Sociolinguistic and Phonetic Perception of Second Language Mandarin Chinese

Robert L. Squizzero

A dissertation

submitted in partial fulfillment of the

requirements for the degree of

Doctor of Philosophy

University of Washington

2022

Reading Committee:

Alicia Beckford Wassink, Chair

Betsy E. Evans

Zev Handel

Richard A. Wright

Program Authorized to Offer Degree:

Linguistics

University of Washington

Abstract

Sociolinguistic and Phonetic Perception of Second Language Mandarin Chinese

Robert L. Squizzero

Chair of the Supervisory Committee:
Alicia Beckford Wassink, Ph.D
Department of Linguistics

Perception of second language (L2) speakers and their speech is known to be influenced both by phonetic and by sociolinguistic factors. The existing body of scholarly research on L2 speech perception, however, is overwhelmingly focused on Indo-European languages, raising doubts about the generalizability of existing sociolinguistic, phonetic, acquisition, and pedagogical theory to other linguistic contexts. This dissertation aims to work towards rectifying this problem through a series of related studies investigating factors affecting the perception of L2 Mandarin Chinese. Chapters 1 and 2 consist of a group of closely related studies investigating the effect of perceived ethnicity on perception of language proficiency - including accentedness - and of speaker personal characteristics. Chapter 1 contains the first known study to establish that there is an effect of perceived ethnicity on perception of personal characteristics of L2 speakers of a non-Indo-European language. Chapter 2 confirms the results of Chapter 1 with a larger and more

diverse sample, allowing for an investigation of differences between listeners that may affect perception of speakers. The excursus to Chapter 2 expands on the findings of the main study by providing evidence for mediating and moderating factors of gendered and racialized judgments of L2 Mandarin speakers. Chapter 3 is a phonetic production study of L2 Mandarin consonants and vowels under the influence of L1 English, with careful attention paid to English and Mandarin dialectal variation. In a study analyzing the time-varying properties of vowels using generalized additive mixed models, Chapter 3 demonstrates that five phonemic vowels significantly differ between first language Mandarin speakers and advanced and intermediate L2 Mandarin speakers. Chapter 4 uses speech analyzed in Chapter 3, the procedures used in Chapters 1 and 2, and additional procedures from the intelligibility-comprehensibility-accentedness literature to establish sociolinguistic and phonetic factors interact to influence perception of adult L2 Mandarin Chinese. Together, these studies demonstrate that both beliefs about a speaker's perceived social group membership and the acoustic properties of speech indeed affect intelligibility, comprehensibility, accentedness, and perceived personal characteristics of L2 Mandarin Chinese speakers.

ACKNOWLEDGEMENTS

This dissertation would not have been possible without the contributions of many individuals. My adviser, Alicia Beckford Wassink, has believed in me and advocated for me since I was an undergraduate. Alicia taught me to be a meticulous scientist and a tough constructive critic, and she has been a role model for navigating the academy while remaining focused on social justice. I also extend thanks to the readers on my committee. Betsy Evans has been incredibly generous with her time and support. The perceptual experiments in this dissertation, especially the first three, were shaped during long conversations in Betsy's office, and she has repeatedly pulled me back from biting off more than I could chew. Richard Wright has kept me grounded by encouraging realistic expectations, he has been able to answer questions on an impressive array of topics, and he has always kept his sense of humor. Zev Handel has provided me with incredibly insightful comments and saved me from making countless errors in experiment design, interpretation, and writing. My GSR, Suhanthie Motha, encouraged me to pursue research combining theoretical and applied linguistics back when I was pursuing a master's degree in her department. Other linguistics faculty and postdocs have provided valuable advice, skills, and support over the past five years as well. I especially appreciate everything I've learned from Sharon Hargus, Qi Cheng, Ellen Kaisse, Sapna Cheryan, Charlotte Vaughn, Marina Oganyan, Alex Panicacci, and Matt Kelley.

I am also indebted to many of my fellow students. Prior to his untimely passing, Jiahui Huang provided many hours of assistance with the design, translation, and proofreading of experimental materials, abstracts, and conference presentations, and always offered me a smile and a chance to be filled in on the latest gossip. My sincere thanks go to Yadi Peng for stepping into Jiahui's shoes to help with the design, translation, and proofreading of the last two experiments and their

numerous pilot studies. Anna Moroz, Naomi Tachikawa Shapiro, Ben Jones, Nicole Chartier, Sara Ng, Ella Deaton, and Monica Jensen participated in reading groups that shaped the direction of this dissertation, and they have all been there to lend an ear as friends when I've needed them. The members of the Sociolinguistics Brown Bag and Phonetics Lab listened to many of my practice talks with kind attention, and they indulged me on all manner of topics relating to my research. Lastly, several of my former colleagues at National Taiwan University, St. John's University, and South Seattle College gave me the confidence to pursue a PhD and gave helpful advice about the process.

I would like to thank those who provided resources for the studies in this dissertation. The department's Linguistics Award in Diversity Scholarship gave me the opportunity for me to finish this work in a timely manner. Yang Chunsheng graciously shared audio recordings from his own dissertation study for use in the final two chapters of this dissertation. And the study participants provided their ears, eyes, judgments, and attention – 谢谢您们！

Finally, I express my sincere gratitude and love to my family and friends. I'm especially grateful to my parents for encouraging me to get up every day and do something that I'm passionate about, for fostering my curiosity, and for raising me to be an independent thinker. I could not possibly understate how much I appreciate my partner Ou Jie Zhao, who has always been there to answer countless language and culture questions, help with programming, proofread, listen to my ideas, and support and care for me as I worked on my degree.

Dissertation Outline

CHAPTER 1 has been *published* as:

Squizzero, Robert. 2020. Attitudes toward L2 Mandarin Speakers of Chinese and non-Chinese Ethnicity. In Kaidi Chen (ed.), *Proceedings of the 32nd Meeting of the North American Conference on Chinese Linguistics*, 521–538. Storrs, CT.

CHAPTER 2 has been *submitted* as:

Squizzero, Robert. Under review. Perceived ethnicity, gender, and attitudes towards second language Mandarin speakers. *Journal of Multilingual and Multicultural Development.*

CHAPTER 3 has been *accepted* as:

Squizzero, Robert. 2022, September. *The role of dialectology in L2 vowel acquisition; evidence from Mandarin Chinese*. Paper to be presented at the 4th International Symposium on Applied Phonetics, Lund, Sweden.

CHAPTER 4 will be *submitted* as:

Squizzero, Robert. (in preparation). The effects of perceived ethnicity and prosodic accuracy on intelligibility, comprehensibility, and accentedness in L2 Mandarin Chinese. *Language and Speech*.

TABLE OF CONTENTS

GENERAL INTRODUCTION

Accented second language (L2) speakers often face specific negative consequences in society, including adverse treatment in the media, courts, housing, and employment (Gluszek & Dovidio 2010; Lippi-Green 2012). This discrimination is often based on a listener's subjective feeling that an accented speaker is difficult to understand, even though research has shown that the strength of a speaker's accent is only weakly correlated with comprehensibility, the ease with which speech is understood, and it is also only weakly correlated with intelligibility, the proportion of speech which is understood (Munro & Derwing 1995). Furthermore, subjective evaluation of the strength of a speaker's accent can have little or nothing to do with the speech itself, but rather can be significantly influenced by language-external cues, such as a speaker's perceived ethnicity (Rubin 1992; Hanulíková 2018; 2021) and gender (Lu & Gnevsheva 2021).

While perceived accentedness has been shown to unfairly and adversely affect the way that L2 speakers are treated, a recent meta-analysis shows that most studies in which this phenomenon has been described have been conducted in English-speaking societies (Fuertes et al. 2012). In fact, Western social, cultural, and linguistic contexts are the settings in which the majority of research has been conducted in the fields of language attitudes, second language pronunciation, and social psychology (Markus & Kitayama 1991; Fuertes et al. 2012; Levis 2021). Of these research studies, and even including studies conducted in non-Western social or cultural contexts, respondents are often speaking or acquiring English; according to a recent editorial in the *Journal of Second Language Pronunciation*, for example, over half of the studies published in the seven-year history of the journal were conducted on English. The focus in these fields on Western, often English-speaking communities is consistent with a broader trend in the

social and behavioral sciences (Henrich, Heine & Norenzayan 2010). Western populations are often presented as being composed of 'standard' or 'unmarked' individuals, even though there are cultural factors, such as the notion of the independent self, which are unique to Western cultures (Markus & Kitayama 1991). This focus on Western populations is especially evidenced by the paucity of studies conducted in China, the most populous country in the world, and on Mandarin Chinese, which is second only to English in number of L2 speakers worldwide (Eberhard, Simons & Fennig 2021). One cultural factor that has been argued to differ psychologically between members of Chinese and Western societies is the perception of ethnicity (Liu, Li & Yue 2010), which has been shown to influence listener impressions of a speaker's accent strength in Western societies.

Research has shown that pronunciation, rather than lexical or morphosyntactic features, is most closely correlated with perceived L2 accentedness (Saito, Trofimovich & Isaacs 2016). However, compared to English, there is a lack of research on features of pronunciation which differ systematically between L1 (first language) and L2 Mandarin speech production. Much existing research on L2 Mandarin pronunciation has been conducted with beginner, rather than intermediate or advanced speakers. Furthermore, much existing research on L2 Mandarin has focused on lexical tone, with a relatively small amount of scholarly work having been conducted on consonants, and an even smaller amount of work having been conducted vowels or on utterance-level prosody. In addition to this relatively narrow focus, dialectal and sociolectal differences both in Mandarin learners' L1s and in Mandarin itself as a target language are often overlooked in Mandarin language pronunciation research and teaching. While a strictly phonemic approach to language learning and teaching has clear benefits in terms of simplicity, sub-phonemic differences between language varieties, such as dialectal differences, can prevent

the faithful acquisition of L2 sounds, adversely affecting speaker intelligibility (Porretta & Tucker 2015) and accentedness (McCullough 2013).

This dissertation examines sociolinguistic and phonetic factors affecting the perception of adult second language Mandarin Chinese, as well as perception of the speakers themselves, in a series of studies submitted for publication as separate manuscripts to peer-reviewed journals. Chapters 1 and 2 consist of a group of closely related studies investigating the effect of perceived ethnicity on perception of language proficiency - including accentedness - and of speaker personal characteristics. Chapter 1 contains the first known study to investigate effects of perceived ethnicity on perception of speakers of a non-Indo-European language. Chapter 2 confirms the results of Chapter 1 with a larger and more diverse sample, allowing for an investigation of differences between listeners that may affect perception of speakers. The excursus to Chapter 2 expands on the findings of the main study by employing the ambivalent sexism inventory (Glick & Fiske 1996) to reveal mediating and moderating factors of gendered and racialized judgments of L2 Mandarin speakers. Chapter 3 is a phonetic production study of L1 and L2 Mandarin consonants and vowels under the influence of L1 English, grounded in the Perceptual Assimilation Model (Best 1995; Best & Tyler 2007) and with careful attention paid to English and Mandarin dialectal variation. Chapter 4 uses speech analyzed in Chapter 3, the procedures used in Chapters 1 and 2, and additional procedures from the intelligibility-comprehensibility-accentedness literature to investigate the interaction of sociolinguistic and phonetic factors in perception of adult L2 Mandarin Chinese.

Throughout this dissertation, the term “sociolinguistic factors” refers to two separate - but related – traditions of research. Grounded in the tradition of language attitudes research (Lambert et al. 1960), the term “sociolinguistic factors” in Chapters 1 and 2 refers to listener beliefs based

on the perceived social group membership of speakers, specifically the speakers' ethnicities. In Chapter 3, the term "sociolinguistic factors" refers to linguistic forms, in both production and perception, as emblematic of a particular social group, consistent with variationist sociolinguistics (Labov, Weinreich & Herzog 1968). Speakers in Chapter 3 differ in terms of their L1/L2 status, and for the L2 speakers, their level of Mandarin proficiency. In Chapter 4, both of these meanings are used; effects of speaker ethnicity rely on the former definition, whereas effects of Mandarin language proficiency rely on the latter definition. Chapter 4 attempts to bridge the two different traditions of sociolinguistics to provide a comprehensive view of the effects on perception of L2 Mandarin as discussed in Chapters 1-3.

This dissertation aims to answer the following research questions:

Overall: How do sociolinguistic and phonetic factors affect perception of L2 Mandarin speech?

Chapter 1: What is the effect of perceived ethnicity on accentedness and perceived personal characteristics of L2 Mandarin speakers?

Chapter 2: What are the effects of perceived speaker ethnicity, speaker gender, and listener gender on accentedness and perceived personal characteristics of L2 Mandarin speakers? What is the mechanism underlying the interaction between perceived speaker ethnicity, speaker gender, and listener gender on accentedness and perceived personal characteristics of L2 Mandarin speakers?

Chapter 3: How can attention to subphonemic detail help identify linguistic factors in L2 speech that contribute to deviation from L1 pronunciation norms?

Chapter 4: What are the effects of perceived ethnicity and prosodic accuracy on intelligibility, comprehensibility, and accentedness in L2 Mandarin Chinese?

REFERENCES

Best, Catherine T. 1995. A direct realist view of cross-language speech perception. In Winifred

Strange (ed.), *Speech perception and linguistic experience: Theoretical and methodological issues in cross-language speech research.*, 167–200. York Press.

Best, Catherine T. & Michael D. Tyler. 2007. Nonnative and second-language speech perception. In *Second language speech learning: The role of language experience in speech perception and production.*, 13–34. Amsterdam: John Benjamins. DOI: https://doi.org/10.1075/lllt.17.07bes

Eberhard, David M., Gary F. Simons & Charles D. Fennig (eds.). 2021. *Ethnologue: Languages of the World.* Twenty-four. Dallas, Texas: SIL International.

Fuertes, Jairo N., William H. Gottdiener, Helena Martin, Tracey C. Gilbert & Howard Giles. 2012. A meta-analysis of the effects of speakers' accents on interpersonal evaluations. *European Journal of Social Psychology* 42(1). 120–133. DOI: https://doi.org/10.1002/ejsp.862

Glick, Peter & Susan T. Fiske. 1996. The Ambivalent Sexism Inventory: Differentiating Hostile and Benevolent Sexism. *Journal of Personality and Social Psychology* 70(450). 491–512.

Gluszek, Agata & John F. Dovidio. 2010. The way they speak: A social psychological perspective on the stigma of nonnative accents in communication. *Personality and Social Psychology Review* 14(2). 214–237. DOI: https://doi.org/10.1177/1088868309359288

Hanulíková, Adriana. 2018. The effect of perceived ethnicity on spoken text comprehension under clear and adverse listening conditions. *Linguistics Vanguard* 4(1). 1–9. DOI: https://doi.org/10.1515/lingvan-2017-0029

Hanulíková, Adriana. 2021. Do faces speak volumes? Social expectations in speech comprehension and evaluation across three age groups. *PLoS ONE* 16(10 October). DOI: https://doi.org/10.1371/journal.pone.0259230

Henrich, Joseph, Steven J. Heine & Ara Norenzayan. 2010. The weirdest people in the world? *Behavioral and Brain Sciences* 33(2–3). 61–83. DOI: https://doi.org/10.1017/S0140525X0999152X

Labov, William, Uriel Weinreich & Marvin I. Herzog. 1968. Empirical foundations for a theory of language change. In W.P Lehmann & Y Malkiel (eds.), *Directions for historical linguistics: a symposium*, 95–195. Austin: University of Texas Press.

Lambert, W.E., R.C. Hodgson, R.C. Gardner & S. Fillenbaum. 1960. Evaluational reactions to spoken languages. *Journal of Abnormal and Social Psychology* 60(1). 44–51.

Levis, John. 2021. L2 pronunciation research and teaching the importance of many languages. *Journal of Second Language Pronunciation* 7(2). 141–153. DOI: https://doi.org/10.1075/jslp.21037.lev

Lippi-Green, Rosina. 2012. *English with an accent : language, ideology and discrimination in the United States* 2nd ed. Routledge. Retrieved from https://www.routledge.com/English-with-an-Accent-Language-Ideology-and-Discrimination-in-the-United/Lippi-Green/p/book/9780415559119

Liu, James H., Mei Chih Li & Xiao Dong Yue. 2010. Chinese social identity and inter-group relations: The influence of benevolent authority. In Michael Harris Bond (ed.), *Oxford Handbook of Chinese Psychology*,. Oxford University Press. DOI: https://doi.org/10.1093/oxfordhb/9780199541850.013.0035

Lu, Yao & Ksenia Gnevsheva. 2021. Accentedness and personality evaluation of Asian and Caucasian second language speakers of English by Asian second language English listeners. *Journal of Multilingual and Multicultural Development*. DOI: https://doi.org/10.1080/01434632.2021.1959598

Markus, Hazel Rose & Shinobu Kitayama. 1991. Culture and the self: Implications for cognition, emotion, and motivation. *Psychological Review* 98(2). 224–253. DOI: https://doi.org/10.1037/0033-295x.98.2.224

McCullough, Elizabeth A. 2013. *Acoustic correlates of perceived foreign accent in non-native English*. The Ohio State University dissertation.

Munro, Murray J. & Tracey M. Derwing. 1995. Foreign Accent, Comprehensibility, and Intelligibility in the Speech of Second Language Learners. *Language Learning* 45(1). 73–97. DOI: https://doi.org/10.1111/j.1467-1770.1995.tb00963.x

Porretta, Vincent & Benjamin V Tucker. 2015. Intelligibility of foreign-accented words : Acoustic distances and gradient foreign accentedness. *Proceedings of the 18th International Congress of Phonetic Sciences* 1–4.

Rubin, Donald L. 1992. Nonlanguage factors affecting undergraduates' judgments of nonnative English-speaking teaching assistants. *Research in Higher Education* 33(4). 511–531. DOI: https://doi.org/10.1007/BF00973770

Saito, Kazuya, Pavel Trofimovich & Talia Isaacs. 2016. Second language speech production: Investigating linguistic correlates of comprehensibility and accentedness for learners at different ability levels. *Applied Psycholinguistics* 37. 217–240.

Proceedings of the 32nd North American Conference on Chinese Linguistics (NACCL-32). 2020.
Edited by Kaidi Chen. University of Connecticut, Storrs, CT. Pages 521-538.

Attitudes toward L2 Mandarin Speakers of Chinese and Non-Chinese Ethnicity

Robert Squizzero
University of Washington

Perceived ethnicity is known to bias perception of accentedness of non-native speakers, but most existing research demonstrating this has been done on majority languages of North American and European countries. In the present study, a group of L1 Mandarin listeners were asked to rate the personalities and language abilities of highly proficient L2 Mandarin speakers of Chinese and non-Chinese ethnicity after listening to short statements by each speaker. Results show an effect of ethnic ingroup favoritism on several personality traits but no difference in perceived language proficiency. In the same way that ethnic biases against L2 speakers of European languages have been confirmed by research, this study confirms that ethnic biases can be present in an East Asian context. The study also provides evidence that achieving nativelike pronunciation in a second language does not eliminate discrimination against members of an outgroup.

0. Introduction

Studies of perceived comprehensibility, intelligibility and accentedness[1] of second language (L2) speakers are often done on English in the US, where multilingualism isn't the norm and ethnic diversity is comparatively common. However, in much of the world, such as in many Chinese-speaking communities, multilingualism is the norm, yet ethnic diversity is not. Mandarin is spoken as a second language in greater China by many first language (L1) speakers of other Chinese languages, so exposure to accented Mandarin is not unusual. However, mutually unintelligible Chinese languages are popularly considered to be dialects of the same language, at least in mainland China, which results in a distinction between L2 Mandarin speakers of L1 Chinese languages and those of other L1s. While there are probably fewer proficient Mandarin speakers of non-Chinese L1s, many Mandarin speakers encounter these speakers through media, such as in unscripted television programs or internet videos, and some encounter these speakers in daily life. This exposure is likely to increase in the future with the growth in interest in Mandarin language

[1] *Intelligibility, comprehensibility,* and *accentedness* are used here as defined by Munro, Derwing, & Morton (2006). *Intelligibility*: Actual understanding of the meaning of a word or utterance, i.e. more understanding on the part of the listener is equivalent to more intelligibility of a part of a word or utterance. *Comprehensibility*: Ease with which a word or utterance can be understood, i.e. less effort on the part of the listener is equivalent to more comprehensibility on of a word or utterance. *Accentedness*: The degree to which the pronunciation of an utterance sounds different from an expected production pattern.

programs. In the United States, Mandarin is the 4th most popular foreign language by enrollment in K-12 education (American Councils for International Education 2017) and 7th most popular in higher education (Looney & Lusin 2018).

L2 speakers face specific consequences in society, such as negative treatment in the media, courts, housing, and employment (Lippi-Green 2012). While comprehensi-bility of L2 speech is influenced by morphosyntactic and lexical features, the factor most associated with L2 accentedness is pronunciation (Saito, Trofimovich, & Isaacs 2016). Both L2 speakers and their listeners often believe that it is possible to eliminate an L2 accent and may even believe that it is laziness on the part of the L2 speaker that prevents elimination of their accent (Gluszek & Dovidio 2010). There are conflicting opinions as to whether it is possible for adult L2 language learners to achieve nativelike pronunci-ation, but there is evidence that comprehensibility is improved if pronunciation instru-ction is more explicitly included in language class-room activities (Derwing & Munro 2015, Yang 2016). There is also evidence, outlined below, that perceived characteristics of a speaker's identity influence processing of the language input received by listeners. Some scholars argue that L2 speakers should not be responsible for "improvement" of their pronunciation to improve communication between themselves and L1 (or other L2) speakers, and that listeners should make an effort to better comprehend accented speech (Gluszek & Dovidio 2010, Kang, Rubin, & Lindemann 2015). Indeed, there is evidence that practice in listening to multiple accented L2 speakers of a language can improve speaker-independent intelligibility of accented speech (Baese-Berk, Bradlow, & Wright 2013. Regardless, students of second languages can make better-informed decisions about how much time and effort to spend on pronunciation with accurate information on how their accented speech is perceived.

1. Literature Review

Perceived speaker ethnicity is known to influence processing of and opinions about speakers and their speech. In a well-known study (Rubin 1992), college-student subjects listened to identical speech samples presented by a university instructor, but some subjects were presented with a photograph of a Caucasian woman while others were presented with a photograph of an Asian woman. Participants shown the photograph of the Asian woman performed worse on an intelligibility test and reported hearing a stronger accent. In a verbal guise study, Yook & Lindemann (2013) demonstrated that knowing a speaker's ethnicity can impact opinions about a speaker's character by surveying Korean college students who listened to speakers of five varieties of English. One group of participants were informed of the speakers' nationalities and ethnicities while the other group was not. In comparison to uninformed listeners, informed listeners tended to rate European-American and Korean English speakers higher than the others on status/competence traits, and they rated British Australian and African American Verna-cular English speakers lower on social attractiveness traits than did uninformed listeners.

SOCIAL COGNITION AND DUAL PROCESSING. The Dual Processing theory of social cognition, which developed over many decades and is not credited to one particular individual (but see Fiske & Neuberg 1990 for one major Dual Processing model and Fiske & Taylor 2017 for a general overview), can be used to better understand evaluative reactions to language input. The dual modes of reacting to and processing stimuli are 1) automatic and 2) conscious, and the model allows for investigation of the interface between the two. This section discusses relevant psychology literature along with its application to language attitudes research.

Macrae & Bodenhausen's (2000) explanation of categorical thinking illustrates many of the details of automatic processing. The categorical thinking process is divided into the three potential steps of activation, application, and inhibition. Automatic category activation refers to the initial, subconscious process that occurs when first encountering a stimulus, like the auditory stimulus of L2-accented speech, or being told that the ethnicity of the person who produced that speech matches the ethnicity of the listener. Automatic category activation may or may not take place; for the purpose of efficiency, the encountered person is associated with a relevant social category if and only if the social meaning of the encountered person is perceived as relevant to current information processing concerns. For the example of speaker ethnicity and L2-accented speech, the speaker will be categorized according to the listener's beliefs about L2 speakers and/or people of the L2 speaker's ethnicity. This categorization then triggers other representations of the category as well as stored information about members of such a social group. Category application is the next step in the categorical thinking process. Once a recognized individual is categorized into a social group, as long as the information perceived is consistent with the stereotypes held by the perceiver, the stereotype category is applied (Fiske & Taylor 2017). This application affects the subsequent interaction between individuals in the same way that any other prime would (Macrae & Bodenhausen 2000).

Categorical thinking, however, is not always applied; certainly, a perceiver could process a speaker as an individual person instead of a representative of a social category. According to a great deal of work cited by Macrae & Bodenhausen (2000), categorical stereotypes are most likely used when a perceiver lacks motivation or is burdened with a shortage of time or a high cognitive load (e.g. due to increased effort required to process accented speech). In such a situation, categorical thinking allows for efficiency and less expended effort in the processing of input. The perceiver can shift focus and free up mental resources for parsing of potential unexpected information. If the automatic perce-ption, however, is inconsistent with the held stereotype, or if multiple social stereotypes are simultaneously activated, the perceiver may either attempt to classify the perceived in terms of a subtype that contains unique categories differentiating the perceived from the stereotype(s), or the perceiver may instead try to access exemplars of members of the stereotype category (Fiske & Taylor 2017). The last possibility, which could happen after either form of category application, is that the perceiver proceeds with conscious processing as executive function takes over (Macrae & Bodenhausen 2000).

According to Fiske & Taylor (2017), conscious processing can inhibit and control the associations made during the category activation above. Fiske & Taylor outline several potential motives for determining whether processing rises to the level of con-sciousness or not; two of these, self-enhancement and trusting ingroup, encourage use of automatic processes. People usually expect good things from their ingroup members, and negativity stands out, cognitively speaking, much more quickly than positivity. An L2 speaker may be more likely to be considered to be an outgroup member by an L1 speaker, as would a person of a different ethnicity. Macrae & Bodenhausen (2000) argue that people are not always successful at inhibiting stereotypical thoughts; a person may fail to do so out of a lack of awareness that the thoughts occurring in the mind are stereo-typical. If a person is, in fact, aware, they must employ a cognitively demanding process to identify and replace the stereotypic thoughts. Should a person be unable to cope with such demands, the person has now created a situation in which the stereotypical thoughts are highly accessible yet not inhibited, resulting in more stereotypic thoughts. Given that cognitive load is increased when processing accented speech (Derwing & Munro 2015), inhibitory failure in this situation may be expected.

THE POTENTIATED RECRUITMENT FRAMEWORK FOR IMPLICIT AND EXPLICIT ATTITUDES. Bassili & Brown's (2005) *potentiated recruitment* framework has been used successfully to model language attitudes in terms of input, processing and response (Preston 2017). In the context of the dual processing model, potentiated recruitment is appealing as it allows components of attitude creation and activation to be accessed both consciously and auto-matically. Bassili & Brown (2005) argue that "connectionist networks" of stored experi-ences, evaluations, attitudes and exemplars are a good fit for the parts that underlie impli-cit and explicit attitudes. Rosenberg's (1968) *attitudinal cognitorium* is used as a means of understanding the structure of these networks in relation to stimulus processing and attitudinal response. The *cognitorium* is described as a connectionist network of *microconcepts*, each of which represents information sensitive to the context at the time it was stored in memory. A microconcept, a node in this network, is retrieved from memory in a new context each time it is accessed. It is then associated with other microconcepts, meaning that each implicit or explicit attitude a person holds can change in context depending on which microconcepts have been activated. This is why context is key to understanding attitudes, according to Bassili & Brown. These authors go as far as to say that "features of the context are just as involved in the potentiation features of the attitude object" (Bassili & Brown 2005, p. 555-556). Bassili & Brown's model is uniquely effect-tive in accounting for attitude malleability, as no node is likely to be recalled in the exact same way as when it was originally stored. Potentiation, or level of activation, of micro-concepts is influenced by a perceiver's recent information processing experiences, curr-ently perceived information about the attitude object, activation between linked concepts in the cognitorium, and, for explicit attitudes, cognitive activity in working memory. The influence from currently perceived information "exert[s] powerful

potentiating influences on the attitudinal cognitorium" (Bassili & Brown 2005, p. 553) and appears to be easiest to manipulate by a researcher in an experiment in the form of a prime.

SOCIAL COGNITION IN THE CHINESE CONTEXT. To apply a social cognition framework to perception of second-language Mandarin speakers by contemporary Chinese, certain concepts must be modified, or at least qualified, to fit the Chinese context. The concept of membership in an East Asian (typically Chinese, Japanese, or Korean) "collectivist" culture can be misinterpreted as loyalty to or identification with one's ingroup from the point of view of a western, "individualistic" culture. Yuki (2003) found that for small social groups of students in the US and Japan, Americans identified more with and were more loyal to their ingroups than their Japanese counterparts, according to psychological measures of ingroup loyalty and identity. Yuki also found that while knowledge of role relations within a group, as measured by a scale of subjective sociometric knowledge, was correlated with ingroup loyalty and identity for both Japanese and American respondents, perceived ingroup homogeneity was only correlated with ingroup identity for Americans. Yuki's experiment was set up from the viewpoint of traditional Confucian notions of social harmony as resulting from fulfillment of one's particular role in the family and not from loyalty to or identification with one's country, community, or society. In fact, according to Liu, Li, & Yue (2010), the idea of asserting one's identity with a broader social group, at least in China's Confucian culture, did not present itself until the forced establishment of European treaty ports in China. Ingroup favoritism with respect to membership in a social group broader than the family first became relevant and widespread in China via the adoption of western nationalism, done as a defense mechan-ism against Europeans and their stratified society, and eventually as "benevolent paternal-ism" towards outgroup members perceived as inferior (Liu et al. 2010). Western-style nationalism in modern China, however, is not necessarily a direct reaction to Western nations, given, for example, the widespread knowledge of the atrocities committed by Japan in China in the early twentieth century. Ethnicity, then, in China, has not historic-ally been divorced from nationality in terms of ingroup/outgroup categorization, and can serve as a source of discrimination in contemporary Chinese society. It has yet to be demonstrated, though, that ethnicity can interact with language in such categorization and subsequent discrimination.

LANGUAGE STEREOTYPES. Existing research on language stereotypes predicts that listeners evaluate speakers along dimensions of *status*, *solidarity*, and *dynamism*. These three affective dimensions are related to the types of affective meaning in social psychology that can be applied to evaluation of any object: *potency*, *evaluation*, and *activity* (Osgood, May, & Miron 1975). According to Zahn & Hopper (1985) and Fiske, Taylor, Cuddy, & Xu (2002) (the latter for *status* and *solidarity* only), individuals rated highly on *status* are thought to be of high competence, high social class, and high language fluency, people rated highly on *solidarity* are thought to be warmer and more socially and aesthetically attractive, and those rated highly on *dynamism* are thought to have higher levels of activity and social power. While Osgood et al. (1975) found that there is some

cultural variation in dimensions of affective meaning, they also found that evaluation, potency, and activity were significant for Cantonese speaker-listeners. Wible & Hui (1985) and White & Li (1991) found that evaluation, potency, and activity were significant for L1 Mandarin speakers listening to L2 Mandarin speech.

The Matched Guise Technique (Lambert, Hodgson, Gardner, & Fillenbaum 1960) traditionally presents a group of listeners with several recordings of the same text, each ostensibly produced by a different speaker. In actuality, some of the recordings are sets of two with each set produced by the same speaker using a different language or language variety. After listening to each recording, respondents are asked to rate each speaker's voice on semantic differential scales of personality traits and language impressions that are expected to represent listeners' attitudes of the groups thought to use these different language varieties. Use of the same speaker removes the possibility of ratings based on idiosyncratic variation naturally found between speakers. In contrast to traditional matched guise studies, the language samples presented to respondents are not varied in the present study (respondents rated the exact same language samples twice); rather, different information about the speaker's ethnicity is presented to respondents for each "guise." A similar technique been used with speaker nationality in previous matched guise studies (Hay, Nolan, & Drager 2006; Niedzielski 1999).

2. Research Questions

This research study aims to answer the following questions: Do ethnically Chinese native Mandarin-speaking listeners find ethnic Chinese L2 speakers to be more socially attract-ive and/or less accented due to simple ingroup bias, i.e. ethnic solidarity? Conversely, do these listeners find highly proficient ethnic outgroup members to be more socially attract-ive and/or less accented due to listener ideologies that ethnic ingroup members should know how to speak the language? Or, is the effect of ethnicity more nuanced than this?

3. Methods

STIMULI. L1 Mandarin stimuli were used to test the hypothetical situation in which an L2 speaker has acquired a nativelike pronunciation of Mandarin. Stimuli were selected from the Mandarin Affective Speech Corpus (Yang, Li, & Wu 2007), which contains recordings of 68 students (males: 45, mean age = 21.7) at Zhejiang University, all of whom had lived in Mainland China since birth and the "majority [of which] were trained to speak in standard Mandarin from early childhood" (p. 3). This corpus was chosen because it contains a large number of speakers reading the same sentences and is control-led for emotional state. The present study used 20 "neutral" recorded utterances from 20 different speakers (males: 10), chosen by the author to be relatively consistent in terms of prosody (length of utterance) and recording quality (presence of clipping, signal-to-noise ratio). In addition to the author, two L1 speakers of Mandarin from two different regions of China and one other highly proficient L2 Mandarin speaker screened the stimuli for regionally

marked speech features (such as merger of alveolar and retroflex fricatives, vowel diphthongization, etc.). Four speakers were presented as matched guises, while the remaining sixteen utterances were included as distractors. The example speakers and sentences for the matched guises are shown in Table 1. Different target sentences were used to make the task more engaging and better maintain rater attention. Half of the recordings were (falsely) presented to respondents as having been produced by ethnically Chinese L2 speakers and the other half of the recordings were (falsely) presented to respondents as produced by ethnically non-Chinese L2 speakers, but there were no actual linguistic differences between the distractor sentences in each group presented to listen-ers, and there were no differences at all between the target (matched guise) sentences.

Speaker ID	Sex	Sentence - Chinese	Sentence – English Translation
F1	Female	今天晚上会下雨。	It's going to rain tonight.
M1	Male		
F2	Female	我们室友总是把寝室弄得很脏。	Our roommates always make the dorm room very messy.
M2	Male		

Table 1: Target sentences and speakers

RESPONDENTS. Respondents were required to have been raised in Mainland China, of self-reported native or near-native Mandarin proficiency, at least 18 years of age, and with no known speech or hearing disorders. Because the web survey was administered using Google Cloud Platform, which is often inaccessible in China, respondents were recruited in Seattle through emails disseminated to Chinese student groups and students in introductory linguistics classes, flyers posted on the University of Washington campus, and through researcher contacts, though trained sociolinguists and first-order researcher contacts were asked to refrain from taking the survey. The survey was distributed from 9/27/2019 – 11/06/2019. 30 of the 58 respondents evaluated all 24 audio stimuli, and 28 of the 30 respondents provided their demographic information. Analyses include the 30 respondents who evaluated all 24 audio stimuli, and the completion rate is calculated to be 52%. An additional four respondents completed enough of the survey to evaluate both guises of at least one speaker; these observations were also included in the analyses. Respondents ranged in from 18-36 years of age, with a mean of 22.9 years, a variance of 22.5, and a standard deviation of 4.7. 13 of the respondents were women, 15 were men, and none were of other genders. Reported ethnicities were Han (n=24), Manchu (n=2), Hui (n=1), and no response (n=1). 11 respondents reported being an L1 speaker of at least one language or dialect in addition to Mandarin. Respondents were raised in a mixture of hometowns from different geographic areas of China (see Online Supplemental Materials).

PRESENTATION. Respondents were told (in Mandarin) that researchers were interested in getting their impressions on how second language Chinese speech sounds. Prior to beginning the study, respondents were asked for informed consent. Respondents

were then told that they would be listening to 24 different speech samples, each recorded by a different, highly proficient, L2 Mandarin speaker. As to not give the impression that some stimuli were produced by heritage speakers, respondents were told that none of these speakers were raised in a Chinese-speaking country, none began studying Chinese before the age of 14, and none learned Chinese from their families. Respondents were told that half of the speakers were ethnically Chinese (华裔 *huáyì)*, and half were of non-Chinese ethnicity (非华裔 *fēi huáyì)*. Respondents were accurately told that they would randomly be assigned to first hear the group of twelve ethnic Chinese speakers (n=21) or the group of twelve non-ethnic Chinese speakers (n=13). After rating each block, respon-dents were advised to take a break of a few minutes before being presented with the remaining twelve speech samples. Respondents were told that they could listen to each sample as many times as they wanted but were asked to quickly respond to each question according to their first impression of each speaker. Respondents were asked to wear earphones for the study and to adjust the volume to a suitable level before beginning.

Respondents were told that they would be asked to answer several questions about each speech sample. In order to familiarize respondents with the task, they were provided with an example question containing 7 radio buttons on a scale of 1 to 7, with two example adjectives that would later be included in the actual rating task: 冷淡的 *lěngdàn de* "cold (personality)" written to the left of the scale and 热情的 *rèqíng de* "warm (personality)" to the right of the scale. They were told the colder their impression of a speaker, the closer to "cold" they should click, whereas the warmer their impression of a speaker, the closer to "warm" they should click. They were told that if they were unsure or had no impression of the speaker that they should select the radio button in the center, labeled as 没什么印象 *méi shénme yìnxiàng* "no impression." At the top of the next page, the speaker number and ethnicity of the speakers in the current block were displayed. For the first recording, respondents clicked a "play" button to listen to each recording. Below the media player, they were presented with twelve semantic differential scales corresponding to nine personality traits and three language measures, described in the following section of this paper. Respondents rated the speaker on each trait and then clicked the "next" button, which submitted their response, cleared the form, and played the next recording. After listening to and rating all stimuli, respondents were asked for demographic information and thanked for their participation.

Participants rated each guise on nine personality traits and three impressions of language proficiency along seven-step semantic differential scales, in keeping with Osgood et al. (1975). The speaker traits included on the questionnaire were chosen after consulting language attitudes work for L2 speakers of English (Yook & Lindemann 2013) and L1 speakers of Mandarin (Liao 2008, Lin 2018, Peng 2016, Tan 2016, Yang 2014). Raters were also asked to rate each speaker's standardness, fluency, and foreign-accentedness on seven-level scales. Four rating scales of speaker traits and language proficiency were selected for each evaluative dimension to which they were predicted to correspond: *status:* competent-incompetent, smart-stupid, standard-not standard, and fluent-not fluent,

solidarity: deep-shallow, warm-cold, likeable-unlikeable, and accented-not accented, and *dynamism:* strong-weak, polite-rude, modern-conservative, and kind-arrogant. All seven values of the semantic differential scales were used by one respondent in a pilot study, so the seven options were left in. All semantic differential scales were arranged by placing the trait with negative valence on the left and positive valence on the right, except for the accented/not accented scale, which was reversed in order to detect straightlining, a strategy in which respondents simply select the same answer to all questions in order to finish the survey faster. No straightlining was observed, and accented/not accented scores were transformed to match the other scales, i.e. a score of 1 was converted to 7, a score of 2 was converted to 6, and a score of 3 was converted to 5.

4. Results

Table 2 shows summary statistics in the form of medians and interquartile ranges (IQRs) by semantic differential scale, pooling ratings from all respondents on all target speakers (distractor recordings were excluded). These descriptive statistics establish that ratings differed in terms of perceived speaker personality and perceived speaker language pro-ficiency, as found by Wible & Hui (1985). While the median ratings for personality trait scales were either 5 (for arrogant/kind, rude/polite and unlikeable/likeable) or 4 (for the other seven personality traits), the median ratings for each of the language proficiency scales (accented/not accented, fluent/not fluent, standard/not standard) was 7. The IQRs for conservative/modern and weak/strong show that a very high proportion of respond-ents were unwilling or unable to evaluate the speakers on those scales. Visual distribu-tions of ratings are presented in Figure 1. An exploratory factor analysis was conducted via principal components analysis with varimax rotation and Kaiser normalization. As expected, this analysis yielded three factors with eigenvalues above 1, accounting for 65.89% of the variance. The factor loadings are shown in Table 3. Rotated component 2 clearly consists of the three scales on which language proficiency was rated. As has been observed in previous studies on social stereotyping, it may be the case that *dynamism* is not a dimension at play here. If so, it is likely that the scales intended to measure dynam-ism actually measured *status* (strong-weak) and *solidarity* (the remaining three), and that rotated component 1 represents *solidarity* and rotated component 3 represents *status*. The principal components analysis and the vastly different distributions in ratings indicate that it would be illogical to group the twelve scales into the original three indices of *status*, *solidarity* and *dynamism*, in which ratings of language proficiency and personality traits were combined. Further, the non-normal distribution of the ratings as shown in Figure 1 indicates that non-parametric statistical procedures should be used, and that combining scales into indices representing dimensions of affective meaning may obscure individual differences among respondents in their interpretation of the points on each semantic differential scale. The ratings in this study, then, are conceptualized as follows: each respondent k generates score x_{ijk} by rating speaker i on bipolar scale j. Raters evaluated

the same speaker on the same bipolar scales twice, one in each ethnicity condition, so their scores were compared as:

$$H_0: Chinese_{ijk} = Non\text{-}Chinese_{ijk}.$$

Scale (English)	Scale (Mandarin)	Median	IQR (25%-75%)
Not Accented / Accented	没/有洋腔洋调	7*	6-7*
Arrogant / Kind	高傲的/和爱的	5	4-5
Cold / Warm	冷淡的/热情的	4	4-5
Conservative / Modern	保守的/前卫的	4	4-4
Not Fluent / Fluent	不/流利的	7	6-7
Incompetent / Competent	没/有才干的	4	4-5
Rude / Polite	没/有礼貌的	5	4-6
Shallow / Deep	肤浅的/有深度的	4	4-5
Not Standard / Standard	不/标准	7	6-7
Stupid / Smart	愚蠢的/聪明的	4	4-5
Unlikeable / Likeable	不/招人喜欢的	5	4-6
Weak / Strong	不/坚强的	4	4-4

Table 2: Medians and interquartile ranges for target stimuli ratings, across raters and speakers.* indicates a transformed rating

Scale / Component	Predicted dimension of affective meaning	Rotated Comp 1	Rotated Comp 3	Rotated Comp 2
Rude / Polite	Dynamism	.73	.32	
Stupid / Smart	Status	.47	.63	
Shallow / Deep	Solidarity		.87	
Cold / Warm	Solidarity	.79		
Incompetent / Competent	Status		.78	
Weak / Strong	Dynamism		.69	
Conservative / Modern	Dynamism	.53		
Arrogant / Kind	Dynamism	.88		
Unlikeable / Likeable	Solidarity	.77	.34	
Not Standard / Standard	Status			.86
Not Fluent / Fluent	Status			.88
Accented / Not Accented	Solidarity			.74

Table 3: Principal component analysis with varimax rotation and Kaiser normalization

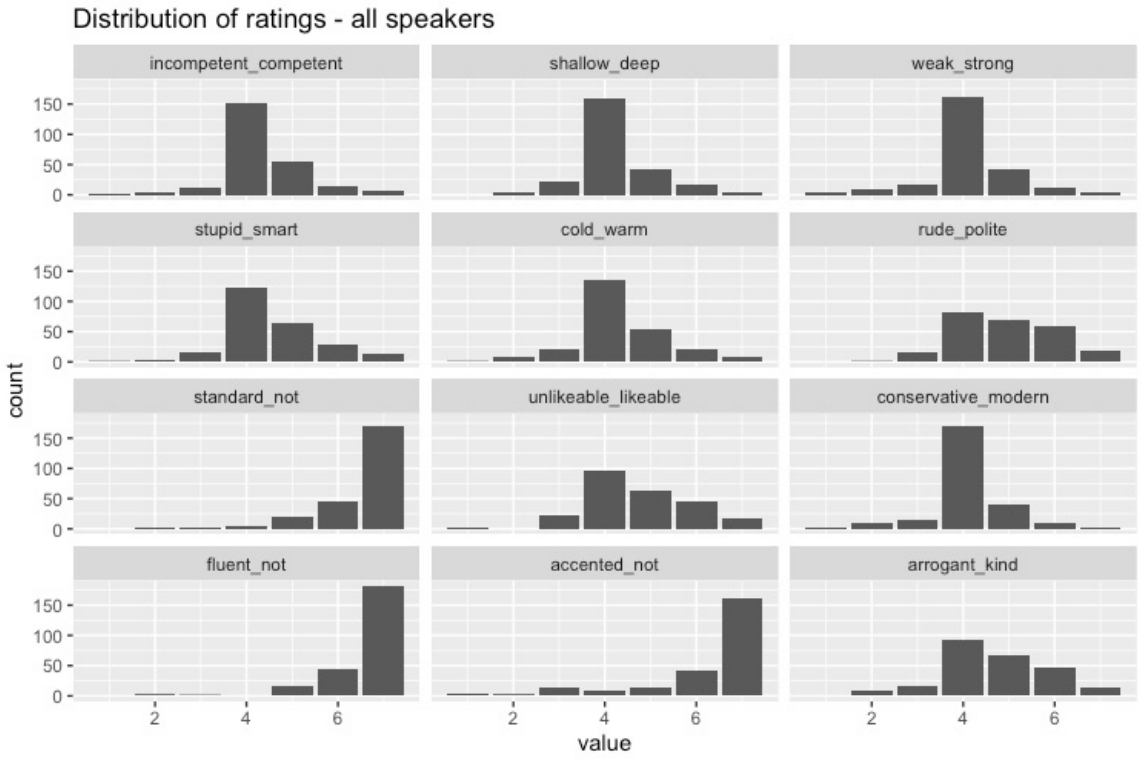

Figure 1 Counts of scores for target stimuli by rating scale, across raters and speakers. The columns, from right to left, represent status, solidarity, and dynamism measures.

Scale / Speaker	**F1**	**M1**	**F2**	**M2**
Not Accented / Accented	n.s.	n.s.	n.s.	n.s.
Arrogant / Kind	0.0094**	n.s.	n.s.	n.s.
Cold / Warm	0.0178*	n.s.	n.s.	n.s.
Conservative / Modern	n.s.	n.s.	n.s.	n.s.
Not Fluent / Fluent	n.s.	n.s.	n.s.	n.s.
Incompetent / Competent	0.0397*	n.s.	n.s.	n.s.
Rude / Polite	n.s.	n.s.	n.s.	n.s.
Shallow / Deep	0.0057**	n.s.	n.s.	n.s.
Not Standard / Standard	n.s.	n.s.	n.s.	n.s.
Stupid / Smart	0.0210**	n.s.	n.s.	n.s.
Unlikeable / Likeable	0.0013**	n.s.	n.s.	n.s.
Weak / Strong	n.s.	n.s.	n.s.	n.s.

Table 4: *p* values for each scale, by speaker. *$p < .05$, **$p < .01$

Scale	Pseudo-median	95% CI	Scale	Pseudo-median	95% CI
Arrogant / Kind	1.00	0.5-1.5	Shallow / Deep	1.00	0.00-2.00
Cold / Warm	1.00	0.00-2.00	Stupid / Smart	1.00	0.00-2.00
Incompetent / Competent	1.00	0.00-2.00	Unlikeable / Likeable	1.50	0.50-2.00

Table 5: Estimated location for the median of the difference from a sample of Chinese v. non-Chinese guises (pseudomedian) for speaker F1, all significant scales

To test whether these paired ratings were significantly different, a two-tailed Wilcoxon signed rank test was conducted, as in Bender (2004). This test has the advantage of not assuming that ratings on the semantic differential scales are continuous (interval or ratio) nor that the ratings are distributed normally in any of dimensions *i, j,* or *k*. In other words, it compares individual differences across conditions, so there is no potential for bias introduced by individual differences between raters' uses of the different scales on the different speakers. Results are shown in Table 4.

The *p* value of each two-tailed Wilcoxon signed rank test does not indicate directionality; it only indicates that there is a significant difference between rankings of each guise. To calculate directionality, an estimated location for the median of the difference between a sample from each guise is calculated based on a confidence interval, represented in Table 5. The sign of the estimated location indicates the direction in which rankings are higher; a positive value indicates that rankings in the first group (Chinese) are higher whereas a negative value indicates that the rankings in the second group (non-Chinese) are higher. All estimated location values for significant scale/speaker combinations were positive, indicating that the Chinese guise was rated more favorably for all statistically significant differences.[2]

3. Discussion

This study has provided strong evidence that L1 Mandarin-speaking listeners find ethnic Chinese L2 Mandarin speakers to be of higher status and more socially attractive than non-Chinese L2 Mandarin speakers due to ethnic ingroup implicit bias. In contrast to the results for character traits, no strong evidence is found for an effect of ethnicity on fluency, accentedness, or standardness. All observed differences in ratings for personality traits, statistically significant or not, are in the direction of favoring those who are ethnically Chinese. These findings confirm that ethnic biases present against L2 speakers of European languages can be extended to an East Asian context, despite claims by Liu et al. (2010) that such favoritism is only displayed by the Chinese as a defensive reaction to an outside threat or as a benevolent paternalism. While neither the presence of an outside

[2] Wilcoxon signed rank tests were also run to investigate ordering effects; no significant ordering effects were found. For more information, see the online supplemental materials.

threat nor the supposed inferiority of outgroup members was directly manipulated in the present experiment, no evidence for either is found in respondents' qualitative comments (see supplementary materials), and the semantic differential scales did not converge in a way comparable to paternalistic prejudice as previously demonstrated (Fiske et al. 2002).

The results, while robust, were only observed for the first of the four target speakers in this experiment that respondents heard. While interactions between demo-graphic variables such as speaker sex and ethnicity, or interactions between sentence length and demographic variables were suspected, it was not the case that an effect was shown for speakers of one sex or reading one sentence – it was shown only for a single speaker. One possible interpretation of this is that it reflects that social biases are not consistently applied; Fiske & Taylor (2017) write that categorical thinking does not happen consistently; rather, categorical thinking tends to occur in times of high cognitive load or low motivation (Macrae & Bodenhausen 2000). Higher cognitive load at the begi-nning of the survey due to unfamiliarity with the survey task is one possible explanation for why speaker F1 was the only speaker that was rated differently across guises.

The difference in distributions of personality and language proficiency ratings is also of interest to a theory of language attitudes; there is a high proportion of "no im-pression" responses found in all nine of the personality trait scales but a high proportion of "not accented," "not fluent," and "not standard" scores for language proficiency traits. Particularly high rates of half or more of the judgements being "no impression/don't know" were found for six of the nine personality trait scales: conservative/modern (169 of 246 ratings), weak/strong (161), shallow/deep (158), incompetent/competent (152), cold/warm (135), and stupid/smart (122). While the most common rating was "no impression/don't know" for the remaining three personality trait judgments, the median rating for rude/polite and unlikeable/likeable were "somewhat polite" and "somewhat likeable," whereas the median rating was still "no impression/don't know" for the arro-gant/kind scale. The most common rating and median rating on the language proficiency scales, in contrast, were the maximum ratings of "very fluent," "very standard," and "very unaccented." This means that respondents were not hesitant in judging a speaker's ability in the same way that they were hesitant about judging character. There are several possible interpretations for this difference. One interpretation is that social desirability bias prevents categorical thinking from applying to ratings of ability, i.e. it may be con-sidered fair, legally or morally, to judge language proficiency, as opposed to character. Alternatively, since most of the respondents are L2 learners of English who live in an English-speaking cultural context, they may be demonstrating empathy towards L2 learn-ers of Chinese. Another possible interpretation, supported by the qualitative comments, is that this population believes that language proficiency, unlike personality, can be judged by hearing a single sentence, such as those respondents heard in this study; this in itself is potentially relevant for future studies on ratings of L2 speakers. This is the opposite result, though, of what Wible & Hui (1985) found when L2 speakers of six different pro-ficiency levels were judged, which is that linguistic areas were less salient than person-ality traits. It may be the case, then, that

language proficiency is seen as something that can be controlled (Gluszek & Dovidio 2010), unlike personality traits, which may be perceived as innate. This would indicate that any preferential treatment of ingroup members is based on biased perceptions of character rather than biased perceptions of ability, at least for highly proficient second-language speakers. The fact that language proficiency ratings did not differ by ethnicity is of interest; while an effect was present for ethnic ingroup bias in rating speaker characteristics, no such effect was present for ratings of language proficiency for these highly proficient speakers in this study. This is particularly relevant for second language learners of Mandarin; it should be known that while speaking with a nativelike accent may result in being perceived as very fluent, standard, and unaccented, regardless of ethnicity, there may be still be negative treatment if a speaker is a member of the ethnic outgroup. Researchers investigating attitudes towards L2 speakers of any language should therefore be cautious about directly comparing ratings of language proficiency to those of personality. Chinese language teachers should also keep in mind that stereotypes against speakers may influence their proficiency. An L2 speaker may worry about how an outgroup member may stereotype them or may believe that an outgroup member may feign inability to understand the speaker. This stress may cause them to avoid communicating in the accented language, leading to less practice and less fluency, and arouse feelings of frustration or insecurity (Gluszek & Dovidio 2010).

The present study's methodology demonstrates a stricter adherence to the assumptions made by statistical procedures as compared to most existing research on language and social psychology. Responses along the semantic differential scales were treated as ordinal, resulting in reporting of median ratings and use of the Wilcoxon signed rank test. Treatment of data as ordinal eliminates the possibility of calculating mean ratings of speakers along the dimensions of affective meaning probed in this study, as well as the use of certain inferential statistical techniques, such as linear regression and ANOVA. Some researchers have argued that it is appropriate under certain circumstances to treat semantic differential scales as interval data (Carfino & Perla 2007); they are often considered to be special, separate from Likert scales or other opinion continua not validated for cross-cultural meaning. This view is contested, however, due to the potential for individual differences in use of the scales; such variation can be conceptualized as a three-dimensional space: variation between respondents in interpretation of the different anchor points on a scale, variation in interpretations of the distance between the seven points of different seven-point scales (scale variation), and variation in applying different scales to different stimuli (stimulus variation) (Murakami & Kroonenberg 2003). This individual variation can have the effect of obscuring, exaggerating, or overgeneralizing results when left unanalyzed. The present study's results should be interpreted with the understanding that individual differences in the ratings of different objects are often left uninvestigated. It may be the case that the result of the present study, in which only one of the four matched guise speakers is rated differently across guises, is not at all unusual; rather, it may be a result that is commonly obscured in other research by the pooling of ratings of different

stimulus speakers. More research in statistics and psychometrics would clarify best practices in quantitative analysis for research in language and social psychology; if semantic differential data can be safely treated as interval, it simplifies the analysis and increases inferential power. Nevertheless, if those who argue that it is only safe to treat these kinds of data as ordinal are correct, social psychologists and sociolinguists would do well to employ available statistical techniques for ordinal data, at least for within-subject study designs. The Wilcoxon signed rank test is an example of a powerful statistical procedure that allows for better understanding of individual distances while offering generalizable results.

While efforts were made to limit sampling bias, it is possible that some bias was introduced by the method by which respondents were recruited. The population sampled is one that lives in a multi-ethnic society, which is not the case for most Mandarin speakers, and is a young and educated population. Bias may also have been introduced by the dropout rate of 48%; reducing the length of the survey and using the semantic differential scales with fewer "no impression" ratings could assist in reducing this bias (Dillman, Smyth, & Christian 2014).

In addition to the short length of the utterances presented, it is possible that giving no information about each speaker other than ethnicity to respondents is too abstract to relate to a real-world situation and may not actually activate social categories or exemplars similar to the speaker being judged. Indeed, it has been shown that perceptions of L2 speech can vary even down to personae within a group (D'Onofrio 2019, Zhang 2008). It may be ideal to provide speaker demographic information in addition to ethnicity, even if kept constant across guises, such as the nationality of the respondents. Giving information about nationality was avoided, though, in order to avoid possible interactions between ethnicity and nationality. Another option, as suggested by one of the respondent's comments, is to tell respondents the context in which the example sentences were spoken; for instance, for the stimulus sentence about the weather, was this sentence produced in casual conversation, or was it produced by a meteorologist?

Future work measuring implicit, rather than explicit, attitudes or differences in processing of L2 Mandarin speech from speakers of different ethnicities may also help get to the root of the bias the population of L1 speakers from China. Such work would help determine whether social desirability bias is present, for example, by looking at differences in processing time when respondents believed they were listening to speech by ethnic Chinese and non-Chinese speakers.

4. Conclusion

The present work has demonstrated the potential for Chinese raters to express ingroup favoritism based only on knowledge of an outgroup member's ethnicity. For this population of raters, ingroup favoritism does not apply to an L2 speaker's language ability; rather, it applies to the speaker's personality. Chinese language students and teachers should be aware of the existence of this ingroup favoritism when considering how much

time and energy to devote to pronunciation. L2 Mandarin speakers of any ethnicity should know that positive or negative bias may very well occur based on their outward appearance or their interlocutors' understanding of their background, even if nativelike pronunciation is achieved, just as in the Western societies in which much research at the intersection of linguistics and social psychology is carried out. This is not to say that ethnic non-Chinese never experience preferential treatment in Chinese society; however, motivations for this treatment are probably not feelings of envy or admiration towards an outgroup (Fiske et al. 2002), nor are they an illustration of the Chinese concept of *chóng yáng mèi wài* (崇洋媚外), which refers to blind worship of foreign things. A detailed discussion of the source of the ethnic ingroup bias presently observed is beyond the scope of this investigation, though the nationalist ideology promoted by contemporary Chinese state media, as part of the global resurgence of nationalism also observed in the United States, Europe, and the Middle East, would not be expected to curtail ingroup favoritism nor outgroup derogation. Whatever the source of this bias, social scientists aiming to produce research which helps improve intergroup relations would benefit from a greater focus on the behavior of populations outside of the Western societies which are overrepresented in our disciplines.

SUPPLEMENTAL MATERIALS. Supplemental material is available at https://zeos.ling.washington.edu/publications/supplemental/squizzero/NACCLsupplemental.html. This material includes respondent demographic information, block-order analysis, and qualitative comments.

REFERENCES

American Councils for International Education. 2017. The National K-12 Foreign Language Enrollment Survey Report (June). Online: https://www.americancouncils.org/sites/default/files/FLE-report-June17.pdf

Baese-Berk, Melissa M., Ann R. Bradlow, and Beverly A. Wright. 2013. Accent-independent adaptation to foreign accented speech. *The Journal of the Acoustical Society of America*, *133*(3), EL174–EL180 https://doi.org/10.1121/1.4789864

Bassili, John N., and Rick D. Brown. 2005. Implicit and Explicit Attitudes: Research, Challenges and Theory. *The Handbook of Attitudes*, ed. by Dolores Albarracin, Blair T. Johnson, and Mark P. Zanna, 543–574. Mahwah: Lawrence Erlbaum.

Bender, Emily M. 2004. On the boundaries of linguistic competence: Matched-guise experiments as evidence of knowledge of grammar. *Lingua*, *115*(11 SPEC. ISS.), 1579–1598. https://doi.org/10.1016/j.lingua.2004.07.005

Carfino, J., & Perla, R. J. (2007). Ten Common Misunderstandings, Misconceptions, Persistent Myths and Urban Legends about Likert Scales and Likert Response Formats and their Antidotes. *Journal of Social Sciences*, *3*(3), 106–116.

D'Onofrio, Annette. 2019. Complicating categories: Personae mediate racialized expectations of non-native speech. *Journal of Sociolinguistics*, 1–21.

Derwing, Tracey, and Murray Munro. 2015. *Pronunciation Fundamentals: Evidence-based perspectives for L2 teaching and research*. Amsterdam: John Benjamins.

Dillman, Don A.; Smyth, Jolene D.; and Christian, Leah Melani. 2014. Reducing People's Reluctance to Respond to Surveys. In *Internet, phone, mail, and mixed-mode survey: The tailored design method.* (pp. 19–55). John Wiley & Sons.

Niedzielski, Nancy A. 1999. The Effect of Social Information on the Phonetic Perception of Sociolinguistic Variables. *Journal of Language and Social Psychology*, *18*(1), 62–85.

Osgood, Charles E.; William H. May; and Murray S. Miron. 1975. *Cross-Cultural Universals of Affective Meaning*. Urbana, IL: University of Illinois Press.

Peng, Chun-Yi. 2016. *Talk Shows and Language Attitudes: A Sociolingusitic Investigation of Language Attitudes towards Taiwan Mandarin among Chinese Mainlanders*. The City University of New York.

Preston, Dennis. R. 2017. The cognitive foundations of language regard. *Poznan Studies in Contemporary Linguistics*, 53(1), 17–42. https://doi.org/10.1515/psicl-2017-0002

Rosenberg, Milton J. 1968. Hedonism, inauthenticity, and other goads toward expansion of a consistency theory. *Theories of cognitive consistency: A sourcebook*, ed. by Robert P. Abelson; Elliot Aronson; William J. McGuire; Theodore M. Newcomb; Milton J. Rosenberg; & Percy H. Tannenbaum, 73-111. Chicago: Rand McNally.

Rubin, Donald L. 1992. Nonlanguage factors affecting undergraduates' judgments of nonnative English-speaking teaching assistants. *Research in Higher Education*, *33*(4), 511–531.

Rubin, Donald. L., and Kim A. Smith. 1990. Effects of accent, ethnicity, and lecture topic on undergraduates' perceptions of nonnative English-speaking teaching assistants. *International Journal of Intercultural Relations, 14*(3), 337–353.

Saito, Kazuya; Pavel Trofimovich; and Talia Isaacs. 2016. Second language speech production: Investigating linguistic correlates of comprehensibility and accentedness for learners at different ability levels. *Applied Psycholingusitics*, *37*, 217–240.

Tan, Lekun. 2016. A Comparison of Language Awareness and Attitudes between Students in the Department of Taiwanese Literature and Other Departments. *Journal of Taiwanese Languages and Literature*, *11*(2).

Wible, David S, and C. Harry Hui. 1985. Perceived Language Proficiency and Person Perception. *Journal of Cross-Cultural Psychology*, *16*(2), 206–222.

White, Michael J., and Yun Li. 1991. Second-Language Fluency and Person Perception in China and the United States. *Journal of Language and Social Psychology*, *10*(2), 99–113.

Yook, Cheongmin, & Stephanie Lindemann. 2013. The role of speaker identification in Korean university students' attitudes towards five varieties of English. *Journal of Multilingual and Multicultural Development*, *34*(3), 279–296.

Yang, Chunsheng. 2014. Language attitudes toward Northeastern Mandarin and Putonghua (PTH) by young professionals. *Language and Discourse: An International and Interdisciplinary Journal*, *5*(2), 211–230.

Yang, Chunsheng. 2016. *The acquisition of L2 Mandarin prosody: From experimental studies to pedagogical practice*. Amsterdam: John Benjamins.

Yang, Yingchun; Li Dongdong; & Wu Zhaohui. 2007. Mandarin Affective Speech Corpus. Philadelphia: Linguistic Data Consortium. Online: https://catalog.ldc.upenn.edu/LDC2007S09.

Yuki, Masaki. 2003. Intergroup comparison versus intragroup relationships: A cross-cultural examination of social identity theory in North American and East Asian cultural contexts. *Social Psychology Quarterly*, *66*(2), 166–183.

Zahn, Christopher, & Robert Hopper. 1985. Measuring language attitudes: the speech evaluation instrument. *Journal of Language and Social Psychology*, *4*(2), 113–123.

Zhang, Qing. 2008. Rhotacization and the 'Beijing Smooth Operator': *Journal of Sociolinguistics*, *12*(2), 201–222. https://doi.org/10.1386/jmpr.7.1.43/1

Perceived ethnicity, gender, and attitudes towards second language Mandarin speakers

Robert Squizzero
University of Washington, Seattle, WA, USA
ORCiD: https://orcid.org/0000-0002-2542-9933

Corresponding author

Robert Squizzero, Department of Linguistics, Guggenheim Hall 4th Floor, Box 352425, University of Washington, Seattle, WA, USA, 98195
rsquizz@uw.edu

Word count: 5662

Abstract

It is known that evaluative beliefs related to the perceived personal characteristics, accentedness, and language proficiency of second language speakers are affected by speaker ethnicity, speaker gender, and listener gender. However, the overwhelming majority of language attitudes research has been conducted in Western cultural and linguistic contexts. In a matched guise study, ethnically Chinese first language Mandarin listeners in mainland China were told that they were listening to statements by highly proficient second language Mandarin speakers. Male listeners rated the personal characteristics of female speakers significantly higher when they believed that they were listening to ethnically Chinese speakers, whereas female listeners rated the personal characteristics of a female speaker significantly higher when they believed that they were listening to an ethnically non-Chinese speaker. Different from previous studies, significant effects were neither observed for male speakers nor for language proficiency. Also differing from previous studies, rather than evaluating speakers along two or three dimensions of affective meaning, listeners evaluated speakers' personal characteristics along a single dimension. Findings indicate that much remains to be learned about the way second language speakers are perceived, both by studying new cultural and linguistic contexts and by directly incorporating speakers' multiplex social identities into study designs.

Keywords: language attitudes, social psychology, Chinese, matched guise

Introduction

Mandarin Chinese is second only to English in number of worldwide second language (L2[i]) speakers (Eberhard, Simons & Fennig 2021). Despite the large number of L2 Mandarin speakers worldwide, studies of attitudes towards and judgments about these speakers are rare; according to a meta-analysis, the vast majority of scholarly work investigating effects of accentedness on interpersonal evaluations has been conducted on accented English (Fuertes et al. 2012). While having an L2 accent has been shown to adversely affect the way that an individual is treated in society (Lippi-Green 2012), most societies in which this phenomenon has been described are English-speaking societies.

The focus on Western, often English-speaking communities is not limited to studies involving accentedness but is also consistent with a broader trend in the social and behavioral sciences (Henrich, Heine & Norenzayan 2010). Western populations are often presented as 'standard' or 'unmarked' individuals, even though there are cultural factors, such as the notion of the independent self, which are likely at play exclusively in Western cultures (Markus & Kitayama 1991). This focus on Western populations has excluded China, the most populous country in the world, even though Western universities enroll many Chinese students in postgraduate research programs (Mei & Brown 2018).

Prior work on L2 English, Dutch and German has shown that perceived speaker ethnicity can affect judgments of both speaker accentedness and personal characteristics, with speakers of the majority ethnicity generally receiving lower accentedness ratings and higher personal characteristics ratings (Rubin 1992; Gnevsheva 2018; Hanulíková 2018; 2021; Lu & Gnevsheva 2021). Yet, in the only known study on L2 Mandarin speakers and ethnicity (Squizzero 2020), accentedness ratings did not differ by perceived speaker ethnicity, even though speakers of the

majority ethnicity received higher personal characteristics ratings. Personal characteristics results were also gendered in that they only varied for a female speaker, consistent with the novel findings of a recent study on L2 English which used Chinese raters (Lu & Gnevsheva 2021). The present study aims to fill a gap in the literature by following up on the author's exploratory study of L2 Mandarin speakers, but by surveying a larger population of individuals living in mainland China rather than by surveying a smaller population of Chinese nationals living in the United States.

Background

Ethnicity and speaker evaluation

Listener opinions both of speakers and their speech are known to be influenced by the listener's perception of the speaker's ethnicity. Kang & Rubin (2009) coined the term *reverse linguistic stereotyping* to refer to the phenomenon in which a listener's judgement of a speaker's language ability is influenced by information that identifies the speaker as a member of a particular social group, including ethnicity information. In Rubin's (1992) Study 1, participants heard identical speech samples produced by a first language (L1) English speaker from Ohio, but participants who were led to believe that the speech came from an ethnically Chinese instructor rated the sample as more accented than participants who were led to believe that the speech came from an ethnically Caucasian instructor. In a study comparing perception of L1 speakers of Canadian English of White versus Chinese ethnicity, Babel & Russell (2015) found that when participants were shown a photograph of a speaker's face, ethnically Chinese speakers were rated as having more of a foreign accent, whereas ethnically White speakers were rated as having less of a foreign accent. In an Australian study comparing ratings of English spoken by L1 German, Korean, and English speakers in which participants could only hear audio, only see video, or

both see video and hear audio, study participants rated speakers who were phenotypically Asian in appearance as more accented than speakers with who were phenotypically Caucasian in appearance (Gnevsheva 2018). In a verbal guise study of English in South Korea, Yook & Lindemann (2013) demonstrated that knowing a speaker's ethnicity can impact opinions about the speaker's status/competence and social attractiveness. In Yook & Lindemann's study, one group of participants was informed of the speakers' ethnicities while the other group was not; informed listeners rated African Americans lower than European Americans on both status/competence and social attractiveness traits.

Speaker judgments and ethnicity in a Chinese cultural and linguistic context

The majority of studies involving perception of L2 speakers based on perceived speaker ethnicity have been conducted on speakers of English. With the exception of one prior study on Mandarin (Squizzero 2020), the only published studies involving perceived ethnicity on languages other than English, to the author's knowledge, were conducted on other West Germanic languages: Dutch (Hanulíková 2018) and German (Hanulíková 2021). In both of these studies, Hanulíková found that non-majority ethnicity speakers were rated as more accented than majority ethnicity speakers under certain experimental conditions, but no such effect was found for Mandarin (Squizzero 2020).

Markus & Kitayama's (1991) seminal work on construal justifies the importance of testing theories about how others, including multilinguals, are perceived in non-Western cultural contexts. According to Markus & Kitayama's review, Western cultures typically view the self as an independent entity, separate from other individuals. Conversely, East Asian cultures typically view the self as an interdependent entity, meaning that one's identity is viewed within the context of one's relations with other people. This difference in construal of the self can lead, for

example, to different ideas about gender roles and stereotypes; in a study of cross-cultural gender stereotypes, Americans rated women to be more collectivistic than men, whereas Koreans rated women as less collectivistic than men (Cuddy et al. 2015).

Studies on attitudes towards L2 English speakers do occasionally include non-Western participants. In one such study, Lu & Gnevsheva (2021) asked multilingual participants from China living in a Western context, in Australia, to rate L2 English speakers of Korean and German ethnicity and L1 backgrounds. However, recruiting participants from a non-Western national or cultural origin may not be enough to produce substantially different results from existing research on English, since the language in which a study is conducted has been shown to influence stereotype content. In a study on gender stereotypes, bicultural Korean Americans were randomly assigned to complete a survey in English or Korean, and participants who completed the survey in English rated individuals more in line with American gender stereotypes whereas participants who completed the same survey in Korean rated individuals more in line with Korean gender stereotypes (Cuddy et al. 2015). Cuddy and colleagues concluded that the difference in results based on the language of the survey can be attributed to the ability of a language itself to prime a culture's norms and values.

Ethnicity, gender, and speaker evaluation

Intersectionality theory (Crenshaw 1989) holds that discrimination can occur along multiple simultaneous dimensions of social group membership. Under intersectionality theory, individuals who belong to multiple marginalized social groups may face discrimination in the way that members of a single marginalized group face discrimination, or they may face discrimination in a unique way that is the result of their multiplex group membership. Crenshaw's first work on the topic was a critique of the American legal system's conception of race or sex-based

discrimination at the time, which specifically made it difficult for Black American women, as opposed to Black American men or white American women, to seek legal recourse for discrimination. Since 1989, intersectionality theory has been extended in the social sciences, not only as a framework for understanding marginalization along a broad variety of social categories but also for understanding how the identities of multiply marginalized individuals are co-constituted, including in linguistics (Levon 2015) and in social psychology (Purdie-Vaughns & Eibach 2008). Intersectionality theory can therefore be extended to help account for attitudes towards L2 Mandarin speakers whose perceived identities reflect different combinations of gender and ethnicity, each with its own potentially unique social perception.

Studies on accentedness and perceived personal characteristics that analyze results based on the ethnicity of the speaker and the gender of both the speaker and the listener are rare. In one example of such a study, Lu & Gnevsheva (2021) compared accentedness and personal characteristics ratings of speakers of Korean-accented and German-accented English across two experimental conditions: 1) listeners presented with audio of speakers saying utterances only and 2) listeners presented with video of speakers saying the same utterances. Results indicated that male listeners, on average, rated Korean-accented speakers as more accented than German-accented speakers. Furthermore, male speakers were rated as more competent when they could be seen by listeners. Lastly, Lu and Gnevsheva also found that while listeners rated both male and female speakers as less accented when they could be seen, that the difference between the two conditions was greater for male speakers than for female speakers. In other words, there was a greater decrease in perceived accentedness for male speakers, relative to female speakers, when listeners were presented with videos as opposed to audio recordings. In a study of Spanish-accented English vs. North American-accented English, male listeners were more likely than

female listeners to rate Spanish-accented speakers as less knowledgeable, and Spanish-accented female speakers were rated as less knowledgeable than Spanish-accented male speakers (Nelson, Signorella & Botti 2016), though the authors of the study used accent as a proxy for speaker ethnicity. Lastly, in an exploratory study, participants rated the personal characteristics of a female L2 Mandarin speaker significantly higher when they believed that she was ethnically Chinese (华裔 *huáyì)* as opposed to when they believed that she was not ethnically Chinese (非华裔 *fēi huáyì)*, but ratings did not differ accordingly for male L2 Mandarin speakers (Squizzero 2020).

Dimensions of affective meaning

The paucity of attitudinal studies conducted on L2s other than English calls into question whether existing knowledge about language attitudes is applicable to humans in general or only to people in English-speaking contexts. One domain in which language attitudes research may be restricted to English speakers is whether listeners evaluate speakers along multiple distinct dimensions, as predicted by a great deal of research in the social psychology and language attitudes traditions. In social psychology, the widely-cited stereotype content model proposes two distinct dimensions of affective meaning: competence and warmth (Fiske, Cuddy, Glick & Xu, 2002), whereas in language attitudes research, three distinct dimensions of affective meaning: superiority, attractiveness, and dynamism (Zahn & Hopper 1985) are often investigated, with Zahn & Hopper's superiority and attractiveness analogous to Fiske and colleagues' competence and warmth, respectively. While Zahn & Hopper caution researchers that their speech evaluation instrument had not been cross-culturally validated at the time of publication, their three proposed dimensions of affective meaning are analogous to those that have been argued to generally apply, cross-culturally, to evaluation of any object: potency,

evaluation, and activity (Osgood, Suci & Tannenbaum 1957; Osgood, May & Miron 1975). According to subsequent research, individuals or groups of people rated highly on superiority or competence are thought to be of high social class, high ability, and high language proficiency, whereas people rated highly on attractiveness or warmth are thought to be more socially and aesthetically attractive, and warmer in personal characteristics. However, these findings were not borne out in a study of L2 Mandarin speakers (Squizzero 2020), which showed only a single dimension of affective meaning for perceived personal characteristics, as well as a dimension of affective meaning for perceived language proficiency. While the separation of language proficiency traits into their own dimension may be due to the methodological choice of using L1 stimuli (as in Rubin 1992), the nonconvergence of the remaining personal characteristics evaluations into the three dimensions of superiority, attractiveness, and dynamism or into the two dimensions of competence and warmth cannot be explained by the selection of stimuli, nor by the selection of semantic differential scales. A broad review of social psychology research on the nature of attitudes showed that while social judgments have been shown to be separable in some instances into multiple dimensions, other research suggests that social judgments vary along a single dimension (Banaji & Heiphetz 2010). Other previous studies on L2 Mandarin offer mixed evidence: White & Li (1991) reported that perceived speaker fluency was positively associated with traits typically linked both to superiority/competence and to attractiveness/warmth, but Wible & Hui (1985) reported that the general impression of a speaker's Mandarin was correlated with traits that classically index superiority/competence, such as intelligence and ambition, and not with traits typically associated with attractiveness/warmth, such as honesty and friendliness.

Materials and Procedure

Participants

96 participants (gender: 50 women, 46 men, 0 neither, ethnicity: 95 Han, 1 Hui, age range: 22-50 yrs, mean age: 30 yrs) were recruited via Wenjuan, a Chinese online survey platform similar to Amazon Mechanical Turk, in April 2020. Participants were required to have been raised in mainland China, of self-reported native or near-native Mandarin proficiency, at least 18 years of age, and with no known speech or hearing disorders. Participants were physically located in China at the time of the survey, and they were raised in a mixture of hometowns from different geographic areas of China.

Materials

L1 Mandarin stimuli were used to test the hypothetical situation in which an L2 speaker has acquired a pronunciation of Mandarin similar to that of L1 speakers and to remove proficiency as a potential confound. Stimuli were selected from the Mandarin Affective Speech Corpus (MASC) (Y. Yang, Li & Wu 2007) which contains recordings of 68 students (sex: 45 males, 23 females[ii], mean age: 21.7) at Zhejiang University, all of whom had lived in mainland China since birth and the "majority [of which] were trained to speak in standard Mandarin from early childhood" (p. 3). This corpus was chosen because it contains many speakers reading the same sentences and is controlled for emotional state. 16 recorded utterances of "neutral" affect from 16 different speakers (males: 8), were chosen to be relatively consistent in terms of prosody (length of utterance) and recording quality (presence of clipping, signal-to-noise ratio). Four speakers were presented as matched guises, while the remaining utterances were included as distractors. The example speakers and sentences for the matched guises are shown in Table 1. Different target sentences were used to make the task more engaging and better maintain rater attention. Half of the recordings were falsely presented to participants as having been produced

by huayi (ethnic Chinese) L2 Mandarin speakers and the other half of the recordings were falsely presented to participants as produced by non-huayi L2 speakers, but there were no actual linguistic differences between the distractor sentences in each group presented to listeners, and there were no differences at all between the target (matched guise) sentences.

Participants rated each guise on personal characteristics and impressions of language proficiency along six-point semantic differential scales. The speaker traits included on the questionnaire were chosen after consulting language attitudes work for L2 speakers of English (Yook & Lindemann 2013; McKenzie, Kitikanan & Boriboon 2016) and L1 speakers of Mandarin (Liao 2008; C. Yang 2014; Peng 2016; Tan 2016; Lin 2018). Raters were also asked to rate each speaker's standardness, fluency, and foreign accentedness. All semantic differential scales were arranged by placing the trait with negative valence on the left and positive valence on the right, except for the accented/not accented scale, the scores of which were transformed to match those of the other scales. The valence for accented/not accented was deliberately arranged in the opposite order of the other scales to detect straightlining, a strategy in which participants select the same answer to all questions for the purpose of finishing the survey faster. Responses of eight of the participants were excluded for straightlining, and all other responses (n=88) were analyzed. After listening to and rating all stimuli, participants were asked for demographic information and thanked for their participation.

Speaker ID	Sex	Sentence - Chinese	Sentence – English Translation
F1	Female	今天晚上会下雨。	It's going to rain tonight.
M1	Male		
F2	Female	我们室友总是把寝室弄得很脏。	Our roommates always make the dorm room very messy.
M2	Male		

Table 1: Target sentences and speakers

Presentation

The Matched Guise Technique (Lambert, Hodgson, Gardner, & Fillenbaum 1960) traditionally presents a group of listeners with several recordings of the same text, each ostensibly produced by a different speaker. In actuality, some of the recordings are sets of two with each set produced by the same speaker using a different language or language variety. After listening to each recording, participants are asked to rate each speaker's voice on semantic differential scales of personal characteristics and language impressions that are expected to represent listeners' attitudes toward the groups thought to use these different language varieties. Use of the same speaker removes the possibility of ratings based on idiosyncratic variation naturally found between speakers. In contrast to traditional matched guise studies, the language samples presented to participants are not varied in the present work - participants rated the exact same language samples twice, but different information about the speaker's ethnicity is presented to participants for each "guise." A similar technique has been used in other matched guise studies with speaker ethnicity (McGowan 2015) and nationality (Hay, Nolan, & Drager, 2006; Niedzielski, 1999).

Participants were told, in written Mandarin, that researchers were interested in getting their impressions on how second language Chinese speech sounds. Prior to beginning the study, participants gave informed consent. Participants were then told that they would be listening to different speech samples, each recorded by a different, highly proficient, L2 Mandarin speaker. As to not give the impression that some stimuli were produced by heritage speakers, participants were told that none of these speakers were raised in a Chinese-speaking country, none began studying Chinese before the age of 14, and none learned Chinese from their families. Participants were told that half of the speakers were huayi, and half were non-huayi. Participants were accurately told that they would randomly be assigned to first hear the group of huayi speakers or

the group of non-huayi speakers. After rating each block, participants were advised to take a break of a few minutes before being presented with the remaining speech samples. Participants were told that they could listen to each sample as many times as they wanted but were asked to quickly respond to each question according to their first impression of each speaker. Participants were asked to wear earphones and to adjust the volume to a suitable level before beginning.

Participants heard ten sentences in each block. The ten sentences consisted of four matched guise sentences and six distractor sentences. Each sentence was presented as a video, with the visual component of the video indicating only the ethnicity of the speaker and the speaker number in black text against a white background. Participants rated each speaker on six personal characteristics and three language proficiency measures, listed in Table 2.

Participants were told that they would be asked to answer several questions about each speech sample. Prior to beginning, participants were provided with an example containing two adjectives that would later be included in the actual rating task: 没有礼貌的 *méiyǒu lǐmào de* "rude" and 有礼貌的 *yǒu lǐmào de* "polite." Participants were told the more polite their impression of a speaker, the closer to "polite" they should click, whereas the ruder their impression of a speaker, the closer to "rude" they should click.

In the actual rating task, participants clicked a "play" button to listen to each recording. Below the media player, they were presented with semantic differential scales corresponding to personal characteristics and language proficiency measures. Participants rated the speaker on each trait and then clicked the "next" button, which submitted their response, and brought them to the next page, which was identical to the previous page except for the embedded media.

Results

An exploratory factor analysis was conducted via principal component analysis in R (R Core Team 2021) on the semantic differential ratings; the analysis yielded two factors with eigenvalues above 1. A principal component analysis with varimax rotation and Kaiser

Scale / Component	**Rotated Comp 1**	**Rotated Comp 2**
Rude / Polite	**.75**	
Stupid / Smart	**.80**	
Shallow / Deep	**.83**	
Incompetent / Competent	**.82**	
Arrogant / Kind	**.74**	
Unlikeable / Likeable	**.77**	
Not Standard / Standard	.37	**.78**
Not Fluent / Fluent	.35	**.75**
Accented / Not Accented		**.72**

Table 2: Principal component analysis with varimax rotation and Kaiser normalization

normalization was then conducted using the psych package for R (Revelle 2020); the factor loadings are shown in Table 2. Rotated component 1 accounts for 45% of the variance, and rotated component 2 accounts for an additional 21% of the variance. An analysis of the factor loadings does not indicate grouping ratings along the dimensions of competence and warmth; rather, the semantic differential scales that have higher factor loadings for rotated component 1 include all of the scales related to personal characteristics, while the scales that have higher factor loadings for rotated component 2 include the scales related to language proficiency. Based on this analysis, personal characteristics and language proficiency ratings were aggregated and averaged for each speaker and separated by listener gender before cross-guise comparisons were conducted using paired t-tests.

Table 3 shows the full results of all paired t-tests. Significant differences were observed for speaker personal characteristics. Male participants preferred the huayi guises of both of the female speakers (Speaker F1: $M_{huayi} = 4.18$, $M_{non\text{-}huayi} = 4.02$, $t = 2.21$, $p < 0.05$; Speaker F2: $M_{huayi} = 3.90$, $M_{non\text{-}huayi} = 3.67$, $t = 3.42$, $p < 0.001$ for speaker F2). Conversely, the female

participants preferred the non-huayi guise, but only for speaker F1 (M_{huayi} = 4.01, $M_{non\text{-}huayi}$ = 4.27, t = -4.55, $p < 0.001$). Regardless of listener gender, mean language proficiency ratings are always higher than mean personal characteristics ratings for all speakers and guises. Also regardless of listener gender, differences in language proficiency ratings were not significant for any of the speakers, and differences in personal characteristics were not significant for either of the two male speakers.

Male listeners

Speaker	Factor	M_{huayi}	$M_{non\text{-}huayi}$	t	Df	p
F1	Personal characteristics	4.18	4.02	2.21	232	.0281*
	Language Proficiency	4.45	4.4	0.59	114	.5547
F2	Personal characteristics	3.90	3.67	3.42	232	.0007***
	Language Proficiency	3.97	4.01	-0.31	115	.7563
M1	Personal characteristics	3.80	3.81	-0.11	233	.9101
	Language Proficiency	4.14	4.17	-0.37	114	.7097
M2	Personal characteristics	3.78	3.70	1.05	232	.2932
	Language Proficiency	4.13	4.11	0.17	115	.8637

Female listeners

Speaker	Factor	M_{huayi}	$M_{non\text{-}huayi}$	t	Df	p
F1	Personal characteristics	4.01	4.27	-4.55	292	< .0001***
	Language Proficiency	4.35	4.38	-0.31	146	.7588
F2	Personal characteristics	3.74	3.72	0.40	293	.6912
	Language Proficiency	3.96	4.08	-1.19	144	.2341
M1	Personal characteristics	4.01	3.92	1.53	292	.126
	Language Proficiency	4.35	4.24	1.18	146	.2384
M2	Personal characteristics	3.91	3.92	-0.22	292	.8292
	Language Proficiency	4.16	4.13	0.34	146	.7355

Table 3: Paired t-tests, personal characteristics and language proficiency ratings, by perceived speaker ethnicity.
* $p < .05$ *** $p < .001$

Discussion

This study compared attitudes towards male and female L2 Mandarin speakers of huayi and non-huayi ethnicity by male and female Chinese raters. Results revealed that judgments of personal characteristics - which varied only along a single dimension - differed based on speaker ethnicity, speaker gender, and listener gender, but that judgments language proficiency ratings did not differ according to any of those factors. Male listeners rated both of the matched guise female speakers significantly higher in their huayi guises, while female listeners rated one of the two female speakers significantly higher in her non-huayi guise.

The difference in ratings based on the gender of the listener contrasts with the findings of Lu & Gnevsheva (2021), who recruited Chinese listeners as participants but who found no effect for listener gender in evaluations of their speakers' personal characteristics. A likely explanation for this contrast is the languages used in each study, consistent with the *cultural moderation of gender stereotypes hypothesis* (Cuddy et al. 2015). Following this hypothesis, English would have activated participants' Australian cultural norms and values in Lu & Gnevsheva's study, but Mandarin would have activated participants' Chinese cultural norms and values in the present study. The language of the study is a more likely explanation than a difference in study participants; while both Lu & Gnevsheva's participants and Squizzero's (2020) participants were multilinguals living in a Western country at the time of the survey, a mixture of monolingual and multilingual speakers residing in mainland China were surveyed here.

Ratings of language proficiency, including ratings of accentedness, did not differ by listener gender, nor did they differ by speaker gender or ethnicity. This result is consistent with Squizzero's (2020) study on L2 Mandarin, but it contrasts with results found in numerous studies

conducted on European languages in which speakers belonging to a minority ethnicity were rated as more accented than speakers belonging to a majority ethnicity (Rubin 1992; Babel & Russell 2015; Gnevsheva 2018; Lu & Gnevsheva 2021). One possible explanation for this contrast is that the experimental stimuli did not have enough variation in accentedness for participants to provide meaningful differences in ratings, although it bears mentioning that there was also no variation in accentedness between the auditory stimuli presented to participants in Rubin's study. Another, perhaps more likely explanation is that ethnicity does not cue expectations of foreign accentedness in Mandarin in the same way that it cues expectations of foreign accentedness in English, since L2-accented Mandarin is commonly spoken in China by ethnically Chinese individuals, often with other Sino-Tibetan L1s.

Gender and ethnicity

The findings of this study replicate Lu & Gnevsheva's (2021) work showing that gender and ethnicity can interact to affect the perception of L2-accented speakers. These results can be interpreted via intersectionality theory's idea that multiply minoritized individuals may face discrimination in ways different from their peers who are minorities only in one aspect of their identities (Crenshaw 1989). Future attitudinal studies should therefore be conscious of possible interactions between perceived ethnicity and perceived gender, among other perceptible social dimensions in which speakers may vary.

While the results show that male listeners rated female huayi L2 Mandarin speakers higher than their female non-huayi counterparts, it is unclear if the difference in ratings is due to positive feelings towards female huayi, negative feelings towards female non-huayi, or both. One promising framework for investigating the mechanism underlying the interaction between listener gender, speaker gender, and speaker ethnicity is ambivalent sexism (Glick & Fiske 1996;

2011). Ambivalent sexism proposes two separate but related attitudinal components: hostile sexism and benevolent sexism. Hostile sexism consists of negative attitudes about and stereotypes towards nontraditional women; these attitudes and stereotypes are based on a perception that nontraditional women pose a threat to men's power and identity. On the other hand, benevolent sexism, thus named because the individuals expressing this form of sexism consider their expressions to be benevolent in nature, consists of positive attitudes towards and stereotypes about traditional women. Benevolent sexist attitudes and stereotypes praise women's abilities to perform social roles traditionally filled by women (e.g. childcare) and therefore entail that women should be limited to these roles (e.g. stay at home parent). Both forms of sexism can be correlated with gender; one study found that Chinese men endorse more benevolent sexist and hostile sexist viewpoints than Chinese women (Lee et al. 2010). If the difference in ratings in the present study is due to positive feelings towards huayi women, it would be expected that participants would endorse more benevolent sexist viewpoints. A benevolent sexist lens would assume that a woman would be the ideal primary provider of care for her children; huayi women could be motivated to teach the language to their children so that they could be in touch with their culture, consistent with ethnolinguistic identity theory (Giles & Johnson 1987). On the other hand, a non-huayi woman would not be motivated to teach Mandarin to her children for cultural reasons; instead, it may be that she is perceived as learning Mandarin for career advancement, suggesting that she is a nontraditional woman. If the difference in ratings is due to negative feelings towards nontraditional, non-huayi women, it would be expected that participants would endorse more hostile sexist viewpoints; Glick & Fiske (1996) found that individuals who agree with hostile sexist viewpoints also tended to agree with racist viewpoints. In contrast to the male listeners, however, female listeners rated one of the two non-huayi

speakers significantly higher than her huayi guise. If female participants assume that the non-huayi speakers have learned Mandarin to a high level of proficiency for the purpose of career advancement, their preference for the non-huayi speaker could reflect appreciation or respect for this nontraditional woman, an antisexist position.

Cultural Moderation

The results provide evidence that the supposedly cross-cultural attitudinal differences (c.f. Osgood et al., 1975) may not apply to second language speakers outside of Western contexts. The apparent unidimensional nature of the stereotype content in terms of personal characteristics is consistent with the *cultural moderation of gender stereotypes hypothesis* (Cuddy et al. 2015). According to this hypothesis, in collectivist cultures, such as Chinese culture, status and solidarity are not necessarily perceived as separate dimensions in the same way that they are in individualist cultures, such as American culture.

Study Populations

Prior exploratory work (Squizzero 2020) was conducted with a small sample of L1 Chinese speakers in the diverse city of Seattle, Washington. The present study provides some evidence for the argument that there were characteristics unique to the population surveyed in that study. In the exploratory study, it was speculated that the participants' demonstration of ethnic bias against only one speaker may have been a product of their empathy for L2 speakers – all participants in that sample were proficient L2 speakers of English who lived outside of the cultural context in which they were raised. However, it was also speculated that this empathy may have been the reason why there was no difference in accentedness ratings of the speakers, yet the present study's result of no significant differences in accentedness based on perceived speaker ethnicity is inconsistent with this explanation. In the context of these new results, it may

be that there was simply not enough variation in the accentedness present in the stimuli; future work could test this by using accented stimuli from actual L2 Mandarin speakers. Lastly, it was speculated that there could have been more social desirability bias in the population of Squizzero (2020), and that this bias, rather than cognitive load, could have contributed to the effects being significant only for one of the four target speakers in that study. This is somewhat supported by the present study, as male raters significantly favored the huayi guise for both female speakers.

Conclusion

In addition to confirming prior work establishing the role of perceived ethnicity and gender in the treatment of second language speakers, this paper has demonstrated the importance of cultural context in the study of the perceived accentedness and personal characteristics of multilinguals. Unlike similar studies conducted in the past which have attempted to replicate Rubin's (1992) work, this study was conducted on L2 Mandarin speakers in Asia, not on speakers of European languages in predominantly White cultural contexts. In this paper, the listeners' cultural context was majority Asian, not White, which reflects reality for a greater share of the world's population for those who live in majority-White cultural contexts. Crucially, unlike in prior studies, only perception of speakers' personal characteristics, not their accentedness, varied based on their ethnicity. While some past studies have found no effect for either accentedness or personal characteristics ratings (Eisenchlas & Michael 2019), there appear to be none which have found only an effect for personal characteristics ratings but not for accentedness.

This effect of speaker ethnicity on perceived personal characteristics was only observed for female speakers, and the direction of the effect depended on listener gender, with men rating huayi speakers higher in personal characteristics than non-huayi speakers, but with women rating

a non-huayi speaker higher in personal characteristics than a huayi speaker. This result suggests that researchers should addition not only investigate non-Western populations and speakers of non-European languages but also should investigate potential effects of both speaker and listener gender. Furthermore, an exploratory factor analysis revealed that listeners only evaluated speakers' personal characteristics along one dimension of affective meaning, rather than two or three dimensions. Rather than assuming a three-factor attitude structure *a priori*, researchers should heed Zahn & Hopper's (1985) warning and test if their raters are truly using multiple dimensions of affective meaning before proceeding with their analyses.

For quite some time, it has been difficult to dispute the fact that social characteristics affect perception of L2 speakers, often with real-world consequences. The effect of social characteristics, however, clearly varies depending on social and cultural context. A truly accurate understanding of perception of L2 speakers and their speech only can be achieved by broadening the share of the world's languages and cultures under investigation, and by doing so in a way that takes into consideration individuals' multiplex social identities.

Acknowledgements
Section to be added upon receiving permission from acknowledged individuals.

Declaration of Interest Statement
None.

Excursus: Ambivalent sexism affects attitudes towards female second language Mandarin speakers of different ethnicities

This excursus is an exploratory study which was conducted to follow up on the main results of this paper. The excursus appears only in the dissertation chapter version of this article.

Chapters 1 and 2 (Squizzero n.d.; 2020) showed that judgments of personal characteristics of L2 Mandarin Chinese speakers are moderated by perceived ethnicity. Chapter 2 also demonstrated that ethnicity as a moderating factor still applies in a Chinese social and

cultural context, in addition to a Chinese linguistic context. In the studies presented in both chapters, however, this ethnic moderation was only present against female speakers, and in Chapter 2, it was shown to have been stronger for male listeners. The discussion section of Chapter 2 called for future identification of the mechanism behind the gendered nature of the ethnicity-based discrimination in evidence, following *intersectionality theory* (Crenshaw 1989). The empirical study that is presented in this excursus grounds itself in *ambivalent sexism theory* (AST; Glick & Fiske, 1996, 2011), exploring ambivalent sexism as a possible mechanism behind the gendered nature of attitudes towards L2 Mandarin speakers. Of particular theoretical importance, this is the first known study conducted on language and social psychology to successfully demonstrate that ambivalent sexism can be a moderating factor in attitude formation.

To reprise and expand upon the discussion section of Chapter 2, in AST, sexism is classified into two separate but related attitudinal components: *hostile sexism* and *benevolent sexism*. Hostile sexism consists of negative attitudes about and stereotypes towards nontraditional women; these attitudes and stereotypes are based on a perception that nontraditional women pose a threat to men's power and identity. Benevolent sexism, in contrast, consists of positive attitudes towards and stereotypes about traditional women. Benevolent sexist attitudes and stereotypes praise women's abilities to perform social roles traditionally filled by women (e.g. childcare) and therefore entail that women should be limited to these roles (e.g. stay at home parent). An example of benevolent sexism is agreement with the statement, "Women, as compared to men, tend to have a superior moral sensibility," (Glick & Fiske, 1996, p. 512) and an example of hostile sexism is agreement with the statement, "Women exaggerate problems they have at work" (Glick & Fiske, 1996, p. 512).

AST was originally developed in the United States for use in a Western context, not a Chinese context. However, ambivalent sexism theory has been used successfully in studies of Chinese populations; in fact, one study of American and Chinese undergraduate students found that Chinese undergraduates more strongly endorsed hostile and benevolent sexist viewpoints than their American counterparts (Lee, Fiske, Glick & Chen 2010). In both populations in the study by Lee and colleagues, men more strongly endorsed both hostile and benevolent sexist viewpoints, on average, relative to women.

AMBIVALENT SEXISM AND FOREIGNNESS. Any adult who has attempted to learn a second language to a high degree of proficiency will attest that it is a time-consuming endeavor; it takes an average adult learner of a second (or additional) language hundreds or thousands of hours of study and practice just to reach just professional working proficiency (Foreign Service Institute - United States Department of State 2022). Reaching a high level of proficiency could therefore be viewed an impressive accomplishment and indicative of a woman's non-fulfillment of traditional social roles. But if it is simply the case that there is something threatening about a woman, as opposed to a man, who has learned to speak Mandarin well despite having been raised outside of a Chinese-speaking country, then there should be no difference in how ethnic ingroup women are perceived relative to ethnic outgroup women; all women in the study should be perceived equally. Why, then, did the sex of the speaker appear to interact with the perceived ethnicity of a speaker in earlier work (Squizzero n.d.; 2020)?

One possible explanation is the association between language, culture, and ethnicity. Because ethnic Chinese (huáyì) women are learning the language associated with their ethnic group, attempting to get in touch with their culture may be a perceived motivation for learning Mandarin that would not exist for ethnic non-Chinese (non-huáyì) women, consistent with

ethnolinguistic identity theory (Giles & Johnson 1987). Further, from a benevolent sexist viewpoint, the idea that a woman should be the primary provider of care for her children could also point to a motivation of wanting to teach the language to her children so that the children can be in touch with their cultural background. Another possible explanation is that there is an interaction between of ethnocentrism and hostile sexism in operation. Glick & Fiske (1996) found a correlation between modern racism and sexism, especially hostile sexism. A lack of ethnocentrism against men could indicate that the bias against women and the bias against ethnic outgroup members is additive, and that the sample size in Chapters 1 and 2 was too small to identify either effect in isolation. Additionally, the results in the main experiment of Chapter 2 showed that female participants actually rated the non-huáyì guise significantly higher than the huáyì guise for one of the two female speakers in that study's experiment. This preference for the non-huáyì guise could reflect appreciation or respect for a nontraditional woman if female participants believed that the non-huáyì speakers have learned Mandarin to a high level of proficiency for the purpose of career advancement.

Using a similar methodology to the main experiment in Chapter 2, the present study investigates a potential interaction between type of sexism and ethnocentrism. Benevolent sexism would predict ingroup favoritism via positive expectations about huáyì women via their motivations for learning the language. Hostile sexism, on the other hand, would predict an additive effect of outgroup derogation, with non-huáyì women being especially different from the default person in Chinese society, considering that the default person in Chinese society is assumed to be male and ethnically Chinese.

MATERIALS. The same female L1 Mandarin stimuli used in the main experiment of Chapter 2, selected from the Mandarin Affective Speech Corpus (MASC)(Y. Yang et al. 2007), were used

to test the hypothetical situation in which an L2 speaker has acquired a pronunciation of Mandarin similar to that of L1 speakers and to remove proficiency as a potential confound. A total of 8 recorded utterances of "neutral" affect from 8 different female speakers were used. Male speakers were excluded from the study because the studies in Chapter 1 and 2 did not show an effect for ethnicity in judgments of male speakers. As in Chapters 1 and 2, two speakers were presented as matched guises, while the remaining utterances were included as distractors. Ten recordings were presented to participants in total; five of the recordings were falsely presented to participants as having been produced by huáyì L2 Mandarin speakers and the other five recordings were falsely presented to participants as produced by non-huáyì L2 speakers, but there were no actual linguistic differences between the distractor sentences in each group presented to listeners, and there were no differences at all between the target (matched guise) sentences.

PERSONALITY TRAITS AND ACCENTEDNESS. Participants rated each guise on impressions of personal characteristics along six-point semantic differential scales. The semantic differential scales in the present study were: arrogant/kind, rude/polite, rarely nags/nags often, unambitious/ambitious, and for target sentence 2 only ("Our roommates always make the dorm room very messy"), messy/neat. Participants also rated each speaker's foreign accentedness. All semantic differential scales were arranged by placing the trait with negative valence on the left and positive valence on the right, except for the accented/not accented and rarely nags/often nags scale, the scores of which were transformed to match those of the other scales. After listening to and rating all stimuli, participants were asked for demographic information and thanked for their participation.

THE AMBIVALENT SEXISM INVENTORY. A shortened, 12-item version of the ambivalent sexism inventory (ASI), translated into Chinese by Lee et al. (2010), was completed by all participants after rating the language samples but before completing the demographic questionnaire. Participants were asked to rate their level of agreement on a six-point scale (0 = *disagree strongly*, 5 = *agree strongly*), and hostile and benevolent sexism scores were subsequently calculated, with a minimum score of 0 and a maximum score of 30 for each type of sexism.

In addition to the ASI, participants indicated their agreement/disagreement on a scale of 0 to 5 on the following items, which were included to directly test possible rationales for personality ratings that vary by ethnicity: 1) A woman should have ambition towards her career. 2) A foreign huáyì should understand Chinese culture. 3) A mother should teach her native language to her children. 4) A huáyì mother should teach Chinese language to her children. 5) A woman who learns a foreign language to a high level of proficiency is an ambitious person. 6) When someone from another culture learns your language, you feel a sense of pride.

PARTICIPANTS. A total of 157 first language Mandarin speakers were recruited via the online survey platform *Wenjuan* (https://www.wenjuan.com; 问卷网). Participants were required to have been raised in Mainland China, of self-reported native Mandarin proficiency, and have no known speech or hearing disorders. Participants were raised in a mixture of hometowns from a range of geographic areas of China.

ATTENTION AND MANIPULATION CHECKS. After reading the initial instructions, participants were given a multiple-choice question in which they were asked to the indicate the age of acquisition of Mandarin language of the speakers they were about to listen to. 62 participants failed this attention check and were excluded from subsequent analyses. Later, after listening to each block of speakers, participants were asked to indicate the ethnicity of the speakers that they had just

listened to. 63 participants failed one or both of the manipulation checks and were excluded from subsequent analyses.

PARTICIPANT DEMOGRAPHIC INFORMATION. After excluding participants who failed the attention and/or manipulation checks, 29 participants remained. Another participant was excluded for indicating that they were under 18 years of age. The remaining 28 participants were all monolingual Mandarin speakers of Han ethnicity, ranging in age from 20-52 years (mean: 28.25), and consisting of 13 men and 15 women (0 neither).

ANALYSIS. An index, hereafter referred to as a matched guise index (MGI), was calculated for accentedness and personality traits. MGI for accentedness (MGI_{Accent}) is the sum of the accentedness scores (on a scale of 1-6) for each speaker in their huáyì guise minus the accentedness scores for each speaker in their non-huáyì guise:

$$MGI_{Accent} = Speaker1_{huáyì} + Speaker2_{huáyì} - Speaker1_{non\text{-}huáyì} - Speaker2_{non\text{-}huáyì}$$

MGI for personality traits is calculated the same way as it is for accentedness, except that the scores for the five personality traits were added together for each speaker and guise. Personality traits are combined into a single factor based on the exploratory factor analysis presented in the main experiment of Chapter 2.

RESULTS. Mean ambivalent sexism index scores, separated by respondent gender, are presented in Table 1. The ratings indicate that endorsement of sexist views is gendered, with men more strongly endorsing both benevolent and hostile sexist viewpoints as compared to women, consistent with the findings of Lee et al (2010). T-tests revealed that gender differences are significant for both hostility (t=2.08, df=25.69, p=0.048, 95% CI=0.04:8.87) and benevolence (t=2.29, df=23.584, p=0.031, 95% CI=0.43:8.29). Inconsistent with Lee et al., however, is that

respondents in the present study reported greater endorsement of benevolent sexist viewpoints, as opposed to hostile sexist viewpoints, regardless of gender.

Means for each scale and guise are presented in Table 2 and are represented graphically in Figure 1. Ratings were similar for both guises and slightly higher for the non-huáyì guise. The messy/neat scale showed the greatest discrepancy between guises, with the non-huáyì guise rated one half-point higher than the huáyì guise.

Attitudes towards personality and accentedness for each speaker were examined using two linear models, with $MGI_{Personality}$ as the dependent variable in one model and MGI_{Accent} as the dependent variable in the other. Independent variables included the respondent's gender, hostile sexism score, benevolent sexism score, and their agreement with each of six possible statements.

	Women	Men	Overall
Benevolence	15.3	19.7	17.4
Hostility	13.5	17.9	15.5
Ambivalent	28.8	37.6	32.9

Table 4 Mean benevolent, hostile, and ambivalent sexism scores, by respondent gender. Ambivalent sexism scores represent the sum of the benevolent and hostile sexism scores for each column.

Semantic differential scale	huáyì	non-huáyì
Arrogant/kind	3.48	3.51
Messy/neat	3.61	4.11
Rarely nags/often nags*	3.31	3.47
Rude/polite	3.56	3.61
Unambitious/ambitious	3.44	3.49
Unaccented/accented*	3.23	3.24

Table 5 Mean ratings, by semantic differential scale and guise.
** indicates that a scale has been transformed*

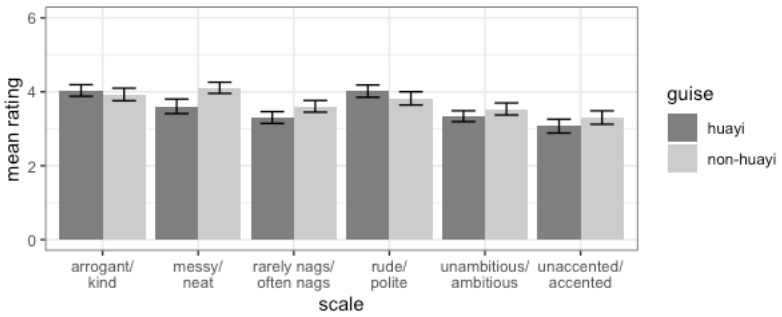

Figure 1 Mean ratings by semantic differential scale with error bars, by guise.

A likelihood ratio test comparing null and saturated models for $MGI_{Personality}$ indicated a significant improvement in model fit for the saturated model (χ^2=16.92, df=9, p=0.049). The coefficients for this linear model are reported in Table 3. According to the model, the significant factors affecting preference for one guise or another are hostile sexism and the belief that a foreign huáyì should understand Chinese culture. Gender was not a significant factor in the model. Respondents with higher hostile sexism scores rated the huáyì guise higher, with each single-point increase on the hostile sexism index associated with a 1.2-point increase in $MGI_{Personality}$. Conversely, respondents who more strongly agreed that a foreign huáyì should understand Chinese culture rated the huáyì guise lower, with each single-point increase in agreement with this statement associated with a 4.2-point decrease in $MGI_{Personality}$.

A likelihood ratio test comparing null and saturated models for MGI_{Accent} did not show a significant improvement in model fit (χ^2=10.315, df=9, p=0.3256).

	Estimate	Std. Error	t value	Pr(>|t|)
(Intercept)	-28.6055	20.6789	-1.383	0.18349
gender	9.1701	4.9302	1.860	0.07931 .
hostile.sexism.index	1.2172	0.4041	3.012	0.00749 **
benevolent.sexism.index	-0.3107	0.4687	-0.663	0.51583
career.aspiration.women	0.8062	1.9839	0.406	0.68925
understand.culture	-4.1896	1.7513	-2.392	0.02786 *
mother.children.L1	2.2793	2.8925	0.788	0.44096
huayi.mother.children.Chinese	-3.4417	3.4678	-0.992	0.33413
advanced.L2.women.career.aspiration	0.7999	2.6092	0.307	0.76271
L2.honor.culture	3.0422	2.8011	1.086	0.29179

*Table 6: Summary for model of personality ratings. Significance codes: ** p < .01, * p < .05, p < .1*

CONCLUSION. Results indicate that hostile sexism, rather than benevolent sexism or respondent gender, significantly contributes to differences in perception of the personal characteristics of L2 Mandarin speakers of Chinese and non-Chinese ethnicity. The significant effect of hostile sexism suggests that the mechanism underlying ethnicity as a moderating factor in attitudes towards L2 Mandarin-speaking women is outgroup derogation towards non-huáyì women. If the underlying

mechanism were ingroup favoritism through the fulfillment of traditional gender roles, a significant effect for benevolent sexism would be expected.

The addition of ambivalent sexism theory to the present line of research has an added benefit in that it moves beyond an essentialist conceptualization of gender. The fact that hostile sexism, not listener gender, mediated bias in favor of ethnic ingroup women, means that it should not be concluded that Chinese men simply discriminate against ethnic outgroup women whereas Chinese women do not. Future interventions aimed at reducing hostile sexism among listeners could therefore have a benefit to L2 Mandarin-speaking women, if not L2 speakers in general.

References

Babel, Molly & Jamie Russell. 2015. Expectations and speech intelligibility. *The Journal of the Acoustical Society of America* 137(5). 2823–2833. DOI: https://doi.org/10.1121/1.4919317

Banaji, Mazharin R. & Larisa Heiphetz. 2010. Attitudes. In Susan T. Fiske, Daniel T Gilbert, & Gardner Lindzey (eds.), *Handbook of Social Psychology*, 353–393. Hoboken, NJ: John Wiley & Sons.

Canagarajah, Suresh. 2015. Clarifying the relationship between translingual practice and L2 writing: Addressing learner identities. *Applied Linguistics Review* 6(4). 415–440. DOI: https://doi.org/10.1515/applirev-2015-0020

Crenshaw, Kimberle. 1989. Demarginalizing the intersection of race and sex: A black feminist critique of antidiscrimination doctrine, feminist theory, and antiracist politics. *Living With Contradictions: Controversies In Feminist Social Ethics* 1. 139–167. DOI: https://doi.org/10.4324/9780429499142-5

Cuddy, Amy J.C., Elizabeth Baily Wolf, Peter Glick, Susan Crotty, Jihye Chong & Michael I. Norton. 2015. Men as cultural ideals: Cultural values moderate gender stereotype content. *Journal of Personality and Social Psychology* 109(4). 622–635. DOI: https://doi.org/10.1037/pspi0000027

Eberhard, David M., Gary F. Simons & Charles D. Fennig (eds.). 2021. *Ethnologue: Languages of the World.* Twenty-fou. Dallas, Texas: SIL International.

Eisenchlas, Susana A. & Rowan B. Michael. 2019. What's in a face? The impact of nonlinguistic 'ethnic' facial features on accent perception. *Journal of Multilingual and Multicultural Development* 40(10). 879–891. DOI: https://doi.org/10.1080/01434632.2019.1587445

Foreign Service Institute - United States Department of State. 2022. Foreign Language Training. (https://www.state.gov/foreign-language-training/) (Accessed 2022-5-13)

Fuertes, Jairo N., William H. Gottdiener, Helena Martin, Tracey C. Gilbert & Howard Giles. 2012. A meta-analysis of the effects of speakers' accents on interpersonal evaluations. *European Journal of Social Psychology* 42(1). 120–133. DOI: https://doi.org/10.1002/ejsp.862

Giles, Howard & P. Johnson. 1987. Ethnolinguistic identity theory: A social psychological approach to language maintenance. *International Journal of the Sociology of Language* 68. 66–99.

Glick, Peter & Susan T. Fiske. 1996. The Ambivalent Sexism Inventory: Differentiating Hostile and Benevolent Sexism. *Journal of Personality and Social Psychology* 70(450). 491–512.

Glick, Peter & Susan T. Fiske. 2011. Ambivalent sexism revisited. *Psychology of Women Quarterly* 35(3). 530–535. DOI: https://doi.org/10.1177/0361684311414832

Gnevsheva, Ksenia. 2018. The expectation mismatch effect in accentedness perception of Asian and Caucasian non-native speakers of English. *Linguistics* 56(3). 581–598. DOI: https://doi.org/10.1007/BF00973770

Hanulíková, Adriana. 2018. The effect of perceived ethnicity on spoken text comprehension under clear and adverse listening conditions. *Linguistics Vanguard* 4(1). 1–9. DOI: https://doi.org/10.1515/lingvan-2017-0029

Hanulíková, Adriana. 2021. Do faces speak volumes? Social expectations in speech comprehension and evaluation across three age groups. *PLoS ONE* 16(10 October). DOI: https://doi.org/10.1371/journal.pone.0259230

Hay, Jennifer, Aaron Nolan & Katie Drager. 2006. From fush to feesh: Exemplar priming in speech perception. *Linguistic Review* 23(3). 351–379. DOI: https://doi.org/10.1515/TLR.2006.014

Henrich, Joseph, Steven J. Heine & Ara Norenzayan. 2010. The weirdest people in the world? *Behavioral and Brain Sciences* 33(2–3). 61–83. DOI: https://doi.org/10.1017/S0140525X0999152X

Kang, Okim & Donald L. Rubin. 2009. Reverse linguistic stereotyping: Measuring the effect of listener expectations on speech evaluation. *Journal of Language and Social Psychology* 28(4). 441–456. DOI: https://doi.org/10.1177/0261927X09341950

Lambert, W.E., R.C. Hodgson, R.C. Gardner & S. Fillenbaum. 1960. Evaluational reactions to spoken languages. *Journal of Abnormal and Social Psychology* 60(1). 44–51.

Lee, Tiane L., Susan T. Fiske, Peter Glick & Zhixia Chen. 2010. Ambivalent sexism in close relationships: (Hostile) power and (benevolent) romance shape relationship ideals. *Sex Roles* 62(7–8). 583–601. DOI: https://doi.org/10.1007/s11199-010-9770-x

Levon, Erez. 2015. Integrating Intersectionality in Language, Gender, and Sexuality Research. *Language and Linguistics Compass* 9(7). 295–308. DOI: https://doi.org/10.1111/LNC3.12147

Liao, Silvie. 2008. A perceptual dialect study of Taiwan Mandarin: Language attitudes in the era of political battle. *Proceedings of The 20th North American Conference on Chinese Linguistics (NACLL-20). Vol. 1* 1. 391–408.

Lin, Yuhan. 2018. *Stylistic variation and social perception in second dialect acquisition*. The Ohio State University dissertation.

Lippi-Green, Rosina. 2012. *English with an accent : language, ideology and discrimination in the United States* 2nd ed. Routledge. Retrieved from https://www.routledge.com/English-with-an-Accent-Language-Ideology-and-Discrimination-in-the-United/Lippi-Green/p/book/9780415559119

Lu, Yao & Ksenia Gnevsheva. 2021. Accentedness and personality evaluation of Asian and Caucasian second language speakers of English by Asian second language English listeners. *Journal of Multilingual and Multicultural Development*. DOI: https://doi.org/10.1080/01434632.2021.1959598

Markus, Hazel Rose & Shinobu Kitayama. 1991. Culture and the self: Implications for cognition, emotion, and motivation. *Psychological Review* 98(2). 224–253. DOI: https://doi.org/10.1037/0033-295x.98.2.224

McGowan, Kevin B. 2015. Social Expectation Improves Speech Perception in Noise. *Language and Speech* 58(4). 502–521. DOI: https://doi.org/10.1177/0023830914565191

McKenzie, Robert M., Patchanok Kitikanan & Phaisit Boriboon. 2016. The competence and warmth of Thai students' attitudes towards varieties of English: the effect of gender and perceptions of L1 diversity. *Journal of Multilingual and Multicultural Development* 37(6). 536–550. DOI: https://doi.org/10.1080/01434632.2015.1083573

Mei, Bing & Gavin T.L. Brown. 2018. Conducting Online Surveys in China. *Social Science Computer Review* 36(6). 721–734. DOI: https://doi.org/10.1177/0894439317729340

Nelson, Larry R., Margaret L. Signorella & Karin G. Botti. 2016. Accent, Gender, and Perceived Competence. *Hispanic Journal of Behavioral Sciences* 38(2). 166–185. DOI: https://doi.org/10.1177/0739986316632319

Niedzielski, Nancy A. 1999. The Effect of Social Information on the Phonetic Perception of Sociolinguistic Variables. *Journal of Language and Social Psychology* 18(1). 62–85.

Osgood, Charles E., William H. May & Murray S. Miron. 1975. *Cross-cultural universals of affective meaning*. University of Illinois Press.

Osgood, Charles E., G.J. Suci & P.H. Tannenbaum. 1957. *The measurement of meaning*. Urbana: University of Illinois Press.

Peng, Chun-Yi. 2016. *Talk Shows and Language Attitudes: A Sociolinguistic Investigation of Language Attitudes towards Taiwan Mandarin among Chinese Mainlanders*. The City University of New York dissertation.

Purdie-Vaughns, Valerie & Richard P. Eibach. 2008. Intersectional invisibility: The distinctive advantages and disadvantages of multiple subordinate-group identities. *Sex Roles* 59(5–6). 377–391. DOI: https://doi.org/10.1007/s11199-008-9424-4

R Core Team. 2021. R: A Language and Environment for Statistical Computing. Vienna, Austria: R Foundation for Statistical Computing.

Revelle, William. 2020. psych: Procedures for Psychological, Psychometric, and Personality Research.

Rubin, Donald L. 1992. Nonlanguage factors affecting undergraduates' judgments of nonnative English-speaking teaching assistants. *Research in Higher Education* 33(4). 511–531. DOI: https://doi.org/10.1007/BF00973770

Squizzero, Robert. n.d. Perceived ethnicity, gender, and attitudes towards second language Mandarin speakers. *Journal of Multilingual and Multicultural Development*.

Squizzero, Robert. 2020. Attitudes toward L2 Mandarin Speakers of Chinese and non-Chinese Ethnicity. In Kaidi Chen (ed.), *Proceedings of the 32nd Meeting of the North American Conference on Chinese Linguistics*, 521–538. Storrs, CT.

Tan, Lekun. 2016. A Comparison of Language Awareness and Attitudes between Students in the Department of Taiwanese Literature and Other Departments. *Journal of Taiwanese Languages and Literature* 11(2).

Yang, Chunsheng. 2014. Language attitudes toward Northeastern Mandarin and Putonghua (PTH) by young professionals. *Chinese Language and Discourse* 5(2). 211–230. DOI: https://doi.org/10.1075/cld.5.2.04yan

Yang, Yingchun, Dongdong Li & Zhaohui Wu. 2007. Mandarin Affective Speech Corpus. Philadelphia: Linguistic Data Consortium. Retrieved from

https://catalog.ldc.upenn.edu/LDC2007S09
Yook, Cheongmin & Stephanie Lindemann. 2013. The role of speaker identification in Korean university students' attitudes towards five varieties of English. *Journal of Multilingual and Multicultural Development* 34(3). 279–296. DOI: https://doi.org/10.1080/01434632.2012.734509
Zahn, Christopher J. & Robert Hopper. 1985. Measuring language attitudes: the speech evaluation instrument. *Journal of Language and Social Psychology* 4(2). 113–123.

Availability of data

The data that support the findings of this study are available from the corresponding author upon reasonable request.

Ethics approval

The University of Washington Human Subjects Division has determined that this study, #00007723, is exempt from Institutional Review Board approval.

Tables

Speaker ID	Sex	Sentence - Chinese	Sentence – English Translation
F1	Female	今天晚上会下雨。	It's going to rain tonight.
M1	Male		
F2	Female	我们室友总是把寝室弄得很脏。	Our roommates always make the dorm room very messy.
M2	Male		

Table 7: Target sentences and speakers

Scale / Component	**Rotated Comp 1**	**Rotated Comp 2**
Rude / Polite	**.75**	
Stupid / Smart	**.80**	
Shallow / Deep	**.83**	
Incompetent / Competent	**.82**	
Arrogant / Kind	**.74**	
Unlikeable / Likeable	**.77**	
Not Standard / Standard	.37	**.78**
Not Fluent / Fluent	.35	**.75**
Accented / Not Accented		**.72**

Table 8: Principal component analysis with varimax rotation and Kaiser normalization

Male listeners

Speaker	Factor	M_{huayi}	$M_{non\text{-}huayi}$	t	Df	p
F1	Personal characteristics	4.18	4.02	2.21	232	.0281*
	Language Proficiency	4.45	4.4	0.59	114	.5547
F2	Personal characteristics	3.90	3.67	3.42	232	.0007***
	Language Proficiency	3.97	4.01	-0.31	115	.7563
M1	Personal characteristics	3.80	3.81	-0.11	233	.9101
	Language Proficiency	4.14	4.17	-0.37	114	.7097
M2	Personal characteristics	3.78	3.70	1.05	232	.2932
	Language Proficiency	4.13	4.11	0.17	115	.8637

Female listeners

Speaker	Factor	M_{huayi}	$M_{non\text{-}huayi}$	t	Df	p
F1	Personal characteristics	4.01	4.27	-4.55	292	< .0001***
	Language Proficiency	4.35	4.38	-0.31	146	.7588
F2	Personal characteristics	3.74	3.72	0.40	293	.6912
	Language Proficiency	3.96	4.08	-1.19	144	.2341
M1	Personal characteristics	4.01	3.92	1.53	292	.126
	Language Proficiency	4.35	4.24	1.18	146	.2384
M2	Personal characteristics	3.91	3.92	-0.22	292	.8292
	Language Proficiency	4.16	4.13	0.34	146	.7355

Table 9: Paired t-tests, personal characteristics and language proficiency ratings, by perceived speaker ethnicity.
* $p < .05$ *** $p < .001$

	Women	Men	Overall
Benevolence	15.3	19.7	17.4
Hostility	13.5	17.9	15.5
Ambivalent	28.8	37.6	32.9

Table 10 Mean benevolent, hostile, and ambivalent sexism scores, by respondent gender. Ambivalent sexism scores represent the sum of the benevolent and hostile sexism scores for each column.

Semantic differential scale	huáyì	non-huáyì
Arrogant/kind	3.48	3.51
Messy/neat	3.61	4.11
Rarely nags/often nags*	3.31	3.47
Rude/polite	3.56	3.61
Unambitious/ambitious	3.44	3.49
Unaccented/accented*	3.23	3.24

Table 11 Mean ratings, by semantic differential scale and guise.
* indicates that a scale has been transformed

[i] The distinction between first and second language speakers in this paper is made solely because certain elements of spoken language almost always systematically vary between individuals who have begun to acquire Mandarin during the critical period of language acquisition and individuals who have acquired Mandarin after the critical period. The distinction between first and second language speakers in many contexts is problematically reductionist in today's transnational, translingual world. For a thorough discussion of these issues, see (Canagarajah 2015).

[ii] The authors of MASC elicited respondent sex, not respondent gender.

The role of dialectology in L2 vowel acquisition; evidence from Mandarin Chinese

Robert Squizzero
University of Washington

Department of Linguistics, Guggenheim Hall 4th Floor, Box 352425
University of Washington, Seattle, WA, USA, 98195
rsquizz@uw.edu

1.1 ABSTRACT

Deciding which segments to teach in the second language (L2) pronunciation classroom is often accomplished by quick comparison of phonemic inventories between the learners' first (L1) and second languages. These decisions, however, often overlook salient dialectal and sociolectal variation in both L1 and L2. Auditory and time-varying acoustic analyses of 5942 Mandarin Chinese segments indicate that significant sub-phonemic differences exist between L1 English and L1 Mandarin speakers; generalized additive mixed models (GAMMs) show that /a/ is lower and backer and /ei u/ are fronter for L2 speakers despite the ostensible similarity of the vowels between Mandarin and American English. The GAMMs also show that there are significant differences in the dynamics of /a y iŋ/ between L1 and L2 speakers. Given that sub-phonemic differences have been shown to interfere with speakers' intelligibility and result in increased perceived accentedness, teachers of Mandarin pronunciation should consider adjusting their curricula to teach pronunciation of the vowels mentioned in this study. More importantly, the results of this study demonstrate that second language learners and teachers would benefit from taking a critical view of the sounds that differ between the learners' L1s and L2s, a view that goes beyond simple comparison of phonemic inventories. (199 words)

1.2 Introduction and Background

Dialectal and sociolectal differences both in learners' L1s and in their target L2s are often overlooked in second language pronunciation research and teaching. While a purely phonemic approach to pronunciation learning and teaching, based on the Contrastive Analysis Hypothesis (Lado 1957) has benefits in terms of simplicity, sub-phonemic differences between language varieties, such as dialectal differences, can prevent the faithful acquisition of L2 sounds, adversely affecting speaker intelligibility (Porretta & Tucker 2015), accentedness (Wu 2011; McCullough 2013), and comprehensibility [Squizzero, next chapter]. Part of the reason why sub-phonemic differences are overlooked is because of the uncritical use of traditional transcription symbols of L1 and L2 phonemes. One widely cited description of American English lists its vowel inventory as /i ɪ e ɛ æ a ɔ o ʊ u ʌ ɝ/ (Hillenbrand et al. 1995), and one widely cited description of Mandarin lists its monophthongal vowel inventory as /i y u ə ɤ a/ (Lee & Zee 2003). The phonetic realizations of phonemic vowels, however, often differ from the symbols chosen to represent them, due to regional and social differences affecting language varieties, as well as the passage of time. Conceiving of sounds simply in terms of the phonetic values assigned to traditional transcription symbols instead of considering the actual phonetic values present both in the L1 of the learner and the target variety of the L2 risks the obscuring of salient sub-phonemic differences. Based on a comparison of the aforementioned vowel inventories of English and Mandarin, one could reasonably conclude that the phonemic monophthongs common to both languages are /a i u/. However, the vast majority of speakers recorded in the study by Hillenbrand, Getty, Clark, and Wheeler were raised in geographic areas in which the English spoken is characterized by changes associated with the Northern Cities Chain shift, which involves fronting of the low vowel which is transcribed in other sources as the /ɑ/, a geographically restricted phenomenon (Labov 1994),

but does not involve fronting of /u/, a geographically widespread phenomenon (Labov, Ash & Boberg 2006). While Hillenbrand and colleagues very careful to point out that the realization of phonemes in their study represent productions characterized by the time and place in which their speakers were raised, their results are often overgeneralized by others as if they represent the speech of all Americans, perhaps due to the misleading nature of the article's title, "Acoustic characteristics of American English vowels."

1.2.1 *Existing work on L2 Mandarin vowels and consonants*

Although most work in acquisition of L2 phonology has been conducted on consonants and vowels, most of the work that has been done on L2 Mandarin phonology has focused on lexical tone (Zhang 2018). A handful of acoustic studies, however, have been conducted on L2 Mandarin vowels.

One acoustic study examined Mandarin and English vowels produced both by L1 English L2 Mandarin speakers and L1 Mandarin L2 English speakers (Shi & Wen 2009). Of the four L1 English speakers included in the study, two were from Utah, and the other two were from Kansas. Shi and Wen noticed that the tokens of English /u/ produced by the Kansans were fronted relative to those produced by the Utahns, and the authors attributed this to dialectal variation. L1 English speakers' productions of Mandarin /u/ had lower mean F2 relative to their English /u/, but the F2 of their Mandarin /u/ productions were still lower than /u/ produced by L1 Mandarin speakers. L1 English speakers also had lower mean F1 productions of /a/ in both languages relative to L1 Mandarin speakers. The L1 English speakers also had slightly lower mean F2 productions of /a/ in English relative to their /a/ productions in Mandarin. These comparisons, unfortunately, are questionable, as the vowel measurements were not normalized, and the Mandarin speakers seemed to have a larger vowel space overall, which could be due to

differences in speaker vocal tract length. There were also differences in the phonological environments in which the vowels in each language were produced. All of the Chinese vowels were produced without onset consonants, but English /u/ was pronounced following alveolar onset stops, which would favor coarticulatory fronting on the vowel, and English /a/ was pronounced in words with labial and velar initials, both of which favor coarticulatory raising on the vowel. Even more problematically, the tokens of English /a/ were pronounced in words with coda /r/, an environment which favors raising and fronting.

Another study of the production of a subset of Mandarin vowels /i a u ə y ɿ ʅ/ by L1 English speakers and L1 Mandarin speakers revealed significant differences in both F1 and F2 for all vowels based on L1 status (Xie 2013). However, the study suffered from several shortcomings. First, the F1 and F2 values of the vowels were compared using t-tests, but only *t* and *p* values, not effect sizes, were not reported. Summary statistics for vowel measurements were also not reported, and there was no mention of normalization. Because of these shortcomings, it is not possible to judge if the differences in vowel height and backness by L1 status are likely to be perceptible.

Wu (2011) studied the production of Mandarin vowels by L1 English speakers and L1 Mandarin speakers, finding that L1 English L2 Mandarin speakers did not faithfully acquire Mandarin /y ɿ ʅ/. In Wu's study, L2 Mandarin speakers' productions of [y] were farther back and lower than L1 Mandarin speakers, and their productions of /ɿ ʅ/ (transcribed as /ɨ/) were not as high as L1 Mandarin speakers. L2 speakers also did not faithfully acquire the backness alternation of Mandarin /a/ (described in section 1.13); L2 speakers either produced both /a/ allophones without variation, distinguished them in height instead of backness, or produced each allophone in the opposite environment from L1 speakers.

1.2.2 *Models of L2 speech acquisition and perception*

While the Contrastive Analysis Hypothesis (Lado 1957) itself has largely been supplanted by newer theories, these newer theories still, in some way, incorporate the idea of phonetic and/or phonological transfer from a speaker's first language to their target language. The three most-cited of these newer theories are the Speech Learning Model (SLM) (Flege 1995; 2007) the Perceptual Assimilation Model (PAM) (Best 1995), including the specific predictions it makes about L2 speech (PAM-L2; Best & Tyler 2007), and the L2 Linguistic Perception model (L2LP) (Escudero 2005). These theories, which often make similar or identical predictions, are frequently used to predict the difficulty or likelihood of acquisition, both in production and perception, of various features of the phonology of a second or additional language. Many papers in acquisition and perception of L2 phonology set out to confirm or falsify one or more of these theories. While this paper does not aim to test any of these theories, all of them, especially PAM-L2, have influenced the selection of segments to include in this paper's investigations of L1 transfer effects.

1.2.3 *Potential L1 transfer effects*

A total of fourteen Mandarin segments will be analyzed in this paper. The rationale for the inclusion of each segment is justified below. There are other potential candidates ripe for segmental comparison based on an L1 transfer analysis, such as [ɻ ɻ̩], but it was not possible to obtain or compile a corpus of data of L2 Mandarin speech that included a sufficient number of these segments for comparison.

a. [u] and [y]

In many, if not most dialects of American English, /u/ is fronted from its cardinal position (Labov et al. 2006), and may be more accurately transcribed using the IPA symbol [ʉ]. In terms of acoustic

backness, [ʉ] is in an intermediate position between Mandarin /u/ and /y/, which are typically realized in their cardinal positions (Lee & Zee 2003). Yet L2 Mandarin curricula for English learners generally include only /y/ and not /u/, potentially resulting in an unfaithfully fronted pronunciation of /u/ and less distance between these two high rounded vowels in the L2.

b. [ou] and [ei]

In some dialects of American English, [ou] has been shown to front, [ei] has been shown to raise, and either or both vowels have been shown to monophthongize (Labov et al. 2006). A fronted [ou], a raised [ei], or a monophthongal variant of either vowel transferred from L1 English would be phonetically distinct from Mandarin [ei ou], which have not been shown to be subject to dialectal variation in these dimensions.

c. [a] and [an]

The Mandarin productions of the vowel /a/ (which would be more accurately transcribed as the central vowel /ɐ/ but will be referred to as /a/ in this paper for typographical convenience) could differ between L1 and L2 speakers in backness because English /ɑ/ is likely to be retracted relative to Mandarin /a/. This difference in vowel quality, however, would not be taught under a traditional approach based on contrastive analysis because the difference should not interfere with any phonological contrast in Mandarin. Furthermore, there might not be a perceived contrast at all if the English vowel is thought of as [a], as transcribed in a widely-cited study describing the qualities of American English vowels (Hillenbrand et al. 1995). Mandarin /a/, while normally realized in a central position, becomes a low front vowel when occurring before coda /n/ (Wu 2011). While /ɑ/ is not subject to this conditioning in any known variety of English, it is possible that L1 English learners of Mandarin may pick up on this distinction, as the fronted quality of Mandarin /a/ before /n/ resembles English /æ/. On the other hand, many dialects of English raise /æ/ before nasals

(Labov et al. 2006), so it may be the case that L1 English speakers raise Mandarin /a/ in addition to fronting it when it occurs before /n/.

d. [iŋ ~ iᵊŋ]

In most dialects of Mandarin, including Beijing Mandarin, /iŋ/ is traditionally realized as [iᵊŋ], even though the presence of a schwa offglide is considered by some to be nonstandard or uneducated (C. W.-C. Li 2004). This traditional pronunciation is not always realized; one study showed that only about 40% of speakers maintain it (Mou 2006). The common production of Mandarin /iŋ/ as [iᵊŋ] is also often omitted from L2 Mandarin curricula, despite its possible perceptual relevance for L2 listeners as a way to distinguish [iŋ] from [in] in cases of nasal coda consonant deletion (see [n# ŋ# ~ Ṽ]).

e. [Vn# Vŋ# ~ Ṽ]

The two Mandarin consonants that can appear in coda position in a syllable, [n ŋ], are often deleted, especially when followed by a vowel or a fricative (J. Li & Cheng 2014). Deletion is considered standard, and occurs even in the speech of news and radio broadcasters (C. W.-C. Li 2004). The contrast between the two nasals, however, is preserved based on the quality of the previous vowel, as all vowels have separate allophones when preceding [n] vs. [ŋ], with the exception of speakers who produce [iŋ] instead of [iᵊŋ], discussed above.

f. [ɕ t͡ɕ t͡ɕʰ]

The alveolopalatal fricative and affricates in Mandarin form part of a typologically uncommon three-way distinction with the sets [s t͡s t͡sʰ] and [ʂ t͡ʂ t͡ʂʰ]. English lacks alveolopalatals and only has [s] and [ʃ t͡ʃ t͡ʃʰ] in its inventory of distinctive consonants. Because Mandarin [ɕ t͡ɕ t͡ɕʰ] and [s t͡s t͡sʰ] can always be distinguished based on the quality of the following vowel or glide, it may be

the case that L1 English speakers simply transfer [s] and/or [ʃ t͡ʃ t͡ʃʰ] instead of learning this new set of sounds.

g. [ɹ]

Mandarin [ɹ] is often produced with frication in prevocalic position (Chen, Pik & Mok 2021), but this frication is not known to occur in English [ɹ].

h. [l ~ ɬ]

English [l] is often pronounced with velarization or pharyngealization, especially in syllable-final position (Sproat & Fujimura 1993). This variation is not known to occur in Mandarin.

1.3 AIMS

This paper investigates transfer effects from L1 North American English into L2 Mandarin vowels and consonants. Relative to most literature investigating acquisition of second language phonology, closer attention is paid to interspeaker phonetic variation. In other words, this paper utilizes an approach that recognizes variation among speakers of a given language in order to accurately describe vowels and consonants of intermediate and advanced L2 Mandarin Chinese spoken by L1 English learners. An accurate description of the vowels and consonants both in a given L1 and a given target language form is necessary for a solid foundation of second language pronunciation teaching and learning. To achieve these aims, this paper includes an auditory investigation (Study 1) and an acoustic investigation (Study 2). Each of these studies builds on Yang's (2011) research by investigating differences in vowels and consonants between L1 and L2 Mandarin speakers, using the same corpus compiled by Yang.

1.4 STUDY 1: AUDITORY ANALYSIS

The primary purpose of Study 1 is to identify linguistic forms produced in the corpus that are likely to be subject to L1 transfer effects. Such forms will be analyzed acoustically in Study 2.

1.4.1 *Methodology: Data*

The corpus analyzed consisted of 44.1 kHz, 24-bit recordings made in a sound-attenuated studio by Yang Chunsheng at the Ohio State University. Recordings were made by 10 L1 Mandarin speakers (4 men, 6 women) who were born and raised in Beijing and 21 American L1 English, L2 Mandarin speakers studying Chinese as a Foreign Language at the Ohio State University. Of the L2 speakers, 11 were of advanced proficiency (9 men, 2 women) and 10 were of intermediate proficiency (6 men, 4 women). Speakers ranged from 19 to 35 years of age (mean: 25.64). Speakers read three repetitions of 24 target sentences embedded in conversational scenarios (see Appendix A). For more information about the corpus, see (Yang 2011). 2103 sentence-length audio files were automatically segmented into words and phones using the Montreal Forced Aligner (McAuliffe et al. 2017) using a custom dictionary consisting only of the words that appeared in the 24 target sentences. A total of 5942 segments were included in the auditory analysis.

1.4.2 *Methodology: Procedure*

Auditory analysis was conducted in Praat (Boersma & Weenink 2020). A Praat script displayed a waveform and spectrogram in the sound editor window, then played each target token, which was extracted from its embedded sentence. The author, a trained phonetician, listened to each token in a quiet room using Sennheiser HD 559 headphones and visually inspected the waveform and spectrogram. Vowel quality was recorded relative to the cardinal position of each vowel on an IPA

vowel chart. Backness was not recorded for [ei] because in a preliminary analysis, it universally covaried with height. If there was a clear discrepancy between the recording played and its transcription due to a segmentation error occurring during the alignment process, it was excluded.

Segment (pinyin) [IPA]	Parameters	Tokens (pinyin)
(u/wu) [u]	Fronted/not fronted (Diphthongal/monophthongal)	dú lù wū mǔmǎ
(yu) [y]	Backed/not backed (Diphthongal/monophthongal)	yùmiáo yùnnán yùnyào
(ou) [ou]	Fronted/not fronted (Diphthongal/monophthongal)	yóu làròu
(ei) [ei]	Diphthongal/monophthongal	fēijī wèi
(an) [an]	Raised/not raised Fronted/canonical/backed (Diphthongal/monophthongal)	ān yúnnán mànyòng ránliào tán
(a#) [a#]	Fronted/not fronted	ná làròu
(ing) [iŋ]	Lowered/not lowered Diphthongal/monophthongal	yīng qīng míng
(n#/ng#) [n#/ŋ#]	Closure/no closure	yīng qīng míng ān yúnnán mǐn yányào mànyòng ránliào tán liànxí yángmáo mèng xiǎng yóng yàn yán
(l) [l]	Velarization (clear/dark)	lái liú lì luó lùn lù liànxí
(x j q) [ɕ tɕ tɕ^{h}]	Sounds alveolopalatal/sounds alveolar	xiū jī qīng xiǎng liànxí
(r_) [ɹ_]	Frication/no frication	làròu, ránliào

Table 1 Segments, parameters, and tokens included in the auditory analysis

1.4.3 *Methodology: Analysis*

Separate logistic mixed models were fit for by parameter and by phone, except for the height of /ei/, and the backness of /an/, for which linear mixed models were run. Each model included the

value associated with a phone and parameter as the dependent variable with level of proficiency as a fixed effect and speaker and word as random effects, except for the VISC models for [a#] and [iŋ], the height model for [ei], and the backness model for [an] which could not include speaker as a random effect due to insufficient variability. Data summarization was completed with the tidyverse packages (Wickham et al. 2019) for R (R Core Team 2021). Statistical models were fit using the lmerTest (Kuznetsova, Brockhoff & Christensen 2017) package for R.

1.4.4 *Results and discussion*

A summary of the auditory analysis of vowels is represented graphically in Figure 1. The top panel shows that the majority of vowels were categorized as canonical in terms of vowel height, with the most deviations occurring with [ei], which tended to be classified as raised for L1 speakers and lowered for intermediate L2 speakers. The center panel demonstrates that there was greater variation in vowel backness relative to vowel height for all analyzed vowels with the exception of [ou]. The greatest amount of variation across proficiency levels occurred for the vowels [a#], [u], and [y]. [a#] sounded backed for approximately 42% of productions by advanced L2 speakers and 21% of productions by intermediate L2 speakers, but only 1% of productions by L1 speakers. [u] was categorized as fronted for approximately 38% of productions by advanced L2 speakers and 2147 of productions by intermediate L2 speakers, but only 7% of productions by L1 speakers. Lastly, [y] sounded backed for approximately 42% of productions by advanced L2 speakers and 65% of productions by intermediate L2 speakers, but only 19% of productions by L1 speakers. The bottom panel reflects that [ei] and [ou] were

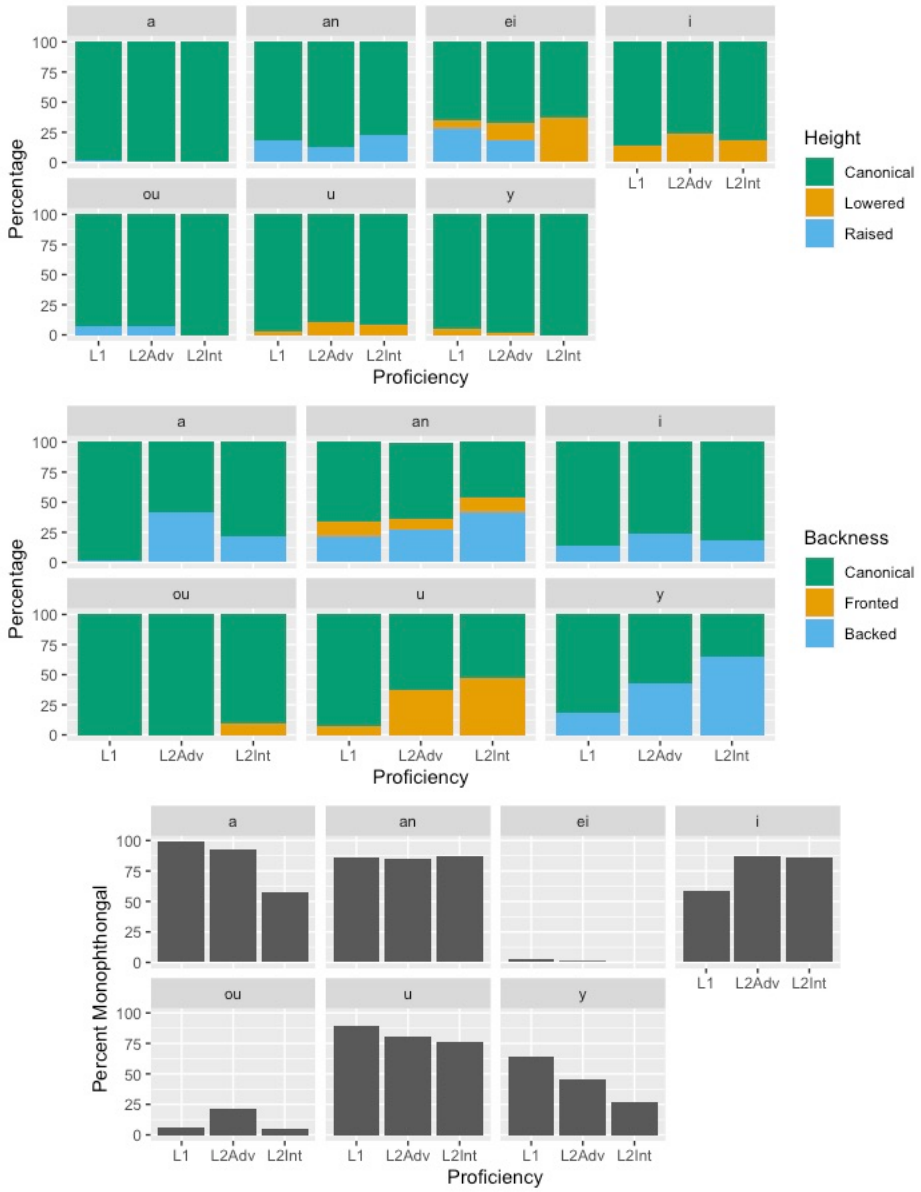

Figure 1 A summary of the auditory analysis for vowels, grouped by parameter
Top: height, Center: backness, Bottom: monophthongal/diphthongal

generally heard as diphthongs, with less than 25% of tokens sounding monophthongal for any level of proficiency, despite [ei] and [ou] monophthongization being common in American English (Labov et al. 2006). The greatest differences across proficiency levels for nominal monophthongs were for the vowels [a# iŋ y]. 43% of productions of [a#] by intermediate L2 speakers sounded diphthongal, but this was the case for only 8% of productions by advanced L2 speakers and 1% of productions by L1 speakers. 13% of productions of [iŋ] by L2 intermediate

speakers and 14% of productions by L2 advanced speakers sounded diphthongal, but 41% of productions by L1 speakers sounded diphthongal. Lastly, more than half of the productions of [y] by intermediate and advanced L2 speakers sounded diphthongal, at 73% and 54% of productions, respectively, but this was the case for only 36% of productions by L1 speakers.

To assess the statistical significance of differences described, results of the logistic mixed models run for each vowel are listed below.

a. u

Models showed significant differences in production of [u] between L1 speakers and both advanced and intermediate L2 speakers for backness (L2 advanced $p = 0.006$, L2 intermediate $p = 0.002$). No significant differences were found for height (L2 advanced $p = 0.075$, L2 intermediate $p = 0.084$) or VISC (L2 advanced $p = 0.140$, L2 intermediate $p = 0.053$).

b. y

Models showed significant differences in production of [y] between L1 speakers and intermediate L2 speakers but not for advanced L2 speakers, for backness (L2 advanced $p = 0.058$, L2 intermediate $p < 0.001$) and for VISC (L2 advanced $p = 0.217$, L2 intermediate $p = 0.008$). No model was run for vowel height because all of the tokens produced by L2 intermediate speakers were observed as canonical.

c. ou

Models showed no significant differences in production of [ou] between L1 speakers and advanced or intermediate L2 speakers for height (L2 advanced $p = 0.717$, L2 intermediate $p = 0.962$), backness (L2 advanced $p = 1.00$, L2 intermediate $p = 0.995$) or VISC (L2 advanced $p = 0.077$, L2 intermediate $p = 0.569$).

d. ei

A linear mixed model showed significant differences in production of [ei] between L1 speakers and both advanced and intermediate L2 speakers for backness (L2 advanced $p = 0.008$, L2 intermediate $p < 0.001$). No model for VISC was fit because there were no monophthongal-sounding tokens for L2 intermediate speakers and only a single monophthongal-sounding token for L2 advanced speakers.

e. an

Logistic mixed models showed no significant differences in production of [an] between L1 speakers and advanced or intermediate L2 speakers for height (L2 advanced $p = 0.178$, L2 intermediate $p = 0.572$) or VISC (L2 advanced $p = 0.856$, L2 intermediate $p = 0.970$). A linear mixed model showed a significant difference in backness between L1 speakers and intermediate L2 speakers, but not between L1 speakers and advanced L2 speakers (L2 advanced $p = 0.339$, L2 intermediate $p < 0.001$).

f. a#

Models showed significant differences in production of [a#] between L1 speakers and both advanced and intermediate L2 speakers for backness (L2 advanced $p < 0.001$, L2 intermediate $p = 0.001$) and for VISC (L2 advanced $p = 0.043$, L2 intermediate $p < 0.001$). No model was run for height because all tokens were canonical except for a single raised token produced by an L1 speaker.

g. iŋ

The VISC model showed significant differences in production of [iŋ] between L1 speakers and both advanced and intermediate L2 speakers (L2 advanced $p < 0.001$, L2 intermediate $p < 0.001$). Models showed no significant differences in production of [iŋ] between L1 speakers and

advanced or intermediate L2 speakers for height (L2 advanced p = 0.345, L2 intermediate p = 0.392) or backness (L2 advanced p = 0.345, L2 intermediate p = 0.392).

Figure 2 shows a graphical representation of the auditory analysis of consonants. The top left panel shows that L1 speakers have the fewest tokens of nasal coda consonants with audible oral closure, followed by L2 advanced speakers, then L2 intermediate speakers. The logistic mixed model for nasal coda consonants showed a significant difference in audible oral closure between L1 speakers and advanced L2 speakers ($p < 0.001$) and between L1 speakers and intermediate L2 speakers ($p < 0.001$). Despite expectations that (r) [ɹ] would have audible frication present, especially for L1 speakers, the top right panel shows that L1 speakers actually produced the smallest proportion of tokens with audible frication. The proportion was comparable across proficiency levels as confirmed by a logistic mixed model; no significant differences in audible frication were observed between L1 speakers and advanced L2 speakers (p = 0.906) or intermediate L2 speakers (p = 0.842). The bottom left panel of Figure 2 also

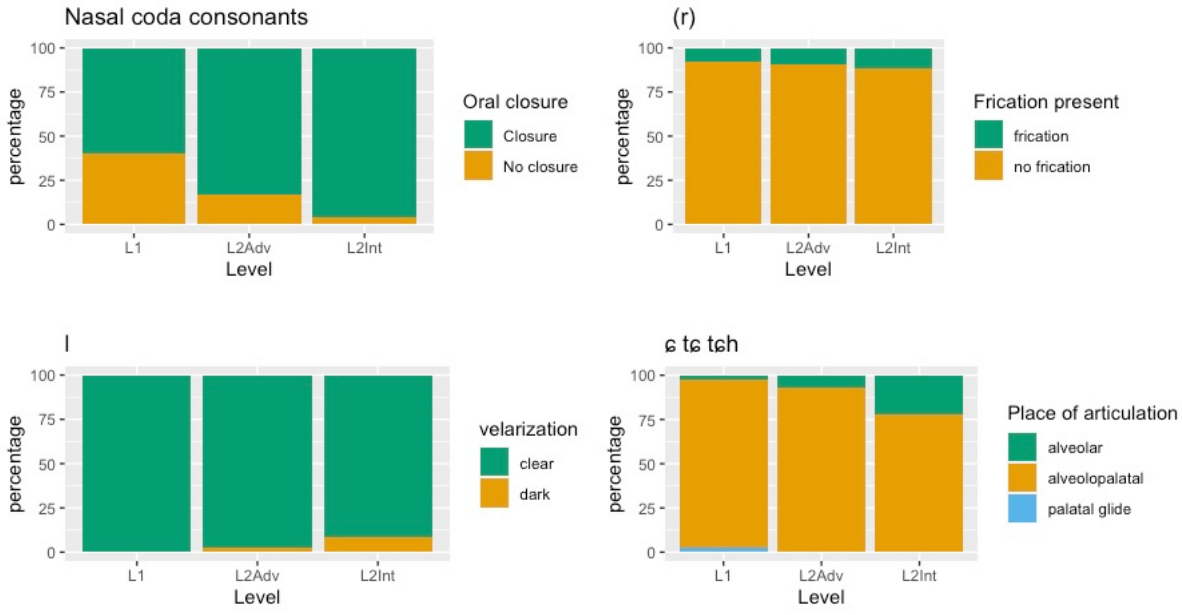

Figure 2 A summary of the auditory analysis for consonants, grouped by parameter

indicates similarity in the presence of velarization across proficiency levels, but a logistic mixed model did show a significant difference in place of articulation for [l] between L1 speakers and intermediate L2 speakers ($p = 0.014$), though not between L1 speakers and advanced L2 speakers ($p = 0.2036$). Lastly, the bottom right panel shows that L2 intermediate speakers have a higher proportion of underlying alveolopalatal fricatives that sounded like alveolar fricative. The logistic mixed model confirmed this, showing a significant difference in place of articulation for [ɕ tɕ tɕʰ] between L1 speakers and intermediate L2 speakers ($p = 0.003$) but not between L1 speakers and advanced L2 speakers ($p = 0.697$); the small number of tokens that sounded like palatal glides were excluded from this model.

Segments (pinyin) [IPA]	Parameters with significant differences across proficiency levels
(u/wu) [u]; (yu) [y]	Backness and VISC
(an) [an]; (a#) [a#]	Backness
(ing) [iŋ]	VISC
(ei) [ei]	Height
(n#/ng#) [n#/ŋ#]	Closure
(l) [l]	Darkness (velarization)
(x j q) [ɕ tɕ tɕʰ]	Place of articulation
(ou) [ou], (r_) [ɹ_]	None

Table 2 Summary of Study 1 results

1.5 STUDY 2: ACOUSTIC ANALYSIS

A subset of the segments included in the auditory analysis in Study 1 was selected for confirmatory acoustic analysis in Study 2: [u y ei a# an i/iŋ ɕ]. The affricates [tɕ tɕʰ] were excluded because there was only one word in the data set for each. [ɹ ou] were not included

because the auditory analysis did not reveal any significant differences across proficiency levels. Oral closure of nasal codas was not included because oral closure is very difficult to measure acoustically. Lastly, [l] was measured in the acoustic study, but it was excluded from summarization and statistical analysis due to doubts about the reliability of measurements for [l], which often appeared on spectrograms as transients, especially for the L2 Mandarin speakers. The auditory analysis has identified both velarization of [l] and closure of final nasals in L1 English L2 Mandarin as ripe for future articulatory study; alveolar and velar oral closure as well as [l] velarization and pharyngealization can be observed either with ultrasound tongue imaging or with MRI video.

1.5.1 *Methodology: Data*

The data used in Study 2 consisted of a subset of the data used in Study 1. The subset included a total of 3513 segments. In addition to the exclusions mentioned above, the word 母马 mǔmǎ “mare” was excluded from acoustic analysis because of concerns that the laryngealization associated with its lexical tones could cause a large number of formant tracking errors.

1.5.2 *Methodology: Procedure*

The first, second, and third formants of each vowel and of /l/ were measured using the Fast Track plugin (Barreda 2021) for Praat in order to quantify vowel quality. Fast Track offers three suggested formant tracking settings based on speaker height; because speaker height was not measured, the lowest and highest analysis frequencies for male speakers were set to 4500 Hz and 6500 Hz and for female speakers were set to 5000 Hz and 7000 Hz. 20 measurement points were taken per vowel, which were then grouped into 5 bins. Fast Track’s algorithm was set to use 3 coefficients per formant to detect the best analysis. Outliers were detected using a Mahalanobis

distance-based routine (Squizzero & Wassink 2022) and were manually corrected. Formant measurements were normalized in phonR (McCloy 2016) using Nearey's shared log-mean technique (Nearey 2) (Nearey 1978), which is speaker-intrinsic and formant extrinsic. Center of gravity (CoG), a measure of the locus of high energy in an acoustic signal, was used to quantify place of articulation of [ɛ]. CoG was measured using six 15-millisecond windows over the central 80% of each fricative, then time-averaged, using a Praat script (DiCanio 2013).

1.5.3 *Methodology: Analysis*

While vowels are often described using formant frequency measurements taken at a single time point or averaged over the course of the vowel in acoustic phonetic research, scientific evidence has shown that North American English vowels are more accurately characterized, both in production and perception, by describing spectral change occurring over their time courses (Morrison & Assmann 2013). To investigate possible transfer effects from L1 North American English to L2 Mandarin Chinese, generalized additive mixed models (GAMMs) (Wood 2011) were therefore constructed for each vowel and formant to examine both their normalized formant values and their time-varying properties. Models were constructed using the mgcv (Wood 2011) and itsadug (van Rij et al. 2020) packages in R, following Sóskuthy (2017). Model summaries, smooth plots, and difference plots were produced using the gamm_hacks.R script by Márton Sóskuthy. Each model included F1 or F2 as the dependent variable, a fixed parametric term for L1/L2 status[i], a fixed smooth term for temporal measurement point, an interaction smooth term for measurement point and L1/L2 status, and a random smooth term for speaker. Model summaries are reported only for segment and formant combinations with significant results in at least one model term; no differences were significant for [an].

For the only consonant in the acoustic study [ɛ], a linear mixed model was fit using the lmerTest package for R (Kuznetsova et al. 2017). The model included CoG as the dependent variable, gender, proficiency, and the interaction between gender and proficiency as fixed effects, and speaker as a random effect.

1.5.4 *Results and discussion: fricatives*

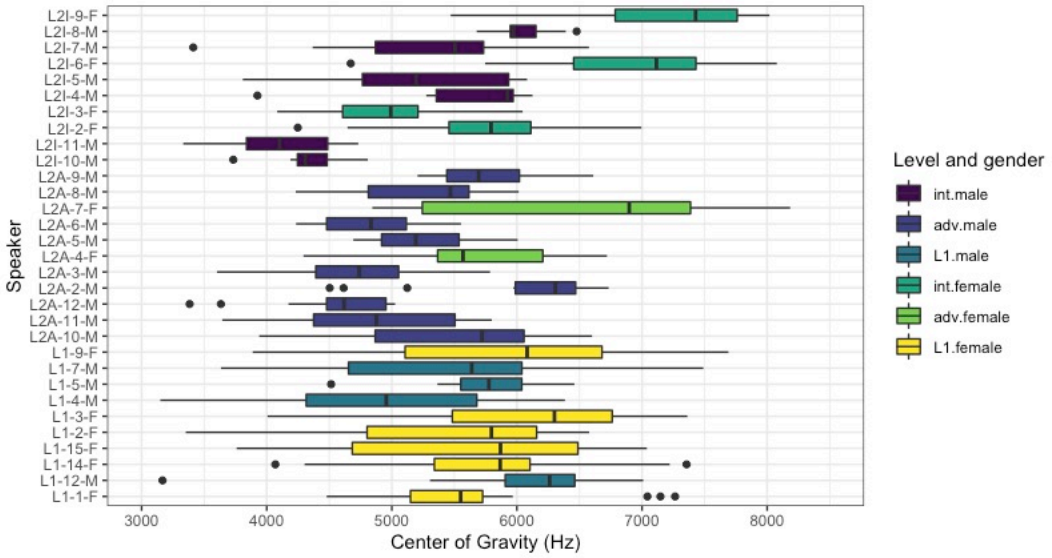

Figure 3 Box plot for center of gravity of [ɛ], by speaker

Proficiency	Gender	Mean CoG (Hz)	SD (Hz)
L2I	female	6213	1131
L2A	female	6194	1213
L1	female	5731	947
L2I	male	5107	865
L2A	male	5156	716
L1	male	5513	1039

Table 3 Mean and standard deviations for center of gravity of [ɛ], by proficiency and gender

	Estimate	Std. Error	df	t value	Pr(>|t|)
(Intercept)	6195.22	314.01	25.42	19.729	<2e-16 ***
leveladv	-33.19	543.62	25.37	-0.061	0.9518
levelL1	-469.07	402.24	24.64	-1.166	0.2547
gendermale	-1093.84	405.46	25.44	-2.698	0.0122 *
leveladv:gendermale	122.70	636.57	25.39	0.193	0.8487
levelL1:gendermale	866.70	567.62	24.43	1.527	0.1396

Table 4 Model summary for [ɛ]. Signif. codes: 0 '***' 0.001 '*' 0.05 ' ' 1

Although neither the main effect of proficiency nor the interaction effect of proficiency and gender in the linear mixed model for [ɛ] (Table 4) reached statistical significance, inspection of by-speaker CoG data (Figure 1) reveals some interesting trends. CoG means for all L1 speakers except for L1-M-4 fall within a relatively narrow range of 5396 (L1-F-2) to 6112 Hz (L1-F-3). L1-M-4's lower mean CoG of 4651 Hz the lowest of the L1 speakers because of two outlier tokens; these tokens were short in duration and appeared to be voiced, under coarticulatory influence from the surrounding vowels. The advanced speakers have a wider range in by-speaker means, ranging from 4556 (L2A-12-M) to 6698 Hz (L2A-7-F). The intermediate speakers have an even wider range in by-speaker mean CoG, ranging from 4113 (L2I-11-M) to 7195 Hz (L2I-9-F). Impressionistically, lower-CoG tokens produced by the L2 speakers sounded more [ʃ]-like, and higher-CoG tokens produced by the L2 speakers sounded more [s]-like.

1.5.5 *Results and discussion: vowels*

L1/L2 status	Vowel	F1 Mean (logHz)	F1 SD	F2 Mean (logHz)	F2 SD
L1	a	0.700	0.117	1.210	0.133
L2	a	0.672	0.079	1.140	0.123
L1	an	0.614	0.113	1.250	0.119
L2	an	0.621	0.086	1.240	0.166
L1	i	0.320	0.076	1.850	0.309
L2	i	0.338	0.073	1.880	0.324
L1	u	0.303	0.048	0.688	0.152
L2	u	0.296	0.042	0.818	0.220
L1	y	0.243	0.042	1.750	0.157
L2	y	0.245	0.038	1.790	0.331

Table 5 Mean and standard deviations for F1 and F2 at midpoint, by vowel and L2 status

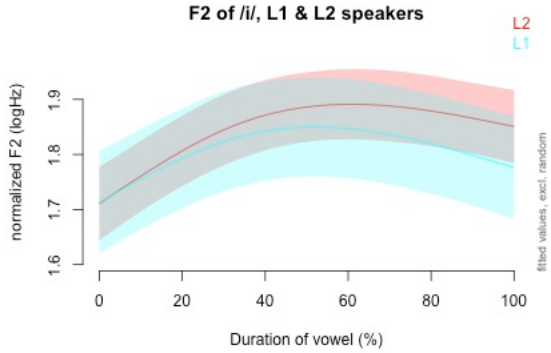

Figure 4 Smooth plot, generalized additive mixed model for F2 of /i/ before /ŋ/. Includes means and 95% confidence intervals; excludes random effects. L1 speakers: cyan; L2 speakers: red

Parametric coefficients:

	Estimate	Std. Error	t value	Pr(>|t|)
(Intercept)	1.79532	0.04528	39.653	<2e-16 ***
L1statusL2	0.03647	0.05518	0.661	0.509

Approximate significance of smooth terms:

	edf	Ref.df	F	ep-value
s(Interval)	2.795	2.96	25.814	<2e-16 ***
s(Interval):L1statusL2	1.000	1.00	5.484	0.0192 *
s(Interval,speaker)	36.005	120.00	5.858	<2e-16 ***

Table 6 Generalized additive mixed model summary for F2 of /i/ before /ŋ/

Significant effects were observed in the GAMM constructed with the F2 of /i/ preceding /ŋ/ as the dependent variable. The significant fixed interaction smooth term for /i/, shown in Table 7, indicates an overall difference in the shape of the trajectory of the vowel based on L1/L2 status. The difference in shape can be observed in the Figure 4, where F2 of /i/ initially rises for both L1 and L2 speakers, but then decreases towards the end of the vowel's duration for L1 speakers (bottom trajectory) only.

The significant fixed parametric term for L1/L2 status in the GAMM shown in Table 8 indicates an overall difference in the F2 of /u/ based on L1/L2 status. The difference in backness can be observed in the left panel of Figure 5, where there is a visible gap between the L2 speakers (top trajectory) and the L1 speakers. The higher F2 for the L2 speakers indicates a fronter

pronunciation of /u/ relative to the L1 speakers. The right panel of Figure 4 shows an estimate of the overall difference between L1 and L2 speakers, with the F2 of /u/ for L2 speakers higher than the L1 speakers by an average of roughly 0.11 logHz. The right panel of Figure 4, in addition to allowing for easier viewing of the magnitude of the difference between the two trajectories, also shows that this difference decreases over the time course of the production of the vowel, though the interaction smooth term for L1/L2 status in the model summary indicates that the shape of the trajectory for the L1 speakers is not significantly different from the L2 speakers' trajectory shape.

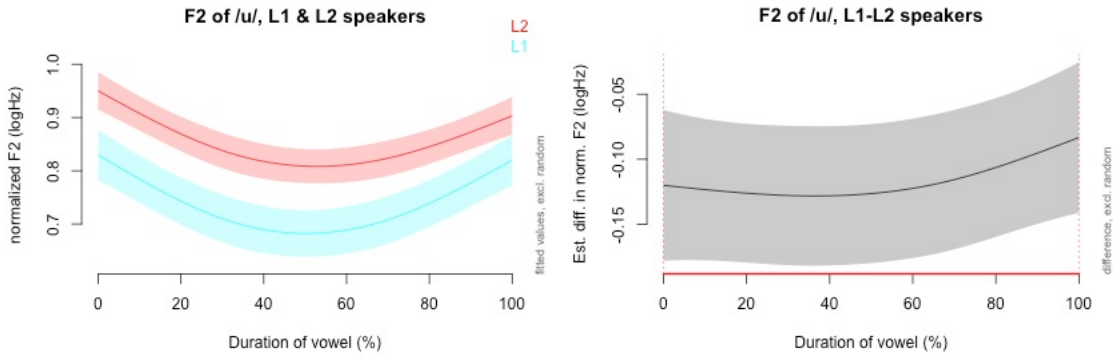

Figure 5 Generalized additive mixed model plots for F2 of /u/. Left: smooth plot, includes means and 95% confidence intervals. L1 speakers: bottom, cyan; L2 speakers: top, red. Right: difference plot, includes the mean difference and 95% confidence intervals. Both panels exclude random effects.

Parametric coefficients:

	Estimate	Std. Error	t value	Pr(>|t|)
(Intercept)	0.75590	0.02117	35.698	< 2e-16 ***
L1statusL2	0.11361	0.02594	4.379	1.23e-05 ***

Approximate significance of smooth terms:

	edf	Ref.df	F	p-value
s(Interval)	2.891	2.976	37.690	< 2e-16 ***
s(Interval):L1statusL2	1.822	2.184	2.074	0.11
s(Interval,speaker)	25.562	120.000	1.819	< 2e-16 ***

Table 7 Generalized additive mixed model summary for F2 of /u/

The significant interaction smooth term for the F2 of /y/ in Table 9 indicates an overall difference in the trajectory shape. Figure 6 shows that there is more change over the course of the trajectory for L2 speakers, with F2 both peaking higher (at around the 40% point) and ending lower

relative to the L1 speakers. The more diphthongal production for some L2 speakers is often realized as [ju] or [iu], possibly due to orthographic interference because syllable-initial /y/ in Hanyu pinyin is spelled as “yu.”

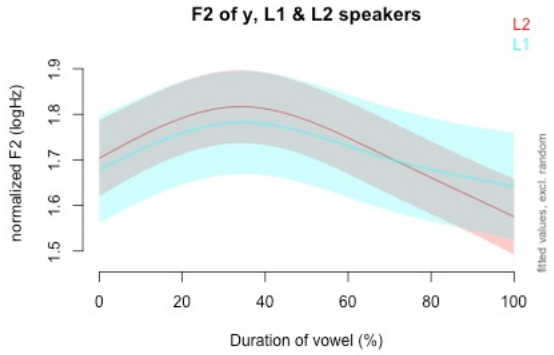

Figure 6 Smooth plot, generalized additive mixed model for F2 of /y/. Includes means and 95% confidence intervals; excludes random effects. L1 speakers: cyan; L2 speakers: red.

Parametric coefficients:

	Estimate	Std. Error	t value	Pr(>|t|)
(Intercept)	1.708368	0.056095	30.455	<2e-16 ***
L1statusL2	0.002693	0.068308	0.039	0.96

Approximate significance of smooth terms:

	edf	Ref.df	F	p-value
s(Interval)	2.827	2.926	12.305	5.28e-06 ***
s(Interval):L1statusL2	1.992	2.283	3.030	0.0362 *
s(Interval,speaker)	54.783	120.000	7.989	< 2e-16 ***

Table 8 Generalized additive mixed model summary for F2 of /y/

The significant fixed parametric term for L1/L2 status in the GAMM constructed with the F2 of /ei/ as the dependent variable, shown in Table 10, indicates an overall difference in the backness of /ei/ based on L1/L2 status. The difference in backness can be observed in Figure 7, which shows an estimate of the overall difference between L1 and L2 speakers of roughly 0.18 logHz. The negative value in the difference plot indicates a fronter pronunciation among the L2 speakers.

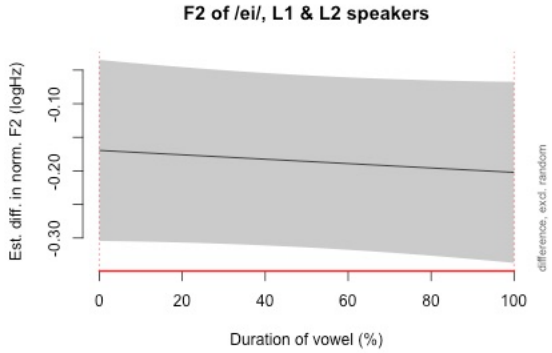

Figure 7 Generalized additive mixed model difference plot for F2 of /ei/. Includes the mean difference and 95% confidence intervals. Excludes random effects.

Parametric coefficients:				
	Estimate	Std. Error	t value	Pr(>|t|)
(Intercept)	1.64893	0.05361	30.760	<2e-16 ***
L1statusL2	0.18590	0.06538	2.843	0.0046 **

Approximate significance of smooth terms:				
	edf	Ref.df	F	p-value
s(Interval)	2.468	2.792	20.044	<2e-16 ***
s(Interval):L1statusL2	1.000	1.000	0.582	0.446
s(Interval,speaker)	27.185	120.000	4.694	<2e-16 ***

Table 9 Generalized additive mixed model summary for F2 of /ei/

/a/ produced in open syllables is the only vowel for which significant differences based on L1/L2 status are observed in both backness and height. In terms of backness, /a/ differs significantly both in overall F2 value, as evidenced by the fixed parametric term value for L1/L2 status in the GAMM shown in Table 11, and in trajectory shape, as evidenced by the interaction term in the same table. The lower F2 for the L2 speakers indicates a backer pronunciation of /a/ relative to the L1 speakers. The magnitude of the backness increases over the time course of the vowel, as can be observed in the top right panel of Figure 8. The difference in trajectory shape can be observed in the top left panel of Figure 8. The L1 speakers (top, cyan) maintain a relatively flat F2 trajectory until F2 starts to rise, approximately 60% of the way through the duration of the vowel, while the L2 speakers show a decrease in F2 from 0% to 60% before rising in a similar pattern to the L1 speakers. As for vowel height, the interaction smooth term for L1/L2 status in

the model summary (Table 12) indicates that the shape of the trajectory for the L1 speakers is significantly different from the L2 speakers' trajectory shape. The significant difference in shape can be observed in the bottom left panel of Figure 8, with the L1 speakers (top trajectory, cyan) raising at a steeper angle over the first half of the vowel as compared to the L2 speakers. The bottom right panel of Figure 8 shows that F1 is significantly different over the course of the last third of the vowel, but the non-significant parametric term in the model suggests that the vowel height difference difference between L1 and L2 speakers in height may not actually be significant (Sóskuthy 2017).

Parametric coefficients:

	Estimate	Std. Error	t value	Pr(>|t|)
(Intercept)	1.24447	0.02017	61.711	<2e-16 ***
L1statusL2	-0.05960	0.02474	-2.409	0.0161 *

Approximate significance of smooth terms:

	edf	Ref.df	F	p-value
s(Interval)	2.738	2.910	5.922	<2e-16 ***
s(Interval):L1statusL2	1.595	1.888	3.740	0.0168 *
s(Interval,speaker)	37.765	120.000	1.894	<2e-16 ***

Table 10 Generalized additive mixed model summary for F2 of /a/

Parametric coefficients:

	Estimate	Std. Error	t value	Pr(>|t|)
(Intercept)	0.65159	0.01802	36.156	<2e-16 ***
L1statusL2	-0.03157	0.02196	-1.438	0.15

Approximate significance of smooth terms:

	edf	Ref.df	F	p-value
s(Interval)	2.959	2.992	127.155	< 2e-16 ***
s(Interval):L1statusL2	1.001	1.002	10.346	0.00132 **
s(Interval,speaker)	55.556	120.000	6.711	< 2e-16 ***

Table 11 Generalized additive mixed model summary for F1 of /a/

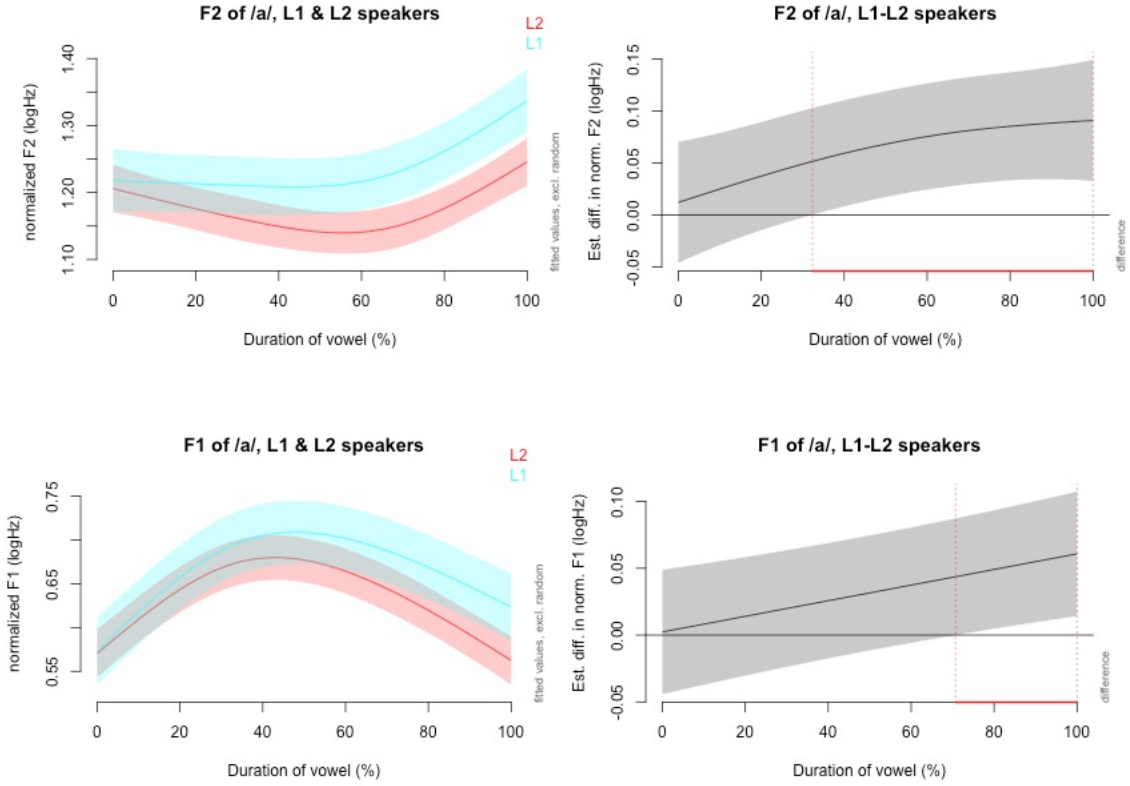

Figure 8 Generalized additive mixed model plots for F1 (top) and F2 of /a/. Left: smooth plots, include means and 95% confidence intervals. L1 speakers: cyan; L2 speakers: red. Right: difference plots, include mean difference and 95% confidence intervals. All panels exclude random effects.

1.6 General Discussion

Vowel	Overall Height	Overall Backness	Vowel-inherent spectral change
a#	L2 lower	L2 backer	Difference in height and backness
an	N.S.	N.S.	N.S.
ei	N.S.	L2 fronter	N.S.
i/ŋ	N.S.	N.S.	Difference in backness
u	N.S.	L2 fronter	N.S.
y	N.S.	N.S.	Difference in backness

Table 12 Summary of acoustic results for vowels by L1 status

A summary of the results from Study 2 is presented in Table 12. In general, L2 speakers' productions of [a#] are lower and backer relative to L1 speakers, and their productions of [ei] and [u] are fronter relative to L1 speakers. Furthermore, L1 and L2 speakers differ with respect

to the time-varying qualities of the vowels [a#], [i/ŋ] and [y]. As more vowels in the present study differed significantly in backness relative to height, results are consistent with prior work showing that backness is the most relevant dimension of the acoustic space for L1 English learners of Mandarin (Wu 2011).

However, the present study also showed that vowels differ between L1 and L2 speakers of Mandarin in terms of their time-varying qualities just as often as they differ in backness. Prior research has demonstrated the importance of vowel dynamics on speech perception; when vowels' formant trajectories are smoothed, intelligibility suffers, but when steady state portions of vowels are removed, intelligibility is not affected (Hillenbrand 2013). Future work investigating L1 transfer effects on L2 vowel production should therefore incorporate vowel-inherent spectral change in its analysis.

In addition to the use of methods that account for the time-varying qualities of vowels, this study has demonstrated the importance of considering sub-phonemic variation in investigation of L2 transfer effects. This study confirms prior reports that L1 English and L1 Mandarin speakers differ in their production of Mandarin /u/ (Shi & Wen 2009; Wu 2011; Xie 2013) /a/ (Wu 2011), and establishes that there are overall differences in production of [i] before [ŋ]. None of these differences would be obvious by a cursory inspection of the phonemic inventories of the two languages.

While future work is needed to confirm effects in Mandarin specifically, accurate L2 pronunciation is important in part because deviation from L1 pronunciation norms, even below the level of phonemic contrast, is associated with reductions in intelligibility (Porretta & Tucker 2015) and comprehensibility [Squizzero, next chapter] of L2 speech. It is unlikely that accurate L2 pronunciation will be achieved, however, unless the sounds of a learner's first language and

their target language are rigorously described. Besides making themselves intelligible and comprehensible to other speakers, language learners can also benefit in perception from understanding sub-phonemic differences. One example of how L2 Mandarin speakers can benefit is by learning to perceive the differences between [in] and [iᵊŋ]; since there is a tendency for coda consonants to be deleted, recognizing this difference can allow L2 Mandarin speakers to improve their listening comprehension. Additionally, language teachers should be aware that learners may manipulate vowel dynamics as a strategy to produce an unfamiliar L2 vowel; [jʉ] and [y] may not differ significantly in mean F2, but the difference between the two sounds can certainly be perceived auditorily.

In addition to the significant results for vowels, the results for [ɕ] are also potentially telling. Although L1/L2 status in the statistical model for CoG of [ɕ] was not significant, the by-speaker box plot (Figure 3) shows that several of the L2 speakers tended to produce [ɕ] with abnormally high or abnormally low CoG relative to the L1 speakers. These CoG values, as well as the auditory quality of these tokens, suggest that these speakers' productions of [ɕ] are closer to [s] or [ʃ], and therefore, these speakers have not truly acquired the segment [ɕ]. The fact that these fricatives are distinctive in L1 varieties of Mandarin suggests that neutralization of their contrast may negatively affect intelligibility or comprehensibility of spoken Mandarin. Future work is needed to understand how neutralization of the contrast between [ɕ] and [s] or [ɕ] and [ʃ] may affect speech perception.

1.7 Conclusion

Language transfer effects from L1 North American English into L2 Mandarin vowels and consonants were investigated via auditory and acoustic phonetic analysis of speech produced by L1 Mandarin speakers as well as advanced and intermediate L2 Mandarin speakers. Statistically

significant acoustic differences were found between L1 and L2 speakers' productions of Mandarin [a# ei i/ŋ u y], as were inter-learner differences in productions of Mandarin [ɐ]. The importance of both L1 and L2 dialectal variation and of time-varying cues for reliable description of L2 speech production were demonstrated. The benefits of having correct descriptions of L1 transfer effects for Mandarin teaching and learning pronunciation teaching and learning were discussed, as were the benefits of accurate L2 pronunciation.

1.8 REFERENCES

Barreda, Santiago. 2021. Fast Track: fast, (nearly) automatic formant-tracking using Praat. *Linguistics Vanguard* 7(1). Retrieved from https://doi.org/10.1515/lingvan-2020-0051

Best, Catherine T. 1995. A direct realist view of cross-language speech perception. In Winifred Strange (ed.), *Speech perception and linguistic experience: Theoretical and methodological issues in cross-language speech research.*, 167–200. York Press.

Best, Catherine T. & Michael D. Tyler. 2007. Nonnative and second-language speech perception. In *Second language speech learning: The role of language experience in speech perception and production.*, 13–34. Amsterdam: John Benjamins. DOI: https://doi.org/10.1075/lllt.17.07bes

Boersma, Paul & David Weenink. 2020. Praat: doing phonetics by computer [Computer program] Version 6.1.11. Retrieved from http://www.praat.org

Chen, Shuwen, Peggy Pik & Ki Mok. 2021. Articulatory and Acoustic Features of Mandarin /ɹ/: A Preliminary Study. DOI: https://doi.org/10.1109/ISCSLP49672.2021.9362070

DiCanio, Christian. 2013. Spectral moments of fricative spectra. Retrieved from https://www.acsu.buffalo.edu/~cdicanio/scripts/Time_averaging_for_fricatives_4.0.praat

Escudero, Paola. 2005. *Linguistic perception and second language acquisition: Explaining the*

attainment of optimal phonological categorization. *LOT, The Netherlands*,. Retrieved from http://wwwlot.let.uu.nl/

Flege, James E. 1995. Second language speech learning: Theory, findings, and problems. In Winifred Strange (ed.), *Speech perception and linguistic experience: Theoretical and methodological issues in cross-language speech research.*, 233–277. York Press.

Flege, James E. 2007. Language contact in bilingualism: Phonetic system interactions. In Jennifer Cole & Jose Ignacio Hualde (eds.), *Laboratory Phonology 9*, 353–381. Berlin: Mouton de Gruyter.

Hillenbrand, James M. 2013. Static and Dynamic Approaches to Vowel Perception. *Vowel Inherent Spectral Change* 9–30. DOI: https://doi.org/10.1007/978-3-642-14209-3_2

Hillenbrand, James M., Laura A. Getty, Michael J. Clark & Kimberlee Wheeler. 1995. Acoustic characteristics of American English vowels. *The Journal of the Acoustical Society of America* 97(5). 3099–3111. DOI: https://doi.org/10.1121/1.411872

Kuznetsova, Alexandra, Per B. Brockhoff & Rune H.B. Christensen. 2017. lmerTest Package: Tests in Linear Mixed Effects Models. *Journal of Statistical Software* 82(13). 1–26. DOI: https://doi.org/10.18637/JSS.V082.I13

Labov, William. 1994. *Principles of Linguistic Change. Volume 1: Internal Factors*. Malden, MA: Blackwell Publishing Ltd.

Labov, William, Sharon Ash & Charles Boberg. 2006. *The Atlas of North American English: Phonetics, Phonology and Sound Change* De Gruyter. Berlin.

Lado, Robert. 1957. *Linguistics across cultures: Applied linguistics for language teachers.* Ann Arbor: University of Michigan Press.

Lee, Wai Sum & Eric Zee. 2003. Standard Chinese (Beijing). *Journal of the International*

Phonetic Association 33(1). 109–112. DOI: https://doi.org/10.1017/S0025100303001208

Li, Chris Wen-Chao. 2004. Conflicting notions of language purity: The interplay of archaising, ethnographic, reformist, elitist and xenophobic purism in the perception of Standard Chinese. *Language and Communication* 24(2). 97–133. DOI: https://doi.org/10.1016/j.langcom.2003.09.002

Li, Jian & Le Cheng. 2014. The Acoustics Properties of the Nasals and Nasalization in Standard Chinese. *Information Technology Journal* 13(11). 1793–1799. DOI: https://doi.org/10.3923/itj.2014.1793.1799

McAuliffe, Michael, Michaela Socolof, Sarah Mihuc, Michael Wagner & Morgan Sonderegger. 2017. Montreal Forced Aligner.

McCloy, Daniel R. 2016. phonR: Tools for Phoneticians and Phonologists. Retrieved from https://cran.r-project.org/web/packages/phonR/index.html

McCullough, Elizabeth A. 2013. *Acoustic correlates of perceived foreign accent in non-native English*. The Ohio State University dissertation.

Morrison, Geoffrey Stewart & Peter F. Assmann. 2013. *Vowel inherent spectral change*. Springer Berlin Heidelberg. DOI: https://doi.org/10.1007/978-3-642-14209-3

Mou, Xiaomin. 2006. *Nasal codas in Standard Chinese : a study in the framework of the distinctive feature theory*. Massachusetts Institute of Technology dissertation.

Nearey, T.M. 1978. *Phonetic Feature Systems for Vowels*. Indiana: Indiana University Linguistics Club.

Porretta, Vincent & Benjamin V Tucker. 2015. Intelligibility of foreign-accented words : Acoustic distances and gradient foreign accentedness. *Proceedings of the 18th International Congress of Phonetic Sciences* 1–4.

R Core Team. 2021. R: A Language and Environment for Statistical Computing. Vienna, Austria: R Foundation for Statistical Computing.

Shi, Feng & Baoying Wen. 2009. Study of Language Transfer in Vowel Articulation by Chinese and American Students. *Journal of the Chinese Language Teachers Association* 44(2). 17–32.

Sóskuthy, Marton. 2017. *Generalised Additive Mixed Models for Dynamic Analysis in Linguistics*. Retrieved from https://eprints.whiterose.ac.uk/113858/2/1703_05339v1.pdf

Sproat, Richard & Osamu Fujimura. 1993. Allophonic variation in English /l/ and its implication for phonetic implementation. *Journal of Phonetics* 21. 291–311.

Squizzero, Robert & Alicia Beckford Wassink. 2022. *A comparison of three methods for identifying formant tracking errors via outlier detection*.

van Rij, Jacolien, Martin Wieling, R. Harald Baayen & Hedderik van Rijn. 2020. itsadug: Interpreting Time Series and Autocorrelated Data Using GAMMs.

Wickham, Hadley, Mara Averick, Jennifer Bryan, Winston Chang, Lucy McGowan, Romain François, … Hiroaki Yutani. 2019. Welcome to the Tidyverse. *Journal of Open Source Software* 4(43). 1686. DOI: https://doi.org/10.21105/joss.01686

Wood, Simon N. 2011. Fast stable restricted maximum likelihood and marginal likelihood estimation of semiparametric generalized linear models. *Journal of the Royal Statstical Society (B)* 73(1). 3–36.

Wu, Chen-huei. 2011. *The Evaluation of Second Language Fluency and Foreign Accent*. University of Illinois at Urbana-Champaign dissertation.

Xie, Yan. 2013. The acquisition of Mandarin basic vowels by American students: A comparison study of monosyllabic and disyllabic words. *Journal of the Chinese Language Teachers*

Association 48(1). 91–108.

Yang, Chunsheng. 2011. *The Acquisition of Mandarin Prosody by American Learners of Chinese as a Foreign Language (CFL)*. The Ohio State University dissertation. https://doi.org/10.1007/s13398-014-0173-7.2

Zhang, Hang. 2018. Current trends in research of Chinese sound acquisition. In Chuanren Ke (ed.), *The Routledge Handbook of Chinese Second Language Acquisition*, 217–233. DOI: https://doi.org/10.1201/b17627

1.9 ACKNOWLEDGEMENTS

The author would like to thank Alicia Beckford Wassink and Richard Wright for their input on the design of this study, and Chunsheng Yang for the use of his data.

1.10 APPENDIX A

Chinese, English, and pinyin reproduced from Yang (2011) p. 190-191. IPA added by the author.

1. 邬安英修飞机。 “Wu Anying repairs planes.”	Wū Ānyīng xiū fēijī. [u˥ an˥jiŋ˥ ɕjou˥ fei˥tɕi˥]
2. 邬安应修飞机。 “Wu An should repair planes.”	Wū Ān yīng xiū fēijī. [u˥ an˥ jiŋ˥ ɕjou˥ fei˥tɕi˥]
3. 殷安青摸猫咪。 “Yin Anqing pets a kitty.”	Yīn Ānqīng mō māomī. [jin˥ an˥tɕʰiŋ˥ mʷo˥ mau˥mi˥]
4. 殷安轻摸猫咪。 “Yin An gently pets a kitty.”	Yīn Ān qīngmō māomī. [jin˥ an˥ tɕʰiŋ˥ mʷo˥ mau˥mi˥]
5. 王明来拿羊毛。 “Wang Ming comes to get wool.”	Wáng Míng lái ná yángmáo. [waŋ˧ miŋ˧ lai˧ na˧ jaŋ˧mau˧]

6. 王明涞拿羊毛。
“Wang Minglai gets wool.”
Wáng Mínglái ná yángmáo.
[waŋ˥ ssssmiŋ˥lai˥ na˥ jaŋ˥mau˥]

7. 刘明莱游云南。
“Liu Minglai travels in Yunnan.”
Liú Mínglái yóu Yúnnán.
[ljou˥ miŋ˥lai˥ jou˥ yn˥nan˥]

8. 刘明来游云南。
“Liu Ming comes to travel in Yunnan.”
Liú Míng lái yóu Yúnnán.
[ljou˥ miŋ˥ lai˥ jou˥ yn˥nan˥]

9. 李敏响买母马。
“Li Minxiang buys a mare.”
Lǐ Mǐnxiǎng mǎi mǔmǎ.
[li˩ min˩ɕjaŋ˩ mai˩ mu˩ma˩]

10. 李敏想买母马。
“Li Min wants to buy a mare.”
Lǐ Mǐn xiǎng mǎi mǔmǎ.
[li˩ min˩ ɕjaŋ˩ mai˩ mu˩ma˩]

11. 李伟想买野鸟。
“Li Wei wants to buy a wild bird.”
Lǐ Wěi xiǎng mǎi yěniǎo.
[li˩ wei˩ ɕjaŋ˩ mai˩ jɛ˩njau˩]

12. 李伟响买野鸟。
“Li Weixiang buys a wild bird.”
Lǐ Wěixiǎng mǎi yěniǎo.
[li˩ wei˩ɕjaŋ˩ mai˩ jɛ˩njau˩]

13. 陆蔚用慢用药。
“Lu Wei uses the slow medicine.”
Lù Wèi yòng mànyòng yào.
[lu˥˩ wei˥˩ joŋ˥˩ man˥˩joŋ˥˩ jau˥˩]

14. 陆卫用卖孕药。
“Lu Weiyong uses the contraceptive medicine.”
Lù Wèiyòng mài yùnyào.
[lu˥˩ wei˥˩joŋ˥˩ mai˥˩ yn˥˩jau˥˩]

15. 魏丽要卖腊肉。
“Wei Li wants to sell bacon.”
Wèi Lì yào mài làròu.
[wei˥˩ li˥˩ jau˥˩ mai˥˩ la˥˩ɹou˥˩]

16. 魏立耀卖腊肉。
“Wei Liyao sells bacon.”
Wèi Lìyào mài làròu.
[wei˥˩ li˥˩jau˥˩ mai˥˩ la˥˩ɹou˥˩]

17. 南梦来卖燃料。
“Nan Meng comes to sell fuels.”
Nán Mèng lái mài ránliào.
[nan˥ məŋ˥˩ lai˥ mai˥˩ ɹan˥˩ljau˥˩]

18. 南梦莱卖燃料。 Nán Mènglái mài ránliào.
“Nan Menglai sells fuel.” [nanˊ məŋꓦlaiˊ maiꓦ ɹanˊljauꓦ]

19. 罗燕谈论名利。 Luó Yàn tánlùn mínglì.
“Luo Yan talks about fame and profit.” [lwoˊ jɛnꓦ tʰanˊlwənꓦ miŋˊliꓦ]

20. 罗彦坛论名利。 Luó Yàntán lùn mínglì.
“Luo Yantan talks about fame and profit.” [lwoˊ jɛnꓦtʰanˊ lwənꓦ miŋˊliꓦ]

21. 陆岩练习育苗。 Lù Yán liànxí yùmiáo.
“Lu Yan practices growing seeds.” [luꓦ jɛnˊ ljɛnꓦɕiˊ yꓦmjauˊ]

22. 陆言练学育苗。 Lù Yánliàn xué yùmiáo.
“Lu Yanlian practices growing seeds.” [luꓦ jɛnˊljɛnꓦ ɕwɛˊ yꓦmjauˊ]

23. 孟岩爱读外文。 Mèng Yán ài dú wàiwén.
“Meng Yan likes studying foreign languages.” [məŋꓦ jɛnˊ aiꓦ tuˊ waiꓦwənˊ]

24. 孟言艾读外文。 Mèng Yán’ài dú wàiwén
“Meng Yanai studies foreign languages.” [məŋꓦ jɛnˊ aiꓦ tuˊ waiꓦwənˊ]

---

[i] L1/L2 status was chosen instead of proficiency for the sake of simplicity; models comparing intermediate and advanced L2 speakers did not produce any significant results, with the exception of the F1 of i/ŋ.

The effects of perceived ethnicity and prosodic accuracy on intelligibility, comprehensibility, and accentedness in L2 Mandarin Chinese

Robert Squizzero

Abstract

Separate traditions of research have examined the impact of linguistic factors and social factors on the intelligibility, comprehensibility, and accentedness of second language (L2) speech, but studies that simultaneously investigate social and linguistic factors are rare, and studies in both traditions are rarely conducted on languages other than English and outside of Western social and cultural environments. This study explores the effects of utterance-level prosody and of speaker ethnicity on perception of L2 Mandarin Chinese speech. 292 first language (L1) Mandarin listeners were asked to select the correct transcriptions of each of six sentences spoken by two male L2 Mandarin speakers who differed in their prosodic accuracy. While listening to each set of sentences, a picture of an Asian face or a White face was displayed on the listener's screen. Results indicate that participants were significantly more likely to select the correct transcription of each sentence when they heard the speaker with high prosodic accuracy and believed that the speaker was ethnically Chinese. Listeners also rated speakers' comprehensibility, accentedness, and perceived personal characteristics; listeners found the speaker with high prosodic accuracy to be more comprehensible and less accented, but listeners also rated a speaker as more comprehensible and less accented when they believed the speaker was ethnically Chinese. This study demonstrates that a link between linguistic and social factors exists in processing L2 speech, even outside of the social, cultural, and linguistic environments typically used as a setting for investigation of L2 speech perception. (244 words)

Introduction

The end goal of L2 pronunciation teaching and learning has been a matter of scholarly debate for over 70 years, ever since Abercrombie's (1949) assertion that L2 speakers should aim for their pronunciation to be of sufficient quality to produce intelligible and reasonably comprehensible speech, rather than to be of the same quality as L1 speakers[1]. Research has shown that the majority of L2 speakers will never pass for L1 speakers (Flege, Munro & MacKay 1995) regardless of how diligent they may be in their efforts to improve their pronunciation. Despite its unfeasibility, 'native-like' L2 pronunciation today remains a commonly sought after goal, whether by teachers and students in classrooms (Munro & Derwing 2020), developers and users of language learning apps and software, or coaches and clients in the burgeoning accent reduction industry (Derwing & Munro 2015b).

To define terms that will frequently be used in this paper: *intelligibility, comprehensibility,* and *accentedness* are used as defined by Munro, Derwing & Morton (2006). *Intelligibility* refers to actual understanding of the meaning of a word or utterance, in other words, more understanding on the part of the listener is equivalent to more intelligibility on the part of a word or utterance. *Comprehensibility* refers to the ease with which a word or utterance can be understood, in other words, less effort on the part of the listener is equivalent to more comprehensibility on the part of a word or utterance. *Accentedness* refers to the degree to which the pronunciation of an utterance sounds different from an expected production pattern.

[1] The distinction between first and second language speakers in this paper is made solely because certain elements of spoken language almost always systematically vary between individuals who have begun to acquire a language during the critical period of language acquisition and individuals who have acquired a language after the critical period. The distinction between first and second language speakers in many contexts is problematically reductionist in today's transnational, translingual world. For a thorough discussion of these issues, see (Canagarajah 2015).

Phonetic factors and intelligibility, comprehensibility, accentedness, and personality

The intelligibility-comprehensibility-accentedness framework has problematized the popular belief that all linguistic deviations from L1 speaker norms should be corrected in the L2 pronunciation classroom. The study first proposing this tripartite framework demonstrated that accented speech is often highly intelligible and comprehensible (Munro & Derwing 1995). In a commentary on their original work, Munro and Derwing (2020) note that some, though not enough, research has been conducted with the goal of identifying the features of accent that are likely to have a negative impact on listener understanding. Several of these studies have identified prosody as an important factor in perceived L2 comprehensibility and accentedness. Yang (2016), for example, found that L1 English L2 Mandarin speakers with greater prosodic accuracy were rated as more comprehensible and less accented than speakers with less prosodic accuracy. In a study of L1 Japanese L2 English speakers, Saito, Trofimovich & Isaacs (2016) found that prosodic factors, including intonation and word stress, were closely related to comprehensibility and accentedness ratings. Another study found that word stress was important in predicting comprehensibility ratings among L1 French L2 English speakers (Isaacs & Trofimovich 2012). Lastly, in a study of L2 English speakers of various L1s, Kang (2010) found that the strongest predictor of listeners' accentedness ratings was overall pitch range.

Outside of the intelligibility-comprehensibility-accentedness framework, research has been conducted examining the effects of language proficiency, fluency, and accentedness on the perceived personal characteristics of L2 speakers. One study found that listeners who rated American L2 speakers of Mandarin Chinese as more proficient in the language also considered the more proficient speakers to be more competent, intelligent, and ambitious relative to less proficient speakers (Wible & Hui 1985). In another study, more fluent American L2 Mandarin

speakers received higher ratings in evaluation, potency, and activity relative to less fluent speakers (White & Li 1991). Dragojevic & Giles (2016) found that comprehensibility (a.k.a. *processing fluency*) resulted in lower ratings in traits pertaining to solidarity for a speaker with an Punjabi accent relative to a speaker of Standard American English, and Dragojevic et al. (2017) and Vaughn & Whitty (2020) found that a more heavily-accented L2 English speaker was given lower ratings in traits pertaining to status than a less heavily accented speaker.

Despite the importance of phonology in L2 speech perception, a recent meta-analysis reveals that little scholarly research has been conducted on acquisition of L2 Mandarin phonology relative to morphosyntax (Zhang 2018). Zhang also writes that most of the work that has been done on L2 Mandarin phonology has focused on lexical tone. In one study, researchers demonstrated the importance of accurate productions of lexical tone by flattening the f0 contour of naturally-produced Mandarin speech; this flattening led to a significant reduction in intelligibility (Chen & Yang 2021).

One of the few studies that has been conducted on the acquisition of utterance-level prosody is Yang's (2011) research, which investigated the productions of L2 Mandarin L1 American English. Yang found that L2 Mandarin speakers made larger breaks between the subject and predicate portions of sentences, and he also found that they tended to break up longer prosodic phrases into smaller units. Additionally, Yang found that f0 patterns differed between L1 and L2 speakers, with more tonal coarticulation present in L1 speech relative to L2 speech, specifically via undershoot of the targets of lexical tones, with L2 speakers tended to hyperarticulate.

Social factors and intelligibility, comprehensibility, accentedness, and personality

Parallel to the study of phonetic factors affecting intelligibility, comprehensibility, accentedness, and perceived personal characteristics of L2 speakers has been a line of research on the effects of social factors. Rubin (1992) authored the first scholarly investigation to demonstrate that social factors affect the intelligibility (which he referred to as 'comprehension') and perceived accentedness of L2 speech. In Rubin's study, undergraduate students listened to two lectures recorded by a female L1 speaker of Standard American English. While listening to each lecture, a photograph of either an Asian woman or a Caucasian woman was projected onto the front of the room. When the photograph of the Asian woman was projected, participants understood significantly less of the lecture material and rated the speaker as significantly more accented relative to when the photograph of the Caucasian woman was projected. Rubin interpreted this result as listeners having stereotyped the ethnically Asian instructor, assuming that her speech would be more accented and less comprehensible. Rubin goes on to argue that it can be stereotypes about a speaker's ethnicity, rather than a speaker's actual adherence to L1 pronunciation norms, that result in listeners trying less hard to comprehend a lecture and even in listeners hearing a non-existent accent, a phenomenon eventually named *reverse linguistic stereotyping* (Kang & Rubin 2009). Therefore, Rubin argues, rather than placing a burden on the L2 speaker by asking her to improve her otherwise intelligible pronunciation, listeners should attempt to improve their listening ability, and interventions aimed at reducing bias towards L2 speakers should be explored[2]. While the intervention attempted in Rubin's (1992) Study 3 was unsuccessful, later studies showed that it is possible for an intervention to reduce negative

[2]This argument is laid out in greater detail in later publications; see (Gluszek & Dovidio 2010; Lippi-Green 2012).

attitudes towards L2 speakers and to improve their perceived comprehensibility (Baese-Berk, Bradlow & Wright 2013; Kang, Rubin & Lindemann 2015).

In the thirty years that have passed since the publication of Rubin's work, numerous studies have explored the mechanism by which social factors – chiefly ethnicity – can affect intelligibility of L2 speech. An alternative theory to the mechanism of *reverse linguistic stereotyping* is the *expectation mismatch effect*. The expectation mismatch effect (Babel & Russell 2015; McGowan 2015; Gnevsheva 2018; Hanulíková 2021) asserts that when listeners are presented with social information about a speaker, that the attentional weights that listeners give to particular phonetic cues will change based on listener expectations about the speech of members of the social category to which the speaker belongs (Pierrehumbert 2003)[3]. Difficulty in comprehension occurs when listeners are presented with information that is incongruent with their expectations, such as when seeing an Asian speaker but hearing Standard American English. McGowan (2015) demonstrated the expectation mismatch effect in a study in which listeners heard Chinese-accented English while being presented with an Asian face (congruent with assumed listener stereotypes) or a Caucasian face (incongruent with assumed listener stereotypes); listeners performed better in the congruent condition relative to the incongruent condition.

All known studies investigating the effects of social factors on intelligibility and/or comprehensibility since 1992, regardless of the mechanism proposed to account for these effects, have been conducted on English, with the exception of two studies which were also conducted on Germanic languages spoken in Western social and cultural contexts (Hanulíková 2018; 2021).

[3] For an overview of the process by which individuals are categorized into social groups, see (Freeman & Ambady 2011). For an overview of how stereotype activation occurs in general, see (Bassili & Brown 2005), or for an overview specific to the present context, see (Squizzero 2020).

This focus has persisted despite work in social psychology showing that individuals in Western cultural contexts differ from individuals in other cultural contexts due to factors such as the concept of the independent self, thought to be unique to Western cultures (Markus & Kitayama 1991). Three separate studies on Mandarin Chinese (Squizzero n.d.; 2020; 2022a) have cast doubt on the generalizability of social factors on perceived accentedness; unlike in Rubin's (1992) study, the listeners in the three studies on Mandarin did not rate the speech ostensibly recorded by individuals of non-Chinese ethnicity as more accented than the same speech when listeners believed that it had been recorded by individuals of Chinese ethnicity. While some scholars have claimed that this effect is due to a general lack of ethnic bias in Chinese society (Liu, Li & Yue 2010), significant effects favoring the personalities of ethnically Chinese women relative to ethnically non-Chinese women were observed in all three of the aforementioned studies conducted on Mandarin (Squizzero n.d.; 2020; 2022a).

The three studies conducted on Mandarin, mentioned in the previous paragraph, are not the only language attitudes investigations of the effects of ethnicity on perceived personal characteristics of L2 speakers to have been carried out. In a study of attitudes towards L1 Korean-L2 English and L1 German-L2 English speakers, L1 Chinese-L2 English participants in Australia rated L1 Korean speakers lower on personality characteristics relative to L1 German speakers, but only when they were able see the speaker (Lu & Gnevsheva 2021). There exist other studies (e.g. Nelson, Signorella & Botti 2016) that claim to have looked at the effects of ethnicity on perceived personal characteristics of L2 speakers, but these studies conflated ethnicity and L1 speaker status by assuming that L2-accented speech is racialized in the minds of listeners.

The present study

In the conclusion of their recent volume on L2 pronunciation research, Derwing & Munro (2015) call for more research investigating situational factors that may affect intelligibility and comprehensibility in L2 speech. The present study responds to that call by exploring perceived speaker ethnicity as a relevant situational factor for intelligibility and comprehensibility. Derwing and Munro also call for work on languages other than English, specifically identifying prosodic phenomena in Mandarin as an area ripe for future research. This study responds to that call by replicating prior work on Mandarin prosody (Yang 2016), but with a larger and more diverse sample of respondents. This study also responds to Yang's (2016) call for further studies investigating the relationship between comprehensibility and accentedness in Mandarin to confirm the importance of teaching prosody in L2 Mandarin pronunciation instruction. The present study investigates whether the perceived ethnicity and prosodic accuracy of a second language (L2) speaker of Mandarin Chinese can affect the intelligibility and perceived comprehensibility of their speech as well as perception of their accentedness and personal characteristics. Specifically, this study aims to answer the following research questions:

1. Does ethnicity affect perception of intelligibility, comprehensibility, accentedness, and perceived personal characteristics of L2 Mandarin speakers?
2. Does a speaker's prosodic accuracy affect perception of intelligibility, comprehensibility, accentedness, and perceived personal characteristics of L2 Mandarin speakers?
3. How does the interrelatedness of intelligibility, comprehensibility, and accentedness in a Chinese social, cultural, and linguistic environment compare to their interrelatedness in Western social cultural, and linguistic environments?

Methodology and Materials

Respondents. A total of 1203 L1 Mandarin speakers in Mainland China were recruited via the online survey platform Wenjuan (https://www.wenjuan.com; 问卷网). 814 respondents were excluded for indicating that they had not used headphones during the study, and an additional 97 respondents were excluded from analysis for failing one or both manipulation checks, leaving a total of 292 respondents. The remaining 292 respondents consisted of 128 men, 150 women, and 4 neither or both. The sample had a mean self-reported age of 28.3 years (median: 27, range: 18 to 55 years). 66 respondents (22.6%) reported having been abroad. 3 respondents reported being L1 speakers of another language in addition to Mandarin (Baotou dialect of Jin Chinese: 1, Hmong: 1, Sichuanese: 1). As for ethnicity, 245 respondents identified as Han, 23 Zhuang, 16 Uyghur, 13 Hui, 11 Manchu, 1 Hmong, 1 Mongolian, 1 Tujia, and 1 Yi. The sum of individuals per ethnic group exceeds 292 because 14 individuals identified as belonging to more than one ethnic group.

Auditory Stimuli. Auditory stimuli were selected from a corpus of Mandarin speech recorded both by L1 Mandarin speakers and L1 English speakers (Yang 2011). In addition to the prosodic differences analyzed by Yang, an acoustic study (Squizzero 2022b) identified several differences between L1 Mandarin and L1 English speakers. Relative to L1 Mandarin speakers, L1 English speakers tend to have a fronter /u/, a backer /a/ in open syllables, and a fronter /ei/. Also, the vowel dynamics of /y/, /a/ in open syllables, and /i/ before /ŋ/ differ significantly between L1 Mandarin and L1 English speakers. Lastly, there are differences between L1 Mandarin and L1 English speakers in the center of gravity (CoG) of /ɕ/, with L1 English speakers tending to produce /ɕ/ either with a lower CoG, closer to English [ʃ], or a higher CoG, closer to English [s].

Six sentences from two different L2 speakers from were selected, for a total of twelve sentences. Twelve sentences were chosen in order to be consistent with the number of target sentences in an earlier perception study (Yang 2016). The sentences were chosen because they contain the greatest number of vowel and consonant features that have been shown in an acoustic study (Squizzero 2022b) to differ between L1 Mandarin and L1 English speakers, and because each sentence has a counterpart that forms a minimal pair for utterance-level prosody. The two speakers, one of advanced Mandarin proficiency and one of intermediate Mandarin proficiency, were categorized by Yang as advanced and intermediate based on their prosodic accuracy. The specific two speakers whose recordings were used in this study were chosen based on the difference in acoustic backness between the phonemic high rounded vowels /y u/, based on the results of prior work (Squizzero 2022b); this coarse measure of accentedness was selected because it is the only vowel pair that is expected to cause ambiguity at the phonemic level and therefore is expected to be the most salient. The advanced speaker had the greatest normalized F2 distance between /y u/ of all advanced speakers, and the intermediate speaker had the third least normalized F2 distance between /y u/ of all male intermediate speakers. The intermediate speakers who had the least and second least normalized F2 distance between /y u/ of all male intermediate speakers were not selected because neither of these speakers had at least six recorded utterances without one or more syllable-level lexical tone errors, as determined by the author's auditory analysis.

Visual stimuli. To operationalize ethnicity, two photographs, one of a White man and one of an Asian man (Fig. 1) were selected from the Chicago Face Database (Ma, Correll & Wittenbrink 2015)(CFD), a repository of standardized, normed photographs of individuals. The two photos were chosen by analyzing the CFD norming data as follows: 1) each photo

Figure 1 The White (left) and Asian faces used in the experiment.

Speaker 1: Intermediate proficiency, White face Speaker 2: Advanced proficiency, Asian face	Speaker 1: Advanced proficiency, Asian face Speaker 2: Intermediate proficiency, White face
Speaker 1: Intermediate proficiency, Asian face Speaker 2: Advanced proficiency, White face	Speaker 1: Advanced proficiency, White face Speaker 2: Intermediate proficiency, Asian face

Table 1 Experimental groups

produced high gender perception accuracy, i.e. above 90% of respondents identified each of the two photos as containing a man, 2) each photo produced high racial perception accuracy, i.e. above 90% of respondents identified the photo of the White person as containing a White person and above 90% of respondents identified the photo of the Asian person as containing an Asian person, 3) the two photos were rated comparably for age (36.82 years for the White photo vs. 33.84 years for the Asian photo), and 4) the two photos were rated comparably along all norming adjectives used in the CFD norming study except for the adjective "prototypic," (difference between the two faces: 2.34 points, absolute difference: 3.47 points, maximum possible difference: 84 points). The author judged "prototypic" to be non-informative for the purposes of this study and therefore excluded it *a priori* from the analysis of norming data.

Procedure. The experiment used a 2 x 2 factorial design, with the ethnicity of the face and the proficiency of the speaker as within-participant independent variables. Speaker proficiency

was determined by Yang (2016) based on the speaker's prosodic accuracy; any further reference to a stimulus speaker's proficiency and prosodic accuracy are interchangeable for the remainder of this paper. The order of presentation of the stimuli was used as a between-participants variable. Table 1 shows the four experimental groups to which respondents were randomly assigned.

Participants first gave consent to participate in the study, which included agreement that they were at least 18 years of age, free of any known speech or hearing disorders, and that they would use headphones during the study. After giving consent, participants were shown a photograph of a White man or an Asian man. Although actual photographs of the speakers were not used, respondents were told that they were about to listen to six sentences recorded by the person in the photo. Respondents were informed that they would be permitted to replay the sentences, but they were asked to respond to the study's questions as quickly as possible. As a manipulation check, respondents were asked, prior to listening, if the person in the photograph looked like Chinese person, a foreign person, or an overseas Chinese person, the last of which refers to a person of Chinese ethnicity or descent but not necessarily of Chinese nationality (Simplified Chinese: 外籍华人; pinyin: wàijí huárén). While the English term 'foreign' typically refers to nationality rather than ethnicity, the Chinese term for 'foreign' (Simplified Chinese: 外国人; pinyin: wàiguórén) was used because it is the Mandarin term that is commonly used for identifying individuals who are not perceived as being of Chinese ethnicity. For the purposes of this study, participants were considered to have passed the manipulation check for the White face if they indicated that the person in the photo looked foreign (pass rate: 1017/1203, 84.5%), and they were considered to have passed the manipulation check for the Asian face if they indicated that the person in the photo looked

Chinese or overseas Chinese (Chinese: 753/1203 62.5%, overseas Chinese: 295/1203, 24.5%, either Chinese or overseas Chinese: 1048/1203, 87.0%). After the manipulation check, participants were taken to the next survey page, which also displayed the same photograph of the speaker. Respondents listened either to six sentences spoken by a L2 Mandarin speaker of intermediate proficiency or by a L2 Mandarin speaker of advanced proficiency. After listening to each sentence, participants chose between two transcriptions, which differed only in utterance-level prosody, of the sentence that they had just heard. The order in which the six sentences were presented was randomized, and the order of the two answer choices were also randomized for each sentence. After listening to all six sentences, participants rated the speaker's overall comprehensibility and foreign accentedness. Participants were also asked to rate four of the speaker's personal characteristics based on what they heard. The four personal characteristics scales (rude/polite, unlikeable/likeable, stupid/smart, and kind/arrogant) were chosen based on prior work (Squizzero n.d.; 2020). Participants were then taken to the next survey page, which displayed the same photograph of the speaker, and were asked to indicate whether they believed that the speaker was a native speaker of Mandarin. Participants then repeated the experiment, but with the photograph and recordings that they had yet to listen to. Finally, participants were asked to confirm that they had used headphones during the experiment and to provide sociodemographic information about themselves.

Analysis. Each intelligibility item was scored as correct or incorrect, and overall accuracy rates for each speaker and photograph were calculated. Comprehensibility, accentedness, and personality characteristics scores were also calculated for each speaker and photograph. Data analysis was conducted using the tidyverse suite of packages (Wickham et al. 2019) for R (R Core Team 2021), and visualization was done using ggplot2 (Wickham 2016).

Three linear mixed effects models were fit to the comprehensibility, accentedness, and personality trait data[4]. Comprehensibility and accentedness ratings were used as the dependent variable in each of their respective models. In the personality trait model, the four personality trait measures were pooled for analysis based on earlier studies (Squizzero n.d.; 2020). Two logistic mixed effects models were fit to the intelligibility and perceived L1 status data, with intelligibility and perceived L1 status used as the dependent variable in each of their respective models. All linear and logistic models included speaker ethnicity, speaker proficiency, and order as main effects. The perceived L1 status model also included intelligibility, comprehensibility, and accentedness as main effects. Experience abroad was not included as an effect in any model, as its inclusion did not improve the fit for any of the statistical models according to likelihood ratio tests. Aside from main effects, all models included participant as a random effect, and, the model for intelligibility included sentence as a random effect. Inclusion of two-way interaction effects among the main independent variables were selected separately for each model based on visual inspection of interaction plots (Fig 2); each interaction was included in a model if the lines in the corresponding interaction plot intersected. Because of the number of main effects in the perceived L1 status model, interaction plots are not shown, but no lines crossed in any of the interaction plots for perceived L1 status. All models were fit using the lmerTest package (Kuznetsova, Brockhoff & Christensen 2017). The interrelatedness of intelligibility and comprehensibility,

[4] Because comprehensibility, accentedness, and personal characteristics ratings are ordinal data, not interval data, cumulative link mixed models (CLMMs) were also run for each of the linear mixed models (LMMs). All significant effects within each of the three LMMs were also significant in the CLMMs, and all non-significant effects in LMMs were also non-significant in the CLMMs. The directions and relative magnitude of the effects were the same across LMMs and CLMMs. LMMs are reported instead of CLMMs for ease of interpretability.

intelligibility and accentedness, and accentedness and comprehensibility were compared using Spearman's rho.

Results

Intelligibility

In the logistic mixed model for intelligibility (Table 2), the main effect for order was significant, with participants less likely to choose the correct transcription of the second set of sentences that they listened to (first block: 65% correct, second block: 62% correct, n=3504). Based on the interaction plots, interaction effects for proficiency and ethnicity and for ethnicity and order were included in the model. The interaction effect for ethnicity and order was significant; when participants saw an Asian face for the first speaker and a White face for the second speaker, they were less likely to choose the correct transcription for the second speaker (White first: 66% correct, White second: 58% correct, Asian second: 65% correct). Lastly, the interaction effect for proficiency and ethnicity was significant, reflecting participants' tendency to choose the correct transcription when they were presented with an Asian face and the advanced guise (Asian + advanced: 73%, White+advanced: 63%, Asian+intermediate: 59%, White+intermediate: 60%).

	Estimate	Std. Error	z value	Pr(>|z|)
(Intercept)	0.6081	0.1820	3.341	0.0008 ***
proficiencyadvanced	0.1585	0.1163	1.363	0.1730
ethnicityasian	-0.2458	0.1484	-1.656	0.0977 .
order	-0.3250	0.1161	-2.799	0.0051 **
proficiencyadvanced:ethnicityasian	0.5401	0.1844	2.929	0.0034 **
ethnicityasian:order	0.3613	0.1828	1.976	0.0481 *

*Table 2 Summary of logistic mixed model for intelligibility. Significance codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1*

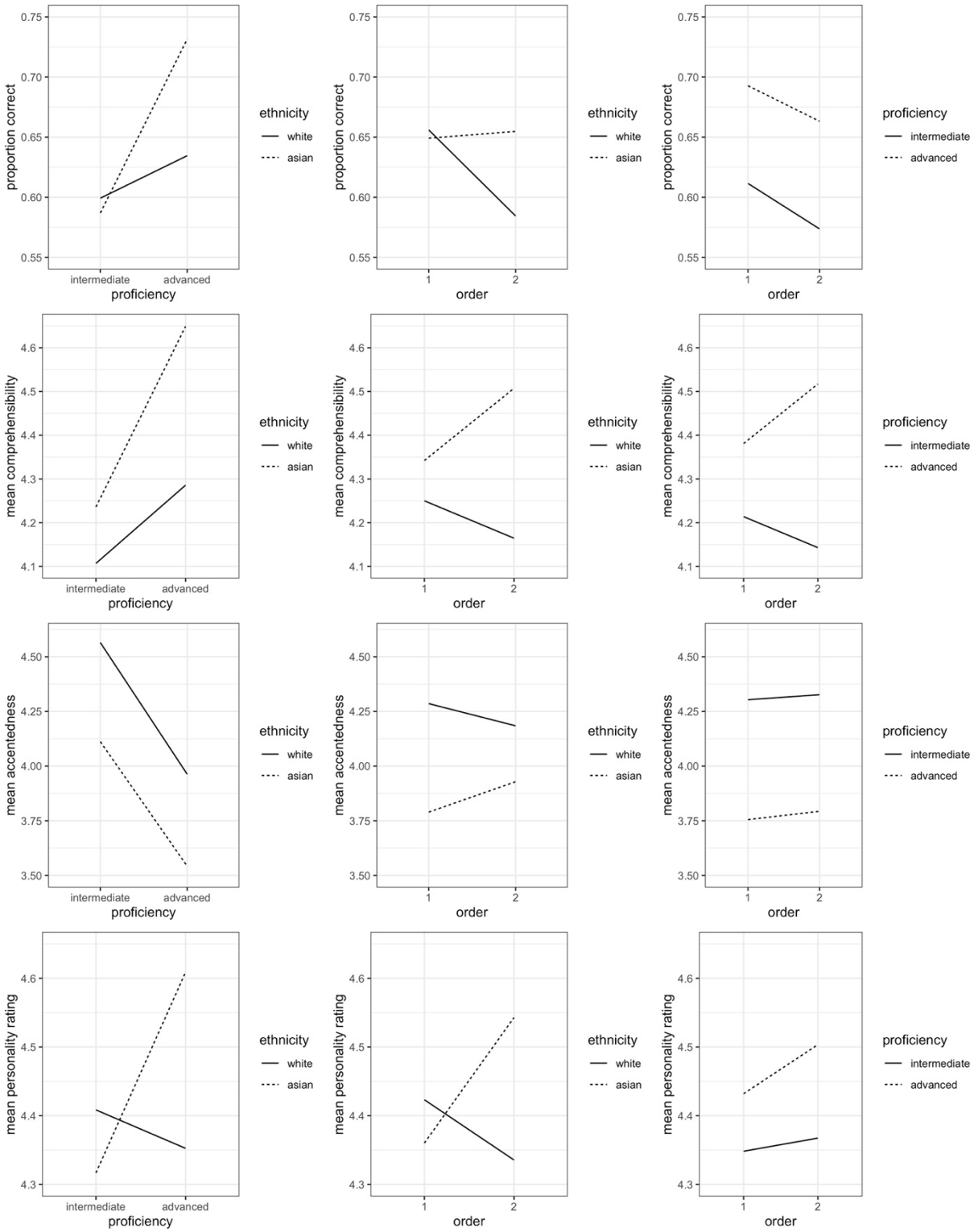

Figure 2 Interaction plots between proficiency and ethnicity, order and ethnicity, and order and proficiency. Ratings for comprehensibility, accentedness, and personality are on a scale of 1-6.

Comprehensibility

A summary of the linear mixed model for comprehensibility is shown in Table 3. The main effect for proficiency was significant and positive in the model, indicating that the advanced guise was rated as more comprehensible than the intermediate guise (mean advanced: 4.45, mean intermediate: 4.18, n = 584). The main effect for ethnicity was also significant and positive, showing that the speaker represented by an Asian face was rated as more comprehensible (mean: 4.42) than the speaker represented by a White face (mean: 4.21).

Accentedness

Table 4 displays a summary of the linear mixed model for accentedness. The main effect for proficiency was significant and negative, indicating that the advanced guise was rated as less accented than the intermediate guise (mean advanced: 3.77, mean intermediate: 4.32, n = 584). The main effect for ethnicity was also significant and negative, reflecting lower ratings for the speaker represented by the Asian face (mean: 3.86) as compared to the White face (mean: 4.23).

	Estimate	Std. Error	df	value	Pr(>|t|)
(Intercept)	3.97665	0.12926	410.44466	30.764	0.0000 ***
ethnicityasian	0.24797	0.06924	289.00000	3.581	0.0004 ***
proficiencyadvanced	0.29632	0.06918	289.00000	4.283	0.0000 ***
order	0.04304	0.06887	289.00000	0.625	0.5326

*Table 3 Summary of linear mixed model for comprehensibility. Significance codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1*

	Estimate	Std. Error	df	value	Pr(>|t|)
(Intercept)	4.5372	0.1809	345.3075	25.081	0.0000 ***
ethnicityasian	-0.4364	0.1021	289.0000	-4.273	0.0000 ***
proficiencyadvanced	-0.5858	0.1020	289.0000	-5.741	0.0000 ***
order	0.0123	0.1016	289.0000	0.121	0.9040

*Table 4 Summary of linear mixed model for accentedness.. Significance codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1*

	Estimate	Std. Error	df	value	Pr(>|t|)
(Intercept)	4.54324	0.17293	331.59745	26.272	0.0000 ***
ethnicityasian	-0.50246	0.31321	293.16091	-1.604	0.1097
proficiencyadvanced	-0.05690	0.10223	331.59746	-0.557	0.5782
order	-0.08832	0.10177	331.59745	-0.868	0.3861
ethnicityasian:proficiencyadvanced	0.35099	0.19751	289.00000	1.777	0.0766 .
ethnicityasian:order	0.27424	0.19664	288.99999	1.395	0.1642

*Table 5 Summary of linear mixed model for personality. Significance codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1*

	Estimate	Std. Error	z value	Pr(>|z|)	
(Intercept)	-0.52499	0.60714	-0.865	0.3872	
proficiencyadvanced	0.34239	0.22323	1.534	0.1250	
ethnicityasian	1.82136	0.26188	6.955	0.0000	***
order	-0.02826	0.20809	-0.136	0.8920	
comprehensibility	0.37464	0.11038	3.394	0.0007	***
accentedness	-0.60070	0.10399	-5.777	0.0000	***
intelligibility	-0.72226	0.50045	-1.443	0.1490	

*Table 6 Summary of logistic mixed model for perceived L1 status. Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1*

Personality

A linear mixed model summary for personality is shown in Table 5. Based on inspection of the interaction plots (Fig. 1), interaction effects were included in the model for ethnicity and proficiency and for ethnicity and order. While none of the main or interaction effects in the model reached significance, a trend is observed for the interaction of ethnicity and proficiency, indicating that the advanced guise paired with the Asian face was rated more favorably than the intermediate guise paired with the White face.

Perceived L1 status

Lastly, a logistic mixed model summary for perceived L1 status is shown in Table 6. Ethnicity and comprehensibility were highly significant and positive, and accentedness was highly significant and negative. In other words, respondents were more likely to guess that the speaker was an L1 Mandarin speaker if they saw the Asian photo, if they found the speaker easier to understand, or if they found the speaker to be less accented. The largest effect size, by far, was ethnicity (1.82), followed by accentedness, (-0.60), then comprehensibility (0.37). In fact, 55.5% of respondents indicated that they believed that the speaker's L1 was Mandarin if they saw the Asian photo, while only 19.2% of respondents believed that the speaker's L1 was Mandarin if they saw the White photo. Neither language proficiency nor intelligibility were significant in the model, though more respondents did believe that the advanced speaker was an L1 Mandarin speaker (41.4%) relative to the intermediate speaker (33.2%).

	Spearman's rho	*P*-value
intelligibility-comprehensibility	0.30	0.0000 ***
intelligibility-accentedness	-0.11	0.0086 **
accentedness-comprehensibility	0.09	0.0304 *

*Table 7 Spearman's rank correlation rho, with p-values. Significance codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1*

Relationships between intelligibility, comprehensibility, and accentedness

Correlations between intelligibility and comprehensibility, intelligibility and accentedness, and accentedness and comprehensibility were assessed by calculating Spearman rank correlation coefficients (Spearman's rho). Spearman's rho is an indication of the strength and direction of a relationship between two ordinal variables. The statistic ranges from -1 to 1, with -1 indicating a perfect negative monotonic relationship, 1 indicating a perfect positive monotonic relationship, and 0 indicating no monotonic relationship between the two variables. The low *p* values (Table 7) indicate that there is at least some association between the two variables of each of the three pairs assessed. Various rules of thumb exist for interpreting the strength of the association based on the value of Spearman's rho (Akoglu 2018), but results indicate a moderate-to-weak positive correlation between intelligibility and comprehensibility, a weak-to-negligible negative correlation between intelligibility and accentedness, and a weak-to-negligible positive correlation between accentedness and comprehensibility. The statistically significant correlation between accentedness and comprehensibility is likely to be a spurious result; the positive correlation coefficient indicates that *more* accented speakers are also rated as more comprehensible.

Discussion

In a study using the same corpus of recordings as auditory stimuli, Yang (2016) found that foreign accentedness and comprehensibility were highly correlated; however, results from the present study showed that the correlation between foreign accentedness and comprehensibility was weak to negligible, and possibly spurious. The present study's results, therefore, are more in line with the original study on intelligibility, comprehensibility, and accentedness (Munro &

Derwing 1995). The source of the difference in results between the present study and Yang's study is likely rooted in methodology; Yang's study included stimuli with lexical tone errors.

When invoking the term prosody in lexical tone languages like Mandarin, it is important for authors to be explicit as to whether they consider lexical tones to be part of the prosody; several scholars have argued that tones function as autonomous phonological units and should be analyzed as segments (Lin 1989; Duanmu 1990; 1994; Hyman 2011). If a researcher working on a lexical tone language believes that lexical tones should indeed be considered part of the prosody, the researcher should still be deliberate in distinguishing lexical tone from prosodic intonation. Such a distinction is highly germane to a discussion of intelligibility, comprehensibility, and accentedness in a language like Mandarin Chinese because accentedness implies linguistic differences at a subphonemic level, and lexical tone in Mandarin clearly operates at the phonemic level. While the subset of stimuli that Yang used reflects greater ecological validity in that lexical tone errors are common in L2 Mandarin speech (Zhang 2018), selecting sentences with lexical tone errors in an experiment on Mandarin is akin to selecting stimuli in an experiment on English with errors that involve substitutions of phonemic vowels or consonants. The present study therefore casts doubt upon Yang's conclusion that reduction of foreign accent, in general, is more important in L2 Mandarin than in L2 English; rather, it would be more accurate to say that lexical tone is a feature that is of the utmost importance for comprehensibility in Mandarin (and almost certainly in other lexical tone languages by extension). Yang is correct, however, in saying that it is critical for L2 Mandarin speakers to acquire prosody similar to that used by L1 speakers (assuming prosody includes lexical tone). Lexical tone has a very high functional load in Mandarin, and functional load errors have been shown to have a large negative impact on comprehensibility ratings (Munro & Derwing 2006).

Results show that prosodic accuracy, including intonation, affects both comprehensibility and accentedness of L2 Mandarin speakers, consistent with prior work on English (Saito et al. 2016). This result is also consistent with prior perceptual work using the same corpus of stimuli (Yang 2016). Being that the stimuli in the present study were free of lexical tone errors, the present study's results suggest that segmental, intonational, and/or sentence break deviations from L1 norms also contribute to the perceived ease of understanding a Mandarin speaker. However, the present study is unable to determine whether one, two, or all of these factors contribute to comprehensibility; a future study using synthetic stimuli to manipulate the segmental properties, intonational properties, and break properties would allow for an understanding of which factors should be prioritized in L2 Mandarin pronunciation instruction.

Unlike comprehensibility and accentedness, results indicate that intelligibility was not affected by adherence to L1 Mandarin utterance-level prosody per se. This is the case even though the intelligibility task was designed around this linguistic feature, presenting participants with two transcriptions to choose from differing only in utterance-level prosody. Instead of exerting its own independent influence, prosodic accuracy interacted with perceived ethnicity to affect the intelligibility of L2 Mandarin speakers; participants performed best when they were presented with an Asian face and the advanced guise. The most likely explanation for this phenomenon is based on *reverse linguistic stereotyping*; participants' stereotyped expectations of high intelligibility based on the appearance of the speaker were confirmed by the first recording in the block of sentences that they listened to, encouraging the participants to sustain their effort for the remainder of the block.

In addition to the significant interaction between ethnicity and language proficiency, the interaction between ethnicity and block order was significant in the statistical model for

intelligibility. This result is somewhat consistent with the spirit of two existing studies: one study on L2 English in which listeners downgraded the second of two speakers when the second, more accented speaker was presented without subtitles (Vaughn & Whitty 2020), and another study in which less fluent speakers of L2 English and L2 Chinese were rated lower on evaluation, potency, and activity, but only if they were presented after a more fluent speaker (White & Li 1991). In both of these studies, the authors interpreted their results as listeners penalizing the more accented or less fluent speaker in the context of having just rated a less accented or more fluent speaker. In the present study, participants who believed that the first block of sentences was recorded by an ethnically Chinese person were significantly less likely to choose the correct transcription of a sentence in the second block, which they believed was recorded by a foreign person. This result also favors a *reverse linguistic stereotyping* interpretation; participants who believed that they were about to hear a foreign person after just having listened to an ethnically Chinese person may have assumed that the foreign person would be less intelligible, and therefore put in less effort towards understanding the foreign-looking speaker. The main effect for order in the intelligibility model, independent of ethnicity, is likely due to respondent fatigue. as respondents' intelligibility decreased in the second block relative to the first block.

Results also showed that ethnicity affects accentedness and comprehensibility of L2 Mandarin speakers. This is in contrast to earlier studies on Mandarin that used L1 stimuli (Squizzero n.d.; 2020; 2022a), in which differences in accentedness ratings based on perceived speaker ethnicity were not observed. The results in these earlier studies differed from work by Rubin (1992), who also used L1 stimuli. One possible reason for this discrepancy is a difference in the populations of the two studies; it could be that Mandarin speakers require more variation in the accentedness of a study's auditory stimuli in order for factors such as ethnicity to play a

role. Another possible reason for this discrepancy is methodological difference. One major methodological difference is the difference in stimulus length between the two studies; the earlier studies on Mandarin asked participants to rate a speaker's accentedness after hearing a single sentence, as opposed to the earlier study on English in which participants rated a speaker's accentedness after hearing a short lecture. Another potentially relevant methodological difference is that the prior studies on Mandarin used a more direct prime for ethnicity; in those studies, ethnicity was operationalized by informing the listeners, in text, of the speaker's ethnicity, rather than by the use of photographs.

Unlike for comprehensibility and accentedness, ethnicity alone did not affect perception of personal characteristics of L2 Mandarin-speaking men. This is consistent with prior studies showing that ethnicity is a not a moderating factor in perception of personal characteristics of men (Squizzero n.d.; 2020; 2022a). Prosodic accuracy also had no significant effect on the perception of personal characteristics in this study. While there are more L2 speakers of Mandarin than any language other than English (Eberhard, Simons & Fennig 2021), many of these L2 Mandarin users speak related languages as their L1s, so English-accented Mandarin and its prosodic deviations may not be familiar enough to L1 Mandarin listeners for listeners to have formed ideas, as Labovian stereotypes (Labov 1972), about the users of these linguistic forms.

The last statistical model presented in the study found that ethnicity was the strongest predictor of whether or not a person is believed to be an L1 Mandarin speaker. While this result does not directly address one of the main research questions of this study, it is relevant because studies researching ethnicity and L2 speech perception in the reverse linguistic stereotyping framework (Kang & Rubin 2009) do so based on the assumption that ethnicity is a clue to the language ability of a given speaker, under a model of person perception (e.g. Freeman &

Ambady 2011). The same statistical model found that accentedness and comprehensibility also play a role in determining whether a person is a L1 Mandarin speaker, but intelligibility does not; in other words, if a learner's goal were to be to pass for a L1 speaker, it would be more important for the learner to speak in a way that is easier for listeners to understand and in a way that matches listener expectations – how much of the learner's speech is actually understood is not a significant factor. Further research is needed, however, to better understand the generalizability of these effects on perceived L1 status.

Conclusion

This paper examined the effects of prosodic accuracy, a phonetic factor, and perceived ethnicity, a social factor, on intelligibility, comprehensibility, accentedness, and perceived personal characteristics of L2 Mandarin speakers. Both prosodic accuracy and perceived ethnicity separately influenced comprehensibility and accentedness ratings, but these factors only influenced intelligibility in combination with each other or, in the case of ethnicity, with the order of stimulus presentation. Neither ethnicity nor prosodic accuracy affected perceived personal characteristics, confirming prior work demonstrating that men are not subject to adverse perception of their personal characteristics based on their perceived ethnicity. The interrelatedness of intelligibility, comprehensibility, and accentedness more closely resembled the study which first proposed the framework (Munro & Derwing 1995) than a recent study investigating the framework in L2 Mandarin (Yang 2016). While the results of the present study do not support Yang's conclusion that reducing foreign accentedness is more important for increasing comprehensibility in L2 Mandarin relative to L2 English, this study, when viewed together with Yang's study, supports an instructional emphasis on accurate prosody in L2 Mandarin in order to ensure that L2 Mandarin users produce comprehensible speech.

References

Abercrombie, David. 1949. Teaching Pronunciation. *ELT Journal* III(5). 113–122. DOI: https://doi.org/10.1093/ELT/III.5.113

Akoglu, Haldun. 2018. User's guide to correlation coefficients. *Turkish Journal of Emergency Medicine* 18(3). 91. DOI: https://doi.org/10.1016/J.TJEM.2018.08.001

Babel, Molly & Jamie Russell. 2015. Expectations and speech intelligibility. *The Journal of the Acoustical Society of America* 137(5). 2823–2833. DOI: https://doi.org/10.1121/1.4919317

Baese-Berk, Melissa M., Ann R. Bradlow & Beverly A. Wright. 2013. Accent-independent adaptation to foreign accented speech. *The Journal of the Acoustical Society of America* 133(3). EL174–EL180. DOI: https://doi.org/10.1121/1.4789864

Bassili, John N. & Rick D. Brown. 2005. Implicit and Explicit Attitudes: Research, Challenges and Theory. In Dolores Albarracin, Blair T. Johnson, & Mark P. Zanna (eds.), *The Handbook of Attitudes*, 543–574. Mahwah: Lawrence Erlbaum.

Canagarajah, Suresh. 2015. Clarifying the relationship between translingual practice and L2 writing: Addressing learner identities. *Applied Linguistics Review* 6(4). 415–440. DOI: https://doi.org/10.1515/applirev-2015-0020

Chen, Kaidi & Chunsheng Yang. 2021. The Effect of Fundamental Frequency on Mandarin Intelligibility by L2 Learners in Quiet and Noise Environments: A Pilot Study 213–232. Springer, Singapore. DOI: https://doi.org/10.1007/978-981-15-3809-4_10

Derwing, Tracey M. & Murray J. Munro. 2015a. Future directions. In *Pronunciation Fundamentals: Evidence-based perspectives for L2 teaching and research*, 167–173. John Benjamins.

Derwing, Tracey M. & Murray J. Munro. 2015b. The ethics of second language accent reduction.

In *Pronunciation Fundamentals: Evidence-based perspectives for L2 teaching and research*, 153–166. Amsterdam: John Benjamins.

Dragojevic, Marko & Howard Giles. 2016. I Don't Like You Because You're Hard to Understand: The Role of Processing Fluency in the Language Attitudes Process. *Human Communication Research* 42(3). 396–420. DOI: https://doi.org/10.1111/hcre.12079

Dragojevic, Marko, Howard Giles, Anna Carrie Beck & Nicholas T. Tatum. 2017. The fluency principle: Why foreign accent strength negatively biases language attitudes. *Communication Monographs* 84(3). 385–405. DOI: https://doi.org/10.1080/03637751.2017.1322213

Duanmu, San. 1990. *A formal study of syllable, tone, stress and domain in Chinese languages*. Massachusetts Institute of Technology dissertation.

Duanmu, San. 1994. Against contour tone units. *Linguistic Inquiry* 25(4). 555–608.

Eberhard, David M., Gary F. Simons & Charles D. Fennig (eds.). 2021. *Ethnologue: Languages of the World.* Twenty-fou. Dallas, Texas: SIL International.

Flege, James Emil, Murray J. Munro & Ian R.A. MacKay. 1995. Factors affecting strength of perceived foreign accent in a second language. *Journal of the Acoustical Society of America* 97(5). 3125–3134. DOI: https://doi.org/10.1121/1.413041

Freeman, Jonathan B. & Nalini Ambady. 2011. A Dynamic Interactive Theory of Person Construal. *Psychological Review*. DOI: https://doi.org/10.1037/a0022327

Gluszek, Agata & John F. Dovidio. 2010. The way they speak: A social psychological perspective on the stigma of nonnative accents in communication. *Personality and Social Psychology Review* 14(2). 214–237. DOI: https://doi.org/10.1177/1088868309359288

Gnevsheva, Ksenia. 2018. The expectation mismatch effect in accentedness perception of Asian and Caucasian non-native speakers of English. *Linguistics* 56(3). 581–598. DOI:

https://doi.org/10.1007/BF00973770

Hanulíková, Adriana. 2018. The effect of perceived ethnicity on spoken text comprehension under clear and adverse listening conditions. *Linguistics Vanguard* 4(1). 1–9. DOI: https://doi.org/10.1515/lingvan-2017-0029

Hanulíková, Adriana. 2021. Do faces speak volumes? Social expectations in speech comprehension and evaluation across three age groups. *PLoS ONE* 16(10 October). DOI: https://doi.org/10.1371/journal.pone.0259230

Hyman, Larry M. 2011. Do tones have features? *Tones and Features: Phonetic and Phonological Perspectives* 50–80. DOI: https://doi.org/10.1515/9783110246223.50/MACHINEREADABLECITATION/RIS

Isaacs, Talia & Pavel Trofimovich. 2012. Deconstructing Comprehensibility: Identifying the Linguistic Influences on Listeners' L2 Comprehensibility Ratings. *Studies in Second Language Acquisition* 34. 475–505. DOI: https://doi.org/10.1017/S0272263112000150

Kang, Okim. 2010. Relative salience of suprasegmental features on judgments of L2 comprehensibility and accentedness. *System* 38(2). 301–315. DOI: https://doi.org/10.1016/j.system.2010.01.005

Kang, Okim & Donald L. Rubin. 2009. Reverse linguistic stereotyping: Measuring the effect of listener expectations on speech evaluation. *Journal of Language and Social Psychology* 28(4). 441–456. DOI: https://doi.org/10.1177/0261927X09341950

Kang, Okim, Donald L. Rubin & Stephanie Lindemann. 2015. Mitigating U.S. Undergraduates' Attitudes Toward International Teaching Assistants. *TESOL Quarterly* 49(4). 681–706. DOI: https://doi.org/10.1002/tesq.192

Kuznetsova, Alexandra, Per B. Brockhoff & Rune H.B. Christensen. 2017. lmerTest Package:

Tests in Linear Mixed Effects Models. *Journal of Statistical Software* 82(13). 1–26. DOI: https://doi.org/10.18637/JSS.V082.I13

Labov, William. 1972. Some principles of linguistic methodology. *Language in Society* 1(1). 97–120. Retrieved from doi:10.1017/S0047404500006576%0A

Lin, Yen Hwei. 1989. *Autosegmental treatment of segmental processes in Chinese phonology*. University of Texas at Austin dissertation.

Lippi-Green, Rosina. 2012. *English with an accent : language, ideology and discrimination in the United States* 2nd ed. Routledge. Retrieved from https://www.routledge.com/English-with-an-Accent-Language-Ideology-and-Discrimination-in-the-United/Lippi-Green/p/book/9780415559119

Liu, James H., Mei Chih Li & Xiao Dong Yue. 2010. Chinese social identity and inter-group relations: The influence of benevolent authority. *Oxford Handbook of Chinese Psychology* (January 2015). DOI: https://doi.org/10.1093/oxfordhb/9780199541850.013.0035

Lu, Yao & Ksenia Gnevsheva. 2021. Accentedness and personality evaluation of Asian and Caucasian second language speakers of English by Asian second language English listeners. *Journal of Multilingual and Multicultural Development*. DOI: https://doi.org/10.1080/01434632.2021.1959598

Ma, Debbie S., Joshua Correll & Bernd Wittenbrink. 2015. The Chicago face database: A free stimulus set of faces and norming data. *Behavior Research Methods* 47(4). 1122–1135. DOI: https://doi.org/10.3758/s13428-014-0532-5

Markus, Hazel Rose & Shinobu Kitayama. 1991. Culture and the self: Implications for cognition, emotion, and motivation. *Psychological Review* 98(2). 224–253. DOI: https://doi.org/10.1037/0033-295x.98.2.224

McGowan, Kevin B. 2015. Social Expectation Improves Speech Perception in Noise. *Language and Speech* 58(4). 502–521. DOI: https://doi.org/10.1177/0023830914565191

Munro, Murray J. & Tracey M. Derwing. 1995. Foreign Accent, Comprehensibility, and Intelligibility in the Speech of Second Language Learners. *Language Learning* 45(1). 73–97. DOI: https://doi.org/10.1111/j.1467-1770.1995.tb00963.x

Munro, Murray J. & Tracey M. Derwing. 2006. The functional load principle in ESL pronunciation instruction: An exploratory study. *System* 34(4). 520–531. DOI: https://doi.org/10.1016/j.system.2006.09.004

Munro, Murray J. & Tracey M. Derwing. 2020. Foreign accent, comprehensibility and intelligibility, redux. *Journal of Second Language Pronunciation* (August). 1–27. DOI: https://doi.org/10.1075/jslp.20038.mun

Munro, Murray J., Tracey M. Derwing & Susan L. Morton. 2006. The mutual intelligibility Of L2 speech. *Studies in Second Language Acquisition* 28(1). 111–131. DOI: https://doi.org/10.1017/S0272263106060049

Nelson, Larry R., Margaret L. Signorella & Karin G. Botti. 2016. Accent, Gender, and Perceived Competence. *Hispanic Journal of Behavioral Sciences* 38(2). 166–185. DOI: https://doi.org/10.1177/0739986316632319

Pierrehumbert, Janet B. 2003. Phonetic Diversity, Statistical Learning, and Acquisition of Phonology. *Language and Speech* 46(2–3). 115–154. DOI: https://doi.org/10.1177/00238309030460020501

R Core Team. 2021. R: A Language and Environment for Statistical Computing. Vienna, Austria: R Foundation for Statistical Computing.

Rubin, Donald L. 1992. Nonlanguage factors affecting undergraduates' judgments of nonnative

English-speaking teaching assistants. *Research in Higher Education* 33(4). 511–531. DOI: https://doi.org/10.1007/BF00973770

Saito, Kazuya, Pavel Trofimovich & Talia Isaacs. 2016. Second language speech production: Investigating linguistic correlates of comprehensibility and accentedness for learners at different ability levels. *Applied Psycholingusitics* 37. 217–240.

Squizzero, Robert. n.d. Perceived ethnicity, gender, and attitudes towards second language Mandarin speakers. *Journal of Multilingual and Multicultural Development*.

Squizzero, Robert. 2020. Attitudes toward L2 Mandarin Speakers of Chinese and non-Chinese Ethnicity. In Kaidi Chen (ed.), *Proceedings of the 32nd Meeting of the North American Conference on Chinese Linguistics*, 521–538. Storrs, CT.

Squizzero, Robert. 2022a. *Ambivalent sexism affects attitudes towards female second language Mandarin speakers of different ethnicities*.

Squizzero, Robert. 2022b. The role of dialectology in L2 vowel acquisition; evidence from Mandarin Chinese. In *4th International Symposium on Applied Phonetics*,. Lund, Sweden.

Vaughn, Charlotte & Aubrey Whitty. 2020. Investigating the relationship between comprehensibility and social evaluation. *Journal of Second Language Pronunciation*. DOI: https://doi.org/https://doi.org/10.1075/jslp.20022.vau

White, Michael J. & Yan Li. 1991. Second-Language Fluency and Person Perception in China and the United States. *Journal of Language and Social Psychology* 10(2). 99–113.

Wible, David S. & C. Harry Hui. 1985. Perceived Language Proficiency and Person Perception. *Journal of Cross-Cultural Psychology* 16(2). 206–222.

Wickham, Hadley. 2016. *ggplot2: Elegant Graphics for Data Analysis*. New York: Springer-Verlag. DOI: https://doi.org/10.18637/jss.v035.b01

Wickham, Hadley, Mara Averick, Jennifer Bryan, Winston Chang, Lucy McGowan, Romain François, … Hiroaki Yutani. 2019. Welcome to the Tidyverse. *Journal of Open Source Software* 4(43). 1686. DOI: https://doi.org/10.21105/joss.01686

Yang, Chunsheng. 2011. *The Acquisition of Mandarin Prosody by American Learners of Chinese as a Foreign Language (CFL)*. The Ohio State University dissertation. https://doi.org/10.1007/s13398-014-0173-7.2

Yang, Chunsheng. 2016. Intelligibility, comprehensibility, and foreign accent in L2 Mandarin Chinese. In *The acquisition of L2 Mandarin prosody: From experimental studies to pedagogical practice*,. John Benjamins.

Zhang, Hang. 2018. Current trends in research of Chinese sound acquisition. In Chuanren Ke (ed.), *The Routledge Handbook of Chinese Second Language Acquisition*, 217–233. DOI: https://doi.org/10.1201/b17627

Acknowledgements

The author would like to thank Chunsheng Yang for the use of his corpus of recorded Mandarin speech, and Yadi Peng for her contributions to the translation of the survey. The author would also like to thank Alicia Beckford Wassink, Betsy Evans, Richard Wright, and Zev Handel for their assistance with the design of this study.

GENERAL CONCLUSION

This general conclusion returns to the main research questions asked throughout the dissertation and summarizes the findings of the several studies.

Perceived speaker ethnicity can affect perception of personal characteristics of L2 Mandarin speakers.

Chapters 1 and 2 are the first known studies to demonstrate that perceived speaker ethnicity can affect perception of personal characteristics of L2 Mandarin speakers. This finding is of theoretical importance because all prior studies investigating perceived ethnicity on L2 speakers were conducted on Indo-European languages and in Western social and cultural contexts.

Effects of perceived speaker ethnicity on perception of personal characteristics of L2 Mandarin speakers differ based on speaker and listener gender.

Perception of personal characteristics of L2 Mandarin speakers varied with perceived speaker ethnicity, but across four studies, this effect was only observed in evaluations of female speakers. Listeners tended to rate a speaker's personal characteristics higher when they believed that she was huáyì, but some listeners rated a speaker higher when they believed that she was not huáyì; specifically, female listeners preferred the non-huáyì guise of one of the female speakers in the main study presented in Chapter 2.

The excursus to Chapter 2 revealed two mechanisms underlying the gendered nature of the effect for ethnicity present in the preceding two studies. The first mechanism is hostile sexism; more agreement with hostile sexist viewpoints was associated with an increase in the ratings of the huáyì guise's personal characteristics relative to the non-huáyì guise. This first mechanism is interpreted as an additive effect of sexism and ethnocentrism; both attitudes must be present in a listener in order for there to be a statistically reliable preference for a huáyì speaker. The second

mechanism was the belief that a huáyì of foreign nationality should understand Chinese culture; more agreement with such a statement was associated with an increase in the ratings of the non-huáyì guise's personal characteristics relative to the huáyì guise. This second mechanism is interpreted as listeners being more impressed by a non-huáyì having learned to speak Mandarin to a high degree of proficiency than by a huáyì of foreign nationality having done so. The mechanisms identified in this study lay groundwork for further research, and they show that ambivalent sexism theory can be a useful framework for understanding gender effects in language attitudes studies. A shortcoming of this study is that only men and women were included; future work could provide a fuller and more inclusive description of the interaction of ethnicity and gender effects by investigating attitudes towards speakers identifying outside of the gender binary.

Transfer effects from L1 North American English into L2 Mandarin Chinese are in evidence for speakers of intermediate and advanced proficiency, with specific regard to consonants and vowels.

The acoustic analysis in Chapter 3 revealed the presence of L1 transfer effects. Specifically, significant differences existed in L1/L2 status for the Mandarin vowels [a ei iŋ u y] with regard to the height, backness, or time-varying qualities of height and/or backness. Chapter 3 found that relative to L1 Mandarin speakers, L1 English speakers produced a significantly lower and backer [a] in open syllables, significantly fronter [ei u], [y] with significantly more time-varying movement in backness, and [iŋ] with significantly less time-varying movement in backness. With respect to consonants, Chapter 3's auditory analysis showed significant differences in L1/L2 status and/or proficiency for the presence of oral closure in coda nasal consonants, for velarization of [l], and for the place of articulation the alveolopalatal fricative and affricates. The acoustic analysis did not show significant differences in the production of consonants based on L1/L2 status or

proficiency, but many L2 speakers appeared to produce the alveolopalatal fricative [ɕ] with a center of gravity consistent with a fronted or retracted tongue. These analyses have identified subphonemic differences between Mandarin and English that can inform Mandarin pronunciation teaching. Future phonetics research could use articulatory methods, such as ultrasound tongue imaging or MRI, to investigate oral closure of coda nasal consonants and velarization of [l].

Both perceived speaker ethnicity and deviation from L1 Mandarin Chinese pronunciation norms in vowels and utterance-level prosody can affect accentedness and comprehensibility ratings of L2 Mandarin speakers, as well as intelligibility rates of L2 Mandarin speech.

Linear mixed-effects models in Chapter 4 showed direct effects of both perceived speaker ethnicity and Mandarin language proficiency (operationalized by prosodic accuracy) on accentedness and comprehensibility ratings; more advanced speakers and ethnically Chinese speakers were rated as more comprehensible and less accented. A logistic mixed-effects model in Chapter 4 showed that perceived speaker ethnicity and Mandarin language proficiency interact to affect intelligibility of L2 Mandarin speech; respondents were significantly more likely to choose the correct transcription of a sentence only when the speaker was both of higher language proficiency and ethnically Chinese in appearance. The intelligibility results support a reverse linguistic stereotyping explanation over an expectation mismatch explanation; under an expectation mismatch explanation, intelligibility should have been higher when the foreign-looking face was paired with the intermediate guise than when it was paired with the advanced guise, but this was not the case.

The interrelatedness of intelligibility, comprehensibility, and accentedness in a Chinese social, cultural, and linguistic environment is similar to their interrelatedness in Western social cultural, and linguistic environments.

Comparisons reported in Chapter 4 showed a moderate-to-weak positive correlation between intelligibility and comprehensibility, a weak-to-negligible negative correlation between intelligibility and accentedness, and a weak-to-negligible (and possibly spurious) positive correlation between accentedness and comprehensibility. These results confirm that intelligibility, comprehensibility, and accentedness are related, but partially independent, and suggest that a stronger foreign accent may not be associated with a strong reduction in the intelligibility or comprehensibility of L2 Mandarin speech.

Concluding thoughts

It is hoped that future descriptive work investigating sociolinguistic and phonetic effects on perception of second language speech of other non-Indo-European languages spoken in non-Western sociocultural contexts is carried out. Such work can serve to confirm or cast doubt upon the generalizability of existing theories beyond a Chinese context. In particular, questions remain about the cross-linguistic generalizability of the intelligibility-comprehensibility-accentedness framework, whether reverse linguistic stereotyping or expectation mismatch is a more consistent explanation for effects of perceived speaker social group membership on L2 speech perception, and the intersection between gender and ethnicity in language attitudes studies.