005>
- Darcy, I., & Rocca, B. (2023). Comprehensibility improvements in integrated pronunciation instruction: A comparison of instructional methods and task effects. *Journal of Second Language Pronunciation* [Advanced online publication]. <https://doi.org/10.1075/jslp.21035.dar>
- DeKeyser, R. (2013). Age effects in second language learning: Stepping stones toward better understanding. *Language Learning*, 63(s1), 52–67. <https://doi.org/10.1111/j.1467-9922.2012.00737.x>
- Delogu, F., Lampis, G., & Belardinelli, M. O. (2010). From melody to lexical tone: Musical ability enhances specific aspects of foreign language perception. *European Journal of Cognitive Psychology*, 22(1), 46–61. <https://doi.org/10.1080/09541440802708136>

- Derwing, T. M., & Munro, M. J. (1997). Accent, intelligibility, and comprehensibility: Evidence from four L1s. *Studies in Second Language Acquisition*, 19(1), 1–16.
<https://doi.org/10.1017/S0272263197001010>
- Derwing, T. M., & Munro, M. J. (2009). Putting accent in its place: Rethinking obstacles to communication. *Language Teaching*, 42(4), 476–490.
<https://doi.org/10.1017/S026144480800551X>
- Derwing, T. M., & Munro, M. J. (2015). *Pronunciation fundamentals. Evidence-based perspectives for L2 teaching and research*. John Benjamins.
- Derwing, T. M., Munro, M. J., & Thomson, R. I. (2008). A longitudinal study of ESL learners' fluency and comprehensibility development. *Applied Linguistics*, 29(3), 359–380.
<https://doi.org/10.1093/applin/amm041>
- 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.
<https://doi.org/10.1111/0023-8333.00047>
- Derwing, T. M., & Rossiter, M. J. (2003). The effects of pronunciation instruction on the accuracy, fluency, and complexity of L2 accented speech. *Applied Language Learning*, 13(1), 1–17.
- Derwing, T. M., Rossiter, M. J., Munro, M. J., & Thomson, R. I. (2004). Second language fluency: Judgments on different tasks. *Language Learning*, 54(4), 655–679.
<https://doi.org/10.1111/j.1467-9922.2004.00282.x>
- Derwing, T.M., Waugh, E., Munro, M. J. (2021). Pragmatically speaking: Preparing adult ESL students for the workplace. *Applied Pragmatics*, 3(2), 105–135.
<https://doi.org/10.1075/ap.20001.der>
- Dewaele, J. M. (2008). “Appropriateness” in foreign language acquisition and use: Some theoretical, methodological and ethical considerations. *International Review of Applied Linguistics*, 46, 245–265. <https://doi.org/10.1515/IRAL.2008.011>
- Díez, F. G. (2012). Trouble Spots in the Learning of English Intonation by Spanish Speakers. Tonality and Tonicity. *Pragmatics and Prosody in English Language Teaching*, 219-230.
https://doi.org/10.1007/978-94-007-3883-6_13
- Dixon, T., Christison, M., Dixon, D. H., & Palmer, A. S. (2021). A Meta-Analysis of Hybrid Language Instruction and Call for Future Research. *The Modern Language Journal*, 105(4), 792-809. <https://doi.org/10.1111/modl.12732>
- Ducasse, A. M., & Brown, A. (2009). Assessing paired orals: Raters' orientation to interaction. *Language Testing*, 26(3), 423–443.
<https://doi.org/10.1177/0265532209104669>

- Duffin, E. (2023). *Number of international students in the U.S. 2021/22, by country of origin*. Statista. <https://www.statista.com/statistics/233880/international-students-in-the-us-by-country-of-origin/>
- Eslami-Rasekh, Z. (1993). A cross-cultural comparison of the requestive speech act realization patterns in Persian and American English. In L. Bouton & Y. Kachru (Eds.), *Pragmatics and language learning*, 4 (pp. 85–103). Division of English as an International Language.
- Faul, F., Erdfelder, E., Lang, A.-G., & Buchner, A. (2007). G*Power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behavior Research Methods*, 39(2), 175–191. <https://doi.org/10.3758/BF03193146>
- Félix-Brasdefer, J. C. (2003). *Refusals in Spanish and English: A cross-cultural study of politeness strategies among speakers of Mexican Spanish, American English, and American learners of Spanish as a foreign language*. [Unpublished doctoral dissertation]. University of Minnesota, MN, USA.
- Félix-Brasdefer, J. C. (2008). *Politeness in Mexico and the United States: A contrastive study of the realization and perception of refusals (Vol. 171)*. John Benjamins.
- Félix-Brasdefer, J. C. (2013). Refusing in L2 Spanish: The effects of the context of learning during a short-term study abroad program. In P. Salazar-Campillo & O. Martí-Arnandiz (Eds.), *Refusals in Instructional Contexts and Beyond*, (pp.147–173). Rodopi.
- Fetzer, A. (2004). *Recontextualizing context: Grammaticality meets appropriateness*. John Benjamins.
- Fleiss, J. L., Levin, B., & Paik, M. C. (1981). *Statistical Methods for Rates and Proportions* (3rd ed). John Wiley & Sons.
- Flege, J. E. (1995). Second language speech learning: Theory, findings, and problems. *Speech perception and linguistic experience: Issues in cross-language research*, 92, 233–277.
- Flege, J. E., & Bohn, O. S. (2021). The revised speech learning model (SLM-r). In R. Wayland (Ed.), *Second language speech learning: Theoretical and empirical progress*, (pp. 3–84). Cambridge University Press.
- Fox, J., & Weisberg, S. (2011). Multivariate linear models in R. *An R Companion to Applied Regression*. Thousand Oaks.
- Gabriel, C., & Kireva, E. (2014). Prosodic transfer in learner and contact varieties: Speech rhythm and intonation of Buenos Aires Spanish and L2 Castilian Spanish produced by Italian native speakers. *Studies in Second Language Acquisition*, 36(2), 257–281. <https://doi.org/10.1017/S0272263113000740>
- Geyang, Z. (2007). A pilot study on refusal to suggestions in English by Japanese and Chinese EFL learners. *Bulletin of the Graduate School of Education, Hiroshima University Part 2 Arts and Science Education*, 56, 155–163.

- Gibson, L. Y., Hogben, J. H., & Fletcher, J. (2006). Visual and auditory processing and component reading skills in developmental dyslexia. *Cognitive Neuropsychology*, 23(4), 621–642. <https://doi.org/10.1080/02643290500412545>
- Godoy, J.L., Vega, J.R., & Marchetti, J. (2014). Relationships between PCA and PLS–regression. *Chemometrics and Intelligent Laboratory Systems*, 130, 182–191. <https://doi.org/10.1016/j.chemolab.2013.11.008>
- Gordon, E. (1995). Manual: Musical aptitude profile. *Creativity. American Psychologist*, 5(9), 444–454.
- Gordon, E. (2001). *A three-year study of the Musical Aptitude Profile*. GIA Publications.
- Gordon, J., & Darcy, I. (2016). The development of comprehensible speech in L2 learners: A classroom study on the effects of short-term pronunciation instruction. *Journal of Second Language Pronunciation*, 2(1), 56–92. <https://doi.org/10.1075/jslp.2.1.03gor>
- Gordon, J., Darcy, I., & Ewert, D. (2013). Pronunciation teaching and learning: Effects of explicit phonetic instruction in the L2 classroom. In J. Levis & K. LeVelle (Eds.). *Proceedings of the 4th Pronunciation in Second Language Learning and Teaching Conference*. Aug. 2012. (pp. 194–206). Iowa State University.
- Grabowski, K. (2013). Investigating the construct validity of a role-play test designed to measure grammatical and pragmatic knowledge at multiple proficiency levels. In S. Ross & G. Kasper, (Eds.), *Assessing second language pragmatics* (pp. 149–171). Palgrave Macmillan.
- Graham, C., & Post, B. (2018). Second language acquisition of intonation: Peak alignment in American English. *Journal of Phonetics*, 66, 1–14. <https://doi.org/10.1016/j.wocn.2017.08.002>
- Grice, H. P. (1975). Logic and conversation. In *Speech acts* (pp. 41–58). Brill.
- Grosser, W., James, A., & Leather, J. (1997). On the acquisition of tonal and accentual features of English by Austrian learners. In J. Leather & A.R. James (Eds.), *Second language speech: Structure and process*, (pp. 211–228). Mouton De Gruyter.
- Guara–Tavares, M. D. G. (2008). *Pre-task planning, working memory capacity, and L2 speech performance*. [Unpublished doctoral dissertation]. Federal University of Santa Catarina, Florianopolis, Brazil.
- Gumperz, J. J. (1982). *Discourse strategies* (No. 1). Cambridge University Press.
- Gutierrez Díez, F. (2012). Trouble Spots in the Learning of English Intonation by Spanish Speakers. Tonality and Tonicity. In *Pragmatics and prosody in English language teaching* (pp. 219–230). Springer, Dordrecht.

- Halliday, M.A.K. (1967). Notes on transitivity and theme in English, Part 2. *Journal of Linguistics*, 3, 199–244. <https://doi.org/10.1017/S0022226700016613>
- Hahn, L. D. (2004). Primary stress and intelligibility: Research to motivate the teaching of suprasegmentals. *TESOL Quarterly*, 38(2), 201–223. <https://doi.org/10.2307/3588378>
- Hallé, P. A., Chang, Y. C., & Best, C. T. (2004). Identification and discrimination of Mandarin Chinese tones by Mandarin Chinese vs. French listeners. *Journal of Phonetics*, 32(3), 395–421. [https://doi.org/10.1016/S0095-4470\(03\)00016-0](https://doi.org/10.1016/S0095-4470(03)00016-0)
- Harding, L. (2012). Accent, listening assessment and the potential for a shared-L1 advantage: A DIF perspective. *Language Testing*, 29(2), 163–180. <https://doi.org/10.1177/0265532211421161>
- Hardison, D. M. (2004). Generalization of computer assisted prosody training: Quantitative and qualitative findings. *Language Learning & Technology*, 8(1), 34–52.
- Hardison, D. M. (2010). Visual and auditory input in second-language speech processing. *Language Teaching*, 43(1), 84–95. <https://doi.org/10.1017/S0261444809990176>
- He, X., van Heuven, V. J., & Gussenhoven, C. (2012). The selection of intonation contours by Chinese L2 speakers of Dutch: Orthographic closure vs. prosodic knowledge. *Second Language Research*, 28(3), 283–318. <https://doi.org/10.1177/0267658312439668>
- Heroku (2021). *Heroku: Cloud Application Platform*. Retrievable online from <https://www.heroku.com/>.
- Herrero, C., & Devís, E. (2020). Unintentional impolite intonation in L2 Spanish requests produced by Chinese workers living in Madrid. *Proceedings of the 10th International Conference on Speech Prosody 2020*, Tokyo, Japan, 848–852.
- Hewings, M. (1995). Tone choice in the English intonation of non-native speakers. *IRAL: International Review of Applied Linguistics in Language Teaching*, 33(3), 251.
- Hincks, R. (2015). Technology and learning pronunciation. In M. Reed & J.M. Levis (Eds.), *The handbook of English pronunciation*, (pp. 505–519). Wiley.
- Hirschberg, J. (2004). Pragmatics and intonation. In L. Horn & G. Ward (Eds.), *The handbook of pragmatics*, (pp. 515–537). Wiley.
- Hirst, D., & Di Cristo, A. (1998). A survey of intonation systems. In D. Hirst & A. Di Cristo (Eds.), *Intonation systems: A survey of twenty languages* (pp. 1–45). Cambridge University Press.
- Hori, T., Toyama, M., & Akatsuka, M. (2021). Perception of English Intonation Patterns by Japanese Learners of English. *Language Education & Technology*, 58, 1–16. https://doi.org/10.24539/let.58.0_1

- House, J., & Kasper, G. (1987). Interlanguage pragmatics: requesting in a foreign language. In W. Lörchner & R. Schulze (Eds.), *Perspectives on language in performance* (pp. 1250–1288). Gunter Narr.
- Hu, X., Ackermann, H., Martin, J. A., Erb, M., Winkler, S., & Reiterer, S. M. (2013). Language aptitude for pronunciation in advanced second language (L2) learners: Behavioural predictors and neural substrates. *Brain and Language*, 127(3), 366–376. <https://doi.org/10.1016/j.bandl.2012.11.006>
- Hu, X., Kuo, L. J., & Dixon, L. Q. (2022). Effectiveness of interventions on improving English language learners' intelligibility and comprehensibility: A meta-analytic review. *Language Teaching Research* (advanced online publication). <https://doi.org/10.1177/1362168822110683>
- Hudson, T., Detmer, E., & Brown, J. D. (1992). *A framework for testing cross-cultural pragmatics* (Vol. 2). National Foreign Language Resource Center.
- Hudson, T., Brown, J. D., & Detmer, E. (1995). *Developing prototypic measures of cross-cultural pragmatics* (Vol. 7). National Foreign Language Resource Center.
- Iizuka, T., Nakatsukasa, K., & Braver, A. (2020). The efficacy of gesture on second language pronunciation: An exploratory study of handclapping as a classroom instructional tool. *Language Learning*, 70(4), 1054–1090. <https://doi.org/10.1111/lang.12415>
- Ikeda, N. (2017). *Measuring L2 oral pragmatic abilities for use in social contexts: Development and validation of an assessment instrument for L2 pragmatics performance in university settings* [Unpublished doctoral dissertation]. University of Melbourne, Australia.
- Isaacs, T., & Thomson, R. I. (2013). Rater experience, rating scale length, and judgments of L2 pronunciation: Revisiting research conventions. *Language Assessment Quarterly*, 10(2), 135–159. <https://doi.org/10.1080/15434303.2013.769545>
- Isaacs, T., & Trofimovich, P. (2010). Falling on sensitive ears? The influence of musical ability on extreme raters' judgments of L2 pronunciation. *TESOL Quarterly*, 44(2), 375–386. <https://doi.org/10.5054/tq.2010.222214>
- Isaacs, T., & Trofimovich, P. (2012). Deconstructing comprehensibility: Identifying the linguistic influences on listeners' L2 comprehensibility ratings. *Studies in Second Language Acquisition*, 34(3), 475–505. <https://doi.org/10.1017/S0272263112000150>
- Isbell, D. (2017). Assessing L2 pronunciation for research purposes with listener-based numerical scales. In O. Kang and A. Ginther (Eds.), *Assessment in second language pronunciation* (pp. 89–111). Routledge.
- Ishihara, N. (2013). Teacher-based assessment of L2 Japanese pragmatics: Classroom applications. In *Assessing second language pragmatics* (pp. 124–148). Palgrave Macmillan.

- Iwasaki, N. (2008). L2 Japanese acquisition of the pragmatics of requests during a short-term study abroad. *Japanese Language Education in Europe*, 12, 51–58.
- Jacquemot, C., & Scott, S. K. (2006). What is the relationship between phonological short-term memory and speech processing?. *Trends in Cognitive Sciences*, 10(11), 480–486. <https://doi.org/10.1016/j.tics.2006.09.002>
- Jadoul, Y. (2020). Building bridges between software ecosystems: Parselmouth, a Python interface for PRAAT. *The Journal of the Acoustical Society of America*, 148(4), 2791–2791. <https://doi.org/10.1121/1.5147768>
- Jia, G., Strange, W., Wu, Y., Collado, J., & Guan, Q. (2006). Perception and production of English vowels by Mandarin speakers: Age-related differences vary with amount of L2 exposure. *The Journal of the Acoustical Society of America*, 119(2), 1118–1130. <https://doi.org/10.1121/1.2151806>
- Jianda, L. (2006). Assessing EFL learners' interlanguage pragmatic knowledge: Implications for testers and teachers. *Reflections on English Language Teaching*, 5(1), 1–22.
- Jiang, Y. (2017). *Development and effectiveness of online intonation training modules to improve Chinese Speakers' English Speech* (Publication No. 10637970) [Doctoral dissertation, University of California, Santa Barbara]. Santa Barbara ProQuest Dissertations Publishing.
- Jiang, Y., & Chun, D. (2021). Web-based intonation training helps improve ESL and EFL Chinese students' oral speech. *Computer Assisted Language Learning*, 1–29. <https://doi.org/10.1080/09588221.2021.1931342>
- Jimenez-Vilches, R. (2017). Analysing Four Prosodic Forms to Convey the Pragmatic Function of Dominance and Control in L2 Monologic Discourse. *EPiC Series in Language and Linguistics*, 2, 199–213. <https://doi.org/10.29007/rs8q>
- Johnson, J. S., & Lim, G. S. (2009). The influence of rater language background on writing performance assessment. *Language Testing*, 26(4), 485–505. <https://doi.org/10.1177/0265532209340186>
- Juffs, A., & Harrington, M. (2011). Aspects of working memory in L2 learning. *Language Teaching*, 44(2), 137–166. <https://doi.org/10.1017/S0261444810000509>
- Kachlicka, M., Saito, K., & Tierney, A. (2019). Successful second language learning is tied to robust domain-general auditory processing and stable neural representation of sound. *Brain and Language*, 192, 15–24. <https://doi.org/10.1016/j.bandl.2019.02.004>
- Kang, O. (2008). *Ratings of L2 oral performance in English: Relative impact of rater characteristics and acoustic measures of accentedness* [Unpublished doctoral dissertation]. University of Georgia, GA, USA.

- Kang, O. (2010). Relative salience of suprasegmental features on judgments of L2 comprehensibility and accentedness. *System*, 38(2), 301–315. <https://doi.org/10.1016/j.system.2010.01.005>
- Kang, O., & Kermad, A. (2017). *Assessment in second language pronunciation*. Routledge.
- Kang, O., & Kermad, A. (2019). Prosody in L2 Pragmatics Research. In N. Taguchi (Ed.), *The Routledge Handbook of Second Language Acquisition and Pragmatics* (pp. 78–92). Routledge.
- Kang, O., Kermad, A., & Taguchi, N. (2021). The effect of study abroad and proficiency on speech acts. *Journal of Second Language Pronunciation*, 7(3), 343–369. <https://doi.org/10.1075/jslp.20024.kan>
- Kang, O., Rubin, D., & Kermad, A. (2019). The effect of training and rater differences on oral proficiency assessment. *Language Testing*, 36(4), 481–504. <https://doi.org/10.1177/0265532219849522>
- Kang, O., Rubin, D. O. N., & Pickering, L. (2010). Suprasegmental measures of accentedness and judgments of language learner proficiency in oral English. *The Modern Language Journal*, 94(4), 554–566. <https://doi.org/10.1111/j.1540-4781.2010.01091.x>
- Kang, O., Thomson, R. I., & Moran, M. (2018). Empirical approaches to measuring the intelligibility of different varieties of English in predicting listener comprehension. *Language Learning*, 68(1), 115–146. <https://doi.org/10.1111/lang.12270>
- Kang, O., Thomson, R. I., & Moran, M. (2020). Which features of accent affect understanding? Exploring the intelligibility threshold of diverse accent varieties. *Applied Linguistics*, 41(4), 453–480. <https://doi.org/10.1093/applin/amy053>
- Kang, O., & Yaw, K. (2021). Social judgement of L2 accented speech stereotyping and its influential factors. *Journal of Multilingual and Multicultural Development*, 1–16. <https://doi.org/10.1080/01434632.2021.1931247>
- Kang, O., Yaw, K., & Kostromitina, M. (2023). The effects of situational contexts and occupational roles on listeners' judgements on accented speech. *Psychology of Language and Communication*, 27(1), 1–22. <https://doi.org/10.58734/plc-2023-0001>
- Kecskes, I. (2004). *The role of salience in processing pragmatic units*. Acta Linguistica.
- Kennedy, S., & Trofimovich, P. (2010). Language awareness and second language pronunciation: A classroom study. *Language Awareness*, 19(3), 171–185. <https://doi.org/10.1080/09658416.2010.486439>
- Kermad, A. A. D. (2018). *Speaker and listener variability in the perception of non-native speech* [Unpublished doctoral dissertation]. Northern Arizona University, AZ, USA.

- Kermad, A. (2021). " From the Sound, it Look Like He Said it from the Deep in his Heart": How do English Learners Make Judgments of Pragma–Prosodic Meaning?. *TESL–EJ*, 25(1).
- Kermad, A., & Bogorevich, V. (2022). Using statistical transformation methods to explore speech perception scale lengths. *Language Teaching Research Quarterly*.
<https://doi.org/10.32038/ltrq.2022.29.05>
- Kimura, T., Sensui, H., Takasawa, M., Toyomaru, A., & Atria, J.J. (2010, September). *A pilot study on perception of Spanish stress by Japanese learners of Spanish*. [Conference presentation]. Interspeech 2010 Satellite Workshop, Tokyo, Japan.
- Kissling, E. M. (2014). What predicts the effectiveness of foreign–language pronunciation instruction? Investigating the role of perception and other individual differences. *Canadian Modern Language Review*, 70(4), 532–558. <https://doi.org/10.3138/cmlr.2161>
- Kissling, E. M. (2018). Pronunciation instruction can improve L2 learners’ bottom-up processing for listening. *The Modern Language Journal*, 102(4), 653–675.
<https://doi.org/10.1111/modl.12512>
- Knowles, G. (2016). *Patterns of spoken English: An introduction to English phonetics*. Routledge.
- Koester, A. (1990). The intonation of agreeing and disagreeing in English and German. In M. Hewings (Ed.), *Papers in discourse intonation*, (pp. 83–101). English Language Research.
- Kohn, K. (2009). Computer assisted foreign language learning. In K. Knapp & B. Seidlhofer (Eds.), *Handbook of applied logistics (Vol. 6)*, (pp. 573–602). Mouton de Gruyter.
- Koike, D. A. (1989). Pragmatic competence and adult L2 acquisition: Speech acts in interlanguage. *The Modern Language Journal*, 73(3), 279–289.
<https://doi.org/10.2307/327002>
- Komar, S. (2017). The relationship between musical training and the acquisition of intonation in foreign learners of English Musical training and the acquisition of intonation. In A. Onysko, E. Graf, W. Delanoy, N. Dobric, & G. Sigott (Eds.), *The Polyphony of English Studies: A Festschrift for Allan James*, (pp. 26–47). Narr.
- Kormos, J., & Sáfár, A. (2008). Phonological short–term memory, working memory and foreign language performance in intensive language learning. *Bilingualism: Language and Cognition*, 11(2), 261–271. <https://doi.org/10.1017/S1366728908003416>
- Kostromitina, M. (in press). An exploratory study of intonational variation in L1 and L2 English speakers’ pragmatic production of high imposition requests and refusals. *Applied Pragmatics*.

- Kostromitina, M., & Kang, O. (2021). The Effects of ESL Immersion and Proficiency on Learners' *Pronunciation Development*. *Frontiers in Communication*, 6. <https://doi.org/10.3389/fcomm.2021.636122>
- Kostromitina, M., & Kermad, A. (in press). Prosody and pragmatics in applied linguistics. In C. Chapelle (Ed.), *Encyclopedia of Applied Linguistics*, 2nd edition. Wiley.
- Kostromitina, M., & Plonsky, L. (2022). Elicited imitation tasks as a measure of L2 proficiency: A meta-analysis. *Studies in Second Language Acquisition*, 44(3), 886–911. <https://doi.org/10.1017/S0272263121000395>
- Kreidler, C. W. (1989). *Intonation in the Pronunciation of English*. Basil Blackwell.
- Kumle, L., Vö, M.L.H. & Draschkow, D. (2021). Estimating power in (generalized) linear mixed models: An open introduction and tutorial in R. *Behavior Research Methods*, 53, 2528–2543. <https://doi.org/10.3758/s13428-021-01546-0>
- Kushch, O., Igualada, A., & Prieto, P. (2018). Prominence in speech and gesture favour second language novel word learning. *Language, Cognition and Neuroscience*, 33(8), 992–1004. <https://doi.org/10.1080/23273798.2018.1435894>
- Ladd, D. R. (2001). Intonational universals and intonational typology. In M. Haspelmath, E. König, W. Oesterreicher, & W. Raible (Eds.), *Language typology and language universals: An international handbook*, (pp. 1380-1390). De Gruyter Mouton.
- Lambert, W. E., Hodgson, R. C., Gardner, R. C., & Fillenbaum, S. (1960). Evaluational reactions to spoken languages. *The Journal of Abnormal and Social Psychology*, 60(1), 44–51. <https://doi.org/10.1037/h0044430>
- Law, L. N., & Zentner, M. (2012). Assessing musical abilities objectively: Construction and validation of the Profile of Music Perception Skills. *PloS one*, 7(12), e52508. <https://doi.org/10.1371/journal.pone.0052508>
- Lee, J., Jang, J., & Plonsky, L. (2015). The effectiveness of second language pronunciation instruction: A meta-analysis. *Applied Linguistics*, 36(3), 345–366. <https://doi.org/10.1093/applin/amu040>
- Lee, B., Plonsky, L., & Saito, K. (2020). The effects of perception–vs. production–based pronunciation instruction. *System*, 88, (advanced online publication). <https://doi.org/10.1016/j.system.2019.102185>
- Lesaux, N. K., Lipka, O., & Siegel, L. S. (2006). Investigating cognitive and linguistic abilities that influence the reading comprehension skills of children from diverse linguistic backgrounds. *Reading and Writing*, 19(1), 99–131. <https://doi.org/10.1007/s11145-005-4713-6>

- Lev-Ari, S., & Keysar, B. (2010). Why don't we believe non-native speakers? The influence of accent on credibility. *Journal of Experimental Social Psychology*, 46(6), 1093–1096. <https://doi.org/10.1016/j.jesp.2010.05.025>
- Levelt, W. J. (1992). Accessing words in speech production: Stages, processes and representations. *Cognition*, 42(1-3), 1-22. [https://doi.org/10.1016/0010-0277\(92\)90038-J](https://doi.org/10.1016/0010-0277(92)90038-J)
- Levis, J. M. (1999). Intonation in theory and practice, revisited. *TESOL Quarterly*, 33(1), 37–63. <https://doi.org/10.2307/3588190>
- Levis, J. (2007). Computer technology in teaching and researching pronunciation. *Annual Review of Applied Linguistics*, 27, 184–202. <https://doi.org/10.1017/S0267190508070098>
- Levis, J., & Pickering, L. (2004). Teaching intonation in discourse using speech visualization technology. *System*, 32(4), 505–524. <https://doi.org/10.1016/j.system.2004.09.009>
- Li, S. (2017). An exploratory study on the role of foreign language aptitudes in instructed pragmatics learning in L2 Chinese. *Chinese as a Second Language Research*, 6(1), 103–128. <https://doi.org/10.1515/caslar-2017-0005>
- Li, W., & Jiang, W. (2019). Requests made by Australian learners of Chinese as a foreign language. *Journal of Language Teaching and Research*, 10(1), 23–34. <http://dx.doi.org/10.17507/jltr.1001.03>
- Li, A., & Post, B. (2014). L2 acquisition of prosodic properties of speech rhythm: Evidence from L1 Mandarin and German learners of English. *Studies in Second Language Acquisition*, 36(2), 223–255. <https://doi.org/10.1017/S0272263113000752>
- Li, S., & Taguchi, N. (2014). The effects of practice modality on pragmatic development in L2 Chinese. *The Modern Language Journal*, 98(3), 794–812. <https://doi.org/10.1111/modl.12123>
- Li, S., Taguchi, N., & Xiao, F. (2019). Variations in rating scale functioning in assessing pragmatic performance in L2 Chinese. *Language Assessment Quarterly*, 16(3), 271–293. <https://doi.org/10.1080/15434303.2019.1648473>
- Lin, P. M. (2013). The prosody of formulaic expression in the IBM/Lancaster Spoken English Corpus. *International Journal of Corpus Linguistics*, 18(4), 561–588. <https://doi.org/10.1075/ijcl.18.4.05lin>
- Linacre, J. M. (1999). Investigating rating scale category utility. *Journal of Outcome Measurement*, 3(2), 103–122.
- Linacre, J. M. (2002). Optimizing rating scale category effectiveness. *Journal of Applied Measurement*, 3(1), 85–106.
- 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.
[https://doi.org/10.6503/THJCS.2016.46\(4\).04](https://doi.org/10.6503/THJCS.2016.46(4).04)
- Lumley, T., & McNamara, T. F. (1995). Rater characteristics and rater bias: Implications for training. *Language Testing*, 12(1), 54–71. <https://doi.org/10.1177/026553229501200104>
- Lundell, F. F., & Erman, B. (2012). High-level requests: A study of long residency L2 users of English and French and native speakers. *Journal of Pragmatics*, 44(6–7), 756–775.
doi:10.1016/j.pragma.2012.02.010
- Madella, P. (2020). *Prosodic pointing: From pragmatic awareness to pragmatic competence in Chinese hearers of L2 English* (Unpublished doctoral dissertation). University of Brighton, Brighton, UK.
- Magne, C., Jordan, D. K., & Gordon, R. L. (2016). Speech rhythm sensitivity and musical aptitude: ERPs and individual differences. *Brain and Language*, 153, 13–19.
<https://doi.org/10.1016/j.bandl.2016.01.001>
- Martin, K.I., & Ellis, N.C. (2012). The roles of phonological short-term memory and working memory in L2 grammar and vocabulary learning. *Studies in Second Language Acquisition* 34, 379–413. <https://doi.org/10.1017/S0272263112000125>
- Martínez-Flor, A. & Usó-Juan, E. (Eds.). (2010). *Speech act performance: Theoretical, empirical and methodological issues* (Vol. 26). John Benjamins.
- McAndrews, M. (2019). Short periods of instruction improve learners’ phonological categories for L2 suprasegmental features. *System*, 82, 151–160.
<https://doi.org/10.1016/j.system.2019.04.007>
- McAndrews, M. (2021). The effects of prosody instruction on listening comprehension in an EAP classroom context. *Language Teaching Research*, (advanced online publication).
<https://doi.org/10.1177/1362168821990346>
- McDonough, K., Trofimovich, P., Tekin, O., & Sato, M. (2022). Exploring linguistic stereotyping of international students at a Canadian university. *Journal of Multilingual and Multicultural Development*, 1–16. <https://doi.org/10.1080/01434632.2022.2115049>
- McGregor, A., & Reed, M. (2018). Integrating pronunciation into the English language curriculum: A framework for teachers. *CATESOL Journal*, 30(1), 69–94.
- McGregor, A., Zielinski, B., Meyers, C., & Reed, M. (2016). An exploration of teaching intonation using a TED Talk. In *Proceedings of the 7th Pronunciation in Second Language Learning and Teaching Conference* (pp. 143–49).
- McIntyre, P. N. (1993). *The importance and effectiveness of moderation training on the reliability of teacher assessments of ESL writing samples*. [Masters Research Thesis]. The University of Melbourne, Australia.

- McKinnon, S. (2017). TBLT instructional effects on tonal alignment and pitch range in L2 Spanish imperatives versus declaratives. *Studies in Second Language Acquisition*, 39(2), 287–317. <https://doi.org/10.1017/S0272263116000267>
- McQueen, J. M., & Cutler, A. (2010). Cognitive Processes in Speech Perception. In W. J. Hardcastle, J. Laver, & F. E. Gibbon (Eds.), *The Handbook of Phonetic Sciences*, (pp. 489–520). Wiley-Blackwell.
- Mennen, I. (2015). Beyond segments: Towards a L2 intonation learning theory. In E. Delais-Roussaire, M. Avanzi, & S. Herment (Eds.), *Prosody and language in contact* (pp. 171–188). Springer.
- Menjo, S. (2018). *The development of L2 prosody in Japanese speakers of English: A quasi-longitudinal study*. [Unpublished doctoral dissertation]. Texas A&M, TX, USA.
- Menzel, U. (2012). CCP: Significance tests for canonical correlation analysis (CCA). *R package version, 1(1)*. <https://CRAN.R-project.org/package=CCP>
- Miyake, A., & Friedman, N. P. (1998). Individual differences in second language proficiency: Working memory as language aptitude. In A. F. Healy & L. E. Bourne, Jr. (Eds.), *Foreign language learning: Psycholinguistic studies on training and retention*, 339–364. Routledge.
- Moore, B. C. (2012). *An introduction to the psychology of hearing*. Brill.
- Moyer, A. (1999). Ultimate attainment in L2 phonology: The critical factors of age, motivation, and instruction. *Studies in Second Language Acquisition*, 21(1), 81–108. <https://doi.org/10.1017/S0272263199001035>
- Müller, F. E. (1996). Affiliating and disaffiliating with continuers: Prosodic aspects of reciprocity. In M. Selting (Ed.), *Prosody in conversation: Interactional studies*, (pp. 131–176). Cambridge University Press.
- Munro, M. J. (2017). Dimensions of pronunciation. In O. Kang, R. I. Thomson, & J. M. Murphy (Eds.), *The Routledge handbook of contemporary English pronunciation* (pp. 413–431). Routledge.
- Munro, M. J., & Derwing, T. M. (2001). Modeling perceptions of the accentedness and comprehensibility of L2 speech the role of speaking rate. *Studies in Second Language Acquisition*, 23(4), 451–468. <https://doi.org/10.1017/S0272263101004016>
- Nagle, C. L. (2018). Examining the temporal structure of the perception–production link in second language acquisition: A longitudinal study. *Language Learning*, 68(1), 234–270. <https://doi.org/10.1111/lang.12275>
- Nagle, C. L., & Baese-Berk, M. M. (2021). Advancing the state of the art in L2 speech perception–production research: revisiting theoretical assumptions and methodological

- practices. *Studies in Second Language Acquisition*, 44(2), 580–605.
<https://doi.org/10.1017/S0272263121000371>
- Nathoo, N., Nair, D., Phillips, M., & Vogelbaum, M. A. (2005). Mapping prosody: correlation of functional magnetic resonance imaging with intraoperative electrocorticography recordings in a patient with a right-sided temporooccipital glioma: Case illustration. *Journal of Neurosurgery*, 103(5), 930–930. <https://doi.org/10.3171/jns.2005.103.5.0930>
- Ngo, A. P., & Setter, J. (2011, August 17–21). Evaluating the Intonation Perception Ability of Vietnamese Learners of English [Paper presentation]. ICPHX XVII, Hong Kong.
- Nikolić, D. (2018). Empirical analysis of intonation activities in EFL student's books. *International Journal of Applied Linguistics and English Literature*, 7(3), 181–187.
<https://doi.org/10.7575/aiac.ijalel.v.7n.3p.181>
- O'brien, C., Hartshorne, R., Beattie, J., & Jordan, L. (2011). A comparison of large lecture, fully online, and hybrid sections of introduction to special education. *Rural Special Education Quarterly*, 30(4), 19-31. <https://doi.org/10.1177/875687051103000404>
- O'Connor, J. D., & Arnold, G. F. (1973). *Intonation of colloquial English* (2nd ed.). Longman.
- Ortega-Llebaria, M., & Colantoni, L. (2014). L2 English intonation: Relations between form–meaning associations, access to meaning, and L1 transfer. *Studies in Second Language Acquisition*, 36(2), 331–353. <https://doi.org/10.1017/S0272263114000011>
- Osaka, M., Osaka, N., Kondo, H., Morishita, M., Fukuyama, H., Aso, T., & Shibasaki, H. (2003). The neural basis of individual differences in working memory capacity: an fMRI study. *NeuroImage*, 18(3), 789–797. [https://doi.org/10.1016/S1053-8119\(02\)00032-0](https://doi.org/10.1016/S1053-8119(02)00032-0)
- Papagno, C., Valentine, T., & Baddeley, A. (1991). Phonological short-term memory and foreign-language vocabulary learning. *Journal of Memory and Language*, 30(3), 331–347. [https://doi.org/10.1016/0749-596X\(91\)90040-Q](https://doi.org/10.1016/0749-596X(91)90040-Q)
- Pastuszek-Lipińska, B. (2004). An overview of a research project and preliminary results of two experiments on perception and production of foreign language sounds by musicians and nonmusicians. *TMH-QPSR*, 46(1), 61–74.
- Patel, A. D. (1998). Syntactic processing in language and music: different cognitive operations, similar neural resources?. *Music Perception*, 16(1), 27–42.
<https://doi.org/10.2307/40285775>
- Patel, A. D., Wong, M., Foxton, J., Lochy, A., & Peretz, I. (2008). Speech intonation perception deficits in musical tone deafness (congenital amusia). *Music Perception*, 25(4), 357–368.
<https://doi.org/10.1525/mp.2008.25.4.357>
- Pei, Z., & Ding, Y. (2013). An empirical study on the effects of musical aptitude on English phonological skills of Chinese university students. *Foreign Language World*, 1, 36–44.

- Pei, Z., Wu, Y., Xiang, X., & Qian, H. (2016). The Effects of Musical Aptitude and Musical Training on Phonological Production in Foreign Languages. *English Language Teaching*, 9(6), 19–29. <http://dx.doi.org/10.5539/elt.v9n6p19>
- Pell, M. D., & Skorup, V. (2008). Implicit processing of emotional prosody in a foreign versus native language. *Speech Communication*, 50(6), 519–530. <https://doi.org/10.1016/j.specom.2008.03.006>
- Pelzl, E., Lau, E. F., Guo, T., & DeKeyser, R. (2019). Advanced second language learners' perception of lexical tone contrasts. *Studies in Second Language Acquisition*, 41(1), 59–86. <https://doi.org/10.1017/S0272263117000444>
- Pennington, M. C. (1999). Computer-aided pronunciation pedagogy: Promise, limitations, directions. *Computer Assisted Language Learning*, 12(5), 427–440. <https://doi.org/10.1076/call.12.5.427.5693>
- Peppé, S., & McCann, J. (2003). Assessing intonation and prosody in children with atypical language development: the PEPS-C test and the revised version. *Clinical Linguistics & Phonetics*, 17(4–5), 345–354. <https://doi.org/10.1080/0269920031000079994>
- Pfordresher, P. Q., & Brown, S. (2009). Enhanced production and perception of musical pitch in tone language speakers. *Attention, Perception, & Psychophysics*, 71(6), 1385–1398. <https://doi.org/10.3758/APP.71.6.1385>
- Pickering, L. (2001). The role of tone choice in improving ITA communication in the classroom. *TESOL Quarterly*, 35(2), 233–255. <https://doi.org/10.2307/3587647>
- Pickering, L. (2009). Intonation as a pragmatic resource in ELF interaction. *Intercultural Pragmatics*, 6(2), 235–255. <https://doi.org/10.1515/IPRG.2009.013>
- Pickering, L. (2018). *Discourse intonation: A discourse-pragmatic approach to teaching the pronunciation of English*. University of Michigan Press.
- Pickering, L., Hu, G. G., & Baker, A. (2012). The pragmatic function of intonation: Cueing agreement and disagreement in spoken English discourse and implications for ELT. In J. Romero-Trillo (Ed.), *Pragmatics and prosody in English language teaching* (pp. 199–218). Springer.
- Pierrehumbert, J. B. (1980). *The phonology and phonetics of English intonation*. [Unpublished doctoral dissertation]. Massachusetts Institute of Technology, MA, USA.
- Pierrehumbert, J., & Hirschberg, J. (1990). The meaning of intonational contours in English. In P.R. Cohen, J. Morgan, & M.E. Pollack (Eds.), *Intentions in communication*, (pp. 271–311). MIT Press.
- Pinto, D., & Raschio, R. (2007). A comparative study of requests in heritage speaker Spanish, L1 Spanish, and L1 English. *International Journal of Bilingualism*, 11(2), 135–155. <https://doi.org/10.1177/13670069070110020101>

- Plonsky, L., & Oswald, F. L. (2014). How big is “big”? Interpreting effect sizes in L2 research. *Language Learning*, 64(4), 878-912. <https://doi.org/10.1111/lang.12079>
- Plonsky, L., & Zhuang, J. (2019). A meta-analysis of L2 pragmatics instruction. In *The Routledge handbook of second language acquisition and pragmatics* (pp. 287–307). Routledge.
- Puga, K., Fuchs, R., Setter, J. and Mok, P. (2017, August 20–24). *The perception of English intonation patterns by German L2 speakers of English* [Conference Presentation]. *Interspeech 2017*, Stockholm, Sweden.
- Purpura, J. E. (2004). *Assessing grammar* (Vol. 8). Cambridge University Press.
- R Core Team (2021). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna, Austria. Retrievable from <http://www.R-project.org/>.
- Rafiei, M., Fakhraee Faruji, L., & Azad, M. (2019). The Relationship between Working Memory, Speaking Accuracy and Length of Utterances of Iranian EFL Learners. *Iranian Journal of Learning and Memory*, 2(6), 59–67. <https://doi.org/10.22034/IEPA.2019.95554>
- Rallo Fabra, L., & Romero, J. (2012). Native Catalan learners' perception and production of English vowels. *Journal of Phonetics*, 40(3), 491-508. <https://doi.org/10.1016/j.wocn.2012.01.001>
- Ramírez Verdugo, D. (2008). A cross-linguistic study on the pragmatics of intonation in directives. In J. Romero-Trillo (Ed.), *Pragmatics and Corpus Linguistics*, (pp. 205–233). De Gruyter Mouton.
- Ramírez Verdugo, D. (2006). A Study of Intonation Awareness and Learning in Non-native Speakers of English. *Language Awareness*, 15(3), 141–159. <https://doi.org/10.2167/la404.0>
- Rasier, L., & Hiligsmann, P. (2007). Prosodic transfer from L1 to L2. Methodological issues and description. *Nouveaux Cahiers de Linguistique Française*, 28, 41–66.
- Rockwell, P. (2000). Lower, slower, louder: Vocal cues of sarcasm. *Journal of Psycholinguistic Research*, 29(5), 483–495. <https://doi.org/10.1023/A:1005120109296>
- Romero-Trillo, J. (2019). Prosodic pragmatics and feedback in intercultural communication. *Journal of Pragmatics*, 151, 91–102. <https://doi.org/10.1016/j.pragma.2019.02.018>
- Romero-Trillo, J., & Newell, J. (2012). Prosody and feedback in native and non-native speakers of English. In J. Romero-Trillo (Ed.), *Pragmatics and prosody in English language teaching* (pp. 117–131). Springer.
- Rota, G., & Reiterer, S. M. (2009). *Cognitive aspects of pronunciation talent* (pp. 67–96). De Gruyter Mouton.

- Sag, I., & Liberman, M. (1975, April). The intonational disambiguation of indirect speech acts. In G. Lakoff (Ed.), *Proceedings of the Chicago Linguistics Society* (Vol. 11, pp. 487–497).
- Sağdıç, A. (2022). *Learning by simulating: Second language pragmatic development in a technology-mediated task-based simulation with feedback* [Doctoral dissertation, Georgetown University]. Georgetown University Library.
- Saito, K. (2012). Effects of instruction on L2 pronunciation development: A synthesis of 15 quasi-experimental training studies. *TESOL Quarterly*, 46(4), 842–854. <https://doi.org/10.1002/tesq.67>
- Saito, K. (2019). The role of aptitude in second language segmental learning: The case of Japanese learners' English /ɪ/ pronunciation attainment in classroom settings. *Applied Psycholinguistics*, 40(1), 183–204. <https://doi.org/10.1017/S0142716418000528>
- Saito, K., & Plonsky, L. (2019). Effects of second language pronunciation teaching revisited: A proposed measurement framework and meta-analysis. *Language Learning*, 69(3), 652–708. <https://doi.org/10.1111/lang.12345>
- Saito, Y., & Saito, K. (2017). Differential effects of instruction on the development of second language comprehensibility, word stress, rhythm, and intonation: The case of inexperienced Japanese EFL learners. *Language Teaching Research*, 21(5), 589–608. <https://doi.org/10.1177/1362168816643111>
- Saito, K., Sun, H., & Tierney, A. (2020a). Domain-general auditory processing determines success in second language pronunciation learning in adulthood: A longitudinal study. *Applied Psycholinguistics*, 41(5), 1083–1112. <https://doi.org/10.1017/S0142716420000491>[Opens in a new window]
- Saito, K., Sun, H., & Tierney, A. (2020b). Brief report: Test-retest reliability of explicit auditory processing measures. *BioRxiv*. <https://doi.org/10.1101/2020.06.12.149484>
- Salgado, E. F. (2011). *The pragmatics of requests and apologies: Developmental patterns of Mexican students* (Vol. 212). John Benjamins Publishing.
- Savić, M. (2015). “Can I very please borrow it”: Request development in young Norwegian EFL learners. *Intercultural Pragmatics*, 12(4), 443–480. <https://doi.org/10.1515/ip-2015-0023>
- Schielzeth, H., Dingemanse, N. J., Nakagawa, S., Westneat, D. F., Allogue, H., Teplitsky, C., ... & Araya-Ajoy, Y. G. (2020). Robustness of linear mixed-effects models to violations of distributional assumptions. *Methods in Ecology and Evolution*, 11(9), 1141–1152. <https://doi.org/10.1111/2041-210X.13434>
- Schmid, C., & Moosmüller, S. (2017, December). L2 acquisition and L1 attrition in Bosnian-German late bilinguals-Phonetic observations at the segmental level. In E. Babatsouli

- (Ed.), *Proceedings of the International Symposium on Monolingual and Bilingual Speech* (p. 251–257). Institute of Monolingual and Bilingual Speech.
- Schön, D., Magne, C., & Besson, M. (2004). The music of speech: Music training facilitates pitch processing in both music and language. *Psychophysiology*, 41(3), 341–349. <https://doi.org/10.1111/1469-8986.00172.x>
- Schulze, K., & Koelsch, S. (2012). Working memory for speech and music. *Annals of the New York Academy of Sciences*, 1252(1), 229–236. <https://doi.org/10.1111/j.1749-6632.2012.06447.x>
- Schwab, S., & Dellwo, V. (2019, August 5–9). *Music and L2 prosody: the role of musical aptitude on the discrimination of stress contrasts* [Paper presentation]. 19th International Congress of Phonetic Sciences, Melbourne, Australia. <https://doi.org/10.5167/uzh-184768>
- Searle, J. R. (1976). A classification of illocutionary acts. *Language in Society*, 1, 1–23. <https://doi.org/10.1017/S0047404500006837>
- Sedivy, J. (2014). *Language in mind: An introduction to psycholinguistics*. Oxford University Press.
- Setter, J., & Jenkins, J. (2005). Pronunciation. *Language Teaching*, 38(1), 1–17.
- Shen, X. (1990). The prosody of Mandarin Chinese. *University of California Publications in Linguistics* (Vol. 118). University of California Press.
- Sheppard, C., Hayashi, C., & Ohmori, A. (2007, August 6–10). *Factors accounting for attainment in foreign language phonological competence* [Paper presentation], *International Congress of Phonetic Sciences*, Saarbrücken, Germany.
- Shih, C. (1988). Tone and intonation in Mandarin. *Working Papers of the Cornell Phonetics Lab*, 3, 83–109.
- Shin, J. (2020). A meta-analysis of the relationship between working memory and second language reading comprehension: Does task type matter?. *Applied Psycholinguistics*, 41(4), 873–900. <https://doi.org/10.1017/S0142716420000272>
- Shin, J., Dronjic, V., & Park, B. (2019). The interplay between working memory and background knowledge in L2 reading comprehension. *TESOL Quarterly*, 53(2), 320–347. <https://doi.org/10.1002/tesq.482>
- Shively, R. L., Menke, M. R., & Manzón-Omundson, S. M. (2008). Perception of irony by L2 learners of Spanish. *Issues in Applied Linguistics*, 16(2), 101–132. <https://doi.org/10.5070/L4162005096>
- Skehan, P. (1989). *Individual differences in second-language learning*. Edward Arnold.

- Slevc, L. R., & Miyake, A. (2006). Individual differences in second-language proficiency: Does musical ability matter?. *Psychological Science*, 17(8), 675–681.
- Snow, D. (1998). Children's imitations of intonation contours: Are rising tones more difficult than falling tones?. *Journal of Speech, Language, and Hearing Research*, 41(3), 576–587. <https://doi.org/10.1044/jslhr.4103.576>
- Snow, D., & Balog, H. L. (2002). Do children produce the melody before the words? A review of developmental intonation research. *Lingua*, 112(12), 1025–1058. [https://doi.org/10.1016/S0024-3841\(02\)00060-8](https://doi.org/10.1016/S0024-3841(02)00060-8)
- So, C. K., & Best, C. T. (2010). Cross-language perception of non-native tonal contrasts: Effects of native phonological and phonetic influences. *Language and Speech*, 53(2), 273–293. <https://doi.org/10.1177/0023830909357156>
- Son, J. (2018). *Acquisition of Spanish intonation by Native Korean speakers*. [Unpublished doctoral dissertation]. University of California, Los Angeles, CA, USA.
- Southwood, H., & Flege, J. (1999). The validity and reliability of scaling foreign accent. *Clinical Linguistics & Phonetics*, 13, 335–349.
- Spencer-Oatey, H. (1997). Unequal relationships in high and low power distance societies: A comparative study of tutor–student role relations in Britain and China. *Journal of Cross-Cultural Psychology*, 28(3), 284–302. <http://dx.doi.org/10.1177/0022022197283005>
- Staples, S. (2015). *The discourse of nurse–patient interactions: Contrasting the communicative styles of US and international nurses (Vol. 72)*. John Benjamins Publishing.
- Stepanov, A., Kodrič, K. B., & Stateva, P. (2020). The role of working memory in children's ability for prosodic discrimination. *Plos One*, 15(3), e0229857. <https://doi.org/10.1371/journal.pone.0229857>
- Sweet, H. (1980). *A Primer of Phonetics*. Clarendon.
- Suh, J. S. (1999). Pragmatic perception of politeness in requests by Korean learners of English as a second language. *International Review of Applied Linguistics in Language Teaching* (37)3, 195–213. <https://doi.org/10.1515/iral.1999.37.3.195>
- Sun, S. (2011). Meta-analysis of Cohen's kappa. *Health Services and Outcomes Research Methods*, 11, 145–163. <https://doi.org/10.1007/s10742-011-0077-3>
- Sun, H., Saito, K., & Tierney, A. (2021). A longitudinal investigation of explicit and implicit auditory processing in L2 segmental and suprasegmental acquisition. *Studies in Second Language Acquisition*, 43(3), 551–573. <https://doi.org/10.1017/S0272263120000649>
- Suzuki, S., & Kormos, J. (2020). Linguistic dimensions of comprehensibility and perceived fluency: An investigation of complexity, accuracy, and fluency in second language

- argumentative speech. *Studies in Second Language Acquisition*, 42(1), 143–167.
<https://doi.org/10.1017/S0272263119000421>
- Suzukida, Y. (2021). The contribution of individual differences to L2 pronunciation learning: Insights from research and pedagogical implications. *RELC Journal*, 52(1), 48–61.
<https://doi.org/10.1177/0033688220987655>
- Suzukida, Y., & Saito, K. (2019). Which segmental features matter for successful L2 comprehensibility? Revisiting and generalizing the pedagogical value of the functional load principle. *Language Teaching Research*, 25(3), 431–450.
<https://doi.org/10.1177/1362168819858246>
- Sydorenko, T., Maynard, C., & Guntly, E. (2014). Rater behaviour when judging language learners’ pragmatic appropriateness in extended discourse. *TESL Canada Journal*, 32(1), 19–41. <https://doi.org/10.18806/tesl.v32i1.1197>
- Sykes, J. M. (2008). *A dynamic approach to social interaction: synthetic immersive environments & Spanish pragmatics*. [Unpublished doctoral dissertation]. University of Minnesota, Minneapolis, MN, USA.
- Sykes, J. M., & Cohen, A. D. (2008). Observed learner behavior, reported use, and evaluation of a website for learning Spanish pragmatics. In M.A. Bowles, R. Foote, S. Perpiñán, & R.M.Bhatt (Eds.), *Selected Proceedings of the 2007 Second Language Research Forum* (pp. 144–157). Cascadia Press.
- Taguchi, N. (2006). Analysis of appropriateness in a speech act of request in L2 English. *Pragmatics*, 16(4), 513–533. <https://doi.org/10.1075/prag.16.4.05tag>
- Taguchi, N. (2008). The effect of working memory, semantic access, and listening abilities on the comprehension of conversational implicatures in English as a foreign language. *Pragmatics and Cognition*, 16(3), 517–538. <https://doi.org/10.1075/pc.16.3.05tag>
- Taguchi, N. (2011a). Rater variation in the assessment of speech acts. *Pragmatics*, 21(3), 453–471. <https://doi.org/10.1075/prag.21.3.08tag>
- Taguchi, N. (2011b). Teaching pragmatics: Trends and issues. *Annual Review of Applied Linguistics*, 31, 289–310. <https://doi.org/10.1017/S0267190511000018>
- Taguchi, N. (2012). *Context, individual differences, and pragmatic competence*. Multilingual Matters.
- Taguchi, N. (2013). Refusals in L2 English: Proficiency effects on appropriateness and fluency. In O. Marti–Arandiz & P. Salazar–Campillo (Eds.), *Refusals in instructional contexts and beyond* (pp. 101–119). Brill.
- Taguchi, N. (2015). “Contextually” speaking: A survey of pragmatic learning abroad, in class, and online. *System*, 48, 3–20. <https://doi.org/10.1016/j.system.2014.09.001>

- Taguchi, N. (2018). Advanced pragmatic competence. In P. A. Malovrh & Benati, A. (Eds.), *The handbook of advanced proficiency in second language acquisition* (pp. 505–526). Wiley Blackwell.
- Taguchi, N. (Ed.). (2019). *The Routledge handbook of second language acquisition and pragmatics*. Routledge.
- Taguchi, N. (2020). Digitally mediated remote learning of pragmatics. *Foreign Language Annals*, 53(2), 353–358. <https://doi.org/10.1111/flan.12455>
- Taguchi, N., Gomez-Laich, M. P., & Arrufat-Marques, M. J. (2016). Comprehension of indirect meaning in Spanish as a foreign language. *Foreign Language Annals*, 49(4), 677–698. <https://doi.org/10.1111/flan.12230>
- Taguchi, N., Hirschi, K., & Kang, O. (2021). Longitudinal L2 development in the prosodic marking of pragmatic meaning: Prosodic changes in L2 speech acts and individual factors. *Studies in Second Language Acquisition*, 44(3), 843–858. <https://doi.org/10.1017/S0272263121000486>
- Taguchi, N., & Li, S. (2020). Contrastive pragmatics and second language (L2) pragmatics: Approaches to assessing L2 speech act production. *Contrastive Pragmatics*, 2(1), 1–23. <https://doi.org/10.1163/26660393-BJA10014>
- Taguchi, N., Naganuma, N., & Budding, C. (2015). Does Instruction Alter the Naturalistic Pattern of Pragmatic Development? A Case of Request Speech Act. *TESL-EJ*, 19(3), n3.
- Taguchi, N., & Roever, C. (2017). *Second language pragmatics*. Oxford University Press.
- Taguchi, N., & Sykes, J. M. (Eds.). (2013). *Technology in interlanguage pragmatics research and teaching* (Vol. 36). John Benjamins Publishing.
- Takahashi, S. (1996). Pragmatic transferability. *Studies in Second Language Acquisition*, 18(2), 189–223. <https://doi.org/10.1017/S0272263100014881>
- Takahashi, S. (2019). Individual learner considerations in SLA and L2 pragmatics. In N. Taguchi (Ed.), *The Routledge handbook of second language acquisition and pragmatics* (pp. 429–444). Routledge.
- Takahashi, T., & Beebe, L. M. (1987). The development of pragmatic competence by Japanese learners of English. *JALT Journal*, 8(2), 131–155.
- Takahashi, S., & DuFon, M. A. (1989). *Cross-Linguistic Influence in Indirectness: The Case of English Directives Performed by Native Japanese Speakers*, (ERIC Document Reproduction Service No. ED 370439). University of Hawaii at Manoa.
- Tajeddin, Z., & Alemi, M. (2014). Pragmatic rater training: Does it affect non-native L2 teachers' rating accuracy and bias. *Iranian Journal of Language Testing*, 4(1), 66–83.

- Tanck, S. (2002). Speech act sets of refusal and complaint: A comparison of native and non-native English speakers' production. *TESOL Working Papers*, 2, 65–81.
- Tareva, E., & Polushkina, T. (2020). Discourse-Based Approach to Practicing L2 Prosody in a Workplace Communication Course: A Review of Recent Research and Instruction. In L. Khalyapina (Ed.), *Examining Content and Language Integrated Learning (CLIL) Theories and Practices*, (pp. 226–246). Hershey.
- Tatton, H. (2008). “Could You, Perhaps, Pretty Please?”: Request Directness in Cross-Cultural Speech Act Realization. *Studies in Applied Linguistics and TESOL*, 8(2), 1–4.
<https://doi.org/10.7916/salt.v8i2.1479>
- Tatsuki, D. (2019). Instructional material development in L2 pragmatics. In N. Taguchi (Ed.), *The Routledge handbook of second language acquisition and pragmatics* (pp. 322–337). Routledge.
- Tickle, A.L. (1991). Japanese refusals in a business setting. *Papers in Applied Linguistics* 6(2), 84–108.
- Tierney, A., White-Schwoch, T., MacLean, J., & Kraus, N. (2017). Individual differences in rhythm skills: Links with neural consistency and linguistic ability. *Journal of Cognitive Neuroscience*, 29(5), 855–868. https://doi.org/10.1162/jocn_a_01092
- Thomson, R. I., & Derwing, T. M. (2015). The effectiveness of L2 pronunciation instruction: A narrative review. *Applied Linguistics*, 36(3), 326–344.
<https://doi.org/10.1093/applin/amu076>
- Trofimovich, P., & Baker, W. (2007). Learning prosody and fluency characteristics of second language speech: The effect of experience on child learners' acquisition of five suprasegmentals. *Applied Psycholinguistics*, 28(2), 251–276.
[doi:10.1017/S0142716407070130](https://doi.org/10.1017/S0142716407070130)
- Trofimovich, P., & Gatbonton, E. (2006). Repetition and focus on form in processing L2 Spanish words: Implications for pronunciation instruction. *The Modern Language Journal*, 90(4), 519–535. <https://doi.org/10.1111/j.1540-4781.2006.00464.x>
- Trofimovich, P., Kennedy, S., & Foote, J. A. (2015). Variables Affecting L2 Pronunciation Development. In M. Reed & J.M. Levis (Eds). *The Handbook of English Pronunciation*, (pp. 363 – 373). Wiley Blackwell.
- Trosborg, A. (2011). *Interlanguage pragmatics: Requests, complaints, and apologies (Vol. 7)*. Walter de Gruyter.
- Trott, S., & Bergen, B. (2019). Individual differences in mentalizing capacity predict indirect request comprehension. *Discourse Processes*, 56(8), 675–707.
<https://doi.org/10.1080/0163853X.2018.1548219>

- Trouvain, J., Jügler, J., & Laprie, Y. (2016, March 2–4). *On the assessment of computer-assisted pronunciation training tools* [Paper presentation], Studententexte zur Sprachkommunikation: Elektronische Sprachsignalverarbeitung, Leipzig, Germany.
- Usó–Juan, E. (2010). Requests. In A. Martínez–Flor & E. Usó–Juan (Eds.), *Speech act performance: Theoretical, empirical, and methodological issues* (Vol. 26). John Benjamins Publishing.
- Van Hedger, S. C., & Nusbaum, H. C. (2018). Individual differences in absolute pitch performance: Contributions of working memory, musical expertise, and tonal language background. *Acta Psychologica*, 191, 251–260. <https://doi.org/10.1016/j.actpsy.2018.10.007>
- Walters, F. S. (2007). A conversation–analytic hermeneutic rating protocol to assess L2 oral pragmatic competence. *Language Testing*, 24(2), 155–183. <https://doi.org/10.1177/0265532207076362>
- Wang, C., & Zhao, H. (2020). The impact of COVID–19 on anxiety in Chinese university students. *Frontiers in Psychology*, 11, 1–8. <https://doi.org/10.3389/fpsyg.2020.01168>
- Waters, G. S., & Caplan, D. (1996). The measurement of verbal working memory capacity and its relation to reading comprehension. *The Quarterly Journal of Experimental Psychology Section A*, 49(1), 51–79. <https://doi.org/10.1080/713755607>
- Wen, X. (2014). Pragmatic development: An exploratory study of requests by learners of Chinese. In Z. Han (Ed.), *Studies in second language acquisition of Chinese* (pp. 30–56). Multilingual Matters.
- Wennerstrom, A. (1994). Intonational meaning in English discourse: A study of non–native speakers. *Applied Linguistics*, 15(4), 399–420. <https://doi.org/10.1093/applin/15.4.399>
- Wennerstrom, A. (1998). Intonation as cohesion in academic discourse: A study of Chinese speakers of English. *Studies in Second Language Acquisition*, 20(1), 1–25. <https://doi.org/10.1017/S0272263198001016>
- Wennerstrom, A. (2001). *The music of everyday speech: Prosody and discourse analysis*. Oxford University Press.
- Wichmann, A. (2000). *Intonation in text and discourse*. Longman.
- Wichmann, A. (2015). Functions of Intonation in Discourse. In M. Reed & J.M. Levis (Eds.), *The handbook of English pronunciation*, (pp. 175–189). Wiley.
- Williamson, V. J., Baddeley, A. D., & Hitch, G. J. (2010). Musicians' and nonmusicians' short–term memory for verbal and musical sequences: Comparing phonological similarity and pitch proximity. *Memory & Cognition*, 38(2), 163–175. <https://doi.org/10.3758/MC.38.2.163>

- Winans, M. (2020). Email requests: Politeness evaluations by instructors from diverse language backgrounds. *Language Learning & Technology*, 24(2), 104–118. <http://hdl.handle.net/10125/44728>
- Wing, H. D. (1948). *Standardised Tests of Musical Intelligence*. City of Sheffield Training College.
- Winke, P., Gass, S., & Sydorenko, T. (2013). Factors influencing the use of captions by foreign language learners: An eye-tracking study. *The Modern Language Journal*, 97(1), 254–275. <https://doi.org/10.1111/j.1540-4781.2013.01432.x>
- Won, J. H., Tremblay, K., Clinard, C. G., Wright, R. A., Sagi, E., & Svirsky, M. (2016). The neural encoding of formant frequencies contributing to vowel identification in normal hearing listeners. *The Journal of the Acoustical Society of America*, 139(1), 1–11. <https://doi.org/10.1121/1.4931909>
- Wong, P. C., & Perrachione, T. K. (2007). Learning pitch patterns in lexical identification by native English-speaking adults. *Applied Psycholinguistics*, 28(4), 565–585. <https://doi.org/10.1017/S0142716407070312>
- Woodfield, H. (2008). Interlanguage requests: a contrastive study. In M. Pütz & JoAnne Nef-van Aertselaer (Eds.), *Developing Contrastive Pragmatics. Interlanguage and Cross-Cultural Perspectives*, (pp. 231–264). De Gruyter Mouton.
- Xu, Y. (2015). Intonation in Chinese. In (W. S–Y. Wang & C. Sun (Eds.), *The Oxford Handbook of Chinese Linguistics*, (pp. 490–502). Oxford Academic.
- Yan, X. (2014). An examination of rater performance on a local oral English proficiency test: A mixed-methods approach. *Language Testing*, 31(4), 501–527. <https://doi.org/10.1177/0265532214536171>
- Yan, X., & Ginther, A. (2017). Listeners and raters: Similarities and differences in evaluation of accented speech. In O. Kang & A. Ginther, (Eds.), *Assessment in second language pronunciation*, (pp. 67–88). Routledge.
- Yan, X. (2020). Unpacking the relationship between formulaic sequences and speech fluency on elicited imitation tasks: Proficiency level, sentence length, and fluency dimensions. *TESOL Quarterly*, 54(2), 460–487. <https://doi.org/10.1002/tesq.556>
- Yaw, K. (2022). *Impact of Accent Familiarity Training on Listener Processing of Second-Language Accented Speech* (Publication No. 29206203) [Doctoral dissertation, Northern Arizona University]. ProQuest Dissertations Publishing.
- Youn, S. J. (2015). Validity argument for assessing L2 pragmatics in interaction using mixed methods. *Language Testing*, 32(2), 199–225. <https://doi.org/10.1177/0265532214557113>
- Yuan, J. (2011). Perception of intonation in Mandarin Chinese. *The Journal of the Acoustical Society of America*, 130(6), 4063–4069. <https://doi.org/10.1121/1.3651818>

- Yuan, C., González–Fuente, S., Baills, F., & Prieto, P. (2019). Observing pitch gestures favors the learning of Spanish intonation by Mandarin speakers. *Studies in Second Language Acquisition*, 41(1), 5–32. <https://doi.org/10.1017/S0272263117000316>
- Youn, S. J. (2018). Task–based needs analysis of L2 pragmatics in an EAP context. *Journal of English for Academic Purposes*, 36, 86–98. <https://doi.org/10.1016/j.jeap.2018.10.005>
- Zalbidea, J., & Sanz, C. (2020). Does learner cognition count on modality? Working memory and L2 morphosyntactic achievement across oral and written tasks. *Applied Psycholinguistics*, 41(5), 1171–1196. <https://doi.org/10.1017/S0142716420000442>[Opens in a new window]
- Zembrzuski, D., Marecka, M., Otwinowska, A., Zajbt, E., Krzemiński, M., Szewczyk, J., & Wodniecka, Z. (2020). Bilingual children do not transfer stress patterns: Evidence from suprasegmental and segmental analysis of L1 and L2 speech of Polish–English child bilinguals. *International Journal of Bilingualism*, 24(2), 93–114. <https://doi.org/10.1177/1367006918810957>
- Zhang, R., & Yuan, Z. M. (2020). Examining the effects of explicit pronunciation instruction on the development of L2 pronunciation. *Studies in Second Language Acquisition*, 42(4), 905–918. <https://doi.org/10.1017/S0272263120000121>
- Zhao, C., Xiong, Z., & Li, A. (2020, November 5–7). *Using Multimodal Methods in L2 Intonation Teaching for Chinese EFL Learners* [Conference presentation], 23rd Conference of the Oriental COCOSDA International Committee for the Co–ordination and Standardisation of Speech Databases and Assessment Techniques (O–COCOSDA), IEEE.
- Zheng, C., Saito, K., & Tierney, A. (2022). Successful second language pronunciation learning is linked to domain–general auditory processing rather than music aptitude. *Second Language Research*, 38(3), 477–497. <https://doi.org/10.1177/0267658320978493>
- Zhuang, Y. (2015). *Evaluating the effectiveness of teaching intonation to learners in an intensive English program*. [Unpublished doctoral dissertation]. Northern Arizona University, AZ, USA.
- Zoom Video Communications (2021). *Zoom video conferencing tool*. Retrievable from <https://zoom.us/>.
- Zybert, J., & Stępień, S. (2009). Musical intelligence and foreign language learning. *Research in Language*, 7(1), 99–111. <https://doi.org/10.2478/v10015-009-0007-4>

Appendix A: Learner Background Questionnaire

Background questionnaire

1. What is your name? (*input*)
2. What is your email? (*input*)
3. What is your gender?
 - ☐ Female
 - ☐ Male
 - ☐ Nonbinary / third gender
 - ☐ Prefer not to say
4. Age (*dropdown*)
5. What is your first language? (The language you learned first when you were a baby).
(*input*)
6. What other languages do you speak? (*input*)
7. Have you ever lived in an English-speaking country?
 - ☐ Yes
 - ☐ No
8. If yes, where and how long? (*input*)

Proficiency

9. If your first language is NOT English, have you taken a proficiency test (e.g., TOEFL, IELTS, Duolingo)? If so, choose which one. If not, chose "Not taken".
 - ☒ TOEFL iBT
 - ☐ IELTS
 - ☐ Duolingo
 - ☐ TOEIC
 - ☐ Not taken
10. What was your score in that proficiency test? (*input*)

Musical Training

11. Can you play a musical instrument?
 - ☐ Yes

- No

12. If yes, when did you start playing it (year)? (*input*)

13. Have you ever taken music classes?

- Yes

- No

14. If yes, what classes and when (years)? (*input*)

15. Have you received your musical education at a school?

- Yes

- No

16. If so, what grade did you achieve? (*input*)

Appendix B: Rater Background Questionnaire

1. Email (*input*)
2. Gender you identify with (*drop-down*):
 - a. Female
 - b. Male
 - c. Nonbinary / third gender
 - d. Prefer not to say
3. Age (*dropdown*)
4. What country were you born in? (*input*)
5. What is your first language? (*input*)
6. What other languages do you speak? (*input*) Please choose your proficiency level for each language.

_____ level: beginner, intermediate, advanced (*multiple choice*)

_____ level: beginner, intermediate, advanced (*multiple choice*)

_____ level: beginner, intermediate, advanced (*multiple choice*)

_____ level: beginner, intermediate, advanced (*multiple choice*)

Linguistic training

1. Have you ever taken a course related to L2 speech (e.g., Listening and Speaking, Phonetics and Phonology, etc.)? (*multiple choice*)
 - a. yes
 - b. no
2. If yes, which course(s) did you take? (*input*)

Teaching experience

1. How many years of teaching experience at the college level do you have (e.g., community college, university)? (*input*)
2. What classes have you taught? (*input*)

Musical training

1. Can you play a musical instrument?
 - Yes
 - No
2. If yes, when did you start playing it (year)? (*input*)
3. Have you ever taken music classes?
 - Yes
 - No
4. If yes, what classes and when (years)? (*input*)

5. Have you received your musical education at a school?

☐ Yes

☐ No

6. If so, what grade did you achieve? (*input*)

Appendix C: Elicited Imitation Task

(adopted from Yan, 2020)

1. Final exams are based on the content covered in the lectures.
2. Students can buy textbooks at the bookstores on campus.
3. There is a small gift for you after you complete the survey.
4. Many students worry that they might not do well on their exams.
5. Sometimes it's helpful to ask questions in class as opposed to keeping them to yourself.
6. Most students declare their major at the end of their sophomore year in college.
7. College students can ride the bus for free as long as they have a valid student ID.
8. Students can keep the books that they borrow from the library for a semester.
9. When you look at the course schedule, you will see the dates for midterm and final exams.
10. Regular workouts benefit the body as well as the mind.
11. You should come up with a topic for your project by midterm.
12. Students who enjoy working in groups are more likely to succeed.
13. It doesn't matter if you work alone or in a group on your homework.
14. You can also talk to senior students about selecting courses.
15. It's hard to express your ideas if your language skills are low.
16. Many students live off campus because the rent is much lower.
17. If you have a morning class, you should go to bed early the night before.
18. When you take courses here, attendance often counts as a part of the final grades.
19. Meeting people and making friends should be an important part of your college life.
20. Students can take courses that have nothing to do with their major areas of study.
21. You can tell me what questions you have on the final project during my office hours.
22. Students should know that they can also borrow books from libraries of other schools.
23. Taking a part-time job on campus has been shown to help students succeed in college.
24. You should talk to your advisor if you are not sure what courses to take next semester.

Appendix D: Instructional Training Lesson Plans

Appendix D1: Day 1

Introduction to prosody and appropriateness

Materials (Introductory Module 1 on the training website):

- Introductory video
- Video Lecture 1 explaining the main concepts in the course: prosody (tone choices, prominence) and politeness in speech acts; Demonstration of Charts.js (how to interpret the prosodic contour)
- Sample sound files to practice identifying tone choices and prominence

Sound files

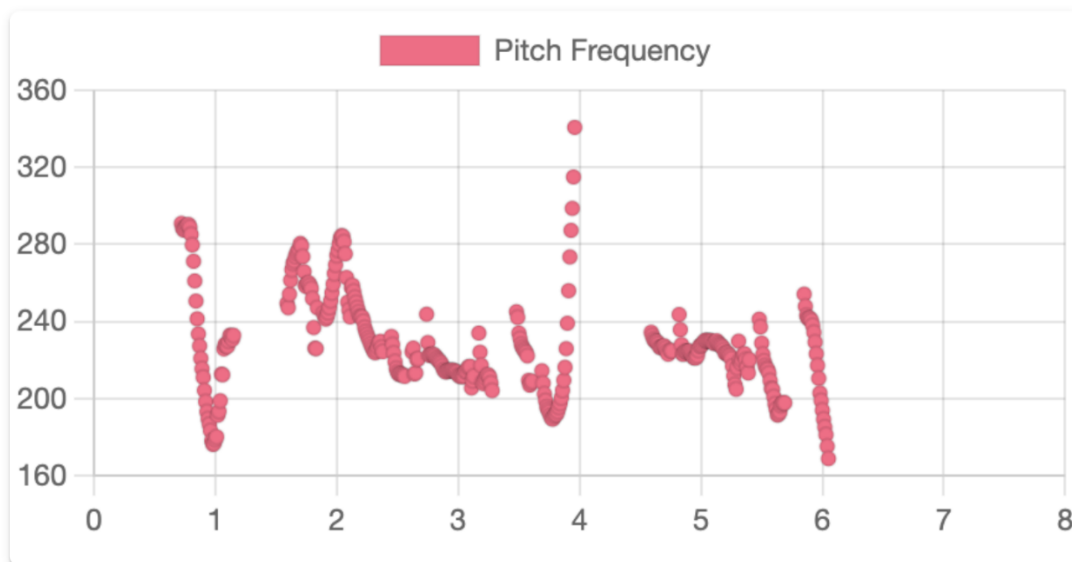
Sound file ID	Lexico-grammatical content
W1D1-1	Hi, I was wondering if I could rearrange my schedule and throw in English 105.
W1D1-2	Hi professor Smith, I was wondering if I can get a letter of recommendation for a scholarship I'm applying for.
W1D1-3	I have a very important test on Friday. I was hoping I could take a day off on Thursday.

Sequence:

1. Ss log in to the training website following a given link and watch the first introductory video. Video highlights:
 - Instructor (researcher) introduction, purpose and overview of the course, overview of the learning platform (where to find exercises, how to submit work)
2. Ss watch Lecture 1. Lecture highlights:
 - Prosody/intonation
 - i. Definitions: prosody *or* intonation = the melody of language. The way we use our voice to convey information.
 - Introduction of Brazil's framework:
 - i. Utterances can be divided into thought groups (intonation phrases). Each thought group has its own intonation. It contains one or more prominent words and has a tone choice.
 - Prominence – definition, how it is marked in the course (using the example sentence).
 - i. Prominence = words that a speaker highlights/emphasizes in discourse, usually by lengthening them or making them louder.
 - Tone choices – definition, how it is marked in the course (using the example sentence).
 - i. Tone choice = movement of the voice at the last prominent word in a thought group (falling rising, level).
 - Speech acts – definition, speech acts of interest in the course

- i. Requests = asking someone to perform a real-life action that would benefit you in some way
- Factors affecting politeness, directness, sincerity, and formality in SAs: requests may differ in these characteristics:
 - i. Who are you talking to? Is it your boss or employer or a friend? (relative power)
 - ii. Do you know them well (a closer friend) or not really (a sales assistant)? (social distance)
 - iii. Are you asking for a big favor or something very small? (ranking of imposition)
- 3. Ss work with a sample sound file and its prosodic contour.
 - Reading the pitch contour
 - Splitting a sound file into thought groups, identifying the tone choice on the last prominent word in a thought group (demonstration on one sound file)
- 4. Ss work with the same example sound file (W1D1-1) from Lecture 1

[Hi ↓] [I was WONDERING if I could rearrange my SCHEDULE↑][and throw in English one-oh-FIVE↓].



- 5. Ss practice identifying the prominence and tone choices on two sample sound files, W1D1-2 and W1D1-3 (see table above) and submit their answers in a quiz on the website. Quiz questions (answer for each sound file).

e.g., Sound file 1:

- How many intonation phrases (or thought groups) does the sound file have?
- How many prominent words does the sound file have? How many words total?

- What is the tone choice on the last prominent word in each thought group?

Appendix D2: Day 2

Module 2: Prominence and tone choices in requests (perceptual identification skills)

Materials:

- Video Lecture 2 – prominence in requests
- Sound files for analyzing

Sound files

Sound file ID	Lexico-grammatical content
W1D2-1 (overuse of prominence)	Hi, I was wondering if you could write me a letter of recommendation.
W1D2-2 (normal prominence)	Hi, I was wondering if you could write me a letter of recommendation.
W1D2-3	Hey I'm feeling a little under the weather (→) I'm sick obviously by my voice (→) and I was wondering (↓) if I can get an extension on the paper we have due tomorrow (↓).
W1D2-6 (falling)	Hey Mr. Steele, I caught a cold, and I was wondering if I could have two extra days.
W1D2-7 (rising)	Hey Mr. Steele, I caught a cold, and I was wondering if I could have two extra days.
W1D2-8 (level)	Hey Mr. Steele, I caught a cold, and I was wondering if I could have two extra days.
Prominence Identification Practice Sentences	
W1D2k	Hi, I was wondering if there is any way I can take a class that I was going to take originally in the spring semester during my fall semester (normal prominence).
W1D2l	Hey, I was wondering if I could drop one of my classes and possibly add a new one (normal prominence).
W1D2m	
W1D2n	I was wondering if I could get a later extension (inappropriate prominence).
W1D2o	But I was just wondering because I'm sick if there is any way I can have two extra days to do it (normal prominence).
W1D2p	I was wondering if there is any way I could do it at a later time (inappropriate prominence).
W1D2q	Hi, I was just wondering if there is any way you could write me a recommendation letter (inappropriate prominence).
W1D2r	Hi professor Smith, I was hoping I could get a letter of recommendation for a scholarship I'm applying for (normal prominence).
W1D2s	Hi, I need a letter of recommendation and I was wondering if you could write it, it's for a scholarship (normal prominence).
W1D2t	I was wondering if I could reschedule when I'm taking the test (inappropriate prominence).
W1D2u	I was wondering if you can help me transfer that credit so I don't have to take it again this year (normal prominence).

Tone Choice Identification Practice Sentences	
W1D2a	I was hoping I could get a letter of recommendation for a scholarship I'm applying for (rising).
W1D2b	Hey, I was wondering if I could drop one of my classes and possibly add a new one (falling).
W1D2c	I was hoping I could get a day off tomorrow (level).
W1D2d	I was wondering if you could write me a letter of recommendation (falling).
W1D2e	Hey Mr. Steele, I caught a cold and I was wondering if I could get two extra days (falling).
W1D2f	I was wondering if I could get you to write me a letter of recommendation (rising).
W1D2g	I was wondering if it's possible to transfer my grades from my old university to this one (rising).
W1D2h	Hi professor, I was wondering if I could change my schedule and include English 105 this semester (level).
W1D2i	I was wondering if I could get my shift covered tomorrow (level)
W1D2j	I was wondering if I could get my shift covered tomorrow (rising)

Sequence:

1. Students watch Lecture 2. Lecture highlights:
 - Explain the nature of requests – on-campus high imposition
 - i. Talking to someone who is “higher” than you in rank – professor, advisor, employer.
 - ii. You don't know them well.
 - iii. You are asking for a big, serious favor (e.g., a letter of recommendation, an extension, a change in grade, an unscheduled day off).
 - How to use prominence (with examples & visuals)
 - i. Prominence in English can be used to highlight certain information in speech, place an emphasis.
 - ii. If we highlight too much, the listener may get confused and distracted. If we don't highlight enough, our speech will sound monotone.
 - iii. Comparison of prominence in sound files **W1D2-1** and **W1D2-2**. Both sound files contain the same utterance, but it is spoken with different prominence.
 - How to use tone choices in requests (with examples & visuals)
 - i. Meaning of falling tones in high-imposition requests = preferred tone to be used in requests. The speaker does not come across as dominant, and it does not assume common ground (the interlocutor does not know what the request is about), sounds sincere. (↓, ↑, →).

Prominence Identification

Students listen to practice sound files. They are asked to select (click on) the words that they think are prominent in a request. In the response box, they are asked to choose the words they heard with marked prominence and share their impressions about the prominence. The learners reach can only move on to the next lesson when they reach 90% accuracy on the task.

- Is the speaker stressing too many/too few words?
- How does it affect your impression about the request?

The screenshot shows a web application interface for a listening exercise. At the top, there is a dark navigation bar with a menu icon, the word 'Home', and user information 'Hello, admin' and 'Logout'. Below the navigation bar, there are icons for settings (gears) and help (question mark). The main content area is divided into two panels. The left panel, titled 'Request 11.', contains the text 'Asking for a change in the schedule.' and a media player showing '0:09 / 0:09'. The right panel, titled 'Choose prominent words in the request.', shows 'You've selected: Hilway clas takes spring fall'. Below this, a text box displays the audio transcript: 'Hi I was wondering if there is any way I can take a class that I was going to take originally in the spring semester during my fall semester'. The words 'Hi', 'way', 'class', 'take', 'spring', and 'fall' are highlighted in green, indicating they were selected by the user.

Tone Choice Identification

Sample sound file: **W1D2-6**

[Hey Mr. Steele↓] [I caught a cold↓] [and I was wondering if I could have two extra days↓]

- Meaning of rising tones = inappropriate and insincere when talking to someone in higher power because it implies dominance. It may also come across as if you are assuming that the listener will agree. It could be used if you are in the middle of an utterance and are finishing a thought.

Sample sound file: **W1D2-7**

[Hey Mr. Steele↑] [I caught a cold↑] [and I was wondering if I could have two extra days↑]

- ii. Meaning of level tones = used only to provide an explanation behind a request (ground for the request). Inappropriate for making the request itself.

Sample sound file: **W1D2-8**

[Hey Mr. Steele→] [I caught a cold→] [and I was wondering if I could have two extra days→]

- iii. Ss listen to short requests and identify the tone choice they hear. They check the box that represents the tone they hear. The learners reach can only move on to the next lesson when they reach 90% accuracy on the task.

 Home

Hello, admin

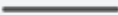
Logout





Request 1.

Asking for a change in the schedule

 0:00 / 0:04   

What tone choice did you hear at the end of the request?

You've selected:

☐ Rising

☐ Falling

☐ Level

Appendix D3: Day 3

Module 3: Prominence and tone choices in requests (perceptual discrimination skills)

Materials:

- Video Lecture 3 – recap on prominence and tone choices in requests
- Sound files for analyzing

Sound file ID	Lexico-grammatical content
W1D2-1 (inappropriate use of prominence; rising intonation)	Hi, I was wondering if you could write me a letter of recommendation.
W1D2-2 (normal prominence; falling intonation)	Hi, I was wondering if you could write me a letter of recommendation.
W1D2-6 (normal prominence; falling)	Hey Mr. Steele, I caught a cold, and I was wondering if I could have two extra days.
W1D2-7 (inappropriate use of prominence; rising)	Hey Mr. Steele, I caught a cold, and I was wondering if I could have two extra days.
W1D2-8 (inappropriate use of prominence; level)	Hi professor, I was wondering if I could change my schedule and include one-oh-five this semester?
W1D2-9 (normal prominence; falling intonation)	Hi professor, I was wondering if I could change my schedule and include one-oh-five this semester?
W1D2-10 (normal prominence; falling intonation)	Hi professor Smith, I was hoping I can get a letter of recommendation for a scholarship I'm applying for?
W1D2-11 (inappropriate use of prominence; rising intonation)	Hi professor Smith, I was hoping I can get a letter of recommendation for a scholarship I'm applying for?
W1D2-12 (normal prominence; falling intonation)	Hey Mr. Steele I caught a cold and I was wondering if I could have two extra days?
W1D2-13 (inappropriate use of prominence; level)	Hey Mr. Steele I caught a cold and I was wondering if I could have two extra days?
W1D2-14 (normal prominence; falling intonation)	I was wondering if I could somehow get in an English 105 this semester?
W1D2-15 (inappropriate use of prominence; rising intonation)	I was wondering if I could somehow get in an English 105 this semester?
W1D2-16 (normal prominence; falling intonation)	Hello professor, I was wondering if I could have two extra days?
W1D2-17 (inappropriate use of prominence; level)	Hello professor, I was wondering if I could have two extra days?

Sequence:

1. Students watch Lecture 3. Lecture highlights:
 - Explain the nature of requests – on-campus high imposition
 - i. Talking to someone who is “higher” than you in rank – professor, advisor, employer.
 - ii. You don’t know them well.
 - How to use prominence (with examples)
 - i. Prominence in English can be used to highlight certain information in speech, place an emphasis.
 - ii. If we highlight too much, the listener may get confused and distracted. If we don’t highlight enough, our speech will sound monotone.
 - How to use tone choices (with examples)
 - i. To sound more polite, falling tones are more appropriate in requests.
 - ii. Level tones are appropriate if you are giving a reason for a request.
 - iii. Rising tones are less appropriate as they imply informality and directness.
2. Learners listen to pairs of requests that are controlled for the lexico-grammatical contents but are different in the tone choices and amount and distribution of prominence. The learners are asked to evaluate the appropriateness of the use of tone choices and the distribution of prominence. The learners reach can only move on to the next lesson when they reach 90% accuracy on the task.

 Home Hello, admin Logout

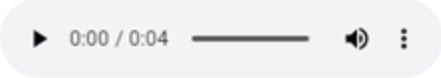
 

Asking for a letter of recommendation.

Request 1.



Request 2.



Which request has more appropriate tone choices?

You've selected: Request 2

Request 1

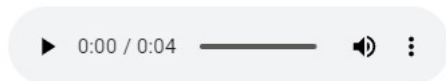
 or

Request 2

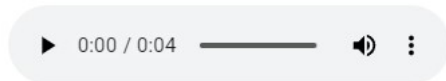


Asking for a letter of recommendation.

Request 1.



Request 2.



Which request has more appropriate prominence?
(Remember that it is appropriate to only have 2-3 prominent words in a thought group).

You've selected: Request 1



Appendix D4: Day 4

Module 4: Prominence and tone choices in requests (controlled production skills)

Materials:

- Video Lecture 7 – a refresher on prominence and tone choices meaning in requests as well as the use of pitch contours.
- Model sound files for comparison with Ss' own production

Equipment needed:

- Speakers / headphones
- Microphone

Sequence:

1. Ss watch Lecture 7 that combines the information on prominence and tone choices in requests from Lectures 3 and 4.
2. Controlled practice 1: Ss read request utterances out loud following marked prominence and tone choices (↓, ↑, →).
 - a. Hi PROFESSOR ↓, I was hoping I can get a letter of RECOMMENDATION for a scholarship I'm APPLYING for ↓.
 - b. I was WONDERING if there is any way I can take a CLASS ↑ that I was going to take in the SPRING ↓ semester.
 - c. HI ↓ I'm HOPING I could take English 105 THIS semester instead of NEXT ↓ semester.
 - d. I NEVER procrastinate ↓ but I procrastinated THIS paper ↓ I was WONDERING if I could get an EXTENSION ↓
 - e. Hi professor STEELE ↓ I was hoping I could talk to you about possibly rescheduling the TEST ↓

For each sentence, learners are asked to look at the pitch contour that is produced and analyze whether they were able to produce the required intonation.

Home

Hello, admin

Logout

Record yourself saying the request below. You can record the request multiple times until you are happy with the recording.

Request 1.

HI, I was WONdering if I could rearrange my SCHEdule and add English one-oh-FIVE (↓).

Press "START" when you are ready to record. Press "STOP" when you are done. If you want to record again, press "START".

Pitch Frequency

PLAY START STOP

Note. The learners' audio recordings are submitted to the researcher. The researcher analyzes the prosodic composition of learners' recordings to see whether they were able to follow the given intonation. If a learner's production meets the requirements of the task, they are allowed to move on to the next lesson. If not, they are required to repeat the lesson.

Appendix D5: Day 5

Module 5: Prominence and tone choices in requests (guided production skills)

Materials:

- Video Lecture 7 – a refresher on prominence and tone choices meaning in requests.
- Model sound files for comparison with Ss' own production

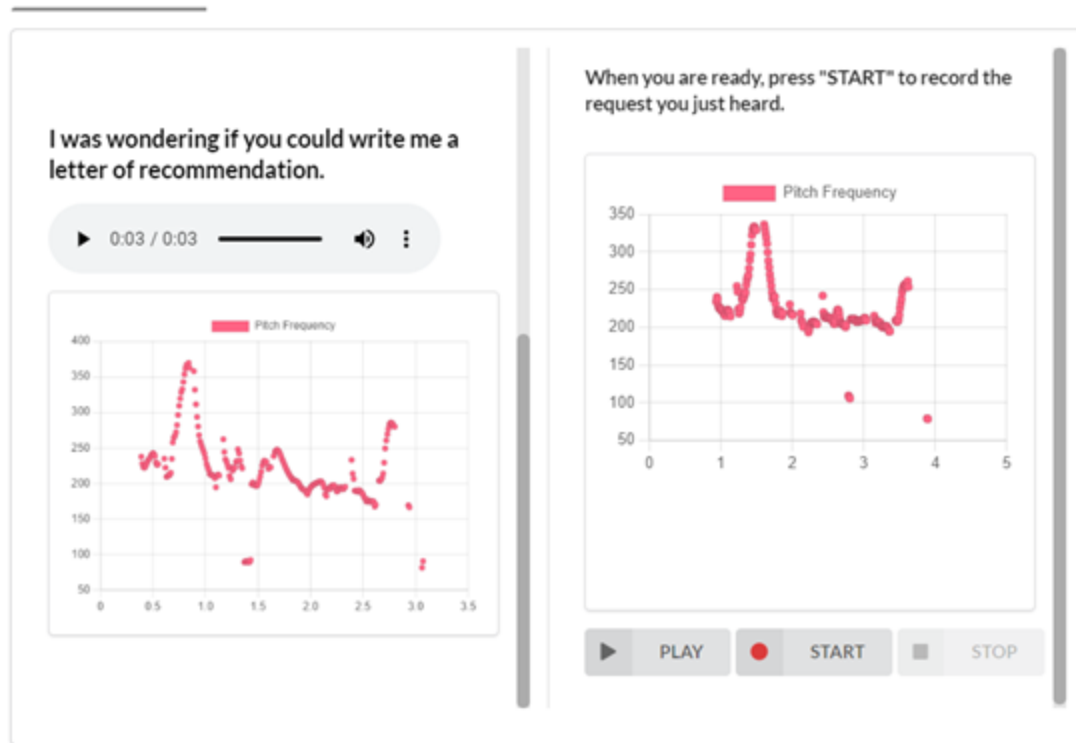
Sound file ID	Lexico-grammatical content
W1D2-2 (normal prominence)	Hi, I was wondering if you could write me a letter of recommendation.
W1D2-3	Hey I'm feeling a little under the weather (→) I'm sick obviously by my voice (→) and I was wondering (↓) if I can get an extension on the paper we have due tomorrow (↓).
W1D2-4	Hey I was wondering if I could drop one of my classes (↓) and possibly add a new one (↓)
W1D2-5	Hi I have an opportunity for a scholarship (→) and I was wondering (↑) if I could get you to write me a letter of recommendation (↓).
W1D2-6 (falling)	Hey Mr. Steele, I caught a cold, and I was wondering if I could have two extra days.

Equipment needed:

- Speakers / headphones
- Microphone

Sequence:

1. Guided practice 1: Listen to a model sentence and repeat it trying to mimic the intonation contour of the model. Compare the intonation contour of the model and your recording to see if they match. If not, record again. Sound files: **W1D2-2 - W1D2-6.**



2. Guided practice 2: Ss complete one video-based request DCT that they received in the immediate pre-test on Gorilla (Appendix E).

Note. The learners' audio recordings are submitted to the researcher. The researcher analyzes the prosodic composition of learners' recordings to see whether they were able to follow the given intonation. If a learner's production meets the requirements of the task, they are allowed to move on to the next lesson. If not, they are required to repeat the lesson.

Appendix D6: Day 6

Prominence and tone choices in requests (communicative practice)

Objective:

- To continue developing Ss' use of prominence and tone choices in requests
- To provide Ss with communicative practice using prominence and tone choices in requests

Materials and equipment:

- Zoom (installed on the learners' computers)
- Role play situations (distributed by the instructor during Zoom)

Sequence:

1. Ss meet with the instructor (researcher) individually on Zoom.
2. Student and instructor role-play two requests. The instructor takes on the higher power role (professor/advisor/employer), while the student performs the lower power role. The student is the one making requests.
3. The role plays are recorded. After the session, the instructor prosodically analyzes each learner's performance on the role plays and provides feedback to the learner.

Role play situations:

The learners will role-play the same situations as they received in the pre-test (Appendix E); however, the format of the role play will be free (not scripted).

Appendix E: Pre- and Post-training DCT Scripts

Situation 1: Ask for a recommendation letter (Request: student – professor)

Instructions: You want to apply for a scholarship offered by your department for the next semester. The applications are due in a month. To apply for the scholarship, you need a letter of recommendation from one of your professors. You want to ask Professor Smith to write the letter of recommendation for you. What do you say to Professor Smith?

* Note that the professor will speak first.

Professor: Hello, what can I do for you?

Your response: I was wondering if you would agree _____.

Professor: Sure, I definitely can write a letter of recommendation for you. When is it due?

Your response: _____.

Professor: Alright, that sounds good. Please email me the detailed information about the scholarship and I will write the letter as soon as I can.

Your response: _____.

Situation 2: Requesting an unscheduled day off (Request: student – employer)

Instructions: You are scheduled to work tomorrow; however, your friend, whom you haven't seen for five years, is coming to visit you on the same day. You want to ask your employer for a day off tomorrow. What do you say to your boss?

* Note that the advisor will speak first.

Boss: So what can I do for you?

Your response: Well, I know I am scheduled to work tomorrow, but _____ . I was wondering if _____.

Boss: Alright, let me see what we can do. I'll have a look at this week's schedule and let you know, okay?

Your response: _____.

Boss: Great, I'll be in touch soon.

Your response: _____.

Situation 3: Change the class registration (Request: student – advisor)

Instructions: You just found out that your best friend is going to take English 105 this semester. You had already registered for all of your classes and were not planning to take English 105 until next year. However, now you want to change your class registration to take English 105 with your friend. You meet with your advisor, Ms. Black. What do you say to her?

* Note that the advisor will speak first.

Advisor: Hey there! What can I do for you today?

Your response: Hi Ms. Black. I was hoping I could change my class registration to take _____.

Advisor: Alright, let me see what I can do. If you want to take English 105 this semester, you won't be able to take Anthropology 102 until next semester. Is that something you want to do?

Your response: _____.

Advisor: Okay, great. You are all set then. Make sure to email the professor about the required textbooks.

Your response: _____.

Situation 4: Ask for an extension (Request: student – professor)

Instructions: Your paper in Professor Smith's paper is due tomorrow. You caught a cold, and you've written only two pages so far. You want to ask for two extra days. What do you say to Professor Steele?

* Note that the professor will speak first.

Professor: Hello! How can I help you?

Your response: Hi professor, I've caught a cold, so I was wondering if I could have _____.

Professor: Well, you are allowed one four-day extension in this class. Are you sure you want to use it on this paper?

Your response: _____.

Professor: Okay then, just make sure to send me the paper by Friday. I hope you'll get better soon!

Your response: _____.

Situation 5: Transfer classes from another university (Request: student-advisor)

Instructions: Your advisor is telling you that you need to take Biology 101 this semester. However, you already took this class last year at a different university. You want to transfer the credits for this class so that you didn't have to take it again. What do you say to your advisor?

* Note that the advisor will speak first.

Advisor: Good morning! How can I help you?

Your response: Hello. I was wondering if you could transfer my credits from
_____.

Advisor: Okay, let me check your transcripts. It does look like you took Biology last year before coming here. Is that what you mean?

Your response: _____.

Advisor: Hmm, I'll have to check with my director to see if we can count this class instead of Bio 101. I will give you a call as soon as I have more information.

Your response: _____.

Situation 6: Reschedule a shift at work (Request: student – employer)

Instructions: You are scheduled to work at the campus bookstore tomorrow; however, you just realized that you had scheduled a doctor's appointment on the same day. You want to reschedule your work shift for the day after tomorrow instead. You've asked all your co-workers, but none

of them can switch with you., so you decided to ask the manager if you could still reschedule your shift. What do you say to him?

* Note that the manager will speak first.

Manager: Hey, how can I help you?

Your response: Hi, I was hoping I could
reschedule_____.

Manager: I see. Well, have you asked your co-workers yet? Maybe someone could switch with you.

Your response: _____.

Manager: Well alright, let me look at the schedule and let you know.

Your response: _____.

Appendix F: Listener Perception Scales

(Partially adapted from Taguchi, 2011)

1. Lexico-Grammatical Appropriateness Scale (not included in the analyses)

Score	Description	Definition
5	(Almost) perfectly appropriate	Effective request in the level of directness, politeness, and formality. - The speaker used a bi-clausal request formula (e.g., I was wondering if.../ I was hoping that...) - The speaker gave a clear, extended explanation for the request
4	Adequately appropriate	Mostly effective request in the level of directness, politeness, and formality. - The speaker used a bi-clausal request formula (e.g., I was wondering if.../ I was hoping that...) - The speaker might not have given an explanation for the request
3	Somewhat appropriate	Somewhat effective request in the level of directness, politeness, and formality. - The speaker did not use a bi-clausal formula, but the request still sounded relatively appropriate - The speaker might not have given an explanation for the request
2	Mostly NOT appropriate	The request is mostly NOT effective in the level of directness, politeness, and formality. - The speaker did not use a bi-clausal formula, and the request still sounded mostly NOT appropriate - The speaker did not give an explanation for the request
1	Clearly NOT appropriate	The request sounds either almost rude or too demanding. The speaker did not use a bi-clausal formula, the request was too direct, informal, and impolite

2. Prosodic Appropriateness Scale

Score	Description	Definition
5	(Almost) perfectly appropriate	Effective request in the level of directness, politeness, and formality. • The speaker's prominence was perfectly appropriate (i.e., not every other word was stressed) • The speaker's intonation was perfectly appropriate (i.e., rising tones are not overused, level tones are used only for grounding the request)
4	Adequately appropriate	Mostly effective request in the level of directness, politeness, and formality. • The speaker's prominence was mostly appropriate (i.e., not every other word was stressed) • The speaker's intonation was mostly appropriate (i.e., rising tones were not overused, level tones were used only for grounding the request)
3	Somewhat appropriate	Somewhat effective request in the level of directness, politeness, and formality. • The speaker's prominence was somewhat appropriate (i.e., over-stressing in some places) • The speaker's intonation was somewhat appropriate (i.e., rising tones are somewhat overused, especially in the actual request and/or level tones are used in the actual request)
2	Mostly NOT appropriate	The request is mostly NOT effective in the level of directness, politeness, and formality. • The speaker's prominence was mostly NOT appropriate (i.e., over-stressing in a lot of places) • The speaker's intonation was mostly NOT appropriate (i.e., rising tones are overused, level tones are used in the actual request, failure to use falling tones in the main request)
1	Clearly NOT appropriate	The request sounds either almost rude or too demanding. • The speaker's prominence was NOT appropriate (i.e., over-stressing throughout) • The speaker's intonation was NOT appropriate (i.e., overuse of rising intonation, lack of variation in the use of tone choices, monotonous speech)

3. Accentedness scales

The speaker was

<i>Has a strong accent</i>	1__2__3__4__5	<i>Has no accent</i>
<i>Effortless</i>	1__2__3__4__5	<i>Effortful</i>

4. Comprehensibility scales

The speaker was

<i>Difficult to understand</i>	1__2__3__4__5	<i>Easy to understand</i>
<i>Sounds like a nonnative English speaker</i>	1__2__3__4__5	<i>Sounds like a native English speaker</i>

Appendix G: Rater Training Session

Objectives:

- To familiarize the raters with the constructs they will be measuring and the measurement scales
- To introduce the raters to the speaking tasks they will be rating and sample benchmark responses
- To allow raters to practice using the scale for evaluating learners' production and calibrate their scoring

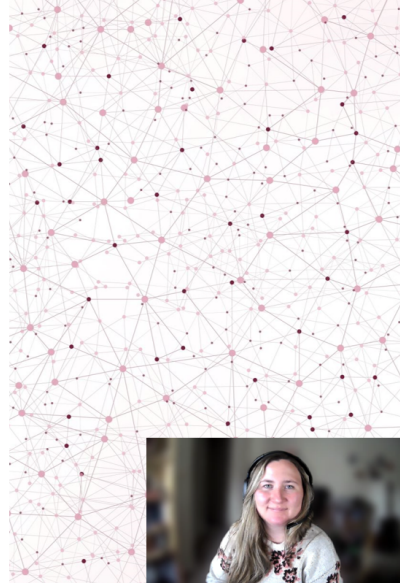
Materials:

- PowerPoint presentation (see below).
- Rating scales (see Appendix F).
- Sound files for the exposure and calibration parts (The samples were taken from the participant recordings collected during the pre- and post-tests. The audios for the Exposure stage were embedded in the PPT; the audios for the Calibration stage were embedded in the Qualtrics survey).

PowerPoint slides

Rater training

Effectiveness of pragma-prosodic instruction for spoken production of requests



Training session outline

1. Introduction of the main constructs
 - a. Pragmatic appropriateness: Politeness, Directness, Formality
 - b. Accentedness & Comprehensibility
2. Demonstration of the role-play tasks
3. Explanation of the rating rubric
4. Exposure to sample speech files



What will you be rating?

- Requests produced by L2 English learners from China.
- Communicative situation: a student making a request to a professor (e.g., extension on a paper, a letter of recommendation), an advisor (e.g., switching classes, transferring credits), or an employer (asking for a day off, asking to reschedule a shift)
 - High imposition (asking for an extension vs. asking for a pen)
 - High relative power (talking to a professor vs. talking to a peer)
 - High social distance (talking to a professor vs. talking to a peer)



Role plays completed by the learners

- Multiple-turn video role plays of high imposition requests (on-campus situations)

- Professor
- Advisor
- Employer

Participants recorded audio responses to those situations (three turns per situation – you will only be rating the first turn i.e., the actual request)



You are being recorded. Click on "Stop recording" when you are done.

Stop Recording

Record the response



Appropriateness

- Learners' ability to produce a target speech act with a proper degree of formality, directness, and politeness.

- Depends on sociolinguistic variables in the context of speech act production
 - Who is the interlocutor (social distance, relative power)?
 - What is being asked for (imposition)?



Appropriateness in requests

- Lexico-grammatical strategies
 - Bi-clausal indirect requests
 - I was wondering if...
 - I was hoping that...
 - Use of grounders
 - Explaining why the request is performed, giving some background to it
- Prosody (prominence & intonation)
 - Falling intonation at the end of the request itself
 - Level intonation is possible, but only when the speaker is enumerating reasons for a request (i.e., in grounders)
 - Rising intonation is not common, especially in the request itself
 - Speech is not monotonous
 - Prominence is not extensive (i.e., only 2-3 content words stressed in a sentence)



Appropriate vs. inappropriate prominence

Prominence = sentence stress

- Speaker stressing only certain content words in a sentence (2-3 per sentence)

L2 speakers oftentimes overuse prominence and stress every other word (or every word)



Appropriate vs. inappropriate intonation

Falling intonation in requests = the speaker is not assuming that the listener will agree

Level intonation = appropriate only when the speaker is enumerating reasons for the request

Rising intonation = appropriate only when the speaker is in the middle of an utterance and has not finished the request yet



Accentedness

- listeners' perception of how strongly L2 speech is different from that of L1 speakers

- Mild accent = speech is slightly different from that of L1 speakers
- Strong accent = speech is extremely different from that of L1 speakers



Comprehensibility

• listeners' perception of how difficult it was to understand the speech and what is being said

- Comprehensible= speech is easy to understand
- Not comprehensible = speech is hard to understand



Scales

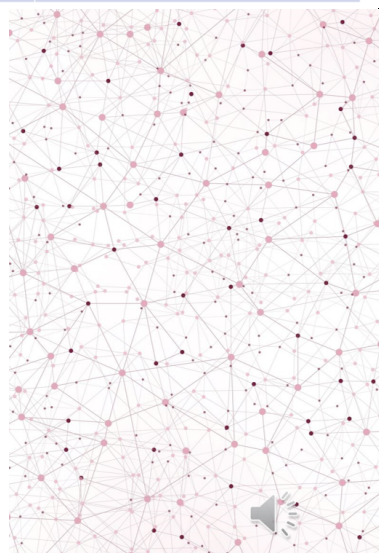
Lexico-grammatical appropriateness

Score	Description	Definition
5	(Almost) perfectly appropriate	Effective request in the level of directness, politeness, and formality. • The speaker used a bi-clausal request formula (e.g., I was wondering if.../ I was hoping that...) • The speaker gave a clear, extended explanation for the request
4	Adequately appropriate	Mostly effective request in the level of directness, politeness, and formality. • The speaker used a bi-clausal request formula (e.g., I was wondering if.../ I was hoping that...) • The speaker gave some explanation for the request
3	Somewhat appropriate	Somewhat effective request in the level of directness, politeness, and formality. • The speaker did not use a bi-clausal formula, but the request still sounded relatively appropriate • The speaker might not have given an explanation for the request
2	Mostly NOT appropriate	The request is mostly NOT effective in the level of directness, politeness, and formality. • The speaker did not use a bi-clausal formula, and the request still sounded mostly NOT appropriate • The speaker did not give an explanation for the request
1	Clearly NOT appropriate	The request sounds either almost rude and too demanding . • The speaker did not use a bi-clausal formula, the request was too direct, informal and impolite

Prosodic appropriateness

Score	Description	Definition
5	(Almost) perfectly appropriate	• The speaker's prominence was perfectly appropriate (i.e., not every other word was stressed) • The speaker's intonation was perfectly appropriate (i.e., raising tones are not overused, level tones are used only for grounding the request)
4	Adequately appropriate	• The speaker's prominence was mostly appropriate (i.e., not every other word was stressed) • The speaker's intonation was mostly appropriate (i.e., raising tones were not overused, level tones were used only for grounding the request)
3	Somewhat appropriate	• The speaker's prominence was somewhat appropriate (i.e., over-stressing in some places) • The speaker's intonation was somewhat appropriate (i.e., raising tones are somewhat overused, especially in the actual request and/or level tones are used in the actual request)
2	Mostly NOT appropriate	• The speaker's prominence was mostly NOT appropriate (i.e., over-stressing in a lot of places) • The speaker's intonation was mostly NOT appropriate (i.e., raising tones are overused, level tones are used in the actual request, failure to use falling tones in the main request)
1	Clearly NOT appropriate	• The speaker's prominence was NOT appropriate (i.e., over-stressing throughout) • The speaker's intonation was NOT appropriate (i.e., overuse of rising intonation, lack of variation in the use of tone choices, monotonous speech)

Sample requests and their ratings



Sample 1



Score	Description	Definition
5	(Almost) perfectly appropriate	Effective request in the level of directness, politeness, and formality. - The speaker used a bi-clausal request formula (e.g., I was wondering if.../ I was hoping that...) - The speaker gave a clear, extended explanation for the request
4	Adequately appropriate	Mostly effective request in the level of directness, politeness, and formality. - The speaker used a bi-clausal request formula (e.g., I was wondering if.../ I was hoping that...) - The speaker gave some explanation for the request
3	Somewhat appropriate	Somewhat effective request in the level of directness, politeness, and formality. - The speaker did not use a bi-clausal formula, but the request still sounded relatively appropriate - The speaker might not have given an explanation for the request
2	Mostly NOT appropriate	The request is mostly NOT effective in the level of directness, politeness, and formality. - The speaker did not use a bi-clausal formula, and the request still sounded mostly NOT appropriate - The speaker did not give an explanation for the request
1	Clearly NOT appropriate	The request sounds either almost rude and too demanding . The speaker did not use a bi-clausal formula, the request was too direct, informal and impolite

Score	Description	Definition
5	(Almost) perfectly appropriate	- The speaker's prominence was perfectly appropriate (i.e., not every other word was stressed) - The speaker's intonation was perfectly appropriate (i.e., raising tones are not overused, level tones are used only for grounding the request)
4	Adequately appropriate	- The speaker's prominence was mostly appropriate (i.e., not every other word was stressed) - The speaker's intonation was mostly appropriate (i.e., raising tones were not overused, level tones were used only for grounding the request)
3	Somewhat appropriate	- The speaker's prominence was somewhat appropriate (i.e., over-stressing in some places) - The speaker's intonation was somewhat appropriate (i.e., raising tones are somewhat overused, especially in the actual request and/or level tones are used in the actual request)
2	Mostly NOT appropriate	- The speaker's prominence was mostly NOT appropriate (i.e., over-stressing in a lot of places) - The speaker's intonation was mostly NOT appropriate (i.e., raising tones are overused, level tones are used in the actual request, failure to use falling tones in the main request)
1	Clearly NOT appropriate	- The speaker's prominence was NOT appropriate (i.e., over-stressing throughout) - The speaker's intonation was NOT appropriate (i.e., overuse of rising intonation, lack of variation in the use of tone choices, monotonous speech)

Sample 2

Lexico-grammatical appropriateness

Score	Description	Definition
5	(Almost) perfectly appropriate	Effective request in the level of directness, politeness, and formality. - The speaker used a bi-clausal request formula (e.g., I was wondering if.../ I was hoping that...) - The speaker gave a clear, extended explanation for the request
4	Adequately appropriate	Mostly effective request in the level of directness, politeness, and formality. - The speaker used a bi-clausal request formula (e.g., I was wondering if.../ I was hoping that...) - The speaker gave some explanation for the request
3	Somewhat appropriate	Somewhat effective request in the level of directness, politeness, and formality. - The speaker did not use a bi-clausal formula, but the request still sounded relatively appropriate - The speaker might not have given an explanation for the request
2	Mostly NOT appropriate	The request is mostly NOT effective in the level of directness, politeness, and formality. - The speaker did not use a bi-clausal formula, and the request still sounded mostly NOT appropriate - The speaker did not give an explanation for the request
1	Clearly NOT appropriate	The request sounds either almost rude and too demanding . The speaker did not use a bi-clausal formula, the request was too direct, informal and impolite

Prosodic appropriateness



Score	Description	Definition
5	(Almost) perfectly appropriate	- The speaker's prominence was perfectly appropriate (i.e., not every other word was stressed) - The speaker's intonation was perfectly appropriate (i.e., raising tones are not overused, level tones are used only for grounding the request)
4	Adequately appropriate	- The speaker's prominence was mostly appropriate (i.e., not every other word was stressed) - The speaker's intonation was mostly appropriate (i.e., raising tones were not overused, level tones were used only for grounding the request)
3	Somewhat appropriate	- The speaker's prominence was somewhat appropriate (i.e., over-stressing in some places) - The speaker's intonation was somewhat appropriate (i.e., raising tones are somewhat overused, especially in the actual request and/or level tones are used in the actual request)
2	Mostly NOT appropriate	- The speaker's prominence was mostly NOT appropriate (i.e., over-stressing in a lot of places) - The speaker's intonation was mostly NOT appropriate (i.e., raising tones are overused, level tones are used in the actual request, failure to use falling tones in the main request)
1	Clearly NOT appropriate	- The speaker's prominence was NOT appropriate (i.e., over-stressing throughout) - The speaker's intonation was NOT appropriate (i.e., overuse of rising intonation, lack of variation in the use of tone choices, monotonous speech)

Samples 3 and 4

Lexico-grammatical appropriateness

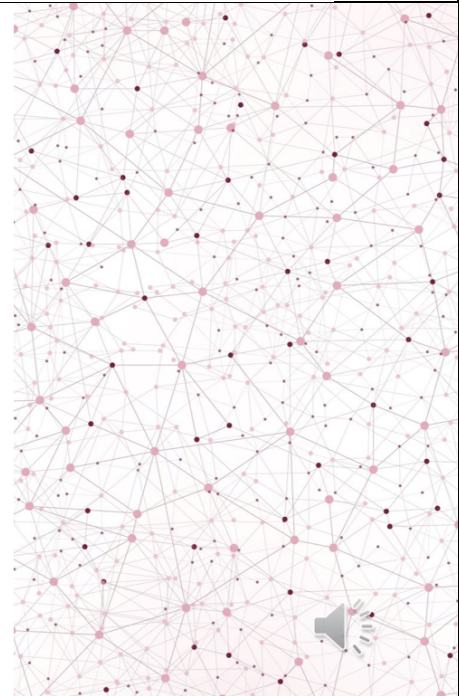
Score	Description	Definition
5	(Almost) perfectly appropriate	Effective request in the level of directness, politeness, and formality. - The speaker used a bi-clausal request formula (e.g., I was wondering if.../I was hoping that...) - The speaker gave a clear, extended explanation for the request
4	Adequately appropriate	Mostly effective request in the level of directness, politeness, and formality. - The speaker used a bi-clausal request formula (e.g., I was wondering if.../I was hoping that...) - The speaker gave some explanation for the request
3	Somewhat appropriate	Somewhat effective request in the level of directness, politeness, and formality. - The speaker did not use a bi-clausal formula, but the request still sounded relatively appropriate - The speaker might not have given an explanation for the request
2	Mostly NOT appropriate	The request is mostly NOT effective in the level of directness, politeness, and formality. - The speaker did not use a bi-clausal formula, and the request still sounded mostly NOT appropriate - The speaker did not give an explanation for the request
1	Clearly NOT appropriate	The request sounds either almost rude and too demanding . The speaker did not use a bi-clausal formula, the request was too direct, informal and impolite

Prosodic appropriateness



Score	Description	Definition
5	(Almost) perfectly appropriate	- The speaker's prominence was perfectly appropriate (i.e., not every other word was stressed) - The speaker's intonation was perfectly appropriate (i.e., raising tones are not overused, level tones are used only for grounding the request)
4	Adequately appropriate	- The speaker's prominence was mostly appropriate (i.e., not every other word was stressed) - The speaker's intonation was mostly appropriate (i.e., raising tones were not overused, level tones were used only for grounding the request)
3	Somewhat appropriate	- The speaker's prominence was somewhat appropriate (i.e., over-stressing in some places) - The speaker's intonation was somewhat appropriate (i.e., raising tones are somewhat overused, especially in the actual request and/or level tones are used in the actual request)
2	Mostly NOT appropriate	- The speaker's prominence was mostly NOT appropriate (i.e., over-stressing in a lot of places) - The speaker's intonation was mostly NOT appropriate (i.e., raising tones are overused, level tones are used in the actual request, failure to use falling tones in the main request)
1	Clearly NOT appropriate	- The speaker's prominence was NOT appropriate (i.e., over-stressing throughout) - The speaker's intonation was NOT appropriate (i.e., overuse of rising intonation, lack of variation in the use of tone choices, monotonous speech)

Thank you!



Appendix H: Working Memory Task Stimuli

(adapted from the reading span task developed by Shin *et al.*, 2019 and Yaw, in preparation)

Level	Order	Sentences	Final word	Semantic judgment
2-1	1	A pencil died in August in a horrible car accident.	accident	NO
	2	She fell in love when she was seventeen years old.	old	YES
2-2	3	After 3 years, a book left the army and studied farming.	farming	NO
	4	When Jenna's father died, he left her a lot of money.	money	YES
3-1	5	She enjoyed swimming and going for picnics on the beach.	beach	YES
	6	Sarah ran home as fast as she could because she was late.	late	YES
	7	The diamonds were a present to the cake from the king.	king	NO
3-2	8	A table went to a summer camp with its sister last year.	year	NO
	9	People from all over the world like the taste of chocolate.	chocolate	YES
	10	She was not interested in singing but was good at dancing.	dancing	YES
4-1	11	It was a beautiful house, but it was also very cold.	cold	YES
	12	John asked his box if it would make his favorite meal.	meal	NO
	13	When Sally came back, Stephen was ready to leave the town.	town	YES
	14	Jack started talking, and the truck listened carefully and nodded often.	often	NO
4-2	15	The three birds saw little men flying in the sky.	sky	NO

	16	Bill's dad was going to drive Kevin to his computer's house.	house	NO
	17	Jenna was surprised when she saw the inside of a watch.	watch	YES
	18	Anna put her hand over her eyes to hide her tears.	tears	YES
5-1	19	People were eating ice cream although it was quite angry outside.	outside	NO
	20	Michael needed some money and wanted to sell his sofa.	sofa	YES
	21	After 5 years, Jack made enough money to support his father.	father	YES
	22	Healthy muscles help the mountain enjoy fun activities like swimming.	swimming	NO
	23	The calendar must go again and catch another fish for his dinner.	dinner	NO
5-2	24	Four men were riding on cell phones along a country road.	road	NO
	25	Nick saw a few grey hairs after a shower that night.	night	NO
	26	A little microwave with wide eyes was holding onto his hand.	hand	NO
	27	The armchair stood by the window waving until Stephen disappeared.	disappeared	NO
	28	Jack saw many paintings on the wall by the chair.	chair	YES
6-1	29	The garden was not big, but it was a wonderful place.	place	YES
	30	Stephen thought the living room sounded a bit bored yesterday.	yesterday	NO
	31	Sam talked to his cat when the door was lonely.	lonely	NO

	32	The sun was burning hot and the air was getting warm.	warm	YES
	33	When the wallet opened, the moon fell upon the ground.	ground	NO
	34	It took hours to get there, but the wall was happy.	happy	NO
6-2	35	Nicole heard Sam talking to buildings on the phone.	phone	NO
	36	Now, the highway is ready to start its own business.	business	NO
	37	They walked down the narrow street to get to the market.	market	YES
	38	I came to have drinks with the swimming pool's parents.	parents	NO
	39	Everyone was very polite, and I thanked them for the invite.	invite	YES
	40	I spoke to Sally, and trees walked to the mall.	mall	NO
7-1	41	Nicole worked very hard but was worried about her exam.	exam	YES
	42	It was only 4 o'clock, but the Chinese restaurant was closed.	closed	YES
	43	The shop was full of green beans in their blue uniforms.	uniforms	NO
	44	They spent 2 hours sitting in the traffic in the morning.	morning	YES
	45	John was surprisingly happy to be studying again after the holidays.	holidays	YES
	46	Eliot sat down and watched the rain running down his window.	window	YES
	47	Michael took his country to his parents' house for two days.	days	NO

7-2	48	Mary tried to call several times, but there was no answer.	answer	YES
	49	Nick soon realized that the field was tired after the run.	run	NO
	50	Nicole walked to the market and bought a lot of vegetables.	vegetables	YES
	51	Dogs tried to build many apartments in a lot of cities.	cities	NO
	52	After we got out, the lake started to go up a hill.	hill	NO
	53	The balloon booked a table for 8 o'clock for two people.	people	NO
	54	It started raining, and there was a huge crash of thunder.	thunder	YES
8-1	55	He sent ships to the hat to trade with foreign lands.	lands	NO
	56	The storm had passed but the sky was still grey.	grey	YES
	57	Milton kindly gave all of his money to the school.	school	YES
	58	The history of music goes back to the beginning of culture.	culture	YES
	59	The police station was crowded with people when the ocean arrived.	arrived	NO
	60	You should be careful with the sun in the hot season.	season	YES
	61	Jack tried to get up, but he could not move his legs.	legs	YES
8-2	62	A woman came in when the stone was still cleaning.	cleaning	NO
	63	The tree stood in the yard, ready for slippers to climb.	climb	NO
	64	I walked along the cow before anyone else was awake.	awake	NO

	65	There was nothing to do except wait, as the snow continued.	continued	YES
	66	Anna suggested going for a walk around the harbor after lunch.	lunch	YES
	67	There were books all over the floor in Bernard's room.	room	YES
	68	John finally opened his closet and pulled out the blue suitcase.	suitcase	YES
	69	A group of young tables was playing basketball at the park.	park	NO
	70	Jack just went into his juice to watch television and rest.	rest	NO

Appendix I: Final Questionnaire
(partially adapted from Ramirez Verdugo 2006)

1. Do you think the platform is useful for practicing your intonation skills?
2. What features of the training did you find useful?
3. What drawbacks did you notice when working with the training?
4. What do you feel you have accomplished in terms of your intonation in English?
5. Do you feel you can apply what you have learned in the training to more natural and spontaneous conversations? Why and how?
6. What do you feel you would need to practice more now?

Appendix J: Instructional Content and Activities for the Control Group

Activity 1. TED Talk “Less Stuff, More Happiness” by Graham Hill [5:33].

Video link: https://www.ted.com/talks/graham_hill_less_stuff_more_happiness?language=en

Writer and designer Graham Hill asks: Can having less stuff, in less room, lead to more happiness? He makes the case for taking up less space and lays out three rules for editing your life.

Instructions: Watch the video (you can watch it up to two times) and answer the questions about the video. There will be six open-ended questions for your to answer.

During the video, take notes about the contents of the talk AND about the speaker’s skills in presenting in front of an audience.

Questions

1. What is the industry Graham is talking about in the beginning of the video?
2. What are the consequences of Americans using more space and having more stuff?
3. What examples of situations does Graham give when people can experience “the joys of less”?
4. What is the name of Graham’s project and what is its goal?
5. What three strategies does Graham name to start “living little”?
6. What, in your opinion, was effective in the way the speaker was communicating with the audience? What was not effective?

Activity 2. TED Talk “Underwater Astonishments” by David Gallo [5:09]

Video link: https://www.ted.com/talks/david_gallo_underwater_astonishments

David Gallo shows jaw-dropping footage of amazing sea creatures, including a color-shifting cuttlefish, a perfectly camouflaged octopus, and a Times Square's worth of neon light displays from fish who live in the blackest depths of the ocean. This short talk celebrates the pioneering work of ocean explorers like Edith Widder and Roger Hanlon.

Instructions: Watch the video (you can watch it up to two times) and answer the questions about the video. There will be six open-ended questions for you to answer.

During the video, take notes about the contents of the talk AND about the speaker's skills in presenting in front of an audience.

Questions

1. How much of the ocean have the humanity explored so far?
2. How have the scientists been able to capture sea creatures that live deep down under the ocean surface?
3. How does an octopus sense its surroundings?
4. What does discoloration on squids mean when they fight?
5. Why can we learn a lot about sea creatures from the shallow water?
6. What, in your opinion, was effective in the way the speaker was communicating with the audience? What was not effective?

Activity 3. TED Talk “Try Something New for Thirty Days” by Matt Cutts [3:27]

Video link:

https://www.ted.com/talks/halla_tomasdottir_a_feminine_response_to_iceland_s_financial_crash?language=en

We all know the arguments that being vegetarian is better for the environment and for the animals -- but in a carnivorous culture, it can be hard to make the change. Graham Hill has a powerful, pragmatic suggestion: Be a weekday veg.

Instructions: Watch the video (you can watch it up to two times) and answer the questions about the video. There will be six open-ended questions for your to answer.

During the video, take notes about the contents of the talk AND about the speaker’s skills in presenting in front of an audience.

Questions

1. What is the industry Graham is talking about in the beginning of the video?
2. What are the consequences of Americans using more space and having more stuff?
3. What examples of situations does Graham give when people can experience “the joys of less”?
4. What is the name of Graham’s project and what is its goal?
5. What three strategies does Graham name to start “living little”?
6. What, in your opinion, was effective in the way the speaker was communicating with the audience? What was not effective?

Activity 4. TED Talk “How We Could Teach Our Bodies to Heal Faster” by Kaitlyn Sadtler [4:48]

Video link: https://www.youtube.com/watch?v=UNP03fDSj1U&t=2s&ab_channel=TED-Ed

What if we could help our bodies heal faster and without scars, like Wolverine in X-Men? TED Fellow Kaitlyn Sadtler is working to make this dream a reality by developing new biomaterials that could change how our immune system responds to injuries. In this quick talk, she shows the different ways these products could help the body regenerate.

Instructions: Watch the video (you can watch it up to two times) and answer the questions about the video. There will be six open-ended questions for you to answer.

During the video, take notes about the contents of the talk AND about the speaker’s skills in presenting in front of an audience.

Questions

1. What is Kathryn’s job and what does it involve?
2. What is the name of the cell in a human body that is critical for wound healing?
3. How does the immune system respond to foreign materials in our body (e.g., insulin pumps)?
4. What does our body accumulate instead of muscle cells if it gets the wrong signal?
5. What are some of the future medical products that the advances in the immune system can help us create in the future?
6. What, in your opinion, was effective in the way the speaker was communicating with the audience? What was not effective?

Activity 5. TED Talk “Why Videos Go Viral” by Kevin Allocca [7:04]

Video link: https://www.ted.com/talks/kevin_allocca_why_videos_go_viral

Kevin Allocca is YouTube's Head of Culture & Trends, and he has deep thoughts about silly web video. In this talk from TEDYouth, he shares the 4 reasons a video goes viral.

Instructions: Watch the video (you can watch it up to two times) and answer the questions about the video. There will be six open-ended questions for your to answer.

During the video, take notes about the contents of the talk AND about the speaker’s skills in presenting in front of an audience.

Questions

1. How many hours of video are uploaded to YouTube every minute?
2. What three things make a video viral?
3. Who helped the Double Rainbow video become famous?
4. How does the community become part of a viral phenomenon?
5. What new culture does Kevin talk about in his talk?
6. What, in your opinion, was effective in the way the speaker was communicating with the audience? What was not effective?

Activity 6. Recording your own monologue.

Instructions:

In this last activity, you will prepare a short 1-minute monologue presentation about the following topic:

“Why Learning English?”

You have watched five short TED talks where five speakers shared their thoughts on a variety of current topics. Think about your favorite speaker of the five. What were the qualities of their talk that you’d like to adopt in your own presentation?

As a reminder, these are the videos you’ve watched [the website will have links to the videos so that participants could rewatch them, if needed]:

1. “Less Stuff, More Happiness” by Graham Hill
2. “Underwater Astonishments” by David Gallo
3. “Try Something New for Thirty Days” by Matt Cutts
4. “How We Could Teach Our Bodies to Heal Faster” by Kaitlyn Sadtler
5. “Why Videos Go Viral” by Kevin Allocca

Once you’ve prepared and rehearsed your presentation and are ready to record, click the button “Record audio” below and begin presenting. [The website will have an in-built voice recorder] When you are done, click “Stop”. Your presentation is now submitted. You will receive an email with the feedback about your presentation from the instructor within a week.

Thank you for your participation!