PRODUCTION AND PERCEPTION OF NATIVE AND NON-NATIVE SPEECH ENHANCEMENTS

by

MISAKI KATO

A DISSERTATION

Presented to the Department of Linguistics
and the Graduate School of the University of Oregon
in partial fulfillment of the requirements
for the degree of
Doctor of Philosophy

September 2020

DISSERTATION APPROVAL PAGE

Student: Misaki Kato

Title: Production and Perception of Native and Non-native Speech Enhancements

This dissertation has been accepted and approved in partial fulfillment of the requirements for the Doctor of Philosophy degree in the Department of Linguistics by:

Melissa M. Baese-Berk	Chairperson
Melissa Redford	Core Member
Tyler Kendall	Core Member
Kaori Idemaru	Institutional Representative

and

Kate Mondloch	Interim Vice Provost and Dean of the Graduate School

Original approval signatures are on file with the University of Oregon Graduate School.

Degree awarded September 2020

DISSERTATION ABSTRACT

Misaki Kato

Doctor of Philosophy

Department of Linguistics

September 2020

Title: Production and Perception of Native and Non-native Speech Enhancements

One important factor that contributes to successful speech communication is an individual's ability to speak more clearly when their listeners do not understand their speech. Though native talkers are able to implement various acoustic-phonetic speech enhancements to make their speech more understandable to their listeners (e.g., by speaking more slowly, loudly, or by articulating sounds more clearly), such goal-oriented adaptations employed by non-native talkers are much less well-understood. This dissertation investigates how talkers' ability to implement speech enhancements is shaped by their target language experience and how these enhancements impact listeners' perception. Specifically, we examine acoustic characteristics of speech enhancements produced by native English talkers and non-native English talkers of higher- and lower-proficiency in different contexts: in a reading task where talkers are explicitly asked to read materials clearly, as well as in a simulated communication task where listeners' communicative needs for enhanced intelligibility are signaled implicitly in the context. We further examine perceptual consequences of speech enhancements in terms of intelligibility (whether listeners understand the speech) and other subjective evaluations of the speech, including perceived degree of comprehensibility (how easy the listeners perceive the speech is to understand).

The results show that native talkers and higher-proficiency non-native talkers generally make larger acoustic modifications than lower-proficiency talkers. However, such effects of talkers' target language experience differ depending on the type of acoustic manipulations involved in the productions. Furthermore, an improvement in intelligibility does not necessarily correspond to an improvement in other subjective evaluations of the speech, suggesting that perceptual benefits resulting from speech enhancements could vary depending on how listeners are asked to evaluate the speech.

The results of this dissertation highlight that talkers have the flexibility to accommodate listeners' communicative needs in a native and non-native language, and suggest that this flexibility is shaped by the combination of talkers' linguistic backgrounds and the focus of adaptation. Furthermore, the current work provides evidence that perceptual consequences of speech enhancements are multi-faceted, and suggest that acoustic features of speech enhancements responsible for an improvement in intelligibility may differ from those influence other types of subjective evaluations.

CURRICULUM VITAE

NAME OF AUTHOR: Misaki Kato

GRADUATE AND UNDERGRADUATE SCHOOLS ATTENDED:

University of Oregon, Eugene
Meiji University, Tokyo, Japan

DEGREES AWARDED:

Doctor of Philosophy, Linguistics, 2020, University of Oregon
Master of Arts, Language Teaching Studies, 2014, University of Oregon
Bachelor of Arts, Global Japanese Studies, 2013, Meiji University

AREAS OF SPECIAL INTEREST:

Speech Perception
Speech Production
Phonetics
Second Language Acquisition

PROFESSIONAL EXPERIENCE:

Graduate Employee (researcher), Center for Applied Second Language Studies, University of Oregon, 2017-2019

Graduate Employee (instructor), American English Institute, University of Oregon, 2016-2017

Graduate Employee (researcher), Department of Linguistics, University of Oregon, 2014-2016

Graduate Employee (teaching assistant), East Asian Languages and Literatures, University of Oregon, 2012-2014

GRANTS, AWARDS, AND HONORS:

National Science Foundation (BCS-1941739). Doctoral Dissertation Research Improvement Grant, "Production and perception of native and non-native speech enhancement", 2020-2021.

Lokey Doctoral Science Fellowship, University of Oregon, 2019-2020.

General University Scholarship, University of Oregon, 2019-2020.

National Federation of Modern Language Teachers Associations Dissertation Support Grant, 2019

Institute of Cognitive and Decision Sciences Dissertation Research Award, University of Oregon, 2019

PUBLICATIONS:

Kato, M., & Baese-Berk, M. The effects of acoustic and semantic enhancements on perception of native and non-native speech. Manuscript submitted for publication.

Idemaru, K., Kato, M., & Tsukada, K. (in press). Foreign accent in L2 Japanese: Cross-sectional study. In R. Wayland (Ed.), *Second Language Speech Learning.* Cambridge: Cambridge University Press.

Kato, M., Kawahara, S., & Idemaru, K. (2020). Speaking rate normalization across different talkers in the perception of Japanese stop and vowel length contrasts. In *Proceedings of the 10th International Conference on Speech Prosody* (pp. 61-65).

Kato, M., & Baese-Berk, M. (2020). The effect of input prompts on the relationship between perception and production of non-native sounds. *Journal of Phonetics*, 79, 100964.

Kato, M., Idemaru, K., & Tsukada, K. (2019). Acoustic correlates of foreign accent in L2 Japanese: A cross-sectional study. In S. Calhoun, P. Escudero, M. Tabain, & P. Warren (Eds.), *Proceedings of the 19th International Congress of Phonetic Sciences* (pp. 3250 – 3254). Melbourne, Australia.

Yerian, K., Mikhaylova, A., Pashby, P., & Kato, M. (2018). Native and non-native teacher candidate perceptions of professional language development in an MA TESOL program. *TESOL Quarterly,* 53(2), 552-565.

Kato, M. (2015). Developing self-evaluation skills through giving peer writing feedback. *ORTESOL Journal, 32*, 70-72.

ACKNOWLEDGMENTS

Thank you, first, to my advisor and mentor Melissa Baese-Berk, for taking me on as your student, and for providing constant support, encouragement, inspiration, and patience throughout this program. You have encouraged me to explore different ways of asking interesting questions, pushed me to think carefully and deeply to build strong arguments, and shown me how to be persistent. Your confidence in me has meant so much to me, and enabled me to keep going during tough times. Thank you for everything.

I have had a pleasure of working closely with a member of my committee, Kaori Idemaru, on several speech perception projects. I am truly grateful for the insightful discussions and the new opportunities that she has introduced me to. Thank you, also, to my other committee members, Lisa Redford and Tyler Kendall, for providing me with important new perspectives and asking me hard questions to refine my arguments.

I would like to express my appreciation to the participants in this dissertation research, and to the instructors who have helped me recruit their students at the American English Institute. Without them, none of this would have been possible. I would also like to thank the funding agencies and grants that have supported this research, including National Science Foundation DDRIG (BCS-1941739) and Lokey Doctoral Science Fellowship.

I wish to thank the brilliant and supportive members of the Speech Perception and Production Lab. Special thanks to Dae-yong Lee, Jonathan Wright, Ellen Gillooly-Kress, and Zack Jaggers, for their helpful input and for the research meetings that have kept me grounded over the years. I also thank so many undergraduate students in the lab for their willingness and capability to jump in and help at various stages of the research projects, especially Aubrey, Brandon, Cydnie, Kayla, Sarah, Tillie, and Zach.

I would have never survived this PhD program without the wonderful graduate students, past and present, in the UO Linguistics department. I especially thank Marie-Caroline Pons for always being there to listen and cheer me up; and Allison Taylor-Adams and Kaylynn Gunter for all the laughs and hardships that we have shared in our office 270. Thank you to Paul Olejarczuk and Julia Trippe for showing me how to do research in my first few years. I am grateful for all the fun and insightful conversations we have had over the years, especially with Jeff Kallay, Amos Teo, Charlie Farrington, Jason McLarty, Jaeci Hall, Zara Harmon, Hideko Teruya, Matt Stave, Amy Smolek, Manuel Otero, Becky Paterson, Zoe Tribur, Shahar Shirtz, and so many more. I would like to also thank my friends, especially Shannon Ball, Hayley Brazier, Russell Moon, Karessa Torgerson, Dustin Crawford, and Shelley Guidrey, for showing me there is life outside grad school.

I feel extremely lucky to have such a supportive family. My in-law parents, brothers, sisters, and nieces have been so supportive and wonderful over the years. Special thanks to my parents Eiichi and Kyoko Kato, my brother Rinichi Kato, my grandparents Satomi and Satoshi Kato, and Koji and Mie Ishikawa, for spoiling me with good food, adventures, and hot springs whenever I go back to Japan, and for always thinking of me and supporting me from a far. Thanks to Casey for keeping me on a strict exercise schedule and for constantly showing me that life is good.

Finally, my most sincere appreciation goes to Isaac Gaines, for being by my side this whole time, for being patient and supportive, and for reminding me that I could do this. This dissertation is as much yours as it is mine.

To my family, for always believing in me.

TABLE OF CONTENTS

Chapter Page

LIST OF FIGURES

Figure Page

LIST OF TABLES

Table Page

Table Page

Table Page

Table Page

CHAPTER I: INTRODUCTION

One important factor that contributes to successful speech communication is an individual's ability to speak more clearly when their listeners do not understand their speech. It has been widely demonstrated that talkers are able to enhance various features of their speech (e.g., by speaking more slowly, loudly, or by articulating sounds more clearly) to make their speech more understandable to their listeners (e.g., Bradlow, Kraus, & Hayes, 2003; Ferguson & Kewley-Port, 2007; Krause & Braida, 2004). However, despite the wealth of information on the speech enhancement strategies that people use in a native language, much less is known about the strategies that are employed in a non-native language. Particularly, our understanding of non-native speech enhancements is limited to those employed by highly proficient talkers (e.g., Smiljanić & Bradlow, 2011), leaving it unclear how speech enhancement strategies could be implemented by non-native talkers who do not produce lengthy, fluent speech with complex structure. The significance of this issue is highlighted by the growing population of non-native English talkers; according to a 2018 United States Census Bureau report (U.S. Census Bureau, 2018), 21.5% of respondents spoke a language other than English at home, and close to 40% of that population stated that they speak English less than very well. Understanding how people with limited language proficiency could enhance speech intelligibility, as well as how this skill may improve as talkers' language proficiency develops, is relevant not only to extending theories of speech production and adaptation to unfamiliar speech, but also to facilitating evidence-based approaches to developing effective assessment and training methods for second language teaching. In order to better understand the dynamics of

speech communication among people with diverse language backgrounds, this dissertation explores how native and non-native English talkers of various English proficiency accommodate listeners' difficulty understanding their speech, and how these talkers' accommodation strategies are perceived by listeners.

In this introductory chapter, we provide an overview of the relevant literature and of the research questions investigated in the dissertation. Specifically, we discuss the previous work examining talkers' speech enhancements aimed to enhance intelligibility for their listeners, as well as the types of acoustic-phonetic modifications that are typically examined in these studies (Section 1.1). Further, we review studies on the possible factors impacting the speech enhancement behavior: talkers' target language experience (i.e., speech enhancements in a non-native language; Section 1.2), and the type of tasks used to elicit talkers' speech enhancements (Section 1.3). Finally, we discuss the impact of speech enhancements in a variety of perceptual aspects including how well listeners understand the speech (i.e., intelligibility; Section 1.4).

1.1. Talkers' speech enhancements

Speech production is a flexible process; talkers modify what they say and how they say it based on a variety of factors. For example, talkers change their speaking patterns, in terms of syntactic structure, lexical items or phrases, as well as acoustic details of pronunciation, to converge to or diverge from those of their conversation partners (e.g., Babel, 2012; Branigan, Pickering, & Cleland, 2000; Brennan & Clark, 1996; Kim, Horton, & Bradlow, 2011). Talkers modify characteristics of their speech not only depending on what their partners say, but also based on what their partners understand or not understand.

Specifically, being able to improve speech intelligibility based on listeners' difficulty understanding the speech is a crucial part of successful message delivery. This goal-oriented mode of speech production has been discussed as variation situated in a dynamic balance between talker- and listener-oriented forces (Hyper-Hypo, or H&H theory; Lindblom, 1990). In this theoretical framework, the talker-oriented force is to minimize articulatory effort, which originates from a biomechanical principle that the speech motor system, like other physical non-speech movements (e.g., movement of an arm), is constrained by the economy of effort. This stands in the opposite end from the listener-oriented requirement for sufficient perceptual discriminability of sounds. Given these two opposing forces, the talker is able to adjust their output depending on the communication context. That is, if the talker interprets that the communication context places extra demand on the listener, decreasing their chance of successfully understanding the message, the talker increases articulatory effort (hyperarticulate or hyperspeech) to make their speech easier to understand for the listener; whereas if the context is favorable for ease of communication, the talker tries to minimize their articulatory effort (hypoarticulate or hypospeech). Thus, variations in phonetic characteristics of speech can be shaped by talkers' goal to minimize effort as well as to ensure ease of perception for listeners, which can change in different listening environments.

One manifestation of the adjustments that talkers make in communicatively challenging contexts is clear speech. Clear speech is a speaking style that talkers adopt when they are aware that the listeners may have difficulty understanding them, possibly because the listeners are hearing-impaired or are non-native listeners of the language (Smiljanić & Bradlow, 2009; Uchanski, 2005). Talkers' clear speech has typically been

examined by having them read the same set of materials twice: once in a plain-speaking style and once in a clear-speaking style (Bradlow & Alexander, 2007; Ferguson & Kewley-Port, 2002; Ferguson 2004; Granlund, Hazan, & Baker, 2012; Picheny, Durlach, & Braida, 1986; Rogers, DeMasi, & Krause, 2010; Schum, 1996; Smiljanić & Bradlow, 2005, 2011). For the plain-speaking style, talkers are instructed to read the materials as if they are talking to someone familiar with their voice and speech patterns; for the clear-speaking style, talkers are instructed to read the same materials as if they are talking to a listener with a hearing loss or to a non-native listener of the language (e.g., Bradlow & Alexander, 2007; Smiljanić & Bradlow, 2011). Here, when referring to the baseline speaking style to compare with clear speech, we use "plain speech" or "plain-speaking style" (Bradlow & Alexander, 2007), instead of "conversational speech" (e.g., Bradlow & Bent, 2002; Ferguson & Kewley-Port, 2002; Picheny et al., 1986; Smiljanić & Bradlow, 2005), in order to better reflect on the speech elicitation condition where talkers read materials in a laboratory setting.

In order to enhance intelligibility of speech for listeners, talkers make a variety of acoustic-phonetic adjustments. For example, in order to increase the overall salience of the speech signal (Bradlow & Bent, 2002), native English talkers speak with higher fundamental frequency (F0), wider F0 range, increased intensity, as well as increased energy in the 1000-3000 Hz range of long-term spectra (e.g., Bradlow et al., 2003; Liu, Del Rio, Bradlow, & Zeng, 2004; Picheny et al., 1986; Smiljanić & Bradlow, 2005). Furthermore, talkers slow their speech by lengthening segments, as well as by inserting more frequent pauses (e.g., Bradlow et al., 2003; Ferguson & Kewley-Port, 2002; Krause & Braida, 2004 Picheny et al., 1986; Smiljanić & Bradlow, 2005). For example, when

reading short sentences and paragraphs in English, native English talkers produced clear speech with longer consonant and vowel durations as well as with more pauses compared to plain speech (Smiljanić & Bradlow, 2008b). The talkers lengthened consonant and vowel intervals to a similar extent, suggesting that relationship between consonant and vowel intervals remained stable across plain and clear speaking styles. Smiljanić and Bradlow (2008b) further suggested that the stable temporal properties of segments across the two speaking styles, along with the increased number of phrasing in clear speech due to increased pauses, could contribute to increased intelligibility.

Talkers also make segmentally-focused modifications in clear speech. For example, native English talkers release word-final consonants more frequently in clear speech compared to plain speech (Bradlow et al., 2003; Krause & Braida, 2004; Picheny et al., 1986). Further, segmentally-focused modifications are often aimed at making the phonological categories of the language more distinct from one another (Lindblom, 1990); for example, talkers increase the voice-onset-timing (VOT) of word-initial voiceless stop consonants, enhancing the difference between voiced and voiceless stops (Krause & Braida, 2004; Smiljanić & Bradlow, 2008a). However, Smiljanić and Bradlow (2008a) also demonstrated that, even though the absolute duration of the VOT of word-initial voiceless stops was longer in clear speech than in plain speech, the proportional measures (i.e., aspiration duration relative to the closure + aspiration duration) did not differ between the two speaking styles. This suggests that the proportional relationship between voiced and voiceless stops was stable across the two speaking styles, and also that speech enhancements could manifest differently for absolute vs. proportional measures of segment durations.

Segmentally-focused modifications are also widely observed for vowels. For example, previous studies unanimously show vowel space expansion in English clear speech as compared to plain speech (e.g., Bradlow, 2002; Bradlow et al., 2003; Johnson, Flemming, & Wright, 1993; Krause & Braida, 2004; Moon & Lindblom, 1994; Smiljanić & Bradlow, 2005). The vowel space expansion is characterized by several features, including increases in vowel space area, and increases in the extent of vowel space dispersion and peripheralization (Ferguson & Kewley-Port, 2002; Smiljanić & Bradlow, 2005). That is, compared to plain speech, vowels in clear speech cover a larger area (defined by the Euclidian area covered by the means of vowel categories) and are more peripheral from the central point of a talker's vowel space. For example, F2 of vowels changes in different directions for front and back vowels; F2 increases for front vowels but decreases for back vowels in clear speech, enhancing the spectral distinction between these vowels (Ferguson & Kewley-Port, 2002).

Characteristics of vowels in clear speech differ from those in plain speech in terms of temporal features as well; studies consistently show that vowels in clear speech have longer durations than those in plain speech (e.g., Ferguson & Kewley-Port, 2002, 2007; Lam, Tjaden, & Wilding, 2012; Moon & Lindblom, 1994; Picheny et al., 1986; Smiljanić & Bradlow, 2008a). Vowel lengthening patterns could differ depending on the phonological structure of the language, as Uchanski (1988) found that English tense vowels were lengthened more than lax vowels, increasing the duration contrast between those vowels in clear speech. However, other studies also suggest that lengthening of vowels does not necessarily change the proportional relations between vowel durations and the surrounding speech across the two speaking styles. For example, though the duration of the

English diphthong /ai/ increased in clear speech, the vowel-to-word proportion remained stable across plain and clear speech (Tasko & Greilick, 2010). Similarly, Smiljanić and Bradlow (2008a) showed that the proportional duration distance between English tense and lax vowels did not differ in plain and clear speech, suggesting relational invariance for vowel duration to maintain length contrast across different speaking styles. Though the findings regarding temporal manipulations of tense and lax vowels are mixed, Leung and colleagues (2016) demonstrated that talkers use different strategies to enhance tense vs. lax vowels. That is, to enhance vowels in clear speech, talkers use duration changes to a greater extent for tense vowels than for lax vowels; whereas talkers use spectral changes to a greater extent for lax vowels than for tense vowels. The researchers suggested that talkers utilize the temporal dimension to enhance tense vowels because the degree of the spectral variation is more limited for tense vowels than for lax vowels. Together, these studies demonstrate that in clear speech, talkers manipulate different acoustic features in order to enhance global features of the speech signal as well as to enhance segmental characteristics of individual sounds.

Though a majority of work examining clear speech enhancements has focused on English speech, studies investigating cross-linguistic clear speech patterns provide insight into the generality and specificity of different enhancement features. That is, cross-language studies suggest that talkers' strategies to enhance overall salience of the speech signal may be rather independent of the specific language (i.e., talkers may use similar global enhancement strategies in different languages); while their strategies to enhance characteristics of individual sounds may be rather specific to the sound system of the language (i.e., talkers may use different segmental enhancement strategies in different

languages). For example, Granlund et al., (2012) compared Finnish-English bilinguals' global speech enhancement strategies in their two languages and demonstrated that the talkers' used similar strategies when communicating to a partner in a situation with a communication barrier (i.e., vocoded speech signal) compared to a situation without a barrier. Specifically, speech produced in a communicatively challenging condition had higher F0 and higher intensity compared to the speech produced in an easy-listening condition, and talkers made these modifications to a similar extent in both languages. Similarly, Smiljanić and Bradlow (2005) compared native English talkers' global enhancement strategies in English and native Croatian talkers' global enhancement strategies in Croatian, and found that talkers of both languages slowed down their speech and spoke with increased F0 range in clear speech compared to plain speech. Furthermore, talkers of both languages expanded the vowel space in clear speech to a similar extent, despite the difference in vowel inventories between the two languages (English has 10+ vowels though Croatian has 5 vowels). Similar results have been reported, where the extent of vowel space expansion was similar for English clear speech and Spanish clear speech, despite that English has larger vowel inventories than Spanish (Bradlow, 2002). These studies suggest that talkers of different languages use similar strategies to enhance overall salience of the speech signal, and their effort to hyperarticulate globally may be implemented regardless of language-specific phonological inventories (e.g., vowel inventories).

Unlike those global strategies, talkers' use of acoustic cues to enhance individual sounds differs depending on the specific language. For example, Smiljanić and Bradlow (2008a) investigated native English and native Croatian talkers' use of duration to enhance

vowel and consonant contrasts in clear speech. The two languages differ in how duration is used when distinguishing vowels and consonants. Specifically, though Croatian distinguishes vowels in duration (e.g., short vs. long vowels), English does not use duration as a primary cue; English tense-lax vowels are distinguished primarily via spectral differences. This was reflected in the enhancement patterns of vowel contrasts in two languages; the extent that native Croatian talkers increased Croatian long vs. short vowel duration difference was larger than the extent that native English talkers increased tense vs. lax vowel duration difference. Further, these talkers' use of duration also differed in consonant contrasts. That is, in order to enhance the difference of a voicing contrast (pre-voiced vs. short lag stops in Croatian and short vs. long lag stops in English), native Croatian talkers lengthened voicing portions of pre-voiced stops, though native English talkers lengthened aspiration of long-lag stops. Thus, native Croatian and native English talkers used duration differently to enhance vowel and consonant contrasts.

Similarly, Granlund et al. (2012) reported that Finnish-English bilinguals manipulated VOTs of initial stop consonants from plain to clear speech differently for two short-lag stops, Finnish /p/ and English /b/. That is, in clear speech, these talkers decreased VOT for English /b/ to a greater extent than for Finish /p/, possibly because English has a voicing counterpart /p/ though Finnish does not. The talkers also increased VOT for English long-lag /p/ in clear speech. Together, these cross-language studies suggest that talkers of different languages use similar acoustic modification strategies to enhance overall salience of the speech signal (global strategies), whereas their use of particular acoustic cues may differ for enhancing individual sounds in different languages (segmental strategies). Further, these studies possibly suggest that making appropriate enhancements at

a segmental level takes extensive experience with the sound structure of the language, thus could be difficult to implement for people who are not familiar with the phonological system of the language.

Taken together, these studies demonstrate that talkers implement various acoustic-phonetic modifications to enhance intelligibility of their speech for listeners who experience perceptual difficulty. Further, cross-language studies have provided unique insight into how talkers' speech enhancement strategies may or may not be influenced by the sound system of the specific language. Though acoustic characteristics of clear speech enhancements could share features with other goal-oriented modes of speech adjustments, such as Lombard speech and infant-directed speech (e.g., Junqua, 1993; Kuhl et al., 1997; Skowronski & Harris, 2006; Summers et al., 1988), clear speech is specifically aimed at enhancing intelligibility for adult listeners with perceptual difficulties, and does not necessarily involve features such as increased affective prosody for attracting children's attention or increased vocal effort for overcoming speaking in noise (Smiljanić & Bradlow, 2009; Uther, Knoll, & Burnham, 2007). Thus, acoustic modifications involved in talkers' clear speech enhancements have a unique purpose of accommodating listeners' communicative needs.

1.2. Non-native talkers' speech enhancements

Though clear speech enhancements have been a subject of considerable research over the past several decades, they have been described largely based on the speech produced by native talkers of the language. This could pose a limitation on the applicability of the theoretical framework used to explain clear speech enhancements, as the nature of

the dynamic balance between talker- and listener-oriented forces on within-talker phonetic variation (Lindblom, 1990) could look different when producing speech in a non-native language vs. in a native language. That is, producing speech in a non-native language can pose a unique challenge to a talker, possibly impacting how they try to minimize articulatory effort (talker-oriented force) as well as how they pay attention to listeners' communicative needs (listener-oriented force). For example, previous literature on bilingual speech production suggests that producing speech in a second language (L2) poses increased processing demands for talkers, thus is more effortful compared to speaking in the first language (L1; e.g., Green, 1998; Hanulová, Davidson, & Indefrey, 2011; Runnqvist, Strijkers, Sada, & Costa, 2011). Particularly, bilingual talkers are slower and less accurate when naming pictures in their L2 compared to native talkers of the language, suggesting that mapping from meaning to phonological forms is less robust in L2 than in L1 (e.g., Gollan, Montoya, Fennema-Notestine, & Morris, 2005; Ivanova & Costa, 2008; Kroll & Stewart, 1994; Roberts, Garcia, Desrochers, & Hernandez, 2002). When producing speech in L2, talkers also have to cope with L1 interference, encoding phonological representations of their less dominant L2 while suppressing phonological representations of their more dominant L1 (e.g., Roelofs & Verhoef, 2006).

Though producing speech in L2 can be generally more effortful than L1, previous studies have also shown that L2 production improves and becomes less effortful as talkers' L2 proficiency develops (e.g., Declerck & Kormos, 2012; Kormos, 2000; Kormos & Dénes, 2004; Nip & Blumenfeld, 2015; Pivneva, Palmer, & Titone, 2012). For example, in picture-naming tasks, highly proficient bilinguals are faster to name L2 words and to translate words from one language to the other language, as well as are able to switch

between languages more efficiently compared to learners of lower L2 proficiency (Costa & Santesteban, 2004; Kroll, Michael, Tokowicz, & Dufour, 2002). When being imposed with dual-task demands (i.e., completing a picture-description task while simultaneously asked to do a finger-tapping task), learners of higher L2 competence produce L2 speech faster and make fewer errors compared to learners of lower L2 competence (Declerck & Kormos, 2012). Further, in spontaneous speech, higher-proficiency learners repair the informational content of the message more frequently than lower-proficiency learners do, indicating that higher-proficiency talkers' L2 encoding at the lexical, grammatical, and phonological levels are more automatized than that of lower-proficiency learners, enabling higher-proficiency learners to monitor the message conceptualization phase of their speech production (Kormos, 2000). These results suggest that increased L2 proficiency is associated with less effort and less cognitive resources required for L2 production, in terms of inhibiting L1 representations, retrieving weaker L2 representations, as well as producing L2 with appropriate phonological specifications (e.g., Green, 1998; Poulisse, 1997; Roelofs & Verhoef, 2006). Such effect of L2 proficiency on L2 production can be observed in fluency characteristics of the speech, where higher-proficiency talkers' speech is produced with increased speed, longer utterance durations (with a greater number of words produced), and shorter and less frequent pauses, compared to lower-proficiency talkers' speech (Kormos & Dénes, 2004; Poulisse, 1997). Furthermore, in terms of speech motor control, higher-proficiency talkers' productions involve less speech movement variability, faster speed, as well as greater ranges of movements compared to lower-proficiency talkers' productions when reading L2 sentences (Nip & Blumenfeld, 2015). These characteristics of increased speech motor control in higher-proficiency talkers' L2 speech

can be associated with greater phonetic specifications (Lindblom, 1990); whereas lower-proficiency talkers' reduced L2 speech motor control may reflect a developing phase of L2 phonological rules and/or their reliance on L1 speech motor planning strategies when producing L2 (Flege, Schirru, & MacKay, 2003).

Therefore, these studies have demonstrated that L2 production can be more effortful than L1 production, requiring cognitively demanding tasks including coping with interference from a more dominant L1, as well as accessing and using grammatical, lexical and phonological systems of a weaker L2. The difficulty associated with producing speech in L2 can be alleviated with increased L2 proficiency, both at the levels of the message formulation (syntactic, grammatical, and lexical levels) as well as at the level of speech motor control. However, how such increased demand associated with L2 production impacts talkers' strategies to increase intelligibility of L2 speech is much less well documented. That is, compared to the wealth of information on the effortful and cognitively demanding nature of L2 production in general, we understand much less about how talkers manipulate features of their L2 production in order to further make their speech more understandable for their listeners. Particularly, within the framework of H&H theory (Lindblom, 1990) suggesting that talkers adjust their articulatory effort based on their listeners' communicative needs, the nature of the balance between the talker-oriented force to minimize the articulatory effort and the listener-oriented force to ensure sufficient intelligibility for listeners may be different when producing speech in a non-native language than in a native language. Thus, examining how goal-oriented speech adaptations are implemented by talkers who are under increased constraints and cognitive demands (compared to L1 productions) may help us better understand how talker-oriented vs.

listener-oriented forces together impact within-talker phonetic variations.

An important question regarding non-native speech adaptation is whether non-native talkers are able to adjust their productions based on their listeners' communicative needs for speech intelligibility, manipulating acoustic properties in a language that is already more difficult to produce compared to their native language. Though there is limited data regarding clear speech enhancement strategies employed by non-native talkers of the language, several studies suggest that clear speech enhancements made by highly proficient non-native talkers are as effective as those made by native talkers. This is shown in the comparable size of intelligibility gains resulting from clear speech enhancements produced by native talkers and by highly proficient learners (Smiljanić & Bradlow, 2011), and by early learners of the language (Rogers et al., 2010). For example, in Smiljanić and Bradlow (2011), highly proficient non-native English (native Croatian) talkers read semantically anomalous sentences (e.g., *"Your tedious beacon lifted our cab"*) in plain- and clear-speaking styles, and the clear speech resulted in a significant intelligibility improvement as compared to plain speech for native English listeners. Further, Granlund et al. (2012) demonstrated that the types of clear speech modifications made by native English talkers and by proficient non-native English (native Finnish) talkers were similar. Specifically, when examining these talkers' spontaneous speech produced in a problem-solving task with a partner, the clear speech elicited by placing a communication barrier (vocoded speech) had higher F0, higher intensity and longer word durations, compared to the speech produced in the context where there was no communication barrier. The extent of these clear speech modifications was similar for native English talkers' and proficient Finnish-English bilinguals' speech. Bradlow (2002) also demonstrated that the extent of

vowel space expansion is similar between the clear speech productions of native English talkers and early Spanish-English bilinguals. Thus, these studies have suggested that highly proficient non-native talkers use similar clear speech strategies as native talkers of the language, and the proficient non-native talkers' clear speech enhancements result in significant intelligibility gains for native listeners.

However, there is little data regarding how relatively inexperienced talkers of the language (e.g., non-native talkers of lower-proficiency) try to make listener-oriented acoustic-phonetic modifications. One study has demonstrated that late learners of English were much less effective at enhancing intelligibility of English vowels than early learners and native English talkers (Rogers et al., 2010). That is, clear speech enhancements of English vowels in /bVd/ syllables produced by monolingual native English talkers and early native Spanish learners of English resulted in a similar size of intelligibility gains, whereas those produced by late native Spanish learners resulted in much smaller intelligibility gains. The late learners' clear speech enhancements sometimes resulted in a decrease in intelligibility for native English listeners; though it is unclear how the late learners' clear speech enhancements differed acoustically from those of early learners and native talkers, because the acoustic analysis of these talkers' productions was not shown in the study. That is, here (and in other studies that report intelligibility measures of speech enhancements), intelligibility gains are used as a metric for acoustic modifications made by talkers. Thus, it may be possible that there are differences between acoustic modifications and their perceptual consequences (e.g., acoustic modifications of vowel durations may not necessarily result in an intelligibility improvement). Further, it is difficult to determine whether late learners' clear speech enhancement strategies differed from those of more

experienced talkers at a more global level than characteristics of the vowels themselves (e.g., changes in speaking rate, F0, and intensity). Thus, in order to better understand how talkers' target language proficiency impacts their ability to enhance intelligibility of their L2 speech, it is critical to examine acoustic characteristics of enhancements of various materials (e.g., words, phrases, sentences), produced by non-native talkers of differing proficiency, and their perceptual consequences.

Furthermore, examining the speech produced in different speaking styles (e.g., plain and clear speech) by non-native talkers of different proficiency levels may help us better understand the relationship between talkers' ability to produce intelligible speech in general vs. their ability to *increase* intelligibility of their speech. For example, given that producing speech in a non-native language becomes more fluent and less effortful as the talkers' proficiency develops (e.g., Kormos & Dénes, 2004; Nip & Blumenfeld, 2015), it is possible that higher-proficiency non-native talkers' speech is generally more intelligible than lower-proficiency talkers' speech. However, the ability to further *increase* intelligibility by manipulating acoustic-phonetic properties of their speech may or may not differ between non-native talkers of differing proficiency levels. That is, if higher-proficiency talkers are able to make acoustic modifications (from plain to clear speech) and increase intelligibility of their speech to a larger extent than lower-proficiency talkers, this would suggest that non-native talkers' increased proficiency is associated with their ability to not only produce generally more intelligible speech but also with their ability to further *increase* intelligibility of their speech. However, if the extent of acoustic modifications is similar between talkers of different proficiency levels, it may suggest that the ability to produce generally intelligible speech and the ability to increase intelligibility are at least

partially independent from one another. Thus, examining patterns of speech enhancements produced by native talkers and non-native talkers of different proficiency levels may help us better understand how talkers of different linguistic backgrounds implement goal-oriented phonetic variations, and this could informative for developing models of second language speech production.

1.3. Speech enhancements in different tasks

Though one way to examine talkers' speech enhancement behavior is to compare their productions between plain- and clear-speaking styles (e.g., Ferguson & Kewley-Port, 2002; Ferguson 2004; Granlund et al., 2012; Picheny et al., 1986), other studies suggest that speech enhancements are not uniform phenomena (e.g., Gilbert, Chandrasekaran, & Smiljanić, 2014; Hazan & Baker, 2011; Scarborough & Zellou, 2013; Tuomainen & Hazan, 2018). That is, talkers' efforts to enhance acoustic-phonetic characteristics of speech can be implemented differently depending on the types of task that they engage in to produce speech enhancements. Particularly, studies examining native talkers' speech enhancements in different contexts suggest that there are differences in acoustic characteristics of speech enhancements produced in read speech with explicit instructions to speak clearly vs. in spontaneous speech during a conversation (e.g., Hazan & Baker, 2011; Scarborough & Zellou, 2013). For example, native talkers' speech enhancements elicited in read speech, using the instruction to speak clearly as if talking to someone who is hearing impaired, result in more extreme changes in some acoustic-phonetic characteristics (e.g., pitch range, speaking rate, vowel duration, vowel space) than speech enhancements elicited in spontaneous speech (e.g., elicited using a fill-in-the-blank

worksheet in a map task: Scarborough & Zellou, 2013; using 'spot the difference' picture tasks with noise: Hazan & Baker, 2011). The acoustic-phonetic modifications in the speech produced for an imaginary hard-of-hearing listener (as compared to the speech produced for a real listener) also involved reduced coarticulation (i.e., less overlap between vowels and nasal consonants; Scarborough & Zellou, 2013).

Variations in the degree of acoustic-phonetic modifications are observed within different types of spontaneous speech as well. For example, Hazan and Baker (2011) demonstrated that talkers make different types of acoustic modifications depending on the type of noise that their listeners are experiencing. In the study, talkers engaged in a "spot the difference" picture task with a partner who heard the speech in different masking conditions, including listening to speech through a three-channel noise-excited vocoder, or with multi-talker babble. When interacting with a partner who was listening to their speech in the multi-talker babble condition, talkers made greater changes in terms of F0, intensity, and vowel formants compared to when they were interactions with a partner who was in the vocoder condition, suggesting that talkers modified their speech differently depending on the type of communicative barrier that their listeners were experiencing. The presence of an actual listener also impacts talkers' acoustic modifications in spontaneous speech. When producing foreigner-directed speech, native talkers employed more extreme changes in durations and vowel space when giving instructions to an imagined non-native listener in a map task, compared to when talking to a real non-native listener (present in the room: Scarborough et al., 2007). Thus, these studies demonstrate that the characteristics of speech enhancements can be greatly influenced by the methods of eliciting the speech.

As demonstrated in these studies, talkers make speech enhancements not only when

they are instructed to speak clearly for an imagined listener but also when listeners' communicative needs are signaled in the communication context. This is further supported by the findings that talkers are able to enhance acoustic features of the speech in a contextually-relevant way. For example, when a listener misunderstands a particular part of an utterance (e.g., a specific word), talkers selectively enhance that part of the utterance to correct the misunderstanding (e.g., Maniwa, Jongman, & Wade, 2009; Ohala, 1994; Oviatt, Levow, Moreton, & MacEachern, 1998; Schertz, 2013; Stent, Huffman, & Brennan, 2008). Specifically, when native English talkers spoke to a simulated speech recognizer and received a feedback that the utterance was misunderstood (e.g., the talker says "pit" but the computer guesses "bit"), the talkers enhanced the misunderstood contrast by manipulating a relevant acoustic feature (e.g, VOTs of the /p/ and /b/) in the second repetition (Schertz, 2013). This type of targeted error correction did not occur when the talker received an open-ended request for repetition (e.g., "???"). Such targeted segmental enhancements in response to listeners' feedback have also been found for a temporal aspect of a vowel contrast (English /i/-/ɪ/: Schertz, 2013) as well as for temporal and spectral aspects of English fricative contrasts (Maniwa et al., 2009).

Furthermore, talkers make contextually-relevant speech enhancements even without feedback from the listener. For example, in a communicative task involving conveying information to a listener, native English talkers exaggerated differences in VOTs of English word-initial consonants (e.g., /p/-/b/) when a target word to communicate (e.g., *pill*) was displayed with another word that is minimally different (e.g., *bill*), compared to when it was not (Baese-Berk & Goldrick, 2009; Buz, Jaeger, & Tanenhaus, 2014; Buz, Tanenhaus, & Jaeger, 2016). Similar types of contextually-relevant hyperarticulation have been

observed for an English word-final fricative voicing contrast (e.g., *dose* vs. *doze*: Seyfarth, Buz, & Jaeger, 2016). Further, it has been suggested that contextually-relevant hyperarticulation of a target word may only occur in the context of other words that are sufficiently similar to the target word (e.g., one major phonological feature away: Kirov & Wilson, 2012). The researchers showed that native English talkers exaggerated VOTs of word-initial voiceless stop consonants (e.g., ***c**ap*) when a word differing in place of articulation (e.g., ***t**ap*) was contextually co-present, but not when a word differing by both place and manner of articulation (e.g., ***k**ilt* vs. ***h**ilt*) was contextually co-present (Kirov & Wilson, 2012). Though the investigation of such contextually-relevant hyperarticualtion has mostly been limited to native talkers' productions, one study demonstrated that highly proficient non-native talkers exaggerated a non-native contrast (e.g., /æ/-/ɛ/) when a target word (e.g., *sat*) was placed next to a similar word (e.g., *set*) in a word-communication task (Hwang, Brennan, & Huffman, 2015). Thus, these studies have demonstrated that experienced talkers (i.e., native talkers and highly proficient non-native talkers) are able to make targeted speech enhancements based not only on listeners' feedback but also on potential communication difficulty signaled in the context.

In sum, these studies have demonstrated that talkers' speech enhancements can be elicited differently using a variety of tasks, ranging from a reading task with explicit instructions to speak clearly, to a communication task where talkers produce unscripted spontaneous speech. Given that characteristics of the elicitation task influence the way native talkers make speech enhancements (e.g., Hazan & Baker, 2011), it is possible that the nature of task influences the types of acoustic-phonetic enhancements made by non-native talkers of different proficiency levels. Thus, in order to better understand how talkers

of different linguistic backgrounds implement speech enhancement strategies, we investigate native and non-native talkers' speech enhancements produced in different contexts, including in clear speech, where they read materials based on explicit instructions to speak clearly, as well as in a more communicative context, where listeners' needs for enhanced speech intelligibility are signaled rather implicitly in the interaction.

1.4. Perceptual consequences of speech enhancements

As the primary goal of speech enhancements is accommodate listeners' communicative needs, it is critical to examine how well the acoustic-phonetic modifications implemented by talkers benefit listeners' perception. One way to investigate perceptual benefits resulting from speech enhancements it to examine an improvement in listeners' understanding of the speech: intelligibility. Previous work has widely demonstrated that English clear speech results in an intelligibility improvement for listeners (e.g., Ferguson, 2004; Ferguson & Kewley-Port, 2002; Picheny, Durlach , & Braida, 1985). In these studies, native English talkers are typically asked to read materials in a plain-speaking style and a clear-speaking style, and listeners evaluate the intelligibility by listening these types of speech with noise, and by transcribing or repeating it (e.g., Bradlow & Bent, 2002; Bradlow & Alexander, 2007). Studies have reported robust intelligibility gains resulting from native English talkers' clear speech enhancements for native English listeners of various characteristics, including hearing-impaired listeners and non-native listeners (e.g., Bradlow & Bent, 2002; Bradlow & Alexander, 2007; Ferguson, 2004; Krause & Braida, 2002; Liu et al., 2004; Picheny et al., 1985; Schum, 1996; Uchanski et al., 1996). A similar clear speech intelligibility benefit has also been reported for native

talkers and native listeners of languages other than English, including Croatian and French (Gagné et al., 1994; Gagné, Rochette, & Charest, 2002; Smiljanić & Bradlow, 2005).

Previous studies also report perceptual benefits associated with speech enhancements that are produced in different contexts than typical clear speech elicitation contexts. That is, native listeners benefit from native talkers' speech enhancements that are produced without explicit instructions to speak clearly. For example, native English listeners made lexical decisions faster when responding to native English talkers' speech that was produced with a real listener present in the room, as compared to when responding to the speech produced for an imagined hard-of-hearing listener (simulated clear speech; Scarborough & Zellou, 2013). Native English listeners understood native English speech better when the speech was produced in a conversation with a non-native English partner than with a native English partner (Lee & Baese-Berk, 2020). Furthermore, native English listeners made word identification responses faster when listening to spontaneous native English speech produced in a situation with a communication barrier (i.e., vocoded speech signal), as compared to the spontaneous speech produced in a situation without a barrier (Hazan, Grynpas, & Baker, 2012). These studies have demonstrated that native talkers' speech enhancements produced in various contexts (e.g., when reading materials based on explicit instructions to speak clearly, when conversing with partners with or without a communication barrier) improve listeners' understanding of the speech as well as how fast they process the information.

However, it is much less well-understood how speech enhancements made by non-native talkers of differing proficiency are perceived by native listeners. Though there is some evidence that speech enhancements made by highly proficient non-native talkers are

as effective as those made by native talkers, data from talkers of lower proficiency are scarce, making it difficult to directly examine whether talkers' ability to improve speech intelligibility in a non-native language improves as their proficiency develops. Specifically, previous work has shown a comparable size of intelligibility gains resulting from clear speech enhancements made by native talkers and by highly proficient or early learners (Rogers et al., 2010; Smiljanić & Bradlow, 2005, 2011). However, it has also been demonstrated that late learners of English were much less effective at enhancing intelligibility of English vowels than early learners and native English talkers (Rogers et al., 2010). Though these studies show some evidence that non-native talkers' target language experience impacts the perceptual benefits resulting from their speech enhancements, it is difficult to generalize such results beyond the level of single sound production (e.g., English vowels in /bVd/ syllables: Rogers et al., 2010). That is, it is unknown how native listeners would benefit from speech enhancements made for longer phrases or sentences, which requires a proficient use of the target language sound system at multiple levels, including phrasing and prominence structure at the sentence level (see Ladd, 2008 for examples), by non-native talkers of different proficiency levels. Furthermore, it is also not clear whether non-native talkers' effort to enhance intelligibility in a communicative context (without explicit instructions to speak clearly) results in perceptual benefits for native listeners. Thus, in order to better understand how speech enhancements made in a non-native language benefit native listeners' understanding, it is critical to examine perception of speech enhancements made for different types of materials (e.g., words, phrases, and sentences), produced in a variety of contexts, including when talkers are explicitly asked to read materials in a clear manner, as well as when talkers adapt their

speech based on listeners' communicative needs signaled implicitly in the context.

Though one way to investigate perceptual benefits of speech enhancements is to examine intelligibility, broader literature on speech perception suggests that listeners' perception of speech is much more diverse than the correct recognition of words or phrases (e.g., Cargile, Giles, Ryan, & Bradac, 1994). That is, listeners not only understand the information communicated in the speech but also subjectively evaluate characteristics of the talker or the speech. For example, literature on perception of accented speech has demonstrated that listeners' subjective evaluations (e.g., intelligence, confidence, communicative ability, or friendliness of the talker) can be impacted by different varieties of regional- or foreign-accented speech (e.g., Adank, Stewart, Connell, & Wood, 2013; Coupland & Bishop, 2007; Kraut & Wulff, 2013; Tsurutani, 2012). Studies on perception of non-native speech has also suggested that listeners' perception could differ depending on how they are asked to evaluate the speech, including measures of comprehensibility (i.e., how easy or difficult listeners perceive the speech is to understand), accentedness (i.e., the degree of foreign accent of the speech that listeners perceive), and credibility (i.e., how credible listeners perceive the information conveyed in non-native speech to be; Munro & Derwing, 1995a, 1999; Lev-Ari & Keysar, 2010; Smiljanić & Bradlow, 2011). These studies demonstrate that listeners' performance on one perception measure does not necessarily correspond to that on another measure. For example, listeners can understand non-native speech (i.e., intelligibility) even if they perceive the same speech to be heavily accented (Munro & Derwing, 1999; Smiljanić & Bradlow, 2011) or not easy to understand (Sheppard, Elliot, & Baese-Berk, 2017). Thus, these lines of work suggest that there can be a gap between what listeners actually understand from the speech and how they perceive

the speech in subjective terms.

There is also some evidence suggesting that perceptual consequences resulting from speech enhancements are multifaceted. For example, perceptual benefits of speech enhancements were observed not only in how fast listeners understood the speech, but also in how clear listeners perceived the speech to be (Hazan et al., 2012). Specifically, when listening to spontaneous native English speech produced in a situation with a communication barrier, native listeners made word identification responses faster, and also perceived the speech to be clearer, as compared to when listening to the speech produced for a listener in an easy listening condition. Furthermore, Smiljanić and Bradlow (2011) examined intelligibility and perceived degree of foreign accent for clear and plain speech produced by non-native (native Croatian) talkers of English. The highly proficient non-native talkers' clear speech was more intelligible than plain speech, but perceived degree of foreign accent was similar between the two styles of the speech for native English listeners, revealing a partial independence of these two perceptual measures from one another. These studies suggest that speech enhancements may be reflected in listeners' perception differently depending on how listeners evaluate the speech. In other words, there may be aspects of perceptual consequences of speech enhancements that we do not necessarily understand by only examining an improvement in intelligibility. Thus, the current work examines how native listeners benefit from speech enhancements produced by native and non-native talkers of different proficiency levels in different contexts, in terms of benefits in intelligibility as well as in subjective terms of perception.

1.5. Current research

The goal of this dissertation is to provide a better understanding of goal-oriented speech accommodation behavior for talkers of different linguistic backgrounds, as well as perceptual consequences of these accommodations. Specifically, in order to better understand how talkers' ability to enhance intelligibility is shaped by their target language experience (e.g., native vs. non-native status; higher- vs. lower-proficiency of the non-native language) in different tasks, we examine native and non-native English talkers' speech enhancements in a reading task where talkers are explicitly asked to read materials clearly (i.e., clear speech enhancements), as well as in a simulated communication task where listeners' needs for enhanced intelligibility for particular sound contrasts are signaled implicitly in the context (i.e., contextually-relevant speech enhancements). We further examine perceptual consequences of these acoustic-phonetic enhancements in terms of intelligibility (whether listeners understand the speech) and subjective evaluations of the speech, including perceived degree of comprehensibility (how easy it is to understand the speech) and perceived degree of talker effort (how hard the talker is trying to speak clearly).

1.5.1. Novel contributions of the current research

Examining speech enhancement strategies used by talkers of varying target language experience and their perceptual consequences has novel contribution in theoretical and practical domains. Particularly, by examining goal-oriented acoustic modifications produced by non-native talkers of differing target language proficiency levels, this work provides insights regarding the applicability of H&H theory (Lindblom, 1990). That is, as discussed earlier, the acoustic modifications implemented along the

continuum of hypo- and hyper-speech have largely been documented based on the productions of the talkers who are fluent in the target language (e.g., native talkers, highly proficient non-native talkers). Thus, it is unclear how the talker-oriented force (i.e., economy of effort) and listener-oriented force (i.e., the need for perceptual discriminability) together impact acoustic adjustments made by talkers of limited target language proficiency. While one study has demonstrated that late English learners are less able to improve intelligibility of English vowels compared to early English learners and native talkers (Rogers et al., 2010), the source of such smaller intelligibility gains for late learners' productions is unclear. That is, it is possible that the small intelligibility gains were associated with small stylistic changes (plain vs. clear speech). It is also possible that late learners made stylistic changes, but they were qualitatively different from those of early learners and native talkers, resulting in smaller intelligibility improvement for native English listeners. The current work examines not only perceptual benefits resulting from speech modifications but also acoustic characteristics of these modifications. This allows us to ask how talkers' target language proficiency is associated with the range of stylistic variations that they are able to implement in their productions, as well as whether such goal-oriented modifications are perceptually effective to native listeners.

Furthermore, by exploring different aspects of perceptual benefits resulting from speech enhancements, the current work highlights multi-faceted nature of speech enhancement perception. Specifically, we examine whether native and non-native talkers' speech enhancements result in perceptual benefits for native listeners, not only in terms of actual understanding of the speech as often examined in previous clear speech studies (e.g., Bradlow & Bent, 2002; Ferguson & Kewley-Port, 2002; Smiljanić & Bradlow, 2011) but

also in terms of listeners' subjective evaluations of the speech. As studies examining perception of non-native speech suggest that intelligibility of the speech is at least partially independent from perceived degrees of comprehensibility or foreign accentedness (Derwing & Munro, 2009; Munro & Derwing, 1995a; Smiljanić & Bradlow, 2011), it is possible that such dissociation is observed in perception of speech enhancements as well. By exploring the potential gap among different measures of perceptual benefits resulting from native and non-native speech enhancements, we aim to understand perceptual consequences of speech enhancements more broadly than previous studies have discussed.

The current work also has practical implications for second language instruction. Particularly, investigating second language learners' ability to implement goal-oriented variations in acoustic-phonetic characteristics of speech could inform the development of pronunciation training methods. Research on second language teaching suggests the importance of pronunciation training that is aimed at achieving mutual intelligibility rather than reducing a foreign accent in learners' productions (e.g., Derwing & Munro, 2005, 2009). By exploring learners' ability to vary their productions to improve intelligibility for listeners, the current work could advocate for the importance of providing explicit pronunciation instruction to help learners achieve intelligible speech production. Such focus on the strategies to improve speech intelligibility implemented by talkers of differing levels of target language proficiency may also inform behavioral therapy techniques for speakers with speech impairments, including speakers with dysarthria and Parkinson's disease (e.g., Duffy, 2005; Hustad & Weismer, 2007; Lam & Tjaden, 2016). Specifically, examining lower-proficiency non-native talkers' speech enhancement patterns may help us better understand the difficulty associated with implementing stylistic variations for talkers

with limited language proficiency, and identify future directions to explore intervention techniques for such population.

1.5.2. Hypotheses explored in the dissertation

In this dissertation, we explore production and perception of speech enhancements. One hypothesis we explore throughout is that talkers' target language experience impacts the acoustic-phonetic enhancements made by native and non-native talkers of different proficiency levels. Given previous results suggesting that L2 production is more effortful than L1 production especially for talkers of lower L2 proficiency (e.g., Kormos & Dénes, 2004; Nip & Blumenfeld, 2015), we expect that increased production difficulty associated with lower target language proficiency will be manifested in general characteristics of talkers' speech. For example, we may observe that speaking rate is generally slower for lower-proficiency non-native talkers' speech than for higher-proficiency non-native talkers' speech, and for higher proficiency talkers' speech than for native talkers' speech. Lower-proficiency talkers may also make smaller acoustic differences between non-native English sound contrasts (e.g., differentiating the word *cab* from *cap*) as compared to higher-proficiency talkers and native talkers do. It is possible that such influence of talkers' target language proficiency level on speech production will extend to their ability to make clear speech enhancements. That is, compared to lower-proficiency talkers, higher-proficiency talkers may make larger modifications to their speech, in acoustic features such as speaking rate, fundamental frequency, and vowel space (e.g., Bradlow et al., 2003; Smiljanić & Bradlow, 2005). Further, the size of higher-proficiency talkers' enhancements could be comparable to those of native English talkers, given that highly proficient non-native

talkers use similar clear speech strategies as native talkers of the language (e.g., Granlund et al., 2012). We also examine talkers' target language experience on acoustic enhancements at the segmental level. Given the previous work suggesting that making appropriate segmental enhancements takes extensive experience with the sound structure of the language (e.g., Granlund et al., 2012; Smiljanić & Bradlow, 2008a), we predict that higher-proficiency talkers are better able to make segmental enhancements than lower-proficiency talkers. However, given that learners' productions of L2 sounds are also influenced by their L1 sound system (e.g., Brière, 1966; Lado, 1957), it is possible that the effect of talkers' target language experience on segmental enhancements varies depending on the relationship between L1 and L2 sound system. Thus, we examine segmental enhancements for different types of non-native sounds contrasts (i.e., non-native contrasts that exist or do not exist in talkers' native language).

The current work also explores perceptual aspects of speech enhancements broadly. Particularly, in addition to listeners' understanding of the speech (i.e., intelligibility), we examine subjective terms of perception, including listeners' perception of how easy the speech is to understand (i.e., perceived degree of comprehensibility) and perception of how hard the talker is trying to speak clearly (i.e., perceived degree of talker effort). We examine these subjective aspects of listeners' perception in order to better understand perceptual consequences of speech enhancements that may not necessarily be manifested in the intelligibility measure alone. Specifically, given that when listening to non-native speech, intelligibility measure does not necessarily correspond to listeners' perception of comprehensibility (Munro & Derwing, 1999; Sheppard et al., 2017), it is possible that non-native talkers' attempt to enhance acoustic characteristics of their speech results in

improvement in intelligibility but not in perception of comprehensibility, or vice versa. Further, especially for lower-proficiency non-native talkers' speech, native listeners may not necessarily understand the clear speech better than plain speech (e.g., late learners in Rogers et al., 2010) or perceive clear speech to be easier to understand than plain speech, but they may still be sensitive to talkers' increased effort to speak clearly. Such potential gap among different perceptual measures of speech enhancements could also be present in native listeners' perception of native talkers' speech. By investigating multiple aspects of listener's perception of speech enhancements produced by native and non-native talkers of differing proficiency, the present work highlights how talkers' attempt to enhance intelligibility translates to different aspects of listeners' perception. We discuss perceptual consequences for listeners in relation with acoustic characteristics of speech enhancements as well as how talkers' speech enhancements are elicited (e.g., via read speech with explicit instructions to speak clearly, via interaction with a conversation partner).

1.5.3. Structure of the dissertation

The four experimental chapters of this dissertation are written as separate research papers, and thus, each has their own introduction, methods, results, and discussion sections. First, we examine production (Chapter 2) and perception (Chapter 3) of clear speech enhancements. Chapter 2 describes acoustic characteristics of clear speech enhancements produced by native English talkers and non-native English (native Mandarin) talkers of higher- and lower-proficiency. These talkers read simple English sentences in a plain- and a clear-speaking style. We carried out a series of acoustic analysis to investigate whether the size of acoustic enhancements differs depending on talkers'

target language experience (i.e., for native vs. non-native talkers, for higher- vs. lower-proficiency non-native talkers). Chapter 3 presents a series of perception experiments investigating how native English listeners benefit from the clear speech enhancements made by the native and non-native English talkers of higher- and lower-proficiency. In the intelligibility task, we examine whether listeners' understanding improves from plain speech to clear speech. We also examine whether the plain-to-clear speech intelligibility gains differ for the speech produced by native talkers and non-native talkers of different proficiency levels. Furthermore, we explore perceptual benefits associated with clear speech enhancements in terms of listeners' subjective evaluations of the speech. Specifically, listeners evaluate the plain and clear speech for perceived degree of comprehensibility (i.e., whether they perceive clear speech to be easier to understand than plain speech) and perceived degree of talker effort (i.e., whether they perceive clear speech to be produced with increased effort than plain speech). Using these multiple measures of perception, we examine whether perceptual benefits of clear speech enhancements are manifested similarly in different aspects of perception.

The second section of this dissertation examines production (Chapter 4) and perception (Chapter 5) of contextually-relevant speech enhancements. Chapter 4 describes acoustic characteristics of contextually-relevant speech enhancements produced by native English talkers and non-native English (native Mandarin) talkers of higher- and lower-proficiency. These talkers participated in the word-naming communication task, where they communicated target words (e.g., *cap*) to a listener when a phonetically similar minimal-pair neighbor (e.g., *cab*) either was or was not present in the context. We examine acoustic characteristics of speech modifications made in these different contexts, asking how

talkers' native language status and non-native talkers' proficiency level impact the size of the modifications, as well as how the effects of talkers' target language experience on the contextually-relevant enhancements differ depending on the talkers' familiarity with the target sound contrast (i.e., a contrast that also exists in non-native talkers' native language vs. a contrast that does not). Chapter 5 explores perceptual consequences of the contextually-relevant speech enhancements by examining native English listeners' understanding of the speech produced in different contexts, as well as subjective evaluations of these types of speech.

In Chapter 6, I summarize the overall findings, and discuss the potential implications and future directions of this research.

CHAPTER II: PRPDUCTION OF CLEAR SPEECH

2.1. Introduction

In order for speech communication to be successful, talkers must be able to deliver their messages clearly to their listeners. A crucial aspect of successful message delivery is the talker's ability to accommodate their speech in different communicative situations in order to make their speech more intelligible for their listeners (Lindblom, 1990). Despite the wide breadth of research on clear speech (Uchanski, 2005) that has demonstrated that native talkers of the language are able to modify various acoustic-phonetic features of their speech to make it more understandable to their listeners (e.g., by speaking more slowly, loudly, or by enunciating individual sounds more clearly: Bradlow et al., 2003; Picheny et al., 1986), much less is known about clear speech strategies used by non-native talkers of the language. Particularly, it is unclear how non-native talkers of differing target language proficiency levels try to enhance intelligibility of their speech, and how their clear speech strategies differ from native talkers' strategies. In order to better understand how non-native talkers' clear speech strategies change as their target language proficiency develops, the present study characterizes acoustic features of clear speech enhancements produced by native English talkers and non-native English talkers of different proficiency levels.

2.1.1. Clear speech

Clear speech is a speaking style that talkers use when they are aware that their listeners may have difficulty understanding them, possibly because the listeners have hearing impairments or are non-native listeners of the language (Smiljanić & Bradlow, 2009; Uchanski, 2005). This speaking style adjustment has often been understood in terms

of Hyper-Hypo (H&H) theory (Lindblom, 1990). According to this theory, speech communication is characterized as a dynamic balance between talker- and listener-oriented forces. Specifically, talkers try to minimize their effort to articulate sounds, which stands in the opposite end from the listener-oriented requirement for sufficient perceptual discriminability of sounds. This balance between the two forces varies depending on the communication context. That is, if the talker interprets that the communication context places extra demand on the listener, decreasing their chance of successfully understanding the message, the talker increases articulatory effort (hyperarticulate or hyperspeech) to make their speech easier to understand for the listener; whereas if the context is favorable for ease of communication, the talker tries to minimize their articulatory effort (hypoarticulate or hypospeech). Thus, talkers adjust their articulatory effort on the continuum between hypo- and hyperspeech, and clear speech enhancements can be characterized as a part of talkers' articulatory adjustments along the hyper articulated end of the continuum.

In order to elicit talkers' clear speech enhancements, researchers have typically asked talkers to read the same set of materials twice: once in a plain- and once in a clear-speaking style (Bradlow & Alexander, 2007; Ferguson & Kewley-Port, 2002; Ferguson 2004; Granlund et al., 2012; Picheny et al., 1986; Rogers et al., 2010; Schum, 1996; Smiljanić & Bradlow, 2005, 2011). For the plain-speaking style, talkers are instructed to read the materials as if they are talking to someone familiar with their voice and speech patterns; for the clear-speaking style, talkers are instructed to read the same materials as if they are talking to a listener with a hearing loss or to a non-native listener of the language (e.g., Bradlow & Alexander, 2007; Smiljanić & Bradlow, 2011). Acoustic characteristics of

the enhance speaking style (i.e., clear speech) are compared to the characteristics of the speech in the baseline speaking style to (i.e., plain speech) to examine talkers' acoustic-phonetic modifications. The current study applies this elicitation method to investigate clear speech strategies employed by native talkers as well as non-native talkers of different proficiency levels.

2.1.2. Acoustic characteristics of clear speech

Previous studies have demonstrated that native talkers of the language make a variety of acoustic-phonetic modifications in clear speech. The modifications include a decrease in speaking rate (characterized by longer segments as well as longer and more frequent pauses), higher pitch (F0), wider F0 range, increased intensity, and increased energy in the 1000-3000 Hz range of long-term spectra (e.g., Bradlow et al., 2003; Liu et al., 2004; Picheny et al., 1986; Smiljanic & Bradlow, 2005). In addition to these global modifications that improve the overall salience of the speech signal (i.e., making the speech more audible in adverse listening conditions; Bradlow & Bent, 2002), talkers also make enhancements at the segmental level. For example, compared to plain speech, native English talkers release word-final consonants more frequently in clear speech (Bradlow et al., 2003; Krause & Braida, 2004; Picheny et al., 1986). Talkers also make modifications to make the phonological categories of the language more distinct from one another (Lindblom, 1990; Johnson et al., 1993); for example, they increase the duration difference in voice-onset-timing (VOT) between voiced and voiceless stops (Krause & Braida, 2004). Furthermore, in clear speech, talkers generally increase duration of vowels (Ferguson & Kewley-Port, 2002; Smiljanić & Bradlow, 2008a), differentiate the duration of tense and

lax vowels (Uchanski, 1988), and expand their vowel space (e.g., Bradlow, 2002; Bradlow et al., 2003; Johnson et al., 1993; Krause & Braida, 2004; Moon & Lindblom, 1994; Smiljanić & Bradlow, 2005). Thus, these studies demonstrate that native talkers make various types of acoustic modifications to make their speech more intelligible to listeners.

Despite the wealth of information on the clear speech enhancement strategies used by native talkers of the language, there is limited data regarding the strategies that are employed by non-native talkers of the language. Several studies suggest that clear speech enhancements made by highly proficient non-native talkers are as effective as those made by native talkers. This is shown in the comparable size of intelligibility gains resulting from clear speech produced by native talkers and by highly proficient or early learners of the language (Rogers et al., 2010; Smiljanić & Bradlow, 2005, 2011). Further, Granlund et al. (2012) have demonstrated that the types of acoustic modifications made by native talkers and by proficient non-native talkers are similar. Specifically, proficient Finnish-English bilinguals and native English talkers used similar clear speech strategies in terms of their modifications of F0, intensity, and mean word duration. Bradlow (2002) has also shown that the extent of vowel space expansion is similar between the clear speech productions of native English talkers and early Spanish-English bilinguals. These studies have suggested that highly proficient non-native talkers use similar clear speech strategies as native talkers of the language.

However, it is not clear what types of clear speech strategies are used by talkers with limited language proficiency (e.g., non-native talkers of lower proficiency). Based on the previous work suggesting that lower proficiency in the target language is associated

with increased effort and increased cognitive resources involved with speech production (e.g., Green, 1998; Kormos & Dénes, 2004; Poulisse, 1997; Roelofs & Verhoef, 2006), as well as with less developed control to use the sound system of the non-native language (e.g., Bohn & Flege, 1992; Fabra & Romero, 2012; Nip & Blumenfeld, 2015), it is possible that such an effect of target language proficiency on non-native speech production in general also impacts talkers' ability to manipulate acoustic-phonetic properties of their speech. However, with the currently available set of data, it is difficult to determine how the ability to make phonetic modifications to accommodate listeners' needs for enhanced intelligibility differ for non-native talkers of higher- and lower-proficiency. Particularly, though one study has demonstrated that late learners of English were much less effective at enhancing intelligibility of English vowels than early learners and native English talkers (Rogers et al., 2010), it is not clear how the late learners' clear speech enhancements differed acoustically from those of early learners and native talkers. Furthermore, it is difficult to determine whether there is a difference in clear speech strategies at a more global level (e.g., changes in speaking rate, F0, and intensity) for native talkers and non-native talkers of different proficiency levels. In order to better understand how second language learners' clear speech strategies improve as their target language proficiency develops, it is critical to examine clear speech strategies produced by learners of differing proficiency levels.

2.1.3. Current study

In the current study, we examine acoustic characteristics of clear speech enhancements produced by native English talkers and non-native talkers of higher- and

lower-proficiency. Previous research has demonstrated that in English, clear speech enhancements involve a range of acoustic-phonetic modifications including a decrease in speaking rate (characterized by more frequent and longer pauses as well as longer segment duration), an increase in overall pitch and pitch range, an increase in intensity, as well as an increase in the vowel space (e.g., Bradlow et al., 2003; Johnson et al., 1993; Krause & Braida, 2004; Moon & Lindbom, 1994; Smiljanić & Bradlow, 2005). Based on these findings, we examine talkers' clear speech enhancements in several features: temporal characteristics (speaking rate, and frequency and duration of silent pauses), fundamental frequency, intensity, and vowel space.

We compare the acoustic measurements of these features between plain- and clear-style productions of the same sentences produced by native talkers and non-native talkers of higher- and lower-proficiency. Given the previous findings that proficient non-native talkers' clear speech modifications are similar to those of native talkers (Bradlow, 2002; Granlund et al., 2012), we expect that native talkers and higher-proficiency non-native talkers will modify the target acoustic features to a similar extent. However, lower-proficiency talkers' acoustic modifications may differ from those of higher-proficiency talkers and native talkers. Specifically, given that non-native speech production can be generally more effortful for talkers of lower-proficiency (e.g., Poulisse, 1997), and also that late English learners' clear speech modifications of English vowels resulted in much smaller intelligibility gains compared to those of early learners and native talkers (Rogers et al., 2010), we expect that the size of plain-to-clear speech modifications of lower-proficiency talkers will be smaller than that of higher-proficiency talkers and native talkers.

In addition to characterizing acoustic features of clear speech enhancements, we

also examine non-native talkers' English proficiency using multiple measures in order to ensure that higher- and lower-proficiency non-native talkers examined here are indeed of different proficiency levels. Specifically, we use the information collected from a language background questionnaire (e.g., information about length of residence in the English-speaking country, standardized English proficiency test score) as well as non-native talkers' perceived accentedness (evaluated by native English listeners) to characterize their English proficiency.

2.2. Method

2.2.1. Participants

Participants were 4 native English talkers (age range = 19 - 22 years, mean = 20) and 8 non-native English talkers whose native language was Mandarin Chinese (age range = 20 - 31 years, mean = 25.3). All talkers identified themselves as female, and reported no history of speech or hearing impairment.

The native English talkers were recruited from the Psychology and Linguistics subject pool at the University of Oregon. In order to recruit non-native English talkers of different proficiency levels, we recruited them from two different instructional settings. Specifically, we recruited 4 higher-level non-native talkers from the graduate student population at the University of Oregon, and 4 lower-level non-native talkers from an intensive English program, who were international students hoping to enter the university as matriculated students. Table 2.1 shows the information regarding non-native talkers' English learning background and proficiency. As shown in the table, lower-proficiency native Mandarin (Native Mandarin-Low) talkers and higher-proficiency native Mandarin

(Native Mandarin-High) talkers have different characteristics, particularly in terms of length of US residence and the Test of English as a Foreign Language (TOEFL) score.

Additionally, 40 native English listeners (13 females, 27 males; age range = 23 - 67 years, mean = 35.8) participated in the foreign accent rating task evaluating the accentedness of the talkers. None of the listeners provided the speech samples. The listeners were recruited via Amazon Mechanical Turk.

Table 2.1. Non-native English (native Mandarin) talkers' English learning background and proficiency.

Talker	Age	Age of onset for English speaking	Years of formal English training	Length of US residence in months	TOEFL score
NativeMandarin-Low (NM-L): Average	**20.5**	**15.5**	**7**	**18.8**	**45.3**
NM-L 103	21	19	9	24	52
NM-L 104	20	15	6	15	35
NM-L 106	20	13	7	19	53
NM-L 107	21	15	6	17	41
NativeMandarin-High (NM-H): Average	**30**	**17.5**	**15.3**	**62.5**	**93.8**
NM-H 302	31	23	12	27	108
NM-H 306	30	13	9	108	91
NM-H 310	28	10	15	19	106
NM-H 311	31	24	25	96	70

2.2.2. Materials

The test materials were 30 sentences chosen from the Revised Bamford-Kowal-Bench Standard Sentence Test (BKB sentences; Bamford & Wilson, 1979), developed by the Cochlear Corporation for use with American children. They were simple English sentences, each sentence consisting of 3 or 4 keywords (e.g., *The shop closed for lunch*), and have been used with non-native English speakers in previous studies (e.g., Bradlow & Bent, 2002). Additionally, 60 English sentences, different from the above 30 BKB

sentences, were included in the test materials to be used for another study.

Further, talkers recorded 15 BKB sentences as practice sentences. None of the practice sentences were part of the test materials discussed above. The recordings of the 10 practice sentences were used as materials for accentedness ratings, as part of characterizing the non-native talkers' English proficiency. The test and practice BKB sentences are provided in Appendix A.

2.2.3. Procedure

All talkers were recorded in a sound booth. The sentences were displayed on the computer screen one at a time; the presentation of each sentence was self-paced. The talkers read into a microphone that fed directly into a desktop computer. Recording was done on a single channel at a sampling rate of 44,100 Hz (16 bit) using the Praat speech analysis software package (Boersma & Weenink, 2001). The talkers first recorded the practice sentences. They were instructed to practice reading sentences to the microphone. Then, the talkers read the test sentences once in a plain-speaking style and once in a clear-speaking style. For the recordings in the plain-speaking style, the talkers were instructed to read as if they were talking to someone who is familiar with their voice and speech patterns. For the recordings in the clear-speaking style, the talkers were instructed to read as if they were talking to a listener who has a hearing loss (Smiljanić & Bradlow, 2011). After the recording, talkers completed a language background questionnaire and other proficiency measuring tasks. The entire session lasted approximately one hour. All speech files were segmented into individual sentence-length files.

The recordings of the practice sentences were evaluated for foreign accentedness by

native English listeners. The sentence-length files of the practice sentences were RMS normalized to 65dB SPL. In the perception task, conducted via Qualtrics, the listeners were told that they would listen to English sentences and evaluate the foreign accent of the speech. In each trial, listeners heard an English sentence without noise and were instructed to rate the accentedness of the speech on a scale of 1 ("a native speaker of English") through 9 ("an extremely strong foreign accent"; similar to Munro & Derwing, 1995a). In order to prevent the accentedness ratings from being influenced by the intelligibility of the speech, the transcript of the sentence was displayed while the listeners were listening to the speech (Gittleman & Van Engen, 2018). Each sentence could not be played more than once, but there was no time limit for responding. Twenty listeners evaluated 6 talkers (i.e., 2 native English, 2 higher-proficiency non-native talkers, 2 lower-proficiency non-native talkers) and another set of 20 listeners evaluated the other 6 talkers. Thus, each listener evaluated 60 sentences (i.e., 10 unique sentences x 6 talkers). The presentation of the sentences was randomized for each listener.

2.2.4. Acoustic analysis

2.2.4.1. Speech rate and pause

In order to examine how temporal characteristics are manifested in the plain and clear speech produced by native talkers and non-native talkers of different proficiency levels, we analyzed two aspects: speech rate and silent pause. In order to examine speech rate in terms of speaking rate (i.e., pause duration included in the calculation) and articulation rate (i.e., pause duration excluded from the calculation), we first counted the number of silent pauses and measured their duration for each sentence. We defined a silent

pause as any silence equal to or longer than 250 ms (de Jong, Groenhout, Schoonen, & Hulstijn, 2015; Goldman-Eisler, 1968; Kahng, 2018). Then, speaking rate was computed by dividing the number of syllables of the sentence by the sentence duration (in seconds). Articulation rate was computed for each sentence by dividing the number of syllables of the sentence by the sentence duration after excluding pause duration (if any). Further, in order to account for individual variability (e.g., some talkers speak faster than others), scaled values of the articulation rate were also computed using the min-max scaling procedure (Gerstman, 1968; Kallay & Redford, 2018). That is, articulation rate for a particular sentence was normalized using the talker's minimum and maximum values of the articulation rate, so that all the values are within the range of 0 (minimum value of that talker) to 1 (maximum value of that talker).

2.2.4.2. Pitch and intensity

In order to examine characteristics of fundamental frequency (F0) and intensity, we measured mean F0, F0 range, and mean intensity for each sentence in each of the plain- and clear-speaking styles for each talker (Bradlow et al., 2003). A Praat script was run to calculate mean F0, maximum F0, minimum F0 (in Hertz), and mean intensity (in dB) for each sentence. F0 range was obtained by subtracting the minimum F0 value from the maximum F0 value for each sentence. The values of mean F0, F0 range, and mean intensity were transformed using the min-max scaling procedure described above.

2.2.4.3. Vowel space

We selected 4 point vowels in order to characterize each talker's vowel space for

clear and plain speech: /i/, /æ/, /ɑ/ and /u/ (Ferguson & Kewley-Port, 2007). Phone-level alignment between sound files and transcripts of the sentence was automated using Montreal Forced Aligner (McAuliffe et al., 2017). Then, automated vowel formant extraction was carried out using Forced Alignment and Vowel Extraction (Rosenfelder et al., 2014). *F1* and *F2* frequencies of the 4 point vowels were taken from the midpoint (i.e., 50% of the vowel duration) of each vowel. Midpoint *F1* and *F2* were then z-score normalized to control for individual differences (i.e., Lobanov method: Nearey, 1977; Thomas & Kendall, 2015). Vowel space area was measured as the Euclidean area covered by the quadrilateral defined by the mean of each of the 4 point vowels, using R package phonR (McCloy, 2016). Vowel space was calculated for each speaking style (plain and clear) for each talker.

2.3. Results

2.3.1. Acoustic characteristics of clear speech enhancements

2.3.1.1. Speech rate and pause

Figure 2.1 shows raw speaking rate (i.e., number of syllables divided by the sentence duration in seconds with pauses; left panel), raw articulation rate (i.e., number of syllables divided by the sentence duration minus pause duration; middle panel), and scaled articulation rate (right panel). The left and middle panels show that Native Mandarin talkers spoke slower than Native English talkers both in terms of the speaking rate and the articulation rate. One-way ANOVA[1] examining the effect of talker groups (i.e., Native English, Native Mandarin-High, and Native-Mandarin-Low talkers) on the raw

[1] We used ANOVA instead of mixed-effects regression models in order to avoid the risk of overfitting (e.g., Crawley, 2002), as there only four talkers in each talker group.

speaking rate was conducted using the afex package (Singmann et al., 2020) within the R computing program (R Core Team, 2020). The results showed a significant effect of talker group [$F(2, 9) = 24.34$, $p < .001$, $\eta^2_p = .84$, $\eta^2_G = .84$[2]]. A post-hoc Tukey test revealed that speaking rate was faster for Native English talkers than for Native Mandarin-High talkers ($\beta = 1.2$, SE = .08, t ratio = 5.25, $p = .001$), for Native English talkers than for Native Mandarin-Low talkers ($\beta = 1.5$, SE = .08, t ratio = 6.6, $p < .001$), but did not significantly differ for the speech of Native Mandarin-High talkers and Native Mandarin-Low talkers ($\beta = .31$, SE = .08, t ratio = 1.35, $p = .4$). Similarly, one-way ANOVA examining the effect of talker groups on the raw articulation rate (i.e., speech rate without pauses) showed a significant effect of talker group [$F(2, 9) = 22.35$, $p < .001$, $\eta^2_p = .83$, $\eta^2_G = .83$]. Post-hoc comparisons confirmed that the articulation rate differed between the speech of Native English vs. Native Mandarin-High talkers ($p = .002$) and between the speech of Native English vs. Native Mandarin-Low talkers ($p < .001$), but not between the speech of Native Mandarin-High and Native Mandarin-Low talkers ($p = .5$). These results confirmed that Native English talkers generally spoke faster than Native Mandarin talkers (both High and Low), but Native Mandarin-High talkers did not speak significantly faster than Native Mandarin-Low talkers.

[2] Following Lakens (2013), we report two types of effect size statistics: partial eta-squared (η^2_p) and generalized eta-squared (η^2_G). Generalized eta-squared (η^2_G) is a recommended effect size statistic for repeated measure designs (used for later analyses in this paper; Bakeman, 2005).

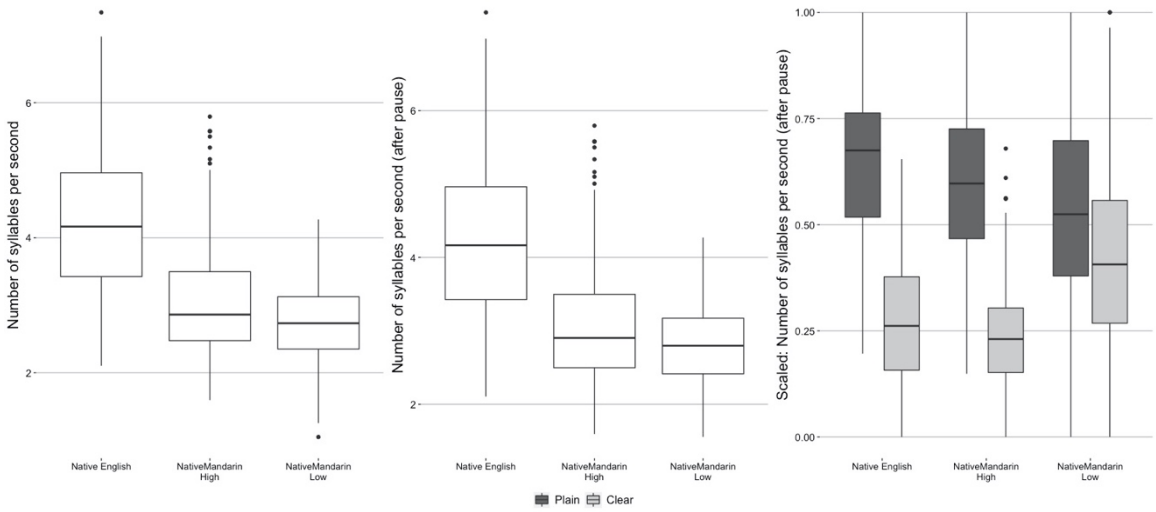

Figure 2.1. Speaking rates for different talker groups (Native English, Native Mandarin-High, and Native-Mandarin-Low talkers) and in two speaking styles (plain and clear): raw speaking rate (number of syllables divided by the sentence duration in seconds; Left panel), raw articulation rate (number of syllables divided by the sentence duration minus pause duration; Middle panel), scaled articulation rate (Right panel).

In the next analysis, we examined whether talkers produced speech more slowly in the clear-speaking style than in the plain-speaking style, and whether this pattern differed for different talker groups' speech. The right panel in Figure 2.1 (the scaled values of the articulation rate) suggests that Native English and Native Mandarin-High talkers spoke more slowly in the clear-speaking style than in the plain-speaking style, but this difference is much smaller for Native Mandarin-Low talkers' speech. A two-way within-subjects ANOVA was conducted examining the effect of speaking style (i.e., clear or plain) on the scaled articulation rate, and whether the effect of speaking style differed for the speech of different talker groups (i.e., Native English, Native Mandarin-High, and Native-Mandarin-Low talkers). Because each talker produced both speaking styles, speaking style was treated as a within-subject factor. The results showed a significant effect of speaking style [$F(1, 9) = 100.11$, $p < .001$, $\eta^2_p = .92$, $\eta^2_G = .87$], as well as a significant interaction between speaking style and talker group [$F(2, 9) = 9.09$, $p = .007$, $\eta^2_p = .67$, $\eta^2_G = .54$].

Post-hoc Tukey pairwise comparisons between speaking styles within each talker group confirmed that the plain-clear difference in the scaled articulation rate was significant for the Native English ($\beta = .36$, SE = .05, t ratio = 7.44, $p < .0001$), Native Mandarin-High ($\beta = .36$, SE = .05, t ratio = 7.59, $p < .0001$), and Native Mandarin-Low group ($\beta = .11$, SE = .05, t ratio = 2.3, $p = .047$).

In order to further examine the interaction between speaking style and talker group, a two-way within-subject ANOVA was conducted for subsets of the data: for Native English and Native Mandarin-High talkers' data, for Native English and Native Mandarin-Low talkers' data, and for Native Mandarin-High and Native Mandarin-Low talkers' data. These tests revealed that the speaking style x talker group interaction was significant for the Native English vs. Native Mandarin-Low talker group comparison [$F(1, 6) = 12.3$, $p = .013$, $\eta^2_p = .67$, $\eta^2_G = .52$] and for the Native Mandarin-High vs. Native Mandarin-Low talker group comparison [$F(1, 6) = 11.85$, $p = .01$, $\eta^2_p = .66$, $\eta^2_G = .56$], but not for the Native English vs. Native Mandarin-High talker group comparison [$F(1, 6) = .01$, $p = .01$, $\eta^2_p = .002$, $\eta^2_G = .001$]. Together, these results demonstrated different patterns for overall speaking rate and for plain-clear differences in speaking rate. That is, in terms of the overall speaking rate, Native English talkers spoke faster than Native Mandarin talkers, but Native Mandarin-High talkers did not speak faster than Native Mandarin-Low talkers. However, in terms the plain-clear differences in speaking rate, Native English and Native Mandarin-High talkers made a larger difference between plain- and clear-speaking styles than Native Mandarin-Low talkers did. The Native English and Native Mandarin-High talkers slowed down their speaking rate from the plain to clear speaking style to a similar extent. This suggests that overall speaking rate is partially independent from clear speech

modifications in speaking rate.

In addition to the speaking rate measures, we also examined frequency and duration of silent pauses in plain and clear speech. Table 2.2 shows the total number of silent pauses and average duration (in milliseconds) of a pause for each talker in clear- and plain-speaking styles. Clear-plain differences, shown in the table, were calculated by subtracting the value of the plain-speaking style from the value of the clear-speaking style. As shown the table, Native English talkers rarely paused either in clear- or plain-speaking styles. Between the two groups of non-native talkers, Native Mandarin-Low talkers produced more pauses than Native Mandarin-High talkers in general. Further, all non-native talkers except for one Native Mandarin-Low talker (i.e., talker 107) produced more pauses in the clear-speaking style than in the plain-speaking style. However, duration of a single pause produced by those talkers was not necessarily longer in clear-speaking style than in plain-speaking style.

A two-way within-subjects ANOVA was conducted examining the effect of talker groups (i.e., Native English, Native Mandarin-High, and Native-Mandarin-Low talkers) and speaking style (i.e., clear or plain) on the total number of pauses produced. The results showed a significant effect of talker group [$F(2, 9) = 6.95$, $p = .015$, $\eta^2_p = .61$, $\eta^2_G = .47$]. However, the effect of speaking style was not significant [$F(1, 9) = 3.1$, $p = .11$, $\eta^2_p = .26$, $\eta^2_G = .13$], nor was the interaction between talker group and speaking style [$F(2, 9) = .65$, $p = .54$, $\eta^2_p = .13$, $\eta^2_G = .06$]. A post-hoc Tukey test was conducted examining the significant effect of talker group. The pairwise comparisons revealed that the total number of pauses produced was different between Native English and Native Mandarin-Low talkers ($\beta = -5.38$, SE = 1.44, t ratio = -3.73, $p = .012$), but not between Native English and Native

Table 2.2. Total number of silent pauses and average pause duration (in milliseconds) of a pause for each talker in two speaking styles.

	Total number of pauses			Average duration of a pause (in milliseconds)		
Talker	Clear	Plain	Diff.	Clear	Plain	Diff.
Native Mandarin-Low (NM-L): Average	**7**	**5**	**2**	**472.6**	**458.4**	**14.3**
NM-L 103	12	5	7	300.5	434.4	-133.9
NM-L 104	10	5	5	494.4	416.4	78.0
NM-L 106	3	2	1	328.7	563.5	-234.8
NM-L 107	1	6	-5	767.0	419.2	347.8
Native Mandarin-High (NM-H): Average	**4**	**1**	**3**	**311.7**	**166.6**	**145.1**
NM-H 302	1	0	1	268.0	0.0	268.0
NM-H 306	6	4	2	334.2	403.5	-69.3
NM-H 310	7	0	7	308.6	0.0	308.6
NM-H 311	3	1	2	336.0	263.0	73.0
Native English (NE): Average	**0.25**	**0**	**0.25**	**111.5**	**0.0**	**111.5**
NE 403	0	0	0	0	0	0
NE 404	0	0	0	0	0	0
NE 405	0	0	0	0	0	0
NE 411	1	0	1	446	0	446

Mandarin-High talkers (β = -2.62, SE = 1.44, t ratio = -1.82, p = .22) or between Native Mandarin-High and Naive Mandarin-Low talkers (β = -2.75, SE = 1.44, t ratio = -1.91, p = .19). Another two-way within-subjects ANOVA was conducted examining the effect of talker groups and speaking style on the average duration of a pause for a talker in each speaking style. The results showed a significant effect of talker group [$F(2, 9) = 14.81$, p = .001, $\eta^2_p = .77$, $\eta^2_G = .61$]. However, the effect of speaking style was not significant [$F(1, 9) = 1.99$, $p = .19$, $\eta^2_p = .18$, $\eta^2_G = .1$], nor was the interaction between talker group and speaking style [$F(2, 9) = .38$, $p = .7$, $\eta^2_p = .08$, $\eta^2_G = .04$]. A post-hoc Tukey test, examining each talker group comparisons, showed that the average duration of a pause was different between Native English and Native Mandarin-Low talkers (β = -410, SE = 75.4, t ratio = -5.43, p = .001), and between Native Mandarin-High and Naive Mandarin-Low talkers (β =

226, SE = 75.4, t ratio = -3.0, p = .036), but not between Native English and Native Mandarin-High talkers (β = -183, SE = 75.4, t ratio = -2.43, p = .088). Thus, the total number of pauses and average duration of pauses differed for different talker groups, but not between the two speaking styles.

2.3.1.2. Pitch and intensity

Figure 2.2 shows the scaled values of F0 range, mean F0, and intensity for the sentences produced by Native English, Native Mandarin-High, and Native-Mandarin-Low talkers in plain- and clear-speaking styles. The figure suggests that there was a general trend for clear-style sentences to have wider F0 range, higher mean F0, and higher intensity than plain-style sentences across the three talker groups. However, the difference between the two speaking styles seems smaller for the sentences produced by the Native-Mandarin-Low talkers compared to those produced by Native English and Native Mandarin-High talkers, particularly for the F0 range and mean intensity values. In order to examine the effect of the speaking style (i.e., plain or clear) and whether it differs for different talker groups, three sets of two-way within-subjects ANOVAs were conducted with the scaled values of F0 range, mean F0 and mean intensity as the dependent variable.

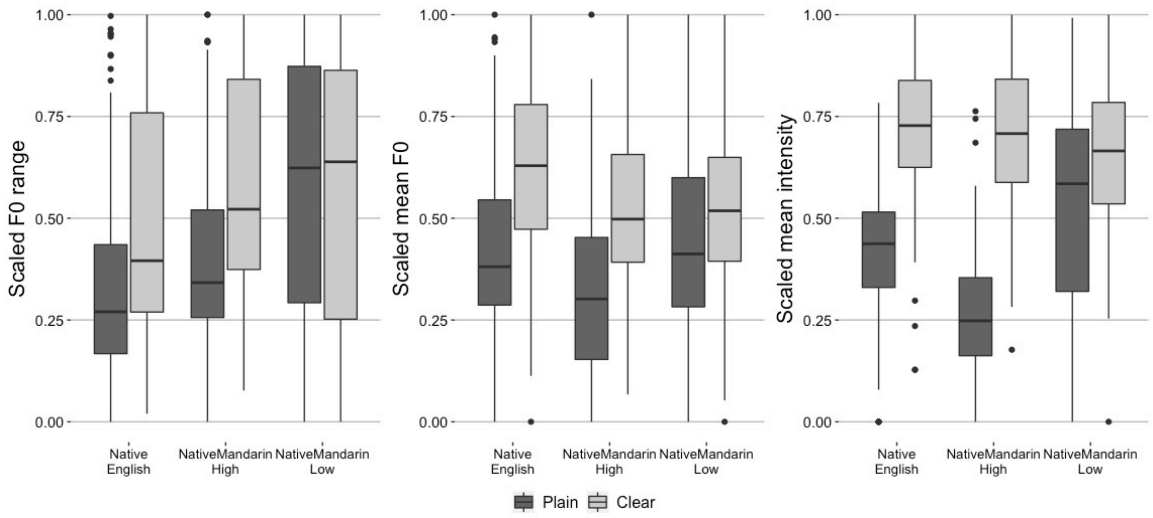

Figure 2.2. Scaled values of F0 range, mean F0, and mean intensity for sentences produced by Native English, Native Mandarin-High and Native Mandarin-Low talkers in plain and clear-speaking styles.

ANOVA results for F0 range showed a significant effect of speaking style [$F(1, 9) = 16.86$, $p = .003$, $\eta^2_p = .65$, $\eta^2_G = .26$] as well as a significant interaction between speaking style and talker group [$F(2, 9) = 5.06$, $p = .034$, $\eta^2_p = .53$, $\eta^2_G = .17$], but not a significant effect of talker group [$F(2, 9) = 2.39$, $p = .15$, $\eta^2_p = .35$, $\eta^2_G = .3$]. Post-hoc Tukey pairwise comparisons examined the effect of speaking style within each talker group; the tests revealed a significant plain-clear difference for the Native English ($\beta = -.16$, SE = .05, t ratio = -3.58, $p = .0059$) and Native Mandarin-High groups ($\beta = -.17$, SE = .05, t ratio = -3.75, $p = .0045$), but not for the Native Mandarin-Low group ($\beta = .01$, SE = .05, t ratio = .23, $p = .83$). In order to further examine the interaction between speaking style and talker group, a two-way within-subject ANOVA was conducted for subsets of the data. These tests revealed that the speaking style x talker group interaction was significant for the Native English vs. Native Mandarin-Low talker group comparison [$F(1, 6) = 8.24$, $p = .028$, $\eta^2_p = .58$, $\eta^2_G = .18$] and for the Native Mandarin-High vs. Native Mandarin-Low talker group comparison [$F(1, 6) = 6.04$, $p = .049$, $\eta^2_p = .5$, $\eta^2_G = .26$], but not for

the Native English vs. Native Mandarin-High talker group comparison [$F(1, 6) = .02$, $p = .898$, $\eta^2_p = .003$, $\eta^2_G < .001$]. Thus, the size of plain-to-clear difference in F0 range was larger for Native English and Native Mandarin-High talkers' speech than for Native Mandarin-Low talkers' speech.

For mean F0, the effect of speaking style was significant [$F(1, 9) = 23.63$, $p < .001$, $\eta^2_p = .72$, $\eta^2_G = .5$], but the effect of talker group was not [$F(2, 9) = 1.71$, $p = .24$, $\eta^2_p = .28$, $\eta^2_G = .19$]. The interaction between speaking style and talker group was not significant [$F(2, 9) = 1.57$, $p = .26$, $\eta^2_p = .26$, $\eta^2_G = .12$]. However, post-hoc Tukey pairwise comparisons showed a significant plain-clear difference for the Native English ($\beta = -.21$, SE = .06, t ratio = -3.58, $p = .006$) and Native Mandarin-High groups ($\beta = -.2$, SE = .06, t ratio = -3.47, $p = .007$), but not for the Native Mandarin-Low group ($\beta = -.08$, SE = .06, t ratio = -1.36, $p = .21$). Two-way within-subject ANOVAs conducted for subsets of the data also showed a significant speaking style x talker group interaction for the Native Mandarin-High vs. Native Mandarin-Low talker group comparison [$F(1, 6) = 6.34$, $p = .045$, $\eta^2_p = .51$, $\eta^2_G = .12$], but not for the Native English vs. Native Mandarin-Low talker group comparison [$F(1, 6) = 1.87$, $p = .22$, $\eta^2_p = .24$, $\eta^2_G = .13$] or for the Native English vs. Native Mandarin-High talker group comparison [$F(1, 6) = .005$, $p = .95$, $\eta^2_p < .001$, $\eta^2_G < .001$]. These results showed that, though there was not a significant difference in the size of plain-clear modifications in mean F0 across different talker groups' speech, there was a tendency that Native English and Native Mandarin-High talkers made a significant plain-clear difference but Native Mandarin-Low talkers did not.

Finally, ANOVA results for mean intensity showed a significant effect of speaking style [$F(1, 9) = 96.04$, $p < .001$, $\eta^2_p = .91$, $\eta^2_G = .79$] as well as a significant interaction

between speaking style and talker group [$F(2, 9) = 10.49$, $p = .004$, $\eta^2_p = .7$, $\eta^2_G = .45$], but not a significant effect of talker group [$F(2, 9) = 3.2$, $p = .09$, $\eta^2_p = .42$, $\eta^2_G = .32$]. Post-hoc Tukey pairwise comparisons revealed a significant plain-clear difference for the Native English ($\beta = -.31$, SE = .06, t ratio = -6.11, $p = .0002$) and Native Mandarin-High groups ($\beta = -.44$, SE = .06, t ratio = -8.65, $p < .0001$), and a marginally significant effect of speaking style for the Native Mandarin-Low group ($\beta = -.11$, SE = .05, t ratio = -2.22, $p = .054$). Two-way within-subject ANOVAs conducted for subsets of the data showed a significant speaking style x talker group interaction for the Native English vs. Native Mandarin-Low talker group comparison [$F(1, 6) = 14.55$, $p = .009$, $\eta^2_p = .71$, $\eta^2_G = .32$] and for the Native Mandarin-High vs. Native Mandarin-Low talker group comparison [$F(1, 6) = 15.04$, $p = .008$, $\eta^2_p = .72$, $\eta^2_G = .49$], but not for the Native English vs. Native Mandarin-High talker group comparison [$F(1, 6) = 2.91$, $p = .14$, $\eta^2_p = .33$, $\eta^2_G = .19$]. Thus, the size of plain-to-clear difference in mean intensity was larger for Native English and Native Mandarin-High talkers' speech than for Native Mandarin-Low talkers' speech.

Together, these results demonstrated that Native English and Native Mandarin-High talkers produced their clear speech with wider F0 range, higher mean F0 and higher intensity compared to their plain speech. The size of plain-clear differences in F0 range, mean F0, and mean intensity was comparable between Native English and Native Mandarin-High talkers' speech. However, the plain-clear differences in these features were much smaller for the Native Mandarin-Low talkers' speech.

2.3.1.3. Vowel space

Figure 2.3 illustrates the vowel space area, covered by the quadrilateral defined by the mean of the 4 point vowels (/i/, /æ/, /ɑ/ and /u/), for each talker in plain- and clear-

speaking styles. The numerical information of the vowel space area is shown in Table 2.3. The figure and table both suggest that the vowel space expansion from plain to clear speech is the largest for the Native English talkers' speech, followed by the Native Mandarin-High talkers' speech. The vowel space expansion is the smallest for the Native Mandarin-Low talkers' speech.

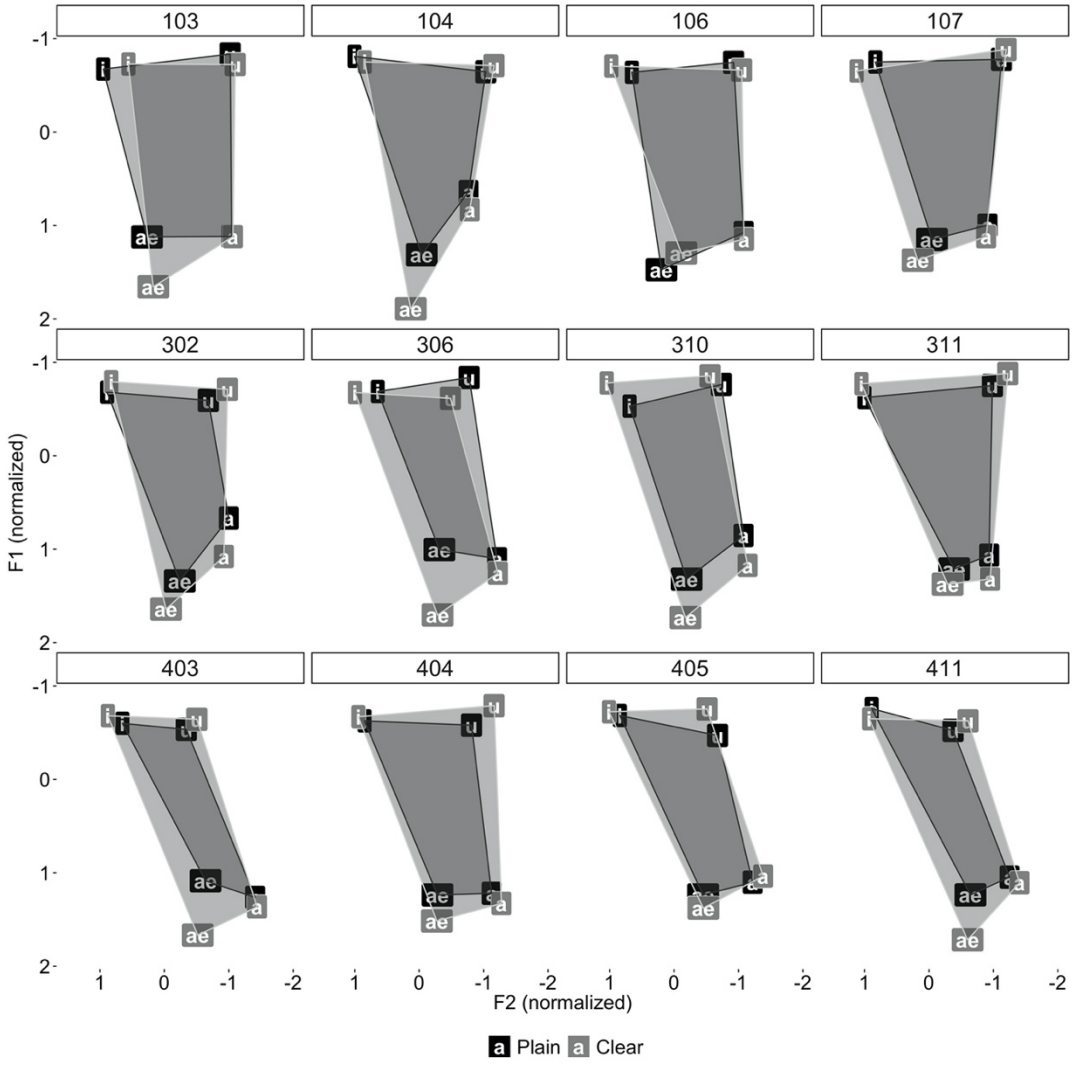

Figure 2.3. Vowel space area measured as the Euclidean area covered by the quadrilateral defined by the mean of the 4 point vowels: /i/, /æ/, /ɑ/ and /u/. Darker area is the vowel space for plain speech and lighter area is the vowel space for clear speech for each talker. The upper four rows (103 - 107) are Native Mandarin-Low talkers, the middle four rows (302 - 311) are Native Mandarin-High talkers, and bottom four rows (403 - 411) are Native English talkers.

Table 2.3. Vowel space area, based on z-score normalized values of midpoint F1 and F2, for each talker in plain- and clear-speaking styles. Talker group averages are also shown in the table. Difference in vowel space area in the two speaking styles was calculated by subtracting the vowel space area for plain-speaking style from the vowel space for clear-speaking style.

Talker	Clear	Plain	Diff.
Native Mandarin-Low (NM-L): Average	**3.152**	**2.768**	**0.385**
NM-L 103	3.103	3.107	-0.003
NM-L 104	3.142	2.455	0.687
NM-L 106	2.883	2.888	-0.005
NM-L 107	3.482	2.622	0.860
Native Mandarin-High (NM-H): Average	**3.035**	**2.216**	**0.820**
NM-H 302	2.945	2.125	0.821
NM-H 306	2.763	2.118	0.645
NM-H 310	3.205	2.216	0.989
NM-H 311	3.229	2.404	0.825
Native English (NE): Average	**2.862**	**1.818**	**1.045**
NE 403	2.662	1.390	1.272
NE 404	3.445	2.281	1.164
NE 405	2.601	1.959	0.642
NE 411	2.740	1.640	1.100

In order to test these observations, a two-way within-subject ANOVA was conducted with the vowel space area (as shown in Table 2.3) as the dependent variable. The effect of speaking style (plain vs. clear), the effect of talker group (Native English, Native Mandarin-High and Native Mandarin-Low), and the interaction between the two on the vowel space area was examined. ANOVA results showed a significant effect of speaking style [$F(1, 9) = 66.68$, $p < .001$, $\eta^2_p = .88$, $\eta^2_G = .69$], talker group [$F(2, 9) = 6.36$, $p = .019$, $\eta^2_p = .59$, $\eta^2_G = .5$], and the interaction between speaking style and talker group [$F(2, 9) = 4.45$, $p = .045$, $\eta^2_p = .5$, $\eta^2_G = .23$]. The post-hoc Tukey comparisons confirmed that the effect of speaking style was significant in all the talker groups' speech: Native English ($\beta = 1.05$, SE = .16, t ratio = 6.57, $p = .0001$), Native Mandarin-High ($\beta = .82$, SE = .16, t ratio = 5.16, $p = .0006$), Native Mandarin-Low ($\beta = .39$, SE = .16, t ratio = 2.42, $p = .039$).

In order to further examine the interaction between speaking style and talker group (as illustrated in Figure 2.4), a two-way within-subject ANOVA was conducted for subsets of the data: for Native English and Native Mandarin-High talkers' data, for Native English and Native Mandarin-Low talkers' data, and for Native Mandarin-High and Native Mandarin-Low talkers' data. These tests revealed that the speaking style x talker group interaction was significant for the Native English vs. Native Mandarin-Low talker group comparison [$F(1, 6) = 6.14$, $p = .048$, $\eta^2_p = .51$, $\eta^2_G = .25$], but not for the Native English vs. Native Mandarin-High talker group comparison [$F(1, 6) = 2.09$, $p = .199$, $\eta^2_p = .26$, $\eta^2_G = .04$] or for the Native Mandarin-High vs. Native Mandarin-Low talker group comparison [$F(1, 6) = 3.35$, $p = .117$, $\eta^2_p = .36$, $\eta^2_G = .23$].

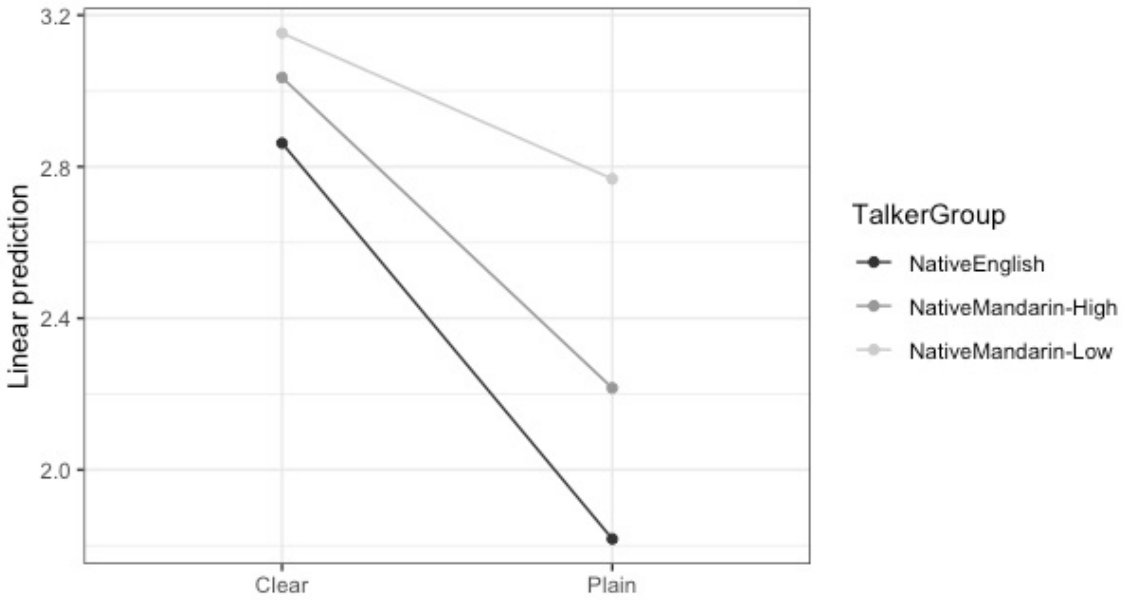

Figure 2.4. Linear prediction for vowel space area for the three talker groups (Native English, Native Mandarin-High, and Native Mandarin-Low) in plain- and clear-speaking styles.

These results demonstrated that the size of vowel space expansion from plain to clear speech differed for Native English, Native Mandarin-High, and Native Mandarin-Low talkers' speech. Specifically, Native English talkers expanded their vowel space the

most, followed by Native Mandarin-High and Native Mandarin-Low talkers. The vowel space area in plain and clear speech further suggested that vowel space expansion is the smallest for Native Mandarin-Low talkers' speech partly because their vowel space for plain speech was already large (i.e., Native Mandarin-Low talkers were using the vowel space that is close to their maximum in plain speech; thus there was not much room to expand in clear speech).

2.3.2. Talkers' perceived foreign-accentedness

Foreign accent ratings were z-score normalized for each listener in order to account for variation in the listeners' use of the nine-point rating scale. Figure 2.5 shows accent ratings by talker group (Native English, Native Mandarin-High, and Native Mandarin-Low). In order to examine whether the accent ratings differed for different taker groups, one-way ANOVA was carried out with z-scored ratings as the dependent variable. The results indicated that the ratings differed significantly by the talker group [$F(2, 9) = 128.29$, $p < .001$, $\eta^2_p = .97$, $\eta^2_G = .97$]. The post-hoc Tukey comparisons confirmed that the all the group comparisons were significant: Native English vs. Native Mandarin-High, Native English vs. Native Mandarin-Low, and Native Mandarin-High vs. Native Mandarin-Low ($p < .0001$ for all). These results demonstrated that there was a clear difference in the perceived accentedness of the talkers. That is, Native English talkers were perceived to be less accented than Native Mandarin talkers, and Native Mandarin-High talkers were perceived to be less accented than Native Mandarin-Low talkers.

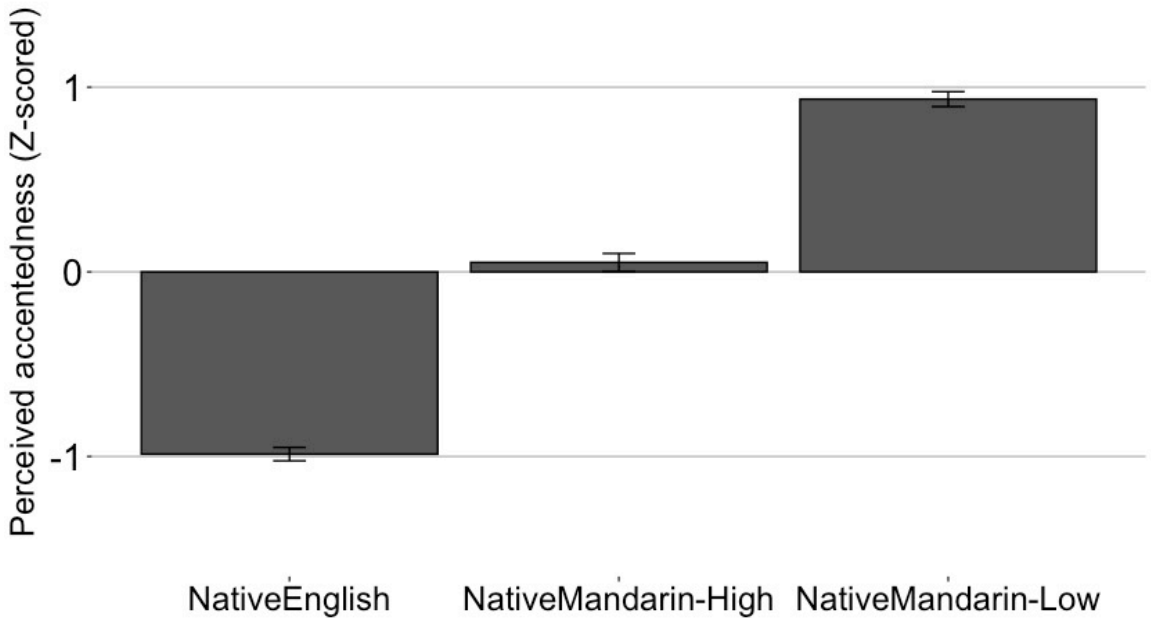

Figure 2.5. Z score-normalized accentedness ratings plotted by talker group. Error bars represent 95% confidence interval of the mean.

2.4. Discussion & conclusion

2.4.1. Summary of the findings

In the present study, we explored acoustic characteristics of clear speech enhancements produced by native English talkers and non-native English (native Mandarin) talkers of higher- and lower-proficiency. Specifically, we examined whether non-native talkers of differing target language proficiency levels employ acoustic-phonetic modifications that are similar or different from those made by native talkers. The difference in proficiency levels between the higher- and lower-proficiency talkers was confirmed by the differences in their English learning background (e.g., TOEFL score) as well as in their perceived accentedness (evaluated by native English listeners).

The native and non-native talkers read English sentences in plain- and clear-speaking styles. We examine acoustic-phonetic modifications from plain to clear speech in

several features: temporal characteristics (i.e., speaking rate, pause frequency, pause duration), fundamental frequency (F0), intensity, and vowel space. Overall, the results demonstrated that talkers generally decreased their speaking rate, increased F0 range, mean F0, mean intensity, and vowel space in clear speech compared to plain speech. However, there were differences in the degrees to which the talkers of different native language backgrounds/proficiency levels modified these acoustic features. That is, the native English talkers and higher-proficiency non-native talkers modified the above-mentioned acoustic features to larger degrees than lower-proficiency non-native talkers did. These results suggest that non-native talkers' clear speech strategies change as their target language proficiency develops; higher-proficiency talkers modify acoustic features to larger degrees than lower-proficiency talkers, and higher-proficiency talkers' strategies are comparable to native talkers' strategies.

2.4.2. The influence of target language proficiency level on non-native clear speech enhancements

A series of comparisons made in the present study, in terms of speaking styles (plain and clear) and talker groups (native English talkers and non-native talkers of different proficiency levels), revealed differences in the characteristics of native and non-native speech in general, as well as in their clear speech modifications. For example, in terms of the difference between native and non-native speech, non-native talkers (higher- and lower-proficiency) spoke more slowly (i.e., slower raw speaking rate and articulation rate) than native English talkers in both plain-and clear-speaking styles. The overall speaking rate was similar between higher- and lower-proficiency talkers' speech. However,

there was a clear difference between the plain-to-clear speech modifications made by higher- and lower-proficiency non-native talkers. That is, higher-proficiency talkers decreased their speaking rate in clear speech to a larger degree than lower-proficiency talkers did, suggesting that non-native talkers' overall speaking rate is partially independent from their ability to make plain-to-clear speech modifications in speaking rate. Furthermore, higher-proficiency talkers made larger plain-to-clear speech modifications in terms of F0 range, mean F0, and mean intensity than lower-proficiency talkers did. Though there was not a statistically significant difference between the degrees of higher- and lower-proficiency talkers' vowel space expansion, higher-proficiency talkers showed a numerically larger vowel space expansion than lower-proficiency talkers did. These findings suggest that non-native talkers' ability to make plain-to-clear speech modifications is influenced by their target-language proficiency level. Furthermore, the present findings demonstrated that higher-proficiency talkers' clear speech strategies, in terms of acoustic characteristics of the plain-to-clear speech modifications, are comparable to those of native English talkers. This is in line with previous studies that have compared clear speech enhancements of native and highly proficient non-native talkers in different research contexts/topics (e.g., examining acoustic characteristics of spontaneous speech: Granlund et al., 2012; examining clear speech intelligibility improvement of read speech: Smiljanić & Bradlow, 2011). Thus, the current results provide further support for the claim that highly proficient non-native talkers' ability to make clear speech modifications approximates that of native talkers.

The lower-proficiency talkers, however, made much smaller plain-to-clear speech modifications than higher-proficiency talkers and native English talkers did. Across

the different measures of acoustic characteristics, lower-proficiency talkers made significant plain-to-clear speech differences in articulation rate (i.e., rate of speech without pauses), mean intensity, and vowel space, but not in F0 range and mean F0. These results suggest that compared to higher-proficiency non-native talkers and native talkers, lower-proficiency non-native talkers were less able to vary acoustic characteristics of their speech along the hypo- and hyperspeech continuum (Lindblom, 1990). There may be several possible explanations to why lower-proficiency talkers have difficulty varying acoustic characteristics of their speech between plain and clear speech. One possibility is that they have difficulty moving from hypospeech (operationalized as plain speech in the current study) to hyperspeech (operationalized as clear speech). That is, while talkers may be able to minimize their articulatory effort in producing speech (hypospeech), they may have difficulty modifying phonetic characteristics to improve intelligibility for listeners (hyperspeech). This is illustrated in the plain-clear variations of the articulation rate for lower-proficiency talkers (right panel in Figure 2.1), where their clear-speech articulation rate did not slow down (i.e., did not move to the slower side within their range of articulation rate), as compared to native English and higher-proficiency non-native talkers.

However, the small range of plain-to-clear speech variations in lower-proficiency talkers' speech could also stem from their difficulty minimizing articulatory effort (hypospeech or plain speech in the current study). Though plain speech, in the current study and previous studies, is elicited by asking talkers to read materials in a laboratory setting (Smiljanić & Bradlow, 2009) and does not typically contain phonetic reductions that speech occurring in a more natural setting does, it is often produced with reduced effort compared to clear speech, as seen in the plain-to-clear speech variations in previous studies

(e.g., Bradlow et al., 2003; Picheny et al., 1986). However, it is possible that lower-proficiency talkers were not necessarily speaking with reduced effort in plain speech, resulting in small plain-to-clear speech variations. In other words, lower-proficiency talkers may already have been exerting substantial effort to produce plain speech (and they may generally hyperarticulate to produce second language speech), thus there was not much room to 'enhance' in clear speech. This is illustrated in plain-to-clear speech variations in mean intensity as well as in vowel space expansion. Specifically, lower-proficiency talkers' mean intensity in plain speech was on the higher side of their range of mean intensity (right panel in Figure 2.2), which is not much different from mean intensity in their clear speech. Further, Figure 2.4 illustrates that the small plain-to-clear variation in lower-proficiency talkers' vowel space originated from the relatively large vowel space in their plain speech; on the other hand, the large plain-to-clear variation in native English talkers' vowel space originated from relatively the small vowel space in their plain speech. These results suggest that lower-proficiency talkers did not necessarily produce the plain speech with reduced effort compared to their clear speech. Together, the current results suggest that the size of plain-to-clear speech modifications could be influenced by the talker's ability to enhance acoustic characteristics in clear speech as well as to reduce articulatory effort in plain speech. This could be compatible with the previous work suggesting that non-native speech production is more effortful for talkers with lower proficiency (e.g., Kormos & Dénes, 2004; Nip & Blumenfeld, 2015; Poulisse, 1997). In conjunction with the current results, it is possible that as non-native talkers' proficiency level develops, their speech production generally becomes less effortful (e.g., in plain speech), and ultimately there is more room for talkers to enhance characteristics of their speech. Furthermore, the influence of non-

native talkers' proficiency level on the size of plain-to-clear speech modifications suggests that knowing how to vary acoustic characteristics of speech (e.g., in plain and clear speech) may be a part of the skill set that contributes to the target language proficiency.

2.4.3. Individual variability within talker groups

Though the present study mainly examined whether the acoustic characteristics of clear speech enhancements differed for different talker groups (i.e., native English, higher- and lower-proficiency non-native talkers), it also revealed some individual differences within talker groups. For example, in terms of vowel space expansion among the lower-proficiency non-native talkers, Talkers 103 and 106 showed either no change or a slight decrease in vowel space from plain to clear speech; whereas Talkers 104 and 107 showed a relatively large expansion (see Table 2.3). In fact, Talker 107's vowel space expansion was larger than that of three higher-proficiency non-native talkers and one native talker. However, Talker 107 did not necessarily make the largest plain-to-clear differences among the lower-proficiency talkers in other acoustic features (e.g., articulation rate), suggesting that the patterns of individual differences may vary depending on the acoustic measures examined. Further, there were individual differences among native English talkers. Particularly, Talker 405's plain-to-clear differences in vowel space and mean F0 were the smallest among the native English talkers *and* higher-proficiency non-native talkers. These results demonstrate that individual differences in the degrees of clear speech modifications are present in both native and non-native talkers' speech. This is in line with previous results demonstrating that the degree of plain-to-clear speech vowel space expansion differs significantly even among native talkers (Ferguson & Kewley-Port, 2007).

However, it is an open question whether the source of individual differences in clear speech enhancements is similar for native talkers and non-native talkers of different proficiency levels. That is, for native talkers and non-native talkers of higher-proficiency, the degrees of acoustic modifications could be influenced by how hard they try to increase intelligibility of their speech. For example, talkers who try less hard to increase intelligibility may make smaller amount of acoustic modifications than other talkers across different acoustic measures (e.g., speaking rate, intensity, vowel space, etc.). However, for non-native talkers of lower-proficiency, the degrees of acoustic modifications may also be influenced by their ability to control their production patterns, including knowing which acoustic features to modify and how to modify these features. For those lower-proficiency talkers, a single talker's ability to make clear speech modifications may differ across different acoustic features. Thus, for example, the talkers who make the smallest amount of plain-to-clear speech modifications in one acoustic measure (e.g., vowel space) may make larger modifications than other talkers in other acoustic features (e.g., speaking rate). These questions regarding how talkers' target language proficiency may relate to individual differences in clear speech enhancements need to be investigated further in future research.

2.4.4. Conclusion

The goal of the current study was to characterize acoustic features of clear speech enhancements produced by native English stalkers and non-native English talkers of different proficiency levels. Specifically, we examined whether non-native talkers of different proficiency levels employ similar strategies to enhance acoustic-phonetic features of their speech as native talkers. The results demonstrated that the talkers generally

decreased their speaking rate, increased F0 range, mean F0, mean intensity, and vowel space in clear-speaking style than plain-speaking style of the same sentences. However, the degrees of plain-to-clear speech modifications were much smaller for lower-proficiency non-native talkers' speech compared to those of higher-proficiency talkers' and native talkers' speech. The higher-proficiency talkers' acoustic modifications were comparable to those of native talkers. These results suggest that second language learners' clear speech strategies become similar to those of native talkers as their target language proficiency develops.

CHAPTER III: PERCEPTION OF CLEAR SPEECH

3.1. Introduction

One of the most common communication difficulties that talkers face when communicating in their non-native language is that their listeners do not understand their speech. Previous research suggests that native talkers' effort to enhance acoustic-phonetic properties of their speech (i.e., clear speech enhancements) results in robust increases in understanding for listeners of various characteristics, including listeners with hearing impairments and non-native listeners (e.g., Bradlow & Alexander, 2007; Picheny et al., 1985, 1986; Schum, 1996). However, much less is known about how non-native talkers' clear speech enhancements are perceived by native listeners. Particularly, it is not clear how non-native talkers' ability to increase intelligibility of their speech improves as their second language (L2) proficiency develops. Furthermore, previous studies demonstrate that listeners' speech perception is more diverse than how well they correctly recognize the words and phrases from the speech (i.e., intelligibility of the speech), suggesting that the intelligibility measure of the speech may not fully represent perceptual benefits of clear speech enhancements (e.g., Hazan et al., 2012). Thus, in order to better understand perceptual consequences of clear speech enhancements, the present study examines multiple aspects of perceptual benefits resulting from native and non-native talkers' clear speech enhancements, with a focus on how native language background and talkers' L2 proficiency influences the way their clear speech enhancements benefit native listeners' perception.

3.1.1. Intelligibility benefits of clear speech enhancements

Perceptual consequences of clear speech enhancements have typically been measured using an intelligibility task, where listeners hear speech materials produced in plain- and clear-speaking styles with noise and transcribe what they hear (e.g., Bradlow & Bent, 2002). Previous studies have found robust intelligibility gains resulting from native English talkers' clear speech enhancements for native English listeners of various characteristics, including hearing-impaired listeners (e.g., Ferguson, 2004; Krause & Braida, 2002; Liu et al., 2004; Picheny et al., 1985; Schum, 1996; Smiljanić & Bradlow, 2005; Uchanski et al., 1996). Furthermore, native English talkers' clear speech enhancements result in intelligibility gains for non-native English listeners (e.g., Bradlow & Bent, 2002; Bradlow & Alexander, 2007). A similar clear speech intelligibility benefit has also been reported for languages other than English, including Croatian and French (Gagné et al., 1994, 2002; Smiljanić & Bradlow, 2005).

While clear speech enhancements made in talkers' native languages have been shown to result in reliable intelligibility gains for a variety of listeners, much less is known about how clear speech enhancements made in a non-native language are perceived by native listeners. Specifically, existing literature regarding intelligibility gains resulting from non-native talkers' clear speech enhancements are mostly limited to those produced by highly proficient non-native talkers. For example, speech enhancements made by highly proficient non-native talkers are as effective as those made by native talkers. This is shown in the comparable size of intelligibility gains resulting from clear speech enhancements made by native talkers and by highly proficient or early learners (Rogers et al., 2010; Smiljanić & Bradlow, 2005, 2011), as well as in the types of acoustic modifications made

by native talkers and by proficient non-native talkers (Granlund et al., 2012). However, data from talkers of lower proficiency are relatively scarce. One study demonstrated that late learners of English were much less effective at enhancing intelligibility of English vowels than early learners and native English talkers (Rogers et al., 2010). Specifically, clear speech enhancements of English vowels in /bVd/ syllables produced by monolingual native English talkers and early native Spanish learners of English resulted in a similar size of intelligibility gains, whereas those produced by late native Spanish learners resulted in much smaller intelligibility gains. The late learners' clear speech enhancements sometimes resulted in a decrease in intelligibility.

These studies suggest that non-native talkers' target language proficiency may affect their ability to increase intelligibility. That is, the more familiar the talkers are with the sound structure of the language, including the system of phonological contrasts and phonetic implementation of those contrasts, the more effective their clear speech enhancements may be at increasing intelligibility for native listeners (Smiljanić & Bradlow, 2011). However, with the data from existing literature, it is difficult to determine the effect of target language proficiency on clear speech enhancements beyond the level of single sound production. That is, while the more experienced L2 learners are better able to increase the intelligibility of English vowels than less experienced L2 learners (Rogers et al., 2010), it is not clear whether the effect of target language experience generalizes to clear speech intelligibility benefits for sentence production. Increasing intelligibility of sentences can be more challenging than doing so for single words because it requires proficient use of the target language sound system at multiple levels, including single sounds or words, in addition to other features such as prosody (e.g., phrasing and

prominence structure at the sentence level; see Ladd, 2008 for examples). Thus, the effect of non-native talkers' proficiency level on clear speech intelligibility benefits could manifest differently at the sentence level compared to the single-word level.

Furthermore, examining the intelligibility of the speech produced in plain- and clear-speaking styles by non-native talkers of different proficiency levels may help us better understand the relationship between talkers' ability to produce intelligible speech in general vs. their ability to *increase* intelligibility of their speech. Specifically, given that producing speech in a non-native language becomes more fluent and less effortful as the talkers' proficiency develops (e.g., Kormos & Dénes, 2004; Nip & Blumenfeld, 2015), it is possible that higher-proficiency non-native talkers' speech (their speech in plain- and clear-speaking styles) is generally more intelligible than lower-proficiency talkers' speech. However, the ability to further *increase* intelligibility (i.e., intelligibility improvement from plain to clear speech) may or may not differ between non-native talkers of differing proficiency levels. That is, if higher-proficiency talkers are able to increase intelligibility of their speech to a larger extent than lower-proficiency talkers, this would suggest that non-native talkers' increased proficiency is associated with their ability to not only produce generally more intelligible speech but also with their ability to further *increase* intelligibility of their speech. However, if the size of intelligibility improvement is similar between the speech of higher- and lower-proficiency talkers, it may suggest that the ability to produce generally intelligible speech and the ability to increase intelligibility are at least partially independent from one another. In order to answer these questions, we examine clear speech intelligibility benefits of English sentences produced by native English talkers and non-native talkers of higher- and lower-proficiency.

3.1.2. Other perceptual benefits of clear speech enhancements

While examining the intelligibility measure is one way to analyze how listeners perceive speech, speech perception literature suggests that listeners' perception of speech is much more diverse than the correct recognition of words or phrases from that speech. For example, listeners make various social evaluations about a talker based on their speech, including social attractiveness, power, and competence (Adank et al., 2013; Bayard, Weatherall, Gallois, & Pittam, 2001; Coupland & Bishop, 2007; Grondelaers, Van Hout, & Steegs, 2010). Further, listeners' perception of a particular talker affects their behavior, including imitation of that talker's speech (Babel, 2012). Perception of non-native speech has also been examined using measures other than intelligibility, including comprehensibility (i.e., how easy or difficult listeners perceive the speech is to understand), accentedness (i.e., the degree of foreign accent of the speech that listeners perceive), and credibility (i.e., how credible listeners perceive the information conveyed in non-native speech to be; Munro & Derwing, 1995a, 1999; Lev-Ari & Keysar, 2010; Smiljanić & Bradlow, 2011). These evaluations are often collected by asking listeners to rate the speech on a Likert scale, such as a comprehensibility scale from 1 ("extremely easy to understand") to 9 ("impossible to understand"; Munro & Derwing, 1999). These studies demonstrate that there can be a gap between what listeners actually understand from the speech (i.e., intelligibility) and how they perceive the speech in more subjective terms. For example, listeners can understand non-native speech even if they perceive the same speech to be heavily accented (Munro & Derwing, 1999; Smiljanić & Bradlow, 2011) or not easy to understand (Sheppard et al., 2017). Thus, these studies suggest that examining different

aspects of listeners' perception, in addition to intelligibility, may provide more insight into how listeners process speech, including listeners' perception of native and non-native talkers' clear speech enhancements.

In addition to this broad set of literature regarding perception of accented speech, studies examining clear speech have also suggested that perception of clear speech enhancements is multifaceted. For example, Smiljanić and Bradlow (2011) have examined intelligibility and perceived degree of foreign accent for clear and plain speech produced by non-native (native Croatian) talkers of English. The highly proficient non-native talkers' clear speech was more intelligible than plain speech, but perceived degree of foreign accent was similar between the two styles of the speech for native English listeners, revealing a partial independence of these two perceptual measures from one another.
Furthermore, Hazan and Baker (2011) and Hazan et al. (2012) measured native English listeners' clarity ratings of native English talkers' spontaneous speech, on the scale of 1 (very clear) to 7 (unclear). The listeners rated the speech produced for a listener in an adverse listening condition (i.e., with noise) to be clearer than the speech produced for a listener in an easy listening condition (i.e., without noise; Hazan et al., 2012). These studies suggest that clear speech enhancements may be reflected in listeners' perception differently depending on how listeners evaluate the speech. In other words, there may be aspects of perceptual consequences that measures of intelligibility alone may not fully capture. This may especially be the case for speech enhancements produced by non-native talkers; their acoustic modifications may not necessarily result in intelligibility improvement (e.g., late learners in Rogers et al., 2010), but may improve other aspects of perception, such as perceived degree of clarity (how clear the speech is) or comprehensibility (how easy the

speech is to understand). In the present study, we examine multiple aspects of clear speech perception to better understand how clear speech enhancements produced by native and non-native talkers are perceived by native listeners.

Taken together, previous studies demonstrate that clear speech enhancements produced by native talkers and highly proficient non-native talkers result in a significant intelligibility benefit for native listeners. However, there is little data for lower-proficiency talkers' clear speech intelligibility benefit, making it difficult to determine how talkers' L2 proficiency level affects their ability to increase intelligibility of L2 sentences. In order to better understand the effect of talkers' L2 proficiency on a clear speech intelligibility benefit, it is critical to examine the intelligibility improvement resulting from clear speech enhancements produced by non-native talkers of various proficiency levels. Furthermore, given that listeners' perception of speech is more diverse than how correctly they recognize the words spoken (e.g., Hazan et al., 2012; Munro & Derwing, 1995a; 1999), intelligibility measures alone may not fully capture perceptual benefits of clear speech enhancements. Examining multiple aspects of clear speech perception may allow us to broaden our understanding of how talkers' attempt to increase their speech intelligibility is perceived by listeners.

3.1.3. Current study

In the current study, we examine three aspects of perceptual benefits of clear speech enhancements produced by native and non-native talkers of English. Experiment 2A examines intelligibility benefits of clear speech enhancements. Specifically, we ask whether native and non-native talkers' clear speech enhancements result in a similar size of

intelligibility improvement for native English listeners. Further, we ask whether clear speech produced by higher-proficiency non-native talkers results in a larger intelligibility improvement compared to that produced by lower-proficiency non-native talkers. Given that non-native talkers' phonological representations of non-native sounds may not be the same as those of non-native talkers (Imai, Walley, & Flege, 2005), non-native talkers may emphasize different acoustic cues than native talkers to enhance intelligibility of non-native sounds. Thus, clear speech modifications made by non-native talkers (including higher- and lower-proficiency talkers) may be much smaller than those made by native talkers for native listeners. However, given previous results demonstrating that clear speech enhancements of English vowels produced by early L2 learners resulted in larger intelligibility gains than those produced by late L2 learners (Rogers et al., 2010), it is also possible that higher-proficiency talkers' intelligibility improvement of sentences approximates that of native talkers, and is much larger than that of lower-proficiency talkers.

In Experiment 2B, we examine other perception aspects of native and non-native clear speech enhancements, namely, perceived degree of comprehensibility (how easy the speech is to understand) and talker effort (how hard the talker is trying to speak clearly). Comprehensibility is one of the major perceptual measures that have been examined in L2 pronunciation research (e.g., Munro & Derwing, 1999; Isaacs & Trofimovich, 2012). Previous studies have claimed comprehensibility to be consistent with the goal of achieving intelligible pronunciation and a primary component of communicative success (Derwing & Munro, 2009). Because the primary goal of clear speech enhancements is to make the speech more intelligible for listeners, it is possible that clear speech enhancements may

impact listeners' perception of comprehensibility as well. That is, talkers' clear speech may be perceived to be easier to understand than plain speech. Another perception measure that we examine is perceived degree of talker effort. We measure this aspect of listener perception because it is possible that talkers' attempt to speak more clearly may not result in ease of understanding, but listeners could still perceive talkers' increased effort in clear speech compared to plain speech. This may especially be the case for native listeners' perception of non-native talkers' clear speech. That is, native listeners may not necessarily perceive non-native talkers' clear speech to be easier to understand than their plain speech (perceived degree of comprehensibility), though native listeners may still perceive the talkers' attempt to speak more clearly (perceived degree of talker effort). Thus, Experiment 2B examines two subjective measures of clear speech perception: whether listeners perceive clear speech to be easier to understand (comprehensibility) and whether listeners perceive clear speech to be spoken with increased effort (talker effort), as compared to plain speech. Further, we compare listeners' responses to these two tasks to examine whether one type of subjective perception is more robustly improved by clear speech enhancements than the other. Finally, we ask whether native listeners' responses to these different tasks (comprehensibility vs. talker effort) pattern similarly for the speech produced by native talkers and non-native talkers of different proficiency levels.

3.2. Experiment 2A

In this experiment, we examine intelligibility benefits of clear speech enhancements produced by native English talkers and non-native English talkers of different proficiency levels. Specifically, we ask whether the clear speech enhancements made by these talkers

result in a similar size of intelligibility improvement for native English listeners. In the perception experiment, native English listeners transcribed English sentences produced by native and non-native English talkers in plain- and clear-speaking styles. We used the native and non-native English sentences analyzed in Chapter 2 as materials for this perception experiment; the information about the talkers and acoustic properties of the materials in the current perception experiment are available in Chapter 2.

Though clear speech intelligibility experiments typically involve having participants listen to speech with noise in order to prevent ceiling performance in transcription (e.g., Bradlow & Alexander, 2007; Bradlow & Bent, 2002, Rogers et al., 2010; Smiljanić & Bradlow, 2011), it can be difficult to determine whether the noise is actually preventing the ceiling performance without knowing the level of listeners' best performance for transcribing different types of speech. That is, the level of listeners' best transcription performance could differ when listening to native talkers and non-native talkers of different proficiency levels (e.g., listeners' best performance for lower-proficiency talkers' speech might be lower than that for higher-proficiency talkers' speech: Rogers, Dalby, & Nishi, 2004). This may make it difficult to determine whether a certain level of noise is limiting the transcription performance to a similar extent for different talkers' speech. Thus, we included a quiet listening condition in order to assess native listeners' best performance for transcribing native and non-native talkers' speech. This helps ensure that any clear speech intelligibility improvement, when listeners are transcribing speech with noise, is not limited by their best transcription performance.

3.2.1. Methods

3.2.1.1. Participants

Participants were 194 native English listeners (94 females, 99 males, 1 declined to provide a gender; age range = 22 - 63 years, mean = 35.9). Participants were recruited using Amazon Mechanical Turk (www.mturk.com). None of the listeners reported a history of speech or hearing impairment. All participants resided in the United States, and self-reported to be native speakers of American English. None of the participants reported experience with Mandarin Chinese.

3.2.1.2. Materials

Materials were the native and non-native speech analyzed in Chapter 2. Specifically, materials consisted of 30 BKB sentences (Bamford & Wilson, 1979) produced by 4 native English (Native English) talkers, 4 higher-proficiency native Mandarin (Native Mandarin-High) talkers, and 4 lower-proficiency native Mandarin (Native Mandarin-Low) talkers in plain- and clear-speaking styles. All speech files were segmented into individual sentence-length files which were then RMS normalized to 65 dB. Silence of 500 ms was then added at the beginning and end of each sound file. Further, in order to create materials for speech-in-noise intelligibility task, we mixed each file with speech-shaped noise (Bradlow & Alexander, 2007) at a signal-to-noise ratio (SNR) of -6 dB for native talkers' items and -2 dB for non-native talkers' items. These SNRs were determined based on a series of pilot testing, where we examined the noise level that would have native and non-native talkers' plain speech intelligibility to be within the range of 45-65% correct (Smiljanić & Bradlow, 2011), so that we could assess the amount of clear speech benefit from a similar baseline level (i.e., plain speech) of recognition accuracy.

Thus, for the intelligibility task in the noise condition, each stimulus file consisted of 500 ms header of noise, followed by the speech-plus-noise portion, and ending with a 500 ms noise-only tail (Bradlow & Alexander, 2007). The noise in the 500 ms header and tail was always at the same level as the noise in the speech-plus-noise portion of the stimulus file. For the intelligibility task in the quiet condition, each stimulus file consisted of 500 ms of silence, followed by speech in quiet, and ending with a 500 ms of silence.

3.2.1.3. Procedure

The experiment was conducted online with Qualtrics (https://www.qualtrics.com/). The participants followed the link posted on the Mechanical Turk to complete the task on Qualtrics. They were told that they would listen to English sentences and transcribe them. They were also instructed to use headphones to complete the task. The experiment began with a consent procedure as well as a sound check to ensure that participants could listen to the audio files at their comfortable volume. After that, participants read the instructions; in each trial, they were asked to listen to an English sentence and type what they heard. They could listen to the sentence only once, but could take as much time as needed to type their answer. They also completed two practice trials with the talkers and sentences that were different from the following 30 test sentences.

During the test trials, each participant listened to 30 unique sentences produced by 6 talkers (i.e., 5 sentences from each of the 2 Native English, 2 Native Mandarin-High, 2 Native Mandarin-Low talkers). They heard half of the sentences (15 sentences) in a plain-speaking style and half in a clear-speaking style. Each participant heard the same number of clear- and plain-style sentences from the three talker groups. That is, they heard 5 clear-

and 5 plain-style sentences produced by Native English talkers, 5 clear- and 5 plain-style sentences produced by Native Mandarin-High talkers, and 5 clear- and 5 plain-style sentences produced by Native Mandarin-Low talkers. The combination of the talker, sentence, and style was counter-balanced for each participant. The presentation order of the sentences was randomized for each participant. After the experimental trials were completed, each participant completed a post-test demographic survey.

3.2.1.4. Analysis

The intelligibility data were analyzed for proportion of keywords that was correctly recognized. Keywords were defined to be content words (e.g., *the shop closed for lunch;* see also Appendix A) and there was a total of 94 keywords scored per participant. Words correct were defined as those that matched the intended target exactly, as well as homophones and/or common misspellings (e.g., *to* for *too* in the sentence *The car is going too fast*). However, words with incorrect, added or deleted morphemes were scored as incorrect (e.g., *ties* for *tied* in the sentence *The man tied his shoes*, or *shoe* for *shoes* in the same sentence*).* In terms of the number of the data points, there were 97 participants who listened to one set of 6 talkers (50 participants in the quiet condition and 47 participants in the noise condition) and 97 participants who listened to the other set of 6 talkers (48 participants in the quiet condition and 49 participants in the noise condition). Thus, there were 18236 data points analyzed (194 participants x 94 keywords). The first author and a research assistant scored these data; both raters scored all of the data. When there was a disagreement (16 instances out of 18236 instances), the two raters discussed discrepancies until they reached agreement.

3.2.2. Results

Figure 3.1 shows the proportion correct of keyword recognition in two listening conditions (noise and quiet) for the speech produced by different talker groups (Native English, Native Mandarin-High, and Native Mandarin-Low) in two speaking styles (plain and clear). The figure shows that listeners recognized keywords correctly more often when listening to the speech in quiet than with noise in general. In order to confirm this, we analyzed the data via logistic mixed-effects regression models using R package lme4 (Bates, Mächler, Bolker, & Walker., 2015). The dependent variable was keyword correct, scored as a 0 for incorrect and 1 for correct for each keyword in the sentence. As fixed effects, Condition (Noise or Quiet), Talker Group (Native English, Native Mandarin-High, Native Mandarin-Low), and interaction between the two were included. Condition was contrast-coded to compare between Noise (-.5) and Quiet (.5) conditions. Talker Group was also contrast-coded to compare between Native English and Native Mandarin (Native Mandarin-High and Native Mandarin-Low) group (.5, -.25, -.25: TalkerGroup1), and between Native Mandarin-High and Native Mandarin-Low group (0, .5, -.5: TalkerGroup2). The maximal random effects structure that would converge was implemented, which included random intercepts for talker, listener, and item. The random effects structure also included by-talker random slope for Condition, by-listener slopes for Talker Group, and by-item slopes for Condition, Talker Group and their interaction. See Table 3.1 for the model syntax.

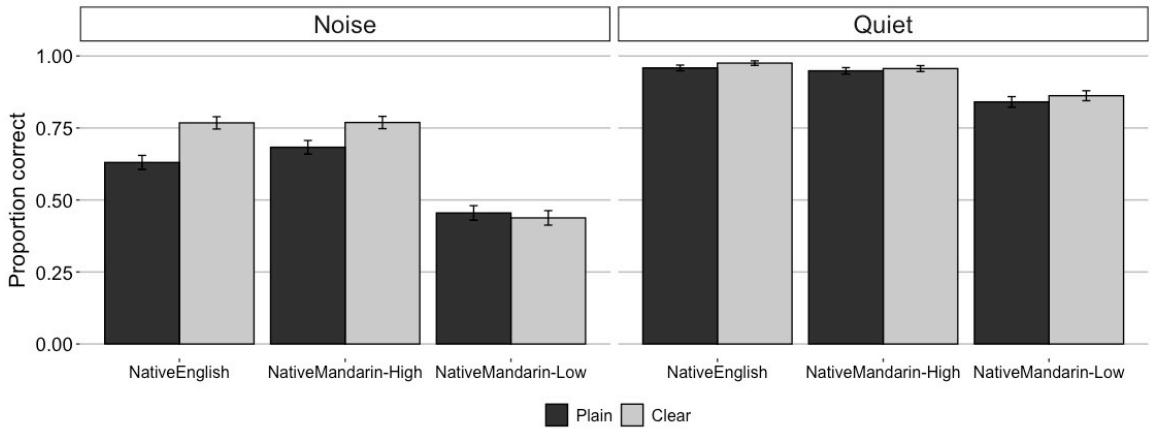

Figure 3.1. Proportion correct of keyword recognition in two listening conditions (Noise and Quiet) for each talker group (Native English, Native Mandarin-High, and Native Mandarin-Low) by speaking style (plain and clear). The error bars represent 95% confidence interval of the mean.

The results of the mixed-effect logistic regression model, which includes significance levels using the Wald z statistic, is summarized in Table 3.1. The results showed that keyword recognition proportion correct was significantly higher in the quiet than in the noise condition ($\beta = 2.86$, $z = 13.78$, $p < .001$). There was a significant interaction between Condition (Noise vs. Quiet) and Talker Group (Native English vs. Native Mandarin; $\beta = 1.16$, $z = 2.64$, $p < .01$). This indicates that the difference in the proportion correct between the two conditions was larger for the Native Mandarin talkers' speech than for the Native English talkers' speech; mean proportion correct in Quiet – mean proportion correct in Noise was .32 (for Native Mandarin) and .27 (Native English). However, a post-hoc Tukey test, conducted using R package lsmeans (Lenth, 2016), confirmed that the effect of Condition (Noise vs. Quiet) was significant in both Native English and Native Mandarin talker groups (see Table 3.1 for the summary of the post-hoc test). This confirms that listeners transcribed keywords correctly more often in the quiet listening condition compared to listening to the same speech with noise. Furthermore,

the intelligibility proportion correct was higher for Native Mandarin-High talkers' speech than for Native Mandarin-Low talkers' speech ($\beta = 1.64$, $z = 5.76$, $p < .001$), and this pattern was similar across the Quiet and Noise condition ($\beta = .18$, $z = .52$, $p = .6$). This indicates that Native Mandarin-High talkers' speech was more intelligible than Native Mandarin-Low talkers' speech in both conditions.

Table 3.1. Summary of the mixed-effects logistic regression model for the intelligibility data in the noise and quiet conditions, as well as the result of the post-hoc Tukey test comparing the effect of Condition (Noise vs. Quiet) for the Native Mandarin and Native English talker groups' speech.

Mixed-effects logistic regression model for Condition & Talker Group
`Response ~ Condition*TalkerGroup1 + Condition*TalkerGroup2`
`+ (1+ Condition | Talker)`
`+ (1+ TalkerGroup1 + TalkerGroup2 | Listener)`
`+ (1+ Condition*TalkerGroup1 + Condition*TalkerGroup2 | Item)`				

Fixed Effects	**Estimate**	**S.E.**	**z-val.**	***p*-val.**
(Intercept)	2.13	.18	12.0	< .001
TalkerGroup1 (Native English vs. Native Mandarin High & Low)	1.3	.36	3.62	< .001 ***
Condition (Noise vs. Quiet)	2.86	.21	13.78	< .001 ***
TalkerGroup2 (Native Mandarin-High vs. Native Mandarin-Low)	1.64	.26	5.76	< .001 ***
TalkerGroup1: Condition	1.16	.44	2.64	.008 **
TalkerGroup2: Condition	.18	.35	.52	.6
Post-hoc Tukey test comparing the effect of Condition (noise vs. quiet)				
Talker Group 1	**Estimate**	**S.E.**	**z-ratio**	**p-val.**
Native Mandarin (High & Low)	-2.57	.207	-12.44	< .0001 ***
Native English	-3.44	.341	-10.09	< .0001 ***

The analysis above suggests that listeners' keyword recognition performance in the noise condition was lower than their performance in the quiet condition for both Native English and Native Mandarin (High and Low) talker groups' speech. In the next analyses, we focus on the data in the noise condition in order to examine whether clear speech enhancements improved listeners' keyword recognition in the challenging listening condition (i.e., with noise). Left panel in Figure 3.1 suggests that listeners' keyword recognition improved from plain to clear speech for Native English (14% increase) and

Native Mandarin-High talkers' speech (9% increase), but not for Native Mandarin-Low talkers' speech (2% decrease). Thus, we examine whether keyword recognition proportion correct was higher in clear speech than in plain speech, and whether this effect of speaking style was different for the speech produced by different talker groups. The effect of speaking style was examined in different sets of talker-group comparisons. First, we report results of a logistic mixed-effects regression model where we compared the effect of speaking style for the Native English vs. Native Mandarin (High and Low) talker groups. Then, we report results from three other models, each examining a subset of the data in the noise condition to compare the effect of speaking style for Native Mandarin-High vs. Native Mandarin-Low, for Native English vs. Native Mandarin-Low, and for Native English vs. Native Mandarin-High groups.

In all the subsequent analyses of the intelligibility data in the noise condition, the same structure of the logistic mixed-effects regression model was used; see the model syntax in Table 3.2. Specifically, the keyword correct (i.e., correct or incorrect) was the dichotomous dependent variable. The fixed effects were Style (plain or clear), Talker Group (different combinations of Native English, Native Mandarin-High, Native Mandarin-Low, depending on each model), and the interaction between the two. Style was contrast-coded to compare between plain (-.5) and clear (.5) speaking styles. Talker Group was contrast-coded in each model differently depending on the combinations of the Talker Groups in each data set/model; the contrast is specified in each section of the analysis below. The maximal random effects structure that would converge was implemented, which included random intercepts for talker, listener, and item. The random effects structure also included by-talker random slope for Style, by-listener slopes for Style, Talker

Group and their interaction, and by-item slopes for Style, Talker Group and their interaction.

The first model included data in the noise condition for all talker groups (Native English, Native Mandarin-High and Native Mandarin-Low). The Talker Group variable was contrast-coded to compare between Native English and Native Mandarin (Native Mandarin-High and Native Mandarin-Low) group (.5, -.25, -.25). The results of the mixed-effect logistic regression model are summarized in Table 3.2. The results showed that listeners recognized keywords correctly more often in the clear speech than in plain speech ($\beta = .44$, $z = 2.46$, $p < .05$). This effect of speaking style did not differ between the Native English and Native Mandarin talker groups ($\beta = .74$, $z = 1.54$, $p = .12$).

Table 3.2. Summary of the mixed-effects logistic regression model for the intelligibility data in the noise condition (for all talker groups: Native English vs. Native Mandarin-High & Low).

Mixed-effects logistic regression model for Style & Talker Group `Response ~ Style*TalkerGroup + (1+ Style| Talker) + (1+ Style*TalkerGroup| Listener) + (1+ Style*TalkerGroup| Item)`				
Fixed Effects	**Estimate**	**S.E.**	**z-val.**	***p*-val.**
(Intercept)	.70	.28	2.55	.01
Speaking style (plain vs. clear)	.44	.18	2.46	.014 *
TalkerGroup (Native English vs. Native Mandarin High & Low)	.82	.67	1.23	.22
Speaking style: TalkerGroup	.74	.48	1.54	.12

Next, we examined the data in the noise condition for Native Mandarin-High and Native Mandarin-Low groups' speech; the Talker Group variable was contrast-coded to compare between Native Mandarin-High (.5) and Native Mandarin-Low (-.5). The results of the mixed-effects logistic regression model are summarized in Table 3.3. The results showed that the proportion correct was significantly higher for Native Mandarin-High talkers' speech than for Native Mandarin-Low talkers' speech ($\beta = 1.64$, $z = 5.04$, $p <$

.001). This shows that in the listening condition of the same noise level (i.e., -2dB SNR), Native Mandarin-High talkers' speech was generally more intelligible than Native Mandarin-Low talkers' speech. The effect of Style was not significant ($\beta = .24$, $z = 1.61$, $p = .11$), but there was a significant interaction between Style and the Native Mandarin-High vs. Native Mandarin-Low talker group comparison ($\beta = .63$, $z = 2.29$, $p < .05$). This indicates that the size of the plain-clear intelligibility increase was different for the two talker groups' speech. To further examine this interaction, a post-hoc Tukey test was conducted. The test showed that the effect of Style on keyword recognition was significant for the speech produced by Native Mandarin-High talkers ($p = .01$) but not for the speech produced by Native Mandarin-Low talkers ($p = .7$); see Table 3.3 for the summary of the post-hoc test. Thus, Native Mandarin-High talkers' speech was generally more intelligible than that of Native Mandarin-Low talkers. Further, the intelligibility improvement from plain to clear speech was larger for Native Mandarin-High talkers' speech than for Native Mandarin-Low talkers' speech.

Table 3.3. Summary of the mixed-effects logistic regression model for the intelligibility data for the Native Mandarin-High and Native Mandarin-Low talkers' speech in the noise condition, as well as the results of the post-hoc Tukey test comparing the effect of Style (plain vs. clear) in each Talker Group.

Mixed-effects logistic regression model for Style & Talker Group				
Fixed Effects	**Estimate**	**S.E.**	**z-val.**	***p*-val.**
(Intercept)	.56	.23	2.44	.015
Speaking style (plain vs. clear)	.24	.15	1.61	.11
TalkerGroup (Native Mandarin High vs. Low)	1.64	.33	5.04	< .0001 ***
Speaking style: TalkerGroup	.63	.28	2.29	.022 *
Post-hoc Tukey test comparing the effect of Style (plain vs. clear)				
Talker Group	**Estimate**	**S.E.**	**z-ratio**	**p-val.**
Native Mandarin-Low	.07	.19	.39	.7
Native Mandarin-High	-.56	.22	-2.55	.01 *

Next, we examined the data in the noise condition for Native English and Native

Mandarin-Low groups' speech; the Talker Group variable was contrast-coded to compare between Native English (.5) and Native Mandarin-Low (-.5). The results of the mixed-effects logistic regression model are summarized in Table 3.4. The effect of Style was significant ($\beta = .36$, $z = 2.07$, $p < .05$), and there was a significant interaction between Style and the Native English vs. Native Mandarin-Low talker group comparison ($\beta = .89$, $z = 2.74$, $p < .01$). This interaction was further examined in a post-hoc Tukey test, and it showed that the effect of Style on keyword recognition was significant for Native English talkers' speech ($p = .001$) but not for Native Mandarin-Low talkers' speech ($p = .72$); see Table 3.4 for the summary of the post-hoc test. Thus, the intelligibility improvement from plain to clear speech was larger for Native English talkers' speech than for Native Mandarin-Low talkers' speech.

Table 3.4. Summary of the mixed-effects logistic regression model for the intelligibility data for the Native English and Native Mandarin-Low talkers' speech in the noise condition, as well as the results of the post-hoc Tukey test comparing the effect of Style (plain vs. clear) in each Talker Group.

Mixed-effects logistic regression model for Style & Talker Group				
Fixed Effects	**Estimate**	**S.E.**	**z-val.**	***p*-val.**
(Intercept)	.43	.23	1.86	.06
Speaking style (plain vs. clear)	.36	.17	2.07	.039 *
TalkerGroup (Native English vs. Native Mandarin-Low)	1.38	.38	3.66	.0003 ***
Speaking style: TalkerGroup	.89	.32	2.74	.006 **
Post-hoc Tukey test comparing the effect of Style (plain vs. clear)				
Talker Group	**Estimate**	**S.E.**	**z-ratio**	**p-val.**
Native Mandarin-Low	.08	.23	.36	.72
Native English	-.8	.25	-3.27	.0011 **

Finally, we examined the data in the noise condition for Native English and Native Mandarin-High groups' speech; the Talker Group variable was contrast-coded to compare between Native English (.5) and Native Mandarin-High (-.5). The results of the mixed-effects logistic regression model are summarized in Table 3.5. The effect of Style was

significant (β = .69, z = 2.92, $p < .01$), though the interaction between Style and the Native English vs. Native Mandarin-High talker group comparison was not significant (β = .24, z = .56, $p = .58$). Thus, the size of the intelligibility improvement from plain to clear speech did not differ for Native English talkers' and Native Mandarin-High talkers' speech (at the noise levels presented in the current study: -6dB SNR for Native English and -2dB SNR for Native Mandarin-High talkers' speech).

Table 3.5. Summary of the mixed-effects logistic regression model for the intelligibility data for the Native English and Native Mandarin-High talkers' speech in the noise condition.

Fixed Effects	**Estimate**	**S.E.**	**z-val.**	***p*-val.**
(Intercept)	1.24	.24	5.27	< .0001
Speaking style (plain vs. clear)	.69	.24	2.92	.003 **
TalkerGroup (Native English vs. Native Mandarin-High)	-.25	.38	-.67	.5
Speaking style: TalkerGroup	.24	.43	.56	.58

3.2.3. Summary of Experiment 2A

Together, these results demonstrated that when listening to speech with noise, clear speech was generally more intelligible than plain speech. However, there was a difference in the size of intelligibility improvement for the speech produced by native English talkers and non-native talkers of higher- and lower-proficiency. That is, the size of intelligibility improvement from plain to clear speech was smaller for lower-proficiency non-native talkers' speech than for higher-proficiency non-native talkers' speech and for native English talkers' speech. However, the size of intelligibility improvement did not differ for higher-proficiency talkers' speech and native English talkers' speech (at the different noise levels: -2dB SNR for higher-proficiency talkers' speech and -6dB SNR for native English talkers' speech). Further, higher-proficiency talkers' speech was generally more intelligible than lower-proficiency talkers' speech at the same noise level (i.e., -2dB SNR). These

results suggest that higher-proficiency non-native talkers not only produce generally more intelligible speech than lower-proficiency non-native talkers do, but also that higher-proficiency talkers are better able to increase intelligibility of their speech than lower-proficiency talkers are.

3.3 Experiment 2B

In this experiment, we explore subjective aspects of perceptual benefits resulting from clear speech enhancements. Specifically, we examine whether listeners perceive clear speech to be easier to understand than plain speech (comprehensibility), and whether listeners perceive that clear speech is produced with increased effort than plain speech (talker effort). We further examine whether one type of subjective evaluation is more robustly improved by clear speech enhancements than the other, as well as whether the subjective evaluations pattern similarly for the speech produced by different talker groups (i.e., native English talkers and non-native talkers of different proficiency levels). As in Experiment 2A, we conducted this experiment in two listening conditions (i.e., listening in quiet and with noise) to examine whether the presence of noise impacts the way listeners respond to these tasks (comprehensibility and talker effort tasks).

3.3.1. Methods

3.3.1.1. Participants

Participants were 483 native English listeners (226 females, 254 males, 1 non-binary, 2 declined to provide a gender; age = 19 - 76 years, mean = 37.1). They were recruited using Amazon Mechanical Turk; none of the participants in this experiment

participated in Experiment 2A. Of the 483 participants, 242 participants completed the comprehensibility task and 241 participants completed the talker effort task. None of the listeners reported a history of speech or hearing impairment. All participants resided in the United States, and self-reported to be native speakers of American English. None of the participants reported experience with Mandarin Chinese.

3.3.1.2. Materials

The materials were the same as those in Experiment 2A. Using the 30 BKB sentences produced by the 4 Native English, 4 Native Mandarin-High, and 4 Native Mandarin-Low talkers, clear- and plain-speaking styles of each sentence (e.g., *the man tied his shoes*) produced by the same talker were concatenated using a Praat script. This was done for the sentences in the quiet and noise conditions; the noise level was -6dB SNR for Native English talkers' sentences and -2dB SNR for Native Mandarin talkers' sentences as in Experiment 2A. For each concatenated sentence, two versions of the clear-plain style order were created (i.e., clear-style sentence preceding plain-style sentence and vice versa). When concatenating the two files, 100 ms silence was added in between the two files. Thus, for the quiet condition and noise condition, each stimulus consisted of the following parts:

```
Quiet condition: 500ms silence + sentence 1 in quiet + 500ms silence +
100ms silence + 500ms silence + sentence 2 in quiet + 500ms silence

Noise condition: 500ms noise + sentence 1 in noise + 500ms noise +
100ms silence + 500ms noise + sentence 2 in noise + 500ms noise
```

Therefore, there were a total of 1440 unique items (30 sentences x 12 talkers x 2 clear-plain style orders x 2 quiet-noise conditions).

3.3.1.3. Procedure

The experiment was conducted online via Amazon Mechanical Turk. Each participant completed either the Comprehensibility or Effort task. Participants were instructed that they would listen to English speech and would be asked to evaluate the speech. In each trial, participants consecutively heard two productions of the same sentence (e.g., *the man tied his shoes*) that were produced by the same talker in a plain- and clear-speaking style, and were asked to evaluate which sentence was easier to understand (Comprehensibility task) or which sentence they thought the speaker was trying to say more clearly (Effort task) by choosing 'first sentence' or 'second sentence' on the screen.

Each participant completed two practice trials (with two talkers and items different from the test trials) followed by 30 test trials. For the 30 test trials, each participant heard the clear-plain pair of all 30 sentences; half of the 30 pairs were presented with the clear-style sentence first, and the other half was presented with the plain-style sentence first. Each participant heard 6 talkers (2 Native English talkers, 2 Native-Mandarin High talkers, 2 Native Mandarin-Low talkers; 5 clear-plain pairs produced by each talker). The combination of the sentence, talker, and clear-plain order was counter-balanced across participants. The order of item presentation was randomized for each participant. Each of the 30 clear-plain pairs for each talker was listened by 4-6 participants. After the experimental trials, participants completed a post-test survey that collected demographic and language background information.

3.3.1.4. Scoring and analysis

Each response in the Comprehensibility and Effort task was given a score of 0 or 1.

That is, if the participant chose the sentence (i.e., first or second sentence) that was produced in the clear-speaking style, the response was scored as 1; if the participant chose the sentence that was produced in the plain-speaking style, the response was scored as 0. There were 7260 data points for the Comprehensibility task (242 participants x 30 clear-plain pairs) and 7230 data points for the Effort task (241 participants x 30 clear-plain pairs). Thus, a total of 14490 data points was analyzed.

3.3.2. Results

Figure 3.2 shows the proportion of 'clear' responses (i.e., the proportion of times a listener chose the sentence spoken in the clear-speaking style as either more comprehensible or more effortful) for each talker group (Native English, Native Mandarin-High, and Native Mandarin-Low) by task (Comprehensibility or Effort) and condition (Noise or Quiet). The figure suggests that listeners chose the clear-style sentence as opposed to the plain-style sentence more often than chance (i.e., .5 proportion clear response) for the Effort task, but not for the Comprehensibility task. This pattern seems similar across the Noise and Quiet conditions and across the three talker groups. In order to confirm these observations, we first analyzed whether the proportion clear response was significantly higher or lower than the chance level (.5). We carried out 12 one-sample two-tailed t-tests (3 talker groups x 2 tasks x 2 conditions; as represented as the 12 individual bars in Figure 3.2) with Bonferroni corrected *p*-values (.05 / 12 = .004). The tests confirmed that proportion clear response for the Effort task was different from the chance level for all talker groups in both noise and quiet conditions ($p < .00001$ for all). The proportion clear response for the Comprehensibility task was not different from the chance

level for all talker groups in both conditions, except for the Native Mandarin-High group in the quiet condition: $t(1199) = -5.97$, $p < .00001$. This indicates that, in the quiet condition, listeners' responses were biased toward Native Mandarin-High group's plain-style sentences when asked which sentence was easier to understand.

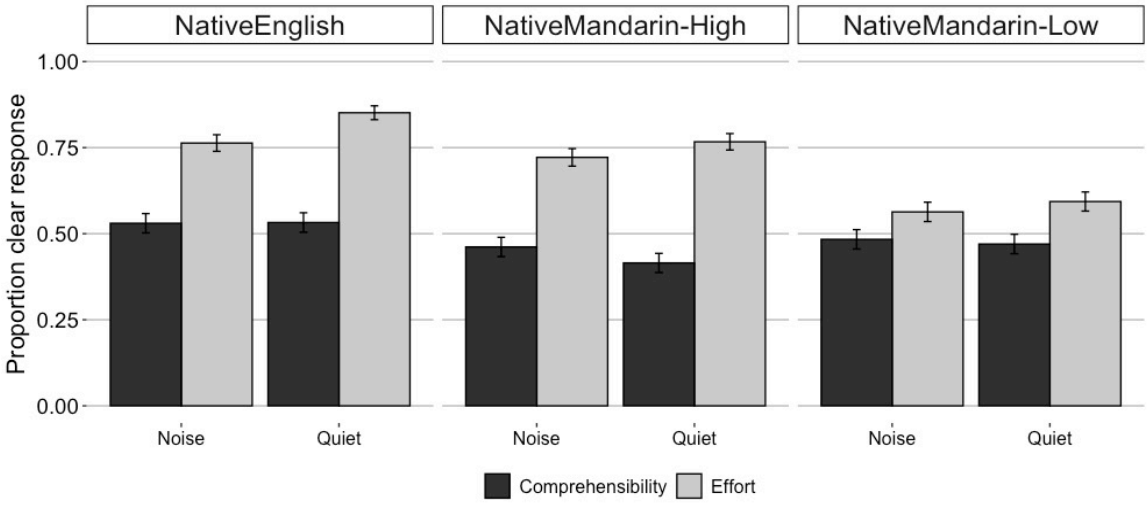

Figure 3.2. Proportion clear response for each talker group (Native English, Native Mandarin-High, and Native Mandarin-Low) by task (Comprehensibility or Effort) and condition (Noise or Quiet). The error bars represent 95% confidence intervals.

The above results indicate that listeners were more likely to choose the clear-style sentence as opposed to the plain-style sentence in the Effort task, but not in the Comprehensibility task. In addition to these results, Figure 3.2 also suggests that proportion clear response was higher in the Effort task than the Comprehensibility task in general (i.e., comparing the proportion clear response between the two tasks, instead of comparing proportion clear response in each task to the chance level as in the above analysis). That is, listeners chose the clear-style sentence more often when asked which sentence was spoken with more effort (Effort task) as compared to when asked which sentence was easier to understand (Comprehensibility task). The figure also suggests that this task-based difference was smaller for Native Mandarin-Low talkers' speech than for Native English

and Native Mandarin-High talkers' speech in both Noise and Quiet conditions. In order to test these observations, the data were analyzed via a series of logistic mixed-effects regression models where clear-style recognition (i.e., 1 or 0) was the dichotomous dependent variable. First, we report results of a logistic mixed-effects regression model where we compared the effect of Task (Comprehensibility vs. Effort) and Condition (Noise vs. Quiet) for the Native English vs. Native Mandarin (High and Low) talker groups. Then, we report results from three other models, each examining a subset of the data comparing the effect of Task and Condition for Native Mandarin-High vs. Native Mandarin-Low, for Native English vs. Native Mandarin-Low, and for Native English vs. Native Mandarin-High groups.

In all the subsequent analyses of the 2AFC response data, the same basic structure of the logistic mixed-effects regression model was used. As fixed effects, Task (Comprehensibility or Effort), Condition (Quiet or Noise), and Talker Group (Native English, Native Mandarin-High, Native Mandarin-Low) were included along with their two- and three-way interactions. Binary predictor variables were contrast-coded: Condition (Noise: -.5, Quiet: .5) and Task (Comprehensibility: -.5, Effort: .5). Talker Group was contrast-coded in each model differently depending on the combinations of the Talker Groups in each data set/model; the contrast is specified in each section of the analysis below. In order to allow for the model to converge with maximal random slopes, we simplified the model by uncorrelating the random effects (Barr, Levy, Scheepers, & Tily, 2013); see the model syntax in each table below. The random effects structure included random intercepts for talker, listener, and item. The random effects structure also included by-talker random slopes for Condition, Task, and their interaction, by-listener slopes for

Talker Group, and by-item slopes for Condition, Task, Talker Group and their interactions. In each model, the random effects that did not account for any variance (e.g., by-talker random slope for Condition x Task) were not included in the model to avoid overfitting of the model; see the model syntax of each model for the specific random effect structure.

The first model included the 2AFC response data for all talker groups' speech (Native English, Native Mandarin-High and Native Mandarin-Low); the model syntax is shown in Table 3.6. The Talker Group variable was contrast-coded to compare between Native English and Native Mandarin (Native Mandarin-High and Native Mandarin-Low) group (.5, -.25, -.25). The results of the mixed-effects logistic regression model, which includes significance levels using the Wald z statistic, are summarized in Table 3.6. In this model and the subsequent model results, we only interpret the results relevant to our questions: whether listeners' perception of the plain and clear speech differed depending on the Task (Comprehensibility vs. Effort) and whether this effect of Task differed for the speech produced by different talker groups. We were also interested in whether the effect of Task differed depending on the listening condition (in Quiet or Noise condition). The results showed that proportion clear response was significantly higher for the Effort task than for the Comprehensibility task ($\beta = 1.17$, $z = 7.19$, $p < .001$). That is, listeners chose the clear-style sentence more reliably when asked which sentence was spoken with more effort (Effort task) as compared to when asked which sentence was easier to understand (Comprehensibility task). This effect of Task interacted with Condition ($\beta = .47$, $z = 3.45$, $p < .001$), indicating that the difference between Comprehensibility vs. Effort task was larger in the Quiet condition than in the Noise condition. This pattern was similar across the Native English and Native Mandarin groups, as the three-way interaction for Condition x

Task x Talker Group did not significantly improve the model fit (β = .44, z = 1.85, *p* = .06). Further, the effect of Task did not significantly interact with the Native English vs. Native Mandarin Talker Group (β = .76, z = 1.77, *p* = .08). This indicates that the effects of Task (Comprehensibility vs. Effort) did not differ across the two talker groups. These results showed that clear speech enhancements improved subjective evaluation of talker effort more than that of comprehensibility, across the speech of Native English and Native Mandarin (High and Low) talkers. This effect of task was larger when listening to the speech in quiet than with noise.

Table 3.6. Summary of the mixed-effects logistic regression model for the 2AFC response data for Native English, Native Mandarin-High, and Native Mandarin-Low groups' speech.

Mixed-effects logistic regression model for Condition, Task & Talker Group

```
Response ~ Condition*Task*TalkerGroup
        + (1+ Condition*Task - Condition:Task || Talker)
        + (1+ TalkerGroup || Listener)
        + (1+ Condition*Task*TalkerGroup - Condition:Task -
            Condition:TalkerGroup || Item)
```

Fixed Effects	**Estimate**	**S.E.**	**z-val.**	***p*-val.**
(Intercept)	.51	.09	5.53	< .001
Condition (Quiet vs. Noise)	.16	.1	1.63	.10
Task (Comprehensibility vs. Effort)	1.17	.16	7.19	< .001 ***
TalkerGroup (Native English vs. Native Mandarin High & Low)	.81	.24	3.32	< .001 ***
Condition: Task	.47	.14	3.45	< .001 ***
Condition: TalkerGroup	.43	.21	2.04	.042 *
Task: TalkerGroup	.76	.43	1.77	.077
Condition: Task: TalkerGroup	.44	.24	1.85	.064

Next, we examined the 2AFC response data for Native Mandarin-High and Native Mandarin-Low groups' speech; the Talker Group variable was contrast-coded to compare between Native Mandarin-High (.5) and Native Mandarin-Low (-.5). The model syntax and the results of the model are summarized in Table 3.7. The results showed a significant effect of Task (Comprehensibility vs. Effort; β = .98, z = 8.28, *p* < .001), and it interacted

with the Native Mandarin-High vs. Native Mandarin-Low talker group comparison (β = 1.04, z = 4.89, $p < .001$). This indicates that the effect of Task differed for the two talker groups' speech. A post-hoc Tukey test confirmed that the effect of Task was significant for both Talker Groups' speech, but the effect was larger for the Native Mandarin-High talkers' speech than for Native Mandarin-Low talkers' speech (see Table 3.7 below). This pattern of the two-way interaction between Task and Talker Group did not differ across the two conditions (Noise vs. Quiet), as the three-way interaction did not improve the model fit (β = .3, z = 1.45, p = .15). Further, a post-hoc Tukey test, examining the effect of Talker Group in each Task, showed that the difference between the Native Mandarin-High and Native Mandarin-Low group was significant in the Effort task but not in the Comprehensibility task (see Table 3.7 below). That is, for native listeners, Native Mandarin-High talkers' increased effort in clear-style sentences (as compared to plain-style sentences) was easier to detect than that of Native Mandarin-Low talkers; though the listeners did not perceive clear-style sentences to be easier to understand than plain-style sentences across the two talker groups' speech. This indicates that the difference between the Effort and Comprehensibility task was smaller for Native Mandarin-Low talkers' speech than for Native Mandarin-High talkers' speech, because Native Mandarin-Low talkers' increased effort in clear speech was more difficult to detect than that of Native Mandarin-High talkers.

Table 3.7. Summary of the mixed-effects logistic regression model for the 2AFC response data for Native Mandarin-High and Native Mandarin-Low groups' speech, as well as the results of the post-hoc Tukey test comparing the effect of Task (Comprehensibility vs. Effort) in each Talker Group, and the post-hoc Tukey test comparing the effect of Talker Group (Native Mandarin-High vs. Native Mandarin-Low) in each Task.

Mixed-effects logistic regression model for Condition, Task & Talker Group
`Response ~ Condition*Task*TalkerGroup`
`+ (1+ Condition*Task – Condition:Task || Talker)`
`+ (1+ TalkerGroup || Listener)`
`+ (1+ Condition*Task*TalkerGroup – Condition:Task:TalkerGroup – Condition:Task || Item)`				

Fixed Effects	**Estimate**	**S.E.**	**z-val.**	***p*-val.**
(Intercept)	.3	.09	3.44	< .001
Condition (Quiet vs. Noise)	.05	.09	.53	.6
Task (Comprehensibility vs. Effort)	.98	.12	8.28	< .001 ***
TalkerGroup (Native Mandarin-High vs. Native Mandarin-Low)	.35	.17	2.03	.04 *
Condition: Task	.37	.13	2.87	.0041 **
Condition: TalkerGroup	.002	.17	.01	.99
Task: TalkerGroup	1.04	.21	4.89	< .001 ***
Condition: Task: TalkerGroup	.3	.21	1.45	.15
Post-hoc Tukey test comparing the effect of Task in each Talker Group				
Talker Group	**Estimate**	**S.E.**	**z-ratio**	**p-val.**
Native Mandarin-Low	-.46	.16	-2.89	.004 **
Native Mandarin-High	-1.5	.16	-9.34	< .0001 ***
Post-hoc Tukey test comparing the effect of Talker Group in each Task				
Comprehensibility task	.18	.2	.88	.38
Effort task	-.87	.2	-4.29	< .0001 ***

Next, we examined the 2AFC response data for Native English and Native Mandarin-Low groups' speech; the Talker Group variable was contrast-coded to compare between Native English (.5) and Native Mandarin-Low (-.5). The model syntax and the results of the model are summarized in Table 3.8. There was a significant effect of Task (Comprehensibility vs. Effort; $\beta = .98$, $z = 8.04$, $p < .001$), and it interacted with the Native English vs. Native Mandarin-Low talker group comparison ($\beta = 1.06$, $z = 4.53$, $p < .001$). A post-hoc Tukey test confirmed that the effect of Task was significant for both Talker

Groups, but the effect was larger for the Native English talkers' speech than for Native Mandarin-Low talkers' speech (see Table 3.8 below). This difference by Task for the two Talker Groups was larger in the Quiet condition than for the Noise condition, as indicated by the significant three-way interaction among Condition, Task, and Talker Group ($\beta = .45$, $z = 2.07$, $p < .05$). The three-way interaction is further illustrated in Figure 3.3. Furthermore, a post-hoc Tukey test, examining the effect of Talker Group in each Task, showed that the difference between the Native English and Native Mandarin-Low group was significant in Effort task but not in Comprehensibility task (see Table 3.8 below). This indicates that Native English talkers' increased effort in the clear-style sentences was easier to detect than that of Native Mandarin-Low talkers; while the listeners did not perceive the clear-style sentences to be easier to understand than the plain-style sentences across the two talker groups' speech.

Finally, we examined the 2AFC response data for Native English and Native Mandarin-High groups' speech; the Talker Group variable was contrast-coded to compare between Native English (.5) and Native Mandarin-High (-.5). The model syntax and the results of model are summarized in Table 3.9. There was a significant effect of Task (Comprehensibility vs. Effort; $\beta = 1.69$, $z = 12.31$, $p < .001$), though it did not interact with the Native English vs. Native Mandarin-High talker group comparison ($\beta = .04$, $z = .19$, $p = .85$). This indicates that the effect of Task was similar between the two Talker Groups, and this pattern was similar across the Noise and Quiet conditions, as the three-way interaction among Condition, Task, and Talker Group did not significantly improve the model ($\beta = .18$, $z = .89$, $p = .37$).

Table 3.8. Summary of the mixed-effects logistic regression model for the 2AFC response data for Native English and Native Mandarin-Low groups' speech, as well as the post-hoc Tukey test comparing the effect of Task (Comprehensibility vs. Effort) in each Talker Group, and the results of the post-hoc Tukey test comparing the effect of Talker Group (Native English vs. Native Mandarin-Low) in each Task.

Mixed-effects logistic regression model for Condition, Task & Talker Group
`Response ~ Condition*Task*TalkerGroup`
`+ (1+ Condition*Task - Condition:Task || Talker)`
`+ (1+ TalkerGroup || Listener)`
`+ (1+ Condition*Task*TalkerGroup – Condition:TalkerGroup – Condition:Task - Condition:Task:TalkerGroup || Item)`				

Fixed Effects	**Estimate**	**S.E.**	**z-val.**	***p*-val.**
(Intercept)	.51	.09	5.75	< .001
Condition (Quiet vs. Noise)	.20	.12	1.75	.08
Task (Comprehensibility vs. Effort)	.98	.12	8.04	< .001 ***
TalkerGroup (Native English vs. Native Mandarin-Low)	.76	.17	4.38	< .001 ***
Condition: Task	.45	.13	3.41	< .001 ***
Condition: TalkerGroup	.31	.20	1.52	.13
Task: TalkerGroup	1.06	.23	4.53	< .001 ***
Condition: Task: TalkerGroup	.45	.22	2.07	.039 *
Post-hoc Tukey test comparing the effect of Task in each Talker Group				
Talker Group	**Estimate**	**S.E.**	**z-ratio**	**p-val.**
Native Mandarin-Low	-.46	.17	-2.73	.006 **
Native English	-1.5	.17	-8.83	< .0001 ***
Post-hoc Tukey test comparing the effect of Talker Group in each Task				
Comprehensibility task	-.24	.21	-1.14	.26
Effort task	-1.29	.21	-6.11	< .0001 ***

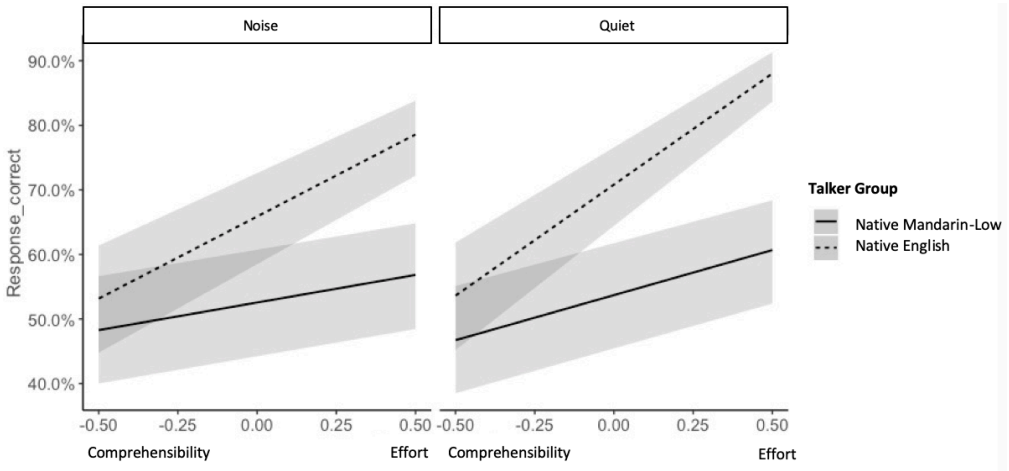

Figure 3.3. Model prediction of response correct for Native English and Native Mandarin-Low talkers' speech by Task (Comprehensibility: -.5, Effort: .5) in two Conditions (Noise and Quiet).

Table 3.9. Summary of the mixed-effects logistic regression model for the 2AFC response data for Native English and Native Mandarin-High groups' speech.

Mixed-effects logistic regression model for Condition, Task & Talker Group
`Response ~ Condition*Task*TalkerGroup`
`+ (1+ Condition*Task - Condition:Task || Talker)`
`+ (1+ TalkerGroup || Listener)`
`+ (1+ Condition*Task*TalkerGroup - Condition:Task - Condition:Task:TalkerGroup || Item)`				

Fixed Effects	**Estimate**	**S.E.**	**z-val.**	***p*-val.**
(Intercept)	.78	.13	6.21	< .001
Condition (Quiet vs. Noise)	.24	.14	1.79	.074
Task (Comprehensibility vs. Effort)	1.69	.14	12.31	< .001 ***
TalkerGroup (Native English vs. Native Mandarin-High)	.46	.23	1.99	.047
Condition: Task	.68	.20	3.34	< .001 ***
Condition: TalkerGroup	.35	.20	1.73	.083
Task: TalkerGroup	.04	.21	.19	.85
Condition: Task: TalkerGroup	.18	.20	.89	.37

3.3.3. Summary of Experiment 2B

In this experiment, we examined listeners' subjective evaluations of plain and clear speech produced by native and non-native talkers, and demonstrated the results of several types of comparisons regarding listeners' performance in two tasks: the comprehensibility task and the talker effort task. Specifically, we examined listeners' choice of plain- vs. clear-style sentences; when listeners were presented with two productions of the same sentence in clear- and plain-speaking styles, they reliably chose the clear-style sentence as opposed to the plain-style sentence in the Effort task (when asked which sentence was spoken with more effort). However, they did not reliably choose the clear-style sentence (as opposed to the plain-style sentence) in the Comprehensibility task (when asked which sentence was easier to understand). Then, we directly compared the likelihood of listeners choosing the clear-style sentence (as opposed to plain-style sentence) between the two tasks (Effort vs. Comprehensibility). Listeners chose the clear-style sentence more often in the Effort task than in the Comprehensibility task, indicating that clear speech enhancements

improved subjective evaluation of talker effort more than that of comprehensibility. The size of this task-based difference in the proportion clear response was smaller for the speech of lower-proficiency non-native talkers than for the other talker groups, and this was because lower-proficiency talkers' increased effort in the clear-style sentences (as compared to plain-style sentences) was more difficult to detect than of native English and higher-proficiency non-native talkers. Further, across the speech of different talker groups, the task-based difference in proportion clear response was larger when listeners evaluated the speech in quiet than with noise.

3.4. Discussion and conclusion

3.4.1. Summary of the results

The present study examined native English listeners' perception of clear speech produced by native English talkers and non-native English talkers of different proficiency levels. Particularly, we investigated perceptual benefits of clear speech in terms of the improvement in intelligibility (Experiment 2A) as well as in listeners' subjective evaluation of the speech, namely, perceived degree of comprehensibility and talkers' effort (Experiment 2B). Experiment 2A demonstrated that listeners generally understood clear speech better than plain speech. However, the size of intelligibility improvement from plain to clear speech differed for the speech produced by native English talkers, higher-proficiency non-native talkers, and lower-proficiency non-native talkers. Specifically, when listening to the speech with noise, the plain-to-clear intelligibility improvement was smaller for the speech of lower-proficiency non-native talkers than that of higher-proficiency talkers and native English talkers, whereas the size of intelligibility improvement did not

significantly differ for the speech of higher-proficiency talkers and native English talkers.

In Experiment 2B, listeners were presented with two productions of the same sentence produced by the same talker in clear- and plain-speaking styles. The listeners reliably chose the clear-style sentence as opposed to the plain-style sentence in the effort task (when asked which sentence was spoken with more effort), but not in the comprehensibility task (when asked which sentence was easier to understand). Between the two tasks, listeners chose the clear-style sentence more often in the effort task than in the comprehensibility task. This task-based difference (i.e., the difference in the proportion of times the listeners chose the clear-style sentence between the effort and comprehensibility task) was generally larger when listeners evaluated the speech in quiet than with noise. However, the size of the task-based difference was smaller for the speech of lower-proficiency talkers than for the speech of higher-proficiency talkers and native talkers. These results suggest that perceptual benefits resulting from clear speech enhancements differed for different types of tasks (i.e., intelligibility, subjective evaluation of comprehensibility and talker effort) as well as for the speech produced by native English talkers and by non-native talkers of different proficiency levels.

3.4.2. Intelligibility improvement based on clear speech enhancements

The results of Experiment 2A demonstrated that non-native talkers' proficiency level impacted their ability to produce intelligible speech in general, as well as to increase intelligibility of their speech. Specifically, the current results showed not only that higher-proficiency talkers' speech was generally more intelligible than lower-proficiency talkers' speech across quiet and noisy listening conditions, but also that higher-proficiency talkers'

clear speech enhancements resulted in larger plain-to-clear speech intelligibility gains than those of lower-proficiency non-native talkers. These results suggest that as non-native talkers' target language proficiency develops, they produce more intelligible speech and are better able to further increase intelligibility of their speech.

It is possible that the difference in the size of intelligibility improvement between higher- and lower-proficiency talkers' clear speech stems from the difference in the types of acoustic modifications made in their clear speech. That is, as shown in Chapter 2, lower-proficiency talkers' acoustic-phonetic modifications made in clear speech were overall smaller than those made in higher-proficiency talkers and native English talkers' clear speech. Thus, it is plausible that small overall changes in acoustic characteristics between plain and clear speech resulted in small changes in intelligibility in perception. However, lower-proficiency talkers did make significant plain-to-clear speech modifications in some aspects, including articulation rate (i.e., rate of speech without pauses). Though previous results suggest a significant contribution of reduced speaking rate in clear speech to intelligibility benefit (Bradlow et al., 2003), it is possible that other types of acoustic modifications in clear speech impact intelligibility more than the changes in speaking rate. For example, a wider pitch range is generally associated with higher intelligibility (Bradlow, Torretta, & Pisoni, 1996). The lower-proficiency talkers in the current study did not make significant modifications in average pitch and pitch range, though higher-proficiency talkers and native English talkers did. Thus, it is possible that the acoustic-phonetic characteristics of clear speech enhancements made by lower-proficiency talkers did not contribute to increasing their speech intelligibility as much as those made by higher-proficiency talkers and native talkers did.

Acoustic characteristics of clear speech enhancements may also explain some of the individual variability observed in the clear speech intelligibility improvement. Though we did not directly examine how different types of acoustic modifications relate to intelligibility gains (as the goal of the study was to examine whether talkers' target language experience at the group level would influence clear speech intelligibility benefits), we describe some of the tendencies of individual variability in intelligibility gains within each talker group's speech, for the purpose of suggesting directions for future work. Particularly, among the native English talkers, Talker 405's clear speech resulted in smaller intelligibility gains (3% increase) compared to the other three native English talkers (15%, 17%, and 15% increase). Also, among the higher-proficiency non-native talkers, Talker 306 and 310's clear speech resulted in relatively large intelligibility gains (19% and 16% increase, respectively), though Talker 302 and 311's clear speech intelligibly gains were small (4% increase and 2% decrease, respectively). These results are in line with substantial individual variability in clear speech intelligibility improvement reported in previous studies, including when native English listeners evaluated native English talkers' plain and clear speech (Ferguson, 2004; Smiljanić & Bradlow, 2005), as well as when native English listeners evaluated highly proficient non-native talkers' plain and clear speech (Smiljanić & Bradlow, 2011).

Given some previous results demonstrating that a larger degree of vowel space expansion is associated with a larger plain-to-clear speech intelligibility improvement (Ferguson, 2004; Ferguson & Kewley-Port, 2007), it is possible that the degree of vowel space expansion may explain some of the individual variability in the size of intelligibility gains. In fact, the native talker 405, who showed the least amount of intelligibility gains,

also showed the smallest vowel space expansion among other native talkers. However, the degree of vowel space expansion may not explain the individual variability in intelligibility gains of non-native clear speech. For example, the higher-proficiency talker 306, who showed the smallest vowel space expansion, showed the largest intelligibility gains among other higher-proficiency talkers. Also, the lower-proficiency talker 107, who showed the largest vowel space expansion among other lower-proficiency talkers, showed 12% decrease in clear speech intelligibility. Thus, it is possible that the relationship between certain acoustic characteristics of clear speech enhancements and intelligibility gains may be different for native listeners' perception of native and non-native clear speech. Further, acoustic features not examined in the materials of the current perception study (e.g., the extent of coarticulation; Bradlow, 2002; frequency of stop-burst releases: Ferguson & Kewley-Port, 2007) may also explain the clear speech intelligibility benefit, in terms of the differences among different talker groups as well as the individual variability within each talker group. A future study with a larger dataset may be better suited to explore the relationship between acoustic-phonetic modifications and variability in clear speech intelligibility improvement, and how it may differ for the speech of native talkers and non-native talkers of different proficiency levels.

3.4.3. The gap among different measures of clear speech perception

The present study also demonstrated that perceptual benefits of clear speech enhancements were manifested differently depending on the type of perception tasks that listeners engaged in. Particularly, despite the clear speech intelligibility benefit shown for native English talkers' and higher-proficiency talkers' speech (Experiment 2A),

native English listeners did not perceive these talkers' clear-style sentences to be easier to understand than their plain-style sentences (Experiment 2B). Listeners even perceived higher-proficiency talkers' *plain*-style sentences to be easier to understand than *clear*-style sentences in the quiet listening condition. However, listeners were sensitive to the increased effort in the clear-style sentences as compared to the plain-style sentences for all the talker groups' speech. Though it may be puzzling to observe such discrepancies among different measures of clear speech perception, it is possible that acoustic features of clear speech enhancements influenced listeners' perception differently for different tasks. For example, some acoustic features, such as decreased speaking rate, may have contributed to improving intelligibility of the speech, but not perceived degree of comprehensibility. Though a decrease in speaking rate has been one of the most prominent features of clear speech enhancements in previous studies (e.g., Bradlow et al., 2003; Picheny et al., 1986) as well as in the current study (see Chapter 2), previous results have also demonstrated that slower speaking rates are associated with poorer comprehensibility ratings for perception of non-native speech (Munro & Derwing, 1998). Specifically, Munro and Derwing (1998, 2001) claimed that very slow speech may place extra processing demand for listeners by requiring them to retain information in working memory for a longer period of time. Further, this slower speaking rate may also allow listeners to notice more phonological errors. Smiljanić and Bradlow (2005) also suggested that there may be a limit to the perceptual benefit influenced by clear speech strategies; modifying acoustic-phonetic features beyond a certain threshold may result in speech sounding unnatural and even less intelligible. The current results contribute to these lines of research by demonstrating that clear speech enhancements that are effective at improving one measure of listeners'

evaluation may not improve other perceptual measures. That is, while larger degrees of acoustic modifications (e.g., slowing down the speech, increasing the pitch range) may contribute to the greater talker effort perceived by listeners, they may not impact other types of perception in the same way. Particularly, for improving intelligibility and perceived degree of comprehensibility, there may be an optimal amount for acoustic-phonetic modification, and modifications exceeding the optimal amount may not result in improvement in these perceptual measures.

The gap between the two subjective measures of clear speech perception observed in the current study (i.e., perceived degree of talker effort vs. comprehensibility) may originate from the particular clear speech elicitation method used here. That is, reading the same sentences once in a plain-speaking style and once in a clear-speaking style, without actual listeners present in the room, may have induced some acoustic-phonetic modifications that sound unnatural to the listeners in the perception experiment. This may have contributed to the clear speech perceived to be not easier to understand than plain speech, but perceived to have been produced with increased effort compared to plain speech. Previous studies have demonstrated that acoustic-phonetic characteristics of clear speech modifications vary depending on how the speech is elicited. For example, clear speech elicited with read speech, using the instruction to speak clearly as if talking to someone who is hearing impaired, involved more extreme changes in some acoustic-phonetic characteristics (e.g., pitch range and speaking rate) than the spontaneous speech produced in a challenging listening condition (Hazan & Baker, 2011). Read clear speech elicited in noise (i.e., the talkers read sentences while simultaneously listening to noise) involved even more extreme changes than the read clear speech produced in quiet (in

speaking rate, pitch mean, energy in the 1-3 kHz range: Gilbert et al., 2014). As suggested by Hazan and Baker (2011), read clear speech likely involves a relatively constant degree of clarification as compared to the spontaneous clear speech produced in a challenging listening condition, which involves a larger variance in degrees of clarification. The varying degrees of clarification may better model the tension between a talker and a listener (i.e., to minimize articulatory effort and to clarify speech to ensure successful communication: Lindblom, 1990), which fluctuates over the course of the interaction, as compared to the constant degree of clarification found in read clear speech. Thus, it is possible that the acoustic-phonetic modifications made in the read clear speech in the current study were perceived to be somewhat unnatural, making the speech perceived to be not easier to understand than plain speech. However, it is an open question to what extent different clear speech elicitation methods influence the way clear speech impacts different aspects of perception (e.g., how intelligibility improvement relates to subjective evaluations of the same speech).

A possible variation in research contexts could also include examining perceptual consequences of clear speech enhancements with different listener populations. That is, listeners with different backgrounds may perceive clear speech produced by native and non-native talkers of English differently than native English listeners. For example, because the talkers in the current study, as in previous studies (e.g., Smiljanić & Bradlow, 2005), were instructed to speak as if talking to a hearing-impaired listener, their clear speech enhancements may be effective at improving perceived degree of comprehensibility for hearing-impaired listeners, in addition to improving intelligibility for these types of listeners as shown in previous studies (e.g., Picheny et al., 1985; Schum, 1996). Further,

given the previous results demonstrating that non-native talkers' speech is better understood by non-native listeners than by native listeners (Hayes-Harb, Smith, Bent, & Bradlow, 2008; Imai et al., 2005; Munro, Derwing, & Morton, 2006), it is possible that non-native listeners, who share an L1 background with the talkers, would benefit from the non-native clear speech enhancements more than native listeners do, not only in terms of intelligibility improvement but also in subjective measures of perception. Particularly, the native English listeners in the current study perceived higher-proficiency non-native talkers' *plain* speech to be easier to understand than their *clear* speech in the quiet listening condition, suggesting that higher-proficiency talkers' clear speech enhancements had a detrimental effect on perceived comprehensibility of their speech. However, these talkers' acoustic modifications may result in an improvement in perceived comprehensibility for native Mandarin listeners, who share the L1 background with the talkers. A future study may investigate how non-native talkers' proficiency level impacts different types of clear speech perception (e.g., intelligibility, subjective evaluations) for native and non-native listeners.

Finally, the current results demonstrated that native listeners' subjective evaluations differed for the speech produced by talkers of different L1 backgrounds and L2 proficiency levels. Specifically, the size of task-based differences in perception (i.e., the difference in the proportion of times the listeners chose the clear-style sentence for the effort vs. comprehensibility task) was much smaller for lower-proficiency talkers' speech compared to higher-proficiency talkers' and native talkers' speech. The results further indicated that detecting the increased effort in the clear-style sentences (as compared to the plain-style sentences) was more difficult for lower-proficiency talkers' speech than for native talkers'

and higher-proficiency talkers' speech. Here, it is important to point out that the current perception task was designed to examine the difference in listeners' perception between plain- and clear-style productions, not their perception of the talkers' speech more generally. That is, the small plain-to-clear difference in perceived talker effort for lower-proficiency talkers' speech does not suggest that listeners did not perceive lower-proficiency talkers' effort producing the clear speech. It is possible that listeners perceived that lower-proficiency talkers' *plain* speech was produced with a relatively high level of effort (perhaps with more effort than higher-proficiency talkers' and native talkers' speech). This may have resulted in the small difference in perceived effort between plain and clear speech for lower-proficiency talkers' speech. Examining listeners' subjective evaluation using a Likert scale (e.g., 1: speech is produced with the least effort; 9: speech is produced with the maximum effort) may be able to test the questions of how perceived degree of talker effort increases from plain to clear speech, in relation with whether foreign-accented speech is generally perceived to be produced with increased effort compared to native speech.

3.4.4. Conclusion

The current study examined multiple aspects of clear speech perception: intelligibility (Experiment 2A) and subjective evaluation of the speech (Experiment 2B). Specifically, Experiment 2A examined whether clear speech enhancements produced by native and non-native talkers of different proficiency levels result in a similar intelligibility improvement for native English listeners. Further, Experiment 2B examined whether native and non-native clear speech enhancements improved native listeners' subjective evaluation,

in terms of perceived degree of comprehensibility (how easy the speech is to understand) and talker effort (how hard the talker is trying to speak clearly). The results of Experiment 2A showed the effect of non-native talkers' target language proficiency level on their speech intelligibility. That is, higher-proficiency talkers' speech was generally more intelligible than lower-proficiency talkers' speech, and higher-proficiency talkers' clear speech resulted in a larger plain-to-clear speech intelligibility improvement than lower-proficiency talkers' clear speech did. The results of Experiment 2B demonstrated that talkers' clear speech enhancements improved listeners' subjective evaluation of talker effort more than comprehensibility across different talker groups' speech. However, lower-proficiency talkers' increased effort was more difficult to detect than that of higher-proficiency talkers and native talkers. Together, these results suggest that non-native talkers' ability to increase intelligibility of their speech improves as their target language proficiency develops. However, for both native and non-native talkers' speech, an improvement in intelligibility does not necessarily correspond to an improvement in subjective evaluations of the speech. A future investigation may examine which acoustic features of clear speech enhancements are responsible for different types of listeners' perception.

CHAPTER IV: PRODUCTION OF CONTEXTUALLY-RELEVANT SPEECH ENHANCEMENTS

4.1. Introduction

One important factor that contributes to successful speech communication is an individual's ability to speak more clearly when their listeners do not understand their speech. It has been widely demonstrated that talkers are able to enhance various features of their speech (e.g., by speaking more slowly, loudly, or by articulating sounds more clearly) to make their speech more understandable to their listeners (e.g., Picheny et al., 1986). While speech enhancement strategies used by native talkers have been examined in a variety of communication contexts (e.g., Hazan & Baker, 2011; Scarborough & Zellou, 2013), our understanding of non-native talkers' speech enhancement strategies is mostly limited to those examined in a context where they read materials as if talking to a hearing-impaired listener (e.g., Smiljanić & Bradlow, 2011). Particularly, though previous work has demonstrated that native talkers make targeted acoustic modifications to enhance characteristics of particular sound contrasts in a communicative task (e.g., Baese-Berk & Goldrick, 2009; Buz et al., 2016; Seyfarth et al., 2016), it is unknown how such targeted enhancements are implemented by non-native talkers of different proficiency levels. Thus, the current study examines native and non-native English talkers' ability to produce speech enhancements when the potential communication difficulty is implicitly signaled in the context. In a word-reading paradigm (Baese-Berk & Goldrick, 2009), talkers communicate target words (e.g., *cap*) to a listener when a phonetically similar minimal-pair neighbor (e.g., *cab*) either is or is not present in the

context. We examine acoustic characteristics of speech modifications made in different contexts, asking how talkers' native language status and non-native talkers' proficiency level impact the size of these modifications, as well as how the effects of talkers' target language experience on the contextually-relevant enhancements differ depending on the talkers' familiarity with the target sound contrast (i.e., a contrast that also exists in non-native talkers' native language vs. a contrast that does not).

4.1.1. Speech enhancements in different tasks

Previous studies have demonstrated that talkers are able to enhance various acoustic-phonetic features of their speech to make it more intelligible to their listeners (e.g., Uchanski, 2005). One way to examine talkers' speech enhancement strategies is to investigate clear speech. Clear speech is a speaking style that talkers use when they are aware that the listeners may have difficulty understanding them, possibly because the listeners have hearing impairments or are non-native listeners of the language (Smiljanić & Bradlow, 2009; Uchanski, 2005). Talkers' clear speech has often been examined by having them read the same set of materials twice: once in a plain- and once in a clear-speaking style (Bradlow & Alexander, 2007; Ferguson & Kewley-Port, 2002; Ferguson 2004; Granlund et al., 2012; Picheny et al., 1986; Rogers et al., 2010; Schum, 1996; Smiljanić & Bradlow, 2005, 2011). For the plain-speaking style, talkers are instructed to read the materials as if they are talking to someone familiar with their voice and speech patterns; for the clear-speaking style, talkers are instructed to read the same materials as if they are talking to a listener with a hearing loss or to a non-native listener of the language (e.g., Bradlow & Alexander, 2007; Smiljanić & Bradlow, 2011). These studies

have shown that in clear speech, native talkers of the language use a range of acoustic-phonetic modifications including a decrease in speaking rate (characterized by longer segments as well as longer and more frequent pauses), higher pitch (F0), wider F0 range, increased intensity, increased energy in the 1-3 kHz range of long-term spectra, as well as expanded vowel space (e.g., Bradlow et al., 2003; Liu et al., 2004; Moon & Lindblom, 1994; Picheny et al., 1986; Smiljanic & Bradlow, 2005). These modifications result in robust intelligibility gains for listeners of various characteristics, including hearing-impaired listeners as well as non-native listeners (e.g., Bradlow & Bent, 2002; Bradlow & Alexander, 2007; Krause & Braida, 2002; Liu et al., 2004; Picheny et al., 1985; Schum, 1996).

However, previous studies have demonstrated that speech enhancements are not uniform phenomena (Tuomainen & Hazan, 2018). That is, a talker's effort to enhance acoustic-phonetic characteristics of their speech can be implemented differently depending on what types of task they engage in to produce speech enhancements. For example, native talkers' speech enhancements elicited in read speech, using the instruction to speak clearly as if talking to someone who is hearing impaired, involved more extreme changes in some acoustic-phonetic characteristics (e.g., pitch range, speaking rate, vowel duration, vowel space) than speech enhancements elicited in spontaneous speech (e.g., elicited using a fill-in-the-blank worksheet: Scarborough & Zellou, 2013; using 'spot the difference' picture tasks with noise: Hazan & Baker, 2011). The read clear speech elicited in noise (i.e., the talkers read sentences while simultaneously listening to noise) involved even more extreme changes than the read clear speech produced in quiet (in terms of speaking rate, pitch mean, energy in the 1-3

kHz range: Gilbert et al., 2014). The presence of an actual listener also impacts talkers' acoustic modifications; when producing foreigner-directed speech, native talkers employ more extreme changes in durations and vowel space, when talking to an imagined non-native listener compared to when talking to a real non-native listener present in the room (Scarborough, 2007). Thus, as demonstrated in these studies, the characteristics of speech enhancements can be greatly influenced by the methods of eliciting the speech.

4.1.2. Speech enhancements in non-native speech

Aside from the influence of the speech elicitation task, factors that are related specifically to the talkers themselves could also impact the quality of speech enhancements. One such factor is the native language background of the talker. Despite the wealth of information on the speech enhancement strategies used by native talkers, investigations of strategies that are employed by non-native talkers are limited. However, previous studies examining non-native clear speech, as well as non-native speech production more broadly, suggest several factors that could possibly impact non-native talkers' ability to enhance intelligibility of their speech. Specifically, non-native talkers' ability to make speech enhancements may differ depending on the talkers' proficiency in the target language, as well as on the focus of the speech enhancements (i.e., which acoustic aspects of the speech the talkers are trying to enhance).

Several studies have shown evidence that non-native talkers' target language proficiency level impacts the types of clear speech strategies that they use. That is, non-native talkers of higher proficiency make clear speech adjustments that are similar to those made by native talkers in terms of modifications of vowel space, F0, intensity, and

temporal characteristics (e.g., word duration, articulation rate of sentences; Bradlow, 2002; Granlund et al., 2012; Chapter 2 of this dissertation). Further, the size of plain-to-clear speech modifications made by non-native talkers of lower proficiency is much smaller compared to those made by non-native talkers of higher proficiency (see Chapter 2). In terms of perceptual benefits, clear speech enhancements made by higher-proficiency non-native talkers result in a larger intelligibility improvement than those made by lower-proficiency talkers for native listeners (enhancements of English vowels: Rogers et al., 2010; enhancements of English sentences: Chapter 3 of this dissertation). Further, the size of intelligibility improvement resulting from higher-proficiency non-native talkers' clear speech is comparable to those resulting from native talkers' clear speech (Rogers et al., 2010; Smiljanić & Bradlow, 2011; Chapter 3 of this dissertation). These studies have suggested that talkers make various acoustic-phonetic modifications to their speech when explicitly asked to read materials clearly, and these modifications result in an intelligibility improvement for listeners. For non-native talkers, their proficiency level affects both the size of plain-to-clear speech acoustic modifications and the size of intelligibility improvement resulting from the modifications.

Though the studies examining non-native clear speech production have demonstrated that non-native talkers' proficiency level impacts the effectiveness of their clear speech, other studies have suggested that such an effect of proficiency level on non-native speech enhancements may differ depending on the focus of the acoustic modifications. Specifically, speech enhancement strategies that non-native talkers can apply from their native language may be easier than those that are different between their native and non-native languages, and talkers' proficiency level might affect the latter

more than the former. The similarity of strategies to enhance speech between native and non-native languages could differ depending on whether the acoustic modifications are made at the global or segmental level. Speech enhancements made at the global level increase the overall salience of the speech signal (e.g., making the speech easier to perceive in an adverse listening condition by speaking with decreased speaking rate, increased intensity and fundamental frequency, as well as expanded pitch range: Bradlow & Bent, 2002). Studies have shown that talkers of different languages use similar global strategies (e.g., Finnish & English: Granlund et al., 2012; Croatian & English: Smiljanić & Bradlow, 2005). However, segmental enhancement strategies are likely to be more specific to the sound system of the specific language because they involve modifications to characteristics of individual sounds. For example, native Croatian talkers manipulated vowel duration to a larger extent in Croatian clear speech than native English talkers did in English clear speech, reflecting the difference in the importance of duration cues between Croatian and English (Smiljanić & Bradlow, 2008a). Further, Finnish-English bilinguals manipulated voice-onset-times (VOTs) of initial stop consonants from plain to clear speech differently for Finnish /p/ and English /p/, possibly because English has a voicing counterpart /b/ though Finnish does not (Granlund et al., 2012). These studies suggest that the types of acoustic adjustments at the segmental level may be more language-dependent (i.e., specific to the sound system of the particular language) than adjustments at the global level. Thus, it may require more extensive experience with the sound system of the language to make appropriate adjustments at the segmental level compared to those at the global level.

Furthermore, particularly at the segmental level, characteristics of non-native segments may also influence the way non-native talkers make acoustic adjustments. Specifically, it has been widely documented that second language (L2) learners' native language influences their learning of L2 (e.g., Lado, 1957, Flege, 1995), and that L2 sounds that exist in learners' native language are easier to learn to produce compared to L2 sounds that do not (e.g., Brière, 1966; Vokic, 2008). In order to be able to produce L2 sounds that do not exist in learners' native language, learners need to establish a sound representation and the articulatory motor control to implement the representation in sound production (Brière, 1966; Flege, 1987). The difficulty learning L2 segments can be manifested in consonants (e.g., learning the English /r/-/l/ contrast as in *room* vs. *loom* for native Japanese talkers: Sheldon & Strange, 1982) and vowels (e.g., learning the English /i/-/ɪ/ contrast as in *sheep vs. ship* for native Mandarin talkers). It is possible that such ease and difficulty associated with non-native sound production extends to enhancements of non-native segments. That is, making segmental enhancements could be more challenging for non-native sounds that do not exist in the talker's native language than the non-native sounds that do. Especially for inexperienced non-native talkers (e.g., non-native talkers of lower-proficiency), their cue weighting strategies in perception and production may differ from those of native talkers (Imai et al., 2005), thus those lower-proficiency talkers may enhance cues that are irrelevant or detrimental to intelligibility improvement for non-native segments that they are not familiar with. Lower-proficiency talkers may also have less established articulatory motor control to produce non-native sounds than higher-proficiency talkers or native talkers. This is partly illustrated in the previous result that late English learners' segmental modifications led to a decreased

intelligibility for an English vowel that does not exist in their native language (Rogers et al., 2010). It has also been demonstrated that in order to signal an English coda voicing contrast (e.g., *bed* vs. *bet*), native Korean talkers manipulated the temporal dimension but not the spectral dimension of the preceding vowels (Choi et al., 2016). However, proficient non-native talkers are able to make segmental modifications to enhance non-native contrasts that do not exist in their native language (Hwang et al., 2015). These results suggest that making non-native acoustic adjustments that talkers are not used to making in their native language can be generally more difficult than those that they are familiar with from their native language experience; though more experienced, higher-proficiency talkers are able to enhance non-native contrasts that exist in their native language as well as those that do not. In other words, non-native talkers' proficiency level (higher- vs. lower-proficiency) could impact the effectiveness of speech enhancements to a larger extent when they are trying to enhance non-native sounds that do not exist in the talker's native language, compared to the non-native sounds that do.

Taken together, previous studies have demonstrated that production of speech enhancements can be affected by multiple factors, including the nature of the speech enhancement elicitation task (e.g., read speech vs. spontaneous speech) and talkers' native language background (e.g., native vs. non-native status). Specifically, for non-native talkers, their target language proficiency level influences the quality of clear speech modifications when explicitly asked to read materials clearly. Furthermore, talkers' experience with the target language sound system (e.g., native vs. non-native status, non-native talkers' proficiency level) may impact speech enhancements differently depending on the focus of the acoustic enhancements (e.g., enhancing global vs.

segmental features, enhancing familiar non-native sounds vs. unfamiliar non-native sounds). However, it is not clear how these factors impact non-native talkers' speech enhancements made in a more ecologically valid communication context. Given that the characteristics of native talkers' speech enhancements differ when the enhancements are elicited in a task similar to a naturalistic talker-listener interaction as compared to those elicited in read speech with explicit instructions to speak clearly (e.g., Hazan & Baker, 2011), it is possible that non-native talkers' speech enhancement behavior also differs in different tasks. Thus, in the current study, we examine how native and non-native talkers make speech enhancements in a task similar to a naturalistic talker-listener interaction. Specifically, we examine how these talkers accommodate their speech when the potential communication difficulty is signaled in the context implicitly, rather than when it is signaled by explicit instructions to read materials clearly as in clear speech.

4.1.3. Contextually-relevant speech enhancements

Previous studies have demonstrated that talkers are able to enhance acoustic features of the speech in a contextually-relevant way. For example, when a listener misunderstands a particular part of an utterance (e.g., a specific word), talkers selectively enhance that part of the utterance to correct the misunderstanding (e.g., Maniwa et al., 2009; Ohala, 1994; Oviatt et al., 1998; Schertz, 2013; Stent et al., 2008). Specifically, when native English talkers spoke to a simulated speech recognizer and received a feedback that the utterance was misunderstood (e.g., the talker says "pit" but the computer guesses "bit"), the talkers enhanced the misunderstood contrast by manipulating a relevant acoustic feature (e.g., VOTs of the /p/ and /b/) in the second

repetition (Schertz, 2013). This type of targeted error correction did not occur when the talker received an open-ended request for repetition (e.g., "???"). Such targeted segmental enhancements in response to listeners' feedback have also been found for a temporal aspect of a vowel contrast (English /i/-/ɪ/: Schertz, 2013) as well as for temporal and spectral aspects of English fricative contrasts (Maniwa et al., 2009).

Talkers make contextually-relevant speech enhancements in a communicative task even without feedback from the listener. For example, in a communicative task involving conveying information to a listener, native English talkers exaggerated differences in VOTs of English word-initial consonants (e.g., /p/-/b/) when a target word to communicate (e.g., *pill*) was displayed with another word that is minimally different (e.g., *bill*), compared to when it was not (Baese-Berk & Goldrick, 2009; Buz et al., 2014, 2016). Similar types of contextually-relevant hyperarticulation have been observed for an English word-final fricative voicing contrast (e.g., *dose* vs. *doze*: Seyfarth et al., 2016). Further, it has been suggested that contextually-relevant hyperarticulation of a target word may only occur in the context of other words that are sufficiently similar to the target word (e.g., one major phonological feature away: Kirov & Wilson, 2012). The researchers showed that native English talkers exaggerated VOTs of word-initial voiceless stop consonants (e.g., ***c****ap*) when a word differing in place of articulation (e.g., ***t****ap*) was contextually co-present, but not when a word differing by both place and manner of articulation (e.g., ***k****ilt* vs. ***h****ilt*) was contextually co-present (Kirov & Wilson, 2012). Though the investigation of such contextually-relevant hyperarticualtion has mostly been limited to native talkers' productions, one study demonstrated that highly proficient non-native talkers exaggerated a non-native contrast (e.g., /æ/-/ɛ/) when a

target word (e.g., *sat*) was placed next to a similar word (e.g., *set*) in a word-communication task (Hwang et al. 2015). Thus, these studies have demonstrated that experienced talkers (i.e., native talkers and highly proficient non-native talkers) are able to make targeted speech enhancements based not only on listeners' feedback but also on potential communication difficulty signaled in the context.

However, it is unknown how such contextually-relevant speech enhancements are made by non-native talkers of differing target language proficiency levels.

Given previous results demonstrating that higher-proficiency non-native talkers make larger clear speech modifications (e.g., in speaking rate, F0, intensity, vowel space) than lower-proficiency non-native talkers when explicitly asked to read materials clearly (Experiment 1 in Chapter 2), it is possible that non-native talkers' proficiency level impacts the degree of contextually-relevant speech enhancements as well. However, previous results have also suggested that non-native talkers' proficiency level may impact speech enhancements differently depending on the focus of the acoustic enhancements (e.g., enhancing global or segmental features, enhancing familiar non-native sounds or unfamiliar non-native sounds). That is, for relatively inexperienced talkers (e.g., lower-proficiency non-native talkers), making global enhancements that are similar across languages (e.g., slowing down the speech, speaking with a louder voice) may be easier than making segmental enhancements that are rather language-specific (e.g., increasing VOTs for word-initial voiceless stop consonants). Similarly, enhancing familiar non-native sounds (sounds that exist in their native language) may be easier than enhancing unfamiliar non-native sounds (sound that do not exist in their native language). Thus, it is possible that non-native talkers' proficiency level influences contextually-relevant speech

enhancements to a larger extent at the segmental level compared to the global level. Specifically, the effect of proficiency level may be larger when talkers are trying to enhance non-native sounds that do not exist in their native language compared to non-native sounds that do. Therefore, in order to better understand how native and non-native talkers make speech enhancements in a contextually-relevant way, we examine whether talkers' target language experience (talkers' native language background, proficiency level) affects the degree of speech enhancements, as well as whether the effect of talkers' target language experience differs depending on the focus of acoustic enhancements (global and segmental levels; for familiar vs. unfamiliar non-native sounds).

4.1.4. Current study

In the current study, we examine acoustic characteristics of contextually-relevant speech enhancements produced by native English talkers and non-native English talkers of higher- and lower-proficiency. We use a word-reading paradigm that has been shown to elicit contextually-relevant speech enhancements (e.g., Baese-Berk & Goldrick, 2009; Buz et al., 2014, 2016; Kirov & Wilson, 2012; Seyfarth et al., 2016). In this task, participants interact with a simulated listener. In each trial, three words appear on the screen. One target word is highlighted on the talker's screen, then they produce the word so that their listener would click on the same word on their own screen. In one type of trials (Context conditions), a target and its minimal-pair neighbor are presented on the screen with a filler as the third word (e.g., ***pill****, bill, send*), so the talker has to be sure that their listener would not confuse the target and the minimal-pair neighbor. In another type of trials (No Context conditions), the target is presented with two fillers (e.g., ***pill****, chair,*

send), so there is no potential for the listener to confuse the target with its minimal-pair neighbor. Using this word-reading paradigm, we ask how talkers' target language experience (native vs. non-native status; non-native talkers' proficiency level) affects the contextually-relevant speech enhancements. Specifically, we ask whether the effect of talkers' language experience differs depending on the type of acoustic enhancements examined (i.e., enhancements at the global and segmental levels; enhancements of non-native sounds that exist in talkers' native language and non-native sounds that do not).

In the current experimental paradigm, talkers instruct the simulated listener which word to choose using the carrier phrase, *"Click on the TARGET now" (e.g., "Click on the pill now."*). We examine how native and non-native talkers' contextually-relevant enhancements (i.e., difference between productions of No Context vs. Context conditions) are manifested at the global level: at the levels of the entire phrase and the target word. Though the potential communication difficulty signaled in the context concerns a particular segmental contrast (e.g., *pill* vs. *bill*) embedded in the target word, talkers' attempt to exaggerate the contrast in Context conditions (as compared to No Context conditions) could be manifested at the global level as well. That is, we may observe some contextually-relevant enhancements at the global level, including increased duration, higher fundamental frequency, and increased amplitude, which are typically found in clear speech (e.g., Bradlow, 2002, Bradlow et al., 2003; Granlund et al., 2012; Picheny et al., 1986; Chapter 2 of this dissertation). However, it is also possible that talkers' enhancements at the global level may be overall less robust as compared to those typically seen in clear speech because the source of potential confusion signaled in the context in the current study is more targeted (i.e., to a sound contrast in the target word),

rather than the clear speech instructions to speak clearly for a hearing-impaired listener. Furthermore, given that global enhancement strategies are used similarly across languages (e.g., Granlund et al., 2012; Smiljanić & Bradlow, 2005), it is possible that the degrees of enhancements at the global level, if any, may not differ for productions of talkers with different levels of target language experience (i.e., native English vs. non-native talkers, higher-proficiency vs. lower-proficiency non-native talkers). We examine these hypotheses in terms of duration, mean fundamental frequency, and mean intensity, which are the features typically examined in previous clear speech studies (e.g., Bradlow et al., 2003; Picheny, 1986), at the phrase level (i.e.,*"Click on the pill now"*) as well as at the target word level (i.e., *pill*).

The current study also examines whether non-native talkers are better able to manipulate acoustic features that enhance a non-native segmental contrast that also exists in their native language (henceforth, L1L2 contrast) than those features that enhance a contrast that does not exist in native language (henceforth, L2-only contrast). Given previous findings that non-native talkers are better able to manipulate an acoustic feature that they are used to manipulating in their native language than a feature that they are not used to manipulating (Choi et al., 2016), it is possible that non-native talkers of higher- and lower-proficiency are better able to enhance an L1L2 contrast than an L2-only contrast. Further, given that late English learners' clear speech enhancements resulted in increased intelligibility for English vowels that exist in their native language, while they led to decreased intelligibility for an English vowel that does not exist in their native language (Rogers et al., 2010), it is possible that lower-proficiency talkers' contextually-relevant segmental enhancements may be much less effective for an L2-only contrast

than for L1L2 contrast. Thus, the effect of non-native talkers' proficiency level (higher- vs. lower-proficiency) on segmental speech enhancements may be larger when talkers are trying to enhance an L2-only contrast than an L1L2 contrast.

We explore these questions for non-native consonant and vowel contrasts for native Mandarin learners of English. In terms of consonant contrasts, we use the English /p/-/b/ contrast in word-initial position (e.g., *pill* vs. *bill*) as the L1L2 consonant contrast, and the /p/-/b/ contrast in word-final position (e.g., *cap* vs. *cab*) as the L2-only consonant contrast. In Mandarin Chinese, voicing distinctions for stop consonants do not occur in word-final position, though they do occur in other word positions (Cheng, 1973; Flege, Munro, & Skelton, 1992; Hayes-Harb et al., 2008; Howie, 1976). Given that learning of a particular L2 phoneme is influenced by the phonetic environment in which the sound occurs (e.g., the sound occurring in word-initial, word-medial, or word-final position: Vokic, 2008), learning to produce the voicing distinction in English may also be impacted by the structural position of the contrast. That is, the voicing distinction may be implemented differently across different structural positions, and for native Mandarin talkers, the distinction may be easier to implement in word-initial position than in word-final position. In fact, native Mandarin talkers have been shown to neutralize the English word-final voicing distinction for stop consonants by devoicing the voiced consonants (Flege et al., 1992).

Here, the underlying assumption for choosing these materials should be explicitly stated; we assume that the set of knowledge and articulatory control that learners need to acquire differs for the English /p/-/b/ contrast in different positions. That is, in order to enhance the /p/-/b/ contrast in word-initial position vs. word-final position, talkers need to

have acquired position-specific phonetic details of the contrast, and implement the details via appropriate articulatory control. This assumption is in line with exemplar-based models (e.g., Goldinger, 1996; Johnson, 1997; Pierrehumbert, 2003), which claim that talkers' sound representations are sensitive to phonetic environments in which they occur, and allophonic variations of a particular phoneme are stored as part of the representations, as opposed to a rule-based phonological system (e.g., Chomsky & Halle, 1968), where allophones are not stored as separate representations because they are predictable from phonemes via rule application. Particularly, an exemplar-based approach could account for different acquisition patterns of phonemes across different positions (e.g., Jusczyk, Goodman, & Baumann, 1999; Zamuner, 2006; Shea & Curtin, 2011). For example, infants showed different discrimination performance for a sound contrast in word-initial vs. word-final position (Zamuner, 2006). Further, Shea and Curtin (2011) demonstrated that, when implementing stop-approximant allophonic alternation (e.g., /b, d, g/-/β, ð, ɣ/), native Spanish talkers and experienced learners of Spanish showed more gradient use of two cues, the position in the word (e.g., word-initial or word-medial) and stress (e.g., stressed or unstressed syllable), as compared to less experienced learners. This suggests that learners with greater target language experience have more gradient representations of L2 sounds, storing fine-grained phonetic details rather than just categorical differentiation of the sounds. Given these studies, it is possible that, when enhancing a phonological contrast in different positions (e.g., English /p/-/b/ in word-initial vs. word-final positions), talkers with greater target language experience are better able to differentiate acoustic modification strategies to reflect position-specific

aspects of the contrast. This could entail that being able to enhance phonetic details of allophonic variations may be part of knowing the target language sound system.

In terms of vowels, Mandarin Chinese has a relatively large number of diphthongs, including /ei/ and /ai/ (Gottfried & Suiter, 1997; Lai, 2010), and these diphthongs are also distinguished phonemically in English (e.g., *say* vs. *sigh*). However, in Mandarin Chinese, vowel tenseness is not a major feature used to distinguish vowels (Lai, 2010). Particularly, though a tense English vowel /i/ exists in Mandarin, a lax vowel /ɪ/ does not (Cheng, 1973; Flege, Bohn, & Jang, 1997; Howie, 1976). Thus, we consider the English /ei/-/ai/ contrast (e.g., *say* vs. *sigh*) as the L1L2 vowel contrast and the English /i/-/ɪ/ contrast as the L2-only vowel contrast for native Mandarin talkers. Using these four types of contrasts (i.e., L1L2 consonant/vowel contrast, L2-only consonant/vowel contrast), we examine whether the effect of talkers' native language status (native English vs. native Mandarin talkers) as well as non-native talkers' target language proficiency level (higher- vs. lower-proficiency) impacts contextually-relevant enhancements of L2-only contrasts to a larger extent than those of L1L2-only contrasts.

4.2. Methods

4.2.1. Participants

Thirty-four native English talkers (29 females, 5 males; age range 18 - 22 years, mean = 19 years) and 44 native Mandarin talkers (34 females, 10 males; age range = 19 - 35 years, mean = 24.7 years) participated. None of the native Mandarin talkers reported a history of speech or hearing impairment. Though 3 native English talkers reported a history of speech or hearing impairment (i.e., childhood speech therapy), their data did not deviate

from other talkers' data. Thus, their data were included in the analyses. Native English talkers were recruited from the Linguistics and Psychology Human Subject Pool at the University of Oregon, and were given partial course credit for their participation. Native Mandarin participants were either paid or given partial course credit for their participation.

Non-native English talkers were recruited from three difference sources. Ten talkers were students at the American English Institute (AEI) at the University of Oregon, which provides academic English support for international students before they enter the university as matriculated students. Eighteen talkers were undergraduate students, and 16 talkers were graduate students at the University of Oregon. For the purpose of data analysis, the 44 non-native English talkers recruited from these sources were classified into lower- and higher-proficiency talkers, based on their most recent English proficiency test score. That is, the talkers who had reported a Test of English as a Foreign Language (TOEFL) score of lower than 72, which was classified to be 'Below Low-Intermediate' or 'Low-Intermediate'[3], were categorized as lower-proficiency native Mandarin (Native Mandarin-Low) talkers (n = 22). The talkers who had reported a TOEFL score of higher than 72, which was classified to be ''High-Intermediate' or 'Advanced', were categorized as higher-proficiency native Mandarin (Native Mandarin-High) talkers (n = 22)[4]. Table 4.1 provides information regarding non-native (native Mandarin) talkers' English learning background and proficiency.

Additionally, 190 native English listeners (71 females, 117 males, 2 Others; age

[3] The classification was done based on the proficiency level classification provided by TOEFL. https://www.ets.org/toefl/institutions/scores/interpret/

[4] Of the 44 native Mandarin talkers, 7 talkers did not report their TOEFL score, but reported their International English Language Testing System (IELTS) score; for these talkers, their IELTS score was converted to a TOEFL score based on the conversion table provided in Educational Testing Service (2010). When a talker provided neither their TOEFL nor IELTS score (n = 4), their perceived accentedness score (see Section 4.2.3) was used as a proxy for their proficiency level.

range = 23 - 74 years, mean = 36.7) participated in the foreign accent rating task evaluating the accentedness of the talkers. The accentedness ratings were used as another measure to characterize the native Mandarin talkers' English proficiency. None of the listeners provided the speech samples. The listeners were recruited via Amazon Mechanical Turk (https://www.mturk.com/).

Table 4.1. Non-native English (native Mandarin) talkers' English learning background and proficiency. Mean and range (in parenthesis) are shown for lower-proficiency (Native Mandarin-Low) and higher-proficiency (Native Mandarin-High) talkers.

Proficiency group	Age	Age of onset for English speaking	Years of formal English training	Length of US residence in months	TOEFL score[2]
NativeMandarin-Low (n=22)	23.2 (19-35)	14.1 (6-23)	10 (5-25)	30.3 (6-96)	55.1 (35-70)
NativeMandarin-High (n=22)	26.1 (19-35)	11.1 (5-23)	13 (6-24)	51.9 (6-144)	89.6 (72-108)

4.2.2. Production experiment

In the experiment session, the participants first completed a context-production task, followed by a sentence-reading task. In the context-production task, target words were 80 English monosyllabic words (see Appendix B for the list of target words). Forty targets consisted of 20 minimal pairs that contrasted consonants and other 40 targets consisted of 20 minimal pairs that contrasted vowels. Of those targets, 20 consonant targets and 20 vowel targets (i.e., 10 minimal pairs each) contained a phonemic contrast that exists in both L1 Mandarin and L2 English for the native Mandarin talkers (henceforth, L1/L2 consonant targets and L1L2 vowel targets). The 20 L1L2 consonant targets contrasted /p/ and /b/ in word-initial position (e.g., *peer* vs. *beer*). The 20 L1L2 vowel targets contrasted /ai/ and /ei/ (e.g. *light* vs. *late*). Other 20 consonant targets and 20 vowel targets (i.e., 10 minimal pairs each) contained a phonemic contrast that exists in English but not in Mandarin (henceforth,

L2-only consonant and L2-only vowel targets). The 20 L2-only consonant targets contrasted /p/ and /b/ in word-final position (e.g., *cap vs. cab*). The 20 L2-only vowel targets contrasted /ɪ/ and /i/ (e.g., *sick* vs. *seek*). The 20 fillers were English monosyllabic words that did not contain the target contrasts.

The context-production task was modeled after the word-reading paradigm used in Baese-Berk and Goldrick (2009) and Buz et al. (2016). The task was administered using E-Prime (Schneider, Eschman, & Zuccolotto, 2002) with Sennheiser HD 202 II headphones and a standing microphone in a sound booth. A simulated partner paradigm was used, where the participant was told that they would interact with a partner online, but it was actually a computer that provided responses (Buz et al., 2016). In order to familiarize the participant with the role that their partner would later play in the context-production task, the task began with 5 perception trials, where the participant saw three words on the screen and heard a male native English speaker say, *"Click on the ___ now."* The participant was instructed to choose one of the three words that the speaker said as quickly and accurately as possible. The 5 perception trials consisted of 2 trials that had a consonant or a vowel minimal pair (i.e., *star, loom,* ***room****; sat,* ***set****, oil*), which were not the target contrasts in the context-production task. Other 3 perception trials did not have minimal pairs. The 2 trials with minimal pairs were included in the perception trials in order to familiarize the participant with the potential difficulty to choose the correct word that their partner might experience in the following part of the context-production task.

After the perception trials, the participants were told that they would now play the role of giving instructions to a partner. Participants read through a short task description with text and images describing their role and their partner's role (following Buz et al.,

2016). Then, they saw the screen showing the message that the computer was searching for their partner online. After a few seconds, participants were told that they had been matched with a partner online and proceeded to the production part of the task.

In the context-production task, on each trial, the participant was presented with three words on the screen. Once they saw all three words, they were instructed to press the space key. Then, one of the three words was highlighted by the computer, and the participant was asked to produce the highlighted word (i.e., the target) in the phrase, *"Click on the TARGET now"*, for a partner, who could also see the three words but did not know which of the three was the target. After a various amount of delay (i.e., 800ms, 1200ms, 2000ms, 4000ms, 6000ms; randomly assigned for items), the participant was informed that their partner made a response but was not informed which word the partner selected, then the trial advanced. In Context conditions, both the target and its minimal pair neighbor were presented on the screen with a filler as the third word (e.g., ***peer,*** *beer, town*). In No Context conditions, the target was presented with two fillers (e.g., *soft,* ***peer,*** *noon*). In Filler trials, three fillers were presented.

After three practice Filler trials, each participant completed 60 test trials. Of the 60 test trials, 20 were in Context conditions (i.e., 5 trials each with L1L2 consonant targets, L1L2 vowel targets, L2-only consonant targets, L2-only targets), and 20 were in No Context conditions (i.e., 5 trials each with L1L2 consonant targets, L1L2 vowel targets, L2-only consonant targets, L2-only vowel targets). Other 20 trials were Filler trials. As shown in Appendix B, some minimal pairs were presented in Context conditions and other minimal pairs were presented in No Context conditions. In order to ensure that one participant did not produce both targets in a minimal pair (e.g., *pad*, *bad*), participants were

divided into two groups. Everyone produced a target from all minimal pairs (i.e., 40 targets) in addition to 20 fillers, but which target of a minimal pair the participant produced was different depending on the group. For example, a participant in one group produced an L1L2 consonant target *pad* in the Context condition, and another participant in a different group produced its minimal pair *bad* in the Context condition. For each type of minimal pair (e.g., pairs with the L1L2 consonant contrast), a participant produced 5 targets with one target phoneme (e.g., word-initial /p/) and 5 targets with the other target phoneme (e.g., word-initial /b/). The combination of a target phoneme (e.g., word-initial /p/ or /b/) and type of trial (Context or No Context conditions) was counter-balanced across participants. At the end of the context-production task, participants answered a few questions regarding the task (e.g., how fast their partner responded; Buz et al., 2016).

Following the context-production task, participants completed a sentence reading task. In the sound booth, participants recorded 15 English sentences selected from the Revised Bamford-Kowal-Bench Standard Sentence Test (BKB sentences: Bamford & Wilson, 1979). The list of the BKB sentences are provided in Appendix A. The sentences were displayed on the computer screen one at a time, and the participants read the sentences at their own pace. The recordings from the sentence-reading task were used as materials for the accentedness judgment task in order to assess non-native talkers' perceived accentedness as a part of their English proficiency measure (Hayes-Harb et al., 2008; Imai et al., 2005; Stibbard & Lee, 2006). After the recording, participants completed a language background questionnaire and other proficiency measuring tasks. The entire session lasted approximately one hour.

4.2.3. Accentedness judgment task

The recordings of the sentence-reading task were evaluated for foreign accentedness by native English listeners. Because the primary purpose of the accentedness judgement task was to assess non-native talkers' English proficiency, sentences produced by all 44 non-native talkers were evaluated in the task. Sentences produced by 6 native English talkers were also included in the accentedness judgement materials to provide the basis of accent comparison for the listeners (Smiljanić & Bradlow, 2011). The sentence recordings were segmented into individual sentence-length files. The sentence-length files of the practice sentences were RMS normalized to 65dB SPL. In the accentedness judgement task, conducted via Qualtrics, the listeners were told that they would listen to English sentences and evaluate the foreign accent of the speech. In each trial, listeners evaluated an English sentence without noise and were instructed to rate the accentedness of the speech on a scale of 1 ("a native speaker of English") through 9 ("an extremely strong foreign accent"; Munro & Derwing, 1995a). In order to prevent the accentedness ratings from being influenced by the intelligibility of the speech, the transcript of the sentence was displayed while the listeners were listening to the speech (Gittleman & Van Engen, 2018). Each sentence could not be played more than once, but there was no time limit for responding. Each listener evaluated 60 sentences: 10 unique sentences produced by 6 talkers (i.e., 2 native English, 2 higher-proficiency non-native talkers, 2 lower-proficiency non-native talkers). Each listener evaluated either 6 female talkers or 6 male talkers. Each non-native talker was evaluated by 13-21 listeners.

Foreign accent ratings were z-score normalized for each listener in order to account for variation in the listeners' use of the nine-point rating scale. Figure 4.1 shows accent

ratings by talker group (Native English, Native Mandarin-High, and Native Mandarin-Low). In order to examine whether the accent ratings differed for different taker groups, one-way ANOVA was carried out with z-scored ratings as the dependent variable. The results indicated that the ratings differed significantly by the talker group [$F(2, 47) = 170.7$, $p < .001$]. The post-hoc Tukey comparisons confirmed that the all the group comparisons were significant: Native English vs. Native Mandarin-High, Native English vs. Native Mandarin-Low, and Native Mandarin-High vs. Native Mandarin-Low ($p < .001$ for all). Thus, these results showed that, for the 44 non-native talkers, higher-proficiency talkers were perceived to be less accented than lower-proficiency talkers.

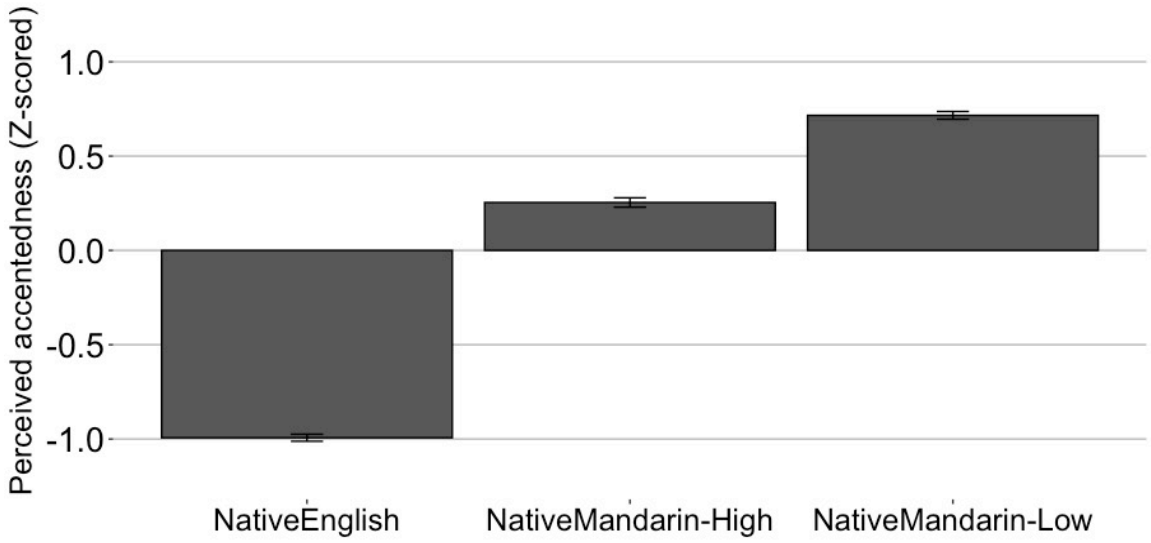

Figure 4.1. Z score-normalized accentedness ratings plotted by talker group. Error bars represent 95% confidence interval of the mean.

4.2.4. Acoustic analysis

In the context-production task, one talker produced a total of 40 target words. That is, each talker produced 5 targets in Context conditions and 5 targets in No Context conditions for each of the four types of contrasts: 10 L1L2 consonant targets (/p/-/b/ in

word initial position), 10 L2-only consonant targets (/p/-/b/ in word final position), 10 L1L2 vowel targets (/ai/-/ei/), 10 L2-only vowel targets (/i/-/ɪ/). Thus, there was a total of 3120 items (78 talkers x 40 targets). Productions with mispronunciation and disfluency (e.g., repetition) were excluded from the acoustic analysis. The excluded items were 46 out of 3120 items (i.e., less than 1% of the total number of items). Thus, we analyzed a total of 3074 items: 773 L1L2 consonant targets, 768 L2-only consonant targets, 762 L1L2 vowel targets (/ai/-/ei/), 771 L2-only vowel targets (/i/-/ɪ/). Praat (Boersma & Weenink, 2001) was used for all measurements of the acoustic analysis.

4.2.4.1. Global measurements

Previous research has demonstrated that global speech enhancements, which enhance the overall salience of the speech signal (Bradlow & Bent, 2002), involve a range of acoustic-phonetic modifications in terms of temporal characteristics, fundamental frequency, and intensity of the speech (Bradlow et al., 2003; Liu et al., 2004; Picheny et al., 1986; Smiljanić & Bradlow, 2005). Specifically, in order to make the speech easier to hear, talkers speak more slowly, speak with higher pitch (F0), as well as with increased intensity. Following these findings, we examine three aspects of global speech enhancements: duration, mean F0, and mean intensity. We examine these features at the phrase-level (i.e., *"Click on the pill now"*) as well as at the target word-level (e.g., *pill*). Phrase durations were measured from the phrase onset until when no visible speech was present in the waveform or spectrogram, which were confirmed acoustically. Target word durations were measured from the word onset (i.e., after the periodicity of the "the" as in *"Click on the pill now"* ended) until when no visible speech was present in the waveform or

spectrogram, or until when the nasal consonant of the "now" in the carrier phrase started. For each phrase and target word, a Praat script was run to measure duration (in msec.) as well as to calculate mean F0 (in Hertz), and mean intensity (in dB). In order to account for individual variability (e.g., some talkers speak faster than others), duration, mean F0, and mean intensity were transformed using the min-max scaling procedure (Gerstman, 1968; Kallay & Redford, 2018). That is, for example, phrase duration of a particular production was normalized using the talker's minimum and maximum values of the phrase duration, so that all the values are within the range of 0 (minimum value of that talker) to 1 (maximum value of that talker). These scaled values of duration, mean F0, and mean intensity were used to examine within-talker variations of these values based on different production conditions (No Context vs. Context) as well as different types of target contrasts (L1L2 vs. L2-only contrasts in target words). We also analyzed raw phrase- and word-durations to examine whether they differ for different Talker Groups' productions (Native English, Native Mandarin-High, Native Mandarin-Low talkers).

4.2.4.2. Segmental measurements

4.2.4.2.1. Consonant targets

For L1L2 consonant targets (i.e., /p/- and /b/-initial words), the voice onset times (VOTs) of /p/ and /b/ were manually annotated and measured (following Baese-Berk & Goldrick, 2009; Buz et al., 2016). VOT was measured from the beginning of the stop burst on the waveform to the onset of the following vowel, which was defined as the left zero crossing of the first complete periodic cycle (Baese-Berk & Goldrick, 2009). Although it is possible that talkers produced /b/-initial targets (e.g., *bill*) with negative VOTs, pre-voicing

was not measured because the word preceding the targets was *"the"* (as in *"Click on the bill now").* That is, it could be ambiguous when deciding whether the voicing before the stop burst was pre-voicing or voicing from *"the"*. Thus, the VOT was measured from the burst to the beginning of voicing for all the files. In order to normalize VOTs for the global speech rate (e.g., slower talkers' VOTs may be inherently longer than faster talkers' VOTs), VOT of each consonant (e.g., /p/) was divided by its word duration (e.g., *pill*; Hirata & Whiton, 2005). These normalized VOTs (i.e., the ratio of the VOT to the duration of the whole target word) were used in the L1L2 consonant segmental analyses.

For L2-only consonant targets (i.e., /p/- and /b/-final words), we annotated and measured the durations of the vowels preceding the target consonants. Vowel durations are one of the primary cues that native and non-native English talkers use to distinguish voiced coda consonants from voiceless coda consonants; vowels before voiced consonants are often longer than those before voiceless consonants (e.g., Chen, 1970; Choi et al., 2016; Goldrick, Vaughn, & Murphy, 2013; Hayes-Harb et al., 2008; Hogan & Rozsypal, 1980; Hwang et al., 2015; Raphael, 1972; Seyfarth et al., 2016). Thus, we measured preceding vowel duration for each L2-only consonant target; vowel duration was measured from the first zero-crossing of a complete periodic cycle to the offset of the last complete periodic cycle in the waveform, with reference to the F2 energy in the spectrogram (Choi et al., 2016; Idemaru & Guion-Anderson, 2010). Duration of a vowel in a target word (e.g., *c**a**p*) was then divided by the duration of the whole word in order to calculate the speech-rate normalized vowel duration. Additionally, previous studies have also shown that talkers use voicing of the target consonants to distinguish voiced and voiceless coda consonants; voiced consonants have longer voicing durations/larger voicing proportions than voiceless

consonants (Hayes-Harb et al., 2008; Hwang et al., 2015; Nittrouer, 2004; Seyfarth et al., 2016). Following these studies, we used voicing proportions of the target consonants (i.e., C2: word-final /p/ and /b/) as another measure to examine production patterns of the coda voicing contrast. To measure C2 voicing proportions, we used Praat to count the total number of voiced 10ms frames in each C2 (Seyfarth et al., 2016). Because the absolute durations of some C2 closures were too short for the Praat script to calculate the voicing proportions, we used C2 closure + burst (release) durations as C2 durations for all L2-only consonant targets (Idemaru & Guion-Anderson, 2010). C2 durations were measured from the offset of the preceding vowel to the end of the stop release, defined as the point where the noise abruptly decreased in intensity in most frequency ranges in spectrogram, which was also confirmed in the waveform (Hwang et al., 2015). Of the 768 L2-only consonant contrast items, we found that 26 items were not released (i.e., about 3%). Because we could not reliably measure the durations of the target consonant closure or burst without the release, we excluded these 26 items from the analyses of normalized vowel durations and C2 voicing proportions.

4.2.4.2.2. Vowel targets

For L1L2 vowel targets (i.e., words with the /ai/-/ei/ contrast), we examined two features: duration of the vowel and spectral (F1 and F2) values at the initial state. Vowels were segmented using the same procedure described above. Duration of a particular vowel (e.g., *l**a**te*) was then divided by the word duration to calculate the normalized vowel duration. Further, based on the finding that English diphthongs /ai/ and /ei/ differ in formant values at the vowel onset (i.e., /ei/ has lower F1, corresponding to higher tongue

position, and higher F2, corresponding to more advanced tongue position, than /ai/ at onset: Lee, Potamianos, & Narayanan, 2013), we measured F1 and F2 values at the initial state of each vowel. In order to obtain formant measures, a Praat script was run to measure F1 and F2 values at 30% of the vowel to avoid the possible influence of the word-initial consonants. These initial F1 and F2 values were then z-score normalized to control for individual differences (i.e., Lobanov method: Nearey, 1977; Thomas & Kendall, 2015).

For L2-only vowel targets (words with the /i/-/ɪ/ contrast), we measured two features: duration and spectral (F1 and F2) values. Native English talkers distinguish /i/- and /ɪ/ primarily with spectral quality, but they also distinguish them with duration (e.g., Bohn & Flege, 1992; House, 1961; Tsukada, 2009). Specifically, /i/ has lower F1 (higher tongue position), higher F2 (more advanced tongue position), as well as longer durations as compared to /ɪ/ (e.g., Hillenbrand, Getty, Clark, & Wheeler, 1995; Strange, Bohn, Trent, & Nishi, 2004; Strange, Bohn, Nishi, & Trent, 2005). Thus, for /i/ and /ɪ/, duration and formant (F1 and F2) values were examined. Vowels were segmented using the same procedure described above. Duration of a particular vowel (e.g., *s**ee**k*) was then divided by the word duration to calculate the normalized vowel duration. In order to obtain formant measures, a Praat script was run to measure F1 and F2 values at the midpoint (50%) of the vowel (following Hwang et al., 2015). Midpoint F1 and F2 were then z-score normalized to control for individual differences.

4.3. Results

We conducted acoustic analyses of the talkers' productions at several different levels: phrase-level (i.e., analysis of the whole phrase, *"Click on the ___ now"*), word-level

(i.e., analysis of the target words: e.g., *pill, bill*) and segmental-level (i.e., analysis of the target contrasts: e.g., /p/-/b/ contrast in word-initial position). Here, we present these results separately for targets that contained consonant contrasts (i.e., L1L2 consonant contrast: /p/-/b/ in word-initial position, L2-only contrast: /p/-/b/ in word-final position) and vowel contrasts (i.e., L1L2 vowel contrast: /ai/-/ei/, L2-only vowel contrast: /i/-/ɪ/). Thus, this results section contains the following components: analyses of the consonant targets (phrase-level, word-level, segmental-level) and analyses of the vowel targets (phrase-level, word-level, segmental-level).

4.3.1. Consonant targets: Global (phrase-, and word-level) analyses

For the analyses of the items that contained consonant target contrasts (i.e., L1L2 consonant contrast: /p/-/b/ in word-initial position, L2-only consonant contrast: /p/-/b/ in word-final position), we examined three aspects at the phrase- and the word-levels: duration, mean F0, and mean intensity. Specifically, we examined whether raw durations of the whole phrases and target words differed for productions of different Talker Groups (e.g., whether Native Mandarin-Low talkers produced the phrases with longer durations than Native Mandarin-High talkers). Further, we examined scaled durations, scaled mean F0, and scaled mean intensity to analyze whether talkers made difference in these measures depending on the Type of the contrast of the target word (L1L2, L2-only) and Condition (No Context, Context), as well as whether the effects of these factors differed for different Talker Groups.

In the linear mixed-effects regression models used to analyze these features, fixed effects were Type (L1L2, L2-only), Condition (Context, No Context), and Talker Group

(Native English, Native Mandarin-High, and Native Mandarin-Low); different combinations of these fixed effects were included in different models (see the description of each model below, as well as the model syntax in each table). Type was contrast coded to compare between the L1L2 contrast (i.e., /p/-/b/ in word-initial position: .5) and the L2-only contrast (i.e., /p/-/b/ in word-final position: -.5). Condition was contrast coded to compare between Context (.5) and No Context (-.5) conditions. Talker Group was contrast coded to compare between Native English and Native Mandarin-High talkers (.5, -.5, 0) and between Native Mandarin-High and Native Mandarin-Low talkers (0, .5, -.5). Models also included the maximal random effects structure that would converge, which included random intercepts for talker and item. P-values were calculated based on Satterthwaite approximations (Luke, 2017), using the lmerTest package for R (Kuznetsova, Brockoff, & Christensen, 2016). When explaining the results of the models, we only interpret the aspects that are relevant to the questions asked in each analysis; see the tables with model summaries for the full results of each model.

4.3.1.1. Phrase-level analyses

4.3.1.1.1. Duration

The left panel in Figure 4.2 shows the raw durations (msec.) of the phrase, *"Click on the ___ now.",* by Talker Group (Native English, Native Mandarin-High, Native Mandarin-Low). The linear mixed-effects regression model, used to analyze raw phrase durations as the dependent variable, included Talker Group (Native English, Native Mandarin-High, or Native Mandarin-Low) as a fixed factor (see Table 4.2 for the model syntax and summary of the results). The model showed significant effects of Talker Group

comparisons: Native English vs. Native Mandarin-High (β = -341.49, t = -2.98, $p < .01$), Native Mandarin-High vs. Native Mandarin-Low (β = -283.48, t = -2.23, $p < .05$). Thus, the results showed that Native Mandarin-Low talkers produced the phrase with longer durations than Native Mandarin-High talkers did; Native Mandarin-High talkers produced the phrase with longer durations than Native English talkers did.

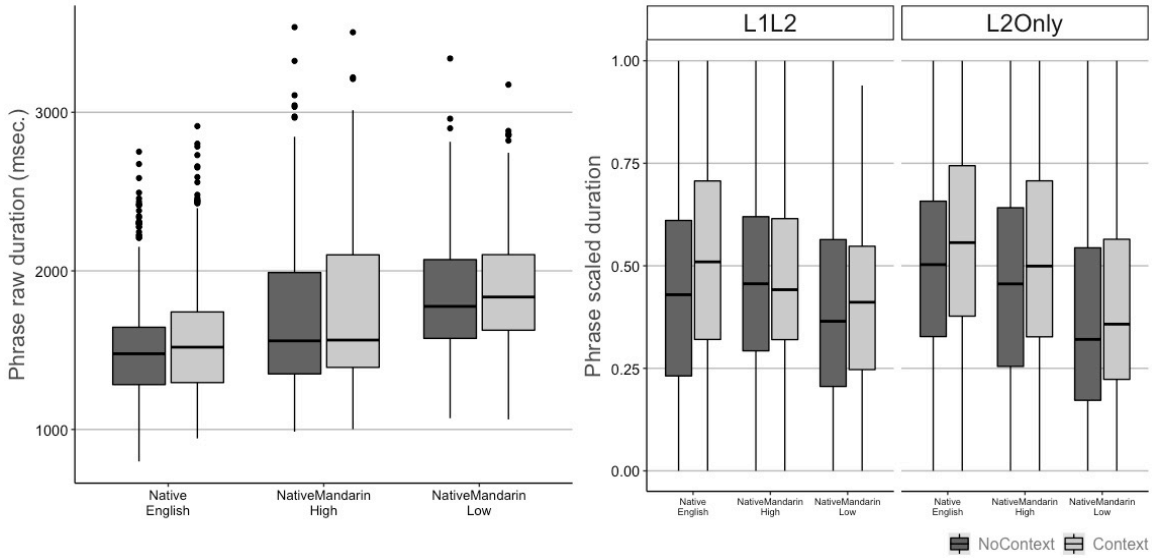

Figure 4.2. Durations of phrases containing consonant targets for different talker groups (Native English, Native Mandarin-High, and Native-Mandarin-Low talkers) in different conditions (No Context and Context): raw durations (Left panel), scaled durations (for phrases containing L1L2 or L2-only contrasts: Right panel).

The right panel in Figure 4.2 shows the scaled phrase durations by Talker Group, Type (L1L2, L2-only), and Condition (No context, Context). The linear mixed-effects regression model, used to analyze scaled phrase durations as the dependent variable, included Talker Group, Type, and Condition, as well as interactions among these factors as fixed effects (see Table 4.2 for the model syntax and summary of the results). Because we are interested in whether talkers made differences in their phrase durations depending on the Type of the contrast of the target word (L1L2, L2-only) and Condition (No Context, Context), as well as whether the effects of these factors differed for different Talker

Groups' productions, we interpret the effects of Type and Condition, and the interactions between these factors and Talker Group. In the model, there was a significant effect of Condition (No Context, Context; $\beta = .05$, $t = 2.75$, $p < .01$). This indicates that talkers produced the phrases with longer durations in Context conditions than in No Context conditions.

Table 4.2. Summary of the linear mixed-effects regression model for raw durations (msec.) and scaled durations of the phrases with consonant targets.

Raw phrase duration Model `Raw duration (msec.) ~ TalkerGroup + (1| Talker) + (1| Word)`				
Fixed Effects	**Estimate**	**S.E.**	**t-val.**	***p*-val.**
(Intercept)	1715.26	43.92	39.06	< .001
TalkerGroup1 (Native English vs. Native Mandarin-High)	-341.49	111.64	-2.98	.004 **
TalkerGroup2 (Native Mandarin-High vs. Native Mandarin-Low)	-283.48	127.19	-2.23	.029 *
Scaled phrase duration Model `Scaled duration ~ TalkerGroup*Type*Condition + (1| Talker) + (1| Word)`				
(Intercept)	.46	.01	31.89	< .001
TalkerGroup1 (Native English vs. Native Mandarin-High)	.08	.03	2.36	.021 *
TalkerGroup2 (Native Mandarin-High vs. Native Mandarin-Low)	.12	.04	3.14	.002 **
Type (L1L2 vs L2-only)	-.03	.02	-1.62	.11
Condition (No Context vs. Context)	.05	.02	2.75	.009 **
TalkerGroup1: Type	-.05	.03	-1.71	.088
TalkerGroup2: Type	-.07	.03	-1.87	.062
TalkerGroup1: Condition	.05	.03	1.56	.12
TalkerGroup2: Condition	.03	.03	.87	.39
Type: Condition	-.009	.03	-.27	.79
TalkerGroup1: Type: Condition	.06	.06	1.01	.31
TalkerGroup2: Type: Condition	-.02	.07	-.32	.75

4.3.1.1.2. Mean F0

The left panel in Figure 4.3 shows scaled values of mean F0 of the phrase, *"Click*

on the ___ now", by Talker Group (Native English, Native Mandarin-High, Native Mandarin-Low), Type (L1L2, L2-only), and Condition (No context, Context). The linear mixed-effects regression model, used to analyze scaled mean F0 as the dependent variable, included Talker Group, Type, and Condition, as well as interactions among these factors as fixed effects (see Table 4.3 for the model syntax and summary of the results). The only significant effect in the model was the interaction among Type (L1L2, L2-only), Condition (No Context, Context), and the Native English vs. Native Mandarin-High group comparison ($\beta = -.13$, $t = -2.41$, $p < .05$). This indicates that the pattern of the Type x Condition interaction was different between Native English and Native Mandarin-High talkers' productions.

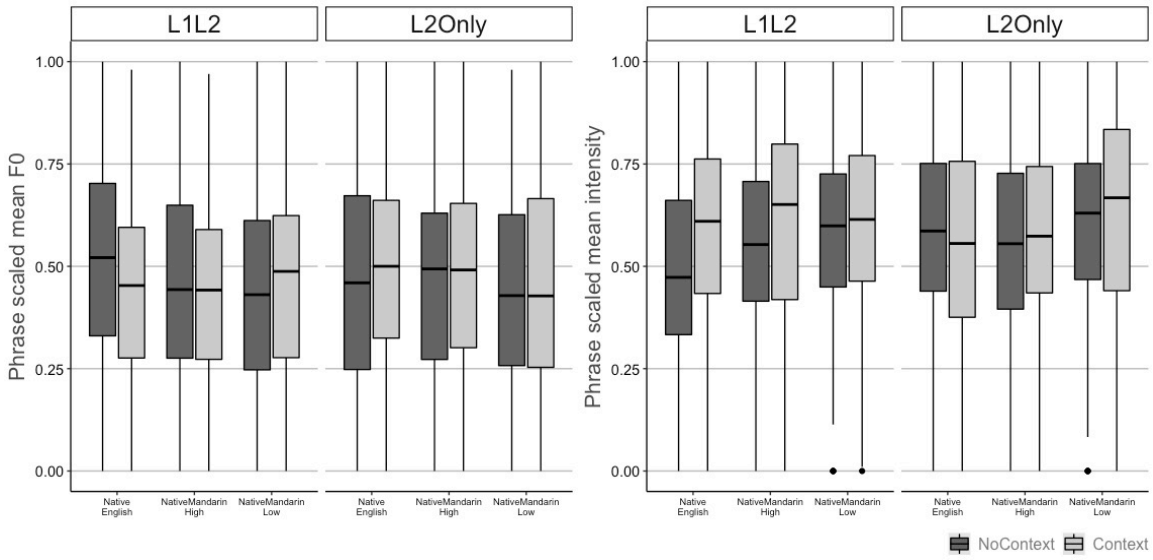

Figure 4.3. Scaled mean F0 (Left panel) and scaled mean intensity (Right panel) for phrases containing consonant targets (with L1L2 and L2-only contrasts) for different talker groups (Native English, Native Mandarin-High, and Native-Mandarin-Low talkers) in different conditions (No Context and Context).

Table 4.3. Summary of the linear mixed-effects regression model for scaled mean F0 and scaled mean intensity of the phrases with consonant targets.

Scaled phrase mean F0 Model `Scaled mean F0 ~ TalkerGroup*Type*Condition + (1| Talker) + (1| Word)`				
Fixed Effects	**Estimate**	**S.E.**	**t-val.**	***p*-val.**
(Intercept)	.47	.02	23.47	< .001
TalkerGroup1 (Native English vs. Native Mandarin-High)	.02	.04	.38	.71
TalkerGroup2 (Native Mandarin-High vs. Native Mandarin-Low)	.02	.04	.45	.65
Type (L1L2 vs L2-only)	-.007	.03	-.25	.81
Condition (No Context vs. Context)	-.001	.03	-.04	.97
TalkerGroup1: Type	.03	.03	1.03	.31
TalkerGroup2: Type	-.006	.03	-.21	.84
TalkerGroup1: Condition	-.03	.03	-1.02	.31
TalkerGroup2: Condition	-.03	.03	-.95	.34
Type: Condition	-.05	.05	-.85	.4
TalkerGroup1: Type: Condition	-.13	.06	-2.41	.016 *
TalkerGroup2: Type: Condition	-.1	.06	-1.7	.089
Scaled phrase mean intensity Model `Scaled mean intensity ~ TalkerGroup*Type*Condition + (1| Talker) + (1| Word)`				
Fixed Effects	**Estimate**	**S.E.**	**t-val.**	***p*-val.**
(Intercept)	.58	.01	42.07	< .001
TalkerGroup1 (Native English vs. Native Mandarin-High)	-.03	.02	-1.53	.13
TalkerGroup2 (Native Mandarin-High vs. Native Mandarin-Low)	-.05	.03	-2.1	.039 *
Type (L1L2 vs L2-only)	-.01	.02	-.47	.64
Condition (No Context vs. Context)	.04	.02	1.7	.097
TalkerGroup1: Type	-.03	.03	-1.06	.29
TalkerGroup2: Type	.01	.03	.33	.74
TalkerGroup1: Condition	-.002	.03	-.07	.94
TalkerGroup2: Condition	.02	.03	.66	.51
Type: Condition	.04	.05	.79	.44
TalkerGroup1: Type: Condition	.16	.06	2.68	.008 **
TalkerGroup2: Type: Condition	.1	.07	1.45	.15

4.3.1.1.3. Mean intensity

The right panel in Figure 4.3 shows scaled values of mean intensity of the phrase, *"Click on the ___ now"*, by Talker Group (Native English, Native Mandarin-High, Native

Mandarin-Low), Type (L1L2, L2-only), and Condition (No context, Context). The linear mixed-effects regression model, used to analyze scaled mean intensity as the dependent variable, included Talker Group, Type, and Condition, as well as interactions among these factors as fixed effects (see Table 4.3 for the model syntax and summary of the results). The figure suggests that, though talkers generally increased mean intensity in Context conditions compared to No Context conditions, this pattern was slightly opposite for Native English talkers' productions of the phrases containing L2-only target words. This was reflected in the significant interaction among Type (L1L2, L2-only), Condition (No Context, Context), and the Native English vs. Native Mandarin-High group comparison ($\beta = .16$, $t = 2.68$, $p < .01$).

Together, acoustic analyses of the whole phrase, *"Click on the ___ now"*, demonstrated different patterns in terms of duration, mean F0, and mean intensity. Raw durations of the phrases differed for different talker groups; Native Mandarin-Low talkers produced the phrases with longer durations than Native Mandarin-High talkers did; Native Mandarin-High talkers produced the phrases with longer durations than Native English talkers did. The difference in the type of production conditions (No Context vs. Context) also affected the phrase durations; talkers produced the phrases in Context conditions with relatively longer durations than in No Context conditions. Though there was a tendency for the phrases to have higher mean intensity in Context conditions than in No Context conditions, this effect was not statistically significant. The type of contrast in the target word (L1L2, L2-only) did not affect duration, mean F0, or mean intensity. These results suggest that at the phrase level, talkers' efforts to communicate target words clearly were manifested in phrase duration, but not in other aspects, such as fundamental frequency or

intensity. Further, phrase durations were also impacted by talkers' target language experience (native vs. non-native status; higher- vs. lower-proficiency).

4.3.1.2. Word-level analyses

4.3.1.2.1. Duration

The left panel in Figure 4.4 shows the raw durations (msec.) of the target words containing consonant contrasts by Talker Group (Native English, Native Mandarin-High, Native Mandarin-Low). The linear mixed-effects regression model, used to analyze raw word durations as the dependent variable, included Talker Group as a fixed factor (see Table 4.4 for the model syntax and summary of the results). The model showed that the effect of Talker Group was significant: Native English vs. Native Mandarin-High (β =61.14, t = 2.62, $p < .05$), Native Mandarin-High vs. Native Mandarin-Low (β = 78.62, t = 3,03, $p < .01$). This indicates that Native English talkers produced longer words than Native Mandarin-High talkers did; Native Mandarin-High talkers produced longer words than Native Mandarin-Low talkers did.

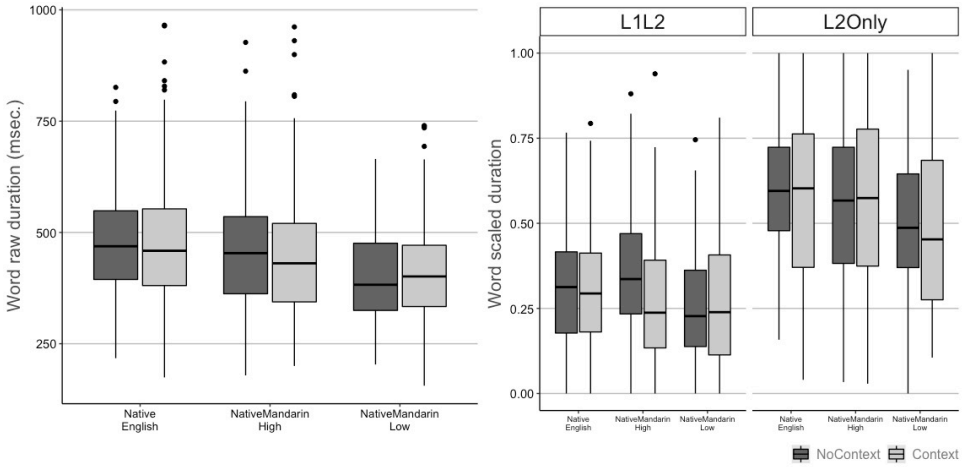

Figure 4.4. Durations of target words with consonant contrasts for different talker groups (Native English, Native Mandarin-High, and Native-Mandarin-Low talkers) in different conditions (No Context and Context): raw durations (Left panel), scaled durations (for targets containing L1L2 or L2-only contrasts: Right panel).

Table 4.4. Summary of the linear mixed-effects regression model for durations (msec.) and scaled durations of the words with consonant contrasts.

Raw word duration Model `Raw duration (msec.) ~ TalkerGroup + (1| Talker) + (1| Word)`				
Fixed Effects	**Estimate**	**S.E.**	**t-val.**	***p*-val.**
(Intercept)	442.5	14.12	31.35	< .001
TalkerGroup1 (Native English vs. Native Mandarin-High)	61.14	23.37	2.62	.011 *
TalkerGroup2 (Native Mandarin-High vs. Native Mandarin-Low)	78.62	25.93	3.03	.003 **
Scaled word duration Model `Scaled duration ~ TalkerGroup*Type*Condition + (1| Talker) + (1| Word)`				
(Intercept)	.42	.02	20.1	< .001
TalkerGroup1 (Native English vs. Native Mandarin-High)	.05	.02	2.33	.022 *
TalkerGroup2 (Native Mandarin-High vs. Native Mandarin-Low)	.08	.02	3.34	.001 **
Type (L1L2 vs L2-only)	-.26	.04	-6.69	< .001 ***
Condition (No Context vs. Context)	-.01	.04	-.3	.77
TalkerGroup1: Type	-.06	.02	-2.78	.006 **
TalkerGroup2: Type	-.04	.02	-1.55	.12
TalkerGroup1: Condition	-.003	.02	-.13	.9
TalkerGroup2: Condition	-.04	.02	-1.56	.12
Type: Condition	-.03	.08	-.33	.75
TalkerGroup1: Type: Condition	.08	.04	1.88	.061
TalkerGroup2: Type: Condition	-.09	.05	-1.8	.071

The right panel in Figure 4.4 shows the scaled word durations by Talker Group, Type (L1L2, L2-only), and Condition (No context, Context). The linear mixed-effects regression model, used to analyze scaled word durations as the dependent variable, included Talker Group, Type, and Condition, as well as interactions among these factors as fixed effects (see Table 4.4 for the model syntax and summary of the results). There was a significant effect of Type (L1L2, L2-only; $\beta = -.26$, $t = -6.69$, $p < .001$). This indicates that talkers produced words with the L2-only contrast (/p/-/b/ in word-final position) with longer durations than those with the L1L2 contrast (/p/-/b/ in word-initial position). This

effect of Type differed for Native English vs. Native Mandarin-High talkers' productions (Type x Native English vs. Native Mandarin-High: β = -.06, t = -2.78, $p < .01$).

4.3.1.2.2. Mean F0

The left panel in Figure 4.5 shows scaled mean F0 of the target words containing consonant contrasts, by Talker Group (Native English, Native Mandarin-High, Native Mandarin-Low), Type (L1L2, L2-only), and Condition (No context, Context). The linear mixed-effects regression model, used to analyze scaled mean F0 as the dependent variable, included Talker Group, Type, and Condition, as well as interactions among these factors as fixed effects (see Table 4.5 for the model syntax and summary of the results). The only significant effect in the model was the interaction between Type (L1L2, L2-only) and the Native English vs. Native Mandarin-High group comparison (β = .1, t = 4.0, $p < .001$). This indicates that the effect of Type was different between Native English and Native Mandarin-High talkers' productions.

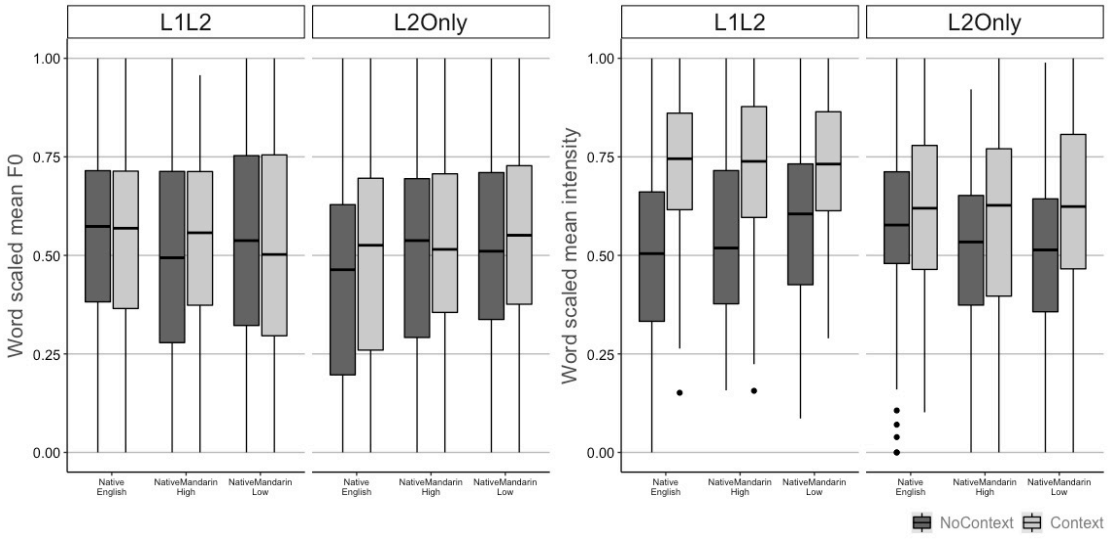

Figure 4.5. Scaled mean F0 (Left panel) and scaled mean intensity (Right panel) for target words containing consonant contrasts (with L1L2 and L2-only contrasts) for different talker groups (Native English, Native Mandarin-High, and Native-Mandarin-Low talkers) in different conditions (No Context and Context).

Table 4.5. Summary of the linear mixed-effects regression model for scaled mean F0 and scaled mean intensity of the target words with consonant contrasts.

Scaled word mean F0 Model `Scaled mean F0 ~ TalkerGroup*Type*Condition + (1| Talker) + (1| Word)`				
Fixed Effects	**Estimate**	**S.E.**	**t-val.**	***p*-val.**
(Intercept)	.51	.02	22.17	< .001
TalkerGroup1 (Native English vs. Native Mandarin-High)	-.02	.05	-.48	.64
TalkerGroup2 (Native Mandarin-High vs. Native Mandarin-Low)	-.02	.05	-.46	.65
Type (L1L2 vs L2-only)	.02	.03	.68	.5
Condition (No Context vs. Context)	.02	.03	.8	.43
TalkerGroup1: Type	.1	.03	4.0	< .001 ***
TalkerGroup2: Type	.04	.03	1.52	.13
TalkerGroup1: Condition	-.005	.03	-.2	.84
TalkerGroup2: Condition	.004	.03	.13	.9
Type: Condition	-.02	.06	-.3	.77
TalkerGroup1: Type: Condition	-.08	.05	-1.54	.12
TalkerGroup2: Type: Condition	-.001	.06	-.03	.98
Scaled word mean intensity Model `Scaled mean intensity ~ TalkerGroup*Type*Condition + (1| Talker) + (1| Word)`				
Fixed Effects	**Estimate**	**S.E.**	**t-val.**	***p*-val.**
(Intercept)	.6	.02	34.92	< .001
TalkerGroup1 (Native English vs. Native Mandarin-High)	.008	.03	.37	.71
TalkerGroup2 (Native Mandarin-High vs. Native Mandarin-Low)	-.02	.02	-.71	.48
Type (L1L2 vs L2-only)	.08	.03	2.41	.02 *
Condition (No Context vs. Context)	.13	.03	3.97	< .001 ***
TalkerGroup1: Type	-.09	.02	-3.74	< .001 ***
TalkerGroup2: Type	-.04	.03	-1.59	.11
TalkerGroup1: Condition	.02	.02	.76	.45
TalkerGroup2: Condition	.002	.03	.08	.94
Type: Condition	.1	.06	1.63	.11
TalkerGroup1: Type: Condition	.18	.05	3.61	< .001 ***
TalkerGroup2: Type: Condition	.16	.05	2.95	.003 **

4.3.1.2.3. Mean intensity

The right panel in Figure 4.5 shows scaled mean intensity of target words with consonant contrasts, by Talker Group (Native English, Native Mandarin-High, Native

Mandarin-Low), Type (L1L2, L2-only), and Condition (No context, Context). The linear mixed-effects regression model, used to analyze scaled mean intensity as the dependent variable, included Talker Group, Type, and Condition, as well as interactions among these factors as fixed effects (see Table 4.5 for the model syntax and summary of the results). There was a significant effect of Type (L1L2, L2-only: $\beta = .8$, $t = 2.41$, $p < .05$). This indicates that mean intensity was higher for the target words containing the L1L2 contrast (/p/-/b/ in word-initial position) than for those containing the L2-only contrast (/p/-/b/ in word-final position). This effect of Type was larger in Native Mandarin-High talkers' productions than for Native English talkers' productions (Type x Native English vs. Native Mandarin-High group: $\beta = -.09$, $t = -3.74$, $p < .001$). There was also a significant effect of Condition (No Context, Context: $\beta = .13$, $t = -3.74$, $p < .001$). This indicates that mean intensity of the target words was higher in Context conditions than in No Context conditions. There were also significant interactions among Type, Condition, and Talker Group comparisons: Type x Condition x Native English vs. Native Mandarin-High ($\beta = .18$, $t = 3.61$, $p < .001$), Type x Condition x Native Mandarin-High vs. Native Mandarin-Low ($\beta = .16$, $t = 2.95$, $p < .01$). These results indicate that the interaction patterns of Type x Condition were different for different Talker Groups' productions.

In sum, acoustic analyses of the target words containing consonant contrasts demonstrated the influence of talkers' target language experience on the word duration; Native Mandarin-Low talkers produced the target words with *shorter* durations than Native Mandarin-High talkers did; Native Mandarin-High talkers produced the target words with *shorter* durations than Native English talkers did. Further, the type of the contrast in the target words (L1L2 contrast: /p/-/b/ in word-initial position, and L2-only contrast: /p/-/b/ in

word-final position) affected multiple aspects of the target word productions (e.g., longer duration and lower mean intensity for the target words with the L2-only contrast than those with the L1L2 contrast). This suggests that the difference in the position of a phonological contrast (i.e., /p/-/b/ in word-initial vs. word-final position) could impact the global acoustic characteristics of the target words. In addition to these general tendencies influenced by the type of segments, talkers' efforts to produce target words clearly in Context conditions were manifested in intensity. Talkers produced the target words with higher mean intensity in the Context conditions than those in the No Context conditions, though their productions did not differ across Conditions in terms of target word durations and fundamental frequency. This may suggest that talkers' contextually-relevant enhancements are targeted such that modifications at the target word level are less obvious compared to acoustic manipulations at the segmental level (examined in Section 4.3.2 below).

4.3.1.3. Summary of the global analyses for consonant targets

Together, analyses of the items with consonant targets at the phrase-level and the word-level showed that global characteristics of these items were influenced by a combination of talkers' target language experience, the type of the target contrast, and the production condition. Specifically, talkers' target language experience generally impacted the temporal properties of the entire phrases and the target words; among the three talker groups, Native Mandarin-Low talkers' phrase durations were the longest, though their target word durations were the shortest. These patterns suggest that Native Mandarin-Low talkers spoke more slowly than Native Mandarin-High talkers and Native English talkers, and they did so at the phrase-level (possibly with more frequent and longer pauses, and/or

longer segment durations), not with durations of target words per se. It is possible that Native English talkers focused on producing the target words clearly (rather than producing the whole phrase clearly). Other aspects of the results (e.g., duration and intensity of target words) suggested that target words containing the same phonological contrast (/p/-/b/) could have different global acoustic properties depending on the position of the contrast (e.g., word-initial, word-final) within the words. Furthermore, talkers' productions differed based on whether a target word was displayed with its minimal-pair neighbor (Context conditions) or not (No Context conditions). However, as the acoustic modifications based on the type of conditions were limited to several aspects of the global characteristics (e.g., longer phrase durations and higher mean intensity of the target words), it is possible that talkers' contextually-relevant enhancements do not manifest widely across different measures (e.g., duration, intensity, fundamental frequency) at the phrase- or word-level.

4.3.2. Consonant targets: Segmental analyses

In order to characterize segmental features of the target consonant contrasts, we analyzed several acoustic features. For the L1L2 consonant contrast (i.e., /p/-/b/ contrast in word-initial position), we analyzed normalized VOTs of /p/ and /b/. For the L2-only consonant contrast (i.e., /p/-/b/ contrast in word-final position), we analyzed the voicing proportions of the /p/- and /b/ consonants (C2 voicing proportions) and normalized durations of the vowels preceding the target consonants (normalized vowel durations). For each acoustic feature, we first present the results of a linear mixed-effects regression model, examining whether talkers made differences in the acoustic measure (e.g., normalized VOTs for the /p/-/b/ contrast in word-initial position) to distinguish one phoneme (e.g., /p/)

from another (e.g., /b/), and whether the size of this difference was larger for one Talker Group's production than for another (e.g., Native Mandarin-High vs. Native Mandarin-Low). Then, we present the results of linear mixed-effects regression models, examining the effect of Condition (No Context vs. Context) separately for the two phonemes (e.g., /p/ and /b/), as well as whether the effect of Condition differed for different Talker Groups' productions.

In the linear mixed-effects regression models presented below, fixed effects were Phoneme (/b/, /p/), Condition (Context, No Context), and Talker Group (Native English, Native Mandarin-High, and Native Mandarin-Low); different combinations of these fixed effects were included in different models (see the description of each model below). Condition and Talker Group were contrast coded as specified above. Phoneme was contrast coded to compare between /b/-targets (.5) and /p/-targets (-.5). Models also included the maximal random effects structure that would converge, which included random intercepts for talker and item. The random effects structure also included a by-talker random slope for Phoneme or Condition (different depending on the model) and a by-word intercept for Talker Group. In each model, the random effects that did not account for any variance was dropped to avoid overfitting; see the description of each model below.

4.3.2.1. L1L2 consonant contrast (/p/-/b/ in word-initial position)

4.3.2.1.1. Normalized VOTs

Figure 4.6 shows the mean normalized VOTs by Phoneme (/b/, /p/), Talker Group (Native English, Native Mandarin-High, Native Mandarin-Low), and Condition (No context, Context). The figure suggests that normalized VOTs were longer for /b/ than for

/p/ for all Talker Groups' productions; mean normalized VOT for /p/ - mean normalized VOT for /b/ was .18 (Native English), .21 (Native Mandarin-High), .21 (Native Mandarin-Low). The linear mixed-effects regression model, used to analyze normalized VOTs as the dependent variable, included Phoneme, Talker Group, as well as interactions between the two factors as fixed effects (see Table 4.6 for the model syntax and summary of the results). There was a significant effect of Phoneme (β = -.2, t = -38.8, $p < .001$). The effect of Phoneme interacted with the Native English vs. Native Mandarin-High comparison (β = .03, t = 2.48, $p < .05$), but not with the Native Mandarin-High vs. Native Mandarin-Low comparison (β = .01, t = .79, p = .43). This indicates that the difference in normalized VOTs between /b/ and /p/ was larger for Native Mandarin-High talkers' productions than for Native English talkers' productions, but the difference was similar for Native Mandarin-High and Native Mandarin-Low talkers' productions. A post-hoc Tukey test confirmed that the effect of Phoneme was significant for all the Talker Groups' productions: Native English (β = .18, SE = .008, t ratio = 23.52, $p < .0001$), Native Mandarin-High (β = .21, SE = .01, t ratio = 21.74, $p < .0001$), and Native Mandarin-Low (β = .21, SE = .01, t ratio = 21.24, $p < .0001$).

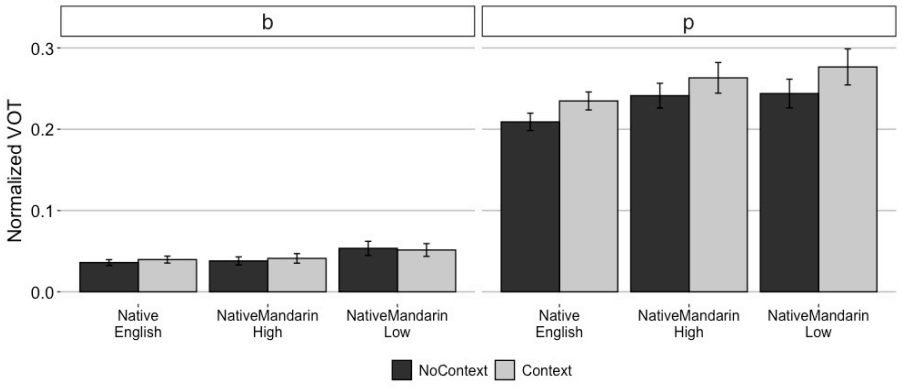

Figure 4.6. Mean VOT of the consonants in the L1L2 contrast by Phoneme (/b/, /p/), Talker Group (Native English, Native Mandarin-High, Native Mandarin-Low), and Condition (No context, Context). The error bars represent the 95% confidence interval of the mean.

Table 4.6. Summary of the linear mixed-effects regression models for normalized VOTs of the consonants in the L1L2 consonant contrast (/p/-/b/ contrast in word-initial position).

Fixed Effects	**Estimate**	**S.E.**	**t-val.**	***p*-val.**
Normalized VOT for /b/ and /p/ Model `Normalized VOT ~ Phoneme*TalkerGroup + (1+ Phoneme| Talker)`				
(Intercept)	.14	.003	46.34	< .001
Phoneme (/b/ vs. /p/)	-.2	.005	-38.8	< .001 ***
TalkerGroup1 (Native English vs. Native Mandarin-High)	-.03	.008	-3.41	.001 **
TalkerGroup2 (Native Mandarin-High vs. Native Mandarin-Low)	-.02	.009	-2.68	.009 **
Phoneme: TalkerGroup1	.03	.01	2.48	.015 *
Phoneme: TalkerGroup2	.01	.02	.79	.43
Normalized VOT for /b/ Model `Normalized VOT ~ Condition*TalkerGroup + (1| Talker)`				
(Intercept)	.04	.002	20.7	< .001
Condition (No Context vs. Context)	.002	.002	1.31	.19
TalkerGroup1 (Native English vs. Native Mandarin-High)	-.01	.006	-2.01	.048 *
TalkerGroup2 (Native Mandarin-High vs. Native Mandarin-Low)	-.02	.006	-3.03	.003 **
TalkerGroup1: Condition	.001	.004	.32	.75
TalkerGroup2: Condition	.003	.005	.64	.52
Normalized VOT for /p/ Model `Normalized VOT ~ Condition*TalkerGroup + (1| Talker)`				
(Intercept)	.24	.005	46.77	< .001
Condition (No Context vs. Context)	.02	.005	4.49	< .001 ***
TalkerGroup1 (Native English vs. Native Mandarin-High)	-.05	.01	-3.29	.002 **
TalkerGroup2 (Native Mandarin-High vs. Native Mandarin-Low)	-.03	.02	-2	.049 *
TalkerGroup1: Condition	.0006	.01	.048	.96
TalkerGroup2: Condition	-.007	.01	-.52	.6

Figure 4.6 also suggests different patterns for normalized VOTs for each consonant (/b/ and /p/). For /b/, talkers did not change VOTs in Context conditions compared to No Context conditions. However, for /p/, normalized VOTs are influenced by both Talker Groups and Conditions. That is, Native Mandarin (both High and Low) talkers produced

/p/ with longer VOTs than Native English talkers did. Further, talkers in all three Talker Groups increased VOTs from No Context to Context conditions. The linear mixed-effects regression models, used to analyze normalized VOTs as the dependent variable separately for /b/-targets and /p/-targets, included Condition (No Context, Context), Talker Group, as well as interactions between the two as fixed effects (see Table 4.6 for the model syntax and summary of the results). For the /b/-target model, there were significant effects of Talker Groups: Native English vs. Native Mandarin-High (β = -.01, t = -2.01, p < .05), Native Mandarin-High vs. Native Mandarin-Low (β = -.02, t = -3.03, p < .01). This indicates that normalized VOTs for /b/ were longer for Native Mandarin-Low talkers' productions than for Native Mandarin-High talkers' productions, and for Native Mandarin-High talkers' productions than for Native English talkers' productions. The effect of Condition was not significant (β = .002, t = 1.31, p = .19). This pattern was similar for different Talker Groups' productions, as the effect of Condition did not interact with the Native English vs. Native Mandarin-High comparison (β = .001, t = .32, p = .75), or with the Native Mandarin-High and Native Mandarin-Low comparison (β = .003, t = .64, p = .52). A post-hoc Tukey test showed that the effect of Condition was not significant for any of the Talker Groups: Native English (β = -.003, SE = .002, t ratio = -1.16, p = .25), Native Mandarin-High (β = -.003, SE = .003, t ratio = -.98, p = .33), Native Mandarin-Low (β = -.0006, SE = .003, t ratio = -.2, p = .84).

For the /p/-target model, there were significant effects of Talker Groups: Native English vs. Native Mandarin-High (β = -.05, t = -3.29, p < .01), Native Mandarin-High vs. Native Mandarin-Low (β = -.03, t = -2, p < .05). This indicates that normalized VOTs for /p/ were longer for Native Mandarin-Low talkers' productions than for Native Mandarin-

High talkers' productions, and for Native Mandarin-High talkers' productions than for Native English talkers' productions. The effect of Condition was significant (β = .02, t = 4.49, $p < .001$). This pattern was similar for different Talker Groups' productions, as the effect of Condition did not interact with the Native English vs. Native Mandarin-High comparison (β = .0006, t = .048, $p = .96$), or with the Native Mandarin-High vs. Native Mandarin-Low comparison (β = -.007, t = -.52, $p = .6$). A post-hoc Tukey test confirmed that the effect of Condition was significant for all Talker Groups' productions: Native English (β = -.02, SE = .007, t ratio = -3.07, $p = .0023$), Native Mandarin-High (β = -.02, SE = .009, t ratio = -1.97, $p = .049$), Native Mandarin-Low (β = -.03, SE = .009, t ratio = -2.83, $p = .005$).

Together, these results demonstrated that talkers distinguished word-initial /b/ and /p/ with VOTs; normalized VOTs were longer for /p/ than for /b/. This difference was larger for Native Mandarin talkers' productions than for Native English talkers' productions. Further, talkers manipulated VOTs in different types of Conditions (No Context vs. Context) only for /p/; they increased VOTs for /p/ from No Context to Context conditions. The size of this Condition-based difference for /p/ was similar for all Talker Groups' productions.

4.3.2.2. L2-only consonant contrast (/p/-/b/ in word-final position)

4.3.2.2.1. Normalized vowel duration

The left panel in Figure 4.7 shows the mean normalized durations of the vowels preceding the target consonants by Phoneme (/b/, /p/), Talker Group (Native English, Native Mandarin-High, Native Mandarin-Low), and Condition (No context, Context). The

figure suggests that normalized durations were longer for the vowels preceding /b/ than those preceding /p/. However, this difference in the preceding vowel durations seems to be larger for Native English talkers' productions than for Native Mandarin-High talkers' productions; mean normalized vowel duration for /b/ - that for /p/ was .11 (Native English), .08 (Native Mandarin-High), .04 (Native Mandarin-Low). The linear mixed-effects regression model, used to analyze normalized durations of the preceding vowels as the dependent variable, included Phoneme, Talker Group, as well as interactions between the two factors as fixed effects (see Table 4.7 for the model syntax and summary of the results). There was a significant effect of Phoneme ($\beta = .08$, $t = 4.98$, $p < .001$). The effect of Phoneme interacted with each Talker Group comparison: Phoneme x Native English vs. Native Mandarin-High ($\beta = .07$, $t = 4.65$, $p < .001$), Phoneme x Native Mandarin-High vs. Native Mandarin-Low ($\beta = .07$, $t = 4.5$, $p < .001$). This indicates that talkers produced longer vowels before /b/ than those before /p/. This difference in normalized vowel durations was larger for Native English talkers' productions than for Native Mandarin-Higher talkers' productions, and for Native Mandarin-Higher talkers' productions than for Native Mandarin-Low talkers' productions. A post-hoc Tukey test showed that the effect of Phoneme was significant for the Native English group ($\beta = -.11$, SE = .02, t ratio = -6.42, $p < .0001$), Native Mandarin-High group ($\beta = -.08$, SE = .02, t ratio = -4.36, $p = .0001$), and Native Mandarin-Low group ($\beta = -.04$, SE = .02, t ratio = -2.21, $p = .033$).

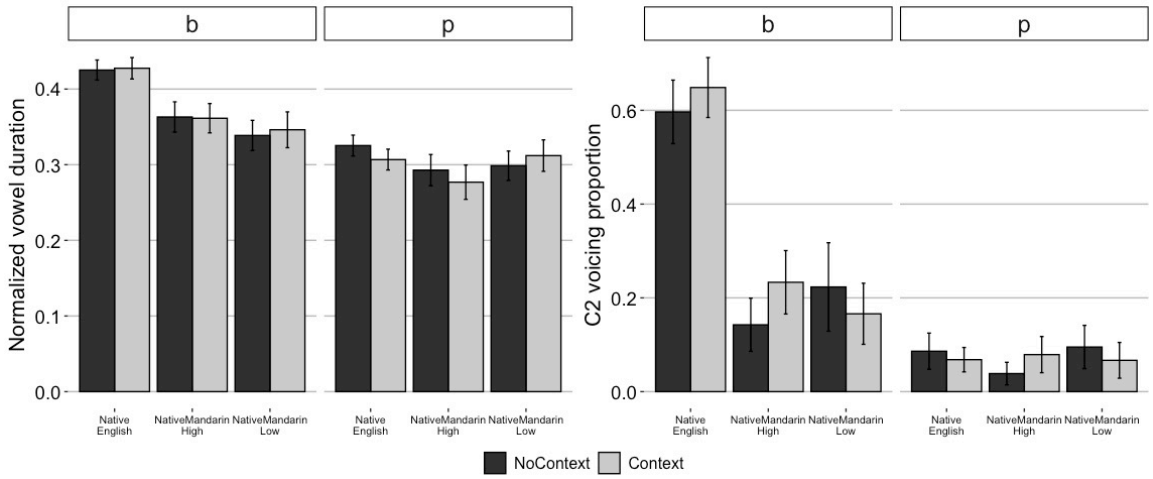

Figure 4.7. Mean normalized durations of the vowels preceding the word-final target consonants (Left panel) and mean voicing proportions of the target consonants (Right panel), by Phoneme (/b/, /p/), Talker Group (Native English, Native Mandarin-High, Native Mandarin-Low), and Condition (No context, Context). The error bars represent the 95% confidence interval of the mean.

The left panel of Figure 4.7 also suggests different patterns for normalized durations of vowels preceding /b/ and /p/. That is, although normalized durations of the vowels preceding /b/ differed for different Talker Groups, this Talker Group-based difference was much smaller for normalized vowel durations preceding /p/. Further, there was a tendency for Native English and Native Mandarin-High talkers to shorten the vowel durations preceding /p/ in Context conditions compared to No Context conditions; whereas talkers generally did not differentiate vowel durations preceding /b/ in different conditions.

The linear mixed-effects regression models, used to analyze normalized vowel durations as the dependent variable separately for /b/-targets and /p/-targets, included Condition (No Context, Context), Talker Group, as well as interactions between the two as fixed effects (see Table 4.7 for the model syntax and summary of the results). For the model with /b/-targets, there were significant effects of Talker Groups: Native English vs. Native Mandarin-High ($\beta = .1$, $t = 7.48$, $p < .001$), Native Mandarin-High vs. Native

Table 4.7. Summary of the linear mixed-effects regression models for normalized durations of the vowels preceding the L2-only consonant contrast (/p/-/b/ contrast in word-final position).

Fixed Effects	**Estimate**	**S.E.**	**t-val.**	***p*-val.**

Normalized vowel duration for /b/ and /p/ Model
`Normalized vowel duration ~ Phoneme*TalkerGroup`
`+ (1+ Phoneme| Talker)`
`+ (1| Word)`				

(Intercept)	.34	.009	39.88	< .001
Phoneme (/b/ vs. /p/)	.08	.02	4.98	< .001 ***
TalkerGroup1 (Native English vs. Native Mandarin-High)	.06	.01	5.13	< .001 ***
TalkerGroup2 (Native Mandarin-High vs. Native Mandarin-Low)	.04	.01	2.4	.019 *
Phoneme: TalkerGroup1	.07	.01	4.65	< .001 ***
Phoneme: TalkerGroup2	.07	.02	4.5	< .001 ***

Normalized vowel duration for /b/ Model
`Normalized vowel duration ~ Condition*TalkerGroup`
`+ (1| Talker)`				

(Intercept)	.38	.005	74.82	< .001
Condition (No Context vs. Context)	.003	.006	.4	.69
TalkerGroup1 (Native English vs. Native Mandarin-High)	.1	.01	7.48	< .001 ***
TalkerGroup2 (Native Mandarin-High vs. Native Mandarin-Low)	.07	.01	4.72	< .001 ***
TalkerGroup1: Condition	-.003	.02	-.15	.88
TalkerGroup2: Condition	-.01	.02	-.71	.48

Normalized vowel duration for /p/ Model
`Normalized vowel duration ~ Condition*TalkerGroup`
`+ (1+ Condition| Talker)`
`+ (1 | Word)`				

(Intercept)	.3	.01	29.26	< .001
Condition (No Context vs. Context)	-.007	.02	-.39	.7
TalkerGroup1 (Native English vs. Native Mandarin-High)	.03	.02	2.04	.045 *
TalkerGroup2 (Native Mandarin-High vs. Native Mandarin-Low)	-.002	.02	-.09	.93
TalkerGroup1: Condition	-.02	.02	-1.53	.13
TalkerGroup2: Condition	-.03	.02	-1.97	.053

Mandarin-Low ($\beta = .07$, $t = 4.72$, $p < .001$). This indicates that normalized durations of vowels preceding /b/ were longer for Native English talkers' productions than for Native Mandarin-High talkers' productions, and for Native Mandarin-High talkers' productions than for Native Mandarin-Low talkers' productions. The main effect of Condition was not

significant (β = .003, t = .4, *p* = .69). The effect of Condition did not interact with the Native English vs. Native Mandarin-High comparison (β = -.003, t = -.15, *p* = .88), or with the Native Mandarin-High vs. Native Mandarin-Low comparison (β = -.01, t = -.71, *p* = .48). This indicates that the effect of Condition did not differ among different Talker Groups. A post-hoc Tukey test confirmed that the effect of Condition was not significant in any of the Talker Groups: Native English (β = -.001, SE = .01, t ratio = -.14, *p* = .89), Native Mandarin-High (β = .003, SE = .01, t ratio = .24, *p* = .81), Native Mandarin-Low (β = -.01, SE = .01, t ratio = -.77, *p* = .44).

For the model with /p/-targets, there was a significant effect of the Native English vs. Native Mandarin-High comparison (β = .03, t = 2.04, *p* < .05). This indicates that normalized vowel durations were shorter for Native Mandarin-High talkers' productions than for Native English talkers' productions. The effect of Condition was not significant (β = -.007, t = -.39, *p* = .7). However, there was a marginally significant interaction between Condition and Native Mandarin-High vs. Native Mandarin-Low comparison (β = -.03, t = -1.97, *p* = .05). This indicates that the effect of Condition was different between the two Talker Groups' productions; Native Mandarin-High talkers tended to decrease the normalized vowel durations in Context conditions compared to No Context conditions, though Native Mandarin-Low talkers did not. A post-hoc Tukey test showed that the effect of Condition was not significant for any of the Talker Groups: Native English (β = .02, SE = .02, t ratio = .86, *p* = .41), Native Mandarin-High (β = .02, SE = .02, t ratio = .55, *p* = .59), Native Mandarin-Low (β = -.008, SE = .02, t ratio = -.38, *p* = .71).

These results suggest that talkers distinguished word-final /b/ and /p/ with preceding vowel durations; normalized durations were longer for the vowels preceding /b/

than for those preceding /p/. This difference was larger for Native English talkers' productions than for Native Mandarin-High talkers' productions, and for Native Mandarin-High talkers' productions than for Native Mandarin-Low talkers' productions. This difference in normalized vowel durations for /b/- vs. /p/-targets among different talker groups was largely influenced by how talkers produced /b/-targets as compared to how they produced /p/-targets. That is, normalized vowel durations preceding /b/ were longer for Native English talkers' productions than for Native Mandarin-High talkers' productions, and for Native Mandarin-High talkers' productions than for Native Mandarin-Low talkers' productions. However, talkers did not manipulate normalized durations of the preceding vowels differently in different conditions (No Context, Context), either for /b/ or /p/. Though Native Mandarin-High and Native Mandarin-Low talkers manipulated the normalized vowel durations for /p/-targets in different directions (i.e., Native Mandarin-High talkers shortened the vowel durations in Context conditions compared to No Context conditions while Native Mandarin-Low talkers slightly increased the vowel durations in Context conditions), these differences between Conditions were not statistically significant.

4.3.2.2.2. C2 voicing proportion

The right panel in Figure 4.7 shows the mean C2 voicing proportions by Phoneme (/b/, /p/), Talker Group (Native English, Native Mandarin-High, Native Mandarin-Low), and Condition (No context, Context). The figure suggests that the C2 voicing proportions were larger for /b/ than for /p/ for Native English talkers' productions, but this difference was much smaller for Native Mandarin talkers' productions; mean C2 voicing proportion for /b/ - mean C2 voicing proportion for /p/ was .55 (Native English), .13 (Native

Mandarin-High), .11 (Native Mandarin-Low). The linear mixed-effects regression model, used to analyze C2 voicing proportions as the dependent variable, included Phoneme, Talker Group, as well as interactions between the two factors as fixed effects (see Table 4.8 for the model syntax and summary of the results). The results showed a significant effect of Phoneme ($\beta = .26$, $t = 11.56$, $p < .001$). The effect of Phoneme interacted with each Talker Group comparison: Phoneme x Native English vs. Native Mandarin-High ($\beta = .3$, $t = 7.65$, $p < .001$), Phoneme x Native Mandarin-High vs. Native Mandarin-Low ($\beta = .13$, $t = 3.11$, $p < .01$). This indicates that the size of difference in C2 voicing proportions between /b/ and /p/ differed for different Talker Groups. The difference was larger for Native English talkers' productions than for Native Mandarin-Higher talkers' productions, and for Native Mandarin-High talkers' productions than for Native Mandarin-Low talkers' productions. A post-hoc Tukey test confirmed that the effect of Phoneme was significant for all the Talker Groups' productions: Native English ($\beta = -.55$, SE = .03, t ratio = -16.06, $p < .0001$), Native Mandarin-High ($\beta = -.13$, SE = .04, t ratio = -2.95, $p = .004$), Native Mandarin-Low ($\beta = -.11$, SE = .04, t ratio = -2.63, $p = .01$).

The right panel of Figure 4.7 also suggests different patterns for C2 voicing proportions for each consonant (/b/ and /p/). For /b/, Native English talkers and Native Mandarin-High talkers increased the C2 voicing proportions from No Context to Context conditions; while Native Mandarin-Low talkers changed C2 voicing proportions in the opposite direction. However, the patterns for /p/ suggest that neither the Talker Group nor Condition influenced C2 voicing proportions.

Table 4.8. Summary of the linear mixed-effects regression models for the voicing proportions of the consonants in the L2-only consonant contrast (/p/-/b/ contrast in word-final position).

Fixed Effects	**Estimate**	**S.E.**	**t-val.**	***p*-val.**
C2 voicing for /b/ and /p/ Model `C2 voicing ~ Phoneme*TalkerGroup + (1+ Phoneme| Talker)`				
(Intercept)	.2	.01	13.68	< .001
Phoneme (/b/ vs. /p/)	.26	.02	11.56	< .001 ***
TalkerGroup1 (Native English vs. Native Mandarin-High)	.3	.04	7.65	< .001 ***
TalkerGroup2 (Native Mandarin-High vs. Native Mandarin-Low)	.13	.04	3.11	.003 **
Phoneme: TalkerGroup1	.57	.06	9.59	< .001 ***
Phoneme: TalkerGroup2	.3	.07	4.48	< .001 ***
C2 voicing for /b/ Model `C2 voicing ~ Condition*TalkerGroup + (1+ Condition| Talker)`				
(Intercept)	.33	.02	13.96	< .001
Condition (No Context vs. Context)	.04	.02	1.76	.08
TalkerGroup1 (Native English vs. Native Mandarin-High)	.58	.06	9.26	< .001 ***
TalkerGroup2 (Native Mandarin-High vs. Native Mandarin-Low)	.28	.07	4.05	< .001 ***
TalkerGroup1: Condition	.06	.06	.96	.34
TalkerGroup2: Condition	.16	.07	2.33	.022 *
C2 voicing for /p/ Model `C2 voicing ~ Condition*TalkerGroup + (1| Talker)`				
(Intercept)	.07	.01	6.36	< .001
Condition (No Context vs. Context)	-.006	.01	-.47	.64
TalkerGroup1 (Native English vs. Native Mandarin-High)	.01	.03	.39	.7
TalkerGroup2 (Native Mandarin-High vs. Native Mandarin-Low)	-.01	.03	-.44	.66
TalkerGroup1: Condition	-.03	.03	-.91	.36
TalkerGroup2: Condition	.05	.04	1.29	.2

The linear mixed-effects regression models, used to analyze C2 voicing proportions as the dependent variable separately for /b/-targets and /p/-targets, included Condition (No Context, Context), Talker Group, as well as interactions between the two as fixed effects (see Table 4.8 for the model syntax and summary of the results). For the /b/-

target model, there were significant effects of Talker Groups: Native English vs. Native Mandarin-High (β = .58, t = 9.26, $p < .001$), Native Mandarin-High vs. Native Mandarin-Low (β = .28, t = 4.05, $p < .001$). This indicates that C2 voicing proportions for /b/ were larger for Native English talkers' productions than for Native Mandarin-High talkers' productions, and for Native Mandarin-High talkers' productions than for Native Mandarin-Low talkers' productions. The effect of Condition was not significant (β = .04, t = 1.76, p = .08); however, it interacted with the Native Mandarin-High vs. Native Mandarin-Low comparison (β = .16, t = 2.33, $p < .05$). This indicates that the effect of Condition was different between the Native Mandarin-High and Native Mandarin-Low groups' productions. The effect of Condition did not interact with the Native English vs. Native Mandarin-High comparison (β = .06, t = .96, p = .34), indicating that the effect of Condition did not differ between the Native English and Native Mandarin-High groups. A post-hoc Tukey test showed that the effect of Condition was significant for the Native English group (β = -.07, SE = .03, t ratio = -2.03, p = .045), and the Native Mandarin-High group (β = -.09, SE = .04, t ratio = -2.09, p = .039), but not for the Native Mandarin-Low group (β = .04, SE = .04, t ratio = .87, p = .39). For the /p/-target model, none of the Context or Talker Group effects were significant (see Table 4.8). A post-hoc Tukey test showed that the effect of Condition was not significant for any of the Talker Groups: Native English (β = .02, SE = .02, t ratio = 1.12, p = .27), Native Mandarin-High (β = -.03, SE = .02, t ratio = -1.43, p = .15), Native Mandarin-Low (β = .03, SE = .02, t ratio = 1.26, p = .21).

These results suggest that talkers distinguished word-final /b/ and /p/ with voicing proportions of the target consonants (C2); C2 voicing proportions were larger for /b/ than

for /p/. This difference was larger for Native English talkers' productions than for Native Mandarin-High talkers' productions, and for Native Mandarin-High talkers' productions than for Native Mandarin-Low talkers' productions. This difference among different talker groups was influenced by how talkers produced /b/, not /p/. That is, voicing proportions of /b/ were larger for Native English talkers' productions than for Native Mandarin-High talkers' productions, and for Native Mandarin-High talkers' productions than for Native Mandarin-Low talkers' productions. Furthermore, Native English talkers and Native Mandarin-High talkers manipulated C2 voicing proportions in different conditions (No Context vs. Context), and they did this differently for /b/ and /p/. That is, these talkers increased voicing proportions from No Context to Context conditions for /b/, but not for /p/. However, Native Mandarin-Low talkers did not manipulate C2 voicing proportions differently in different conditions.

4.3.2.3. Summary of the segmental analyses for consonant targets

The results of the consonant segmental analyses showed that the talkers' ability to manipulate acoustic features of a non-native contrast differed depending on whether the contrast exists in the talkers' native language or not. That is, Native Mandarin (both High and Low) talkers were able to distinguish the L1L2 contrast (/p/-/b/ in word-initial position) better than Native English talkers, and further clarify the distinction as well as Native English talkers. This was evidenced in the larger difference in normalized VOTs between /p/ and /b/ for Native Mandarin talkers' productions than for Native English talkers' productions. Further, talkers increased normalized VOTs of /p/ from No Context to Context conditions, and the size of this increase was similar for Native Mandarin (High and Low)

talkers' and Native English talkers' productions. However, the extent that talkers manipulated acoustic features to distinguish the L2-only consonant contrast (/p/-/b/ in word-final position) differed by talkers' native language as well as their target language proficiency level. That is, talkers generally distinguished word-final /b/ and /p/ with preceding vowel durations and C2 voicing proportions, but the difference in these acoustic features was larger for Native English talkers' productions than for Native Mandarin-High talkers', and for Native Mandarin-High talkers' productions than for Native Mandarin-Low talkers'. Further, Native English and Native Mandarin-High talkers used C2 voicing proportions (for /b/) rather than preceding vowel durations to make contextually-relevant enhancements (i.e., No Context vs. Context conditions). The ability to enhance this contrast also differed for different Talker Groups; Native English and Native Mandarin-High talkers increased C2 voicing proportions for /b/ from No Context to Context conditions; though Native Mandarin-Low talkers did not manipulate C2 voicing proportions. Thus, these results suggest that the effect of talkers' target language experience impacted productions of the English /p/-/b/ contrast in word-initial vs. word-final position differently in terms of their ability to produce the contrast, as well as their ability to further enhance the contrast.

4.3.3. Vowel targets: Global (phrase-, and word-level) analyses

For the items that contained targets with vowel contrasts (i.e., L1L2 vowel contrast: /ai/-/ei/, L2-only vowel contrast: /i/-/ɪ/), we examined duration, mean F0, and mean intensity at the phrase-level and word-level. In the linear-mixed effects regression models used to analyze these features, fixed effects were Type (L1L2, L2-only), Condition (Context, No Context), and Talker Group (Native English, Native Mandarin-High, and

Native Mandarin-Low); these factors were contrast coded as specified above. Models also included the maximal random effects structure that would converge, which included random intercepts for talker and item. The random effects structure also included random slopes (see the model syntax in each table below); in each model, the random effects that did not account for any variance was dropped to avoid overfitting.

4.3.3.1. Phrase-level analyses

4.3.3.1.1. Duration

The left panel in Figure 4.8 shows the raw durations (msec.) of the phrase, *"Click on the ___ now"*, by Talker Group (Native English, Native Mandarin-High, Native Mandarin-Low). The linear mixed-effects regression model, used to analyze raw phrase durations as the dependent variable, included Talker Group as a fixed factor (see Table 4.9 for the model syntax and summary of the results). The model showed significant effects of Talker Group comparisons: Native English vs. Native Mandarin-High (β = -356,66, t = -3.24, $p < .01$), Native Mandarin-High vs. Native Mandarin-Low (β = -294.14, t = -2.41, $p < .05$). This indicates that Native Mandarin-Low talkers produced the phrases with longer durations than Native Mandarin-High talkers did; Native Mandarin-High talkers produced the phrases with longer durations than Native English talkers did.

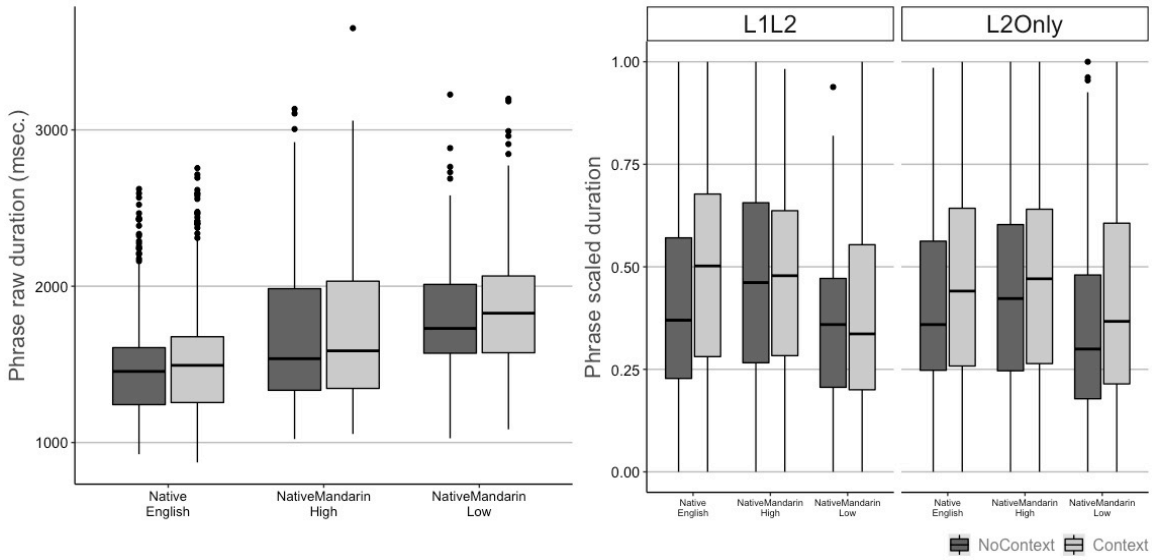

Figure 4.8. Durations of phrases containing vowel targets for different talker groups (Native English, Native Mandarin-High, and Native-Mandarin-Low talkers) in different conditions (No Context and Context): raw durations (Left panel), scaled durations (for phrases containing L1L2 or L2-only contrasts: Right panel).

The right panel in Figure 4.8 shows the scaled phrase durations by Talker Group (Native English, Native Mandarin-High, Native Mandarin-Low), Type (L1L2, L2-only), and Condition (No context, Context). The linear mixed-effects regression model, used to analyze scaled phrase durations as the dependent variable, included Talker Group, Type, and Condition, as well as interactions among these factors as fixed effects (see Table 4.9 for the model syntax and summary of the results). There was a significant effect of Condition (No Context, Context; $\beta = .06$, $t = 3.13$, $p < .01$). This indicates that talkers produced the phrases with longer durations in Context conditions than those in No Context conditions.

Table 4.9. Summary of the linear mixed-effects regression model for raw durations (msec.) and scaled durations of the phrases with vowel targets.

Raw phrase duration Model `Raw duration (msec.) ~ TalkerGroup + (1| Talker) + (1| Word)`				
Fixed Effects	**Estimate**	**S.E.**	**t-val.**	***p*-val.**
(Intercept)	1688.68	42.12	40.1	< .001
TalkerGroup1 (Native English vs. Native Mandarin-High)	-356.66	110.04	-3.24	.002 **
TalkerGroup2 (Native Mandarin-High vs. Native Mandarin-Low)	-294.15	122.1	-2.41	.018 *
Scaled phrase duration Model `Scaled duration ~ TalkerGroup*Type*Condition + (1| Talker) + (1| Word)`				
(Intercept)	.42	.02	28.2	< .001
TalkerGroup1 (Native English vs. Native Mandarin-High)	.03	.03	.81	.42
TalkerGroup2 (Native Mandarin-High vs. Native Mandarin-Low)	.09	.04	2.46	.016 *
Type (L1L2 vs L2-only)	.007	.02	.4	.69
Condition (No Context vs. Context)	.06	.02	3.13	.003 **
TalkerGroup1: Type	.05	.03	1.53	.13
TalkerGroup2: Type	.06	.03	1.78	.076
TalkerGroup1: Condition	.04	.03	1,2	.23
TalkerGroup2: Condition	-.04	.03	-1.08	.28
Type: Condition	-.02	.04	-.52	.61
TalkerGroup1: Type: Condition	.11	.06	1.82	.069
TalkerGroup2: Type: Condition	.07	.07	1.0	.32

4.3.3.1.2. Mean F0

The left panel in Figure 4.9 shows scaled values of mean F0 of the phrase, *"Click on the ___ now"*, by Talker Group (Native English, Native Mandarin-High, Native Mandarin-Low), Type (L1L2, L2-only), and Condition (No context, Context). The linear mixed-effects regression model, used to analyze scaled mean F0 as the dependent variable, included Talker Group, Type, and Condition, as well as interactions among these factors as fixed effects (see Table 4.10 for the model syntax and summary of the results). There was a significant effect of Type (L1L2, L2-only contrast; $\beta = -.11$, $t = -3.98$, $p < .001$). This

indicates that phrases containing target words with the L2-only vowel contrast (/i/-/ɪ/) were produced with higher mean F0 than those containing target words with the L1L2 vowel contrast (/ai/-/ei/). The figure also suggests that Native English and Native Mandarin-High talkers increased the mean F0 from No Context to Context conditions for phrases with the L2-only contrast (/i/-/ɪ/); though this pattern was the opposite for Native Mandarin-Low talkers' productions. This was reflected in the interaction among Type (L1L2, L2-only), Condition (No Context, Context), and the Native Mandarin-High vs. Native Mandarin-Low group comparison ($\beta = -.16$, $t = -2.4$, $p < .05$).

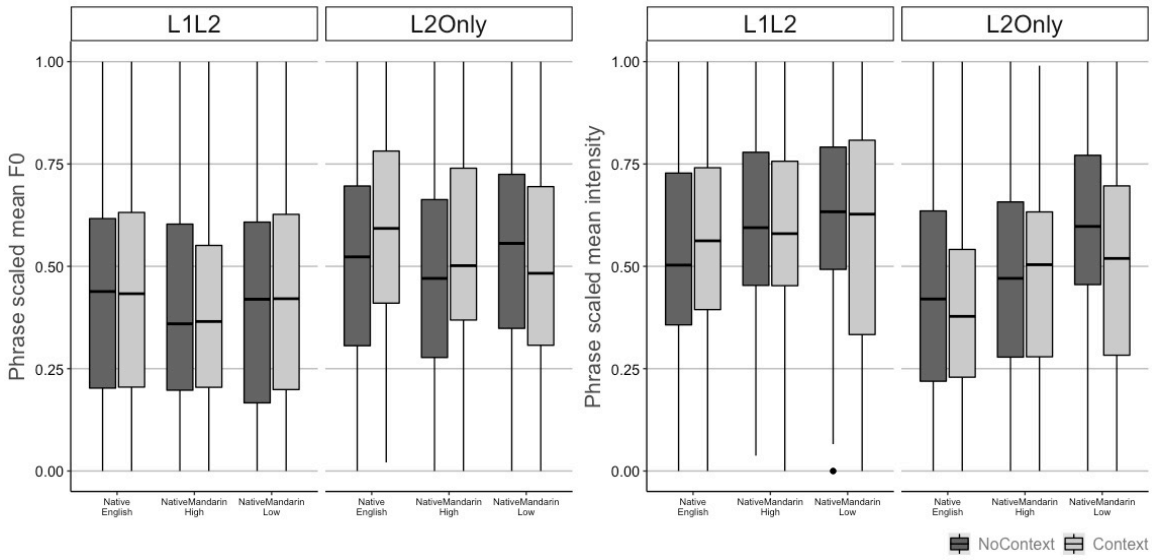

Figure 4.9. Scaled mean F0 (Left panel) and scaled mean intensity (Right panel) for phrases containing vowel targets (with L1L2 and L2-only contrasts) for different talker groups (Native English, Native Mandarin-High, and Native-Mandarin-Low talkers) in different conditions (No Context and Context).

Table 4.10. Summary of the linear mixed-effects regression model for scaled mean F0 and scaled mean intensity of the phrases with vowel targets.

Scaled phrase mean F0 Model `Scaled mean F0 ~ TalkerGroup*Type*Condition + (1| Talker) + (1| Word)`				
Fixed Effects	**Estimate**	**S.E.**	**t-val.**	***p*-val.**
(Intercept)	.47	.02	24.67	< .001
TalkerGroup1 (Native English vs. Native Mandarin-High)	.04	.04	1.06	.29
TalkerGroup2 (Native Mandarin-High vs. Native Mandarin-Low)	-.0004	.04	-.01	.99
Type (L1L2 vs L2-only)	-.11	.03	-3.98	< .001 ***
Condition (No Context vs. Context)	.008	.03	.3	.77
TalkerGroup1: Type	-.009	.03	-.29	.77
TalkerGroup2: Type	-.02	.03	-.59	.56
TalkerGroup1: Condition	.04	.03	1.39	.17
TalkerGroup2: Condition	.05	.03	1.42	.16
Type: Condition	-.04	.06	-.64	.52
TalkerGroup1: Type: Condition	-.1	.06	-1.63	.1
TalkerGroup2: Type: Condition	-.16	.07	-2.4	.016 *
Scaled phrase mean intensity Model `Scaled mean intensity ~ TalkerGroup*Type*Condition + (1| Talker) + (1| Word)`				
(Intercept)	.53	.02	29.47	< .001
TalkerGroup1 (Native English vs. Native Mandarin-High)	-.1	.03	-3.64	< .001 ***
TalkerGroup2 (Native Mandarin-High vs. Native Mandarin-Low)	-.08	.03	-2.75	.007 **
Type (L1L2 vs L2-only)	.11	.03	3.5	.001 **
Condition (No Context vs. Context)	-.03	.03	-.83	.41
TalkerGroup1: Type	.06	.03	2.0	.046 *
TalkerGroup2: Type	.08	.03	2.38	.017 *
TalkerGroup1: Condition	.03	.03	1.12	.26
TalkerGroup2: Condition	.06	.03	1.96	.05
Type: Condition	.04	.06	.66	.51
TalkerGroup1: Type: Condition	.06	.06	1.06	.29
TalkerGroup2: Type: Condition	-.03	.06	-.3	.77

4.3.3.1.3. Mean intensity

The right panel in Figure 4.9 shows scaled values of mean intensity of the phrase, *"Click on the ___ now"*, by Talker Group (Native English, Native Mandarin-High, Native Mandarin-Low), Type (L1L2, L2-only), and Condition (No context, Context). The linear

mixed-effects regression model, used to analyze scaled mean intensity as the dependent variable, included Talker Group, Type, and Condition, as well as interactions among these factors as fixed effects (see Table 4.10 for the model syntax and summary of the results). There was a significant effect of Type (L1L2, L2-only contrast; $\beta = .11$, $t = 3.5$, $p < .01$). This indicates that talkers produced the phrases with higher intensity for those containing target words with the L1L2 vowel contrast (/ai/-/ei/) compared to those containing target words with the L2-only vowel contrast (/i/-/ɪ/). This difference in mean intensity based on Type was larger for Native English talkers' productions than for Native Mandarin-High talkers' productions (Type x Native English vs. Native Mandarin-High: $\beta = .06$, $t = 2.0$, $p < .05$), and for Native Mandarin-High talkers' productions than for Native Mandarin-Low talkers' productions (Type x Native Mandarin-High vs. Native Mandarin-Low: $\beta = .08$, $t = 2.38$, $p < .05$).

Together, acoustic analyses of the phrases containing vowel targets demonstrated that global characteristics of the entire phrases were impacted by talkers' target language proficiency level, the type of the target vowel contrast (L1L2, L2-only vowel contrasts), and the production condition (Context, No Context conditions). Specifically, talkers' target language experience impacted phrase durations (raw durations: Native Mandarin-Low > Native Mandarin-High > Native English). The difference in the type of production conditions (No Context vs. Context) only affected the phrase duration; talkers produced the phrases in Context conditions with longer durations than in No Context conditions. Further, the difference in the types of vowel contrast in the target word (L1L2 vs. L2-only) was manifested in mean F0 and mean intensity of the entire phrase. Specifically, talkers produced the phrases with lower mean F0 and higher mean intensity for those containing

the L1L2 vowel contrast (/ai/-/ei/) compared to those containing the L2-only vowel contrast (/i/-/ɪ/). These results suggest that talkers' efforts to communicate target word clearly were manifested in phrase durations, but not in other aspects such as mean F0 and mean intensity. Global characteristics of mean F0 and mean intensity seemed to reflect the characteristics of the target vowels themselves (e.g., the /ai/-/ei/ contrast and /i/-/ɪ/ contrast) rather than the difference in Talker Groups or production conditions (Context vs. No Context).

4.3.3.2. Word-level analyses

4.3.3.2.1. Duration

The left panel in Figure 4.10 shows the raw durations (msec.) of the target words containing vowel contrasts by Talker Group (Native English, Native Mandarin-High, Native Mandarin-Low). The linear mixed-effects regression model, used to analyze raw word durations as the dependent variable, included Talker Group as a fixed factor (see Table 4.11 for the model syntax and summary of the results). The model showed no significant effects of Talker Group comparisons: Native English vs. Native Mandarin-High ($\beta = 22.84$, $t = .92$, $p = .36$), Native Mandarin-High vs. Native Mandarin-Low ($\beta = 43.61$, $t = 1.58$, $p = .12$). This indicates that word durations did not differ for different Talker Groups.

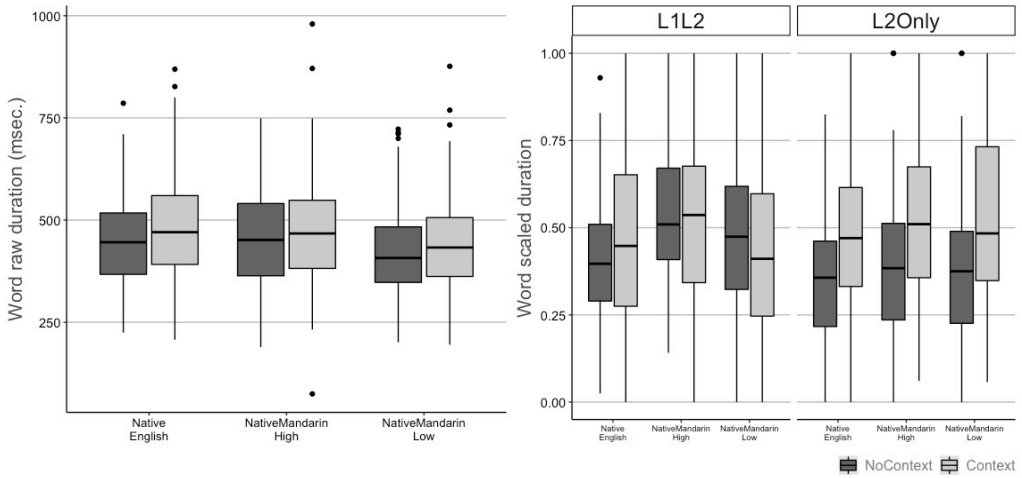

Figure 4.10. Durations of target words with vowel contrasts for different talker groups (Native English, Native Mandarin-High, and Native-Mandarin-Low talkers) in different conditions (No Context and Context): raw durations (Left panel), scaled durations (for targets containing L1L2 or L2-only contrasts: Right panel).

Table 4.11. Summary of the linear mixed-effects regression model for durations (msec.) and scaled durations of the words with vowel contrasts.

Raw word duration Model `Raw duration (msec.) ~ TalkerGroup + (1| Talker) + (1+ TalkerGroup| Word)`				
Fixed Effects	**Estimate**	**S.E.**	**t-val.**	***p*-val.**
(Intercept)	452.77	11.84	38.24	< .001
TalkerGroup1 (Native English vs. Native Mandarin-High)	22.84	24.85	.92	.36
TalkerGroup2 (Native Mandarin-High vs. Native Mandarin-Low)	43.61	27.61	1.58	.12
Scaled word duration Model `Scaled duration ~ TalkerGroup*Type*Condition + (1+ Type+ Condition| Talker) + (1+ TalkerGroup| Word)`				
(Intercept)	.46	.02	22.65	< .001
TalkerGroup1 (Native English vs. Native Mandarin-High)	-.06	.03	-2.18	.031 *
TalkerGroup2 (Native Mandarin-High vs. Native Mandarin-Low)	-.0005	.03	-.02	.99
Type (L1L2 vs L2-only)	.03	.04	.86	.39
Condition (No Context vs. Context)	.07	.04	1.83	.074
TalkerGroup1: Type	-.03	.04	-.75	.46
TalkerGroup2: Type	.06	.05	1.25	.22
TalkerGroup1: Condition	.05	.04	1.34	.19
TalkerGroup2: Condition	.02	.04	.45	.65
Type: Condition	-.15	.07	-2.03	.049 *
TalkerGroup1: Type: Condition	.14	.07	1.94	.06
TalkerGroup2: Type: Condition	.11	.09	1.25	.22

The right panel in Figure 4.10 shows the scaled word durations by Talker Group (Native English, Native Mandarin-High, Native Mandarin-Low), Type (L1L2, L2-only), and Condition (No context, Context). The linear mixed-effects regression model, used to analyze scaled word durations as the dependent variable, included Talker Group, Type, and Condition, as well as interactions among these factors as fixed effects (see Table 4.11 for the model syntax and summary of the results). There was a significant interaction between Type (L1L2, L2-only) and Condition (No Context, Context; $\beta = -.15$, $t = -2.03$, $p < .05$). This indicates that for target words with the L2-only vowel contrast (/i/-/ɪ/), talkers produced the words with longer durations in Context conditions than those in No Context conditions; though they did so less for target words with the L1L2 vowel contrast (/ai/-/ei/).

4.3.3.2.2. Mean F0

The left panel in Figure 4.11 shows scaled mean F0 of the target words containing vowel contrasts, by Talker Group (Native English, Native Mandarin-High, Native Mandarin-Low), Type (L1L2, L2-only), and Condition (No context, Context). The linear mixed-effects regression model, used to analyze scaled mean F0 as the dependent variable, included Talker Group, Type, and Condition, as well as interactions among these factors as fixed effects (see Table 4.12 for the model syntax and summary of the results). There was a significant effect of Type (L1L2, L2-only; $\beta = -.16$, $t = -5.95$, $p < .001$). This indicates that talkers produced target words with the L2-only vowel contrast (/i/-/ɪ/) with higher mean F0 than those with the L1L2 vowel contrast (/ai/-/ei/). The figure also suggests that for target words with the L2-only vowel contrast (/i/-/ɪ/), Native English talkers increased mean F0

from No Context to Context conditions, though Native Mandarin talkers did not. This was reflected in the significant interaction among Type, Condition, and the Native English vs. Native Mandarin-High group comparison (β = -.11, t = -1.98, p < .05).

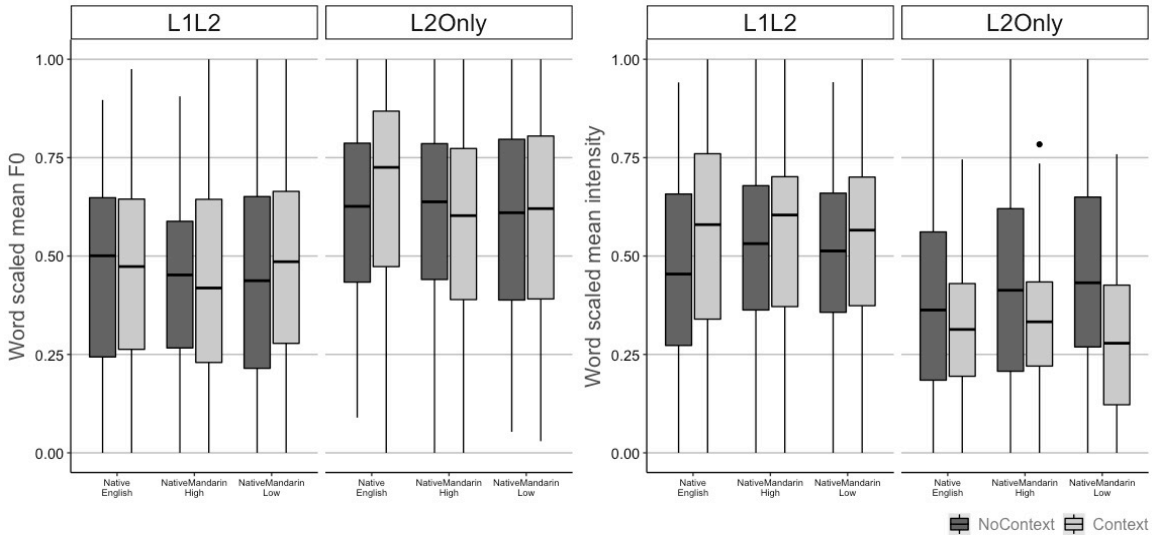

Figure 4.11. Scaled mean F0 (Left panel) and scaled mean intensity (Right panel) for target words containing vowel contrasts (with L1L2 and L2-only contrasts) for different talker groups (Native English, Native Mandarin-High, and Native-Mandarin-Low talkers) in different conditions (No Context and Context).

4.3.3.2.3. Mean intensity

The right panel in Figure 4.11 shows scaled mean intensity of target words with vowel contrasts, by Talker Group (Native English, Native Mandarin-High, Native Mandarin-Low), Type (L1L2, L2-only), and Condition (No context, Context). The linear mixed-effects regression model, used to analyze scaled mean intensity as the dependent variable, included Talker Group, Type, and Condition, as well as interactions among these factors as fixed effects (see Table 4.12 for the model syntax and summary of the results). There was a significant effect of Type (L1L2, L2-only: β = .15, t = 2.92, p < .01). This indicates that mean intensity was higher for target words containing the L1L2 vowel contrast (/ai/-/ei/) than for those containing the L2-only vowel contrast (/i/-/ɪ/). There was

also a significant interaction between Condition (No Context, Context) and the Native English vs. Native Mandarin-High group comparison (β = .06, t = 2.05, $p < .05$), and a marginally significant interaction between Condition and the Native Mandarin-High vs. Native Mandarin-Low group comparison (β = .06, t = 1.88, $p = .068$).

Table 4.12. Summary of the linear mixed-effects regression model for scaled mean F0 and scaled mean intensity of the words with vowel contrasts.

Scaled word mean F0 Model `Scaled mean F0 ~ TalkerGroup*Type*Condition + (1| Talker) + (1| Word)`				
Fixed Effects	**Estimate**	**S.E.**	**t-val.**	***p*-val.**
(Intercept)	.53	.02	25.84	< .001
TalkerGroup1 (Native English vs. Native Mandarin-High)	.04	.04	1.02	.31
TalkerGroup2 (Native Mandarin-High vs. Native Mandarin-Low)	.02	.05	.41	.68
Type (L1L2 vs L2-only)	-.16	.03	-5.95	< .001 ***
Condition (No Context vs. Context)	.008	.03	.31	.76
TalkerGroup1: Type	-.05	.03	-1.79	.074
TalkerGroup2: Type	-.04	.03	-1.38	.17
TalkerGroup1: Condition	.04	.03	1.44	.15
TalkerGroup2: Condition	-.009	.03	-.31	.76
Type: Condition	-.01	.05	-.28	.78
TalkerGroup1: Type: Condition	-.11	.05	-1.98	.048 *
TalkerGroup2: Type: Condition	-.05	.06	-.9	.37
Scaled word mean intensity Model `Scaled mean intensity ~ TalkerGroup*Type*Condition + (1+ Type| Talker) + (1+ TalkerGroup| Word)`				
(Intercept)	.44	.03	16.86	< .001
TalkerGroup1 (Native English vs. Native Mandarin-High)	-.03	.02	-1.38	.17
TalkerGroup2 (Native Mandarin-High vs. Native Mandarin-Low)	-.004	.03	-.16	.88
Type (L1L2 vs L2-only)	.15	.05	2.92	.006 **
Condition (No Context vs. Context)	-.03	.05	-.54	.59
TalkerGroup1: Type	.02	.03	.64	.53
TalkerGroup2: Type	-.005	.03	-.16	.88
TalkerGroup1: Condition	.06	.03	2.05	.048 *
TalkerGroup2: Condition	.06	.03	1.88	.068
Type: Condition	.16	.1	1.63	.11
TalkerGroup1: Type: Condition	.002	.06	.03	.97
TalkerGroup2: Type: Condition	-.05	.07	-.78	.44

These results indicate that Native Mandarin talkers produced target words with lower mean intensity in Context conditions compared to No Context conditions, though Native English talkers did not. This tendency was stronger for Native Mandarin-Low talkers' productions than for Native Mandarin-High talkers' productions.

In sum, acoustic analyses of the target words containing vowel contrasts demonstrated that raw word durations did not differ for different talker groups' productions. Overall, the effect of Condition (No Context vs. Context) somewhat differed for target words with different types of vowel contrasts. That is, for the words with the L2-only vowel contrast (/i/-/ɪ/), talkers increased durations and decreased mean intensity from No Context to Context conditions. For the words with the L1L2 vowel contrast (/ai/-/ei/), talkers did not change durations but increased mean intensity from No Context to Context conditions. In terms of mean F0, conditions did not influence it but the contrast type did; mean F0 was higher for the words with the L2-only vowel contrast (/i/-/ɪ/) than for the words with the L1L2 vowel contrast (/ai/-/ei/). These results suggest that acoustic characteristics of the target words were greatly influenced by the type of vowels contained in the words. Further, talkers' efforts to produce target words clearly manifested differently depending on the type of vowel segments in the target words.

4.3.3.3. Summary of the global analyses for vowel targets

Taken together, phrase-level and word-level analyses of the items containing targets with vowel contrasts demonstrated that the type of vowel contrasts (i.e., /ai/-/ei/ contrast or /i/-/ɪ/ contrast) consistently impacted the global characteristics, and also how Condition (No Context vs. Context) influenced the productions. This was observed in the results where the

effect of Condition (No Context vs. Context) influenced the global characteristics differently for items with the L1L2 vowel contrast (/ai/-/ei/) vs. those with the L2-only vowel contrast (/i/-/ɪ/). That is, talkers increased durations from No Context to Context conditions for phrases and target words with the L2-only vowel contrast (/i/-/ɪ/), but not for those with the L1L2 vowel contrast (/ai/-/ei/). Talkers also increased mean intensity from No Context to Context conditions for target words with the L1L2 vowel contrast (/ai/-/ei/), but decreased intensity from No Context to Context conditions for target words with the L2-only vowel contrast (/i/-/ɪ/). Perhaps, as a result of these patterns in intensity of target words, phrases with the L1L2 vowel contrast (/ai/-/ei/) had higher mean intensity than phrases with the L2-only vowel contrast (/i/-/ɪ/). Mean F0 of the target words and phrases was only influenced by the type of the contrast; words and phrases with the L2-only vowel contrast (/i/-/ɪ/) had higher mean F0 than those with the L1L2 vowel contrast (/ai/-/ei/).

These results suggest that global characteristics of the items containing vowel targets were greatly influenced by the type of the vowel in the item (e.g., duration, vowel height). Particularly, it is possible that F0 and intensity, as well as changes in these values across different production conditions, were influenced by physiological/articulatory aspects of these vowels. This will be further explored in the discussion section.

4.3.4. Vowel targets: Segmental analyses

In order to characterize segmental features of the target vowel contrasts, we analyzed several acoustic features. For the L1L2 vowel contrast (i.e., /ai/-/ei/), we analyzed normalized durations as well as F1 and F2 values at the initial state (30%). For the L2-only vowel contrast (i.e., /i/-/ɪ/), we analyzed normalized durations as well as F1 and F2 values

at the midpoint (50%). For each acoustic feature, we first present the results of a linear mixed-effects regression model, examining whether talkers made differences in the acoustic measure (e.g., normalized durations of /ai/ and /ei/) to distinguish one phoneme (e.g., /ai) from another (e.g., /ei/), and whether the size of this difference was larger for one Talker Group's production than for another (e.g., Native Mandarin-High vs. Native Mandarin-Low). Then, we present the results of linear mixed-effects regression models, examining the effect of Condition (No Context vs. Context) separately for the two phonemes (e.g., /ai/ and /ei/), as well as whether the effect of Condition differed for different Talker Groups' productions.

In the linear-mixed effects regression models presented below, fixed effects were Phoneme (/ai/-/ei/ or /i/-/ɪ/), Condition (Context, No Context), and Talker Group (Native English, Native Mandarin-High, and Native Mandarin-Low); different combinations of these fixed effects were included in different models (see the description of each model below). Condition and Talker Group were contrast coded as specified above. For the L1L2 vowel contrast, Phoneme was contrast coded to compare between /ai/-targets (.5) and /ei/-targets (-.5). For the L2-only vowel contrast, Phoneme was contrast coded to compare between /i/-targets (.5) and /ɪ/-targets (-.5). Models also included the maximal random effects structure that would converge, which included random intercepts for talker and item. The random effects structure also included a by-talker random slope for Phoneme or Condition (different depending on the model) and a by-word intercept for Talker Group. In each model, the random effects that did not account for any variance was dropped to avoid overfitting; see the description of each model below.

4.3.4.1. L1L2 vowel contrast (/ai/-/ei/)

4.3.4.1.1. Normalized vowel duration

Figure 4.12 shows the mean normalized durations of the vowels in the L1L2 vowel contrast by Phoneme (/ai/, /ei/: "AI" and "EI" in the figure), Talker Group (Native English, Native Mandarin-High, Native Mandarin-Low), and Condition (No context, Context). The figure suggests that vowel durations did not differ for different vowels (/ai/, /ei/). In order to statistically examine whether talkers made a difference in durations to distinguish the two vowels, we implemented a linear mixed-effects regression model with normalized vowel duration as the dependent variable. The fixed factors were Phoneme, Talker Group, as well as interactions between the two factors as fixed effects (see Table 4.13 for the model syntax and summary of the results). The effect of Phoneme was not significant ($\beta = .02$, $t = .25$, $p = .8$). This indicates that normalized durations did not differ between /ai/ and /ei/. There were significant effects of Talker Group comparisons: Native English vs. Native Mandarin-High ($\beta = .06$, $t = 3.7$, $p < .001$), Native Mandarin-High vs. Native Mandarin-Low ($\beta = .08$, $t = 3.85$, $p < .001$). These results indicate that vowel durations were longer for Native English talkers' productions than for Native Mandarin-High talkers' productions, and for Native Mandarin-High talkers' productions than for Native Mandarin-Low talkers' productions.

Figure 4.12 also suggests that talkers somewhat increased the vowel durations from No Context to Context conditions, and they did so to similar extents for /ai/ and /ei/. The linear mixed-effects regression models, used to analyze normalized VOTs as the dependent variable separately for /ai/-targets and /ei/-targets, included Condition (No Context, Context), Talker Group, as well as interactions between the two as fixed effects (see Table

4.13 for the model syntax and summary of the results).

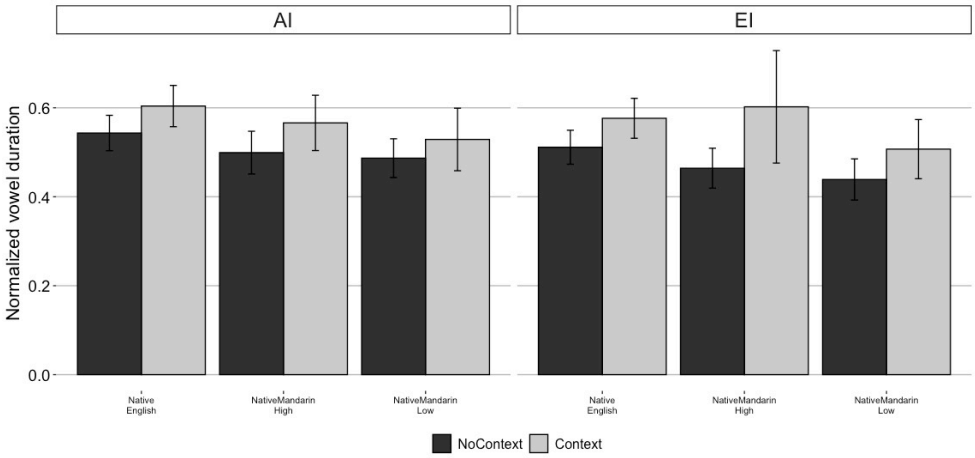

Figure 4.12. Mean normalized durations of the vowels in the L1L2 vowel contrast by Phoneme (/ai/: AI, /ei/: EI), Talker Group (Native English, Native Mandarin-High, Native Mandarin-Low), and Condition (No context, Context). The error bars represent the 95% confidence interval of the mean.

For the model with /ai/-targets, there were significant effects of Talker Group comparisons: Native English vs. Native Mandarin-High ($\beta = .09$, $t = 5.18$, $p < .001$), Native Mandarin-High vs. Native Mandarin-Low ($\beta = .07$, $t = 4.03$, $p < .001$). The effect of Condition was not significant ($\beta = .06$, $t = .5$, $p = .63$). These results indicate that normalized vowel durations of /ai/ were longer for Native English talkers' productions than for Native Mandarin-High talkers' productions, and for Native Mandarin-High talkers' productions than for Native Mandarin-Low talkers' productions. However, normalized vowel durations did not differ in different types of Conditions (No Context, Context). Similarly, for the model with /ei/-targets, there were significant effects of Talker Group comparisons: Native English vs. Native Mandarin-High ($\beta = .06$, $t = 2.19$, $p < .05$), Native Mandarin-High vs. Native Mandarin-Low ($\beta = .09$, $t = 2.0$, $p < .01$). The effect of Condition was not significant ($\beta = .09$, $t = .77$, $p = .46$). Thus, though normalized durations of /ei/ differed for different Talker Groups, they did not differ in different Conditions.

Table 4.13. Summary of the linear mixed-effects regression models for normalized durations of the vowels in the L1L2 vowel contrast (/ai/-/ei/).

Fixed Effects	**Estimate**	**S.E.**	**t-val.**	***p*-val.**
Normalized vowel durations for /ai/ and /ei/ Model `Normalized vowel duration ~ Phoneme*TalkerGroup + (1+ Phoneme| Talker)`				
(Intercept)	.53	.04	12.62	< .001
Phoneme (/ai / vs. /ei/)	.02	.09	.25	.8
TalkerGroup1 (Native English vs. Native Mandarin-High)	.06	.02	3.7	< .001 ***
TalkerGroup2 (Native Mandarin-High vs. Native Mandarin-Low)	.08	.02	3.85	< .001 ***
Phoneme: TalkerGroup1	.02	.03	.7	.29
Phoneme: TalkerGroup2	-.02	.03	-.79	.43
Normalized vowel durations for /ai/ Model `Normalized vowel duration ~ Condition*TalkerGroup + (1+ Condition| Talker)`				
(Intercept)	.54	.06	9.19	< .001
Condition (No Context vs. Context)	.06	.12	.5	.63
TalkerGroup1 (Native English vs. Native Mandarin-High)	.08	.01	5.18	< .001 ***
TalkerGroup2 (Native Mandarin-High vs. Native Mandarin-Low)	.07	.02	4.03	< .001 ***
TalkerGroup1: Condition	-.001	.02	-.05	.96
TalkerGroup2: Condition	.02	.02	1.07	.28
Normalized vowel durations for /ei/ Model `Normalized vowel duration ~ Condition*TalkerGroup + (1| Talker)`				
(Intercept)	.52	.06	9.11	< .001
Condition (No Context vs. Context)	.09	.11	.77	.46
TalkerGroup1 (Native English vs. Native Mandarin-High)	.06	.03	2.19	.032 *
TalkerGroup2 (Native Mandarin-High vs. Native Mandarin-Low)	.09	.03	3.0	.004 **
TalkerGroup1: Condition	-.04	.05	-.89	.37
TalkerGroup2: Condition	.05	.05	.91	.37

These results suggest that talkers did not distinguish /ai/ and /ei/ with vowel durations; normalized vowel durations for these vowels were similar. Normalized vowel durations, though, differed for different Talker Groups; Native English talkers generally produced longer vowels than Native Mandarin-High talkers did, and Native Mandarin-High talkers produced longer vowels than Native Mandarin-Low talkers did. While there

was a tendency for talkers to increase the vowel durations from No Context to Context conditions for both /ai/ and /ei/, this pattern was not statistically significant.

4.3.4.1.2. Initial F1 and F2

Figure 4.13 shows the mean F1 and F2 values at 30% of the vowels in the L1L2 vowel contrast by Phoneme (/ai/, /ei/; AI and EI in the figure), Talker Group (Native English, Native Mandarin-High, Native Mandarin-Low), and Condition (No context, Context). The figure suggests that F1 values may be generally lower (the tongue position is higher) for /ei/ than for /ai/. The linear mixed-effects regression model, used to analyze formant values as the dependent variable, included Phoneme, Talker Group, Formant (F1, F2) as well as interactions among these factors as fixed effects (see Table 4.14 for the model syntax and summary of the results). The Formant variable was contrast coded to compare between F1 (-.5) and F2 (.5); other factors were contrast coded as specified above. The results showed a non-significant effect of Phoneme (/ai/, /ei/; $\beta = .11$, $t = 1.69$, $p = .11$), Formant (F1, F2; $\beta = 0005$, $t = .01$, $p = .99$), and the interaction between the two ($\beta = -.02$, $t = -.22$, $p = .82$). This indicates that between /ei/ and /ai/, F1 and F2 values at the initial state did not significantly differ.

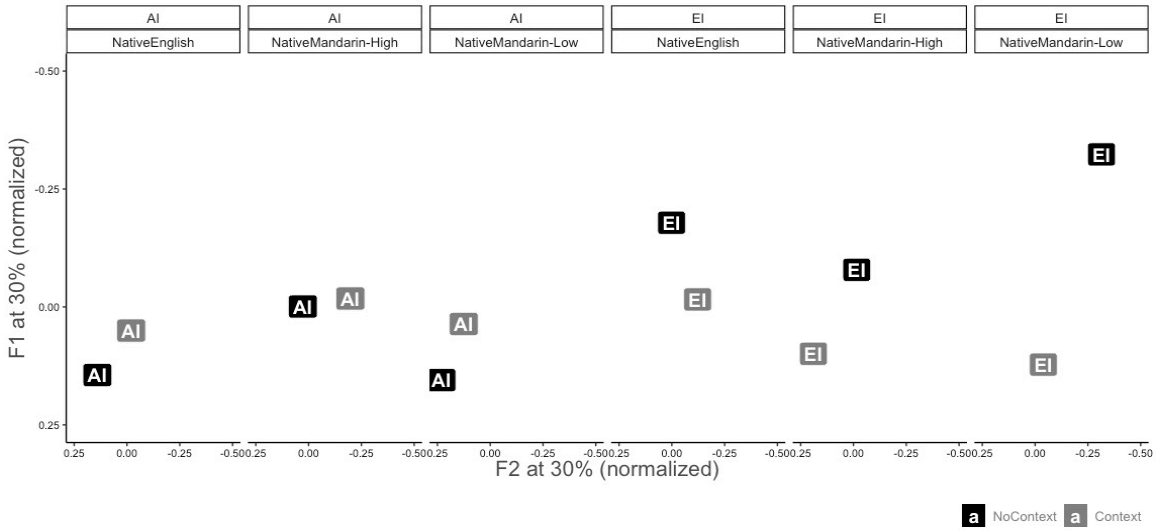

Figure 4.13. Mean F1 and F2 values at 30% of the vowels in the L1L2 vowel contrast by Phoneme (/ai/, /ei/; AI and EI in the figure), Talker Group (Native English, Native Mandarin-High, Native Mandarin-Low), and Condition (No context, Context).

Table 4.14. Summary of the linear mixed-effects regression model for normalized F1 and F2 values at 30% of the vowels in the L1L2 vowel contrast (/ai/-/ei/): Model for both /ai/ and /ei/.

Normalized F1 and F2 at 30% for /ai/ and /ei/ Model `Formant value ~ Phoneme*TalkerGroup*FormantType + (1| Word)`				
Fixed Effects	**Estimate**	**S.E.**	**t-val.**	***p*-val.**
(Intercept)	.0007	.03	.02	.98
Phoneme (/ai / vs. /ei/)	.11	.06	1.69	.11
TalkerGroup1 (Native English vs. Native Mandarin-High)	-.001	.06	-.02	.98
TalkerGroup2 (Native Mandarin-High vs. Native Mandarin-Low)	-.003	.07	-.04	.97
Formant type (F1 vs. F2)	.0005	.05	.01	.99
Phoneme: TalkerGroup1	.1	.13	.76	.45
Phoneme: TalkerGroup2	-.31	.15	-2.14	.033 *
Phoneme: Formant type	-.02	.1	-.22	.82
TalkerGroup1: Formant type	-.001	.13	-.008	.99
TalkerGroup2: Formant type	-.002	.15	-.02	.99
Phoneme: TalkerGroup1: Formant type	-.1	.26	-.4	.69
Phoneme: TalkerGroup2: Formant type	-.37	.3	-1.26	.21

In order to examine the effect of Condition (No Context, Context) on F1 and F2 values separately for /ai/-targets and for /ei/-targets, we analyzed the formant data for these vowels in separate linear mixed-effects regression models. The fixed factors were Formant (F1, F2), Condition (No Context, Context), Talker Group (Native English, Native Mandarin-High, and Native Mandarin-Low), and the interaction among these factors (see Table 4.15 for the model syntax and summary of the results). For the model with /ai/-targets, none of the fixed factors were significant (see Table 4.15). This indicates that F1 and F2 values at the 30% of /ai/ did not significantly differ between Context and No Context conditions. For the model with /ei/-targets, there was a significant effect of Condition (No Context, Context; $\beta = .19$, $t = 2.39$, $p < .05$). This effect of Condition did not interact with Formant (F1, F2; $\beta = .01$, $t = .18$, $p = .11$). This indicates that from No Context to Context conditions, talkers generally increased F1 (lowered the tongue position) and F2 (fronted the tongue position) of /ei/.

These results suggest that talkers did not differentiate F1 and F2 values at the initial state of /ai/ and /ei/. However, talkers manipulated initial F1 and F2 values of /ei/ more than those of /ai/, in order to clarity the distinction between the two vowels in Context conditions. That is, talkers increased F1 (lowered the tongue position) and F2 (fronted the tongue position) of /ei/ from No Context to Context conditions; though they did not change initial F1 and F2 values of /ai/ between the two conditions.

Table 4.15. Summary of the linear mixed-effects regression models for normalized mid-point F1 and F2 values at 30% of the vowels for the L1L2 vowel contrast (/ai/-/ei/): /ai/ Model and /ei/ Model.

Normalized F1 and F2 at 30% for /ai/ Model `Normalized vowel duration ~ Condition*TalkerGroup*FormantType + (1| Word)`				
Fixed Effects	**Estimate**	**S.E.**	**t-val.**	***p*-val.**
(Intercept)	.06	.04	1.47	.17
Condition (No Context vs. Context)	-.12	.08	-1.58	.14
TalkerGroup1 (Native English vs. Native Mandarin-High)	.05	.09	.5	.62
TalkerGroup2 (Native Mandarin-High vs. Native Mandarin-Low)	-.16	.1	-1.54	.12
Formant type (F1 vs. F2)	-.01	.07	-.14	.89
Condition: TalkerGroup1	-.01	.18	-.08	.94
Condition: TalkerGroup2	-.02	.21	-.07	.94
Condition: Formant type	-.09	.14	-.62	.54
TalkerGroup1: Formant type	-.05	.18	-.29	.77
TalkerGroup2: Formant type	-.18	.21	-.89	.37
Condition: TalkerGroup1: Formant type	.04	.37	.12	.91
Condition: TalkerGroup2: Formant type	-.2	.42	-.48	.63
Normalized F1 and F2 at 30% for /ei/ Model `Normalized vowel duration ~ Condition*TalkerGroup*FormantType + (1| Word)`				
(Intercept)	-.06	.04	-1.4	.19
Condition (No Context vs. Context)	.19	.08	2.39	.034 *
TalkerGroup1 (Native English vs. Native Mandarin-High)	-.05	.09	-.51	.61
TalkerGroup2 (Native Mandarin-High vs. Native Mandarin-Low)	.16	.1	1.6	.11
Formant type (F1 vs. F2)	.01	.07	.18	.85
Condition: TalkerGroup1	-.34	.18	-1.88	.06
Condition: TalkerGroup2	-.22	.2	-1.65	.1
Condition: Formant type	-.14	.14	-1.04	.3
TalkerGroup1: Formant type	.05	.18	.27	.79
TalkerGroup2: Formant type	.18	.2	.88	.38
Condition: TalkerGroup1: Formant type	-.29	.36	-.79	.43
Condition: TalkerGroup2: Formant type	.05	.41	.13	.89

4.3.4.2. L2-only vowel contrast (/i/-/ɪ/)

4.3.4.2.1. Normalized vowel duration

Figure 4.14 shows the mean normalized vowel durations by Phoneme (/i/, /ɪ/: "ii" and "I" in the figure), Talker Group (Native English, Native Mandarin-High, Native Mandarin-Low), and Condition (No context, Context). The figure suggests that normalized vowel durations were longer for /i/ than for /ɪ/. This difference in normalized vowel durations seems to be the largest for Native Mandarin-High talkers' productions than for the other two Talker Groups' productions; mean normalized vowel duration for /i/ - that for /ɪ/ was .044 (Native English), .063 (Native Mandarin-High), .023 (Native Mandarin-Low). The linear mixed-effects regression model, used to analyze normalized vowel durations as the dependent variable, included Phoneme, Talker Group, as well as interactions between the two factors as fixed effects (see Table 4.16 for the model syntax and summary of the results). There was a significant effect of Phoneme ($\beta = .04$, $t = 9.52$, $p < .001$). The effect of Phoneme interacted with the Native Mandarin-High vs. Native Mandarin-Low group comparison ($\beta = .04$, $t = 3.0$, $p < .01$). A post-hoc Tukey test showed that the effect of Phoneme was significant for all the Talker Groups: Native English ($\beta = -.04$, SE = .007, t ratio = -6.43, $p < .0001$), Native Mandarin-High ($\beta = -.06$, SE = .009, t ratio = -7.54, $p < .0001$), and Native Mandarin-Low ($\beta = -.02$, SE = .009, t ratio = -2.77, $p = .006$). These results suggest that talkers produced /i/ with longer normalized durations than /ɪ/. This difference was larger for Native Mandarin-High talkers' productions than for Native Mandarin-Low talkers' productions; though the difference was similar for Native English talkers' productions and for Native Mandarin-High talkers' productions.

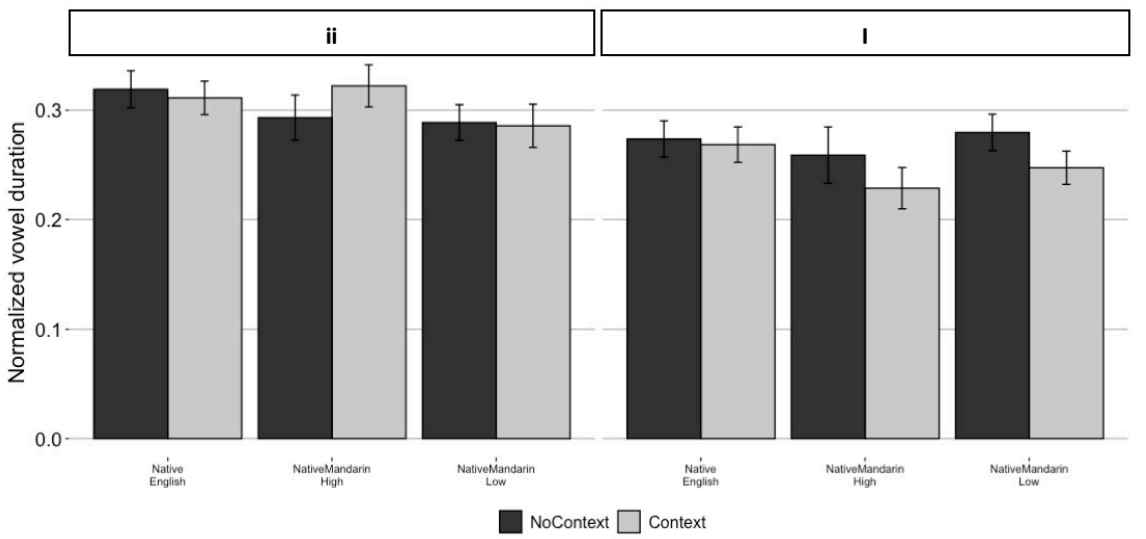

Figure 4.14. Mean normalized durations of the vowels in L2-only vowel contrast by Phoneme (/i/, /ɪ/: "ii" and "I" in the figure), Talker Group (Native English, Native Mandarin-High, Native Mandarin-Low), and Condition (No context, Context). The error bars represent the 95% confidence interval of the mean.

Figure 4.14 also suggests that talkers in different Talker Groups manipulated vowel durations differently for the two Conditions (No Context, Context). That is, Native Mandarin-High talkers increased the durations of /i/ from No Context to Context conditions; they also decreased the durations of /ɪ/ from No Context to Context conditions. Native Mandarin-Low talkers also decreased the durations of /ɪ/ from No Context to Context conditions. However, Native English talkers made a much smaller difference between the two Conditions for both /i/ and /ɪ/. The linear mixed-effects regression models, used to analyze normalized vowel durations as the dependent variable separately for /i/-targets and /ɪ/-targets, included Condition (No Context, Context), Talker Group, as well as interactions between the two as fixed effects (see Table 4.16 for the model syntax and summary of the results)

Table 4.16. Summary of the linear mixed-effects regression models for normalized durations of the vowels in the L2-only vowel contrast (/i/-/ɪ/).

Fixed Effects	**Estimate**	**S.E.**	**t-val.**	***p*-val.**
Normalized vowel durations for /i/ and /ɪ/ Model `Normalized vowel duration ~ Phoneme*TalkerGroup + (1+ Phoneme| Talker)`				
(Intercept)	.28	.005	58.03	< .001
Phoneme (/i/ vs. /ɪ/)	.04	.005	9.52	< .001 ***
TalkerGroup1 (Native English vs. Native Mandarin-High)	.02	.01	1.84	.07
TalkerGroup2 (Native Mandarin-High vs. Native Mandarin-Low)	.01	.01	.79	.42
Phoneme: TalkerGroup1	-.0002	.01	-.02	.99
Phoneme: TalkerGroup2	.04	.01	3.0	.0028 **
Normalized vowel durations for /i/ Model `Normalized vowel duration ~ Condition*TalkerGroup + (1+ Condition| Talker)`				
(Intercept)	.03	.005	59.6	< .001
Condition (No Context vs. Context)	.006	.007	.9	.37
TalkerGroup1 (Native English vs. Native Mandarin-High)	.02	.01	1.77	.08
TalkerGroup2 (Native Mandarin-High vs. Native Mandarin-Low)	.03	.01	2.08	.041 *
TalkerGroup1: Condition	-.02	.02	-1.25	.21
TalkerGroup2: Condition	.02	.02	.83	.41
Normalized vowel durations for /ɪ/ Model `Normalized vowel duration ~ Condition*TalkerGroup + (1| Talker)`				
(Intercept)	26	.006	46.94	< .001
Condition (No Context vs. Context)	-.02	.007	-3.25	.001 **
TalkerGroup1 (Native English vs. Native Mandarin-High)	.02	.01	1.61	.11
TalkerGroup2 (Native Mandarin-High vs. Native Mandarin-Low)	-.008	.02	-.53	.6
TalkerGroup1: Condition	.03	.02	1.87	.06
TalkerGroup2: Condition	.02	.02	1.25	.21

For the model with /i/-targets, there was a significant effect of the Native Mandarin-High vs. Native Mandarin-Low group comparison ($\beta = .03$, $t = 2.08$, $p < .05$). This indicates that normalized durations for /i/ were longer for Native Mandarin-High talkers' productions than for Native Mandarin-Low talkers' productions. The main effect of

Condition was not significant (β = .006, t = .9, *p* = .37). The effect of Condition did not interact with the Native English vs. Native Mandarin-High comparison (β = -.02, t = -1.24, *p* = .21) or with the Native Mandarin-High vs. Native Mandarin-Low comparison (β = .02, t = .83, *p* = .41). This indicates that the effect of Condition did not differ among different Talker Groups' productions. A post-hoc Tukey test showed that the effect of Condition was not significant for Native English (β = .005, SE = .01, t ratio = .49, *p* = .62) and Native Mandarin-Low (β = .002, SE = .01, t ratio = .16, *p* = .87) groups, but was marginally significant for the Native Mandarin-High group (β = -.03, SE = .01, t ratio = -1.99, *p* = .05).

For the model with /ɪ/-targets, there was a significant effect of Condition (No Context, Context; β = -.02, SE = .007, t ratio = -3.25, *p* < .001). This effect of Condition did not interact with Talker Groups: Condition x Native English vs. Native Mandarin-High comparison (β = .03, SE = .02, t ratio = 1.87, *p* = .06), Condition x Native Mandarin-High vs. Native Mandarin-Low comparison (β = .02, SE = .02, t ratio = 1.25, *p* = .21). However, a post-hoc Tukey test showed that the effect of Condition was significant for the productions of Native Mandarin-High (β = .03, SE = .01, t ratio = 2.1, *p* = .037), and Native Mandarin-Low talkers (β = .03, SE = .01, t ratio = 2.72, *p* = .007), but not for those of Native English talkers (β = .005, SE = .01, t ratio = .54, *p* = .59).

These results suggest that talkers generally distinguished /i/ and /ɪ/ with vowel durations; normalized durations were longer for /i/ than for /ɪ/. Furthermore, talkers in different Talker Groups manipulated the vowel durations differently. Specifically, Native Mandarin-High talkers increased the /i/-/ɪ/ difference in normalized durations from No Context to Context conditions, by increasing the durations for /i/ as well as decreasing the

durations for /ɪ/ in Context conditions. Native Mandarin-Low talkers also increased the /i/-/ɪ/ difference by decreasing the durations for /ɪ/ from No Context to Context conditions. However, Native English talkers did not manipulate the vowel durations in different types of conditions.

4.3.4.2.2. Midpoint F1 and F2

The left panel in Figure 4.15 shows the mean mid-point F1 and F2 values of the vowels in the L2-only vowel contrast by Phoneme (/i/, /ɪ/: "ii" and "I" in the figure), Talker Group (Native English, Native Mandarin-High, Native Mandarin-Low), and Condition (No context, Context). The figure suggests that Native English talkers produced the largest difference between the two vowels in F1 and F2, followed by Native Mandarin-High talkers and by Native Mandarin-Low talkers. The linear mixed-effects regression model, used to analyze formant values as the dependent variable, included Phoneme, Talker Group, Formant (F1, F2) as well as interactions among these factors as fixed effects; these factors were contrast coded as specified above (see Table 4.17 for the model syntax and summary of the results). There was a significant interaction between Phoneme (/i/, /ɪ/) and Formant (F1, F2; $\beta = 2.13$, $t = 13.25$, $p < .001$). This indicates that from /ɪ/ to /i/, F1 decreased but F2 increased. This tendency was larger for Native English talkers' productions than for Native Mandarin-High talkers' productions (Formant x Phoneme: x Native English vs. Native Mandarin-High: $\beta = 2.55$, $t = 14.03$, $p < .001$), and for Native Mandarin-High talkers' productions than for Native Mandarin-Low talkers' productions (Formant x Phoneme: x Native Mandarin-High vs. Native Mandarin-Low: $\beta = 2.92$, $t = 14.41$, $p < .001$). These interactions are illustrated in the right panel in Figure 4.15;

confirming that Native English talkers distinguished between /i/ and /ɪ/ to a larger extent than Native Mandarin-High talkers did. Native Mandarin-High talkers distinguished the vowels to a larger extent than Native Mandarin-Low talkers did.

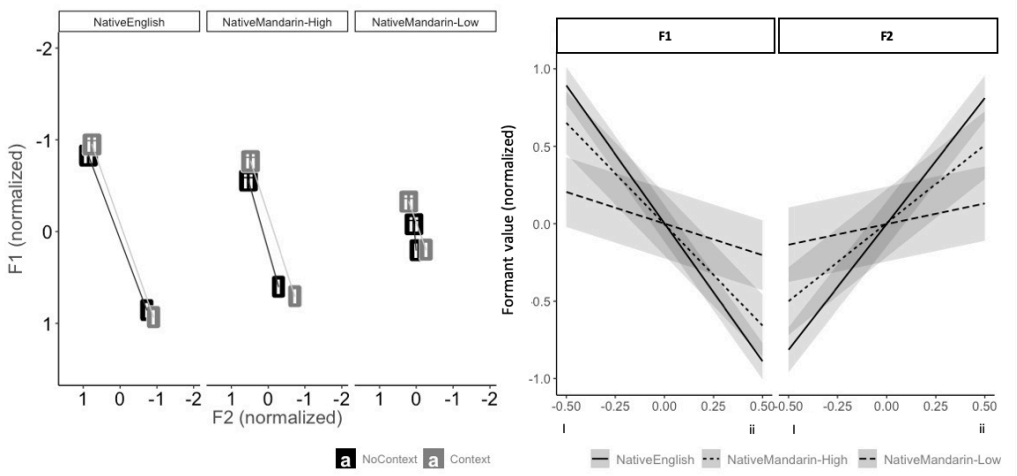

Figure 4.15. Left panel: mean mid-point F1 and F2 values of the vowels in the L2-only vowel contrast by Phoneme (/i/, /ɪ/: "ii" and "I" in the figure), Talker Group (Native English, Native Mandarin-High, Native Mandarin-Low), and Condition (No context, Context). Right panel: linear predictions for mid-point F1 and F2 values of /i/ and /ɪ/ for different Talker Groups' productions.

Table 4.17. Summary of the linear mixed-effects regression model for normalized mid-point F1 and F2 values of the vowels in the L2-only vowel contrast (/i/-/ɪ/): Model for both /i/ and /ɪ/.

Fixed Effects	**Estimate**	**S.E.**	**t-val.**	***p*-val.**
Normalized F1 and F2 for /i/ and /ɪ/ Model `Formant value ~ Phoneme*TalkerGroup*FormantType + (1+ FormantType| Word)`				
(Intercept)	-.00008	.03	0.003	.998
Phoneme (/i/ vs. /ɪ/)	-.1	.05	-1.95	.065
TalkerGroup1 (Native English vs. Native Mandarin-High)	.0009	.05	.02	.98
TalkerGroup2 (Native Mandarin-High vs. Native Mandarin-Low)	.002	.05	.03	.97
Formant type (F1 vs. F2)	.0001	.08	.001	.999
Phoneme: TalkerGroup1	.04	.09	.48	.63
Phoneme: TalkerGroup2	-.06	.1	-.58	.57
Phoneme: Formant type	2.13	.16	13.25	< .001 ***
TalkerGroup1: Formant type	-.008	.09	-.09	.93
TalkerGroup2: Formant type	.007	.1	.07	.95
Phoneme: TalkerGroup1: Formant type	2.55	.18	14.03	< .001 ***
Phoneme: TalkerGroup2: Formant type	2.92	.2	14.41	< .001 ***

In order to further examine whether talkers manipulated F1 and F2 values depending on different types of Conditions (No Context, Context), we implemented separate linear mixed-effects regression models for /i/-targets and /ɪ/-targets with formant values as the dependent variable. Fixed effects were Condition (No Context, Context), Formant (F1, F2), Talker Group, as well as interactions among these factors (see Table 4.18 for the model syntax and summary of the results).

For the model with /i/-targets, there was a significant effect of Formant (F1, F2; $\beta = 1.07$, $t = 21.78$, $p < .001$), and this effect of Formant interacted with the Native English vs. Native Mandarin-High group comparison ($\beta = 1.26$, $t = 9.82$, $p < .001$), and with the Native Mandarin-High and Native Mandarin-Low group comparison ($\beta = 1.47$, $t = 10.21$, $p < .001$). These results, as illustrated in the left panel of Figure 4.16, indicate that for /i/-targets, normalized formant values were different for F1 and F2, and this difference was larger for Native English talkers' productions than for Native Mandarin-High talkers' productions, and for Native Mandarin-High talkers' productions than for Native Mandarin-Low talkers' productions. The effect of Condition (No Context, Context) was not significant ($\beta = -.1$, $t = -1.57$, $p = .15$), but there was a marginally significant interaction between Condition and Formant ($\beta = .19$, $t = 1.92$, $p = .055$). This indicates that there was a tendency for talkers to decrease F1 (raise the tongue position) from No Context to Context conditions, but the talkers did not make a difference in terms of F2 (front/back dimension of the tongue position) between the two conditions.

Table 4.18. Summary of the linear mixed-effects regression models for normalized mid-point F1 and F2 values of the vowels in the L2-only vowel contrast (/i/-/ɪ/): /i/ Model and /ɪ/ Model.

Normalized F1 and F2 for /i/ Model `Normalized vowel duration ~ Condition*TalkerGroup*FormantType + (1| Word)`				
Fixed Effects	**Estimate**	**S.E.**	**t-val.**	***p*-val.**
(Intercept)	-.05	.03	-1.61	.14
Condition (No Context vs. Context)	-.1	.06	-1.57	.15
TalkerGroup1 (Native English vs. Native Mandarin-High)	.02	.06	.36	.72
TalkerGroup2 (Native Mandarin-High vs. Native Mandarin-Low)	-.03	.07	-.39	.7
Formant type (F1 vs. F2)	1.07	.05	21.78	< .001 ***
Condition: TalkerGroup1	-.02	.13	-.16	.87
Condition: TalkerGroup2	-.09	.14	-.66	.51
Condition: Formant type	.19	.1	1.92	.055
TalkerGroup1: Formant type	1.26	.13	9.82	< .001 ***
TalkerGroup2: Formant type	1.47	.14	10.21	< .001 ***
Condition: TalkerGroup1: Formant type	-.33	.26	-1.28	.2
Condition: TalkerGroup2: Formant type	-.4	.29	-1.39	.17
Normalized F1 and F2 for /ɪ/ Model `Normalized vowel duration ~ Condition*TalkerGroup*FormantType + (1| Word)`				
(Intercept)	.05	.03	1.49	.17
Condition (No Context vs. Context)	-.12	.07	-1.8	.1
TalkerGroup1 (Native English vs. Native Mandarin-High)	-.02	.07	-.32	.75
TalkerGroup2 (Native Mandarin-High vs. Native Mandarin-Low)	.03	.07	.42	.68
Formant type (F1 vs. F2)	-1.07	.05	-20.9	< .001 ***
Condition: TalkerGroup1	.13	.13	1.0	.32
Condition: TalkerGroup2	.03	.15	.22	.83
Condition: Formant type	-.36	.1	-3.5	< .001 ***
TalkerGroup1: Formant type	-1.29	.13	-9.61	< .001 ***
TalkerGroup2: Formant type	-1.47	.15	-9.83	< .001 ***
Condition: TalkerGroup1: Formant type	.19	.27	.72	.47
Condition: TalkerGroup2: Formant type	-.19	.3	-.63	.53

For the model with /ɪ/-targets, there was a significant effect of Formant (F1, F2; β = -1.07, t = -20.9, $p < .001$), and this effect of Formant interacted with the Native English vs. Native Mandarin-High group comparison (β = -1.29, t = -9.61, $p < .001$), and with the

Native Mandarin-High and Native Mandarin-Low group comparison (β = -1.47, t = -9.83, p < .001). These results, as illustrated in the right panel of Figure 4.16, indicate that for /ɪ/, normalized formant values were different for F1 and F2, and this difference was larger for Native English talkers' productions than for Native Mandarin-High talkers' productions, and for Native Mandarin-High talkers' productions than for Native Mandarin-Low talkers' productions. There was also a significant interaction between Condition (No Context, Context) and Formant (β = -.36, t = -3.5, p < .001). This indicates that, as illustrated in the right panel of Figure 4.16, talkers increased F1 (lowered the tongue position), but decreased F2 (retracted the tongue position) from No Context to Context conditions.

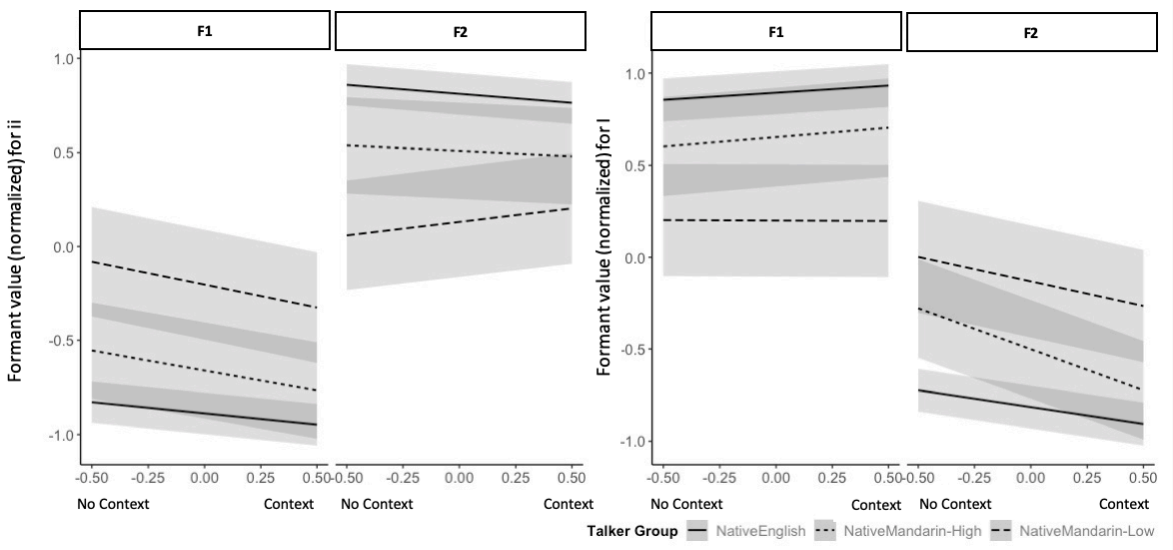

Figure 4.16. Linear predictions for mid-point F1 and F2 values of /i/ (Left panel) and /ɪ/ (Right panel) in different Conditions (No context, Context) for different Talker Groups' productions (Native English, Native Mandarin-High, Native Mandarin-Low).

These results suggest that talkers generally distinguished /i/ and /ɪ/ with formant values; talkers produced /i/ with lower F1 and higher F2 than /ɪ/. Native English talkers distinguished the two vowels to a larger extent than Native Mandarin-High talkers did; Native Mandarin-High talkers also distinguished the two vowels to a larger extent than

Native Mandarin-Low talkers did. Furthermore, talkers made No Context-Context distinctions somewhat differently for /i/ and /ɪ/. That is, for /i/, there was a tendency for talkers to manipulate F1 (i.e., decrease F1; raise the tongue position), but not F2. For /ɪ/, talkers manipulated both F1 and F2 (i.e., increased F1 and decreased F2 from No Context to Context conditions). These patterns did not differ for different talker groups' productions.

4.3.4.3. Summary of the segmental analyses for vowel targets

The results of the segmental analyses suggest that differences in talkers' native language background (English vs. Mandarin) and their target language proficiency levels (Native Mandarin-High vs. Native Mandarin-Low talkers) influenced productions of the L2-only vowel contrast (/i/-/ɪ/), but not those of the L1L2 vowel contrast (/ai/-/ei/). Specifically, for the L2-only vowel contrast (/i/-/ɪ/), talkers in different Talker Groups generally distinguished the two vowels in terms of vowel durations and formant values, but the size of this distinction was different depending on the Talker Group: duration difference for /i/-/ɪ/: Native English = Native Mandarin-High > Native Mandarin-Low, spectral difference for /i/-/ɪ/: Native English > Native Mandarin-High > Native Mandarin-Low. Talkers further enhanced the contrast to different degrees in Context conditions (compared to No Context conditions) as well. Native Mandarin (High and Low) talkers further increased the /i/-/ɪ/ duration difference to a larger extent than Native English talkers did, by increasing the durations of /i/ (High talkers) and by decreasing the durations of /ɪ/ (High and Low talkers). Talkers in all groups increased the contrast in formant values to a similar extent; they decreased F1 for /i/ (raised the tongue position) and increased F1 and

decreased F2 for /ɪ/ (raised and retracted the tongue position) in Context conditions compared to No Context conditions. For the L1L2 vowel contrast (/ai/-/ei/), differences in Talker Groups did not influence their productions. That is, talkers in different groups did not distinguish /ai/ from /ei/ with vowel durations or with initial F1 and F2 values. In Context conditions (as compared to No Context conditions), talkers tended to increase durations for both /ai/ and /ei/ (though the difference was not statistically significant); they also increased F1 and F2 for /ei/ but did not manipulate formant values for /ai/. Together, these results suggest that talkers' contextually-relevant segmental enhancements of English vowel contrasts differed depending on talkers' target language experience, as well as the type of acoustic cues used to enhance the contrasts. This will be further discussed in the section below.

4.4. Discussion & conclusion

4.4.1. Summary of the main findings

In the present study, we examined acoustic characteristics of contextually-relevant speech enhancements made by native English talkers and non-native English talkers of higher- and lower-proficiency. Specifically, in a simulated communication task, we examined how talkers enhance acoustic-phonetic characteristics of their speech when a target word and its minimal-pair neighbor were present in the same context (Context conditions) compared to when the minimal-pair neighbor was not present in the context (No Context conditions). At the global level (i.e., modifications made to the characteristics of the entire phrases and target words), native English and non-native talkers of higher- and lower-proficiency made similar speech enhancements. For example, talkers produced entire

phrases (e.g., *"Click on the pill now"*) with longer durations in the Context condition than in the No Context condition. Talkers across different groups also made similar types of acoustic modifications at the target word-level (e.g., longer durations and higher mean intensity in Context conditions than in No Context conditions). Thus, differences in talkers' target language experience (i.e., native vs. non-native talkers; higher- vs. lower-proficiency non-native talkers) did not affect modifications made at the global level.

However, at the segmental level, the effects of talkers' target language experience on speech enhancements manifested differently depending on the type of the English contrast examined. Specifically, segmental enhancement patterns did not differ across the speech of native and non-native English talkers, for the English contrasts that also exist in non-native talkers' native language, Mandarin (i.e., L1L2 consonant contrast: /p/-/b/ in word-initial position; L1L2 vowel contrast: /ai/-/ei/). However, talkers' ability to enhance an English contrast that does not exist in non-native talkers' native Mandarin (i.e., the L2-only consonant contrast: /p/-/b/ in word-final position) differed among different talker groups; native English and higher-proficiency talkers were better able to enhance the contrast than lower-proficiency talkers. Further, for the L2-only vowel contrast (/i/-/ɪ/), non-native (both higher- and lower-proficiency) talkers enhanced the vowel duration in the contrast to a larger extent than native English talkers. Further, non-native talkers enhanced the vowel formants of the contrast to a similar extent as native English talkers. A close examination of these results (as discussed in detail below) suggests that talkers' contextually-relevant segmental enhancements are influenced by their target language experience as well as by the type of acoustic cues (e.g., manipulations in duration, spectral values) used to enhance these particular contrasts.

4.4.2. Talkers' production patterns at the global level

In the word-reading paradigm used in the current study, talkers' production patterns at the global level were influenced by several factors. Specifically, talkers' target language experience influenced duration of the speech; lower-proficiency non-native talkers spoke the slowest (i.e., with the longest phrase duration, possibly with more frequent and longer pauses as well as longer segment durations), followed by higher-proficiency talkers and by native English talkers. This is in line with the previous results showing that lower-proficiency talkers' speech is slower than that of higher-proficiency talkers and native talkers (e.g., Kormos & Dénes, 2004), suggesting that talkers' target language experience impacts global temporal characteristics when producing simple phrases, such as *"Click on the ____ now"*.

Furthermore, talkers across different proficiency levels modified global characteristics of their speech in a similar way, depending on whether a phonetically similar word was present in the same context as the target word (Context condition) or not (No Context condition). These acoustic modifications included longer phrase or target word durations and higher mean intensity of the target words in Context conditions than in No Context conditions (manifested somewhat differently depending on the type of the contrast in the target words). These patterns of global enhancements are similar to those demonstrated in clear speech, where native and non-native talkers produce speech with increased duration and amplitude when explicitly instructed to speak clearly (e.g., Granlund et al., 2012; Picheny et al., 1986; Chapter 2 of this dissertation). These results suggest that the production task that encourages talkers to enhance segmental features can also facilitate

enhancements of some features at the global level (i.e., duration, intensity, but not mean F0), impacting the overall salience of the signal.

The current results further suggest the influence of the articulatory nature of particular segments on the global characteristics of the native and non-native talkers' productions. Specifically, target words with the L1L2 vowel contrast (/ai/-/ei/) may have had lower F0 than those with the L2-only vowel contrast (/i/-/ɪ/) because the L1L2 vowel contrast involved lower vowels compared to the L2-only vowel contrast. That is, given previous studies suggesting that high vowels, such as /i/ and /u/, tend to have higher F0 values than low vowels, such as /a/ (Van Hoof & Verhoeven, 2011; Whalen & Levitt, 1995), it is possible that such intrinsic relationship between vowel height and F0 were present in the productions of native and non-native talkers in the current study.

The vowel results also suggest some relationship between vowel height and intensity. Specifically, from No Context to Context conditions, talkers increased mean intensity for target words with the L1L2 vowel contrast (/ai/-/ei/), but decreased mean intensity for target words with the L2-only vowel contrast (/i/-/ɪ/). These differences at the target word level also impacted characteristics at the phrase level; phrases with the L1L2 vowel contrast had higher mean intensity as well as lower mean F0 than those with the L2-only vowel contrast. This is line with previous results suggesting that higher vowels tend to have lower intensity than lower vowels because the enlarged pharyngeal cavity during production of higher vowels results in a stronger dampening of the excitation signal (Lehiste & Peterson, 1959; Möbius, 2003). It is possible that, in Context conditions, talkers' efforts to enhance higher vowels in the L2-only contrast (/i/-/ɪ/) resulted in decreased intensity, and at the same time, their efforts to enhance lower vowels in the L1L2

vowel contrast (/ai/-/ei/) resulted in increased intensity. These patterns may have impacted the intensity of the target words as well as the whole phrases. Thus, the production patterns in the current results could partly be explained by unique characteristics of the sound contrasts, specifically, physiological aspects of vowel productions.

4.4.3. The effect of target language experience on contextually-relevant segmental enhancements

The present study demonstrates somewhat different effects of talkers' target language experience on segmental enhancements of English sound contrasts, depending on whether or not the sound contrast exists in non-native talkers' native language (i.e., Mandarin). That is, talkers' target language experience (native vs. non-native; higher- vs. lower-proficiency) did not impact how the talkers enhanced the English contrasts that also exist in non-native talkers' native language (i.e., L1L2 contrasts) but it did for the English contrasts that do not exist in non-native talkers' native language (i.e., L2-only contrasts). In the next several sections, we discuss native and non-native talkers' segmental enhancement patterns in relation with the type of acoustic cues involved in the enhancements.

4.4.3.1. Non-native sound contrasts that exist in talkers' native language: L1L2 contrasts

For the L1L2 consonant contrast (/p/-/b/ in word-initial position), non-native talkers of higher- and lower-proficiency enhanced the contrast by increasing the VOTs of /p/ as well as native English talkers did. These production patterns are in line with previous studies demonstrating that native English talkers exaggerate voiceless onset plosive VOTs when a voiced competitor is contextually present (Baese-Berk & Goldrick, 2009; Buz et

al., 2014, 2016; Kirov & Wilson, 2012). The current results further extend these findings to non-native talkers, demonstrating that non-native talkers of higher- and lower-proficiency use the same strategy to enhance the contextually-relevant voicing contrast in word-onset position, and they do so to a similar extent as native talkers do. Similar patterns were found for the L1L2 vowel contrast (/ai/-/ei/); native and non-native talkers' strategies of enhancing the contrast did not differ. Though there was a tendency for talkers to produce both /ai/ and /ei/ with longer durations in Context conditions compared to No Context conditions, this increase in normalized vowel durations was not statistically significant for either of /ai/ or /ei/. However, talkers manipulated initial F1 values of /ei/ more than those of /ai/ in order to clarify the distinction between the two vowels in Context conditions. That is, talkers increased F1 (lowered the tongue position) of /ei/ in Context conditions. It is possible that talkers manipulated initial formant values of /ei/ to a larger extent than those of /ai/, because /e/ and /i/ are closer in vowel space than between /a/ and /i/. By lowering /e/, they may have tried to exaggerate the movement from /e/ to /i/, though lowering /e/ made it closer to the space of /a/ in /ai/. These results demonstrated that talkers' target language experience (native vs. non-native; higher- vs. lower-proficiency) did not impact the size of modifications implemented to enhance the L1L2 consonant and vowel contrasts. These results suggest that manipulating acoustic properties to enhance a familiar non-native contrast is possible even for talkers with limited language proficiency (e.g., lower-proficiency talkers), especially when the production task is simple (i.e., the target word to communicate is embedded in a simple phrase, such *as "Click on the ___ now"*).

4.4.3.2. Non-native sound contrasts that do not exist in talkers' native language: L2-only contrasts

Unlike the results of L1L2 consonant contrast (/p/-/b/ in word-initial position), the ability to enhance the L2-only consonant contrast (/p/-/b/ in word-final position) differed for higher- vs. lower-proficiency non-native talkers. Specifically, lower-proficiency talkers showed a tendency to decrease voicing proportions of /b/ in Context conditions as compared to No Context conditions; whereas native English and higher-proficiency talkers increased the voicing proportions of /b/ from No Context to Context conditions to enhance the coda /p/-/b/ contrast. It is possible that lower-proficiency talkers rather strengthened their general production pattern of devoicing the voiced stop consonant in coda position (e.g., Broselow et al, 1998; Flege et al., 1992). These results demonstrated that among the non-native talkers, higher-proficiency talkers were better able to enhance the non-native consonant contrast that does not exist in their native language compared to lower-proficiency talkers, possibly suggesting that as talkers' target language proficiency develops they are better able to use the acoustic-enhancement strategy that they are not necessarily familiar with from their native language experience.

The current results provide support for the exemplar-based representations of sounds (e.g., Goldinger, 1996; Johnson, 1997; Pierrehumbert, 2003) rather than rule-based representations (e.g., Chomsky & Halle, 1968; Prince & Smolensky, 1993). Specifically, by demonstrating that native talkers and higher-proficiency talkers differentiate strategies to enhance a contrast (/p/-/b/ contrast) in different positions (word-initial vs. word-final), we suggest that these talkers have acquired the context-dependent knowledge of this contrast, and the articulatory control to further enhance the position-specific details of the

particular contrast. In other words, the results demonstrate that these talkers do not treat the word-initial /p/-/b/ contrast and word-final /p/-/b/ contrast the same way. These results also have implications for the relatively understudied area of inquiry regarding second language acquisition of allophonic variations. Particularly, previous work has shown that second language learners' use of allophonic variations is different from that of native talkers, but experienced learners' use of acoustic cues to implement allophonic variations become similar to that of native talkers as compared to that of less experienced learners (e.g., Barlow, 2014; Shea, 2014; Shea & Curtin, 2011; Tajima, Kitahara, & Yonayama, 2015; Vaughn, Baese-Berk & Idemaru, 2019; Vokic, 2010). The current results further extend the effect of learners' target language experience to the enhancements of allophonic variations. That is, by demonstrating that higher-proficiency talkers are able to enhance the English /p/-/b/ contrast in both word-initial and word-final positions (though lower-proficiency talkers were only able to enhance the contrast in word-initial position), the current results highlight the role of talkers' target language experience in their ability to manipulate important acoustic cues to enhance allophonic variations of the same phonemic category.

It should also be pointed out that 'proficient' talkers' (i.e., native English and higher-proficiency non-native talkers) use of acoustic cues to enhance the coda voicing contrast was different from those used to produce the contrast in general. That is, these talkers differentiated the coda /p/-/b/ contrast using both preceding vowel durations (i.e., longer vowels before /b/ than before /p/) and target consonant voicing proportions (i.e., larger voicing proportions for /b/ than for /p/); while they used preceding vowel durations to a much lesser extent than target consonant voicing proportions to further enhance the contrast. Previous results have also reported a lack of enhancement in vowel durations

when exaggerating coda voicing plosive contrasts. For example, lexically-mediated enhancements of an English coda voicing contrast were not found in preceding vowel durations. That is, native English talkers did not change durations of vowels preceding voiceless stops when they read target words that had minimal-pair neighbors (e.g., coat - code) as compared to words that did not have minimal-pair neighbors (e.g., vote - *vode); they even decreased durations of vowels preceding voiced stops for words with minimal-pair neighbors compared to those without minimal-pair neighbors, reducing the voicing distinction (Goldrick et al., 2013). Further, the size of duration difference between vowels before voiced and voiceless coda consonants was not larger when the competitors were of phonological focus (e.g., bed - bet), compared to when the competitors were of lexical/semantic focus (e.g., bed - chair: Choi et al., 2015), or compared to when the competitors were of voicing-irrelevant focus (e.g., bed - bad: de Jong, 2004).

Though it is still an open question why vowel durations preceding the target consonants are not utilized to enhance the coda voicing contrast, the current results showed a tendency that talkers relied on the preceding vowel duration to different degrees for enhancing different segments involved in the coda voicing contrast. Specifically, higher-proficiency talkers and native talkers used the strategy of decreasing the vowel durations before /p/ (though the decrease in preceding vowel durations was not statistically significant), while they used the strategy of increasing the voicing proportions for /b/. This is in line with native English talkers' hyperarticulation patterns of a coda fricative voicing contrast (/s/-/z/: Seyfarth et al., 2016), suggesting that the way coda voicing contrast is implemented could differ for voiced vs. voiceless end of the contrast across different manners of articulation. The current results contribute to these lines of studies by

suggesting that higher-proficiency non-native talkers could implement strategies for enhancing word-final voiced vs. voiceless sounds as native English talkers do, though these talkers may rely on one type of strategy more than the other in order to enhance the contrast.

The results with the L2-only vowel contrast (/i/-/ɪ/) also showed that the effects of talkers' target language experience on speech enhancement patterns differed depending on the type of acoustic cues used to enhance the contrast: the spectral or temporal dimensions. For example, from No Context to Context conditions, native and non-native (both higher- and lower-proficiency) talkers increased F1 (lowered the tongue position) and decreased F2 (retracted tongue position) for /ɪ/, and they slightly decreased F1 (raised the tongue position; marginally significant result) for /i/. This suggests that native and non-native talkers used the spectral cues to a similar extent to enhance the /i/-/ɪ/ contrast. However, in terms of the normalized vowel durations (i.e., raw vowel durations/whole word durations), higher-proficiency talkers enhanced the contrast the most by both increasing the durations for /i/ and decreasing the durations for /ɪ/ from No Context to Context conditions. Lower-proficiency talkers also decreased the durations for /ɪ/; though native English talkers did not manipulate the normalized vowel durations in different conditions. These results demonstrate that, though the extent of enhancing the contrast was similar between non-native and native English talkers' productions in terms of spectral features, non-native talkers enhanced the contrast even better than native talkers in terms of temporal features. This is partially in line with previous results demonstrating a greater use of temporal cues over spectral cues for non-native talkers' productions of English tense-lax vowel distinctions. For example, exaggeration of duration differences of English tense-lax vowel

contrasts has been reported for productions of non-native talkers whose native language makes use of duration differences to distinguish sound categories (e.g., Bohn & Flege, 1992; Munro, 1993; Tsukada, 2009) as well as for productions of talkers whose native language does not involve durational contrasts, including Mandarin (e.g., Chen, 2006; Flege et al., 1997). Similar results of non-native speakers' use of temporal cues have also been found in perception of non-native English tense-lax vowel distinctions, where non-native listeners use temporal cues to a greater extent than native listeners (e.g., Bohn, 1995; Minnick-Fox & Maeda, 1999; Wang & Munro, 1999). The current results extend these lines of findings to non-native talkers' enhancements of the English tense-lax vowel contrast by demonstrating that non-native (native Mandarin) talkers modified normalized vowel durations of /i/ and /ɪ/ to a larger extent than native English talkers did.

These results could potentially be explained based on non-native and native talkers' perceptual tendencies. For example, it has been suggested that in order to perceptually distinguish a non-native contrast, listeners are better able to use a contrastive feature that they use in their native language (e.g., duration feature) than the feature that they do not (e.g., spectral feature; Feature Hypothesis: McAllister, Flege, & Piske, 2002). Though Mandarin vowels are not primarily distinguished by temporal cues, there are some duration differences associated with Mandarin contrastive tones (i.e., tone 2 tends to be shorter than tone 3; Blicher, Diehl, & Cohen, 1990), possibly contributing to native Mandarin talkers' use of temporal cues in productions. Alternatively, temporal cues may be inherently more salient than spectral cues regardless of non-native listeners' native language background (Desensitization Hypothsis: Bohn, 1995), and this may possibly make it easier to manipulate temporal cues than spectral cues in productions of non-native vowels as

well. Furthermore, given that native English listeners have been found to rely primarily on spectral cues and little on temporal cues to perpetually distinguish the English /i/-/ɪ/ contrast (Hillenbrand, Clark, & House, 2000), it is possible that native English talkers in the current study mainly used spectral cues to enhance the contrast, resulting in the smaller extent of native talkers' duration enhancements compared to those of non-native talkers.

Though cue weighting tendencies in perceptual discrimination may be one source that impacts talkers' use of acoustic cues to enhance non-native contrasts, this may not be an adequate explanation. That is, given a previous result demonstrating that non-native speakers' cue weighting strategies in perception do not correlate with their production patterns of non-native sounds (Schertz, Cho, Lotto, & Warner, 2015), it is possible that non-native speakers' strategies to discriminate non-native sounds in perception and production do not necessarily match. Furthermore, current results have also demonstrated mixed results regarding the non-native talkers' use of temporal cues; though they enhanced the word-initial /p/-/b/ contrast by increasing VOTs for /p/s, they did not utilize preceding vowel durations to enhance the word-final /p/-/b/ contrast. Thus, it is possible that use of a particular acoustic dimension (e.g., duration) differs depending on how it is implemented (e.g., differentiating vowel durations to enhance the vowels themselves vs. to enhance the contrast of the following consonants, or differentiating VOTs of stop consonants) not only for non-native talkers' but also for native talkers' productions. Therefore, a future investigation may examine to what extent an individual's perceptual cue weighting applies to their productive use of cues in acoustic enhancements, as well as how consistent an individual's use of a particular acoustic dimension is across enhancements of different segments.

4.4.3.3. Talkers' ability to produce a sound contrast vs. further enhance the contrast

Taken together, the current results suggest that the effect of talkers' target language experience (native vs. non-native; higher- vs. lower-proficiency) on contextually-relevant segmental enhancements vary greatly depending not only on the type of non-native contrast (whether or not the contrast exists in talkers' native language) but also on the type of acoustic feature used to enhance the contrast (e.g., duration or spectral cues). Here, it is important to point out that the effect of talkers' target language experience also differed for talkers' ability to produce the contrast vs. talkers' ability to manipulate a particular cue to further enhance the contrast. For example, in order to distinguish the word-final /p/-/b/ contrast, native English talkers made the largest difference (between /p/ and /b/) in normalized durations of the preceding vowels as well as in target consonant voicing proportions, followed by higher-proficiency non-native talkers and by lower-proficiency non-native talkers. However, the size of enhancements in the preceding vowels durations did not differ for native English, higher-proficiency or lower-proficiency talkers' productions; the size of enhancements in the consonant voicing proportions also did not differ for native English and higher-proficiency talkers' productions. Similarly, in order to distinguish /i/ vs. /ɪ/, native English talkers made the largest difference in midpoint F1 and F2, followed by higher-proficiency talkers and by lower-proficiency talkers; though the size of enhancements of formant values did not differ for these talkers.

Though it may seem puzzling that native English talkers' size of segmental enhancements was often times similar to that of higher-proficiency talkers, it is possible that native talkers have already reached close-to-maximum degrees of clarity when

producing items in No Context conditions, leaving little room to enhance from No Context to Context conditions (e.g., in terms of F1 and F2 values of /i/ in the /i/-/ɪ/ contrast). These results also suggest that the ability to distinguish a certain contrast may be partially independent from the ability to further enhance the contrast. This can be observed in some of the higher-proficiency talkers' productions. For example, higher-proficiency talkers made a much smaller difference between voicing proportions of word-final /p/ and /b/ compared to native English talkers, but higher-proficiency talkers enhanced the contrast to a similar extent as native talkers by increasing voicing proportions for /b/. Further, higher-proficiency talkers made a much smaller difference between /i/ and /ɪ/ in F1 and F2 compared to native English talkers, though higher-proficiency talkers enhanced the contrast to a similar extent as native talkers did by lowering and retracting the tongue position for /ɪ/. Thus, it is possible that there is a difference in native vs. non-native talkers' ability to distinguish certain non-native contrasts in production (e.g., non-native talkers may be less able than native talkers to produce a non-native contrast that does not exist in their native language); though non-native talkers' ability to enhance the contrast may become comparable to those of native talkers as non-native talkers' proficiency develops. The current results further suggest that knowing how to enhance a specific sound contrast, in addition to knowing how to produce the contrast in general, may be a part of what characterizes language proficiency.

4.4.4. Nature of contextually-relevant speech enhancements

The current results demonstrated that some of the contextually-relevant enhancements observed here were quite targeted, highlighting the specificity of the acoustic

modifications to enhance contrastive features of non-native sounds. In other words, while contrast-enhancing hyperarticualtion can be associated with greater duration and expansion of vowel space for the entire speech (e.g., in clear speech: Bradlow et al., 2003; Picheny et al., 1986; Smiljanić & Bradlow, 2008a), some of the native and non-native talkers' segmental enhancements in the current study are at least partially independent from the global enhancements of the speech signal (e.g., producing the phrases with longer durations in Context conditions than in No Context conditions). For example, in order to enhance the word-initial /p/-/b/ contrast (L1L2 consonant contrast), native and non-native talkers of higher- and lower-proficiency increased the VOT proportions of /p/ (out of the whole word durations). Further, in order to enhance the /i/-/ɪ/ contrast (L2-only vowel contrast), higher-proficiency talkers increased vowel proportions of /i/ (out of the whole word durations). These results suggest that contextually-driven lengthening of VOTs and vowel durations was not attributable to overall word lengthening.

Furthermore, current results also showed that some of the contextually-relevant enhancements were realized as shortening of durations as well as centralizing of vowels. For example, in order to enhance the word-final /p/-/b/ contrast (L2-only consonant contrast), native English and higher-proficiency talkers tended to shorten the normalized durations of vowels preceding /b/s. For vowel targets, in order to enhance the /i/-/ɪ/ contrast (L2-only vowel contrast), higher- and lower-proficiency talkers shortened the normalized vowel durations for /ɪ/; also, both native and non-native talkers centralized /ɪ/ (lowered and retracted the tongue position). Therefore, these results support the claim that types of contrastive hyperarticulation are not necessarily limited to elongation of segments or peripheralization of vowels, but can also involve shortening of durations or

centralization of vowels in order to enhance specific contrasts (e.g., Leung et al., 2016; Seyfarth et al., 2016; Wedel, Nelson, & Sharp, 2018). Furthermore, the current results provide evidence that such targeted modifications can be found in productions of higher- and lower-proficiency non-native talkers when the potential ambiguity for communication is signaled in the context, which is a much more implicit way of inducing speech enhancements compared to asking talkers to speak clearly (e.g., instructing participants to speak as if talking to a hearing-impaired listener: Granlund et al., 2012; Rogers et al., 2010; Smiljanić & Bradlow, 2011).

Because the context-specific speech enhancements in the current study were examined using a paradigm that signals potential communicative difficulty in the context, one might wonder to what extent the enhancements were listener-driven. That is, native and non-native talkers' speech modifications implemented to enhance certain contrasts may have been driven not only by talkers' intention to be better understood by listeners (listener-oriented) but also by talkers' internal processing of the target lexical items (talker-oriented). Previous studies have suggested several theoretical accounts for contrastive hyperarticulation. One explanation is that contrastive hyperarticulation is based on talkers' modeling of listeners' communicative needs (perceptual monitoring, or communication-based accounts: Baese-Berk & Goldrick, 2009; Buz et al., 2016). That is, talkers modify phonetic characteristics of their productions based on their understanding of what their listeners understand or know in the communication. However, another explanation is that contrastive hyperarticulation is facilitated by lexical competition during production planning (production-internal account; see Baese-Berk & Goldrick, 2009 for detailed discussion). That is, the presence of phonologically similar words in the same context as

the target words increases the difficulty of phonological encoding during planning, and this causes higher activation of the target words, resulting in hyperarticulation. This theoretical account suggests that contrastive hyperarticulation is talker-driven, originating from lexical and phonological planning processes of the talker.

The current study was not designed to differentiate these types of explanations for contrastive hyperarticulation, and we find that the current results could be compatible with both of these explanations. Particularly, it is possible that talkers hyperarticulated the contrasts because target words were presented with their minimal-pair neighbors in the same context in Context conditions, increasing lexical competition between those words. However, because some of the enhancements made by native and non-native talkers were quite targeted to enhance specific contrasts (e.g., lengthening of segment durations that is independent of overall word durations, shortening of segment durations, centralizing a vowel), it is also plausible that speech enhancements observed in the current study were at least to some extent driven by talkers' intention of increasing perceptual distance of the contrasts for the listener. Thus, based on these results, we suggest that talker-oriented and listener-oriented explanations for contrastive hyperarticulation may not be exclusive of one another, and they could work in concert to characterize talkers' production patterns. In fact, some previous results suggest that production-internal processing (e.g., lexical neighborhood-density effects) and listener-oriented processing (e.g., effects of clear speech instructions) can both impact talkers' hyperarticulation, and these effects are independent of one another (e.g., Scarborough, 2010; Scarborough & Zellou, 2013). Thus, native and non-native talkers' contextually-relevant speech enhancements observed in the current study could be explained by combinations of talker-driven and listener-driven processes.

However, some aspects of these explanations may not directly apply to non-native talkers' contrastive speech enhancements. For example, previous studies have suggested that native talkers' contrastive hyperarituclation for words with minimal pairs (e.g., *cod vs. god*; as compared to those without minimal pairs: *cop vs. *gop*) can occur without the overt presence of minimal-pair neighbor in the same context as the target word (e.g., Baese-Berk & Goldrick, 2009; Buz et al., 2016; Wedel et al., 2018). However, such hyperarticulation driven by production-internal lexical processing may not necessarily occur for non-native talkers' productions when the talkers do not realize that some non-native words have minimal pairs and some do not, or that some non-native words have higher-neighborhood density than others. Furthermore, whether non-native talkers are able to implement the intended enhancements via their articulatory control (e.g., increasing voicing proportions of word-final stop or fricative consonants: Seyfarth et al., 2016, current results) would be a separate question from how their hyperarticulation is induced (by lexical processing internal to talkers' production system and/or by talkers' modeling of listeners' communicative needs). Thus, it is an open question whether mechanisms underlying contextually-relevant contrastive enhancements are similar for talkers of different linguistic backgrounds.

4.4.5. Conclusions

This study examined acoustic characteristics of contextually-relevant speech enhancements produced by native English talkers and non-native English talkers of higher- and lower-proficiency. When the potential communication difficulty was signaled in the communication context, talkers made acoustic enhancements at the global and segmental

levels. Characteristics of these enhancements were also affected by talkers' target language experience (native vs. non-native; higher- vs. lower-proficiency) as well as by the type of the enhancements. Particularly, though we did not observe the effect of talkers' target language experience on the size of acoustic modifications for the non-native contrasts that both native and non-native talkers are familiar with (i.e., non-native contrasts that exist in non-native talkers' native language: L1L2 contrast), we found the effect of target language experience for the non-native contrasts that do not exist in non-native talkers' native language (L2-only contrast). Further, the effect of talkers' target language experience was manifested differently depending on the type of acoustic features examined (e.g., manipulation of temporal feature or spectral feature). The current findings add to the growing body of work showing that talkers are able to accommodate phonetic characteristics of their productions based on the potential communication difficulty signaled in the context, and that these findings can be extended to the productions of higher- and lower-proficiency non-native talkers. Furthermore, non-native talkers' ability to enhance a non-native contrast improves as their target language proficiency level develops. Knowing how to enhance a specific sound contrast, in addition to knowing how to produce the contrast in general, may be a part of what characterizes language proficiency, though such effect of proficiency level could differ depending on the type of acoustic manipulations required to enhance the contrast.

CHAPTER V: PERCEPTION OF CONTEXTUALLY-RELEVANT SPEECH ENHANCEMENTS

5.1. Introduction

Talkers are able to make goal-oriented speech adaptations in order to improve the speech intelligibility for listeners in various contexts, including when listeners' needs for enhanced speech intelligibility are signaled implicitly in the communication context (e.g., Baese-Berk & Goldrick, 2009; Buz et al., 2016; Seyfarth et al., 2016; Chapter 4 of this dissertation). Specifically, talkers make targeted speech enhancements of a particular word (e.g., *peer*) depending on whether or not there is a similar-sounding word in the context (e.g., *beer*) that potentially introduces communication difficulty. However, it is not clear whether such contextually-relevant speech enhancements result in perceptual benefits. In this study, we examine perceptual consequences of speech enhancements that were produced by native English talkers and non-native English talkers of different proficiency levels in a context where listeners' communicative needs for distinguishing particular sound contrasts were signaled implicitly. We investigate listeners' perception of these enhancements in terms of target word identification as well as subjective evaluations of the speech (e.g., perceived degree of comprehensibility and talker effort).

Previous studies have demonstrated robust perceptual benefits resulting from acoustic-phonetic modifications made when talkers are asked to read materials with explicit instructions to speak clearly for a hear-impaired listener or a non-native listener (i.e., clear speech enhancements: Picheny et al., 1985, 1986; Schum, 1996; Smiljanić & Bradlow, 2011; Chapter 2 of this dissertation). For example, intelligibility gains resulting from native English talkers' clear speech enhancements have been found for listeners of various

characteristics, including hearing-impaired listeners and non-native listeners (e.g., Bradlow & Bent, 2002; Bradlow & Alexander, 2007; Ferguson, 2004; Krause & Braida, 2002; Liu et al., 2004; Picheny et al., 1985; Schum, 1996; Uchanski et al., 1996). Clear speech enhancements produced by non-native talkers of the language could improve native listeners' perception as well. Specifically, for native English listeners, acoustic-phonetic modifications in clear speech made by higher-proficiency non-native talkers result in a larger intelligibility improvement than those made by lower-proficiency talkers (enhancements of English vowels: Rogers et al., 2010; enhancements of English sentences: Chapter 3 of this dissertation). The size of intelligibility improvement resulting from higher-proficiency non-native talkers' clear speech is comparable to those resulting from native talkers' clear speech (Rogers et al., 2010; Smiljanić & Bradlow, 2011; Chapter 3 of this dissertation). Furthermore, native and non-native talkers' clear speech enhancements can affect listeners' subjective evaluations of the speech; clear speech increases perceived degree of talker effort (i.e., whether listeners perceive clear speech to be produced with increased effort compared to plain speech), but it does not increase perceived degree of comprehensibility (i.e., whether listeners perceive clear speech to be easier to understand than plain speech; Chapter 3 of this dissertation).

Though perceptual consequences of clear speech enhancements have been widely examined, it is less clear how listeners' perception is influenced by speech enhancements produced in contexts that are more similar to naturalistic talker-listener interactions. Particularly, previous studies have demonstrated that talkers produce speech enhancements that are relevant to the specific communication contexts, including when the listener misunderstands a particular part of an utterance (e.g., the talker says *"pit"* but the listener

guesses *"bit"*) during a communication task (e.g., Maniwa et al., 2009; Ohala, 1994; Oviatt et al., 1998; Schertz, 2013; Stent et al., 2008), as well as when the listener's need for enhanced speech intelligibility (e.g., the need to perceptually differentiate similar words such as *pit* and *bit*) is signaled in the communication context (e.g., Baese-Berk & Goldrick, 2009; Buz et al., 2014, 2016; Hwang et al. 2015; Kirov & Wilson, 2012; Seyfarth et al., 2016; Chapter 4 of this dissertation). However, it is not clear whether such targeted speech modifications, aimed at enhancing particular aspects of the speech (e.g., distinction between *pit* and *bit)*, result in perceptual benefits for listeners; particularly, whether non-native talkers' contextually-relevant speech enhancements result in perceptual benefits for native listeners.

There is some evidence suggesting that native talkers' speech enhancements produced without explicit instructions to speak clearly can result in perceptual benefits. For example, native English listeners made lexical decisions faster when responding to native English talkers' speech that was produced with a real listener present in the room, as compared to when responding to the speech produced for an imagined hard-of-hearing listener (simulated clear speech; Scarborough & Zellou, 2013). Native English listeners also made word identification responses faster when listening to spontaneous native English speech produced in a situation with a communication barrier (i.e., vocoded speech signal), as compared to the spontaneous speech produced in a situation without a barrier (Hazan et al., 2012). The speech produced with a communication barrier was also perceived to be clearer than the speech produced without a barrier (Hazan et al., 2012). While these results demonstrate that native listeners benefit from native talkers' speech enhancements that are elicited without explicit instructions to speak clearly, it is unknown

whether such perceptual benefits would extend to those produced by non-native talkers of different proficiency levels. Specifically, it is not clear whether non-native talkers' attempt to enhance acoustic characteristics of contextually-relevant non-native sound contrasts results in an improvement in listeners' understanding of the speech, as well as in subjective evaluation of their speech, such as perceived degree of comprehensibility. In order to better understand the benefits of contextually-relevant speech enhancements produced by talkers of different linguistic backgrounds, it is critical to examine how these enhancements influence listeners' perception.

In the present study, we examine perceptual consequences of contextually-relevant speech enhancements produced by native English talkers and non-native English talkers of different proficiency levels. Using the speech samples analyzed in Chapter 4, we examine whether native English listeners correctly identify the target word communicated by native and non-native talkers (i.e., in the phrase, *"Click on the ___ now"*), and whether listeners' identification accuracy improves as a result of contextually-relevant enhancements made by these talkers. Furthermore, following Chapter 3 (i.e., perception of clear speech enhancements), we examine whether native and non-native talkers' contextually-relevant enhancements improve two types of listeners' subjective evaluation: perceived degree of comprehensibility (how easy it is to understand the speech) and talker effort (how hard the talker is trying to speak clearly). We examine these questions in a series of perception experiments: in a two-alternative forced choice (2AFC) identification task, perceived comprehensibility rating task, and perceived talker effort rating task.

5.2. Methods

5.2.1. Participants

Participants were 420 native English listeners (179 females, 239 males, 1 other, 1 declined to provide a gender; age range = 18 - 72 years, mean = 36.7). Participants were recruited using Amazon Mechanical Turk (www.mturk.com). Of these participants, 139 listeners participated in the 2AFC identification task, 141 listeners participated in the comprehensibility rating task, and 140 listeners participated in the talker effort rating task. None of the listeners reported a history of speech or hearing impairment. All participants resided in the United States, and self-reported to be native speakers of American English. None of the participants reported experience with Mandarin Chinese.

5.2.2. Materials

Materials were the native and non-native speech analyzed in Experiment 3 (Chapter 4). Specifically, materials consisted of the target words produced in the context-production task, by 22 native English talkers (Native English; 18 females, 4 males), 22 higher-proficiency native Mandarin talkers (Native Mandarin-High; 19 females, 3 males), and 22 lower-proficiency native Mandarin talkers (Native Mandarin-Low; 15 females, 7 males). These talkers produced 80 English monosyllabic target words in the phrase, *"Click on the ___ now".* The 80 targets consisted of 4 segment types: 20 targets with the L1L2 consonant contrast (/p/-/b/ in word-initial position as in *peer* vs. *beer),* 20 targets with the L2-only consonant contrast (/p/-/b/ in word-final position as in *cap* vs. *cab*), 20 targets with the L1L2 vowel contrast (/ai/-/ei/ as in *light* vs. *late*), and 20 targets with the L2-only vowel contrast (/i/ vs. /ɪ/ as in *seek* vs. *sick*). Half of these targets were produced in Context

conditions, where both the target and its minimal pair neighbor were presented on the screen with a filler as the third word (e.g., ***peer,*** *beer, town*). Half of the targets were produced in No Context conditions, where the target was presented with two fillers (e.g., *soft,* ***peer****, noon);* see Chapter 4 for further details about the native and non-native talkers, the production materials, and the context-production task.

The items that were excluded from the acoustic analysis (in Chapter 4) due to mispronunciation and disfluency (e.g., repetition), were also excluded from the items in the perception experiments. Thus, there was a total of 2596 unique items (i.e., 878 Native English items, 869 Native Mandarin-High items, 849 Native Mandarin-Low items). These speech files were RMS normalized to 65 dB, then silence of 500 ms was added at the beginning and end of each sound file.

5.2.3. Procedure

The experiment was conducted online with Qualtrics (https://www.qualtrics.com/). The participants followed the link posted on the Mechanical Turk to complete the task on Qualtrics. They were told that they would listen to short English sentences and make responses to them. They were also instructed to use headphones to complete the task. The experiment began with a consent procedure as well as a sound check to ensure that participants could listen to the audio files at their comfortable volume. After that, participants proceeded with the task. In the 2AFC identification task, in each trial, listeners heard an English phrase once (e.g., *"Click on the peer now"*), and were asked to choose the target word from the two options displayed on the computer screen (e.g., *peer* or *beer*). They were asked to respond as accurately and quickly as possible. In the comprehensibility

rating task, in each trial, listeners heard an English phrase (e.g., *"Click on the peer now"*) and were asked to evaluate how easy/difficult the speech is to understand, by choosing a number from 1 (very easy to understand) to 9 (very difficult to understand). In the talker effort rating task, listeners were asked to evaluate how hard the talker is trying to speak clearly, by choosing a number from 1 (the talker is trying extremely hard to speak clearly) to 9 (the talker is not trying at all to speak clearly). They also completed two practice trials in the beginning with the talkers and items that were different from the following test items.

During the test trials, each participant listened to 75-80 items produced by two talkers; the number of items varied because some of the 80 unique items were excluded due to mispronunciations or disfluency for some talkers. They heard two talkers from the same talker group (i.e., each participant heard two Native English talkers, two Native Mandarin-High talkers, or two Native Mandarin-Low talkers). Each listener heard all 4 segment types (20 items with the L1L2 consonant contrast, 20 items with the L2-only consonant contrast, 20 items with the L1L2 vowel contrast, 20 items with the L2-only vowel contrast). Within each segment type, half of them were produced by one talker and the other half was produced by another talker (e.g., of the 20 items with the L1L2 consonant contrast, 10 items were produced by talker A and 10 items were produced by talker B). For each segment type produced by one talker (e.g., 10 items with the L1L2 consonant contrast produced by talker A), half of them were with one phoneme (e.g., 5 items with /p/-initial targets) and the other half was with another phoneme (e.g., 5 items with /b/-initial targets). For each segment type produced by one talker (e.g., 10 items with the L1L2 consonant contrast produced by talker A), some of them were produced in Context conditions and others were produced in No Context conditions (e.g., 3 of the 5 items with /p/-initial targets

in Context conditions; 2 of the 5 items with /b/-initial targets in Context conditions; counter-balanced across listeners). The presentation order of the items was randomized for each listener. After the experiment trials were completed, each listener completed a post-test demographic survey.

5.2.4. Analysis

Each response in the 2AFC identification task was given a score or 0 or 1. For example, if the participant heard the phrase, *"Click on the peer now"*, and chose *'peer'* instead of *'beer*', the response was scored as 1; otherwise, 0 was given. For the comprehensibility and talker effort rating data, raw data were analyzed because we were interested in examining differences among talker groups; for example, whether native English speech was generally perceived to be easier to understand than lower-proficiency non-native talkers' speech. Each item from each talker's speech was evaluated by 3-5 listeners in each task (i.e., 2AFC identification, comprehensibility rating, talker effort rating). There were 10942 data points for the 2AFC identification task, 11105 data points for the comprehensibility rating task, and 11035 data points for the talker effort rating task.

5.3. Results and discussion

Here, we present the results and discussion separately for different perception tasks: 2AFC identification task, comprehensibility rating task, and talker effort rating task. In analyses of the data in these tasks, the same basic structure of the mixed-effects regression model was used. First, we analyzed whether the proportion of correct responses (for 2AFC identification) or rating (for comprehensibility and talker effort) differed for different

Talker Groups' speech (Native English, Native Mandarin-High, Native Mandarin-Low), Condition (Context, No Context), and Segment Type (L1L2 consonant contrast: /p/-/b/ in word-initial position, L1L2 vowel contrast: /ai/-/ei/, L2-only consonant contrast: /p/-/b/ in word-final position, L2-only vowel contrast: /i/-/ɪ/). Talker Group was contrast coded to compare between Native English and Native Mandarin-High (.5, -.5, 0) and between Native Mandarin-High and Native Mandarin-Low (0, .5, -.5). Condition was contrast coded to compare between Context (.5) and No Context (-.5). Segment Type was contrast coded to compare items with the L1L2 consonant contrast vs. L2-only consonant contrast (.5, 0, -.5, 0), items with the L1L2 vowel contrast vs. L2-only vowel contrast (0, .5, 0, -.5), and items with consonant contrasts (L1L2 and L2-only) vs. vowel contrasts (L1L2 and L2-only; .25, -.25, .25, -.25). The interaction among these fixed factors was also included in the model. Models also included the maximal random effects structure that would converge, which included random intercepts for word, talker, and listener, as well as by-word random slope for Talker Group, by-talker random slopes for Condition, Segment Type, and the interaction between the two, and by-listener slopes for Condition, Segment Type, and the interaction between the two. In each model, the random effects that did not account for any variance (e.g., by-talker random slope for Condition x Segment Type) were not included in the model to avoid overfitting of the model; see the tables with model summaries for the model syntax and the specific random effect structure.

We also analyzed whether the proportion of correct responses (for 2AFC identification) or rating (for comprehensibility and talker effort) differed for different segments within each Segment Type (e.g., 2AFC identification proportion correct for /p/-initial targets vs. /b/-initial targets for items with the L1L2 consonant contrast). For each

model with items with each Segment Type (i.e., L1L2 consonant contrast, L1L2 vowel contrast, L2-only consonant contrast, L2-only vowel contrast), fixed effects were Talker Group (Native English, Native Mandarin-High, Native Mandarin-Low), Condition (Context, No Context), and Phoneme. Talker Group and Condition were contrast coded as specified above. For the models with items with the L1L2 and L2-only consonant contrasts, Phoneme was contrast coded to compare between /b/-items (.5) vs. /p/-items (-.5). For the model with items with the L1L2 vowel contrast, Phoneme was contrast coded to compare between /ai/-items (.5) vs. /ei/-items (-.5). For the model with items with the L2-only vowel contrast, Phoneme was contrast coded to compare between /i/-items (.5) vs. /ɪ/-items (-.5). The interaction among these fixed factors was also included in the models. Models also included the maximal random effects structure that would converge, which included random intercepts for word, talker, and listener, as well as by-word random slope for Talker Group, by-talker random slopes for Condition, Phoneme, and the interaction between the two, and by-listener slopes for Condition, Phoneme, and the interaction between the two. In each model, the random effects that did not account for any variance (e.g., by-talker random slope for Condition x Phoneme) were not included in the model to avoid overfitting of the model; see the tables with model summaries for the model syntax and the specific random effect structure.

5.3.1. 2AFC identification task

5.3.1.1. Results

Figure 5.1 shows the mean proportion correct for the 2AFC identification task by Segment Type (items with the L1L2 consonant contrast, L1L2 vowel contrast, L2-only

consonant contrast, L2-only vowel contrast), Talker Group (Native English, Native Mandarin-High, Native Mandarin-Low), and Condition (No Context, Context). The figure suggests that listeners' identification accuracy for Native Mandarin (High and Low) talkers' items was generally lower than the accuracy for Native English talkers' items, and this difference was larger for items with the L2-only consonant and L2-only vowel contrasts compared to those with the L1L2 consonant and L1L2 vowel contrasts. The identification accuracy for Native English talkers' items was at ceiling across different Segment Types. In order to examine whether adding noise to the speech materials would lower listeners' 2AFC identification for native English talkers' items from the ceiling performance, we mixed 10 native English talkers' items with a speech-shaped noise at -6dB SNR and presented them to a different set of native English listeners in a 2AFC identification task. The identification accuracy for different Segment Types was on average 92% correct (range: 87% - 95%). Thus, we analyzed the data for the 2AFC identification task in the quiet condition.

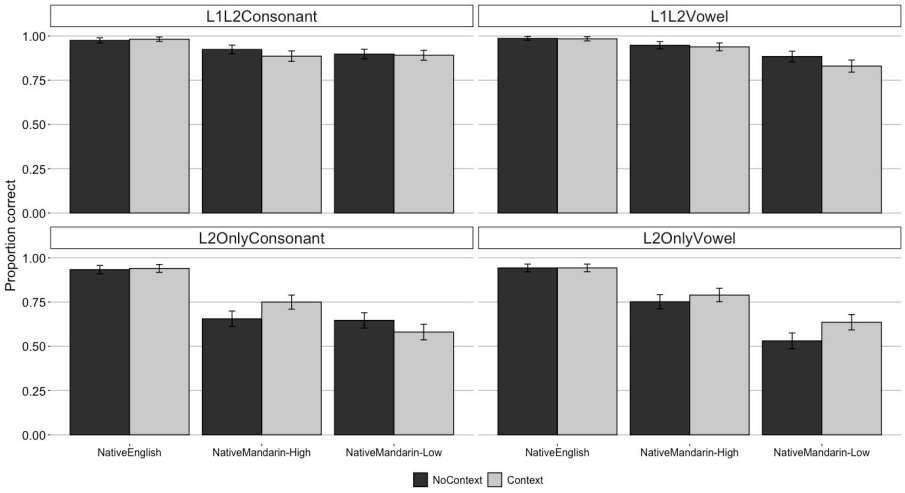

Figure 5.1. Mean proportion correct for the 2AFC identification task by Segment Type (L1L2 consonant contrast: /p/-/b/ in word-initial position, L1L2 vowel contrast: /ai/-/ei/, L2-only consonant contrast: /p/-/b/ in word-final position, L2-only vowel contrast: /i/-/ɪ/), Talker Group (Native English, Native Mandarin-High, Native Mandarin-Low), and Condition (No Context, Context). Error bars represent 95% confidence intervals.

In order to examine whether identification accuracy differed for different Talker Groups' speech, Condition, and Segment Type, we analyzed the data using logistic mixed-effects regression models with listeners' 2AFC identification accuracy (i.e., correct or incorrect) as the dependent variable. The fixed-effect and random-effect structure are specified above (see also Table 5.1 for the model syntax and summary of the results). The model showed significant effects of Talker Group comparisons: Native English vs. Native Mandarin-High ($\beta = 3.12$, $z = 9.31$, $p < .001$), Native Mandarin-High vs. Native Mandarin-Low ($\beta = 2.38$, $z = 8.37$, $p < .001$). This indicates that listeners' identification accuracy was higher for the items produced by Native English talkers than those produced by Native Mandarin-High talkers, and for the items produced by Native Mandarin-High talkers than those produced by Native Mandarin-Low talkers. There was a significant effect of the L1L2 vs. L2-only consonant contrast ($\beta = 1.6$, $z = 6.65$, $p < .001$), indicating that the identification accuracy was higher for items with the L1L2 consonant contrast (/p/-/b/ in word-initial position) than for those with the L2-only consonant contrast (/p/-/b/ in word-final position). There was also a significant effect of the L1L2 vs. L2-only vowel contrast ($\beta = 1.73$, $z = 6.97$, $p < .001$), indicating that the identification accuracy was higher for items with the L1L2-vowel contrast (/ai/-/ei/) than for those with the L2-only vowel contrast (/i/-/ɪ/). Further, these effects of Talker Group and Segment Type interacted in some comparisons. Specifically, the identification accuracy was generally higher for items produced by Native Mandarin-High talkers than those produced by Native Mandarin-Low talkers. This difference was larger for items with the L2-only consonant contrast than for items with the L1L2 consonant contrast (Native Mandarin-High vs. Native Mandarin-Low comparison x L1L2 vs. L2-only consonant contrast: $\beta = -.69$, $z = -2.0$, $p < .05$), as well as

for items with the vowel contrasts than for items with the consonant contrasts (Native Mandarin-High vs. Native Mandarin-Low comparison x L1L2 and L2 only consonant contrasts vs. L1L2 and L2-only vowel contrasts: $\beta = -.69$, $z = -2.0$, $p < .05$). The effect of Condition was not significant ($\beta = .03$, $z = .19$, $p = .85$), indicating that listeners' identification accuracy did not differ for items produced in No Context vs. Context conditions. Thus, overall, the talkers' target language experience impacted listeners' identification accuracy of the target words. Listeners' identification accuracy was also affected by non-native talkers' English proficiency differently depending on whether or not the target English contrast exists in the talkers' native language.

Further, we examined whether listeners' identification accuracy differed for different Phonemes within each Segment Type (e.g., identification accuracy for /p/-initial targets vs. /b/-initial targets for items with the L1L2 consonant contrast). Figure 5.2 is a different illustration of the same identification data, showing the mean proportion correct by Segment Type (items with the L1L2 consonant contrast, L1L2 vowel contrast, L2-only consonant contrast, L2-only vowel contrast), Talker Group (Native English, Native Mandarin-High, Native Mandarin-Low), Condition (No Context, Context), and Phoneme (/b/ and /p/ for the L1L2 consonant and L2-only consonant contrasts; /ai/ and /ei/ for the L1L2 vowel contrast; /i/ and /ɪ/ for the L2-only vowel contrast). We examined the effects of Talker Group, Condition, and Phoneme separately for items with each Segment Type, using logistic mixed-effects regression models with listeners' 2AFC identification accuracy as the dependent variable. The fixed-effect and random-effect structure are specified above (see each table below for the model syntax and summary of the results).

Table 5.1. Summary of the logistic mixed-effects regression model for 2AFC response correct data for all segment types.

2AFC Model for all Segment Types `Response correct ~ TalkerGroup * Condition * SegmentType + (1+ TalkerGroup| Word) + (1+ Condition| Talker) + (1| Listener)`				
Fixed Effects	**Estimate**	**S.E.**	**z-val.**	***p*-val.**
(Intercept)	2.6	.13	20.18	
TalkerGroup1 (Native English vs. Native Mandarin-High)	3.12	.33	9.31	< .001 ***
TalkerGroup2 (Native Mandarin-High vs. Native Mandarin-Low)	2.38	.28	8.37	< .001 ***
Condition (No Context vs. Context)	.03	.18	.19	.85
SegmentType1 (L1L2 consonant vs. L2-only consonant)	1.6	.24	6.65	< .001 ***
SegmentType2 (L1L2 vowel vs. L2-only vowel)	1.73	.25	6.97	< .001 ***
SegmentType3 (L1L2 & L2-only consonant vs. L1L2 & L2-only vowel)	-.35	.34	-1.01	.31
TalkerGroup1: Condition	.21	.36	.59	.56
TalkerGroup2: Condition	.26	.26	.97	.33
TalkerGroup1: SegmentType1	-.7	.47	-1.49	.14
TalkerGroup2: SegmentType1	-.69	.35	-2.0	.047 *
TalkerGroup1: SegmentType2	-.38	.52	-.73	.47
TalkerGroup2: SegmentType2	-.02	.36	-.06	.95
TalkerGroup1: SegmentType3	-.38	.7	-.55	.58
TalkerGroup2: SegmentType3	-1.72	.5	-3.44	< .001 ***
Condition: SegmentType1	-.21	.48	-.44	.66
Condition: SegmentType2	-.52	.49	-1.05	.3
Condition: SegmentType3	.12	.69	.17	.87
TalkerGroup1: Condition: SegmentType1	.81	.94	.87	.39
TalkerGroup2: Condition: SegmentType1	-.89	.69	-1.29	.2
TalkerGroup1: Condition: SegmentType2	.78	1.03	.75	.45
TalkerGroup2: Condition: SegmentType2	1.02	.74	1.41	.16
TalkerGroup1: Condition: SegmentType3	1.12	1.39	.8	.42
TalkerGroup2: Condition: SegmentType3	1.02	1.0	1.02	.31

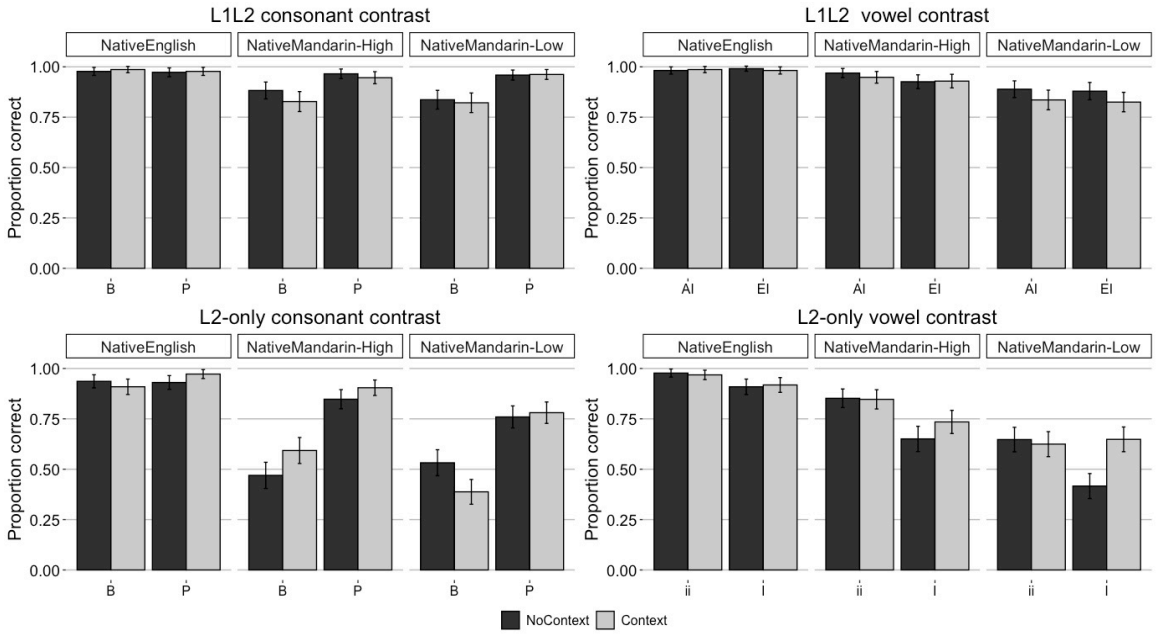

Figure 5.2. Mean proportion correct for the 2AFC identification task by Segment Type (L1L2 consonant contrast, L1L2 vowel contrast, L2-only consonant contrast, L2-only vowel contrast), Talker Group (Native English, Native Mandarin-High, Native Mandarin-Low), Condition (No Context, Context), and Phoneme (/b/ and /p/ for the L1L2 consonant and L2-only consonant contrasts; /ai/ and /ei/ for the L1L2 vowel contrast; /i/ and /ɪ/ for the L2-only vowel contrast). Error bars represent 95% confidence intervals.

For the model with the L1L2 consonant contrast items (see Table 5.2 for the model summary), there was a significant effect of the Native English vs. Native Mandarin-High group comparison ($\beta = 2.46$, $z = 3.22$, $p < .01$), but not the Native Mandarin-High vs. Native Mandarin-Low group comparison ($\beta = 1.17$, $z = 1.7$, $p = .89$). This indicates that for items with the L1L2 consonant contrast, listeners' identification accuracy was higher for Native English talkers' items than for Native Mandarin-High talkers' items, but identification accuracy did not differ for Native Mandarin-High vs. Native Mandarin-Low talkers' items. There was a significant effect of Phoneme ($\beta = -2.1$, $z = -2.78$, $p < .01$), indicating that identification accuracy was higher for /p/-items (e.g. *"Click on the peer now"*) than for /b/-items (e.g., *"Click on the beer now"*). This effect of Phoneme was larger for Native Mandarin-High talkers' items than for Native English talkers' items (Native

English vs. Native Mandarin-High x Phoneme: β = 3.14, z = 2.89, $p < .01$), and for Native Mandarin-Low talkers' items than for Native Mandarin-High talkers' items (Native Mandarin-High vs. Native Mandarin-Low x Phoneme: β = 2.15, z = 2.31, $p < .05$). The effect of Condition did not improve the model fit (β = -.08, z = -.37, $p = .71$). These results suggest that listeners' identification accuracy was impacted by the type of phoneme in the target contrast (/p/ or /b/ in word-initial position), and this effect differed for the perception of different talker groups' items.

Table 5.2. Summary of the logistic mixed-effects regression model for 2AFC response correct data for items with the L1L2 consonant contrast.

2AFC Model for L1L2 consonant contrast `Response correct ~ TalkerGroup * Condition * Phoneme + (1+ Phoneme| Talker) + (1+ Phoneme| Listener)`				
Fixed Effects	**Estimate**	**S.E.**	**z-val.**	***p*-val.**
(Intercept)	4.7	.42	11.22	
TalkerGroup1 (Native English vs. Native Mandarin-High)	2.46	.76	3.22	.0013 **
TalkerGroup2 (Native Mandarin-High vs. Native Mandarin-Low)	1.17	.69	1.7	.089
Condition (No Context vs. Context)	-.08	.23	-.37	.71
Phoneme (/b/ vs. /p/)	-2.1	.75	-2.78	.0054 **
TalkerGroup1: Condition	.99	.76	1.3	.2
TalkerGroup2: Condition	-.12	.59	-.21	.83
TalkerGroup1: Phoneme	3.14	1.09	2.89	.0039 **
TalkerGroup2: Phoneme	2.15	.93	2.31	.021 *
Condition: Phoneme	.02	.46	.04	.97
TalkerGroup1: Condition: Phoneme	.86	1.53	.57	.57
TalkerGroup2: Condition: Phoneme	.78	1.18	.67	.51

For the model with the L2-only consonant contrast items (see Table 5.3 for the model summary), there was a significant effect of the Native English vs. Native Mandarin-High group comparison (β = 3.39, z = 8.97, $p < .001$), and the Native Mandarin-High vs. Native Mandarin-Low group comparison (β = 2.39, z = 7.43, $p < .001$). This indicates that for items with the L2-only consonant contrast, listeners' identification accuracy was higher

for Native English talkers' items than for Native Mandarin-High talkers' items, and for Native Mandarin-High talkers' items than for Native Mandarin-Low talkers' items. There was a significant effect of Phoneme ($\beta = -1.28$, $z = -4.93$, $p < .001$), indicating that identification accuracy was higher for /p/-items (e.g., *"Click on the cap now"*) than for /b/-items (e.g., *"Click on the cab now"*). This effect of Phoneme was larger for Native Mandarin-High talkers' items than for Native English talkers' items (Native English vs. Native Mandarin-High x Phoneme: $\beta = 2.62$, $z = 3.41$, $p < .001$), but it did not differ for Native Mandarin-High vs. Native Mandarin-Low talkers' items (Native Mandarin-High vs. Native Mandarin-Low x Phoneme: $\beta = .73$, $z = 1.11$, $p = .27$). Though the effect of Condition was not significant ($\beta = .29$, $z = 1.94$, $p = .53$), it interacted with the Native Mandarin-High vs. Native Mandarin-Low group comparison ($\beta = 1.07$, $z = 3.05$, $p < .01$). This indicates that identification accuracy was higher for items produced in Context conditions than those produced in No Context conditions for Native Mandarin-High talkers' items, but this pattern was the opposite for Native Mandarin-Low talkers' items. This result was likely influenced by the pattern that accuracy for Native Mandarin-High talkers' items increased from No Context to Context condition for both /p/-items and /b/-items, while accuracy for Native Mandarin-Low talkers' items decreased from No Context to Context condition for /b/-items. This pattern may also have influenced that result that the effect of Condition interacted with Phoneme ($\beta = -.64$, $z = -2.55$, $p < .05$), indicating that identification accuracy for /p/-items increased from No Context to Context condition, while this pattern was the opposite for /b/-items. In sum, these results showed that in addition to the overall effect of talkers' target language experience on listeners' identification accuracy of the targets, the type of phoneme (/p/ or /b/ in word-final position)

and the condition where the items were produced also impacted listener's identification accuracy.

Table 5.3. Summary of the logistic mixed-effects regression model for 2AFC response correct data for items with the L2-only consonant contrast.

2AFC Model for L2-only consonant contrast `Response correct ~ TalkerGroup * Condition * Phoneme + (1+ Phoneme| Talker) + (1+ Condition+ Phoneme| Listener)`				
Fixed Effects	**Estimate**	**S.E.**	**z-val.**	***p*-val.**
(Intercept)	1.7	.13	13.35	
TalkerGroup1 (Native English vs. Native Mandarin-High)	3.39	.38	8.97	< .001 ***
TalkerGroup2 (Native Mandarin-High vs. Native Mandarin-Low)	2.39	.32	7.43	< .001 ***
Condition (No Context vs. Context)	.29	.15	1.94	.053
Phoneme (/b/ vs. /p/)	-1.28	.26	-4.93	< .001 ***
TalkerGroup1: Condition	.29	.49	.59	.55
TalkerGroup2: Condition	1.07	.35	3.05	.002 **
TalkerGroup1: Phoneme	2.62	.77	3.41	< .001 ***
TalkerGroup2: Phoneme	.73	.66	1.11	.27
Condition: Phoneme	-.64	.28	-2.25	.024 *
TalkerGroup1: Condition: Phoneme	-1.35	.94	-1.44	.15
TalkerGroup2: Condition: Phoneme	.36	.67	.54	.59

For the model with the L1L2 vowel contrast items (see Table 5.4 for the model summary), there was a significant effect of the Native English vs. Native Mandarin-High group comparison ($\beta = 2.62$, $z = 6.52$, $p < .001$), and the Native Mandarin-High vs. Native Mandarin-Low group comparison ($\beta = 2.44$, $z = 9.58$, $p < .001$). This indicates that for items with the L1L2 vowel contrast, listeners' identification accuracy was higher for Native English talkers' items than for Native Mandarin-High talkers' items, and for Native Mandarin-High talkers' items than for Native Mandarin-Low talkers' items. No other factors significantly improved the model fit. This suggests that listeners' identification accuracy of the targets with the English /ai/-/ei/ contrast was only impacted by talkers' target language experience.

Table 5.4. Summary of the logistic mixed-effects regression model for 2AFC response correct data for items with the L1L2 vowel contrast.

2AFC Model for L1L2 vowel contrast `Response correct ~ TalkerGroup * Condition * Phoneme + (1| Word)`				
Fixed Effects	**Estimate**	**S.E.**	**z-val.**	***p*-val.**
(Intercept)	3.19	.21	15.08	
TalkerGroup1 (Native English vs. Native Mandarin-High)	2.62	.41	6.52	< .001 ***
TalkerGroup2 (Native Mandarin-High vs. Native Mandarin-Low)	2.44	.25	9.58	< .001 ***
Condition (No Context vs. Context)	-.25	.42	-.6	.55
Phoneme (/ai/ vs. /ei/)	.29	.42	.7	.49
TalkerGroup1: Condition	.24	.8	.29	.77
TalkerGroup2: Condition	.37	.51	.73	.47
TalkerGroup1: Phoneme	-.67	.8	-.83	.41
TalkerGroup2: Phoneme	.28	.51	.55	.58
Condition: Phoneme	.36	.84	.43	.67
TalkerGroup1: Condition: Phoneme	1.87	1.6	1.17	.24
TalkerGroup2: Condition: Phoneme	.52	1.02	.51	.61

For the model with the L2-only vowel contrast items (see Table 5.5 for the model summary), there was a significant effect of the Native English vs. Native Mandarin-High group comparison ($\beta = .35$, $z = 16.86$, $p < .001$), and the Native Mandarin-High vs. Native Mandarin-Low group comparison ($\beta = .36$, $z = 17.67$, $p < .001$). This indicates that for items with the L2-only vowel contrast, listeners' identification accuracy was higher for Native English talkers' items than for Native Mandarin-High talkers' items, and for Native Mandarin-High talkers' items than for Native Mandarin-Low talkers' items. There was a significant effect of Phoneme ($\beta = .11$, $z = 2.94$, $p < .01$), indicating that identification accuracy was higher for /i/-items (e.g., *"Click on the seek now"*) than for /ɪ/-items (e.g., *"Click on the sick now"*). This effect of Phoneme was larger for Native Mandarin-High talkers' items than for Native English talkers' items (Native English vs. Native Mandarin-High x Phoneme: $\beta = -.09$, $z = -2.22$, $p < .05$). Though the effect of Condition

was not significant (β = .05, z= 1.32, p = .2), it interacted with the Native English vs. Native Mandarin-High group comparison (β = -.09, z = -2.25, p < .05), and with the Native Mandarin-High vs. Native Mandarin-Low group comparison (β = -.11, z = -2.71, p < .01). This indicates that the increase in identification accuracy from items produced in No Context conditions to items produced in Context conditions was larger for Native Mandarin-Low talkers' items than for Native Mandarin-High talkers' items, and for Native Mandarin-High talkers' items than for Native English talkers' items. These interactions between Condition and Talker Groups were present for /ɪ/-items but not for /i/-items, as seen in the three-way interaction among Condition, Native English vs. Native Mandarin-High group comparison, and Phoneme (β = .21, z = 2.49, p < .05), as well as in the three-way interaction among Condition, Native Mandarin-High vs. Native Mandarin-Low group comparison, and Phoneme (β = .27, z = 3.29, p < .01). These results showed that in addition to the overall effect of talkers' target language experience on listeners' identification accuracy of the targets, the type of phoneme (/i/ or /ɪ/) and the condition where the items were produced also impacted listener's identification accuracy. This suggests that the acoustic-phonetic modifications made by non-native talkers in different conditions affected native listeners' identification accuracy of the target words.

Table 5.5. Summary of the logistic mixed-effects regression model for 2AFC response correct data for items with the L2-only vowel contrast.

2AFC Model for L2-only vowel contrast `Response correct ~ TalkerGroup * Condition * Phoneme + (1| Word)`				
Fixed Effects	**Estimate**	**S.E.**	**z-val.**	***p*-val.**
(Intercept)	.77	.02	42.54	
TalkerGroup1 (Native English vs. Native Mandarin-High)	.35	.02	16.86	< .001 ***
TalkerGroup2 (Native Mandarin-High vs. Native Mandarin-Low)	.36	.02	17.67	< .001 ***
Condition (No Context vs. Context)	.05	.04	1.32	.2
Phoneme (/i/ vs. /ɪ/)	.11	.04	2.94	.008 **
TalkerGroup1: Condition	-.09	.04	-2.25	.025 *
TalkerGroup2: Condition	-.11	.04	-2.71	.007 **
TalkerGroup1: Phoneme	-.09	.04	-2.22	.027 *
TalkerGroup2: Phoneme	.008	.04	.18	.86
Condition: Phoneme	-.12	.07	-1.71	.1
TalkerGroup1: Condition: Phoneme	.21	.08	2.49	.012 *
TalkerGroup2: Condition: Phoneme	.27	.08	3.29	.001 **

5.3.1.2. Summary of the main findings and discussion

Overall, these results have demonstrated that identification accuracy of target words in the phrase, *"Click on the ___ now"*, differed for items produced by different talker groups. That is, targets in native English talkers' items were correctly identified more often than those in higher-proficiency non-native talkers' items; targets in higher-proficiency talkers' items were correctly identified more often than those in lower-proficiency talkers' items. Identification accuracy also differed for different segment types; accuracy was lower for items with the L2-only consonant and vowel contrasts compared to items with the L1L2 consonant and vowel contrasts. However, as identification accuracy for native English speech was overall at ceiling, the lower identification accuracy for items with the L2-only contrasts (compared to items with the L1L2 contrasts) was likely driven by the perception of non-native talkers' speech. That is, for non-native talkers' speech, the difficulty

associated with producing a non-native contrast that does not exist in native language (i.e., L2-only consonant contrast: /p/-/b/ in word-final position, L2-only vowel contrast: /i/-/ɪ/) may have manifested in the low identification accuracy in native English listeners' perception.

For items with the L2-only contrasts, listeners' identification accuracy was also influenced by individual phonemes of the target contrasts as well as the production condition of the items (No Context, Context). Particularly, for non-native (higher- and lower-proficiency) talkers' items with the L2-only consonant contrast, listeners identified targets in /p/-items (e.g., *"Click on the cap now"*) correctly more often than /b/-items (e.g., *"Click on the cab now"*). These perceptual patterns may originate from acoustic characteristics of these items produced by non-native talkers. Specifically, the acoustic analysis in Chapter 4 showed that preceding vowel durations in /b/-targets (e.g., *cab*) in non-native talkers' items were shorter than those of native talkers, and were more similar to non-native talkers' preceding vowel durations in /p/-targets. Similarly, for non-native talkers' productions, voicing proportions of /b/ were much smaller than those of native English talkers. These production patterns of the word-final /b/-/p/ consonants suggest that acoustic characteristics of the /b/-final targets produced by non-native talkers were similar to those of the /p/-final targets, and this may have biased native English listeners' perception of /b/-final targets produced by non-native talkers toward /p/-final targets. However, interestingly, identification accuracy of /b/-targets improved for those produced in Context conditions compared to those produced in No Context conditions for higher-proficiency talkers' productions, but this pattern was the opposite for lower-proficiency talkers' productions. This suggests that contextually-relevant speech adjustments for /b/-

targets made by higher-proficiency talkers (e.g., increasing voicing proportions of word-final /b/, as shown in Chapter 4) did improve native listeners' identification accuracy, though the adjustments made by lower-proficiency talkers (e.g., not changing or decreasing voicing proportions of word-final /b/) lowered listeners' identification accuracy.

Identification accuracy of items with the L2-only vowel contrast may also reflect acoustic characteristic of talkers' productions, in addition to native English listeners' perceptual tendencies. For example, as shown in Chapter 4, the duration differences between /i/ (e.g., *seek*) and /ɪ/ (e.g., *sick*) were similar between native English talkers' productions and higher-proficiency non-native talkers' productions. However, listeners' identification accuracy was higher for native English talkers' productions than for higher-proficiency talkers' productions, possibly because the spectral difference between /i/ and /ɪ/ was much larger for native English talkers' productions than for higher-proficiency talkers' productions. This relationship between production patterns and identification accuracy suggests that native English listeners may have relied on one type of acoustic cue more heavily than another. That is, the spectral differences between /i/ and /ɪ/ (larger differences for native English than higher-proficiency talkers' productions) may have contributed to listeners' identification accuracy more than duration differences (similar difference for native English and higher-proficiency talkers' productions), which is consistent with previous studies demonstrating that native English listeners use spectral differences as a primary cue to distinguish the tense-lax vowel sounds (e.g., Hillenbrand et al., 2000).

However, the results also suggest that listeners' identification accuracy may have been impacted by non-native talkers' attempt to enhance the /i/-/ɪ/ contrast not only in spectral differences but also in duration. Particularly, as observed in production results

(Chapter 4), lower-proficiency talkers' duration of /ɪ/ was similar to that of /i/ in No Context conditions, though they increased the duration difference between these vowels in Context conditions by shortening the duration of /ɪ/. The identification results showed that native listeners' perception of lower-proficiency talkers' /ɪ/ was biased toward /i/ in No Context conditions, but the /ɪ/ identification accuracy improved in Context conditions. Similarly, higher-proficiency talkers also shortened the duration of /ɪ/ in Context conditions compared to No Context conditions, and their /ɪ/-targets produced in Context conditions were correctly identified more often than those produced in No Context conditions. It is difficult to determine, though, how duration and spectral enhancements impacted listeners' identification accuracy of these vowels as the higher- and lower-proficiency non-native talkers also enhanced the special difference between /i/ and /ɪ/ to some extent. However, the current results suggest that non-native talkers of different proficiency levels could make contextually-relevant acoustic modifications for a non-native vowel to improve native listeners' perceptual identification accuracy. A further analysis is needed to directly investigate the relationship between different types of acoustic modifications used in contextually-relevant segmental enhancements and listeners' identification accuracy, as well as whether native listeners use similar cue weighting strategies for perception of native and non-native speech.

5.3.2. Comprehensibility rating task

5.3.2.1. Results

Figure 5.3 shows the mean comprehensibility rating (1: very easy to understand, 9: very difficult to understand) by Segment Type (items with the L1L2 consonant contrast,

L1L2 vowel contrast, L2-only consonant contrast, L2-only vowel contrast), Talker Group (Native English, Native Mandarin-High, Native Mandarin-Low), and Condition (No Context, Context). The figure suggests that overall, comprehensibility ratings were lower (i.e., perceived to be easier to understand) for Native Mandarin-High talkers' items than for Native Mandarin-Low talkers' items, but this difference is much smaller between Native Mandarin-High talkers' and Native English talkers' items. Further, the figure suggests that the effect of Condition (No Context, Context) was different for different Segment Types. Particularly, for items with the L2-only consonant contrast, listeners perceived items produced in Context conditions to be easier to understand than those produced in No Context conditions, though this pattern was much less clear in other Segment Types.

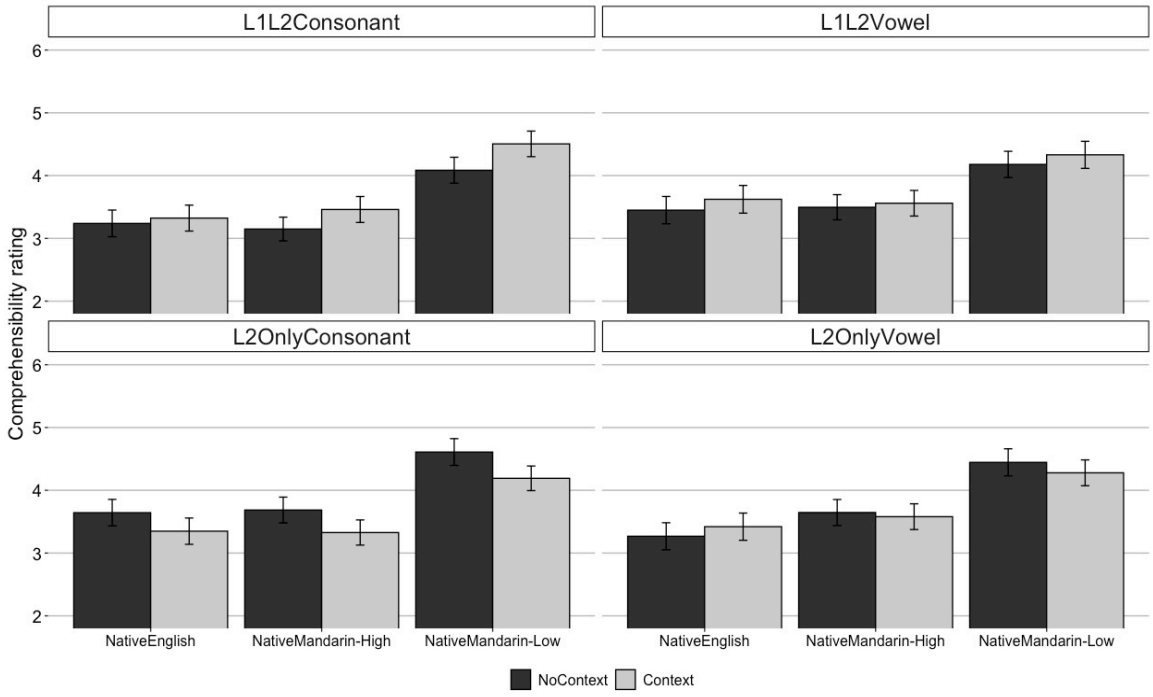

Figure 5.3. Mean comprehensibility rating (1: very easy to understand, 9: very difficult to understand) by Segment Type (L1L2 consonant contrast, L1L2 vowel contrast, L2-only consonant contrast, L2-only vowel contrast), Talker Group (Native English, Native Mandarin-High, Native Mandarin-Low), and Condition (No Context, Context). Error bars represent 95% confidence intervals.

In order to examine whether comprehensibility ratings differed for different Talker Groups' speech, Condition, and Segment Type, we analyzed the data using linear mixed-effects regression models with comprehensibility rating as the dependent variable. The fixed-effect and random-effect structure are specified above (see also Table 5.6 for the model syntax and summary of the results). P-values were calculated based on Satterthwaite approximations (Luke, 2017), using the lmerTest package for R (Kuznetsova et al., 2016). The model showed a significant effect of the Native Mandarin-High vs. Native Mandarin-Low group comparison (β = -1.31, t = -3.93, $p < .001$), but not the Native English vs. Native Mandarin-High group comparison (β = -.63, t = -1.88, p = .65). This indicates that listeners perceived Native Mandarin-High talkers' items to be easier to understand than Native Mandarin-Low talkers' items, but there was not a significant difference in perceived comprehensibility between Native Mandarin-High and Native English talkers' items. There was also a significant interaction between Condition and the L1L2 consonant vs. L2-only consonant contrast comparison (β = .64, t = .22, $p < .01$). This indicates that listeners perceived items produced in Context conditions to be easier to understand than those produced in No Context conditions for the L2-only consonant contrast items, but this pattern was different for the L1L2 consonant contrast items. A post-hoc Tukey test confirmed that the effect of Condition was significant for items with the L2-only consonant contrast (β = .36, SE = .15, z.ratio = 2.35, p = .019), but not for items with other Segment Types: L1L2 consonant contrast (β = -.27, SE = .15, z.ratio = -1.77, p = .08), L1L2 vowel contrast (β = -.13, SE = .16, z.ratio = -.84, p = .41), L2-only vowel contrast (β = .03, SE = .15, z.ratio = .21, p = .083). These results showed that overall, native listeners perceived Native English and Native Mandarin-High talkers' speech to be easier to understand than

that of Native Mandarin-Low talkers' speech. Listeners perceived items produced in Context conditions easier to understand than those produced in No Context conditions, only for the items contained the targets with the /p/-/b/ contrast in the word-final position.

Table 5.6. Summary of the linear mixed-effects regression model for comprehensibility ratings for all segment types.

Comprehensibility Model for all Segment Types `Rating ~ TalkerGroup * Condition * SegmentType + (1| Word) + (1| Talker)`				
Fixed Effects	**Estimate**	**S.E.**	**t-val.**	***p*-val.**
(Intercept)	3.74	.12	30.52	
TalkerGroup1 (Native English vs. Native Mandarin-High)	-.63	.33	-1.88	.065
TalkerGroup2 (Native Mandarin-High vs. Native Mandarin-Low)	-1.31	.33	-3.93	< .001 ***
Condition (No Context vs. Context)	.002	.08	.02	.98
SegmentType1 (L1L2 consonant vs. L2-only consonant)	-.18	.11	-1.66	.1
SegmentType2 (L1L2 vowel vs. L2-only vowel)	.0007	.11	.006	.995
SegmentType3 (L1L2 & L2-only consonant vs. L1L2 & L2-only vowel)	-.12	.15	-.79	.43
TalkerGroup1: Condition	.05	.11	.47	.64
TalkerGroup2: Condition	.04	.11	.4	.69
TalkerGroup1: SegmentType1	-.08	.16	-.48	.63
TalkerGroup2: SegmentType1	-.11	.15	-.7	.49
TalkerGroup1: SegmentType2	.38	.16	2.44	.015 *
TalkerGroup2: SegmentType2	.19	.16	1.23	.22
TalkerGroup1: SegmentType3	.04	.22	.2	.84
TalkerGroup2: SegmentType3	-.35	.22	-1.6	.11
Condition: SegmentType1	.64	.22	2.91	.004 **
Condition: SegmentType2	.16	.22	.74	.46
Condition: SegmentType3	-.19	.31	-.61	.55
TalkerGroup1: Condition: SegmentType1	-.51	.31	-1.62	.11
TalkerGroup2: Condition: SegmentType1	-.42	.31	-1.36	.17
TalkerGroup1: Condition: SegmentType2	-.28	.31	-.9	.37
TalkerGroup2: Condition: SegmentType2	-.32	.31	-1.04	.3
TalkerGroup1: Condition: SegmentType3	-.7	.44	-1.57	.12
TalkerGroup2: Condition: SegmentType3	-.36	.44	-.82	.42

Further, we examined whether listeners' comprehensibility ratings differed for different Phonemes within each Segment Type (e.g., ratings for /p/-initial targets vs. /b/-

initial targets for items with the L1L2 consonant contrast). Figure 5.4 is a different illustration of the same rating data, showing the mean comprehensibility rating by Segment Type (items with the L1L2 consonant contrast, L1L2 vowel contrast, L2-only consonant contrast, L2-only vowel contrast), Talker Group (Native English, Native Mandarin-High, Native Mandarin-Low), Condition (No Context, Context), and Phoneme (/b/ and /p/ for the L1L2 consonant and L2-only consonant contrasts; /ai/ and /ei/ for the L1L2 vowel contrast; /i/ and /ɪ/ for the L2-only vowel contrast). We examined the effects of Talker Group, Condition, and Phoneme separately for items with each Segment Type, using linear mixed-effects regression models with listeners' comprehensibility rating as the dependent variable. The fixed-effect and random-effect structure are specified above (see each table below for the model syntax and summary of the results).

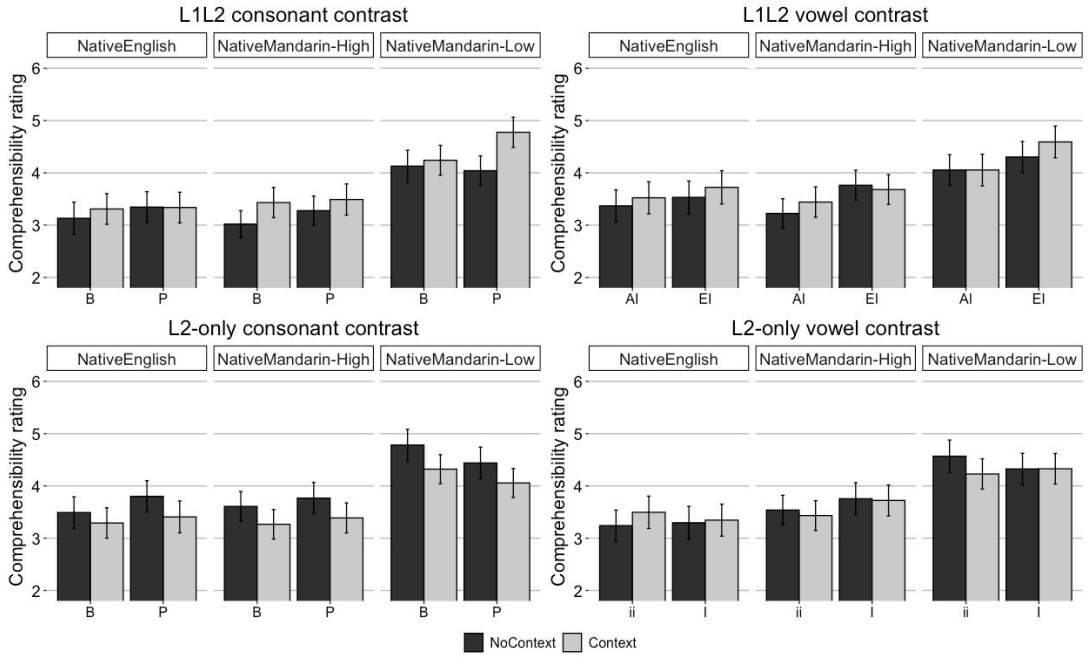

Figure 5.4. Mean comprehensibility rating (1: very easy to understand, 9: very difficult to understand) by Segment Type (L1L2 consonant contrast, L1L2 vowel contrast, L2-only consonant contrast, L2-only vowel contrast), Talker Group (Native English, Native Mandarin-High, Native Mandarin-Low), Condition (No Context, Context), and Phoneme (/b/ and /p/ for the L1L2 consonant and L2-only consonant contrasts; /ai/ and /ei/ for the L1L2 vowel contrast; /i/ and /ɪ/ for the L2-only vowel contrast). Error bars represent 95% confidence interval.

For the model with the L1L2 consonant contrast items (see Table 5.7 for the model summary), there was a significant effect of the Native Mandarin-High vs. Native Mandarin-Low group comparison ($\beta = -1.45$, $t = -3.33$, $p < .01$). This indicates that for items with the L1L2 consonant contrast, listeners perceived Native Mandarin-High talkers' items to be easier to understand than Native Mandarin-Low talkers' items. None of the other factors significantly improved the model fit. Thus, non-native talkers' proficiency level was the only factor that impacted listeners' perceived degree of comprehensibility for the items that contained targets with the word-initial /p/-/b/ contrast.

Table 5.7. Summary of the linear mixed-effects regression model for comprehensibility ratings for items with the L1L2 consonant contrast.

Comprehensibility Model for L1L2 consonant contrast
`Rating ~ TalkerGroup * Condition * Phoneme`
`+ (1+ TalkerGroup| Word) + (1+ Condition*Phoneme| Talker)`
`+ (1| Listener)`				

Fixed Effects	**Estimate**	**S.E.**	**t-val.**	***p*-val.**
(Intercept)	3.64	.16	22.26	
TalkerGroup1 (Native English vs. Native Mandarin-High)	-.72	.44	-1.64	.1
TalkerGroup2 (Native Mandarin-High vs. Native Mandarin-Low)	-1.45	.43	-3.33	.002 **
Condition (No Context vs. Context)	.27	.14	1.93	.066
Phoneme (/b/ vs. /p/)	-.17	.14	-1.26	.22
TalkerGroup1: Condition	-.37	.24	-1.51	.14
TalkerGroup2: Condition	-.26	.25	-1.04	.31
TalkerGroup1: Phoneme	.09	.23	.39	.7
TalkerGroup2: Phoneme	.09	.23	.39	.7
Condition: Phoneme	-.14	.28	-.5	.62
TalkerGroup1: Condition: Phoneme	.31	.46	.68	.51
TalkerGroup2: Condition: Phoneme	.74	.47	1.59	.13

For the model with the L2-only consonant contrast items (see Table 5.8 for the model summary), there was a significant effect of the Native Mandarin-High vs. Native Mandarin-Low group comparison ($\beta = -1.32$, $t = -3.88$, $p < .001$), indicating that for items

with the L2-only consonant contrast, listeners perceived Native Mandarin-High talkers' items to be easier to understand than Native Mandarin-Low talkers' items. There was also a significant effect of Condition (No Context vs. Context; $\beta = -.36$, $t = -2.46$, $p < .05$), indicating that listeners perceived the L2-only consonant contrast items produced in Context conditions to be easier to understand than those produced in No Context conditions. This effect was similar across items produced by different Talker Groups (Condition x Native English vs. Native Mandarin-High group comparison: $\beta = .12$, $t = .55$, $p = .58$, Condition x Native Mandarin-High vs. Native Mandarin-Low group comparison: $\beta = .17$, $t = .78$, $p = .44$). Though the effect of Phoneme was not significant ($\beta = -.02$, $t = -.11$, $p = .92$), it interacted with the Native Mandarin-High vs. Native Mandarin-Low group comparison ($\beta = -.61$, $t = -2.8$, $p < .01$). This indicates that /p/-items (e.g., *"Click on the cap now")* were perceived to be easier to understand than /b/-items (e.g., *"Click on the cab now"*) for Native Mandarin-Low talkers' speech, but not for Native Mandarin-High talkers' speech. In sum, these results showed that listeners perceived higher-proficiency talkers' items to be easier to understand than lower-proficiency talkers' items. The type of production condition (Context or No Context), as well as the type of phoneme in the targets (/p/ or /b/), also impacted listeners' perceived degree of comprehensibility.

Table 5.8. Summary of the linear mixed-effects regression model for comprehensibility ratings for items with the L2-only consonant contrast.

Comprehensibility Model for L2-only consonant contrast `Rating ~ TalkerGroup * Condition * Phoneme + (1| Word) + (1| Talker)`				
Fixed Effects	**Estimate**	**S.E.**	**t-val.**	***p*-val.**
(Intercept)	3.81	.14	28.09	
TalkerGroup1 (Native English vs. Native Mandarin-High)	-.59	.34	-1.73	.09
TalkerGroup2 (Native Mandarin-High vs. Native Mandarin-Low)	-1.32	.34	-3.88	< .001 ***
Condition (No Context vs. Context)	-.36	.15	-2.46	.024 *
Phoneme (/b/ vs. /p/)	-.02	.15	-.11	.92
TalkerGroup1: Condition	.12	.22	.55	.58
TalkerGroup2: Condition	.17	.22	.78	.44
TalkerGroup1: Phoneme	-.41	.22	-1.83	.068
TalkerGroup2: Phoneme	-.61	.22	-2.8	.0052 **
Condition: Phoneme	.08	.3	.28	.78
TalkerGroup1: Condition: Phoneme	.02	.45	.04	.97
TalkerGroup2: Condition: Phoneme	-.11	.44	-.24	.81

For the model with the L1L2 vowel contrast items (see Table 5.9 for the model summary), there was a significant effect of the Native Mandarin-High vs. Native Mandarin-Low group comparison ($\beta = -1.1$, $t = -2.63$, $p < .01$), indicating that for items with the L1L2 vowel contrast, listeners perceived Native Mandarin-High talkers' items to be easier to understand than Native Mandarin-Low talkers' items. There was also a significant effect of Phoneme (/ai/-items vs. /ei/-items; $\beta = -.35$, $t = -2.44$, $p < .05$). This indicates that listeners perceived /ai/-items (e.g., *"Click on the light now"*) to be easier to understand than /ei/-items (e.g., *"Click on the late now"*). Thus, in addition to the effect of non-native talkers' proficiency level on listeners' perceived degree of comprehensibility, the type of phoneme in the target contrast also impacted the ratings; /ai/-targets were perceived to be easier to understand than /ei/-targets across different talker groups' items.

Table 5.9. Summary of the linear mixed-effects regression model for comprehensibility ratings for items with the L1L2 vowel contrast.

Comprehensibility Model for L1L2 vowel contrast `Rating ~ TalkerGroup * Condition * Phoneme + (1| Word) + (1+ Condition*Phoneme| Talker) + (1| Listener)`				
Fixed Effects	**Estimate**	**S.E.**	**t-val.**	***p*-val.**
(Intercept)	3.79	.16	23.28	
TalkerGroup1 (Native English vs. Native Mandarin-High)	-.49	.43	-1.15	.25
TalkerGroup2 (Native Mandarin-High vs. Native Mandarin-Low)	-1.1	.42	-2.63	.0095 **
Condition (No Context vs. Context)	.13	.14	.92	.37
Phoneme (/ai/ vs. /ei/)	-.35	.14	-2.44	.025 *
TalkerGroup1: Condition	.08	.18	.46	.64
TalkerGroup2: Condition	-.03	.18	-.18	.85
TalkerGroup1: Phoneme	.34	.18	1.91	.056
TalkerGroup2: Phoneme	.24	.18	1.36	.17
Condition: Phoneme	.01	.31	.04	.97
TalkerGroup1: Condition: Phoneme	.15	.49	.31	.76
TalkerGroup2: Condition: Phoneme	.76	.49	1.56	.12

For the model with the L2-only vowel contrast items (see Table 5.10 for the model summary), there was a significant effect of the Native English vs. Native Mandarin-High group comparison ($\beta = -.86$, $t = -2.19$, $p < .05$), and the Native Mandarin-High vs. Native Mandarin-Low group comparison ($\beta = -1.19$, $t = -3.01$, $p < .01$). This indicates that for items with the L2-only vowel contrast, listeners perceived Native English talkers' items to be easier to understand than Native Mandarin-High talkers' items; they also perceived Native Mandarin-High talkers' items to be easier to understand than Native Mandarin-Low talkers' items. No other factors significantly improved the model fit. Thus, talkers' target language experience was the only factor that impacted listeners' perceived degree of comprehensibility for the items with the /i/- and /ɪ/-targets.

Table 5.10. Summary of the linear mixed-effects regression model for comprehensibility ratings for items with the L2-only vowel contrast.

Comprehensibility Model for L2-only vowel contrast `Rating ~ TalkerGroup * Condition * Phoneme + (1+ TalkerGroup| Word) + (1| Listener)`				
Fixed Effects	**Estimate**	**S.E.**	**t-val.**	***p*-val.**
(Intercept)	3.78	.16	24.26	
TalkerGroup1 (Native English vs. Native Mandarin-High)	-.86	.4	-2.19	.03 *
TalkerGroup2 (Native Mandarin-High vs. Native Mandarin-Low)	-1.19	.4	-3.01	.003 **
Condition (No Context vs. Context)	-.03	.17	-.19	.85
Phoneme (/i/ vs. /ɪ/)	-.04	.17	-.24	.82
TalkerGroup1: Condition	-.37	.24	1.52	.14
TalkerGroup2: Condition	.29	.28	1.02	.32
TalkerGroup1: Phoneme	.18	.24	.72	.48
TalkerGroup2: Phoneme	-.26	.28	-.92	.37
Condition: Phoneme	-.07	.34	-.21	.84
TalkerGroup1: Condition: Phoneme	.55	.49	1.13	.27
TalkerGroup2: Condition: Phoneme	.57	.56	1.01	.32

5.3.2.2. Summary of the main findings and discussion

Overall, these results have demonstrated that perceived degree of comprehensibility differed for different talker groups' speech. That is, higher-proficiency non-native talkers' items were consistently perceived to be easier to understand than lower-proficiency talkers' items. This may reflect a difference in these talkers' ability to produce generally 'easy-to-understand' speech for native English listeners. That is, the difference in the characteristics of the whole phase, *"Click on the ___ now",* rather than the characteristics of the segmental contrast in the target word (e.g., *peer* vs. *beer*), may have influenced native English listeners' perception of comprehensibility for higher- and lower-proficiency talkers' speech. For example, previous work has demonstrated that perception of poorer comprehensibility of non-native speech is associated with slower speaking rate (Munro & Derwing, 1998) and stronger accentedness (Munro & Derwing, 1995b). The acoustic

analysis of the non-native talkers' items (see Chapter 4), showed that phrase duration was longer for lower-proficiency talkers' speech than for higher-proficiency talkers' speech. Further, lower-proficiency talkers' speech was perceived to be more heavily accented than higher-proficiency talkers' speech (see Chapter 4). Thus, it is possible that acoustic-phonetic properties of the whole phrase impacted the difference in perceived degree of comprehensibility for lower-proficiency talkers' items vs. higher-proficiency talkers' items. That is, native listeners may have based their judgements of how easy or difficult to understand the speech on their general perception of non-native talkers' overall proficiency (as in perceived foreign accentedness or fluency characteristics), in addition to the characteristics of the target words themselves.

However, perceived degree of comprehensibility overall did not differ for native English and higher-proficiency non-native talkers' items. That is, though higher-proficiency talkers' speaking rate was slower and their speech was perceived to be more accented than that of native English talkers (see Chapter 4), native English listeners generally perceived these talkers' items to be comprehensible to a similar degree. It is possible that these global characteristics of their speech were not different enough to the point that influences subjective perception of comprehensibility differently. However, perceived comprehensibility ratings differed for these talkers' speech for items with the L2-only vowel contrast (e.g., *"Click on the seek/sick now"*); native English talkers' items were perceived to be easier to understand than higher-proficiency non-native talkers' items. This may reflect the effect of how the target words (e.g., *seek or sick*), rather than the whole phrase (*"Click on the ___ now"*), were produced. That is, the difficulty associated with producing the English vowel contrast that does not exist in native Mandarin may have

manifested in higher-proficiency talkers' speech being perceived as less easy to understand compared to that of native English talkers' speech.

There were some differences in the comprehensibility ratings specific to individual segments. For example, for items with the L1L2 vowel contrast, /ai/-items (e.g., *"Click on the light now"*) were perceived to be easier to understand than /ei/-items (e.g., *"Click on the late now"*). Because this pattern was found in all talker groups' speech, it is possible that this result is related to the characteristics of the vowels included in these items and how they are perceived. That is, it is possible that listeners generally perceive lower vowels to be easier to understand than higher vowels, due to characteristics such as lower F0 associated with lower vowels (e.g., Van Hoof & Verhoeven, 2011; Whalen & Levitt, 1995). It is also possible that acoustic characteristics not examined in the current study (i.e., Chapter 4) impacted perceived degree of comprehensibility. For example, for items with L2-only consonant contrast (e.g., *"Click on the cap/cab now"*), listeners perceived items produced in Context conditions to be easier to understand than those produced in No Context conditions. This consistent improvement in perceived comprehensibility from No Context to Context conditions across different phonemes (/p/-items and /b/-items) and across different talker groups' speech (native English, higher-proficiency and lower-proficiency non-native talkers) is puzzling because this does not necessarily reflect acoustic characteristics of the items produced in these conditions. That is, as demonstrated in Chapter 4, there were not condition-based differences in global characteristics (e.g., duration, F0, or intensity) that were specific to the items with the L2-only consonant contrast. At the segmental level, the difference between No Context and Context conditions was not present in the duration of the vowels preceding the target consonants, for either of

the phonemes (e.g., *c**a**b* vs. *c**a**p*). Though native English and higher-proficiency non-native talkers increased voicing proportions of /b/ from No Context to Context conditions, lower-proficiency talkers showed the opposite pattern. Given these variable acoustic patterns for items produced in No Context vs. Context conditions, it is difficult to determine what facilitated the consistent improvement in perceived degree of comprehensibility from No Context to Context conditions for these items. However, this result also suggests that native and non-native talkers of differing proficiency levels made some acoustic modifications to their productions of simple phrases improving subjective evaluation of comprehensibility. A future investigation may examine a wider range of acoustic-phonetic modifications to explore what acoustic factors contribute to an improvement in subjective perception of comprehensibility.

5.3.3. Talker effort rating task

5.3.3.1. Results

Figure 5.5 shows the mean talker effort rating (1: the talker is trying extremely hard to speak clearly, 9: the talker is not trying at all to speak clearly) by Segment Type (items with the L1L2 consonant contrast, L1L2 vowel contrast, L2-only consonant contrast, L2-only vowel contrast), Talker Group (Native English, Native Mandarin-High, Native Mandarin-Low), and Condition (No Context, Context). The figure suggests that effort ratings did not differ among different Talker Groups' items. The figure also suggests that for Native English talkers' items, listeners perceived increased effort for items produced in Context conditions compared to those produced in No Context conditions, especially for items with the L1L2 consonant contrast and the L1L2 vowel contrast.

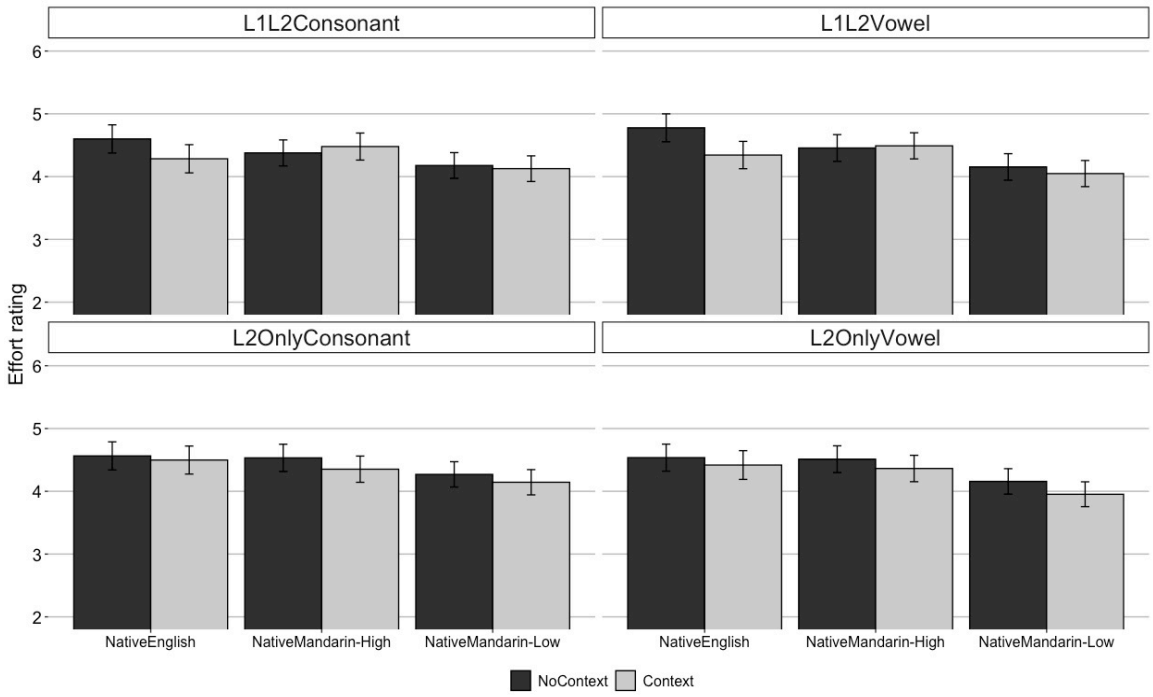

Figure 5.5. Mean talker effort rating (1: the talker is trying extremely hard to speak clearly, 9: the talker is not trying at all to speak clearly) by Segment Type (L1L2 consonant contrast, L1L2 vowel contrast, L2-only consonant contrast, L2-only vowel contrast), Talker Group (Native English, Native Mandarin-High, Native Mandarin-Low), and Condition (No Context, Context). Error bars represent 95% confidence intervals.

In order to examine whether effort ratings differed for different Talker Groups' speech, Condition, and Segment Type, we analyzed the data using linear mixed-effects regression models with effort rating as the dependent variable. The fixed-effect and random-effect structure are specified above (see also Table 5.11 for the model syntax and summary of the results). P-values were calculated as specified above. The model revealed that listeners' effort rating did not significantly differ for different Talker Groups' speech: Native English vs. Native Mandarin-High ($\beta = .32$, $t = .64$, $p = .52$), Native Mandarin-High vs. Native Mandarin-Low ($\beta = .57$, $t = 1.15$, $p = .25$). However, there was a significant effect of Condition (No Context, Context; $\beta = -.14$, $t = -3.84$, $p < .001$). This indicates that listeners perceived increased effort for items produced in Context conditions than for those

produced in No Context conditions. This effect of Condition was larger for Native English talkers' items than for Native Mandarin-High talkers' items (Condition x Native English vs. Native Mandarin-High group comparison: $\beta = -.19$, $t = -2.04$, $p < .05$). This interaction between Condition and the Native English vs. Native Mandarin-High group comparison was present for the L1L2 consonant contrast items, but not for the L2-only consonant contrast items (Condition x Native English vs. Native Mandarin-High group comparison x L1L2 consonant vs. L2-only consonant contrast comparison: $\beta = -.53$, $t = -1.98$, $p < .05$). Similarly, the interaction between Condition and the Native English vs. Native Mandarin-High group comparison was present for the L1L2 vowel contrast items, but not for the L2-only vowel contrast items (Condition x Native English vs. Native Mandarin-High group comparison x L1L2 vowel vs. L2-only vowel contrast comparison: $\beta = -.59$, $t = -2.23$, $p < .05$). These results suggest that listeners perceived increased effort for Native English talkers' items produced in Context conditions compared to those produced in No Context conditions, and this pattern was driven by their items with the L1L2 consonant (/p/-/b/ in word-initial position) and the L1L2 vowel contrasts (/ai/-/ei/).

Further, we examined whether listeners' talker effort ratings differed for different Phonemes within each Segment Type (e.g., ratings for /p/-initial targets vs. /b/-initial targets for items with the L1L2 consonant contrast). Figure 5.6 is a different illustration of the same rating data, showing the mean effort rating by Segment Type (items with the L1L2 consonant contrast, L1L2 vowel contrast, L2-only consonant contrast, L2-only vowel contrast), Talker Group (Native English, Native Mandarin-High, Native Mandarin-Low), Condition (No Context, Context), and Phoneme (/b/ and /p/ for the L1L2 consonant and L2-only consonant contrasts; /ai/ and /ei/ for the L1L2 vowel contrast; /i/ and /ɪ/ for the L2-

only vowel contrast). We examined the effects of Talker Group, Condition, and Phoneme separately for items with each Segment Type, using linear mixed-effects regression models with listeners' comprehensibility rating as the dependent variable. The fixed-effect and random-effect structure are specified above (see each table below for the model syntax and summary of the results).

Table 5.11. Summary of the linear mixed-effects regression model for effort ratings for all segment types.

Effort Model for all Segment Types `Rating ~ TalkerGroup * Condition * SegmentType + (1| Word) + (1| Talker) + (1| Listener)`				
Fixed Effects	**Estimate**	**S.E.**	**t-val.**	***p*-val.**
(Intercept)	4.29	.18	24.34	
TalkerGroup1 (Native English vs. Native Mandarin-High)	.32	.5	.64	.52
TalkerGroup2 (Native Mandarin-High vs. Native Mandarin-Low)	.57	.5	1.15	.25
Condition (No Context vs. Context)	-.14	.04	-3.84	< .001 ***
SegmentType1 (L1L2 consonant vs. L2-only consonant)	-.05	.05	-1.07	.29
SegmentType2 (L1L2 vowel vs. L2-only vowel)	.05	.05	.95	.34
SegmentType3 (L1L2 & L2-only consonant vs. L1L2 & L2-only vowel)	.03	.07	.48	.63
TalkerGroup1: Condition	-.19	.09	-2.04	.042 *
TalkerGroup2: Condition	-.03	.09	-.36	.72
TalkerGroup1: SegmentType1	-.07	.13	-.52	.6
TalkerGroup2: SegmentType1	-.001	.13	-.009	.99
TalkerGroup1: SegmentType2	.07	.13	.53	.6
TalkerGroup2: SegmentType2	.02	.13	.18	.86
TalkerGroup1: SegmentType3	-.2	.19	-1.07	.29
TalkerGroup2: SegmentType3	-.33	.18	-1.8	.07
Condition: SegmentType1	.007	.1	.07	.94
Condition: SegmentType2	-.02	.1	-.21	.83
Condition: SegmentType3	.13	.14	.92	.36
TalkerGroup1: Condition: SegmentType1	-.53	.26	-1.98	.047 *
TalkerGroup2: Condition: SegmentType1	.03	.26	.11	.92
TalkerGroup1: Condition: SegmentType2	-.59	.26	-2.23	.026 *
TalkerGroup2: Condition: SegmentType2	-.25	.26	-.96	.34
TalkerGroup1: Condition: SegmentType3	.09	.37	.24	.81
TalkerGroup2: Condition: SegmentType3	-.12	.37	-.32	.75

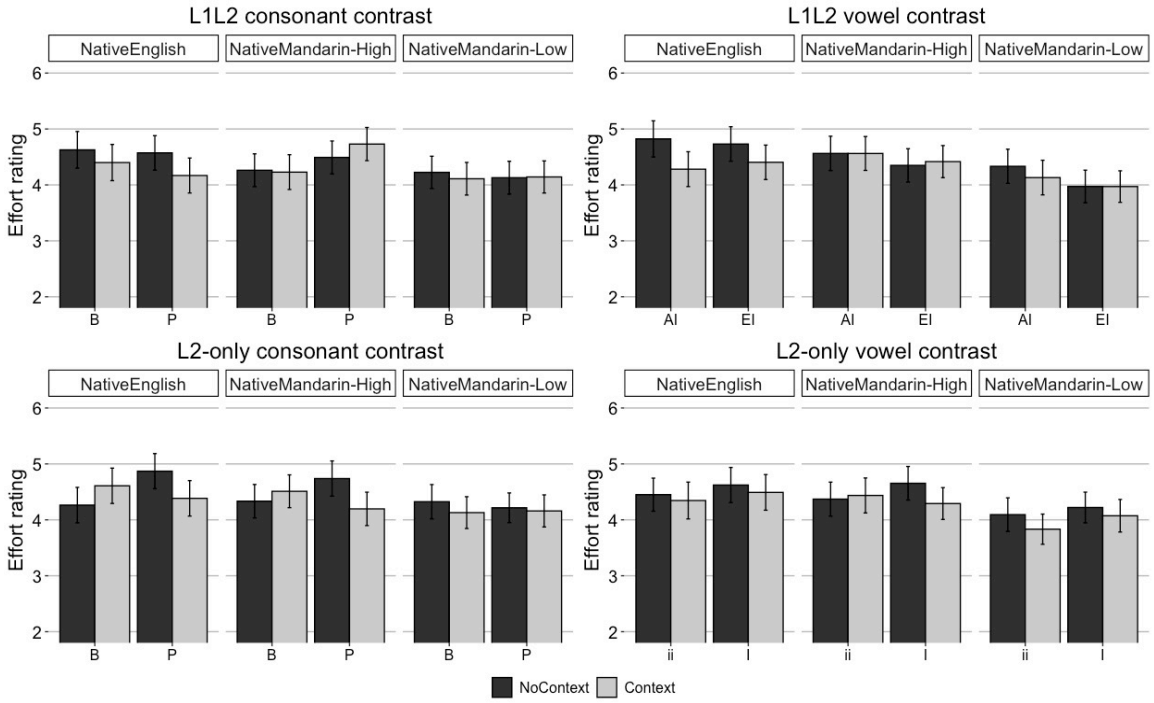

Figure 5.6. Mean talker effort rating (1: the talker is trying extremely hard to speak clearly, 9: the talker is not trying at all to speak clearly) by Segment Type (L1L2 consonant contrast, L1L2 vowel contrast, L2-only consonant contrast, L2-only vowel contrast), Talker Group (Native English, Native Mandarin-High, Native Mandarin-Low), Condition (No Context, Context), and Phoneme (/b/ and /p/ for the L1L2 consonant and L2-only consonant contrasts; /ai/ and /ei/ for the L1L2 vowel contrast; /i/ and /ɪ/ for the L2-only vowel contrast). Error bars represent 95% confidence intervals.

For the model with the L1L2 consonant contrast items (see Table 5.12 for the model summary), there was not a significant effect of Talker Group (Native English vs. Native Mandarin-High; β = .25, t = .5, p = .62, Native English vs. Native Mandarin-High; β = .49, t = .51, p = .34), Condition (No Context vs. Context; β = -.1, t = -1.56, p = .12), or Phoneme (/p/-items vs. /b/-items; β = -.06, t = -.9, p = .37). However, these factors interacted with one another. Particularly, there was an interaction between the Native English vs. Native Mandarin-High group comparison and Condition (β = -.43, t = -2.3, p < .05). This indicates that listeners perceived increased effort for Native English talkers' items produced in Context conditions compared to those produced in No Context

conditions, but this tendency was less strong for Native Mandarin-High talkers' items. The effect of Condition also interacted with Phoneme ($\beta = -.38$, $t = -2.85$, $p < .01$), indicating that listeners perceived increased effort for /b/-items (e.g., *"Click on the beer now"*) produced in Context conditions compared to those produced in No Context conditions, but this tendency was less strong for /p/-items (e.g., *"Click on the peer now"*). There was also an interaction between Phoneme and the Native English vs. Native Mandarin-High group comparison ($\beta = .4$, $t = 2.16$, $p < .05$). This indicates that for Native Mandarin-High talkers' speech, /b/-items were perceived to be produced with increased effort than /p/-items, but this pattern was less strong for Native English talkers' speech. In sum, these results showed that listeners' perceived degree of talker effort for items with the /p/-/b/ initial targets was influenced by the type of production conditions (Context or No Context) and the type of segments (/p/- or /b/-targets) differently for the speech produced by talkers in different groups.

For the model with the L2-only consonant contrast items (see Table 5.13 for the model summary), none of the factors significantly improved the model fit.

For the model with the L1L2 vowel contrast items (see Table 5.14 for the model summary), there was a significant effect of Condition (No Context, Context; $\beta = -.18$, $t = -2.68$, $p < .01$). This indicates that for items with the L1L2 vowel contrast (/ai-/ei/), listeners perceived increased effort for items produced in Context conditions compared to those produced in No Context conditions. This effect of Condition was larger for Native English talkers' speech than for Native Mandarin-High talkers' speech (Condition x Native English vs. Native Mandarin-High group comparison; $\beta = -.49$, $t = -2.52$, $p < .05$). A post-hoc Tukey test confirmed that the effect of Condition was significant for the L1L2 vowel

contrast items produced by Native English talkers (β = .43, SE = .12, t.ratio = 3.57, $p <$ .001), but not for those produced by Native Mandarin-High (β = -.002, SE = .12, t.ratio = -.01, $p = .99$) or Native Mandarin-Low talkers (β = .12, SE = .12, t.ratio = 1.03, $p =$ .31). Thus, listeners perceived increased effort for the items produced in Context conditions than for those produced in No Context conditions, but only for the items produced by Native English talkers.

Table 5.12. Summary of the linear mixed-effects regression model for effort ratings for items with the L1L2 consonant contrast.

Effort Model for L1L2 consonant contrast `Rating ~ TalkerGroup * Condition * Phoneme + (1| Talker) + (1| Listener)`				
Fixed Effects	**Estimate**	**S.E.**	**t-val.**	***p*-val.**
(Intercept)	4.27	.18	23.64	
TalkerGroup1 (Native English vs. Native Mandarin-High)	.25	.51	.5	.62
TalkerGroup2 (Native Mandarin-High vs. Native Mandarin-Low)	.49	.51	.97	.34
Condition (No Context vs. Context)	-.1	.07	-1.56	.12
Phoneme (/b/ vs. /p/)	-.06	.07	-.9	.37
TalkerGroup1: Condition	-.43	.19	-2.3	.022 *
TalkerGroup2: Condition	-.05	.18	-.27	.79
TalkerGroup1: Phoneme	.4	.19	2.16	.031 *
TalkerGroup2: Phoneme	-.19	.18	-1.03	.3
Condition: Phoneme	-.38	.13	-2.85	.004 **
TalkerGroup1: Condition: Phoneme	.19	.38	.49	.62
TalkerGroup2: Condition: Phoneme	-.53	.37	-1.43	.15

Table 5.13. Summary of the linear mixed-effects regression model for effort ratings for items with the L2-only consonant contrast.

Effort Model for L2-only consonant contrast `Rating ~ TalkerGroup * Condition * Phoneme + (1| Talker) + (1| Listener)`				
Fixed Effects	**Estimate**	**S.E.**	**t-val.**	***p*-val.**
(Intercept)	4.33	.17	25.27	
TalkerGroup1 (Native English vs. Native Mandarin-High)	.3	.49	.62	.53
TalkerGroup2 (Native Mandarin-High vs. Native Mandarin-Low)	.46	.48	.96	.34
Condition (No Context vs. Context)	-.11	.07	-1.69	.09
Phoneme (/b/ vs. /p/)	-.08	.07	-1.21	.23
TalkerGroup1: Condition	.1	.19	.52	.6
TalkerGroup2: Condition	-.07	.19	-.39	.7
TalkerGroup1: Phoneme	-.22	.19	-1.17	.24
TalkerGroup2: Phoneme	-.18	.19	-.97	.33
Condition: Phoneme	.19	.14	1.43	.15
TalkerGroup1: Condition: Phoneme	.31	.39	.8	.42
TalkerGroup2: Condition: Phoneme	.61	.38	1.6	.11

Table 5.14. Summary of the linear mixed-effects regression model for effort ratings for items with the L1L2 vowel contrast.

Effort Model for L1L2 vowel contrast `Rating ~ TalkerGroup * Condition * Phoneme + (1+ Condition| Talker) + (1| Listener)`				
Fixed Effects	**Estimate**	**S.E.**	**t-val.**	***p*-val.**
(Intercept)	4.32	.18	24.65	
TalkerGroup1 (Native English vs. Native Mandarin-High)	.4	.5	.8	.43
TalkerGroup2 (Native Mandarin-High vs. Native Mandarin-Low)	.65	.49	1.33	.19
Condition (No Context vs. Context)	-.18	.07	-2.68	.009 **
Phoneme (/ai/ vs. /ei/)	.12	.07	1.76	.08
TalkerGroup1: Condition	-.49	.2	-2.52	.014 *
TalkerGroup2: Condition	-.12	.2	-.63	.53
TalkerGroup1: Phoneme	-.27	.19	-1.43	.15
TalkerGroup2: Phoneme	-.1	.19	-.54	.59
Condition: Phoneme	.12	.14	.91	.37
TalkerGroup1: Condition: Phoneme	.09	.39	.24	.81
TalkerGroup2: Condition: Phoneme	.62	.39	1.61	.11

For the model with the L2-only vowel contrast items (see Table 5.15 for the model summary), there was a significant effect of Condition (No Context vs. Context; β = -.16, t = -2.42, $p < .05$). This indicates that listeners perceived increased effort for the L2-only vowel contrast items produced in Context conditions compared to those produced in No Context conditions. There was also a significant effect of Phoneme (/i/-items vs. /ɪ/-items; β = -.13, t = -2.0, $p < .05$). This indicates that listeners perceived that talkers produced /i/-items (e.g., *"Click on the seek now"*) with more effort than /ɪ/-items (e.g., *"Click on the sick now"*). These effects of Condition and Phoneme did not interact with other factors in the model. Thus, for the items with the /i/-/ɪ/ targets, the production condition (Context, No Context) and the type of segment (/i/- or /ɪ/-targets) independently impacted listeners' perception of talker effort.

Table 5.15. Summary of the linear mixed-effects regression model for effort ratings for items with the L2-only vowel contrast.

Effort Model for L2-only vowel contrast `Rating ~ TalkerGroup * Condition * Phoneme + (1| Talker) + (1| Listener)`				
Fixed Effects	**Estimate**	**S.E.**	**t-val.**	***p*-val.**
(Intercept)	4.27	.18	24.2	
TalkerGroup1 (Native English vs. Native Mandarin-High)	.32	.5	.64	.52
TalkerGroup2 (Native Mandarin-High vs. Native Mandarin-Low)	.64	.5	1.29	.2
Condition (No Context vs. Context)	-.16	.07	-2.42	.016 *
Phoneme (/i/ vs. /ɪ/)	-.13	.06	-2.0	.045 *
TalkerGroup1: Condition	.08	.19	.42	.67
TalkerGroup2: Condition	.11	.18	.62	.54
TalkerGroup1: Phoneme	-.06	.19	-.31	.76
TalkerGroup2: Phoneme	.08	.18	.45	.65
Condition: Phoneme	-.15	.13	-1.12	.26
TalkerGroup1: Condition: Phoneme	-.42	.38	-1.12	.26
TalkerGroup2: Condition: Phoneme	-.27	.37	-.72	.47

5.3.3.2. Summary of the main findings and discussion

Overall, these results have demonstrated that perceived degree of talker effort did not significantly differ among different talker groups' speech (native English, higher-proficiency and lower-proficiency non-native talkers). However, there were some differences specific to individual segments. For example, for items with the L2-only vowel contrast, /i/-items (e.g., *"Click on the seek now"*) were perceived to be produced with more effort than /ɪ/-items (e.g., *"Click on the sick now"*). This suggests that a possible factor that could affect listeners' perception of talker effort is duration of vowel segments. That is, the duration of /i/ was consistently longer than that of /ɪ/ across different conditions and across different talker groups' productions (see Chapter 4), and this may have contributed to the increased effort perceived with /i/-items than with /ɪ/-items. This potential effect of duration on perceived degree of talker effort may also be applicable for listeners' perception of native English talkers' items. That is, listeners perceived increased effort for items produced in Context conditions than those produced in No Context conditions, and this pattern was manifested the most clearly in the perception of native English talkers' items with the L1L2 consonant contrast (/p/-/b/ in word-initial position) and the L1L2 vowel contrast (/ai/-/ei/), but much less clear with higher- and lower-proficiency non-native talkers' items. It is possible that acoustic characteristics of these items influenced the difference in perceived degree of effort. That is, native English talkers' phrase durations were longer in Context conditions than in No Context conditions, but this difference in phrase durations was smaller in non-native talkers' productions (see Chapter 4); this difference may have contributed to the difference in perceived degree of effort between the two conditions for native English talkers' speech but less so for non-native talkers' speech.

Further, given that there was a tendency for native English talkers to increase intensity of the target words from No Context to Context conditions (see Chapter 4), listeners may also have been sensitive to some changes in acoustic properties that were associated with increased intensity (as intensity was normalized for the items used in the perception experiment). Which acoustic factor contributes to listeners' perception of talker effort, for simple speech materials (e.g., *"Click on the ___ now"*) as well as for speech materials involving more variable lexical items and complex syntactic structure, may benefit from a further investigation examining a wider range of acoustic properties.

5.4. Conclusion

The present study examined perceptual consequences of contextually-relevant speech enhancements produced by native English talkers and non-native English talkers of different proficiency levels. Talkers' speech enhancements were produced in a communication task where listeners' needs for enhanced intelligibility for particular sound contrasts (e.g., *peer* vs. *beer*) were signaled in the communication context. We examined whether native English listeners benefited from these contextually-relevant enhancements in terms of accuracy of target word identification (e.g., *Click on the TARGET now*), perceived degree of comprehensibility (how easy it is to understand the speech), and perceived degree of talker effort (how hard the talker is trying to speak clearly). Listeners' responses in these tasks were influenced by combinations of different factors, such as talkers' target language experience (i.e., native vs. non-native talkers; higher- vs. lower-proficiency of non-native talkers), the type of target sound contrast (i.e., whether or not the target English contrast exists in non-native talkers' native language, Mandarin), and

production conditions (i.e., whether or not the target items were produced with a minimal pair neighbor present in the context, such as *peer* for *beer*).

The results from different perception tasks revealed that some aspects of listeners' perception in one task was related to another task. For example, listeners' identification accuracy was higher for native English talkers' items compared to those of higher-proficiency talkers; accuracy was also higher for higher-proficiency talkers' items than those of lower-proficiency talkers. Such effects of talkers' target language experience on listeners' perception were also present for listeners' perceived degree of comprehensibility; listeners perceived native English talkers' items and higher-proficiency talkers' items to be easier to understand that those of lower-proficiency talkers. This suggests that talkers' target language experience generally impacted how well native listeners understood their speech, as well as how easy the listeners perceived the speech to be.

However, the results also revealed some gaps among different perception tasks, suggesting that listeners' identification accuracy and subjective evaluations of the speech may have been influenced by different aspects of talkers' productions. For example, lower accuracy in listeners' target word identification did not necessarily correspond to those items being perceived as more difficult to understand (e.g., see the results for items with the L2-only consonant contrast). Further, talkers' target language experience impacted listeners' identification accuracy and perceived degree of comprehensibility, though it did not affect perceived degree of talker effort. Given these results, it is possible that listeners' target word identification accuracy was more directly impacted by the characteristics of the target words themselves, rather than the characteristics of the whole phrases (e.g., *"Click on the ___ now"*), whereas their subjective evaluations of comprehensibility and talker effort

may have been affected by the perception of the whole phrases. Specifically, it is possible that listeners' subjective perception of comprehensibility was affected more strongly by global characteristics of the speech (e.g., foreign accentedness of the speech, speaking rate) than by the ease of identifying the target words. As for the perception of talker effort, because the production materials were simple phrases with the only difference being the target words (i.e., *"Click on the ___ now"*), difficulty of producing these materials may not have differed significantly for talkers of different target language proficiency levels, resulting in little difference in perceived degree of talker effort among different talker groups' productions. These results provide insight into how speech enhancements produced by talkers of different linguistic backgrounds in a communicative context translate to listeners' perception. A future investigation directly comparing acoustic characteristics of productions and different types of listeners' perception may help us better understand which acoustic features are responsible for improving different types of perceptual evaluations, as well as whether the relationship between acoustic features and perception differ for the talkers and listeners of different linguistic backgrounds.

CHAPTER IV: CONCLUSION

This dissertation sought to understand how talkers of different linguistic backgrounds implement goal-oriented phonetic modifications, and how these modifications impact listeners' perception. Particularly, I examined acoustic characteristics of speech enhancements produced by native English talkers and non-native English talkers of higher- and lower-proficiency. Further, I explored perceptual consequences of these talkers' speech enhancements for native English listeners. In this chapter, I summarize the major findings of each of the four studies, and state novel contributions of the current work. Further, I consider the implications of these findings for understanding the talker- and listener-oriented factors in the effectiveness of speech enhancements, as well as for practices in second language classrooms, and discuss directions for future work.

6.1. Summary of the current research

6.1.1. Main findings of the four studies

The first two studies examined production and perception of clear speech enhancements. In Chapter 2, we observed that acoustic-phonetic characteristics of clear speech enhancements differed for talkers with differing levels of experience with the target language. Specifically, when native and non-native talkers were asked to read English sentences based on explicit instructions to speak clearly, they generally decreased speaking rate, increased F0 range, mean F0, mean intensity, and vowel space in the clear-speaking style as compared to the baseline plain-speaking style. However, the degree of plain-to-clear speech modifications was much smaller for lower-proficiency non-native talkers'

productions compared to those of higher-proficiency talkers and native talkers. Further, the size of acoustic-phonetic modifications produced by higher-proficiency talkers was comparable to that of native talkers.

This difference in the size of clear speech enhancements was reflected in the size of intelligibility improvement as well. Particularly, Chapter 3 demonstrated that for native English listeners, native talkers' and higher-proficiency non-native talkers' clear speech enhancements resulted in a larger intelligibility improvement compared to those of lower-proficiency non-native talkers. Furthermore, across different talker groups (i.e., native English, higher- & lower-proficiency non-native talkers), clear speech enhancements improved listeners' subjective evaluation of talker effort (i.e., how hard the talker is trying to speak clearly), though not evaluation of comprehensibility (i.e., how easy the speech is to understand).

The last two studies examined production and perception of contextually-relevant speech enhancements. In Chapter 4, we demonstrated that the quality of native and non-native talkers' speech enhancements in a simulated communication task was impacted by talkers' target language experience differently depending on the focus of the enhancements. Specifically, when the listener's communicative needs for enhanced intelligibility for particular English contrasts (e.g., *peer* vs. *beer*) were signaled implicitly in the context, native English and non-native talkers made global enhancements similarly (i.e., modifications made to the characteristics of the entire phrases and target words as in *"Click on the ___ now"*); the global modifications included speaking with longer overall phrase durations and higher intensity of the target words. However, particularly for non-native talkers, their ability to enhance a specific sound contrast differed depending on their

familiarity with the target English contrast from their native language experience (Mandarin), as well as their English proficiency level. Specifically, native talkers and non-native talkers of differing proficiency employed similar strategies to enhance the English consonant contrast (/p/-/b/ word-initially as in *peer* vs. *beer*) and vowel contrast (/ai/-/ei/ as in *light* vs. *late*) that also exist in Mandarin. However, native English and higher-proficiency non-native talkers were better able than lower-proficiency talkers to modify acoustic characteristics of an English consonant contrast that does not exist in Mandarin (/p/-/b/ word-finally as in *cap* vs. *cab*). Non-native talkers' acoustic modification behavior was also influenced by the type of acoustic features examined (e.g., manipulation of temporal feature or spectral feature for an English contrast /i/ vs. /ɪ/ as in *seek* vs. *sick*). This suggests that talkers' ability to make acoustic-phonetic enhancements in a contextually-relevant way is influenced by the combination of target language experience and the type of the acoustic modifications required to enhance particular features of the speech.

Chapter 5 examined perceptual consequences of these contextually-relevant speech enhancements. Listeners' identification accuracy of the target words (as in *"Click on the TARGET now"*) was generally influenced by talkers' target language experience (i.e., identification accuracy was the highest for native English talkers' items > higher-proficiency talkers' items > lower-proficiency talkers' items). Furthermore, especially for non-native talkers' productions of target words containing unfamiliar English contrasts (e.g., word-final /p/-/b/ contrast), listeners' identification accuracy was impacted by the segmental characteristics of the target contrasts (e.g., non-native talkers' tendency to devoice word-final /b/, higher-proficiency non-native talkers' effort to enhance the word-final /p/-/b/ contrast by increasing voicing proportions of /b/). However, listeners'

subjective evaluations of the speech did not necessarily correspond to the patterns of identification accuracy. Particularly, perceived degree of comprehensibility and talker effort seemed to be less directly impacted by the segmental characteristics of the target words (as compared to identification accuracy), but were more strongly associated with the overall characteristics of the speech (e.g., accentedness or fluency characteristics of the speech). These results suggest that native listeners focus on different aspects of native and non-native speech depending on how they are asked to evaluate the speech.

6.1.2. Novel contributions of the current research

Taken together, the current work provides novel contributions that inform production and perception of speech enhancements. In terms of production, we demonstrated that talkers' ability to implement goal-oriented acoustic-phonetic modifications was generally impacted by their target language experience; native talkers and higher-proficiency talkers were better able to modify characteristics of their speech and enhance intelligibility for listeners as compared to lower-proficiency talkers. This suggests that lower-proficiency talkers have not acquired the range of stylistic variations as much as higher-proficiency talkers and native talkers have. The results also revealed that a small range of stylistic variations in lower-proficiency talkers' productions partly originated from their difficulty implementing the plain-speaking style rather than the clear-speaking style. It is possible that producing second language speech was generally challenging for lower-proficiency talkers, and they were in a "clear speech" mode at all times.

This provides support for the argument that second language learners' ability to use phonetic specifications of the target language sound system is influenced by their

proficiency (Bohn & Flege, 1992; Fabra & Romero, 2012; Nip & Blumenfeld, 2015; Shea & Curtin, 2011). Specifically, as second language learners' proficiency develops, they are able to implement a wider range of stylistic variations in acoustic-phonetic characteristics of their speech. This proficiency-related improvement in the ability to implement larger speech enhancements for more experienced talkers can originate from their increased knowledge of how to manipulate specific acoustic features in the target language, as well as from less effort required to produce second language speech in general (i.e., as speech production becomes less effortful for more experienced talkers, there is more room to enhance characteristics of the speech from their baseline, plain speech). Together, these results suggest that knowing how to manipulate acoustic-phonetic properties of the speech to *increase* intelligibility, in addition to being able to produce intelligible speech in general, is part of what characterizes talkers' language proficiency. However, based on the findings that lower-proficiency non-native talkers' ability to make acoustic modifications did not differ from that of higher-proficiency and native talkers in some aspects (e.g., enhancing the English word-initial /p/-/b/ contrast: Chapter 4), we also suggest that the effect of target language experience on speech enhancement behavior depends on *what* talkers are trying to enhance and what acoustic feature or speech motor control is involved in enhancing that feature.

These results have implications for how H&H theory (Lindblom, 1990) could be applied for speech production of talkers of limited target language proficiency. Specifically, the current results suggest that the range of acoustic-phonetic variations between hypo- and hyper-speech is generally more limited for lower-proficiency non-native talkers' speech (as compared to higher-proficiency and native talkers' speech), and

this limitation could originate from the difficulty implementing hyper-speech (i.e., knowing how to implement acoustic modifications to maximize perceptual discriminability) as well as hypo-speech (i.e., minimizing the articulatory effort). However, lower-proficiency talkers do implement acoustic modifications to a similar extent that higher-proficiency and native talkers do in some cases (e.g., for a non-native contrast that they are familiar with from their native language experience). Thus, the way talker-oriented force (e.g., economy of effort) and listener-oriented force (i.e., the need for sufficient perceptual discriminability) impact productions differ for talkers of differing levels of target language experience as well as for the target of acoustic modifications.

In terms of perception of speech enhancements, the current work provides a novel contribution by measuring listeners' perceptual benefits in different ways (e.g., measuring listeners' subjective evaluations of the speech in addition to measuring listeners' understanding of the speech), and by demonstrating a dissociation among these measures. That is, perceptual benefits resulting from speech enhancements differed depending on how listeners were asked to evaluate the speech, such that native talkers' and higher-proficiency talkers' clear speech enhancements that resulted in an intelligibility improvement did not improve perceived comprehensibility (i.e., how easy to understand the speech). However, those clear speech enhancements robustly improved listeners' perceived degree of talker effort (i.e., how hard the talker is trying to speak clearly); the robust improvement in perceived degree of talker effort was also observed in perception of lower-proficiency talkers' clear speech enhancements. We showed a similar dissociation in perception of contextually-relevant speech enhancements.

These results contribute to the lines of research suggesting that acoustic features

that impact intelligibility of non-native speech are at least partially independent from those that impact other subjective measures of the speech (e.g., foreign accentedness, comprehensibility; Derwing & Munro, 2009; Munro & Derwing, 1995a; Smiljanić & Bradlow, 2011). Furthermore, as perception of speech enhancements in one measure (i.e., intelligibility) does not necessarily correspond to another (e.g., perceived degree of comprehensibility or talker effort), we suggest that the effects of certain acoustic modifications may be different for different aspects of perceptual benefits. For example, slowing down the speech or increasing the pitch range may contribute to improved intelligibility and/or increased degree of perceived talker effort, but it may not for perceived comprehensibility. Further, the acoustic modifications not examined in the current study (e.g., degree of coarticulation, spectral balance of the voice, and frequency of stop-burst releases: Bradlow, 2002; Ferguson & Kewley-Port, 2007; Hazan et al., 2018; Scarborough & Zellou, 2013) may also impact perception differently depending on how listeners are asked to evaluate the speech. Thus, future work may investigate how different types of acoustic modifications contribute to listeners' perception similarly or differently depending on the type of perceptual measure. This relationship between acoustic characteristics of speech enhancements and perceptual benefits should be examined in relation with how the presence of a non-native accent may or may not impact this relationship. That is, given previous results demonstrating that certain properties of speech could influence listeners' perception differently in native speech vs. non-native speech (e.g., pause, grammatical error: Bosker, Quené, Sanders, & De Jong, 2014; Hanulíková, Van Alphen, Van Goch, & Weber, 2012), it is possible that a particular type of acoustic modification (e.g., decrease in speaking rate) impacts listeners' perception differently for

speech enhancements produced in more accented speech vs. less accented speech. Examining these questions would highlight the multi-faceted nature of listeners' perception of acoustic-phonetic modifications employed by talkers of different linguistic backgrounds.

Further, the dissociation between the two subjective measures of listeners' perceptual benefits (i.e., listeners' perception of comprehensibility and talker effort) for *both* native and non-native speech could entail that there is some inherent relationship between the two measures. That is, the two measures may naturally not pattern together; the speech perceived to be produced with increased effort may be perceived to be less easy to understand. However, examining perception of speech enhancements produced in spontaneous speech with a conversation partner may show a different relationship between these two aspects of perception. Particularly, previous results suggest that acoustic characteristics of speech enhancements differ when the enhancements are elicited in spontaneous conversation as compared to those elicited in read speech with explicit instructions to speak clearly (e.g., Hazan & Baker, 2011), and that listeners benefit more from real-listener directed speech than imagined-listener-directed clear speech on some measures (e.g., faster lexical decisions: Scarborough & Zellou, 2013). Listeners also understand native English speech that was produced in a task-oriented conversation with a non-native English partner better than the speech produced in the same task with a native English partner (Lee & Baese-Berk, 2020). Such perceptual benefits associated with speech enhancements elicited in spontaneous conversation could extend to subjective aspects of perception (e.g., perceived talker effort and comprehensibility). Further, it is possible that acoustic properties of speech enhancements elicited in spontaneous conversation are more natural-sounding than those elicited in simulated clear speech, and this could be manifested

in similar patterns of perceptual benefits in subjective evaluations of the speech (e.g., the speech perceived to be produced with increased effort may also be perceived to be easier to understand).

6.2. Future directions

6.2.1. Production of speech enhancements

The current research extensively examined native and non-native talkers' speech enhancements. By demonstrating that differences in speech enhancement strategies emerge at the group level (e.g., native English talkers and higher-proficiency talkers make larger plain-to-clear speech modifications than lower-proficiency non-native talkers), the findings provide a first step towards characterizing how goal-oriented phonetic modifications are implemented by the talkers with differing levels of target language experience. However, the current results also show individual variability in acoustic modifications and intelligibility improvement, which may suggest the need to examine goal-oriented adaptation at the individual level (e.g., within each native English, higher-proficiency, lower-proficiency group) in relation with other types of production variability. Specifically, talkers' ability to induce conditioned variations in their productions (e.g., modifying acoustic characteristics of their speech to make it more understandable for listeners) could be associated with the degree of within-talker phonetic variability that are not necessarily goal-oriented, or not intended by the talker. For example, a larger degree of within-talker variability in production can be associated with less stable coordination/speech motor control (e.g., variability of lip and jaw movements: Nip & Blumenfeld, 2015; Smith & Zelaznik, 2004; variability in vowel durations; Redford & Oh, 2017). It is possible that,

especially for lower-proficiency non-native talkers, a higher degree of such 'unintended' within-talker variability could make it more difficult to implement controlled variations in their speech. The current results showed a tendency that lower-proficiency talkers' within-talker variability was larger than that of higher-proficiency talkers and native talkers especially in the sentence productions of plain speech, in features such as articulation rate, F0 range, mean intensity (Chapter 2). Higher-proficiency talkers' within-talker variability in these features seemed to be similar to that of native talkers, which is in line with other studies showing a comparable degree of within-talker variability in productions of proficient non-native talkers and native talkers (Redford & Oh, 2017; Vaughn et al., 2019). Given these patterns, the size of speech enhancements (between two different modes of speech, such as plain vs. clear speech in Chapter 2; productions in No Context vs. Context conditions in Chapter 4) may have been influenced by the degree of within-talker variability inherent in the talkers' productions, such that larger within-talker variability in each mode of speech obscured the difference between the two modes of speech.

This type of within-talker variability, possibly signaling lower proficiency or less stable speech motor coordination, needs to be examined in relation with other types of variability associated with native-like control of speech production (e.g., Idemaru, Wei, & Gubbins, 2019; Vaughn et al., 2019). For example, phonetic realizations of Japanese voiced stop consonants were more variable for native Japanese talkers' productions than for Japanese learners' productions (e.g., stops realized as canonical stops, approximates, and fricatives: Vaughn et al., 2019). Further, higher vowel duration variability in native Japanese talkers' productions (as compared to learners' productions) was associated with native listeners' perception of less strong foreign accent (Idemaru et al., 2019). Native

English talkers also showed larger variability in word durations, partly caused by grater function word reduction, compared to non-native English talkers' productions (Baker et al., 2011). These studies suggest that a larger degree of within-talker variability in some aspects of productions is associated with greater control, and that listeners consider such variations as an indication of language proficiency.

Taken together, talkers' ability to implement goal-oriented, conditioned variations (e.g., plain speech vs. clear speech) can be characterized in terms of, as Durham (2014) describes, "learning-related variation" (e.g., variation associated with the developing state of non-native language use) and "target-based variation" (e.g., patterns inherent in the target language system). Thus, how these types of variations are manifested in individual talkers' productions in different styles/modes of speech (e.g., plain and clear speech) may be an important step towards understanding how developmental factors shape talkers' goal-oriented phonetic modifications. Specifically, future work may examine the relationship between within-mode variation (e.g., item-by-item speaking rate variation in plain speech) and between-mode variation (e.g., speaking rate difference between plain and clear speech). For example, do talkers who show a larger degree of within-mode variation of a particular feature also show larger between-mode modifications in that feature, and does this pattern differ for different types of acoustic features (e.g., speaking rate, vowel duration, VOT duration), or for talkers with differing levels of experience with the target language (e.g., non-native talkers of higher- and lower-proficiency; adults and children)? Asking these questions may help us better understand talkers' speech accommodation behavior from a developmental perspective, which could inform a broader model of speech production.

6.2.2. Perception of speech enhancements

Another important avenue for future inquiry concerns how perceptual benefits of native and non-native talkers' speech enhancements may differ for listeners of different characteristics. The current work investigated native English listeners' perception of the speech enhancements made by native English talkers and non-native talkers of higher- and lower-proficiency. However, as successful speech communication relies on contributions from both a talker and a listener (Clark & Wilks-Gibbs, 1986), it is critical to assess the effectiveness of speech enhancements in terms of not only talker-related factors but also listener-related factors. For example, given that intelligibility of native and non-native speech is influenced by the language backgrounds of talkers and listeners (e.g., Bent & Bradlow, 2003), it is possible that listeners of different language backgrounds would benefit from non-native (native Mandarin) talkers' speech enhancements differently than native English listeners did in the current study. Specifically, non-native (native Mandarin) talker' speech enhancements may benefit native Mandarin listeners more than native English listeners, given that a shared language background between a talker and a listener results in intelligibility benefit (i.e., matched interlanguage speech intelligibility benefit: Bent & Bradlow, 2003; Hayes-Harb et al., 2008; Imai et al., 2005; Munro et al., 2006). Furthermore, as the current results showed that the speech enhancement patterns did not often differ between the speech of higher-proficiency talkers and native English talkers (see Chapter 2), it is possible that the intelligibility benefit resulting from the shared native language background (e.g., for native Mandarin listeners perceiving native Mandarin talkers' English speech) will be larger for the perception of lower-proficiency talkers'

enhancements than for the perception of higher-proficiency talkers' enhancements.

The shared language experience between a talker and a listener may also impact native English listeners' ability to take advantage of non-native talkers' speech enhancements. Previous work has shown that, while processing foreign-accented speech can be generally more difficult and more effortful compared to listening to familiar native-accented speech (e.g., Adank, Evans, Stuart-Smith, & Scott, 2009; Rogers et al., 2004; Van Engen & Peelle, 2014), listeners can rapidly adapt and become able to understand accented speech better after a short period of training (e.g., Baese-Berk, Bradlow, & Wright, 2013; Bradlow & Bent, 2008). Given these results, it is possible that native listeners' experience or familiarity with a particular accented talker or a type of accent (e.g., Mandarin-accented English) may help them not only better understand the speech in general, but also take advantage of the speech enhancements made by these talkers. Furthermore, perceptual consequences of non-native speech enhancements for native listeners may also vary depending on the listeners' social attitude or expectation. Particularly, previous studies have demonstrated that listeners' cultural expectations and attitudes towards certain social groups or accented speech impact their language comprehension (e.g., Hay & Drager, 2010; Hay, Warren, & Dragger, 2006; Kang & Rubin, 2009) as well as subjective evaluations of the speech (e.g., perceived degree of comprehensibility: Sheppard et al., 2017). Thus, native listeners' attitude towards accented speech and how hard they are willing to try to understand the speech of an unfamiliar variety could impact multiple aspects of perceptual benefits that they could receive from non-native speech enhancements (e.g., how well they understand the speech, how easy they perceive the speech to understand). Thus, considering listener-related factors, such as their language experience

and social attitude, may reveal what types of listeners are more or less equipped to perceptually benefit from speech enhancements produced by talkers of a particular linguistic background. Exploring the role of both talker- and listener-related factors in benefits of speech enhancements would shed light on how talkers and listeners of diverse linguistic backgrounds could overcome communication barriers together.

6.2.3. Implications for second language instruction

We observed that, while higher-proficiency talkers were generally better than lower-proficiency talkers at implementing acoustic-phonetic modifications to improve intelligibility, lower-proficiency did show significant enhancements in some aspects of their productions. For example, lower-proficiency non-native talkers' ability to enhance a non-native contrast that they are familiar with from their native language did not differ from that of higher-proficiency talkers or native talkers (Chapter 4). These results suggest that talkers with limited language proficiency are capable of implementing intelligibility-enhancing acoustic-phonetic modifications given sufficient practice. Such results provide support for the growing body of work arguing for the importance of pronunciation training incorporated in the second language instruction (e.g., Derwing et al., 1997, 1998; Derwing & Munro, 2009; Derwing & Rossiter, 2003; Levis, 2005). While Communicative Language Teaching approaches had prioritized a focus on meaning in language instruction and de-emphasized explicit, form-focused pronunciation training (Derwing & Munro, 2009), the importance of pronunciation training has been increasingly recognized and examined in recent studies (e.g., Gonzales-Bueno & Quintana-Lara, 2011; Liu & Fu, 2011; Saito & Lyster, 2012). The current work contributes to these lines of research by empirically

demonstrating that while second language learners' speech is generally less intelligible than native talkers' speech, learners could still improve intelligibility of their speech. We suggest that explicit guidance on learners' strategies to enhance acoustic-phonetic properties of their speech has practical implications for their ability to manage communication in a challenging situation (e.g., when a listener does not understand their speech, when communicating in a noisy environment). Such goal of training learners' pronunciation strategies that improve intelligibility, rather than those aimed to reduce accentedness (as these two aspects have been shown to be partially independent of one another: Munro & Derwing, 1995a; Smiljanić & Bradlow, 2011), is compatible with the argument that pronunciation training should focus on the aspects that make second language communication more successful (Derwing & Munro, 2005; Levis, 2005).

Pronunciation training could be implemented both at the suprasegmental and segmental levels (e.g., Derwing et al., 1998; Kartushina, Hervais-Adelman, Frauenfelder, & Golestani, 2016, Saito & Lyster, 2012; see Thomson & Derwing, 2015 for an extensive review). For example, explaining and practicing stress and rhythm patterns in English (e.g., reduction of function words: Baker et al., 2011) in both slow and fast speech may help implement variations in their speaking rate. Such training could promote learners' awareness of different modes/styles of speech in different communicative contexts (e.g., carrying out casual conversation, speaking clearly for listeners who have difficulty understanding) and help them modify characteristics of their speech accordingly. Further, training on individual segments can be effective even for lower-proficiency learners who are not proficient enough to produce fluent speech with complex structures, as shown in a previous result where explicit training on individual segments significantly improved

pronunciation for beginning-level learners (e.g., reduced category variability: Kartushina et al., 2016). Thus, by using a simple but still communicative task, for example, by having learners work on a map task that involves differentiations of specific non-native sound contrasts (e.g., English /p/-/b/ in word-final position), the training may help them become aware of the production difficulty associated with particular non-naive sounds, and practice their speech motor control to produce those sounds in an intelligible manner. A future investigation may explore the effect of pronunciation training on learners' ability to not only produce intelligible speech in general but also to *increase* intelligibility of their speech. Examining the efficacy of pronunciation training on learners' ability to adjust speech in different contexts via a variety of materials/tasks/research settings (e.g., reading simple sentences, conversing with a partner in lab or in classrooms) would promote evidence- and theory-based approach, with attention to ecological validity of the findings, in second language speech communication research.

6.3. Conclusions

In the four studies of this dissertation, we examined how native English talkers and non-native English talkers of different proficiency levels produced speech enhancements, and how these enhancements impacted native English listeners' perception. Throughout, we demonstrated that talkers' target language experience generally impacted the size of acoustic-phonetic modifications made in speech enhancements. However, especially for non-native talkers, their ability to enhance a specific feature of their speech differed depending on the type of acoustic manipulations involved in the productions. We further demonstrated that perceptual benefits resulting from speech enhancements could

vary depending on how listeners are asked to evaluate the speech.

By examining both acoustic properties and perceptual consequences of speech enhancements, the current work is a unique contribution to the growing body of work on speech adaptation. Specifically, the findings suggest that native and non-native talkers have the flexibility to accommodate listeners' communicative needs, and that this flexibility is shaped by the combination of talkers' linguistic backgrounds and the focus of adaptation. Furthermore, multi-faceted nature of speech perception needs to be accounted for when evaluating perceptual benefits of speech enhancements. Thus, as a whole, this dissertation has provided insights into, and has raised new questions about, the mechanisms for goal-oriented phonetic adaptations as well as perceptual consequences of these adaptations. Going forward, considering both talker-related and listener-related factors in speech enhancements is not only relevant but also critical to building successful communication in the increasingly multi-lingual and multi-cultural society.

APPENDICES

APPENDIX A

List of 30 test and 15 practice BKB sentences (marked as 'Test' and 'Practice') recorded by native and non-native English talkers (Chapter 2). In the test sentences, keywords used for intelligibility scoring (Chapter 3) are underlined. The 10 practice sentences used in the accentedness rating task are marked as 'Practice (A)'.

Type	Sentence	Type	Sentence
Test	The shop closed for lunch.	Practice (A)	The three girls are listening.
Test	Some nice people are coming.	Practice	They washed in cold water.
Test	They met some friends.	Practice (A)	The young people are dancing.
Test	Flowers grow in the garden.	Practice (A)	The ball broke the window.
Test	The train stops at the station.	Practice (A)	The boy forgot his book.
Test	The puppy plays with a ball.	Practice (A)	They had two empty bottles.
Test	Mother cut the birthday cake.	Practice (A)	The coat is on the chair.
Test	He closed his eyes.	Practice (A)	The new road is on the map.
Test	The raincoat is very wet.	Practice (A)	The jug is on the shelf.
Test	She is paying for her bread.	Practice (A)	The girl has a picture book.
Test	Some men shave in the morning.	Practice (A)	The orange was very sweet.
Test	The driver lost his way.	Practice	A friend came for lunch.
Test	The oven door was open.	Practice	They heard a funny noise.
Test	The car is going too fast.	Practice	The floor looked clean.
Test	The silly boy is hiding.	Practice	The bus left early.
Test	The apple pie is baking.		
Test	The sky was very blue.		
Test	People are going home.		
Test	She is calling for her daughter.		
Test	He is skating with his friend.		
Test	They painted the wall.		
Test	The dog is eating some meat.		
Test	A boy broke the fence.		
Test	The snow is on the roof.		
Test	The bath water was warm.		
Test	He is reaching for his spoon.		
Test	The boy got into trouble.		
Test	He paid his bill.		
Test	Mother made some curtains.		
Test	The man tied his shoes.		

APPENDIX B

List of targets used in the context-production task (Chapter 4). Targets consisted of four types of English contrasts: L1L2 consonant contrast (/p/-/b/ in word-initial position), L1L2 vowel contrast (ai/-/ei/), L2-only consonant contrast (/p/-/b/ in word-final position), L2-only vowel contrast (/i/-/ɪ/). Half of the pairs were presented in Context conditions, and the other half were presented in No Context conditions.

Condition	L1L2 consonant targets		L2-only consonant targets	
Context	pad	bad	cap	cab
Context	punch	bunch	slop	slob
Context	pay	bay	rope	robe
Context	pea	bee	mop	mob
Context	poll	ball	tap	tab
No Context	path	bath	lap	lab
No Context	park	bark	cop	cob
No Context	pace	base	sop	sob
No Context	peer	beer	slap	slab
No Context	pack	back	crap	crab
Condition	L1L2 vowel targets		L2-only vowel targets	
Context	light	late	sick	seek
Context	sigh	say	fit	feet
Context	rise	raise	sit	seat
Context	height	hate	lick	leak
Context	die	day	list	least
No Context	like	lake	mill	meal
No Context	high	hay	chick	cheek
No Context	right	rate	hit	heat
No Context	fight	fate	mitt	meet
No Context	lie	lay	wit	wheat

REFERENCED CITED

Adank, P., Evans, B. G., Stuart-Smith, J., & Scott, S. K. (2009). Comprehension of familiar and unfamiliar native accents under adverse listening conditions. *Journal of Experimental Psychology: Human Perception and Performance*, *35*(2), 520-529.

Adank, P., Stewart, A. J., Connell, L., & Wood, J. (2013). Accent imitation positively affects language attitudes. *Frontiers in Psychology*, *4*, 1-10.

Babel, M. (2012). Evidence for phonetic and social selectivity in spontaneous phonetic imitation. *Journal of Phonetics, 40*, 177–189.

Baese-Berk, M., Bradlow, A. R., & Wright, B. A. (2013). Accent-independent adaptation to foreign accented speech. *Journal of the Acoustical Society of America*, *133*(3), EL174–EL180.

Baese-Berk, M., & Goldrick, M. (2009). Mechanisms of interaction in speech production. *Language and Cognitive Processes*, *24*(4), 527-554.

Bakeman, R. (2005). Recommended effect size statistics for repeated measures designs. *Behavior Research Methods*, *37*(3), 379-384.

Baker, R. E., Baese-Berk, M., Bonnasse-Gahot, L., Kim, M., Van Engen, K. J., & Bradlow, A. R. (2011). Word durations in non-native English. *Journal of Phonetics*, *39*(1), 1-17.

Barlow, J. A. (2014). Age of acquisition and allophony in Spanish-English bilinguals. *Frontiers in Psychology*, *5*, 1-14.

Barr, D. J., Levy, R., Scheepers, C., & Tily, H. J. (2013). Random effects structure for confirmatory hypothesis testing: Keep it maximal. *Journal of Memory and Language*, *68*(3), 255-278.

Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software, 67*(1), 1-48.

Bayard, D., Weatherall, A., Gallois, C., & Pittam, J. (2001). Pax Americana? Accent attitudinal evaluations in New Zealand, Australia and America. *Journal of Sociolinguistics, 5*(1), 22–49.

Bamford, J., & Wilson, I. (1979). Methodological considerations and practical aspects of the BKB sentence lists. In J. Bench & J. Bamford (Eds.), *Speech-hearing tests and the spoken language of hearing-impaired children* (pp. 148–187). London: Academic Press.

Bent, T., & Bradlow, A. R. (2003). The interlanguage speech intelligibility benefit. *Journal of the Acoustical Society of America*, *114*(3), 1600–1610.

Blicher, D. L., Diehl, R. L., & Cohen, L. B. (1990). Effects of syllable duration on the perception of the Mandarin Tone 2/Tone 3 distinction: Evidence of auditory enhancement. *Journal of Phonetics*, *18*(1), 37-49.

Boersma, P., & Weenink, D. (2001). Praat, a system for doing phonetics by computer. *Glot International* 5:9/10, 341-345.

Bohn, O. S. (1995). Cross-language perception in adults: First language transfer doesn't tell it all. In W. Strange (Ed.), *Speech perception and linguistic experience: Issues in cross-language research* (pp. 379–410). Timonium, MD: York Press.

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

Bosker, H. R., Quené, H., Sanders, T., & De Jong, N. H. (2014). Native 'um's elicit prediction of low-frequency referents, but non-native 'um's do not. *Journal of Memory and Language*, *75*, 104-116.

Bradlow, A.R. (2002). Confluent talker- and listener-related forces in clear speech production. In C. Gussenhoven, & N. Warner (Eds.), *Laboratory phonology VII (phonology and phonetics)* (pp. 241–273). Berlin: Mouton de Gruyter.

Bradlow, A. R., & Alexander, J. (2007). Semantic-contextual and acoustic-phonetic enhancements for English sentence-in-noise recognition by native and non-native listeners. *Journal of the Acoustical Society of America, 121*(4), 2339–2349.

Bradlow, A. R., & Bent, T. (2002). The clear speech effect for non-native listeners. *Journal of the Acoustical Society of America, 112*(1), 272–284.

Bradlow, A. R., & Bent, T. (2008). Perceptual adaptation to non-native speech. *Cognition*, *106*(2), 707-729.

Bradlow, A. R., Kraus, N. & Hayes, E. (2003). Speaking clearly for learning-impaired children: sentence perception in noise. *Journal of Speech, Language, and Hearing Research, 46*, 80–97.

Bradlow, A. R., Torretta, G. M., & Pisoni, D. B. (1996). Intelligibility of normal speech I: Global and fine-grained acoustic-phonetic talker characteristics. *Speech Communication*, *20*(3), 255-272.

Branigan, H. P., Pickering, M. J., & Cleland, A. A. (2000). Syntactic coordination in dialogue. *Cognition, 75*, B13–B25.

Brennan, S. E., & Clark, H. H. (1996). Conceptual pacts and lexical choice in conversation. *Journal of Experimental Psychology: Learning, Memory and Cognition, 6*, 1482–1493.

Brière, E. J. (1966). An investigation on phonological interference. *Language, 42*, 768-796.

Broselow, E., Chen, S. I., & Wang, C. (1998). The emergence of the unmarked in second language phonology. *Studies in Second Language Acquisition*, *20*(2), 261-280.

Buz, E., Jaeger, T. F., & Tanenhaus, M. K. (2014). Contextual confusability leads to targeted hyperarticulation. In *Proceedings of the Annual Meeting of the Cognitive Science Society* (Vol. 36, No. 36).

Buz, E., Tanenhaus, M. K., & Jaeger, T. F. (2016). Dynamically adapted context-specific hyper-articulation: Feedback from interlocutors affects speakers' subsequent pronunciations. *Journal of Memory and Language*, *89*, 68-86.

Cargile, A. C., Giles, H., Ryan, E. B., & Bradac, J. J. (1994). Language attitudes as a social process: a conceptual model and new directions. *Language and Communication, 14*, 211–236.

Clark, H. H., & Wilkes-Gibbs, D. (1986). Referring as a collaborative process. *Cognition*, *22*(1), 1-39.

Costa, A., & Santesteban, M. (2004). Lexical access in bilingual speech production: Evidence from language switching in highly proficient bilinguals and L2 learners. *Journal of Memory and Language*, *50*(4), 491-511.

Chen, M. (1970). Vowel length variation as a function of the voicing of the consonant environment. *Phonetica*, *22*(3), 129-159.

Chen, Y. (2006). Production of tense-lax contrast by Mandarin speakers of English. *Folia phoniatrica et logopaedica*, *58*(4), 240-249.

Cheng, C. C. (1973). *A synchronic phonology of Mandarin Chinese*, The Hague: Mouton.

Choi, J., Cho, T., Kim, S., Baek, Y., & Jang, J. (2015). Phonetic encoding of coda voicing contrast and its interaction with information structure in L1 and L2 speech. In *Proceedings of the 18th International Congress of Phonetic Sciences*.

Choi, J., Kim, S., & Cho, T. (2016). Phonetic encoding of coda voicing contrast under different focus conditions in L1 vs. L2 English. *Frontiers in Psychology*, *7(*624), 1-17.

Chomsky, N., & Halle, M. (1968). *The sound pattern of English*. New York: Harper and Row.

Coupland, J., & Bishop, H. (2007). Ideologised values for British accents. *Journal of Sociolinguistics, 11,* 74–93.

Crawley, M. J. (2002). *Statistical Computing: An Introduction to Data Analysis Using S-Plus*. London: John Wiley & Sons Ltd.

Declerck, M., & Kormos, J. (2012). The effect of dual task demands and proficiency on second language speech production. *Bilingualism: Language and Cognition*, *15*(4), 782-796.

de Jong, K. (2004). Stress, lexical focus, and segmental focus in English: patterns of variation in vowel duration. *Journal of Phonetics*, *32*(4), 493-516.

de Jong, N. H., Groenhout, R., Schoonen, R., & Hulstijn, Y. H. (2015). Second language fluency: Speaking style or proficiency? Correcting measures of second language fluency for first language behavior. *Applied Psycholinguistics, 36*, 223–243.

Derwing, T. M., Munro, M. J., & Wiebe, G. E. (1997). Pronunciation instruction for 'fossilized' learners: Can it help? *Applied Language Learning*, 8, 217–235.

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

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., & Munro, M. (2005). Second language accent and pronunciation teaching: A research–based approach. *TESOL Quarterly, 39*, 379–397.

Derwing, T. M., & Munro, M. J. (2009). Putting accent in its place: Rethinking obstacles to communication. *Language Teaching*, *42*(4), 476-490.

Duffy, J. R. (2005). *Motor speech disorders: Substrates, differential diagnosis, and management* (2nd ed.). New York, NY: Mosby.

Durham, M. (2014). *The Acquisition of Sociolinguistic Competence in a Lingua Franca Context*. Bristol, U.K.: Multilingual Matters.

Educational Testing Service. (2010). Linking TOEFL iBT scores to IELTS scores—A research report. Princeton, NJ: Educational Testing Service. Retrieved from https://www.ets.org/s/toefl/pdf/linking_toefl_ibt_scores_to_ielts_scores.pdf

Fabra, L. R., & Romero, J. (2012). Native Catalan learners' perception and production of English vowels. *Journal of Phonetics*, *40*(3), 491-508.

Ferguson, S. H. (2004). Talker differences in clear and conversational speech: Vowel intelligibility for normal-hearing listeners. *Journal of the Acoustical Society of America*, *116*(4), 2365-2373.

Ferguson, S. H., & Kewley-Port, D. (2002). Vowel intelligibility in clear and conversational speech for normal-hearing and hearing-impaired listeners. *Journal of the Acoustical Society of America, 112*, 259–71.

Ferguson, S. H. & Kewley-Port, D. (2007). Talker differences in clear and conversational speech: acoustic characteristics of vowels. *Journal of Speech, Language, and Hearing Research, 50*, 1241–55.

Flege, J. E. (1987). Effects of equivalence classification on the production of foreign language speech sounds. In A. James & J. Leather (Eds.), *Sound patterns in second language acquisition* (pp. 9-40). Dordrecht: Foris.

Flege, J. E. (1995). Second language speech learning: theory, findings, and problems. In *Speech perception and linguistic experience: Issues in cross-language research* (pp. 233–277). Timonium, MD: York Press.

Flege, J. E., Bohn, O. S., & Jang, S. (1997). Effects of experience on non-native speakers' production and perception of English vowels. *Journal of Phonetics*, *25*(4), 437-470.

Flege, J. E., Munro, M. J., & Skelton, L. (1992). Production of the word-final English /t/-/d/ contrast by native speakers of English, Mandarin, and Spanish. *Journal of the Acoustical Society of America, 92*, 128–143.

Flege, J. E., Schirru, C., & MacKay, I. R. (2003). Interaction between the native and second language phonetic subsystems. *Speech Communication*, *40*(4), 467-491.

Gagné, J. P., Masterson, V., Munhall, K. G., Bilida, N., & Querengesser, C. (1994). Across talker variability in auditory, visual, and audiovisual speech intelligibility for conversational and clear speech. *Journal-Academy of Rehabilitative Audiology*, *27*, 135-158.

Gagné, J. P., Rochette, A. J., & Charest, M. (2002). Auditory, visual and audiovisual clear speech. *Speech Communication*, *37*(3-4), 213-230.

Gerstman, L. (1968). Classification of self-normalized vowels. *IEEE transactions on audio and electroacoustics*, *16*(1), 78-80.

Gilbert, R. C., Chandrasekaran, B., & Smiljanić, R. (2014). Recognition memory in noise for speech of varying intelligibility. *Journal of the Acoustical Society of America*, *135*(1), 389-399.

Gittleman, S., & Van Engen, K. J. (2018). Effects of noise and talker intelligibility on judgments of accentedness. *Journal of the Acoustical Society of America*, *143*(5), 3138-3145.

Goldinger, S. D. (1996). Words and voices: episodic traces in spoken word identification and recognition memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *22*(5), 1166-1183.

Goldman-Eisler, F. (1968). *Psycholinguistics: Experiments in spontaneous speech*. New York: Academic Press.

Goldrick, M., Vaughn, C., & Murphy, A. (2013). The effects of lexical neighbors on stop consonant articulation. *Journal of the Acoustical Society of America*, *134*(2), EL172-EL177.

Gollan, T. H., Montoya, R. I., Fennema-Notestine, C., & Morris, S. K. (2005). Bilingualism affects picture naming but not picture classification. *Memory & Cognition*, *33*(7), 1220-1234.

Gonzales-Bueno, M., & Quintana-Lara, M. (2011). The teaching of L2 pronunciation through processing instruction. *Applied Language Learning*, *21*, 53-78.

Gottfried, T. L., & Suiter, T. L. (1997). Effect of linguistic experience on the identification of Mandarin Chinese vowels and tones. *Journal of Phonetics*, *25*(2), 207-231.

Granlund, S., Hazan, V., & Baker, R. (2012). An acoustic–phonetic comparison of the clear speaking styles of Finnish–English late bilinguals. *Journal of Phonetics*, *40*(3), 509-520.

Green, D. W. (1998). Mental control of the bilingual lexico-semantic system. *Bilingualism: Language and cognition*, *1*(2), 67-81.

Grondelaers, S., Van Hout, R., & Steegs, M. (2010). Evaluating regional accent variation in Standard Dutch. *Journal of Language and Social Psychology*, *29*(1), 101-116.

Hanulíková, A., Van Alphen, P. M., Van Goch, M. M., & Weber, A. (2012). When one person's mistake is another's standard usage: The effect of foreign accent on syntactic processing. *Journal of Cognitive Neuroscience*, *24*(4), 878-887.

Hanulová, J., Davidson, D. J., & Indefrey, P. (2011). Where does the delay in L2 picture naming come from? Psycholinguistic and neurocognitive evidence on second language word production. *Language and Cognitive Processes*, *26*(7), 902-934.

Hay, J., & Drager, K. (2010). Stuffed toys and speech perception. *Linguistics, 48*, 865–892.

Hay, J., Warren, P., & Drager, K. (2006). Factors influencing speech perception in the context of a merger-in-progress. *Journal of Phonetics*, *34*(4), 458-484.

Hayes-Harb, R., Smith, B., Bent, T., & Bradlow, A. R. (2008). The interlanguage speech intelligibility benefit for native speakers of Mandarin: Production and perception of English word-final voicing contrasts. *Journal of Phonetics, 36*, 664–679.

Hazan, V. & Baker, R. (2011). Acoustic-phonetic characteristics of speech produced with communicative intent to counter adverse listening conditions. *Journal of the Acoustical Society of America, 130*(4), 2139-2152.

Hazan, V., Grynpas, J., & Baker, R. (2012). Is clear speech tailored to counter the effect of specific adverse listening conditions?. *Journal of the Acoustical Society of America*, *132*(5), EL371-EL377.

Hazan, V., Tuomainen, O., Kim, J., Davis, C., Sheffield, B., & Brungart, D. (2018). Clear speech adaptations in spontaneous speech produced by young and older adults. *Journal of the Acoustical Society of America*, *144*(3), 1331-1346.

Hillenbrand, J., Getty, L. A., Clark, M. J., & Wheeler, K. (1995). Acoustic characteristics of American English vowels. *Journal of the Acoustical society of America*, *97*(5), 3099-3111.

Hillenbrand, J. M., Clark, M. J., & Houde, R. A. (2000). Some effects of duration on vowel recognition. *Journal of the Acoustical Society of America, 108*(6), 3013–3022.

Hirata, Y., & Whiton, J. (2005). Effects of speaking rate on the single/geminate stop distinction in Japanese. *Journal of the Acoustical Society of America*, *118*(3), 1647-1660.

Hogan, J. T., & Rozsypal, A. J. (1980). Evaluation of vowel duration as a cue for the voicing distinction in the following word-final consonant. *Journal of the Acoustical Society of America*, *67*, 1764–1771.

House, A. S. (1961). On vowel duration in English. *Journal of the Acoustical Society of America*, *33*(9), 1174-1178.

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

Hustad, K. C., & Weismer, G. (2007). Interventions to improve intelligibility and communicative success for speakers with dysarthria. In G. Weismer (Ed.), *Motor speech disorders* (pp. 217–228). San Diego, CA: Plural Publishing.

Hwang, J., Brennan, S. E., & Huffman, M. K. (2015). Phonetic adaptation in non-native spoken dialogue: Effects of priming and audience design. *Journal of Memory and Language*, *81*, 72-90.

Idemaru, K., & Guion-Anderson, S. (2010). Relational timing in the production and perception of Japanese singleton and geminate stops. *Phonetica*, *67*(1-2), 25-46.

Idemaru, K., Wei, P., & Gubbins, L. (2019). Acoustic sources of accent in second language Japanese speech. *Language and Speech*, *62*(2), 333-357.

Imai, S., Walley, A. C., & Flege, J. E. (2005). Lexical frequency and neighborhood density effects on the recognition of native and Spanish-accented words by native English and Spanish listeners. *Journal of the Acoustical Society of America, 117*, 896–907.

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.

Ivanova, I., & Costa, A. (2008). Does bilingualism hamper lexical access in speech production?. *Acta Psychologica*, *127*(2), 277-288.

Johnson, K. (1997). Speech perception without speaker normalization. In K. Johnson, & J. W. Mullennix (Eds.), *Talker variability in speech processing,* (pp. 145–165). San Diego, CA: Academic Press.

Johnson, K., Flemming, E., & Wright, R. (1993). The hyperspace effect: Phonetic targets are hyperarticulated. *Language, 69*, 505–528.

Junqua, J. C. (1993). The Lombard reflex and its role on human listeners and automatic speech recognizers. *Journal of the Acoustical Society of America*, *93*(1), 510-524.

Jusczyk, P. W., Goodman, M. B., & Baumann, A. (1999). Nine-month-olds' attention to sound similarities in syllables. *Journal of Memory and Language*, *40*(1), 62-82.

Kahng, J. (2018). The effect of pause location on perceived fluency. *Applied Psycholinguistics*, *39*(3), 569-591.

Kallay, J. E., & Redford, M. A. (2018). Coarticulatory effects on "the" production in child and adult speech. In *Proceedings of the 9th International Conference on Speech prosody,* (pp. 1004-1007). Poznan, Poland.

Kang, O., & Rubin, D. L. (2009). Reverse linguistic stereotyping: Measuring the effect of listener expectations on speech evaluation. *Journal of Language and Social Psychology*, *28*(4), 441-456.

Kartushina, N., Hervais-Adelman, A., Frauenfelder, U. H., & Golestani, N. (2016). Mutual influences between native and non-native vowels in production: Evidence from short-term visual articulatory feedback training. *Journal of Phonetics*, *57*, 21-39.

Kim, M., Horton, W. S., & Bradlow, A. R. (2011). Phonetic convergence in spontaneous conversations as a function of interlocutor language distance. *Laboratory Phonology*, *2*(1), 125-156.

Kirov, C. & Wilson, C. (2012). The specificity of online variation in speech production. In *Proceedings of the 34th Annual Conference of the Cognitive Science Society* (pp. 587–592). Austin, TX.

Kormos, J. (2000). The role of attention in monitoring second language speech production. *Language Learning*, *50*(2), 343-384.

Kormos, J., & Dénes, M. (2004). Exploring measures and perceptions of fluency in the speech of second language learners. *System, 32,* 146–164.

Krause, J. C., & Braida, L. D. (2002). Investigating alternative forms of clear speech: The effects of speaking rate and speaking mode on intelligibility. *Journal of the Acoustical Society of America*, *112*(5), 2165-2172.

Krause, J. C., Braida, L. D. (2004). Acoustic properties of naturally produced clear speech at normal speaking rates. *Journal of the Acoustical Society of America, 115*, 362-78.

Kraut, R., & Wulff, S. (2013). Foreign-accented speech perception ratings: A multifactorial case study. *Journal of Multilingual and Multicultural Development*, *34*(3), 249-263.

Kroll, J. F., & Stewart, E. (1994). Category interference in translation and picture naming: Evidence for asymmetric connections between bilingual memory representations. *Journal of Memory and Language*, *33*(2), 149-174.

Kroll, J. F., Michael, E., Tokowicz, N., & Dufour, R. (2002). The development of lexical fluency in a second language. *Second Language Research*, *18*(2), 137-171.

Kuhl, P. K., Andruski, J. E., Chistovich, I. A., Chistovich, L. A., Kozhevnikova, E. V., Ryskina, V. L., & Lacerda, F. (1997). Cross-language analysis of phonetic units in language addressed to infants. *Science*, *277*(5326), 684-686.

Kuznetsova, A., Brockoff, P. B., & Christensen, R. H. B. (2016). Tests in Linear Mixed Effects Models. R package lmerTest version 2.0-30. Comprehensive R Archive Network (CRAN).

Ladd, D. R. (2008). *Intonational phonology*. Cambridge: Cambridge University Press.

Lado, R. (1957). *Linguistics across cultures*. Ann Arbor: University of Michigan Press.

Lai, Y. H. (2010). English vowel discrimination and assimilation by Chinese-speaking learners of English. *Concentric: Studies in Linguistics*, *36*(2), 157-182.

Lakens, D. (2013). Calculating and reporting effect sizes to facilitate cumulative science: a practical primer for t-tests and ANOVAs. *Frontiers in Psychology*, *4*, 1-12.

Lam, J., Tjaden, K., & Wilding, G. (2012). Acoustics of clear speech: Effect of instruction. *Journal of Speech, Language, and Hearing Research, 55,* 1807-1821.

Lam, J., & Tjaden, K. (2016). Clear speech variants: An acoustic study in Parkinson's disease. *Journal of Speech, Language, and Hearing Research*, *59*(4), 631-646.

Lee, D. Y., & Baese-Berk, M. M. (2020). The maintenance of clear speech in naturalistic conversations. *Journal of the Acoustical Society of America*, *147*(5), 3702-3711.

Lee, S., Potamianos, A., & Narayanan, S. (2013, June). Developmental aspects of American English diphthong trajectories in the formant space. In *Proceedings of Meetings on Acoustics ICA2013* (Vol. 19, No. 1, p. 060067). Acoustical Society of America.

Lehiste, I., & Peterson, G. E. (1959). Vowel amplitude and phonemic stress in American English. *Journal of the Acoustical Society of America*, *31*(4), 428-435.

Lenth, R.V. (2016). Least-Squares Means: The R Package lsmeans. *Journal of Statistical Software, 69*(1), 1-33.

Leung, K., Jongman, A., Wang, J., & Sereno, Z. (2016). Acoustic characteristics of clearly spoken English tense and lax vowels. *Journal of the Acoustical Society of America, 140*(1), 45–58.

Levis, J. M. (2005). Changing contexts and shifting paradigms in pronunciation teaching. *TESOL Quarterly*, *39*(3), 369-377.

Lindblom, B. (1990). Explaining phonetic variation: a sketch of the H&H theory. In W. J. Hardcastle & A. Marchal (Eds.), *Speech production and speech modelling*, (pp.403–439). Amsterdam: Kluwer Academic Publishers.

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.

Liu, S., Del Rio, E., Bradlow, A. R., & Zeng, F. G. (2004). Clear speech perception in acoustic and electrical hearing. *Journal of the Acoustical Society of America, 116*, 2374–2383.

Liu, Q., & Fu, Z. (2011). The Combined Effect of Instruction and Monitor in Improving Pronunciation of Potential English Teachers. *English Language Teaching*, *4*(3), 164-170.

Luke, S. G. (2017). Evaluating significance in linear mixed-effects models in R. *Behavior Research Methods*, *49*(4), 1494-1502.

Maniwa, K., Jongman, A., & Wade, T. (2009). Acoustic characteristics of clearly spoken English fricatives. *Journal of the Acoustical Society of America*, *125*(6), 3962-3973.

McAllister, R., Flege, J. E., & Piske, T. (2002). The influence of L1 on the acquisition of Swedish quantity by native speakers of Spanish, English and Estonian. *Journal of Phonetics*, *30*(2), 229-258.

McAuliffe, M., Socolof, M., Mihuc, S., Wagner, M., & Sonderegger, M. (2017). *Montreal Forced Aligner [Computer program]. Version 0.9.0*, retrieved 17 January 2017 from http://montrealcorpustools.github.io/Montreal-Forced-Aligner/

McCloy, D. R. (2016). phonR: tools for phoneticians and phonologists. R package version 1.0-7.

Minnick-Fox, M., & Maeda, K. (1999). Categorization of American English vowels by Japanese speakers. In J. J. Ohala, Y. Hasegawa, M. Ohala, D. Granville, & A. Baily (Eds.), *Proceedings of the 14th international congress of phonetic sciences* (pp. 1437–1440). Berkeley, CA: University of California.

Möbius, B. (2003). Gestalt Psychology meets phonetics–An early experimental study of intrinsic F0 and intensity. In *Proceedings of the 15th International Congress of Phonetic Sciences* (pp. 2677-2680).

Moon, S.J. & Lindblom, B. (1994). Interaction between duration, context, and speaking style in English stressed vowels. *Journal of the Acoustical Society of America, 96,* 40-55.

Munro, M. J. (1993). Productions of English vowels by native speakers of Arabic: Acoustic measurements and accentedness ratings. *Language and Speech*, *36*(1), 39-66.

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

Munro, M. J., & Derwing, T. M. (1995b). Processing time, accent, and comprehensibility in the perception of native and foreign-accented speech. *Language and Speech*, *38*(3), 289-306.

Munro, M. J., & Derwing, T. M. (1998). The effects of speaking rate on listener evaluations of native and foreign-accented speech. *Language Learning*, *48*(2), 159-182.

Munro, M. J., & Derwing, T. M. (1999). Foreign accent, comprehensibility, and intelligibility in the speech of second language learners. *Language Learning*, *49*(Supplement 1), 285–310.

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.

Munro, M. J., Derwing, T. M., & Morton, S. L. (2006). The mutual intelligibility of L2 speech. *Studies in Second Language Acquisition, 28*, 111–131.

Nearey, T. M. (1977). *Phonetic feature systems for vowels*. PhD Dissertation. University of Connecticut, USA.

Nip, I. S., & Blumenfeld, H. K. (2015). Proficiency and linguistic complexity influence speech motor control and performance in Spanish language learners. *Journal of Speech, Language, and Hearing Research*, *58*(3), 653-668.

Nittrouer, S. (2004). The role of temporal and dynamic signal components in the perception of syllable-final stop voicing by children and adults. *Journal of the Acoustical Society of America, 115*, 1777–1790.

Ohala, J. (1994). Acoustic study of clear speech: A test of the contrastive hypothesis. In *Proceedings of the International Symposium on Prosody* (pp. 75–89).

Oviatt, S., Levow, G.-A., Moreton, E., & MacEachern, M. (1998). Modeling global and focal hyperarticulation during human-computer error resolution. *Journal of the Acoustical Society of America, 104*, 3080–3098.

Picheny, M. A., Durlach, N. I., & Braida, L. D. (1985). Speaking clearly for the hard of hearing I: Intelligibility differences between clear and conversational speech. *Journal of Speech, Language, and Hearing Research*, *28*(1), 96-103.

Picheny, M. A., Durlach, N.I., & Braida, L.D., (1986). Speaking clearly for the hard of hearing II: acoustic characteristics of clear and conversational speech. *Journal of Speech and Hearing Research, 29*, 434–46.

Pierrehumbert, J. B. (2003). Phonetic diversity, statistical learning, and acquisition of phonology. *Language and Speech, 46,* 115–54.

Pivneva, I., Palmer, C., & Titone, D. (2012). Inhibitory control and L2 proficiency modulate bilingual language production: Evidence from spontaneous monologue and dialogue speech. *Frontiers in Psychology*, *3(*57), 1-18.

Poulisse, N. (1997). Language production in bilinguals. In A. M. B. de Groot & J. F. Kroll (Eds.), *Tutorials in bilingualism: Psycholinguistic perspectives* (pp. 201–224). Mahwah, NJ: Erlbaum.

Prince, A., & Smolensky, P. (1993). Optimality theory: Constraint interaction in generative grammar. *Rutgers University Center for Cognitive Science Technical Report* 2.

R Core Team (2020). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL https://www.R-project.org/.

Raphael, L. J. (1972). Preceding vowel duration as a cue to the perception of the voicing characteristics of word-final consonants in American English. *Journal of the Acoustical Society of America, 51*, 1296–1303.

Redford, M. A., & Oh, G. E. (2017). The representation and execution of articulatory timing in first and second language acquisition. *Journal of Phonetics*, *63*, 127-138.

Roberts, P. M., Garcia, L. J., Desrochers, A., & Hernandez, D. (2002). English performance of proficient bilingual adults on the Boston Naming Test. *Aphasiology*, *16*(4-6), 635-645.

Roelofs, A., & Verhoef, K. (2006). Modeling the control of phonological encoding in bilingual speakers. *Bilingualism: Language and Cognition*, *9*(2), 167-176.

Rogers, C. L., Dalby, J., & Nishi, K. (2004). Effects of noise and proficiency on intelligibility of Chinese-accented English. *Language and Speech*, *47*(2), 139-154.

Rogers, C. L., DeMasi, T. M., & Krause, J. C. (2010). Conversational and clear speech intelligibility of /bVd/ syllables produced by native and non-native English speakers. *Journal of the Acoustical Society of America*, *128*(1), 410-423.

Rosenfelder, I., Fruehwald, J., Evanini, K., Seyfarth, S., Gormon, K., Prichard, H., & Yuan, J. (2014). *FAVE (Forced Alignment and Vowel Extraction) Program Suite v1.2.2*, Available: 10.5281/zenodo.22281

Runnqvist, E., Strijkers, K., Sadat, J., & Costa, A. (2011). On the temporal and functional origin of L2 disadvantages in speech production: A critical review. *Frontiers in Psychology*, *2*, 1-8.

Saito, K., & Lyster, R. (2012). Effects of form-focused instruction and corrective feedback on L2 pronunciation development of /ɹ/ by Japanese learners of English. *Language Learning*, *62*(2), 595-633.

Scarborough, R., Dmitrieva, O., Hall-Lew, L., Zhao, Y., & Brenier, J. (2007). An acoustic study of real and imagined foreigner-directed speech. *Journal of the Acoustical Society of America*, *121*(5), 3044.

Scarborough, R. (2010). Lexical and conceptual predictability: Confluent effects on the production of vowels. In C. Fougeron, B. Kühnert, M.D'Imperio, & N. Vallée (Eds.). *Papers in Laboratory Phonology* (Vol. 10, pp. 557–586). Berlin: de Gruyter.

Scarborough, R., & Zellou, G. (2013). Clarity in communication: "Clear" speech authenticity and lexical neighborhood density effects in speech production and perception. *Journal of the Acoustical Society of America*, *134*(5), 3793-3807.

Schertz, J. (2013). Exaggeration of featural contrasts in clarifications of misheard speech in English. *Journal of Phonetics, 41*(3-4), 249–263.

Schertz, J., Cho, T., Lotto, A., & Warner, N. (2015). Individual differences in phonetic cue use in production and perception of a non-native sound contrast. *Journal of Phonetics*, *52*, 183-204.

Schneider, W., Eschman, A., & Zuccolotto, A. (2002). *E-Prime: User's guide*. Psychology Software Incorporated.

Schum, D. J. (1996). Intelligibility of clear and conversational speech of young and elderly talkers. *Journal-American Academy of Audiology*, *7*, 212-218.

Seyfarth, S., Buz, E., & Jaeger, T. F. (2016). Dynamic hyperarticulation of coda voicing contrasts. *Journal of the Acoustical Society of America*, *139*(2), EL31-EL37.

Sheppard, B. E., Elliott, N. C., & Baese-Berk, M. M. (2017). Comprehensibility and intelligibility of international student speech: Comparing perceptions of university EAP instructors and content faculty. *Journal of English for Academic Purposes*, *26*, 42-51.

Shea, C. E. (2014). Second language learners and the variable speech signal. *Frontiers in Psychology*, *5*, 1-3.

Shea, C. E., & Curtin, S. (2011). Experience, representations and the production of second language allophones. *Second Language Research*, *27*(2), 229-250.

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

Singmann, H., Bolker, B., Westfall, J., Aust, F., & Ben-Shachar, M.S. (2020). afex: Analysis of Factorial Experiments. R package version 0.27-2. https://CRAN.R-project.org/package=afex

Skowronski, M. D., & Harris, J. G. (2006). Applied principles of clear and Lombard speech for automated intelligibility enhancement in noisy environments. *Speech Communication*, *48*(5), 549-558.

Smiljanić, R., & Bradlow, A. R. (2005). Production and perception of clear speech in Croatian and English. *Journal of the Acoustical Society of America, 118*(3), Pt. 1:1677–1688.

Smiljanic, R., & Bradlow, A. R. (2008a). Stability of temporal contrasts across speaking styles in English and Croatian. *Journal of Phonetics*, *36*(1), 91-113.

Smiljanić, R., & Bradlow, A. R. (2008b). Temporal organization of English clear and conversational speech. *Journal of the Acoustical Society of America*, *124*(5), 3171-3182.

Smiljanić, R., & Bradlow, A. R. (2009). Speaking and hearing clearly: Talker and listener factors in speaking style changes. *Language and Linguistics Compass*, *3*(1), 236-264.

Smiljanić, R., & Bradlow, A. R. (2011). Bidirectional clear speech perception benefit for native and high-proficiency non-native talkers and listeners: Intelligibility and accentedness. *Journal of the Acoustical Society of America, 130*(6), 4020-4031.

Smith, A., & Zelaznik, H. N. (2004). Development of functional synergies for speech motor coordination in childhood and adolescence. *Developmental Psychobiology*, *45*(1), 22-33.

Stent, A., Huffman, M., & Brennan, S. (2008). Adapting speaking after evidence of misrecognition: Local and global hyperarticulation. *Speech Communication, 50*(3), 163–178.

Stibbard, R. M., & Lee, J. I. (2006). Evidence against the mismatched interlanguage intelligibility benefit hypothesis. *Journal of the Acoustical Society of America, 120*, 433–442.

Strange, W., Bohn, O. S., Trent, S. A., & Nishi, K. (2004). Acoustic and perceptual similarity of North German and American English vowels. *Journal of the Acoustical Society of America*, *115*(4), 1791-1807.

Strange, W., Bohn, O. S., Nishi, K., & Trent, S. A. (2005). Contextual variation in the acoustic and perceptual similarity of North German and American English vowels. *Journal of the Acoustical Society of America*, *118*(3), 1751-1762.

Summers, W.V., Pisoni, D.B., Bernacki, R.H., Pedlow, R.I., Stokes, M.A., (1988). Effects of noise on speech production: acoustic and perceptual analyses. *Journal of the Acoustical Society of America, 84*(3), 917–928.

Tajima K., Kitahara M., & Yoneyama K. (2015). Production of a non-contrastive sound in a second language. In The Scottish Consortium for ICPhS 2015 (Ed), *Proceedings of the 18th International Congress of Phonetic Sciences*. Glasgow, University of Glasgow, paper No 0802.

Tasko, S., & Greilick, K. (2010). Acoustic and articulatory features of diphthong productions: A speech clarity study. *Journal of Speech, Language, and Hearing Research, 53*, 84–99.

Thomas, E. R., & Kendall, T. (2015). NORM: The vowel normalization and plotting suite. Retrieved from http://lingtools.uoregon.edu/norm/norm1

Thomson, R. I., & Derwing, T. M. (2015). The effectiveness of L2 pronunciation instruction: A narrative review. *Applied Linguistics*, *36*(3), 326-344.

Tsukada, K. (2009). Durational characteristics of English vowels produced by Japanese and Thai second language (L2) learners. *Australian Journal of Linguistics*, *29*(2), 287-299.

Tsurutani, C. (2012). Evaluation of speakers with foreign-accented speech in Japan: The effect of accent produced by English native speakers. *Journal of Multilingual and Multicultural Development*, *33*(6), 589-603.

Tuomainen, O. T., & Hazan, V. (2018). Investigating clear speech adaptations in spontaneous speech produced in communicative settings. In M. Gósy & T. E. Gráczi (Eds.), *Challenges in analysis and processing of spontaneous speech* (pp. 9–25). Research Institute for Linguistics for the Hungarian Academy of Sciences.

Uchanski, R. M. (1988). *Spectral and temporal contributions to speech clarity for hearing impaired listeners*. Unpublished Doctoral Dissertation, Department of Electrical Engineering and Computer Science, Massachusetts Institute of Technology.

Uchanski, R. M. (2005). Clear speech. In D. B. Pisoni & R. E. Remez (Eds.), *Handbook of Speech Perception* (pp. 207-235). Malden, MA: Blackwell Publishers.

Uchanski, R. M., Choi, S. S., Braida, L. D., Reed, C. M., & Durlach, N. I. (1996). Speaking clearly for the hard of hearing IV: Further studies of the role of speaking rate. *Journal of Speech, Language, and Hearing Research*, *39*(3), 494-509.

Uther, M., Knoll, M. A., & Burnham, D. (2007). Do you speak E-NG-LI-SH? A comparison of foreigner-and infant-directed speech. *Speech Communication*, *49*(1), 2-7.

U.S. Census Bureau. (2018). *American Community Survey*. Retrieved from https://www.census.gov/acs/www/about/why-we-ask-each-question/language/

Van Engen, K. J., & Peelle, J. E. (2014). Listening effort and accented speech. *Frontiers in Human Neuroscience*, *8*, 577.

Van Hoof, S., & Verhoeven, J. (2011). Intrinsic vowel F0, the size of vowel inventories and second language acquisition. *Journal of Phonetics*, *39*(2), 168-177.

Vaughn, C., Baese-Berk, M., & Idemaru, K. (2019). Re-examining phonetic variability in native and non-native speech. *Phonetica*, *76*(5), 327-358.

Vokic, G. (2008). The role of structural position in L2 phonological acquisition: Evidence from English learners of Spanish as L2. *Foreign Language Annals*, *41*(2), 347-363.

Vokic, G. (2010). L1 allophones in L2 speech production: The case of English learners of Spanish. *Hispania*, 430-452.

Wang, X., & Munro, M. J. (1999). The perception of English tense-lax vowel pairs by native Mandarin speakers: The effect of training on attention to temporal and spectral cues. In J. J. Ohala, Y. Hasegawa, M. Ohala, D. Granville, & A. Baily (Eds.), *Proceedings of the 14th International Congress of Phonetic Sciences* (pp. 125–129). Berkeley, CA: University of California.

Wedel, A., Nelson, N., & Sharp, R. (2018). The phonetic specificity of contrastive hyperarticulation in natural speech. *Journal of Memory and Language*, *100*, 61-88.

Whalen, D. H., & Levitt, A. G. (1995). The universality of intrinsic F0 of vowels. *Journal of Phonetics*, *23*(3), 349-366.

Zamuner, T. S. (2006). Sensitivity to word-final phonotactics in 9- to 16-month-old infants. *Infancy*, *10*(1), 77-95.