PhD Thesis

The role of job candidate's accent, ethnicity, and vocal competence on social evaluations in a hiring context

By

Yondu Mori

SUBMITTED TO THE FACULTY OF GRADUATE AND POSTDOCTORAL STUDIES

IN PARTIAL FULFILMENT OF THE REQUIREMENTS FOR THE DEGREE OF

DOCTOR OF PHILOSOPHY

in

COMMUNICATION SCIENCES AND DISORDERS

School of Communication Sciences and Disorders

Faculty of Medicine and Health Sciences

McGill University

Montréal, Québec, Canada

March 2023

Table of Contents

Abstract

People who speak with a non-standard accent can experience challenges in their social communication, particularly in a job interview. There is substantive evidence that having a non-standard accent can lead to perceived low competence compared to speakers with standard accents. Conversely, job candidates conveying high competence using their voice (i.e., tone of voice) are perceived as high in competence and are more likely to be hired. However, there is little research on the interactive effects of a job candidate's vocal cues and accent on the competence impressions formed, with most research using speech that does not express a mental or affective state. In this thesis we sought to better understand the impact of a speaker's vocal competence level, accent, and ethnic appearance on their perceived job suitability (i.e., competence and hirability) in two studies. Speakers had standard (American/Canadian English) accent or non-standard (Singaporean English) accents and expressed high (confidence/pride) or low (doubt/shame) vocal competence as a proxy of speaker's varied mental/affective states in a job interview. As a control, we measured speaker's perceived vocal warmth (attractiveness and friendliness). In Study 1, participants heard speech in a standard and non-standard accent, conveying vocal pride and shame, and judged speaker's perceived job suitability. We found that listeners less clearly differentiated between high and low vocal competence by speakers with non-standard accents. Also, vocal friendliness was a stronger mediator between speaker vocal competence level and perceived job suitability for speakers with standard compared to non-standard accents. In Study 2, we explored the interactive effects of speaker's accent and ethnic appearance, with their vocal competence level. Again, we found that vocal warmth measures strongly predicted speaker's perceived job suitability. Also, the ethnic appearance of speakers with non-standard accents did not amplify negative competence perceptions related to their accent. This work contributes to our understanding of the socio-cognitive processes of a speaker's accent when speakers are expressing a social intention. It also

expands on the prominence of competence and warmth dimensions in shaping our social impressions, showing a dominance of perceived warmth when evaluating others in a high-stakes social context.

Résumé

Les personnes qui parlent avec un accent non standard peuvent rencontrer des difficultés dans leur communication sociale, notamment lors d'un entretien d'embauche. Il existe des preuves substantielles que le fait d'avoir un accent non standard peut conduire à une perception de faible compétence par rapport aux locuteurs ayant un accent standard. À l'inverse, les candidats à l'emploi qui transmettent une compétence élevée par leur voix (c'est-à-dire le ton de la voix) sont perçus comme très compétents et ont plus de chances d'être embauchés. Cependant, il existe peu de recherches sur les effets interactifs des indices vocaux et de l'accent d'un candidat à l'emploi sur les impressions de compétence formées, la plupart des recherches utilisant un discours qui n'exprime pas un état mental ou affectif. Dans cette thèse, nous avons cherché à mieux comprendre l'impact du niveau de compétence vocale, de l'accent et de l'apparence ethnique d'un locuteur sur la perception de son aptitude à l'emploi (c'est-à-dire sa compétence et son aptitude à l'embauche) dans deux études. Les locuteurs avaient un accent standard (anglais américain/canadien) ou non standard (anglais singapourien) et exprimaient une compétence vocale élevée (confiance/ fierté) ou faible (doute/honte) comme indicateur de leurs différents états mentaux/affectifs lors d'un entretien d'embauche. Comme contrôle, nous avons mesuré la perception de la chaleur vocale du locuteur (attractivité et amabilité). Dans l'étude 1, les participants ont entendu un discours avec un accent standard et un accent non standard, exprimant la fierté et la honte vocales, et ont jugé la perception de l'aptitude à l'emploi du locuteur. Nous avons constaté que les auditeurs faisaient moins clairement la différence entre les compétences vocales élevées et faibles des locuteurs ayant un accent non standard. De plus, l'amabilité vocale était un médiateur plus fort entre le niveau de compétence vocale du locuteur et la perception de l'adéquation de l'emploi pour les locuteurs avec des accents standards par rapport aux accents non standards. Dans l'étude 2, nous avons exploré

les effets interactifs de l'accent et de l'apparence ethnique du locuteur, avec son niveau de compétence vocale. Une fois de plus, nous avons constaté que les mesures de la chaleur vocale prédisaient fortement la perception de l'aptitude à l'emploi du locuteur. De plus, l'apparence ethnique des locuteurs ayant un accent non standard n'a pas amplifié les perceptions négatives de compétence liées à leur accent. Ce travail contribue à notre compréhension des processus socio-cognitifs de l'accent d'un locuteur lorsque celui-ci exprime une intention sociale. Ils développent également la proéminence des dimensions de compétence et de chaleur dans la formation de nos impressions sociales, en montrant une dominance de la chaleur perçue lors de l'évaluation des autres dans un contexte social de première importance.

Acknowledgements

This work is a culmination of my growth as a researcher. I spent countless hours writing, reading, re-writing, re-reading, learning new skills, and challenging myself to think deeply about this study. Reaching this goal would not have been possible without the notable support of others.

I greatly appreciate the support I received from my supervisor (Dr. Marc D. Pell) along this journey. Thank you for supporting my creative endeavours to pursue this research, and for showing me the level of rigour and thoroughness needed to conduct high quality research. Also, thank you to my committee members (Dr. Meghan Clayards and Dr. Jordan Axt) who provided valuable feedback throughout this process. Thank you to my lab mates (Deirdre, Maël, Shuyi, Luisa, Marcos, Jonathan) and colleagues (Angela, Alvaro, Eve, Dahlia, Mehrgol, Max, Chelsea, Tanya, Haruka, AJ, Steffi), for imparting your worldly perspectives, offering a hand whenever you could, or just checking in. You ensured that I celebrated the small victories along the way, no matter the size or ambiguousness, and you (forcibly) taught me that taking breaks can be productive. Thank you to my friends (Rayanna, Victoria P., Candice, Victoria R., Ranon, John, Huong, Marilyn) and family who kept me grounded and inspired. Thank you to my partner, Keegan, whose support, kindness, and understanding were instrumental to me making it to the finish line.

This thesis was financially supported by my fellowships from the Social Sciences and Humanities Research Council (SSHRC Doctoral fellowship, Joseph Bombardier Doctoral fellowship) and McGill (Graduate Excellence Scholarships) and Dr. Pell's SSHRC Insight grant. Also, thank you to the SCSD/Abelson travel award and the Centre for Research on Brain, Language, and Music (CRBLM) travel award for supporting me to attend research conferences. Lastly, thank you all the participants across Canada who completed my experiments. By sharing your perspective, I am advancing our understanding of intercultural communication in Canada.

Contribution to original knowledge

This thesis is original, unpublished, independent work by the first author. It consists of two novel studies examining the impact of a speaker's vocal affective state and accent on their perceived competence in a job interview. In a secondary study we investigated the interactive effects with a speaker's ethnic appearance to gather a comprehensive picture of the factors that can impact the perceived competence of job candidates who have non-standard accents. The speech stimuli used came from the VENEC corpus (Laukka et al., 2016; Elfenbein et al., 2022) where we specifically focused on the emotions of pride and shame, as well as stimuli collected from the lab where actors conveyed confidence or doubt (Jiang & Pell, 2017). Study 2 used a novel experimental design to investigate the effect of speaker ethnic appearance by first examining the effect of ethnic appearance when all job candidates had a standard accent and comparing these findings to a study where job candidates had standard or non-standard accents. Also, the investigation of a non-standard accent (i.e., Singaporean English accent) that is spoken by native speakers of English who are also visible ethnic minorities, with a Canadian/American English accent is novel. This research may inform hiring practices and policies by highlighting the impact of a job interviewer's social connection with a job candidate and encouraging companies to welcome language and culture diversity in their workforce.

Contribution of Authors

This thesis was conceived, designed, and conducted by the first author, Yondu Mori, under the guidance of her supervisor, Dr. Marc D. Pell. The statistical approach and analysis of results were computed and determined by the first author, with advice from Dr. Pell. The thesis was written in full by the first author, with editorial suggestions from Dr. Pell, Dr. Clayards and Dr. Axt.

List of Figures

Figure

List of Tables

Introduction

Obtaining employment as an immigrant in Canada can greatly foster a sense of belonging and social integration to the country (e.g., Hyndman & Hynie, 2016). It can provide an opportunity for people to build a social network, including people outside their cultural, ethnic, or language group, and adopt new cultural practices/norms (Birman & Simon, 2014; Doucerain, Varnaamkhaasti, Segalowitz, & Ryder, 2015; Sam & Berry, 2010). However, immigrants may also be judged to have a non-standard accent, which can pose a challenge to them securing employment. Qualitative research reports that Canadian immigrants have lost their jobs or were not hired due to accent-based stereotyping (Munro, 2003). Canadian immigrants also report that their accents denied them jobs where they would be speaking with the public (e.g., receptionist, teacher, customer service worker) (Creese, 2010; Creese & Kambere, 2003), and some speakers feel they should get rid of their accent to be treated like a native Canadian English speaker (Branker, 2017). In addition, experimental quantitative research repeatedly shows that speakers with non-standard accents are perceived as less competent based solely on their accent (e.g., Fuertes et al., 2012; Gluszek & Dovidio, 2010). These results demonstrate how prevalent and impactful first-time judgments of a speaker's competence are in daily interactions between speakers with standard and non-standard accents.

Moreover, the job interview can be described as a "culturally specific speech event" (Demo, 2006, p.45). Newcomers may differ in their cultural norms related to communicative behaviour, including non-verbal cues (e.g., tone of voice) (Elfenbein & Ambady, 2003), and other display rules or conventions specific to a job interview (e.g., Campbell & Roberts, 2007; Kerekes, 2007; Manroop, Boekhorst, & Harrison, 2013). These vocal cues, such as how we express high versus low competence to a potential employer can greatly impact a speaker's perceived

competence and likelihood of being hired (e.g., Bradac, Konsky, & Elliott, 1976; Bradac & Street, 1989; DeGroot & Motowidlo, 1999; Gallois, Callan, & Palmer, 1992; Hollandsworth, Kazelskis, Stevens, & Dressel, 1979; Parton, Siltanen, Hosman, & Langenderfer, 2002; Scherer, London, & Wolf, 1973). Employers may also (un)consciously use a speaker's vocal cues to evaluate their level of cultural fit for an organization and/or country (e.g., Bye et al., 2014; Elrick, 2016; Horverak et al., 2013; Sakamoto, Ku, & Wei, 2009; Madziva, McGrath, & Thondhlana, 2016; Sakamoto, Chin, & Young, 2010).

Despite a lot of research examining the perceived competence of speakers with non-standard accents (e.g., Dragojevic & Giles, 2014; Fuse et al., 2018; Hansen et al., 2017; 2018, Hosoda & Stone-Romero, 2010), they have not gathered impressions using speech that expresses a vocal mental or affective state, such as expressing one's competence using their voice, while also considering a speaker's cultural background. Instead, the impressions have largely been based on listeners' ability to identify a speaker as having a standard or non-standard accent.

The main purpose of this study was to better understand why speakers with non-standard accents are often perceived as less competent than speakers with standard accents, particularly in job interviews. We aimed to understand the effects of speech-related cues on a speaker's perceived competence with a secondary goal of connecting this information to visual cues about a speaker's ethnicity. Specifically, we examined how vocal and ethnicity-related information about a speaker can contribute to their perceived competence and hirability (i.e., job suitability) in a job interview setting. Speakers were assessed as job candidates for a human resource manager position. This position was chosen because it typically involves a lot of spoken communication with others and has been used in previous studies examining differences in job interview outcomes by speaker's accent and/or ethnicity (e.g., Bye et al., 2014; Deprez-Sims & Morris, 2010; Hansen et al., 2018).

This study also allowed us to bridge the extant quantitative and qualitative research. Although speakers with non-standard accents may believe that their accent has a large impact on their job interview outcome, other vocal and ethnicity cues, or unrelated vocal trait inferences, such as attractiveness, may be interacting with their accent and contributing to their perceived communicative competence.

In Study 1, we examined the perceived job suitability of speakers with American English or Singaporean English accents expressing vocal pride and shame and did an acoustic analysis of the speech to characterize the vocal cues that contribute to a speaker's general perceived communicative competence when they express a mental or affective state in a high-stakes context. In Study 2, we determined the effect of a speaker's accent, ethnic appearance, and vocal cues on their perceived job suitability. These studies allowed us to determine the impact of a speaker's vocal competence level and accent on their perceived competence, and how these cues interact with cues from a speaker's physical appearance that mark their ethnicity. By the end, we address themes centered around intercultural/interethnic communication as it relates to speech processing and the social impressions we form of others.

Chapter 1

Literature Review

A job interview is a common method used to assess the social competence (or ability to communicate and cooperate with others) and intellectual competence (or ability to complete job-related tasks via education, skills, experience) of job candidates (e.g., Hofhuis et al., 2016; Kristof-Brown, 2002; Macan, 2009). It is also a high-stakes social context for job candidates and job interviewers, who aim to evaluate candidates as accurately as possible and within a limited amount of time so that a highly competent candidate is hired (e.g., Dipboye et al., 2012; Nordstrom, Hall, & Bartels, 1998). Due to this social pressure, interviewers may use heuristics to reduce the cognitive load of evaluating a candidate's behaviour from various communication channels (e.g., linguistic, vocal, and visual cues) and information sources (e.g., resume, assessments) (Derous, Buijsrogge, Roulin, & Duyck, 2016). One type of communicative cue that can (un)consciously and automatically impact the competence impressions formed, is a candidate's vocal cues (e.g., speech rate, pausing, amplitude, pitch) (Schroeder & Epley, 2015).

Job interviewers are more likely to have a positive impression of job candidates who sound competent when responding to interview questions. Specifically, candidates who speak fluently, in a loud volume, with a relatively fast speech rate, few filled (e.g., um, uh) pauses, and pauses of short duration, are more likely to be perceived as competent, confident, trustworthy, and intelligent (e.g., Bradac, Konsky, & Elliott, 1976; Bradac & Street, 1989; DeGroot & Motowidlo, 1999; Gallois, Callan, & Palmer, 1992; Hollandsworth, Kazelskis, Stevens, & Dressel, 1979; Miller, Gayfer & Powell; 2018; Parton, Siltanen, Hosman, & Langenderfer, 2002; Scherer, London, & Wolf, 1973). These vocal cues indicate a candidate's varying mental and emotional states, such as having a high level of certainty in the content of their speech, a confident personality, or feeling

enthusiastic. In contrast, job candidates are likely to be perceived as less competent when they respond to questions in a quieter volume, with a slower speech rate, many filled and unfilled (silent) pauses, and pauses of longer duration. These vocal cues may indicate candidates' uncertainty in their responses, anxiousness from the high-stakes social context (e.g., Feiler & Powell, 2016; Miller, Gayfer & Powell, 2018), or low vocal expressiveness (e.g., McGovern & Tinsley, 1978). Depending on the quantity and intensity of these vocal cues marking low competence, a job candidate may not be hired as they are perceived as uncertain, anxious, and less capable of communicating well with others.

Cognitive and social processing of job candidate's vocal competence level

The competence inferences drawn from a candidate's voice are shaped by job interviewers' engagement in cognitive and social processes. In terms of cognitive processing, when listeners hear highly fluent speech, such as the vocal cues associated with high competence, it is processed with greater ease. With this greater processing fluency, interviewers are more likely to be persuaded by candidates' responses and may perceive them as more knowledgeable and accurate in the information they present (Alter & Oppenheimer, 2009; Guyer et al., 2019; Price & Stone, 2004). Also, high processing fluency is generally associated with positive evaluations towards a target (e.g., a person) (Schwarz, 2004; Winkielman, Schwarz, Fazendeiro, & Reber, 2003). In contrast, speech that vocally expresses anxiousness and uncertainty, can be more challenging to understand. Depending on the quantity and intensity of disfluencies, it may be more difficult to accurately detect and readily activate the corresponding linguistic representations. Based on episodic memory trace theories and exemplar-based models, spoken words are stored in long-term memory with detailed traces (Goldinger, 1996; Johnson, 1997). These traces can also include information about a speaker's voice such as their relative pitch, gender (Goldinger, 1996), speech

rate (Bradlow, Nygaard, & Pisoni, 1999; Pisoni, 1997), accent, emotional state, or speech style (Sumner, Kim, King, & McGowan, 2014). Words are retrieved from memory by activating groups of these detailed traces (Goldinger, 1996). More recent hybrid models of speech processing postulate that vocal expressions of low competence may not necessarily impair speech comprehension. The dual route approach hypothesizes that speech comprehension is facilitated by the simultaneous activation of linguistic and social representations (Sumner et al., 2014). It hypothesizes that socially salient linguistic tokens, such as vocal expressions of low competence, are encoded with greater attention (Sumner et al., 2014). Thus, we may predict that when listeners hear speech conveying low competence, particularly of a mild intensity, they rapidly activate associated social attributes of the speaker. However, there may be additional cognitive effort involved as listeners engage in mentalizing processes to try to understand a speaker's communicative intention (e.g., Jiang & Pell, 2015), which may negatively affect their perceived competence.

In terms of social processing, job interviewers' impressions are influenced by their affective reaction to job candidates (e.g., Howard & Ferris, 1996; Rivera, 2015; Young & Kacmar, 1998). When candidates display various speech and non-verbal cues (e.g., lexical content, amplitude, eye contact) associated with high competence, interviewers are more likely to experience positive affect towards the candidate, positively impacting the competence impression formed (Gallois, Callan, & Palmer, 1992; Howard & Ferris, 1996). This positive affect may be related to an interviewer's likability of the candidate, interest in interacting with them, or perceived level of similarity with them. According to the similarity-attraction paradigm, the greater perceived similarity between the interviewer and candidate, the more likeable the candidate is perceived to be, and the more likely they will be hired (Byrne, 1971). Additionally, by candidates conveying

high competence, they may also be perceived as enthusiastic. The perception of these vocal cues may result in an emotional contagion effect (e.g., Van Kleef, 2009), whereby interviewers may also experience a high activation of energy and positive affect (or excitement and enthusiasm) for the candidate. Additionally, interviewers may demonstrate the halo effect (e.g., Nisbett & Wilson, 1977), by unconsciously perceiving candidates who express warmth or friendliness as being competent. Conversely, when candidates express low competence from anxiety, including speaking slower and producing a greater frequency and duration of pauses, interviewers are more likely to perceive them as less interpersonally warm (e.g., enthusiasm, likeable, warm), negatively affecting their interview performance (Feiler & Powell, 2016). Though not many studies have examined the impact of speakers' vocal cues of competence on listeners' interpersonal attraction to them, these findings suggest that speech cues can affect an interviewers' warmth towards a candidate, subsequently affecting their hiring decision.

Overall, the detection and decoding of a job candidate's vocal competence level can greatly shape their perceived intellectual and social competence due to job interviewer's cognitive and social processing of their speech.

Decoding job candidates' vocal cues within a cultural context

While there has been extensive research on the impact of a job candidate's vocal cues on their perceived competence, less is understood regarding the impact of these vocal cues given the cultural context of the interview. Previous findings assume that job interviewers and candidates are of a similar cultural and language background, and potentially language proficiency level, allowing interviewers to accurately decode a candidate's vocal cues. Job interviewers (un)consciously rely on cues from various communication channels, including vocal cues, to make competence-based inferences. However, in an intercultural/interethnic hiring context, where job

interviewers differ from candidates in their cultural and/or language background, and candidates of various language, cultural, and ethnic backgrounds are compared, competence-based inferences are affected by additional social and cognitive factors (Gluszek & Dovidio, 2010; Goslin et al., 2012; Van Engen & Peelle, 2014). The mere detection of a candidate's language, cultural or ethnic cues can impact their perceived competence level. Moreover, the competence-based inferences drawn from a candidate's vocal cues are affected by interviewers' accuracy in decoding these cues. This diverse hiring context is becoming more common as companies recruit skilled immigrants (e.g., Elrick, 2016; Guererro & Rothstein, 2012; Sakamoto, Chin, & Young, 2010), and are more aware of the benefits of having a culturally diverse workforce for innovation (e.g., Wang, Cheng, Chen, & Leung, 2019). However, less is empirically understood as to how vocal cues marking a candidate's competence level are socially weighed amongst cues marking their language, cultural or ethnic background, to affect their perceived competence.

Specifically, it is unclear how well job interviewers can understand the vocal competence of job candidates who differ from them in their cultural and language background, as evidenced by their manner of pronunciation (or accent) and racial/ethnic appearance. Here, an accent refers to a speaker's production of segmental (e.g., consonants, vowels) and suprasegmental features (e.g., stress, intonation) (Cristia et al., 2012; Giles, 1970; Lippi-Green, 1997; Wells, 1982). Thus, everyone has an accent. A speaker's accent can be categorized in many ways including geographically, or by social class (Foon, 1986). In this study we examined the patterns of pronunciation that are shared by members of various language and/or cultural groups. Here, a standard accent refers to a manner of pronunciation largely produced by native speakers of that language who acquired the language in a given country, such as a Canadian English accent spoken by native speakers of Canadian English (Cargile, 2000; Deprez-Sims & Morris, 2010; Mai &

Hoffmann, 2014; Pantos, 2012). By using the word "standard" we are not implying that this accent is "proper", "correct", "superior" or "prestigious", for which it has historically been used by Caucasian native speakers of English to discriminate against the way that visible ethnic minorities and/or people from colonized countries spoke English (Kachru, 1986; Lippi-Green, 1994; 1997). Instead, "standard" is used to refer to an accent that is typical or average for speakers who speak a native language or dialect in a nation. In contrast, we define a non-standard accent as a manner of pronunciation which broadly deviates from the phonological and intonational speech norms of native speakers in a country or region. At a community/group level, listeners may perceive that a speaker with a non-standard accent does not share the same native language or dialect as them (Van Engen & Peelle, 2014). A non-standard accent may be perceived from three main contexts. One, phonological and prosodic features of a speaker's native language are transferred to their second language (L2) (e.g., Derwing, Fraser, Kang, & Thomson, 2014). For example, an Italian French accent in Canada, may refer to a native speaker of Italian who has a perceived non-standard accent when speaking French. Two, speakers are native speakers of a language, although speaking a different dialect, such as an Australian English accent in Canada. Lastly, a non-standard accent may be perceived from a combination of the former two contexts; a speaker is a native speaker of a language and also speaks other languages that may shape the phonological and intonational characteristics of their speech. For example, a speaker of Singaporean English is a native speaker of English (an official language in Singapore) who may also speak Malay or Mandarin. This final context is the focus of this study.

Here, racial/ethnic appearance refers to a speaker's ethnic heritage based on physiognomical features in the face (e.g., skin tone, the size and shape of facial features). This study specifically focuses on speakers who are visible racial/ethnic minorities. In Canada, visible

minority refers to “persons, other than Aboriginal peoples, who are non-Caucasian in race or non-white in colour” and can include people who identify as South Asian, Chinese, Black, Filipino, Latin American, Arab, Southeast Asian, West Asian, Korean, and Japanese (Statistics Canada, 2016). For speakers who are visible ethnic minorities, their racial/ethnic appearance can additionally activate stereotypes and/or (prejudiced) cultural and language attitudes, when presented in a context with other ethnic groups, especially a dominant ethnic group (e.g., Caucasian) (Derous et al., 2016).

This study investigates the effect of a job candidate’s (standard or non-standard) accent, racial/ethnic appearance, and vocal competence level on their perceived competence in a hiring context. I review the cognitive and social processes, and cultural factors that can additionally shape the perceived competence of job candidates with non-standard accents. I also evaluate results from the few studies that examined the interaction between a speaker’s vocal cues and accent, and the interaction between a speaker’s ethnic appearance and accent on listener's competence-based impressions.

Cognitive processing of speech by job candidates with non-standard accents

The cognitive processing of vocal cues in a non-standard accent can negatively impact a candidate’s perceived competence. Native listeners are found to have greater difficulty processing speech in a non-standard accent compared to a standard accent (Van Engen & Peelle, 2014). It may be assumed that native listeners encode speech in their native (or standard) accent, and when they perceive speech in a non-standard accent, the variation in the acoustic/phonetic signals and prosodic patterns makes it more difficult to retrieve the associated phonological and/or lexical representations and their semantic properties (Goslin et al., 2012). Moreover, listeners are hypothesized to encode speech in their standard accent, with greater strength (or attention)

compared to speech in a non-standard accent (Sumner et al., 2014). For example, when transcribing speech in a non-standard accent, intonation and phonemic errors are significant predictors of native listeners' mistakes (Munro & Derwing, 1999). Also, native listeners are slower at recognizing words presented in an unfamiliar intonation contour compared to a familiar one (Braun, Dainora, & Ernestus, 2011). Thus, when job candidates with non-standard accents are additionally conveying a mental or affective state via their vocal cues, the potentially unfamiliar prosody may make it more challenging for interviewers to decode their communicative intention.

Listeners' processing fluency for speech in a non-standard accent is also affected by subjective characteristics of a speaker's accent. Recent research using mediation analyses show that the reduced competence and/or hirability ratings for speakers with non-standard accents are mediated by how easily their speech is understood (or *comprehensibility*) and/or the strength of their accent (or *accentedness*) (Deprez-Sims & Morris, 2013; Dragojevic, Giles, Beck, & Tatum, 2017; Roessel, Schoel, Zimmermann, & Stahlberg, 2019). Firstly, listener's self-rated ease or difficulty to understand speech (Derwing & Munro, 1997, 2009; Munro & Derwing, 1995b) contrasts with *intelligibility* or how well a listener understands the speech content, as measured by their accuracy in transcribing speech (Derwing & Munro, 1997) or answering comprehension questions (Hansen & Dovidio, 2016). While speech with low intelligibility may be rated as having low comprehensibility (i.e., subjectively more difficult to understand), highly intelligible speech may also be rated as having low comprehensibility (Munro & Derwing, 1995b), suggesting that social and cultural attitudes can contribute to a speaker's perceived comprehensibility. Listeners in a workplace prefer to interact with speakers with non-standard accents who have higher comprehensibility ratings (Derwing & Munro, 2009). Also, comprehensibility ratings are found to mediate the relationship between listener's social attitude towards other language/cultural groups

and their hiring recommendation for a speaker with a non-standard accent (Hansen & Dovidio, 2016).

Secondly, the perceived strength of a speaker's accent can affect a speaker's perceived competence. It is often measured by listeners rating how different/similar the speech is to a native speaker (e.g., "*How different does the speaker sound from a native English speaker*?" or "*How strong is the speaker's accent*?") (Bergeron & Trofimovich, 2017; Munro & Derwing, 1995b; Schmid & Hopp, 2014), or the level of acoustic dissimilarity from speech in a standard accent (see Perceptual Distance Hypothesis, Clarke & Garrett, 2004; Floccia et al., 2006; Porretta et al., 2016). The stronger a speaker's accent, the more difficult it may be to understand it, which may lead to reduced competence. This judgment may be related to the associated negative affect listeners can experience while processing this speech. Though there is little research on the affective reactions towards speakers with non-standards accents, a few studies have found that listeners report more negative affect in general after listening to speech in a non-standard accent compared to a standard accent (Dragojevic et al., 2017; Hosoda, Stone-Romero, & Walter, 2007; Roessel et al., 2019). Thus, when speakers with non-standard accents are vocally expressing high or low competence, it is unclear how the potential increased processing effort and frustration of the listener may further impact speakers' perceived competence.

Effect of cultural communicative style on vocal competence

The difficulty in decoding the vocal cues of speakers with non-standard accents is further compounded by a speaker's cultural background which can shape how they modulate their voice to convey vocal affective/mental states. Listeners may face greater difficulty in decoding if speakers modulate their voice differently to convey competence. This effect is addressed by several theories on culture and emotions. Firstly, the Hofstede cultural dimension of masculinity-

femininity (or assertiveness-modesty) describes the extent to which cultures value aggression, assertiveness, and competitiveness compared to modesty (Hofstede, 1980; 1984; 2001). Collectivistic cultures are thought to be higher in femininity or modesty than individualistic, Western cultures and these values may be reflected in a job candidate's behaviour (Manroop, Boekhorst, Harrison, 2013; Wong & Phooi-Ching, 2000). Secondly, the Affect Valuation theory outlines that in Western individualistic cultures, such as Canada and the United States, people value high-arousal positive (HAP) states (e.g., excitement, enthusiasm) more than people in collectivistic cultures, such as parts of East Asia where low-arousal positive states (e.g., calm, peacefulness) hold greater value (Tsai, Knutson, & Fung, 2006; Tsai, 2007).

These cultural values can be observed in the way that speakers vocally express themselves and in the social perceptions formed. In a Western job interview context, it is more favourable for candidates to express high competence in an assertive, enthusiastic, and self-promoting manner, by speaking loudly, quickly, and fluently (Bencharit et al., 2019; Peng, Zebrowitz, & Lee, 1993; Schmid Mast, Frauendorfer, & Popovic, 2011) and with greater vocal expressiveness (e.g., greater pitch variation) (Semnani-Azad & Adair, 2013). In contrast, in East Asian cultures, high competence is displayed in a more modest manner involving a slower speech rate (Bencharit et al., 2019), "sounding apologetic" (Wong & Phooi-Ching, 2000, p. 20), more unfilled pauses (Wong & Phooi-Ching, 2000), and less vocal expressiveness (e.g., less pitch variation) (Semnani-Azad & Adair, 2013). In a Canadian context, a candidate producing these vocal cues is likely to be perceived as uncertain and anxious (Manroop, Boekhorst, & Harrison, 2013). Thus, when job candidates have a non-standard accent and are from a culture where vocal competence is expressed differently, Canadian interviewers may misinterpret their vocal cues as indicating low competence.

Lastly, the dialect theory of vocal emotion posits that our expression and perception of vocal emotions is shaped by 1) common (or universal) features, 2) culturally influenced display and decoding rules, and 3) culture-specific affect features that result in subtle and stylistic differences across cultures (Elfenbein & Ambady, 2003). Listeners are found to be better at accurately recognizing vocal emotions by members of their own cultural group (i.e., in-group) than by members of a different cultural group (i.e., out-group), an effect known as an in-group advantage (Elfenbein & Ambady, 2002; 2003; Laukka & Elfenbein, 2021). Less is understood regarding whether this in-group advantage also applies to the perception of more complex emotions/mental states (e.g., vocal competence) and the social inferences drawn. Taken together, the degree of cultural distance between a job interviewer and candidate can complexly impact a candidate's perceived competence level, in addition to their accent.

Social processing of job candidates with non-standard accents

The social processing of vocal cues in a non-standard accent can also negatively impact a candidate's perceived competence. There is substantive evidence, across numerous languages, cultures, and countries, that speakers with non-standard accents are perceived as less competent than speakers with standard accents (e.g., Fuertes et al., 2012; Gluszek & Dovidio, 2010). This reduced competence may result in inferences that candidates have lower capabilities/skills (Fiske et al., 2002; Lambert, 1967), are of lower social status (Fuertes et al., 2012), or are less suitability for a job where there is a lot of spoken communication with others (Deprez-Sims & Morris, 2010; Hosoda & Stone-Romero, 2010; Kalin & Rayko, 1978; Roessel et al., 2019; Timming, 2017). This perception is hypothesized to occur because listeners categorize speakers as belonging to a different cultural or language group than themselves (i.e., out-group). This categorization process, as outlined by the social identity theory, allows listeners to develop/reinforce their sense of self as

they determine whether they want to associate with a speaker or not (Billig & Tajfel, 1973; Tajfel & Turner, 1979).

Moreover, interviewers are hypothesized to evaluate their level of similarity to candidates, to assess level of fit at a cultural or organizational level (Bye et al., 2014; Goldberg, 2005; Wolgast et al., 2018). This is done by (un)consciously evaluating cues indicating candidate's perceived demographic similarity to an interviewer, such as their accent or race/ethnicity. Adding to the predictions of the similarity-attraction paradigm (Byrne, 1971), researchers have found that when an interviewer perceives greater similarity with a candidate (e.g., from similar accents), the candidate is perceived to be more likeable, and the more likely they will be hired (Deprez-Sims & Morris, 2010; Hofhuis et al., 2016). Overall, based on these social categorization processes one may posit that by interviewers identifying candidates as having a non-standard accent, they may have a negative impression of their competence.

Perceiving competence from a candidate's accent and vocal cues

The impact of a candidate's vocal competence level and accent on their perceived competence is less empirically understood because of methodological limitations and mixed results from previous studies.

Methodological limitations

Perceptual studies have predominantly elicited speech by having speakers read standard narrative passages (e.g., "Rainbow Passage" (Fairbanks, 1960); "Comma Gets a Cure", (Honorof, McCullough, & Somerville, 2000); "The North Wind and Sun" (International Phonetic Association, 1949)) or scripted sentences (e.g., Harvard sentences, (Egan, 1948; IEEE, 1969)) in a neutral way with a constant speech rate and volume (e.g., Dragojevic & Giles, 2014; Fuse et al., 2018; Hansen et al., 2017; 2018, Hosoda & Stone-Romero, 2010; Wong & Babel, 2017). This

elicitation procedure allows researchers to control the linguistic content between speakers and the level of variability in their manner of speaking. It also ensures the production of relatively fluent speech by speakers with standard and non-standard accents, which is useful when examining the non-standard accents of L2 speakers with varying spoken language proficiency levels (e.g., Munro & Derwing, 1994). At the same time, these speech materials tend to have a reading intonation with a neutral or flat affect.

Although the use of this speech elicitation procedure and speech materials can be beneficial, they do not allow researchers to most accurately understand the social perceptions that listeners form of speakers with non-standard accents, and the implications in real-world contexts. When speakers express a communicative intention with their voice, such as a mental or emotional state, they meaningfully produce various communicative cues including changes to their prosody. This variation in the speech signal can impact listeners' processing of the speech (e.g., Nygaard & Queen, 2008; Pell, Jaywant, Monetta, & Kotz, 2011) and the social inferences formed (Jiang & Pell, 2015; Jiang, Sanford, & Pell, 2018; Jiang, Gossack-Keenan, & Pell, 2020). Thus, previous results that speakers with non-standard accents are perceived as less competent may be due to differences in speakers' speech rate from reading (which may be affected by their language proficiency) or from listeners' merely categorizing speakers as having a non-standard accent. Listeners likely did not engage in the same inferential processing as when candidates are expressing varying mental and emotional states via their vocal cues, or when job interviewers assess the competence of job candidates.

Decoding a speaker's communicative intention

There is evidence that, to some extent, that listeners can accurately decode the vocal cues of speakers with standard and non-standard accents. Bradac and Wisegarver (1984) examined the

effect of lexical diversity on the perceived competence of speakers with standard American English and Mexican English accents. One speaker producing both accents, read scripted responses to job interview questions that were of high or low lexical diversity, as determined by the number of different types of words (e.g., nouns, verbs, adjectives, adverbs) relative to the total words (Bradac, Konsky, & Davies, 1976). In the low lexical diversity condition, the speaker also produced more self-repetitions and pausing, less details, and a slower speech rate, similar to vocal expressions of low competence. Listeners (monolingual native speakers of American English) heard the five-minute excerpt of the speaker's responses in one of the accents and one of the lexical diversity levels and rated the speaker's intellectual competence (e.g., competent, intelligent, knowledgeable). The speaker expressing a high lexical diversity response was rated as more intellectually competent than a low lexical diversity response, regardless of speaker accent (Bradac & Wisegarver, 1984), indicating that participants accurately decoded the speaker's vocal and linguistic cues associated with their competence level. This accurate decoding has also been found when participants are exposed to stimuli shorter in duration and involving only vocal cues.

Jiang, Gossack-Keenan, and Pell (2020) and Jiang, Sanford, and Pell (2018) examined the effect of speaker's vocal confidence level and accent on their perceived believability. Native Canadian English speakers heard the same linguistic statements (4-11 syllables in duration) expressing vocal confidence or doubt by speakers with Canadian English, Quebecois English, or Australian English accents. The confident expressions, compared to the doubtful expressions, were characterized by a faster speech rate, higher amplitude, decreased pitch, and greater variation in amplitude and pitch (Jiang & Pell, 2017). After hearing each statement, participants rated how much they believed the speaker and other social dimensions (e.g., competence). Regardless of speaker accent, utterances expressing vocal confidence were rated as more believable than

utterances expressing vocal doubt (Jiang et al., 2018; Jiang et al., 2020). Thus, even if the candidate has a non-standard accent, interviewers may accurately decode a candidate's vocal cues to make competence-based inferences.

However, the level of accuracy in decoding the vocal competence level of speakers with non-standard accents may be more variable and reduced compared to speakers with standard accents. For example, in Bradac and Wisegarver (1984) a trend is seen for a greater difference in perceived intellectual competence when speech in an American English accent expressed a high versus low lexical diversity response, compared to the same responses in a Mexican English accent. The magnitude of this differentiation may indicate the extent to which participants accurately decode a speaker's vocal expression of high versus low competence and draw differentiated inferences about their competence level. The smaller this differentiation the more likely that vocal cues are misinterpreted.

Even when participants are only exposed to vocal cues associated with a competence level, similar trends may be observed. In validation studies for perceived confidence, there is a trend of a greater mean difference for speakers with Canadian English accents expressing vocal confidence versus doubt, compared to speakers with non-standard English accents (Jiang, Sanford & Pell, 2018; Jiang, Keenan-Gossack, & Pell, 2020). These results may support hypotheses that native listeners engage in different inferential processing of speech in their standard accent compared to a non-standard accent (Bestelmeyer, Belin, & Ladd, 2015; Foucart et al., 2020; Puhacheuskaya & Järvikivi, 2022). This may be due to greater familiarity or experience decoding mental/affective states in their standard accent which can facilitate comprehension (Van Engen & Peelle, 2014) or because speech in a non-standard accent is less socially and behaviourally relevant for people with standard accents (Bestelmeyer et al., 2015; Lev-Ari & Keysar, 2012). Nonetheless, it is unclear

how accurately interviewers can decode a candidate's vocal competence level when speakers have a non-standard accent and are from a culture where their voice is modulated differently to communicate their competence level.

Social weight of speaker's accent and vocal competence level

The extent to which a speaker's vocal cues (of high or low competence) modulate their perceived competence level, varies based on their accent. Bradac and Wisegarver (1984) found that when the speaker expressed a high lexical diversity response, speech in the American English accent was rated as more competent than speech in a Mexican English accent. Similarly, Jiang and colleagues (2020) found that when speakers expressed vocal confidence, speakers with Canadian English and Australian English accents were perceived as more believable than speakers with Quebecois English accents. These results may suggest that having a standard accent and being a native English speaker can result in greater perceived competence than speakers with non-standard accents who also speak a second language. At the same time, comparable believability ratings have also been found when speakers with standard and non-standard accents express vocal confidence (Jiang et al., 2018). Altogether, when candidates with standard and non-standard accents are expressing high vocal competence, it is unclear the extent to which their accent will negatively impact the impression formed. Also, more research is needed to better understand the variability between participants. The degree to which interviewers activate attitudes and stereotypes based on a candidate's accent to impact their competence impression, may vary between interviewers.

When speakers express low vocal competence, two outcomes have been found. One, there is no difference in the perceived competence of speakers with standard or non-standard accents. In Bradac and Wisegarver (1984) when speakers expressed the low lexical diversity responses, speakers with American English and Mexican English accents received comparable competence

ratings. The researchers hypothesize that when speakers produce speech indicating a high anxiety level (or low competence), inferences about a speaker's intellectual ability carries more perceptual weight than social inferences derived from their accent (Bradac & Wisegarver, 1984). This result may occur because speech cues conveying low competence are stronger diagnostic cues for evaluating a speaker's competence level than a speaker's (non-standard) accent (Skowronski & Carlston, 1986), especially when there is a strong speech signal via linguistic and vocal cues.

On the other hand, speakers with standard accents may also be perceived as less competent than speakers with non-standard accents. For example, when speakers expressed vocal doubt, speakers with Canadian English accents were rated as significantly less believable and competent than speakers with non-standard accents (i.e., Australian English and Quebecois English accents) (Jiang, Sanford & Pell, 2018; Jiang, Keenan-Gossack, & Pell, 2020). The researchers suggest that this occurred because vocal doubt is more salient in an in-group accent than an out-group accent (Jiang et al., 2018). This result may imply that native listeners are more familiar with decoding vocal cues in a standard accent and thus engage in different inferential processing of speech in a standard accent compared to a non-standard accent. Unlike the results of Bradac and Wisegarver (1984) when speakers expressed low competence, here the speech signal only involved vocal cues and thus may not have conveyed low competence as strongly. Thus, the intensity of low competence perceived may impact the social weight of a candidate's accent and the competence impression formed.

Overall, a candidate's perceived competence level varies based on their accent and vocal competence cues. The perceived competence of candidates with non-standard accents is affected by the potentially reduced or different inferential processing of their speech, the activation of cultural/language attitudes and stereotypes, and their comparison to candidates with standard

accents. However, more research is needed to understand how a candidate's vocal cues of competence may impact their perceived competence level if they have a non-Western, non-standard accent, and how the inferential processing of speech in a non-standard accent affects the job interview (or hiring) outcome.

Effect of job candidate's visual ethnicity cues on perceived competence

Another social factor that can affect the perceived competence of speakers with non-standard accents is their race/ethnicity. Speakers with standard accents are often members of a dominant racial/ethnic group (e.g., Caucasians) while speakers with non-standard accents are often visible racial/ethnic minorities (e.g., Chinese). This difference in race/ethnicity can affect a speaker's perceived competence as participants may activate concepts (e.g., ethnic appearances) or attitudes towards these racial/ethnic groups by hearing their voice. This activation may be explained by the dynamic interactive theory of person construal (Freeman & Ambady, 2011) which posits that our perception of others is dynamically affected by cues from various communication channels and the activation of associated conceptual representations. For example, when presented speech input people may activate category level concepts related to a speaker's perceived sex, race/ethnicity, and/or vocal emotion, as potentially affected by stereotypes (Freeman & Ambady, 2011). Specifically, there is evidence that when participants are only presented speech, speakers with non-standard accents who are visible racial/ethnic minorities (e.g., Chinese English, Japanese English, Indian English, Mexican English accents) are perceived as less competent than speakers of a dominant racial/ethnic group who have standard or non-standard accents (Acheme & Cionea, 2022; Fuertes et al., 2012; Timming, 2017). This result may suggest that a speaker's perceived race/ethnicity as a visible minority also strongly affects the impression formed from their accent. Speakers with non-standard accents who are visible ethnic minorities

may be rated similarly to speakers with non-standard accents who are part of a dominant racial/ethnic group if listeners are unfamiliar with the speaker's non-standard accent (Deprez-Sims & Morris, 2010), so associated racial/ethnic conceptual representations are not strongly activated.

Though there is not a specific theory addressing this relationship between a speaker's accent and race on speech processing, this effect may occur because the non-standard accents from speakers who are visible ethnic minorities are associated with lower prestige. According to the accent prestige theory, speakers with non-standard accents may be perceived negatively because their accent is associated with a linguistic/cultural group of people who are of a lower social class and prestige (relative to the dominant racial/ethnic group) (Giles, 1970), and thus are of lower competence. Additionally, for speakers who are visible ethnic minorities, their accent and potentially their racial/ethnic appearance, may indicate that they are an immigrant or broadly, an out-group member (Paladino & Mazzurega, 2020; Thai, Szeszeran, Hornsey, & Barlow, 2020). Thus, listener's impressions of speakers with non-standard accents who are visible racial/ethnic minorities may also reflect their impression of people from that racial/ethnic group.

Depending on a job interviewer's experience with people who have non-standard accents, and people of various races/ethnicities, the impact of a candidate's racial/ethnic appearance in combination with their accent and vocal competence level, may vary. However, there is little empirical research examining the individual and interactive impact of a speaker's racial/ethnic appearance and accent on their perceived competence in a hiring context. A few studies have examined the effect of job candidate's ethnic appearance and accent on their perceived competence. However, other competence-relevant information about candidates was also presented via a resume with an ethnically marked name (Huang et al., 2013; Singer & Eder, 1989) or no details regarding candidate's manner of speaking were provided (Hansen et al., 2017; 2018;

Huang et al., 2013; Singer & Eder, 1989). Overall, it is unclear if the competence impressions of speakers with non-standard accents who are visible ethnic minorities will be further attenuated if perceivers are also processing their racial/ethnic appearance.

Impact of candidate's racial/ethnic appearance

During face-to-face interviews, job interviewers may activate conceptual representations about a job candidate's communicative competence based on their racial/ethnic appearance. For example, Gnevsheva (2018) assigned participants (Caucasian native speakers of New Zealand English) to one of three conditions: audio only, visual only, or audiovisual and were presented speech from native New Zealand English speakers and non-native speakers of English (L1 Korean or L1 German). Speakers produced spontaneous speech without any communicative intention or mental/emotional state. After seeing and/or hearing each speaker participants rated the strength of the speakers' accents. Participants in the visual only condition rated the Korean speakers as having stronger foreign accents than the Caucasian (German and New Zealand) speakers even though they were not exposed to any speech (Gnevsheva, 2018). This result may suggest that participants expected the Korean speakers to have a non-standard English accent and/or viewed them as out-group members based on their race/ethnicity. Moreover, the Korean speakers received comparable accent strength ratings regardless of condition (Gnevsheva, 2018). In contrast, the German speakers were rated as having a less foreign accent in the visual only condition compared to the audio only and audiovisual conditions (Gnevsheva, 2018). In other words, participants may not have expected a Caucasian speaker to have a non-standard accent and/or viewed them as in-group members. Though competence judgments were not made, these results demonstrate that for visible racial/ethnic minorities who may have a non-standard accent, their racial/ethnic appearance can contribute to others' perceived communication with them.

Interactive effects of candidate's racial/ethnic appearance and accent

When participants integrate information from a speaker's racial/ethnic appearance and accent, speakers with non-standard accents who are visible racial/ethnic minorities may still experience a cost in their perceived competence. Hansen and colleagues (2018) presented German participants in one condition with speech by speakers with standard German accents and Turkish German accents who looked German or Turkish. Here, Germans (Caucasians) were considered the dominant racial/ethnic group and the Turkish were visible racial/ethnic minorities. Participants were asked to imagine they were helping with the hiring process for a middle-level manager and completed competence ratings after each image. In Block 1 Visual Only, participants saw images of males who had typical German or Turkish appearances. The German and Turkish candidates received comparable competence ratings based on their appearance only (Hansen et al., 2018). This result may suggest that in an interethnic hiring context, job candidate's racial/ethnic appearance alone may not be a strong diagnostic cue for assessing their competence level. However, this effect may vary depending on the racial/ethnic groups compared.

Then during Block 2 Audio and Visual when participants saw the same faces paired with speech uttered in neutral manner, speakers with Turkish German accents and Turkish appearances were perceived as comparably competent to when participants saw only their face (Hansen et al., 2018). In contrast, speakers with standard German accents and German appearances were perceived as more competent compared to when participants only saw their face (Hansen et al., 2018). Since the authors describe people of Turkish heritage as being stereotypically low in competence by Germans (Asbrock, 2010), this result may reflect that negative stereotype (Hansen et al., 2018). Turkish speakers' ethnic appearance may have strongly shaped participants' impressions as hearing the Turkish German accent did not significantly change their judgment.

Conversely, for speakers with standard German accents and German appearances, their standard accent and exposure to their voice increased their perceived competence. Thus, their vocal cues may carry meaningful social information for drawing competence-based inferences and for perceivers to identify them as in-group members. At the same time, speakers did not convey a vocal competence level, which can positively impact their perceived competence. Thus, for job candidates with non-standard accents who are visible racial/ethnic minorities, their racial/ethnic appearance may primarily shape their perceived competence, though results may differ for speakers from other accent and ethnic groups. More research is needed regarding how a job candidate's vocal competence level, accent, and racial/ethnic appearance can contribute to their perceived competence.

Chapter 2

Examining the effect of job candidates' vocal competence level and accent on perceived job suitability

2.1 Introduction

This perceptual study examined how the perceived job suitability of speakers with standard (American English) and non-standard (Singaporean English) accents varies with their vocal affective presentation style (vocal pride vs shame) to determine how the vocal competence level of speakers with standard and non-standard accents contributes to their perceived competence. It can be argued that these vocal affective states are associated with high versus low competence, respectively, by a job candidate. We additionally explored the effect of speaker's perceived vocal friendliness and vocal attractiveness as these may relate to job interviewers' affective reactions of job candidates or more broadly their interpersonal attraction (Feiler & Powell, 2016; Gallois, Callan, & Palmer, 1992; Howard & Ferris, 1996). We also explored the effect of listeners' social, cultural, and language characteristics on a speaker's perceived competence and hirability. Then, to determine whether speaker's cultural background affects the modulation of their voice to convey pride and shame, we conducted an acoustic analysis.

A job candidate's perceived competence can be greatly shaped by the vocal cues they produce while answering interview questions. Due to the high-stakes context of the interview, job candidate's vocal expression of competence may indicate their varying mental and emotional states. The mental states may reflect their level of certainty in recalling information from memory to answer a question (e.g., feeling of knowing) (Smith & Clark, 1993), while the emotional states may reflect their emotional (stress) reaction to the question being asked and relatedly, their perceived ability to answer it (e.g., Saslow et al., 2014). Thus, when expressing high vocal competence, candidates' vocal cues may indicate their high level of certainty in the content of their

speech, feeling enthusiastic, or feeling confident in their ability to answer a question well. In contrast, when expressing low vocal competence, candidates may have been asked questions they did not prepare for, feeling uncertain about their ability to answer a question. Previous studies examining a speaker's accent and vocal cues related to a speaker's competence level have focused on speaker's expression of competence as it relates to their mental state (e.g., vocal confidence vs doubt) (Jiang & Pell, 2017; Jiang et al., 2018; 2020) or level of anxiety (Bradac & Wisegarver, 1984), which may be considered a mental and emotional state. Less is understood regarding how a speaker's affective state contributes to their perceived competence. In this study, we explored the effect of vocal affective states, and aimed to replicate previous effects that involved a vocal mental state (Jiang et al., 2018; 2020).

Here, a speaker's vocal affective presentation style involved two affective states related to the expression of competence. Pride and shame are self-conscious (or social or moral) emotions (Adolphs, 2002; Lewis, 1993; Miller & Leary, 1992; Tracy & Robins, 2004a), meaning that they are experienced when one is consciously aware of their behaviour and evaluates it relative to social norms and standards. These social norms may be shaped by a speaker's cultural background, which may affect how speakers modulate their voice to convey pride or shame (Laukka, Neiberg, & Elfenbein, 2014). However, these emotions are less understood vocally, with most research on facial and posture expressions. Pride is an emotion of positive valence experienced in response to perceived success (Tracy & Matsumoto, 2008). The vocal expression of pride may be confused with the vocal expression of happiness or anger (Elfenbein et al., 2022). Conversely, shame is an emotion of negative valence experienced in response to perceived failure and may relate to feelings of embarrassment (Keltner, 1995; Tracy & Matsumoto, 2008). The vocal expression of shame may be confused with the vocal expressions of sadness, distress, or fear (Elfenbein et al., 2022).

We predicted an effect of vocal affective presentation style, with speakers expressing pride receiving higher competence ratings than when they are expressing shame, regardless of accent. This result would indicate that listeners are accurately perceiving a difference in speakers' vocal affective state, impacting the perceived competence rating. Like the previous trends reported (Bradac & Wisegarver, 1984; Jiang et al., 2018; 2020), we predicted that the mean difference in competence and hirability ratings when speakers convey vocal pride versus shame, will be greater for speakers with standard English accents compared to speakers with non-standard English accents. From the acoustic analysis, we predicted that there may be acoustic differences in how speakers express vocal pride, if cultural differences for expressing high competence affects their speech. We did not have specific predictions about acoustic differences for vocal shame.

Of interest is the potential interaction between speaker vocal affective presentation style and accent. When speakers express vocal pride, speakers with Singaporean English accents may receive lower competence and hirability ratings than speakers with American English accents if they modulate their voice differently to express high competence. Additionally, if a speaker's non-standard accent is perceived to be a more salient cue than their vocal affective state, speakers with Singaporean English accents may receive lower competence ratings. This may occur because cultural/language attitudes and stereotypes towards speakers with perceived low language proficiency are activated. Alternatively, there may be no difference in the perceived competence ratings between speakers with Singaporean English or American English accents if the vocal expression of pride carries more perceptual weight on a speaker's competence than their accent. When speakers express vocal shame, we predict a smaller effect of speaker accent, compared to expressions of vocal pride. Specifically, we predict that there may be no difference in the perceived competence ratings of speakers with Singaporean English or American English accents if the vocal

expression of shame is perceived as a stronger more diagnostic cue of a speaker's competence. Alternatively, speakers with Singaporean English accents may receive higher competence ratings than speakers with American English accents if listeners engage in reduced or different inferential processing of speech in a non-standard accent.

2.2 Experiment 1

2.2.1 Methods

Participants

Participants were 60 native Canadian English speakers (30 males, 30 females, Mean Age = 30.62 years, SD = 4.69 years, Range = 24-40 years) recruited through the online platform Prolific Academic (www.prolific.co) [accessed December 2021]. Sample size was determined a priori from a power analysis in G*Power (Faul, Erdfelder, Lang & Buchner, 2007), based on a repeated measures ANOVA for the within-subject factors of accent and vocal affective presentation style. This sample size allows us to detect medium effects ($\eta_p^2 = 0.06$ or $d = 0.50$) with $\alpha = 0.05$, an assumed correlation of 0.50 between repeated measures, and a statistical power of 0.80, $F(13,312) = 1.75$.

All participants were born in and currently resided in Canada and identified their nationality as Canadian. [1]Approximately half of the participants were monolingual English speakers (n = 28). The remaining participants were bilingual (n = 29) or trilingual (n = 3). Some of these additional languages included French, Spanish, Cantonese, Polish, Tagalog, Korean, and Arabic. Most participants began learning only English at birth (n = 57). The remaining participants were sequential bilinguals (n = 3), who began learning English between the ages of 1-5 years old.

[1] Participants reported their current province of residence as follows: Ontario (50%), British Columbia (15%), Manitoba (11.7%), Alberta (8.3%), Saskatchewan (5%), Quebec (3.3%), Nova Scotia (3.3%), New Brunswick (1.7%), and not indicated (1.7%).

Participants attended school in English (three participants did not specify the language of instruction) and were interested in working in English.

Most participants identified their race/ethnicity as White/European (n = 42). The remaining participants identified as Asian (n = 9, East Asian: n = 7, Southeast Asian = 2), African (n = 2), Middle Eastern/Arab (n = 1), or two races/ethnicities (one of which included White/European) (n = 6). Also, most participants were third or more generation Canadian (n = 39) (i.e., participant and their parents were born in Canada) (Statistics Canada, 2018, July 25). The remaining participants were second generation Canadian (n = 20) (i.e., participants were born in Canada and at least one parent was born outside Canada), or unspecified (n = 1). Participants had a wide range of occupations including student, accountant, software developer, teacher, retail associate, and musician. 58% of participants had experience helping with the interviewing and hiring process (Range = 5 days to 10 years).

Participants reported normal or corrected-to-normal vision and hearing, and no speech, language or hearing disorders, or neurological and psychiatric disorders. Participants provided informed consent prior to the experiment and were compensated 11.25 CAD for their participation. This study was approved by the McGill Faculty of Medicine Institutional Review Board.

Speech stimuli

Vocal affective presentation style

Utterances conveying pride and shame by speakers with American English and Singaporean English accents were obtained from the Vocal Expressions of Nineteen Emotions across Cultures (VENEC) corpus (Laukka et al., 2016; Elfenbein et al., 2022). The affect for the selected utterances was produced with moderate intensity by native speakers of English from the United States (three males, two females) and Singapore (three males, two females) saying

one phrase with emotionally neutral content ("*That is exactly what happened*"). Speech was produced by professional actors (aged 18-30 years old). Speech was elicited by providing speakers with scenarios involving the experience of pride and shame and instructing them to imagine themselves in similar situations. The expression of each emotion was not modeled to speakers. See Laukka et al. (2016) and Cowen et al., (2019) supplementary material for details on the speech elicitation procedure. The American English accent was produced by Caucasian individuals in California, USA, and the Singaporean English accent was produced by Chinese individuals in Singapore, Singapore. Speakers were born and raised in their respective countries and started speaking English at birth or early childhood (Laukka et al., 2016).

Twenty unique utterances were used in the perceptual experiment (1 utterance x 10 speakers with standard and non-standard English accents x 2 vocal affective presentation styles (pride, shame)). Recordings were normalized to a peak intensity of 70 dB in Praat (Boersma & Weenink, 2020) and saved as .wav files. The selected utterances were short, ranging in duration from 1.33 to 3.40 seconds. There was no significant difference in the duration of the pride and shame utterances, $F(1, 16) = 0.32$, $p = 0.58$, 95% CI [-0.36, 0.62], or the duration of utterances produced by speakers with Singaporean English versus American English accents, $F(1, 16) = 0.12$, $p = 0.73$, 95% CI [-0.57, 0.41]. There was also no interaction between vocal affective presentation style and speaker accent for stimulus duration.

Narrative passage

To characterize speaker's accent strength and comprehensibility, participants heard speakers producing The North Wind and Sun (International Phonetic Association, 1949), also obtained from the VENEC corpus (Laukka et al., 2016; Elfenbein et al., 2022). The North Wind and Sun is a standardized passage used to assess speaker accent strength or the pronunciation

ability of English L2 speakers (e.g., Gass & Varonis, 1984; Tsurutani, 2012; Van Engen et al., 2010; Wong & Babel, 2017). The original recordings were cut into separate recordings based on the sentences in the passage, to ease participant fatigue in the perceptual study. The first sentence from the passage was used ("*The North Wind and the Sun were disputing which was the stronger when a traveller came along wrapped in a warm cloak."*) (Appendix B).

Ten narrative utterances were used in the perceptual experiment (1 utterance x 10 speakers with standard and non-standard accents). Recordings were cut and normalized to a peak intensity of 70 dB in Praat (Boersma & Weenink, 2020). There was a marginally significant difference in the duration of the narrative utterances based on speaker accent. Utterances by speakers with Singaporean English accents (M = 7.59 seconds, SD = 0.93) had a marginally longer duration than those by speakers with American English accents (M = 6.29 seconds, SD = 0.85), t (7.94) = 2.30, p = 0.05, 95% CI [-0.003, 2.60], Range = 5-8 seconds.

Acoustic features of speech stimuli

This analysis acoustically characterized the utterances expressing a vocal affective presentation style and the narrative utterances by speakers with Singaporean English and American English accents. This analysis was conducted using the Geneva Minimalistic Acoustic Parameter Set (GeMAPS) package (Eyben et al., 2016) and the publicly available openSMILE toolkit (Eyben et al., 2010) which provides a standardized baseline set of affect-related acoustic measures. These analysis tools were also previously used to acoustically characterize a subset of the vocal pride and shame utterances used (Laukka et al., 2016; Nordström et al., 2017). The acoustic parameters of interest involve frequency, amplitude, and temporal measures that are perceptually relevant to vocally expressing confidence versus doubt (Jiang et al., 2017; 2018; 2020), charisma (Yang et al., 2020), anxiousness (Feiler & Powell. 2016), and assertiveness versus modesty (Wong & Phooi-

Ching, 2000). Spectral measures related to the expression of vocal emotions were also explored (Laukka et al., 2016; Waaramaa et al., 2010). These measures were computed on the entire utterance for the expressions of pride and shame, and the narrative utterances. The acoustic parameters of interest included:

Frequency measures

1. *Mean logarithmic F0* (fundamental frequency): indexing voice pitch on a semitone frequency scale starting at 27.5 Hz
2. *Standard deviation of logarithmic F0*: variability of F0
3. *Jitter*: frequency perturbations or variability in consecutive F0 period lengths

Amplitude measures

4. *Mean Loudness*: mean signal intensity estimated from an auditory spectrum
5. *Shimmer*: difference in peak amplitudes of consecutive F0 periods
6. *Rate of loudness peaks*: the number of loudness peaks per second

Spectral measures

7. *Alpha ratio*: ratio of the summed energy from 50–1000 Hz and 1000–5000 Hz of voiced regions
8. *Hammarberg Index*: ratio of the strongest energy peaks in the 0-2000 Hz and 2000-5000Hz of voiced regions

Temporal measures

9. *Syllable rate*: number of continuous voiced regions per second (pseudo syllable rate)
10. *Pause duration*: mean segment length of unvoiced regions (approximating silent pauses)
11. *Standard deviation of pause duration*

Vocal affective presentation style utterances

The analysis revealed no significant differences between the pride and shame utterances by speakers with American English or Singaporean English accents for mean logarithmic f0, standard deviation of logarithmic f0, mean amplitude, mean jitter, mean shimmer, mean rate of loudness peaks, and mean pseudo syllable rate (see Table 2.1). Two differences for temporal measures were found. There was a significant effect of speaker accent on pause duration, $F(1, 16) = 16.98$, $p < .001$, $\eta_p^2 = 0.51$, 95% CI for η_p^2 [0.15, 0.73]; speakers with American English accents produced longer pauses than speakers with Singaporean English accents. This effect was

moderated by speaker's vocal affective presentation style, F (1, 16) = 10.86, p < .001, η_p^2 = 0.40, 95% CI for η_p^2 [0.06, 0.66]. Based on Tukey pairwise comparisons, when speakers expressed pride there was no significant difference in pause duration, regardless of speaker accent, p = 0.94, 95% CI [-0.04, 0.06]. When speakers expressed shame, speakers with American English accents produced longer pauses than speakers with Singaporean English accents, p < .001, 95% [0.04, 0.13]. Also, there was no significant difference in the pause duration for speakers with Singaporean English accents regardless of the vocal affective presentation style expressed (mean difference = 0.02, p = 0.45). Speakers with American English accents had significantly greater pause durations when they expressed shame compared to pride (mean difference = 0.05, p = 0.03).

There was also a significant effect of speaker accent on the variation in pause duration, F (1, 16) = 7.24, p = 0.02, η_p^2 = 0.31, 95% CI for η_p^2 [0.01, 0.59], with speakers with American English accents having greater variation in pause duration than speakers with Singaporean English accents. This effect was moderated by speaker's vocal affective presentation style, F (1, 16) = 7.62, p = 0.01, η_p^2 = 0.32, 95% CI for η_p^2 [0.02, 0.60]. Based on Tukey pairwise comparisons, when speakers expressed pride there was no significant difference in variation for pause duration, regardless of speaker accent, mean difference = -0.001, p = 1.00, 95% CI [-0.07, 0.04]. When speakers expressed shame, speakers with American English accents produced pauses with a greater variation in duration than speakers with Singaporean English accents, difference = 0.08, p = 0.01, 95% [0.02, 0.14]. Also, there were no significant differences in the variation in pause duration of speakers with Singaporean English accents regardless of the vocal affective presentation style expressed (mean difference = 0.02, p = 0.88). For speakers with American English accents, there was greater variation in their pause duration when they expressed shame compared to pride (mean difference = 0.07, p = 0.03).

Table 2.1. *Major acoustic features of the pride and shame utterances by speaker accent group*

Acoustic parameter	Speaker accent			
	American English		Singaporean English	
	M	*SD*	*M*	*SD*
	Pride			
Mean logarithmic F0	29.47	9.71	32.39	6.16
Standard deviation of logarithmic F0	0.12	0.04	0.13	0.04
Mean Jitter	0.05	0.02	0.05	0.01
Mean Loudness	1.21	0.11	1.13	0.09
Mean Shimmer	1.62	0.46	1.54	0.16
Rate of loudness peaks	4.69	1.12	3.43	0.34
Alpha ratio	-9.93	1.98	-9.63	3.69
Hammarberg index	19.16	2.70	18.78	2.64
Syllable rate	2.70	1.10	2.43	0.37
Pause duration	0.07	0.02	0.07	0.01
Standard deviation of pause duration	0.04	0.01	0.04	0.01
	Shame			
Mean logarithmic F0	22.74	4.73	30.25	3.80
Standard deviation of logarithmic F0	0.19	0.07	0.13	0.06
Mean Jitter	0.07	0.02	0.05	0.01
Mean Loudness	1.16	0.16	1.04	0.17
Mean Shimmer	1.98	0.49	1.57	0.56
Rate of loudness peaks	4.35	0.93	4.35	1.02
Alpha ratio	-10.84	2.37	-14.48	3.57
Hammarberg index	19.90	3.58	23.58	4.41
Syllable rate	2.99	1.49	2.37	1.57
Pause duration	0.13	0.03	0.04	0.03

Standard deviation of pause duration	0.11	0.06	0.03	0.03

Narrative utterances

For the narrative utterances, the analyses revealed no significant differences between speakers with American English and Singaporean English accents, except for the rate of loudness peaks (Table 2.2). Speakers with American English accents produced a greater rate of loudness peaks than speakers with Singaporean English accents, t (4) = -4.14, p = 0.01, 95% CI [-1.94, -0.38].

In summary, the acoustic analysis did not reveal many significant differences between the pride and shame utterances by speakers with American English or Singaporean English accents. A couple of temporal differences were found, particularly when speakers expressed vocal shame, where speakers with American English accents produced longer pauses, and had greater variation in their pause duration compared to speakers with Singaporean English accents. For the narrative utterances, an amplitude difference was found, with speakers with American English accents producing greater loudness peaks.

Table 2.2. *Major acoustic features of the narrative utterance by speaker accent group*

Acoustic parameter	Speaker accent			
	American English		Singaporean English	
	M	*SD*	*M*	*SD*
Mean logarithmic F0	26.42	6.34	28.67	4.65
Standard deviation of logarithmic F0	0.28	0.13	0.17	0.04
Mean Jitter	0.05	0.01	0.05	0.01
Mean Loudness	1.08	0.11	0.99	0.08
Mean Shimmer	1.54	0.35	1.35	0.23

Rate of loudness peaks	5.01	0.74	3.85	0.33
Alpha ratio	-12.16	2.04	-15.38	2.32
Hammarberg index	23.08	2.31	26.06	3.02
Syllable rate	2.51	0.41	2.61	0.69
Pause duration	0.11	0.04	0.11	0.03
Standard deviation of pause duration	0.09	0.06	0.08	0.03

Design and Procedure

The experiment used a 2 x 2 within-subjects design. Participants heard speech produced by speakers with standard and non-standard English accents to investigate the effect of speaker accent (American English, Singaporean English) and vocal affective presentation style (pride, shame) on their perceived competence and hirability (main dependent variables). Participants completed perceptual ratings in two blocks. Participants heard each speaker once in each block, producing one of the vocal affective presentation styles.

The experiment was presented online using jsPsych (de Leeuw, 2015) and Just Another Tool for Online Studies (JATOS) (Lange, Kühn, Filevich, & 2015) via the BRAMS Online Testing Platform (OTP)(https://brams.org/category/online-testing-platform/), where the instructions and task were written in English. Before the experiment, participants were asked to complete the experiment in a quiet environment, preferably using headphones. To ensure that participants could successfully hear audio clips, they first completed an audio test consisting of two trials. In each trial, participants played an utterance of neutral content in a neutral tone of voice and were presented a written transcript of the utterance with one word missing. A choice of four words were presented below this fill-in-the-blank sentence. Participants were asked to click the word that correctly completed the blank. Participants needed 100% accuracy on these trials to proceed to the

experiment. Participants then completed questions regarding their language background, demographic information, employment, and hiring experience.

Participants were asked to imagine that they were working for a company, where they will help to decide who receives a job promotion. They will decide which employee should be promoted to a human resources manager. Participants were then asked to carefully read a job description for the human resources manager position (see Appendix A). After reading the job description, participants completed three multiple-choice questions about the job description to assess their attention level and understanding of the position. Questions included, "Which of the following tasks does a human resource manager not complete? What are two categories of tasks that a human resource manager completes? Which of the following positions does a human resource manager communicate with?" Participants were provided feedback on the accuracy of their responses after answering all three questions. Participants needed 100% accuracy to complete the rest of the experiment. If they did not have 100% accuracy, they were re-directed to the job description and answered the same questions again in the same order.

Then participants evaluated the perceived confidence, competence, and hirability of job candidates in two blocks. Participants were presented 10 job candidates and were told the candidates all have a business degree with a major in human resources and have worked for at least three years as a human resources assistant. In each block, participants heard each speaker express vocal pride or shame. In the first block, one utterance from each speaker was randomly selected. In the second block, the other vocal affective state (pride or shame) for each speaker was presented. Participants were told that they would hear each speaker's voice as they responded to questions about their skills (e.g., What is your greatest accomplishment? Describe one of your strengths.). They were told to focus on the way the speaker talks instead of the words they say

since every speaker will be saying the same phrase. Each trial began with a 500-millisecond fixation cross and then the audio file with the rating scales displayed directly below. Participants completed three 9-point rating scales, 1) the speakers' perceived confidence via, "How confident does the candidate sound?" on a scale from "not at all confident" to "very confident" (Jiang & Pell, 2017; Jiang, Keenan-Gossack, & Pell, 2020), 2) the speaker's competence for the position via, "How competent would this candidate be as a human resources manager?", on a scale from "not at all competent" to "very competent", and 3) their hiring recommendation via, "How likely would you recommend hiring this candidate as a human resources manager?", on a scale from "very likely" to "not at all likely". Each utterance was presented once.

Then participants completed questionnaires regarding their cultural experiences/background, social traits, and explicit attitudes. We measured participants' explicit attitudes towards people with non-standard accents via the Accent Belief Scale (Hansen, 2020). It involves two dimensions, 1) accent diagnosticity, which measures the extent to which a person believes they can infer other traits about a speaker based on their accent, and 2) accent stability, which measures the extent to which a person believes that a speaker's accent is stable and unchangeable (Hansen, 2020). Then for social traits, participants completed the Tolerance of Ambiguity scale (Herman et al., 2010), which measures participants' level of tolerance for ambiguity in cross-cultural settings. Higher scores indicate a higher tolerance for ambiguity and uncertainty. The Social Desirability scale (He et al., 2014), was also completed which measures participants' likelihood of trying to minimize their general explicit negative bias.

Lastly, to control for how participants perceived the job candidates' voices, they rated the job candidates' vocal attractiveness using two items assessing different aspects of social attractiveness (Lambert, 1967; Reysen, 2005; Ryan, Hewstone, & Giles, 1984). Each trial began

with a 500-millisecond fixation cross, the same audio stimuli from the experimental trials (vocal shame and pride utterances) and two 9-point rating scales displayed directly below, 1) "How attractive is this person's voice?", on a scale from "not at all attractive" to "very attractive", and 2) "How friendly does this person sound?", on a scale from "not at all friendly" to "very friendly". The stimuli were presented in a randomized order.

In the last block, listeners heard the narrative utterances to characterise the strength of the speaker's accent and the utterance's comprehensibility. Each trial began with a 500-millisecond fixation cross, the audio stimuli and two ratings scaled displayed directly below, 1) "How easy is it for you to understand the speech?", on 9-point scale from "extremely easy to understand" to "extremely difficult/impossible to understand" (Derwing & Munro, 1997; Derwing & Munro, 2009), and 2) "How different does this speaker sound from a native Canadian English speaker?" measured on a 7-point scale from "Heavily accented" to "No accent at all" (Bergeron & Trofimovich, 2017). The stimuli were presented in a randomized order. The experiment took 30-45 minutes to complete.

Statistical analysis

Analyses were performed and figures were created in RStudio (version 1.4.1717) (R Studio Team, 2021) (R Version 4.1.0, http://cran.r-project.org) (R Core Team, 2021). The effect size (partial eta squared, η_p^2) for the ANOVAs was estimated using the eta_squared function in the effectsize package (Ben-Shachar, Lüdecke, & Makowski, 2020). The linear mixed effects models were estimated using the lme function in the lmerTest package (Kuznetsova, Brockhoff, & Christensen, 2017), where the two-sided t-tests, degrees of freedom and p-values were computed using Satterthwaite approximation. Pairwise comparisons, if computed, were estimated using the emmeans package (Lenth, 2022), where the degrees of freedom were computed using the

Kenward-Roger approximation and the p-values were adjusted using the Tukey method. Adjusted R^2 values for the linear mixed effects models was computed using the rsq package (Zhang, 2021). Figures were created using the ggplot2 package (Wickham, 2016) and emmeans package (Lenth et al., 2020) for visualizing interactions. Variable selection for the linear mixed models was determined by conducting exploratory data analysis (via data visualization, t-tests, and ANOVAs), and then using likelihood ratio tests via an ANOVA function. We always attempted to add random slopes to the random effect terms of the models if signs of model overfitting were not detected. Speaker accent and vocal affective presentation style were effects (or deviation) coded, coding each level as 1 (pride and Singaporean) and -1 (shame and American). Other numerical predictors were rescaled (between 0 and 1) and centered, so the intercepts in the presented models can be interpreted in reference to the lowest rating for that variable (e.g., a rating of not at all friendly for vocal friendliness). The moderated mediation analysis was conducted using the mediate function in the mediation package (Tingley et al., 2014).

Cronbach's alpha (α) was calculated for each of the perceptual rating scales to assess their level of internal consistency. High internal consistency means that the average inter-item correlation between the speech stimuli for a given measure is high. There was high internal consistency for the competence (α = 0.854, 95% CI [0.789, 0.894]), hirability (α = 0.826, 95% CI [0.729, 0.881]), vocal attractiveness (α = 0.891, 95% CI [0.815, 0.929]), vocal friendliness (α = 0.853, 95% CI [0.731, 0.908]), and comprehensibility scales (α = 0.812, 95% CI [0.735, 0.865]). The internal consistency was acceptable for the accent strength scale (α = 0.712, 95% CI [0.557, 0.802]). Thus, the ratings for each scale were averaged across participants and used as dependent measures or predictors in the following analyses.

2.2.2 Results

Initial analyses

Manipulation check of confidence ratings

The pride utterances were perceived as sounding more confident (M = 6.69, SD = 1.80) than the shame utterances (M = 4.28, SD = 2.13), regardless of speaker accent (F (1, 1196) = 464.82, $p < .001$, $\eta_p^2 = 0.28$, 95% CI for η_p^2 [0.24, 0.32]). There was no main effect of speaker accent, F (1, 1196) = 3.42, p = 0.06). The perceived confidence of pride and shame utterances was moderated by speaker accent, (F (1, 1196) = 46.23, $p < .001$, $\eta_p^2 = 0.04$, 95% CI for η_p^2 [0.02, 0.06]). When expressing pride, speakers with American English accents were perceived as more confident than speakers with Singaporean English accents, mean difference = 0.55, p = 0.003, 95% CI [0.15, 0.96]. When expressing shame, speakers with American English accents were perceived as less confident than speakers with Singaporean English accents, mean difference = -0.97, $p < .001$, 95% CI [-1.37, -0.56]. Interpreted another way, when speakers with American English accents expressed pride versus shame, there was a greater difference in their perceived confidence ratings (mean difference = 3.17), compared to when speakers with Singaporean English accents expressed these vocal affective states (mean difference = 1.65) (Figure 2.1).

These results demonstrate that our manipulation of vocal affective presentation style affected the confidence ratings as predicted, with pride utterances conveying higher confidence compared to shame utterances. Moreover, participants perceived a greater change in perceived confidence when speakers with American English accents expressed pride versus shame compared to speakers with Singaporean English accents.

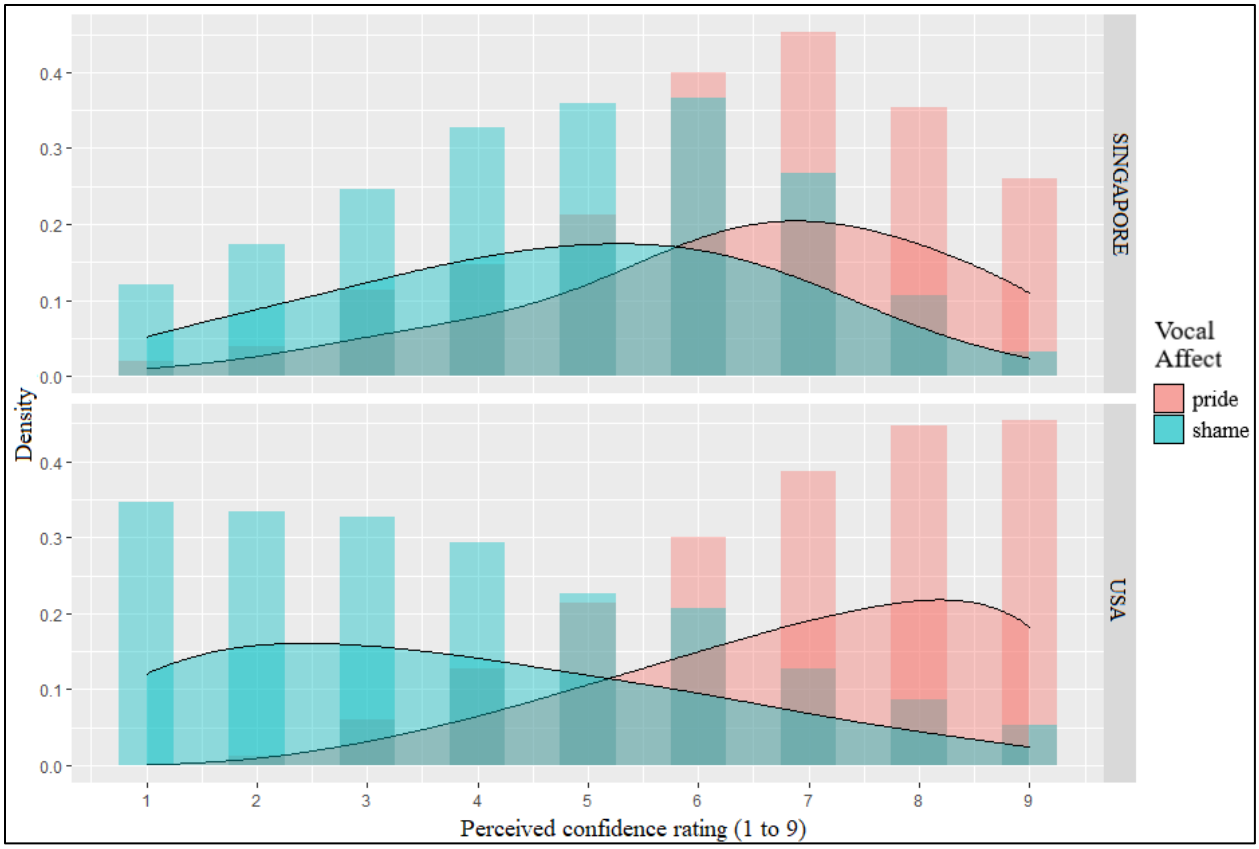

Figure 2.1. Density plot with overlayed histogram of perceived confidence ratings for speakers with American English or Singaporean English accents expressing vocal pride or shame. The perceived confidence ratings range from 1 = "not at all confident" to 9 = "very confident".

Competence and hirability ratings

The competence and hirability ratings were highly correlated, $\rho = 0.94$, $p < .001$. In other words, the more competent speakers were rated for the human resources manager position, the more likely they were rated to be hired. Thus, the competence and hirability ratings were summed (now out of 18) and this composite rating will subsequently be referred to as, "Job suitability", the primary outcome variable. See Table 2.3 for descriptive statistics of the perceptual ratings.

To investigate the extent to which each participant associated expressions of pride with higher job suitability and expressions of shame with lower job suitability, we calculated the difference score of job suitability ratings for each speaker between their expressions of pride versus

shame. As predicted, and similar to the perceived confidence ratings, there was a significantly greater mean difference in job suitability ratings for utterances by speakers with American English accents (M = 4.96, SD = 4.59) compared to speakers with Singaporean English accents (M = 1.81, SD = 4.10), t (590.27) = -8.88, p < .001, 95% CI [-3.85, -2.46].

Table 2.3. *Perceptual ratings based on speaker accent group and vocal affective presentation style.*

Speaker accent	Confidence		Job suitability		Social perception			
					Vocal attractiveness		Vocal friendliness	
	M	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*
				Pride				
American English	6.97	1.71	12.02	3.30	5.10	2.33	5.38	2.32
Singaporean English	6.42	1.85	10.96	3.56	4.92	2.27	5.45	2.33
				Shame				
American English	3.80	2.20	7.06	3.89	4.66	2.29	4.81	2.22
Singaporean English	4.77	1.95	9.15	3.57	4.88	2.29	5.17	2.32

Accent quality ratings

Speakers with Singaporean English accents were rated as having an accent that was significantly different from speakers with American English accents, (t (598) = -4.19, p < .001, 95% CI [-1.06, -0.38]), and more difficult to understand compared to speakers with American English accents, (t (593.87) = -2.79, p = 0.01, 95% CI [-0.65, -0.11]) (Table 2.4). Based on the average accent strength and comprehensibility ratings, the Singaporean English accent may be considered mildly different from speakers with Canadian English accents and relatively easy to understand, respectively.

Table 2.4. *Perceptual ratings of accent quality based on speakers' narrative utterance*

Speaker accent	Accent perception			
	Comprehensibility		Accent strength	
	M	*SD*	*M*	*SD*
American English	7.96	1.61	5.23	2.12
Singaporean English	7.57	1.75	4.50	2.12

Comprehensibility was rated on a 9-point scale from "Extremely easy to understand" to "Extremely difficult/impossible to understand". Accent strength was rated on a 7-point scale from "Heavily accented" to "No accent at all.

Social perception of speaker's voices

For perceived vocal attractiveness, there was a significant effect of vocal affective presentation style, F (1, 1196) = 113.45, $p < .001$, $\eta_p^2 = 0.09$, 95% CI for η_p^2 [0.06, 0.12], and speaker accent, F (1, 1196) = 41.68, $p < .001$, $\eta_p^2 = 0.03$, 95% CI for η_p^2 [0.02, 0.06]. This effect of speaker accent was moderated by speakers' vocal affective presentation style, F (1, 1196) = 59.43, $p < .001$, $\eta_p^2 = 0.05$, 95% CI for η_p^2 [0.03, 0.07]. Based on Tukey pairwise comparisons, when speakers expressed pride, speakers with American English accents were perceived as sounding more attractive than speakers with Singaporean English accents (mean difference = 1.73, $p < .001$, 95% CI [1.28, 2.18]). When speakers expressed shame, there was no significant difference in perceived vocal attractiveness based on speaker accent (mean difference = -0.15, $p = 0.81$, 95% CI [-0.60, 0.29]). Also, speakers with Singaporean English accents were perceived as sounding comparably attractive when they expressed pride or shame (mean difference = -0.36, $p = 0.16$, 95% CI [-0.81, 0.09]). In contrast, speakers with American English accents were perceived as sounding more attractive when they expressed pride compared to shame (mean difference = 2.25, $p < .001$, 95% CI [1.80, 2.69]).

For perceived vocal friendliness, there was a significant effect of vocal affective presentation style, F (1, 1196) = 368.64, p < .001, η_p^2 = 0.24, 95% CI for η_p^2 [0.20, 0.28], and speaker accent, F (1, 1196) = 40.58, p < .001, η_p^2 = 0.03, 95% CI for η_p^2 [0.02, 0.06]. This effect of speaker accent was moderated by speakers' vocal affective presentation style, F (1, 1196) = 50.92, p < .001, η_p^2 = 0.04, 95% CI for η_p^2 [0.02, 0.06]. Based on Tukey pairwise comparisons, when speakers expressed pride, there was no significant difference in perceived vocal friendliness based on speaker accent (mean difference = 0.09, p = 0.95, 95% CI [-0.33, 0.50]). When speakers expressed shame, speakers with American English accents were perceived as sounding less friendly than speakers with Singaporean English accents (mean difference = -1.53, p < .001, 95% CI [-1.94, -1.12]).

Main model – predicting speaker job suitability

A linear mixed effects model was used to determine how speakers' perceived job suitability varied based on their accent and vocal affective presentation style. The fixed effects included speaker accent (American English, Singaporean English), vocal affective presentation style (pride, shame), and their two-way interaction. Vocal attractiveness, vocal friendliness ratings, and their two-way and three-way interactions with speaker accent and vocal affective presentation style were included as control variables. The model also included a random intercept by participant with a correlated random slope to control for variability among participants as a function of the effect of vocal attractiveness and vocal friendliness ratings, and a random intercept by speaker.

The model had an adjusted R^2 of 0.57 in accounting for variation in job suitability ratings (0.21 from random effects, 0.35 from fixed effects) (Table 2.5). The model returned a significant intercept (B = 6.94, SE = 0.48, t (38.23) = 14.46, p < .001, 95% CI [5.97, 7.91], with the by-participant random intercept contributing SD = 1.95, and the by-speaker random intercept

contributing SD = 1.00). There was a significant effect of vocal affective presentation style, B = 1.19, SE = 0.21, t (1126.46) = 5.58, p < .001, 95% CI [0.77, 1.61], controlling for the above-mentioned fixed and random effects. There was no effect of speaker accent, B = 0.59, SE = 0.39, t (20.36) = 1.50, p = 0.15, 95% CI [-0.23, 1.40]. Participants gave higher job suitability ratings for speakers they perceived to sound more attractive, B = 3.46, SE = 0.57, t (62.52) = 6.08, p < .001, 95% CI [2.29, 4.58], or friendly, B = 2.66, SE = 0.59, t (100.10) = 4.50, p < .001, 95% CI [1.48, 3.83], controlling for the above-mentioned fixed and random effects, (by-subject random slope of vocal attractiveness contributing SD = 2.59 and random slope of vocal friendliness contributing SD = 2.52). There was a significant interaction between Speaker accent and Vocal affective presentation style, B = -0.74, SE = 0.21, t (1115.91) = -3.53, p < .001, 95% CI [-1.15, -0.33] (Figure 2.2). This interaction was further moderated by a three-way interaction with speakers' perceived vocal friendliness, B = 1.18, SE = 0.42, t (1130.60) = 2.84, p = 0.005, 95% CI [0.36, 2.01] (Figure 2.3). The pairwise comparisons, revealed that when speakers expressed pride, regardless of their perceived vocal friendliness, there was no difference in the perceived job suitability of speakers with American English or Singaporean English accents. When speakers expressed shame and were perceived as sounding less friendly (a rating of 1 or 2 out of 9), speakers with American English accents had significantly lower job suitability ratings than speakers with Singaporean English accents (rating of 1: mean difference = 2.79, SE = 0.95, t (41.2) = 2.94, p = 0.03; rating of 2: mean difference = 2.36, SE = 0.86, t (26.3) = 2.74, p = 0.05). The more friendly speakers were perceived to sound, there was no difference in the perceived job suitability ratings of speakers with American English accents versus Singaporean English accents.

Table 2.5. *Linear mixed effects model predicting speaker perceived job suitability (out of 18)*

Effect	Estimate	SE	95% CI	p

			[LL, UL]	
Fixed effects				
Intercept	6.94	0.48	[5.97, 7.91]	<.001
Vocal affective presentation style	1.19	0.21	[0.77, 1.61]	<.001
Speaker accent	0.59	0.39	[-0.23, 1.40]	0.15
Vocal friendliness	2.66	0.59	[1.48, 3.83]	<.001
Vocal attractiveness	3.46	0.57	[2.29, 4.58]	<.001
Speaker accent x Vocal affective presentation style	-0.74	0.21	[-1.15, -0.33]	<.001
Vocal affective presentation style x Vocal attractiveness	0.75	0.39	[-0.03, 1.53]	0.06
Vocal affective presentation style x Vocal friendliness	-1.21	0.43	[-2.05, -0.37]	0.005
Speaker accent x Vocal attractiveness	-0.16	0.41	[-0.98, 0.67]	0.71
Speaker accent x Vocal friendliness	-0.57	0.45	[-1.47, 0.33]	0.21
Speaker accent x Vocal affective presentation style x Vocal attractiveness	-0.30	0.42	[-1.06, 0.47]	0.44
Speaker accent x Vocal affective presentation style x Vocal friendliness	1.18	0.42	[0.36, 2.01]	0.005

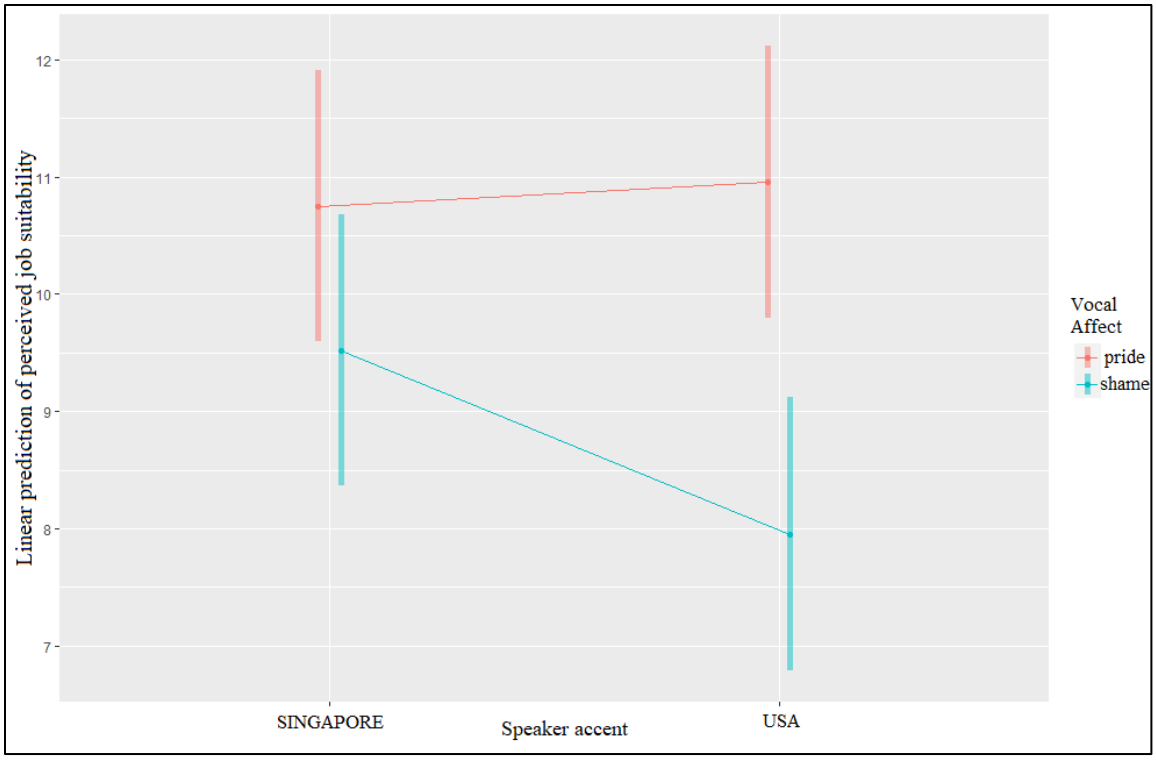

Figure 2.2. Interaction between speaker accent and vocal affective presentation style for average job suitability (out of 18) based on model prediction. The vertical bars indicate a 95% confidence interval.

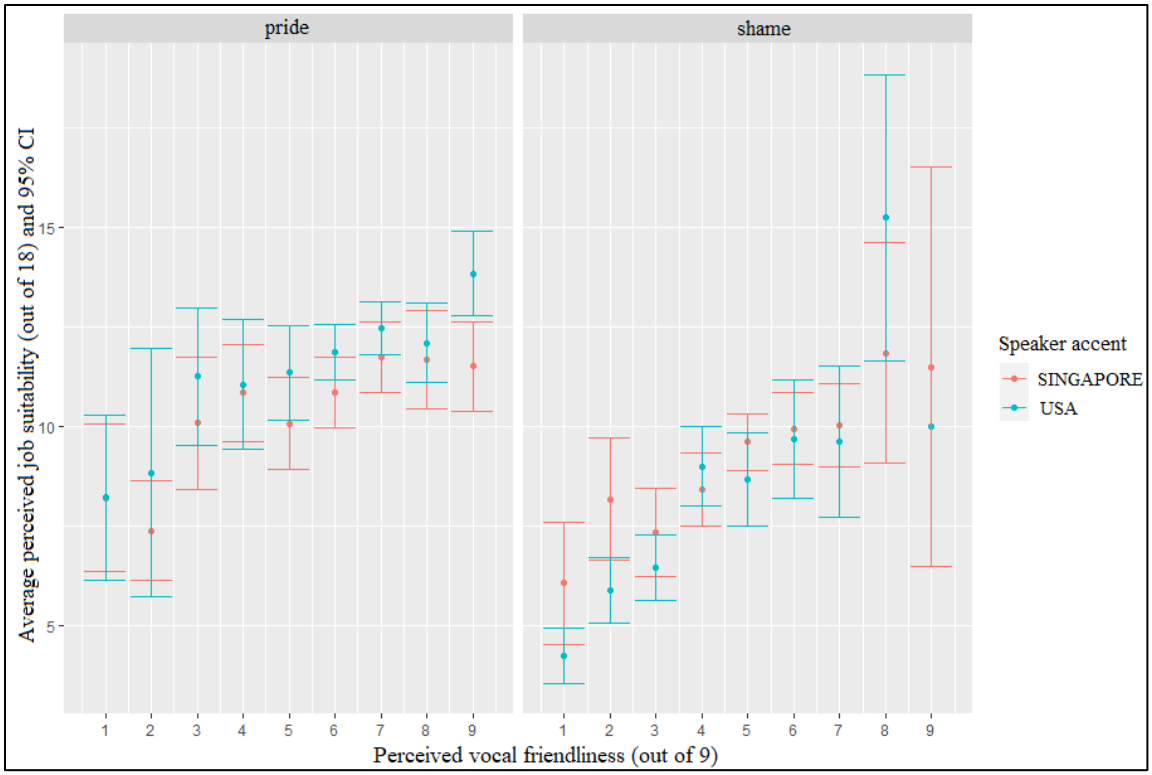

Figure 2.3. Relationship between speaker accent (American English, Singaporean English), vocal affective presentation style (pride, shame) and vocal friendliness (out of 9) on speakers' average job suitability ratings (out of 18).

Moderated mediation analysis of social perception variables

Given the greater magnitude of the effect of vocal friendliness on job suitability and its relationship with speaker accent and vocal affective presentation style, we examined vocal friendliness as a potential mediator. Unlike the interaction terms reported above, this analysis indicates whether vocal friendliness has a causal effect on the relationship between speaker vocal affective presentation style on perceived job suitability, moderated by speaker accent. We computed a moderated mediation model with vocal affective presentation style as an independent variable, speaker accent as a moderator, job suitability as a dependent variable, and vocal friendliness as a mediator. We used 95% bias-corrected and accelerated confidence intervals based

on 5000 bootstrap samples. The categorical variables (vocal affective presentation style, speaker accent) were contrast coded to -0.5 (American English accent, pride), 0.5 (Singaporean English accent, shame), and the continuous predictor (vocal friendliness) was centered and scaled to have a mean of zero.

The mediation analysis for vocal friendliness (Figure 2.4) revealed that speaker accent moderates the effect of speaker vocal affective presentation style on perceived vocal friendliness, $b = 0.70$, $SE = 0.10$, $t(1196) = 7.17$, $p < .001$. Speaker accent also moderates the effect of vocal friendliness on job suitability, $b = -0.64$, $SE = 0.23$, $t(1194) = -2.81$, $p < .01$, and the direct effect of speaker vocal affective presentation style on job suitability, $b = 1.50$, $SE = 0.45$, $t(1194) = 3.34$, $p < .001$. The indirect effect of vocal affective presentation style on job suitability through vocal friendliness was significant for speakers with standard and non-standard accents, $b = -1.56$, 95% CI [-1.82, -1.30], $p < .001$, with this indirect effect explaining 45.5% of the total effect. The indirect effect of vocal affective presentation style on perceived job suitability through vocal friendliness was stronger for speakers with American English accents compared to those with Singaporean English accents, estimate = 1.66, 95% CI [1.11, 2.22], $p < .001$.

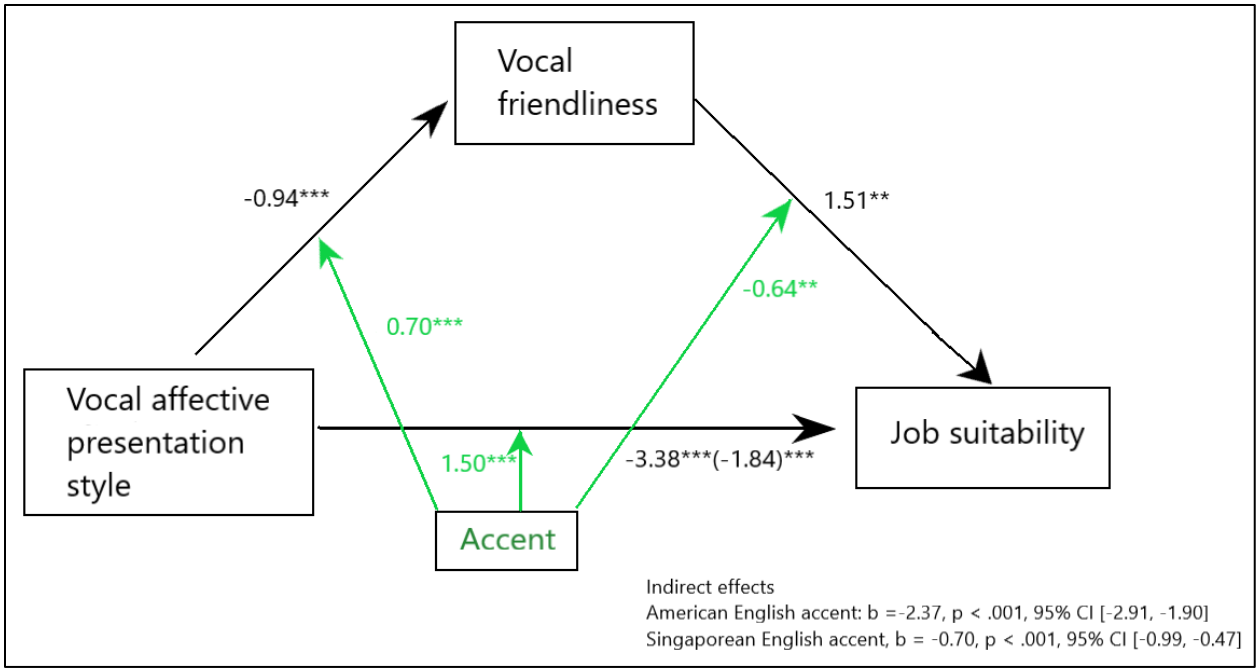

Figure 2.4. Moderated mediation model for the effect of speaker vocal affective presentation style on perceived job suitability, through perceived vocal friendliness and moderated by speaker accent. Significance levels are indicated for βs (***p<.001, **p<0.01).

Examining individual differences

To determine the effect of participants' individual variability on speakers' perceived job suitability, we first examined the intercorrelations between the various questionnaire responses and sociodemographic variables with job suitability ratings (Table 2.6). Then, we ran a linear mixed effects model with the language, cultural and social variables from the questionnaires predicting job suitability ratings. The model included fixed effects for speaker accent (American English, Singaporean English), vocal affective presentation style (pride, shame), and their 2-way interaction, scores on the Social Desirability scale, Tolerance of Ambiguity scale, and Accent Belief scale. Random intercepts by participant and speaker were also included. The model accounted for 44% of the variability in job suitability ratings (0.23 from fixed effects, 0.21

from random effects). None of the variables significantly predicted speakers' job suitability ratings.

Table 2.6. *Correlations for social perception variables, participant characteristics and job suitability*

Variable	1	2	3	4	5	6	7	8	9	10
1.Job suitability	----									
2.Participant age	-0.03	----								
3.Accent belief scale –accent diagnosticity	-0.09**	0.13**	----							
4.Accent belief scale – accent stability	-0.05	0.06**	0.09**	----						
5.Tolerance of ambiguity – Valuing diverse others	0.01	0.16**	-0.06**	-0.13**	-----					
6.Tolerance of ambiguity - Unfamiliarity	-0.02	0.25**	0.11**	-0.28**	0.20**	-----				
7.Tolerance of ambiguity – Coping with change	0.12**	0.01	0.12**	-0.25**	0.18**	0.41	---			
8.Tolerance of ambiguity – Challenging perspectives	0.06*	-0.04	-0.06*	0.03	0.11**	0.10	0.06	----		
9.Social desirability scale - Enhancement	-0.06*	0.01	0.04	0.01	0.07**	0.19	-0.09	0.30	----	
10.Social desirability scale - Denial	0.04	0.01	-0.13**	0.16**	0.34**	-0.04	0.16	-0.17	-0.03	----
11. Hiring/interview experience (days)	0.0	0.24**	0.17**	-0.09**	0.08**	-0.03	-0.08**	-0.03	0.15**	0.10**

***p<.001, **p<0.05, *p=0.05

2.2.3 Discussion

This study investigated how the perceived job suitability (competence and hirability) of speakers is affected by their accent and vocal affective presentation style in a Canadian hiring context. Previous studies examining the impact of a speaker's vocal cues and accent on competence-based impressions have manipulated speaker's vocal cues related to a mental state (e.g., confidence versus doubt) and when comparing speakers of Western cultural backgrounds (Jiang, Sanford, & Pell, 2018; Jiang, Keenan-Gossack, & Pell, 2020). Also, less is understood regarding how differences in speakers' cultural communicative styles for expressing high vocal competence, may affect their perceived suitability for a job. We found similar effects of vocal competence as it relates to an affective state, approximately replicating the effects from Jiang, Sanford, & Pell (2018); and Jiang, Keenan-Gossack, & Pell (2020).

Decoding speakers' vocal affective presentation style

Participants accurately decoded the associated confidence level of speakers' vocal affective presentation style. Moreover, this decoding of a high versus low confident state was more differentiated when speech was in an American English accent compared to a Singaporean English accent (Figure 2.1). A similar pattern was also observed for speaker's perceived job suitability. These results provide empirical evidence that native listeners can accurately decode the speech of speakers with non-standard accents, though with reduced accuracy compared to speech in a standard accent, supporting previous trends (Bradac & Wisegarver, 1984; Jiang et al., 2018; 2020). Participants' ability to discriminate between vocal pride and shame based on a speaker's accent, may indicate their greater ability to recognize vocal affective states in an in-group (standard) accent and draw social inferences (Laukka, Neiberg, & Elfenbein, 2014). This effect may be driven by the culture-specific expression of positive emotions, such as vocal pride, which may be harder to decode across cultures (Laukka et al., 2014; Sauter et al., 2010). Thus, a job candidate's cultural

background via their vocal affective presentation style can impact interviewers' ability to decode their vocal communicative intent and subsequently affect their perceived job suitability.

Listeners' less differentiated perception of vocal pride versus shame in a non-standard accent may broadly apply to the pragmatic inferencing of vocal cues in a non-standard accent, not just when a vocal mental/affective state is conveyed. A recent study by Puhacheuskaya and Järvikivi (2022) had speakers with Canadian English and Mandarin English accents naturally modulate the intonation of their voice to produce ironic and literal statements (criticisms and praise). No explicit instruction was given as to how to convey sarcasm, for example. Native English speakers heard these utterances and rated the speaker's level of irony. Ironic speech by speakers with Canadian English accents was perceived as more ironic than when it was produced by speakers with Mandarin English accents, while there was no effect of speaker accent for literal statements. The researchers suggest that pragmatic inferencing of speech in a non-standard accent may be reduced/differ for several reasons including listeners having difficulty integrating the meaning of speaker's vocal cues with a conceptual representation from memory or inferring a different (or incorrect) meaning from a speaker's vocal cues (Puhacheuskaya & Järvikivi, 2022). Thus, listeners may be more likely to not understand or misunderstand speakers' vocal affective presentation style if they have a non-standard accent. These results also support other proposed mechanisms, such as listeners encoding speech in their standard accent with greater strength (or attention) compared to speech in a non-standard accent (Sumner et al., 2014), and/or there being greater social and behavioural relevance for drawing social inferences from speech by speakers with in-group (standard) accents (Bestelmeyer et al., 2015).

Socially weighing a speaker's accent and vocal affective presentation style

As predicted, we observed a significant interaction between speaker accent and vocal affective presentation style, which was further moderated by speaker's perceived vocal friendliness. When speakers expressed pride, regardless of their perceived vocal friendliness, speakers with Singaporean English or American English accents received comparable job suitability ratings. When speakers expressed shame, if they were also rated as sounding less friendly, speakers with American English accents were rated as lower in job suitability than speakers with Singaporean English accents. In contrast, if speakers were rated as sounding more friendly, speakers with American English or Singaporean English accents received comparable job suitability ratings. This effect of vocal friendliness particularly when speakers are conveying low vocal competence may demonstrate that the vocal expression of shame is a less definitive diagnostic cue of low job suitability (Skowronski & Carlston, 1987). Rather, listeners are concurrently and (un)consciously drawing various social inferences (e.g., perceived friendliness) from a speaker's vocal cues (Freeman & Ambady, 2011) which can reduce the negative judgment of expressing low competence.

Vocal friendliness also significantly mediated the relationship between speaker vocal affective presentation style and perceived job suitability, as moderated by speaker accent. The indirect effect was greater for speakers with American English accents compared to speakers with Singaporean English accents. The perceived vocal friendliness of a speaker relates to interviewers' affective reactions of candidates, including their likability, warmth, perceived level of similarity and general interpersonal attraction (DeGroot & Motowidlo, 1999; Gallois, Callan, & Palmer, 1992; Howard & Ferris, 1996). In vocal friendliness being a stronger mediator for speakers with American English accents compared to Singaporean English accents, listeners may find it easier

to accurately decode vocal friendliness in a standard accent, potentially because of its greater social relevance or their greater experience connecting with in-group speakers.

This result relates to two major theories. Firstly, the similarity-attraction paradigm where participants perceive greater similarity to speakers with in-group accents which subsequently encourages a greater social attraction/liking of them (Byrne, 1971). This increased social attraction may extend to other social attributes of job candidates (Montoya et al., 2008; Montoya & Horton, 2012) such as a speaker's perceived job suitability. Secondly, the uncertainty reduction theory, where strangers interacting try to accurately predict each other's communicative behavior and explain it (Berger & Calabrese, 1975). The more similar a stranger's communicative behavior is, the greater the certainty and subsequent liking of them (Berger & Calabrese, 1975). In our study, speakers' accents and vocal affective presentation styles pose uncertainty as they indicate cultural and linguistic distance between job interviewers and candidates (Gudykunst, 1985). This preference to interact with people we have greater similarity with can reduce the risk of misunderstandings and the amount of cognitive effort needed to be understood (Hebbani & Colic-Peisker, 2012). Thus, the stronger mediating effect of vocal friendliness for job candidates with in-group accents may represent job interviewer's greater perceived certainty during a potential interaction (e.g., few breakdowns in communication, better decoding of vocal affective states), compared to job candidates with out-group accents. Since a job interview is also a high-stakes social context for job interviewers, this perceived reduction in uncertainty may be appealing.

Compared to studies that found a mediating effect of listeners' interpersonal attraction or general affect on the relationship between a speaker's accent and their perceived hirability or competence (Deprez-Sims & Morris, 2010; 2013; Dragojevic et al., 2017), we found similar results by listeners specifically attending to a speaker's vocal affective presentation style and rating the

quality of a speaker's voice. Previous studies used speech produced with a neutral tone or did not characterize the speaker's voice (Deprez-Sims & Morris, 2010; 2013; Dragojevic et al., 2017). Our result may suggest that previously listeners (un)consciously drew social inferences even from a neutral voice. Thus, examining a speaker's vocal cues can allow us to more accurately understand how a speaker's accent can impact their perceived job suitability.

Limitations and Conclusion

One possible limitation to our study is that there was a small number of speech stimuli and number of productions of each vocal affective presentation style by each speaker. This may not have provided sufficient power for the acoustic analyses where we did not observe predicted amplitude and pitch differences between utterances in a standard and non-standard accent. Although, acoustic differences on temporal measures were found supporting the effect of speech rate and pausing on perceived competence (Bradac & Wisegarver, 1984; Jiang et al., 2018; 2020). Also, our hypothesis regarding acoustic differences for vocal pride due to differences in speaker's cultural background and communicative style, was not supported. It is unclear whether vocal pride and shame are more difficult to elicit from one utterance, potentially requiring more context and the production of cues from other communication channels (Tracy& Robins, 2004a;2004b), or if using stimuli elicited with a greater intensity would lead to more pronounced acoustic differences. At the same time, we observed the predicted association between a speaker's perceived vocal confidence level and their vocal expression of pride and shame, adding to our understanding of how these emotions are vocally expressed. Nonetheless, more research is needed to better characterise vocal pride and shame.

Also, the speakers with Singaporean English accents had mild accents and though they were perceived as more difficult to understand compared to speakers with American English

accents, they were not too difficult to understand based on their mean rating. Results may differ if the speakers with non-standard accents were more difficult to understand and had stronger accents. In that case, we may expect a greater difference in job suitability and vocal friendliness and attractiveness ratings between speakers with standard and non-standard accents and greater difficulty decoding their vocal affective state.

It is also important to acknowledge the dynamic nature of job candidates' accents and vocal affective states. The strength of a speaker's accent may be dynamically shaped by their (un)conscious desire to emphasize or reduce the strength of their accent when interacting with speakers with standard English accents (Cheung & Sung, 2016; Gluszek, Newheiser, & Dovidio, 2011). This may occur often in high-stakes contexts where speakers are trying to make a good impression (Raghuram, 2013). Then in terms of a speaker's vocal cues, job candidates may (un)consciously adopt self-presentation management tactics to try to appear as competent as possible or show a level of belonging with a culture or organization (e.g., Jansen et al., 2012; König et al., 2011; Sandal et al., 2014). Future research could explore how these dynamic changes to a speaker's voice can affect their perceived competence.

Overall, in an intercultural hiring context involving job candidates with American English and Singaporean English accents for a human resources manager position, simply having a non-standard accent does not result in reduced competence ratings compared to speakers with standard accents. Listeners use a speaker's vocal cues indicating their accent, competence level, as well as perceived friendliness when forming an impression. For speech in a non-standard accent, listeners may be less accurate at differentiating between opposing vocal affective states and in the social inferences drawn. Thus, it may be advantageous for speakers with non-standard accents to use other speech cues to clearly communicate their mental/affective state (e.g., Bradac and

Wisegarver, 1984) and prevent potential miscommunications. Although competently responding to interview questions is beneficial, regardless of speaker accent, an interviewers' impression of a candidate's competence may be impacted more by their interest in socially connecting with them.

Chapter 3

Effect of job candidate's vocal competence level, accent, and ethnic appearance on their perceived job suitability

3.1 Introduction

This perceptual study examined how the perceived competence and hirability of speakers with standard (Canadian/American English) and non-standard (Singaporean English) accents within a Canadian hiring context, varies with speaker's vocal competence level (high: vocal pride/confidence versus low: vocal shame/doubt) and ethnic appearance (East Asian, European). These vocal affective/mental states represent high versus low competence, respectively, by a job candidate. The impact of these factors was investigated by presenting cues from the visual, and audio and visual communication channels. Like the Study 1 perceptual experiment, we simulated a high-stakes social context for participants by asking them to evaluate job candidates for a human resources manager position. We also continued to explore the effect of speaker's perceived vocal friendliness and attractiveness on participants' impressions, as well as variability in participant's social, cultural, and language characteristics.

This study consists of two perceptual experiments. Experiment 1 examined the interactive effects of speaker ethnicity (East Asian, European) and vocal confidence level (confidence, doubt) by speakers with in-group (Canadian English) accents. Experiment 2 examined the interactive effects of speaker ethnicity (East Asian, European) and vocal affective presentation style (pride, shame) by speakers with in-group (American English) and out-group (Singaporean English) accents. Due to differences in the vocal intention of the stimuli in the experiments, the utterances were compared for their perceived level of confidence. Also, the Canadian English and American English accent were considered in-group accents due to their high accent similarity.

The experiments used a paradigm seen in the impression updating literature (e.g., Bhanji & Beer, 2013; Brambilla, Carraro, Castelli, & Sacchi, 2019), by presenting speaker's ethnic appearance in Block 1, and then combining speaker's ethnic appearance with their voice (accent and vocal competence level) in Block 2. This allowed us to examine how participants temporally integrate visual and vocal cues about a job candidate and update their first impression to form one cohesive evaluation of their competence.

Here, ethnic appearance refers to a speaker's ethnic heritage based on physiognomical features in the face. People of East Asian and European heritage were selected as these broad ethnic groups differ in their social power within the Canadian context. People of European heritage are a dominant racial/ethnic group, socio-politically, economically, and historically. There are nearly 20 million people in Canada reporting English, Scottish, French, and Irish origins (Statistics Canada, Census of Population, 2016). In contrast, people of East Asian heritage (e.g., mainland China, Hong Kong, and Taiwan) constitute a visible ethnic minority group, and account for a large proportion of the immigrant population; also, Mandarin and Cantonese are among the top six non-official languages spoken at home (Statistics Canada, 2021 Census of Population).

3.2 Experiment 1 Perceptual Study (In-group accent only)

3.2.1 Objective

This experiment investigated the effect of a speaker's ethnic appearance on their perceived competence and hirability. Although having a standard (in-group) accent can lead to higher competence ratings compared to having a non-standard accent (e.g., Fuertes et al., 2012; Gluszek & Dovidio, 2010), it is unclear how robust this effect is. Specifically, it is unclear if this standard accent advantage is specific to speakers who are also part of a dominant racial/ethnic group or if it also applies to speakers who are visible ethnic minorities.

People of East Asian heritage, specifically Asian Canadians, may face several stereotypes regarding their competence (Padgett, Lou, Lalonde, & Sasaki, 2020). Most popularly, they may be viewed as model minorities, or high in competence and economic status compared to other immigrant/visible ethnic minority groups (e.g., Cargile, 1997; 2000; Fiske et al., 2002; Kil et al., 2019). On the other hand, they may be perceived as perpetual foreigners who have poor communication skills in English (e.g., Lee, Ottati, Lin, & Chan, 2014), suggesting low competence. Thus, it is unclear whether seeing a person of East Asian heritage will activate either of these stereotypes, particularly in a context involving people with Caucasian appearances. We predicted that when participants assess speaker's competence and hirability via ethnic appearance only (Block 1), if people are strongly activating cultural stereotypes related to high or low competence or expectations that people of East Asian heritage have a non-standard accent, then there will be a difference in competence and hirability ratings between speakers of East Asian versus European heritage. Alternatively, if a speaker's ethnic appearance is not a reliable diagnostic cue for assessing competence, then speakers of East Asian or European heritage may receive comparable competence and hirability ratings.

When participants assess competence via speaker's ethnic appearance, vocal confidence level and standard Canadian English accent (Block 2), we predict that speakers expressing vocal confidence will be perceived as more competent than those expressing vocal doubt, because it is a related to high competence. This result would also indicate that participants are accurately perceiving a speaker's vocal competence level. Of interest is the potential interaction between speaker vocal confidence level and speaker ethnicity. When speakers express vocal confidence, speakers of East Asian and European heritage may be rated comparably if this vocal display of high competence and having a standard accent carries more perceptual weight on speaker

competence than their ethnic appearance. Alternatively, if participants activate cultural/language attitudes and stereotypes based on speaker's ethnic appearance, there may be a difference in competence and hirability ratings between speakers of East Asian and European heritage. When speakers express vocal doubt, we predict no difference in the competence/hirability ratings for speakers of East Asian or European heritage if the vocal expression of low competence is a stronger more diagnostic cue of a speaker's competence. Alternatively, speakers of East Asian heritage may receive lower competence/hirability ratings than speakers of European heritage if the salient cue of vocal doubt and negative attitudes towards East Asians are compounded.

3.2.2 Methods

Participants

Participants were 151 native Canadian English speakers (75 females and 76 males, Mean Age = 30.44, SD = 4.52 years, Range = 24-40 years) recruited through the online platform Prolific Academic (www.prolific.co) [accessed September 2020]. Sample size was determined a priori from a power analysis in G*Power (Faul, Erdfelder, Lang & Buchner, 2007), based on a repeated measures ANOVA for the between-subjects factor of vocal confidence level. This sample size allows us to detect medium effects ($\eta_p^2 = 0.06$ or $d = 0.50$) with $\alpha = 0.05$, an assumed correlation of 0.50 between repeated measures, and a statistical power of 0.81, $F(3,132) = 2.67$. None of the participants completed Study 1.

All participants were born in and currently reside in Canada and identified their nationality as Canadian[2]. Approximately half of the participants were monolingual English speakers (n = 62). The remaining participants were bilingual (n = 72), trilingual (n = 11), or quadrilingual (n = 6).

[2] Participants reported their current residence as follows: Ontario (56.3%), British Columbia (16.6%), Alberta (11.9%), Quebec (7.3%), Manitoba (2.6%), Nova Scotia (2.0%), Saskatchewan (1.3%), New Brunswick (1.3%), and Newfoundland and Labrador (0.7%).

Some of these additional languages included French, Spanish, Mandarin, Vietnamese, Japanese, Cantonese, and Tamil. Most of the participants began learning only English at birth (n = 122). The remaining participants were simultaneous bilinguals (n = 3), who learned English and another language since birth, or sequential bilinguals who began learning English between the ages of 2-6 years old (n =21), or between the ages of 7-10 years old (n = 5).

Most participants identified their race/ethnicity as White/European (n = 88). The remaining participants identified as East Asian or Southeast Asian (n = 32) (e.g., Chinese, Vietnamese, Filipino), South Asian (n = 9), Black/African/Caribbean (n = 9), Latin American/Hispanic (e.g., Mexican, Colombian) (n = 2), or with two races/ethnicities (one of which included White/European) (n = 11). Also, the participants were relatively split in being third or more generation Canadian (n = 79) (i.e., participant and their parents were born in Canada) or second generation Canadian (n = 70) (i.e., participant was born in Canada, and at least one parent was born outside Canada), with the generation status unspecified for two participants (Statistics Canada, 2018, July 25). Participants had a wide range of occupations including registered nurse, student, accountant, software engineer, geologist, and cashier. 48% of participants had experience helping with interviewing and hiring (Range = 2 days to 15 years).

Participants reported normal or corrected-to-normal vision and hearing, and no speech, language or hearing disorders, or neurological and psychiatric disorders. Participants provided informed consent prior to the experiment and were compensated 5 CAD for their participation. This study was approved by the McGill Faculty of Medicine Institutional Review Board.

Stimuli/Materials

Visual stimuli

High resolution colour images of male and female faces of European and East Asian heritage from the MR2 Database (Strohminger et al., 2016) were used. All faces displayed a neutral expression, fair to medium skin tone, dark brown to black hair, and medium to dark brown eyes. See Strohminger and colleagues (2016) for details on the faces and norming procedure.

Faces were selected based on two criteria derived from the MR2 Database validation rating data (Strohminger et al., 2016). First, we assessed the face's perceived race, or the percentage of correct categorization responses when participants were asked, "What is the primary race of this person?" as well as the percentage of responses that perceived the face as multi-racial. We selected faces with a 95.74-100% categorization for the person's primary race and 6.38% or less of responses judged the person to be multi-racial. This criterion ensures there is minimal racial ambiguity when differentiating faces of East Asian or European heritage. Second, we assessed the face's perceived physical attractiveness, as this can affect a speaker's perceived competence (e.g., Hosoda, Stone-Romero, & Coats, 2003). In the norming data, participants rated the physical attractiveness of each face on a 7-point scale (1 = "not at all attractive" and 7 = "very attractive"). We calculated the mean physical attractiveness rating for each gender (e.g., the average physical attractiveness rating for selected female faces of East Asian heritage), and selected faces within 1.5 SD of this mean. The mean physical attractiveness rating for female faces was 3.60 (SD = 0.25, Range = 3.41 – 3.94), and 3.18 for the male faces (SD = 0.42, Range = 2.67 – 3.58). Eight faces were selected (four faces of East Asian heritage (two males, two females) and four faces of European heritage (two males, two females)), to allow for variability in each race/ethnic group.

The selected faces were then divided into two groups, each involving four faces. Each group consisted of one male and female of East Asian heritage, and one male and female of European heritage. See Appendix 1 for descriptive characteristics. The two groups of faces did not

significantly differ in their perceived race, t (6) = -0.10, p = 0.46, or physical attractiveness, t (4) = -0.24, p = 0.41. Overall, the selected faces depicted young adults of average physical attractiveness who were of European or East Asian heritage.

Speech stimuli

Utterances conveying confidence and doubt by speakers with Canadian English accents were obtained from a large inventory of utterances spoken in a confident or doubtful tone of voice (Jiang & Pell, 2017; 2018). Utterances were produced by native speakers of English from Ontario, Canada (two males, two females) saying one phrase ("*I'll be promoted very soon*"). Speakers had lay experience in acting or public speaking (Mean age = 24.5, SD = 2.65 years, Range = 22 – 28 years old). Speech was elicited by speakers responding to questions posed by an examiner with a certain level of confidence. The expression of each confidence level was not modeled to speakers. See Jiang and Pell (2017) for details on the speech elicitation procedure. The selected utterances were previously validated by 12 native speakers of Canadian English for their perceived confidence level (Jiang & Pell, 2017). Specifically, participants rated how confident the speaker sounded on a 5-point scale (1= not at all confident to 5 = very much confident). The confident utterances were perceived as sounding more confident (M = 4.38, SD = 0.43, Range = 3.73-4.64) than the doubtful utterances (M = 2.00, SD = 0.56, Range = 1.22-2.56), t (5.62) = 6.69, 95% CI [1.49, 3.25], p < .001.

Eight unique utterances were used in the perceptual experiment (1 utterance x 4 speakers with standard English accents x 2 vocal confidence levels (confidence, doubt)). Recordings were normalized to a peak intensity of 70 dB in Praat (Boersma & Weenink, 2020) and saved as .wav files. The selected utterances were short, ranging in duration from 1.30 to 2.48 seconds. There was

no significant difference in the duration of the confident and doubtful utterances, t (5.56) = -0.49, p = 0.65, 95% CI [-0.95, 0.64].

Acoustic features of speech stimuli

This analysis acoustically characterized the utterances expressing a vocal confidence level by speakers with Canadian English accents (Jiang, Keenan-Gossack, & Pell, 2020; Jiang & Pell, 2017; Pell et al., 2009). The same acoustic parameters as Study 1 were computed, using the Geneva Minimalistic Acoustic Parameter Set (GeMAPS) package (Eyben et al., 2016) and the publicly available openSMILE toolkit (Eyben et al., 2010). The measures were computed on the entire utterance for the expressions of confidence and doubt.

Table 3.1. *Major acoustic features of the confidence and doubt utterances.*

Acoustic parameter	Vocal confidence level			
	Confidence		Doubt	
	M	*SD*	*M*	*SD*
Mean logarithmic F0	27.70	4.74	29.63	6.26
Standard deviation of logarithmic F0	0.26	0.10	0.22	0.08
Mean Jitter	0.06	0.03	0.04	0.01
Mean Loudness	0.57	0.15	0.49	0.12
Standard deviation of amplitude	0.94	0.09	0.74	0.08
Mean Shimmer	1.49	0.36	1.07	0.23
Rate of loudness peaks	3.62	0.53	3.52	0.40
Alpha ratio	-18.64	3.58	-21.65	6.53
Hammarberg index	26.28	7.12	20.35	9.29
Syllable rate	2.59	0.91	1.95	0.64
Pause duration	0.14	0.06	0.19	0.07
Standard deviation of pause duration	0.15	0.08	0.17	0.01

The analysis revealed no significant differences between the confidence and doubt utterances by speakers with Canadian English accents for mean logarithmic f0, standard deviation of logarithmic f0, mean amplitude, mean jitter, mean shimmer, mean rate of loudness peaks, mean pseudo syllable rate, pause duration and standard deviation of pause duration (see Table 3.1). One difference for an amplitude measure was found. Speakers expressing confidence (M = 0.94, SD = 0.09) had greater variation in their amplitude than speakers expressing doubt (M = 0.74, SD = 0.08), t (5.96) = 3.31, p = 0.02, 95% CI [0.05, 0.34].

Design and Procedure

The experiment used a mixed-factorial design. Participants heard speech by speakers with standard English accents to investigate the effect of speaker ethnicity (East Asian, European heritage) and vocal confidence level (confidence, doubt) on their perceived competence and hirability (main dependent variables). Participants completed perceptual ratings in two blocks. In Block 1 (Visual Only), participants saw faces of East Asian and European heritage (within-subjects factor). In Block 2 (Audio and Visual), participants saw each speaker's face again paired with a voice expressing confidence or doubt (between-subjects factor). Participants rated each speaker once in each block.

The experiment was presented online using jsPsych (de Leeuw, 2015) and Just Another Tool for Online Studies (JATOS) (Lange, Kühn, Filevich, 2015) via the BRAMS Online Testing Platform (OTP)(https://brams.org/category/online-testing-platform/), where the instructions and task were written in English. Participants were randomly assigned to one of four versions of the experiment, created by pairing the confident and doubtful utterances with the two groups of faces (i.e., Group 1 faces-confident utterances, Group 1 faces–doubtful utterances, Group 2 faces–confident utterances, Group 2 faces–doubtful utterances).

A similar procedure to Study 1 was used except participants evaluated the perceived competence and hirability of job candidates in three blocks and with fewer job candidates. Participants were presented four job candidates. In Block 1, participants were randomly presented images of the job candidates. Each trial began with a 500-millisecond fixation cross and then the image with the rating scales displayed directly below. Participants completed two 9-point rating scales, 1) the speaker's competence for the position via, "How competent would this candidate be for the human resources manager position?", on a scale from "not at all competent" to "very much competent", and 2) their hiring recommendation via, "How likely would you recommend hiring this candidate as a human resources manager?", on a scale from "very likely" to "not at all likely". In Block 2, participants were presented the job candidates' face and voice. Each trial began with a 500-millisecond fixation cross, and then the image and audio clip were presented with the same rating scales as Block 1, displayed directly below. The faces from Block 1 were presented in a new randomized order. Across conditions, the male and female speakers were paired with faces of East Asian and European heritage. For example, one of the female voices was paired with a female European face in one condition (e.g., Group 1 faces – confident utterances) and a female East Asian face in another condition (e.g., Group 2 faces – confident utterances). This ensured that the perceptual ratings were not due to a specific voice-ethnicity pairing. Then in Block 3, participants ranked each candidate in terms of their hiring recommendation. Images of the job candidates were presented within a vertical rectangular space. Participants were asked to drag each image and order them into a list from, "most likely to hire" (at the top) to "least likely to hire" (at the bottom), without the images overlapping.

Then, participants completed questions regarding their cultural experiences/background, social traits, and explicit attitudes. Like Study 1, for social traits, participants completed the

Tolerance of Ambiguity scale (Herman et al., 2010), which measures participants' level of tolerance for ambiguity in cross-cultural settings. Participants also completed adapted items from the Scale of Anti-Asian American Stereotypes (SAAAS) (Lin et al., 2005) to measure their explicit cultural attitudes towards Asian Canadians.

Lastly, to better understand how participants perceived the job candidates' voices and faces, they rated the job candidates for accent similarity, vocal attractiveness, and physical attractiveness. The measure of accent similarity allowed us to account for regional accents in Canada and validate speakers' in-group accents. First, participants were presented the job candidates' voices again. Each trial began with a 500-millisecond fixation cross, the same audio stimuli from the experimental trials and two 9-point rating scales displayed directly below, 1) "How similar was this candidate's accent to your accent?", on a scale from "not at all similar" to "very similar" (Dragojevic & Giles, 2013), and 2) "How attractive was this candidate's voice to you?", on a scale from "not at all attractive" to "very attractive". Second, participants were presented the job candidates' faces again. Each trial began with a 500-millisecond fixation cross, the job candidate's face and one rating scale displayed directly below, "How physically attractive was this candidate?", using a 9-point scale from "not at all attractive" to "very attractive". The stimuli were presented in a randomized order. The experiment took 30 minutes to complete.

Statistical analysis

Conditions were combined based on vocal confidence level (i.e., confident conditions: Group 1 faces-confident utterances, Group 2 faces-confident utterances; doubtful conditions: Group 1 faces-doubtful utterances, Group 2 faces-doubtful utterances). This was done because our main hypotheses concerned the effect of speaker ethnicity and vocal confidence level on competence and hirability ratings. The following analyses involved 81 participants in the confident

condition (40 males, 41 females) and 70 participants in the doubtful condition (36 males, 34 females).

Analyses were performed and figures were created in RStudio (version 1.3.1093) (R Studio Team, 2019) (R Version 4.0.2, http://cran.r-project.org) (R Core Team, 2020). The same statistical analysis approach as Study 1 was used. For the linear mixed effects models, speaker ethnicity and vocal confidence level were effects (or deviation) coded, coding each level as 1 (confidence and East Asian) and -1 (doubt and European). Other numerical predictors were rescaled (between 0 and 1) and centered, so the intercepts in the presented models can be interpreted in reference to the lowest rating for that variable (e.g., a rating of not at all attractive for vocal attractiveness).

Cronbach's alpha (α) was calculated for each of the perceptual rating scales to assess their level of internal consistency. High internal consistency means that the average inter-item correlation between the visual or speech stimuli for a given measure is high. For the visual stimuli, there was high internal consistency on the competence scale (confident conditions: α = 0.868, 95% CI [0.704, 0.939]; = 0.896, 95% CI [0.804, 0.946], doubtful conditions: α = 0.867, 95% CI [0.730, 0.945]; = 0.957, 95% CI [0.916, 0.982]) and hirability scale (confident conditions: α = 0.852, 95% CI [0.655, 0.925]; = 0.829, 95% CI [0.627, 0.919], doubtful condition: α = 0.855, 95% CI [0.677, 0.936]; = 0.956, 95% CI [0.911, 0.981]), and acceptable internal consistency for the physical attractiveness scale (group 1 faces: α = 0.747, 95% CI [0.579, 0.842], group 2 faces: α = 0.803, 95% CI [0.660, 0.879]). For the speech stimuli, the internal consistency was high/acceptable for the competence scale (confident conditions: α = 0.781, 95% CI [0.650, 0.854], doubtful conditions: α = 0.844, 95% CI [0.747, 0.898]) hirability scale (confident conditions: α = 0.706, 95% CI [0.539, 0.810], doubtful conditions: α = 0.849, 95% CI [0.760, 0.903]), and accent similarity scale (confident conditions: α = 0.802, 95% CI [0.703, 0.874], doubtful conditions: α = 0.76, 95% CI

[0.648, 0.841]). It was acceptable for the vocal attractiveness scale (confident conditions: α = 0.60, 95% CI [0.388, 0.734], doubtful conditions: α = 0.682, 95% CI [0.516, 0.784]). Thus, the ratings for each scale were averaged across participants and used as dependent measures or predictors in the following analyses.

3.2.3 Results

Initial analyses

Competence and hirability ratings in Block 1 and Block 2

In Block 1, across conditions, faces of East Asian and European heritage were rated as comparably competent, t (600.68) = 1.74, 95% CI [-0.02, 0.42], p = 0.08 (East Asian faces: M = 6.24, SD = 1.35; European faces: M = 6.04, SD = 1.41). Thus, all speakers were perceived to be moderately competent based only on their face. Also, there was a marginal difference in hirability ratings for faces of East Asian and European heritage, t (596.88) = 1.91, p = 0.06, 95% CI [-0.01, 0.46] (East Asian faces: M = 6.14, SD = 1.40; European faces: M = 5.9, SD = 1.54).

In Block 2, faces of East Asian heritage were rated as more competent (F (1,600) = 8.58, p = 0.004, η_p^2 = 0.01, 95% CI for η_p^2 [0.00, 0.04]) and more hirable (F (1,600) = 12.11, p < .001, η_p^2 = 0.02, 95% CI for η_p^2 [0.00, 0.05]) than faces of European heritage regardless of speaker vocal confidence level. Also, speakers expressing confidence were rated as more competent (F (1,600) = 76.643, p < .001, η_p^2 = 0.11, 95% CI for η_p^2 [0.07, 0.16]), and more hirable (F (1,600) = 58.63, p < .001, η_p^2 = 0.09, 95% CI for η_p^2 [0.05, 0.13]), than those expressing doubt regardless of speaker ethnicity. There was no significant interaction between speaker vocal confidence level and ethnicity. These results demonstrate that our manipulation of vocal confidence level affected the competence and hirability ratings as predicted.

The competence and hirability ratings were also highly correlated across both blocks, ρ = 0.93, p < .001, and within each block (Block 1: ρ = 0.91, p < .001; Block 2: ρ = 0.94, p < .001). In

other words, the more competent speakers were rated for the human resources manager position, the more likely they were rated to be hired for the position. Thus, we summed the competence and hirability ratings (now out of 18) and will subsequently refer to this composite rating as, "Job suitability", our primary outcome variable (See Table 3.2 for descriptive statistics).

Social perception of speakers

For perceived accent similarity, speakers expressing confidence (M = 5.99, SD = 2.23) received higher accent similarity ratings than those expressing doubt (M = 5.49, SD = 2.37), t (578.24) = 2.67, p = 0.01, 95% CI [0.13, 0.87]. Based on these mean ratings, participants perceived speakers' voices to sound moderately similar to their own. For perceived vocal attractiveness, speakers expressing confidence (M = 5.38, SD = 1.99) were rated as more vocally attractive than speakers expressing doubt (M = 4.18, SD = 2.19), t (569.4) = 7.04, p < .001, 95% CI [0.87, 1.54]. Based on these mean ratings, participants perceived speakers' voices to be low to moderately attractive. Then for perceived physical attractiveness, speakers of East Asian (M = 4.83, SD = 1.80) and European (M = 4.97, SD = 1.83) heritage were comparable in physical attractiveness, t (601.89) = -0.96, p = 0.34, 95% CI [-0.43, 0.15]. Based on these mean ratings, participants perceived speakers' faces to be moderately attractive.

In summary, when speakers expressed confidence participants perceived their accents to be more like their own and their voices to be more attractive, compared to when speakers expressed doubt. Speakers of East Asian or European heritage received comparable physical attractiveness ratings.

Table 3.2. *Job suitability ratings based on speaker ethnicity and vocal confidence level*

Speaker ethnicity	Job suitability	
	Block 1: Visual Only	Block 2: Audio and Visual

	M	*SD*	*M*	*SD*
	Confidence			
European	11.80	2.82	10.73	3.38
East Asian	12.43	2.45	12.02	2.88
	Doubt			
European	12.13	2.93	8.69	3.98
East Asian	12.32	2.96	9.23	3.99

Job suitability ratings at Block 1 – Visual only

A linear mixed effects model was used to determine how speakers' perceived job suitability varied based on their ethnicity. The fixed effects included speaker ethnicity (European, East Asian). Physical attractiveness ratings, speaker sex, participant sex, and participant SAAAS scores were included as control variables. The model also included a random intercept by participant with a correlated random slope to control for variability among participants as a function of the effect of physical attractiveness ratings. Vocal confidence level was not included as a fixed effect because at Block 1 participants in all conditions were not exposed to speech cues.

The model had an adjusted R^2 of 0.79 in accounting for variation in job suitability ratings (0.65 from random effects, 0.14 from fixed effects). The model returned a significant intercept (B = 11.38, SE = 0.58, t (198.40) = 19.46, $p < .001$, 95% CI [10.22, 12.53], with the by-participant random intercept contributing SD = 2.52. There was an effect of speaker ethnicity, B = 0.27, SE = 0.06, t (449.89) = 4.79, $p < .001$, 95% CI [0.16, 0.38], with East Asian candidates receiving higher job suitability ratings that European candidates. Participants gave higher job suitability ratings for speakers they perceived to look more attractive, B = 3.85, SE = 0.51, t (119.45) = 7.55, $p < .001$, 95% CI [2.84, 4.86], (by-subject random slope of physical attractiveness contributing SD = 3.63). There was also a significant effect of speaker sex, B = -0.48, SE = 0.11, t (427.88) = -4.42,

$p < .001$, 95% CI [-0.69, -0.27], with female faces receiving higher job suitability ratings than male faces. No other effects were significant.

Change in job suitability ratings

To investigate the extent to which hearing a speaker's vocal confidence level changed the job suitability ratings for speakers of East Asian or European heritage, we calculated the difference in job suitability ratings for each speaker's face between Block 1 (Visual only) and Block 2 (Audio and Visual). As seen from Table 3.2, all speakers experienced a decrease in job suitability ratings in Block 2 compared to Block 1.

The difference in job suitability ratings was marginally affected by speaker ethnicity, $F(1, 600) = 3.59$, $p = 0.06$, $\eta_p^2 = 0.006$, 95% CI for η_p^2 [0.00, 0.02], with speakers of European heritage showing a trend towards a greater decrease in job suitability ratings than speakers of East Asian heritage. The difference in job suitability ratings was significantly affected by speaker vocal confidence level, $F(1, 600) = 83.51$, $p < .001$, $\eta_p^2 = 0.12$, 95% CI for η_p^2 [0.08, 0.17], with speakers expressing doubt showing a greater decrease in job suitability ratings. There was no significant interaction between speaker ethnicity and vocal confidence level.

Job suitability ratings at Block 2 – Audio and visual

A linear mixed effects model was used to determine how speakers' perceived job suitability varied based on their ethnicity and vocal confidence level. The fixed effects included speaker ethnicity (European, East Asian) and vocal confidence level. Accent similarity ratings, vocal attractiveness ratings, and their two-way and three-way interactions with speaker ethnicity and vocal confidence level were included as control variables. Also, speaker sex, participant sex, physical attractiveness ratings, and participant SAAAS scores were included as control variables. The model also included a random intercept by-participant and by-speaker.

The model had an adjusted R^2 of 0.69 in accounting for variation in job suitability ratings (0.31 from random effects, 0.38 from fixed effects). The model returned a significant intercept (B = 7.78, SE = 0.74, t (132.80) = 10.51, p < .001, 95% CI [6.32, 9.25], with the by-participant random intercept contributing SD = 2.02 and the by-speaker random intercept contributing SD = 0.52. Participants gave higher job suitability ratings if they heard the candidate express confidence compared to the doubt, B = 1.57, SE = 0.40, t (591.22) = 3.95, p < .001, 95% CI [0.79, 2.36]. Like Block 1, participants gave higher job suitability ratings for candidates they perceived to be more physically attractive, B = 1.42, SE = 0.54, t (596.76) = 2.65, p = 0.008, 95% CI [0.37, 2.47] and higher job suitability ratings to candidates of East Asian heritage compared to candidates of European heritage, B = -1.65, SE = 0.45, t (473.12) = -3.64, p < .001, 95% CI [-2.54, -0.76]. Participants gave higher job suitability ratings for speakers they perceived to sound more attractive, B = 5.32, SE = 0.64, t (530.50) = 8.31, p < .001, 95% CI [4.05, 6.58].

There was a significant interaction between speaker ethnicity and vocal confidence level, B = -0.96, SE = 0.45, t (473.61) = -2.12, p = 0.03, 95% CI [-1.86, -0.07] (Figure 3.1). The pairwise comparisons revealed that when speakers expressed confidence, speakers of East Asian heritage received higher job suitability ratings than speakers of European heritage, t (465) = 5.10, SE = 0.25, p < .001. When speakers expressed doubt, speakers of East Asian or European heritage received comparable job suitability ratings, t (469) = 0.41, SE = 0.27, p = 0.98. Also, speakers of European heritage received comparable job suitability ratings regardless of vocal confidence level, t (249) = 1.94, SE = 0.43, p = 0.22.

There was also a significant interaction between speaker vocal confidence level and vocal attractiveness, B = -1.63, SE = 0.62, t (534.62) = -2.64, p = 0.008, 95% CI [-2.85, -0.42] (Figure 3.2). The pairwise comparisons revealed that for speakers who were perceived to sound less

attractive (rating of 1 – 5 out of 9), they received higher job suitability ratings when they expressed confidence compared to doubt, (rating of 1: t (469) = 5.26, SE = 0.59, $p < .001$, rating of 2: t(359) = 5.31, SE = 0.51, $p < .001$, rating of 3: t(251) = 5.18, SE = 0.44, $p < .001$, rating of 4: t(183) = 4.70, SE = 0.40, $p < .001$, rating of 5: t(168) = 3.76, SE = 0.39, p = 0.03). The more attractive speakers' voices were perceived to sound (rating of 6 -9 out of 9), there was no difference in job suitability rating between speakers expressing confidence or doubt (rating of 6: t(204) = 2.53, SE = 0.41, p = 0.51, rating of 7: t (292) = 1.35, SE = 0.46, p = 1.00, rating of 8: t(406) = 0.40, SE = 0.54, p = 1.00, rating of 9: t(507) = -0.31, SE = 0.63, p = 1.00). No other effects were significant.

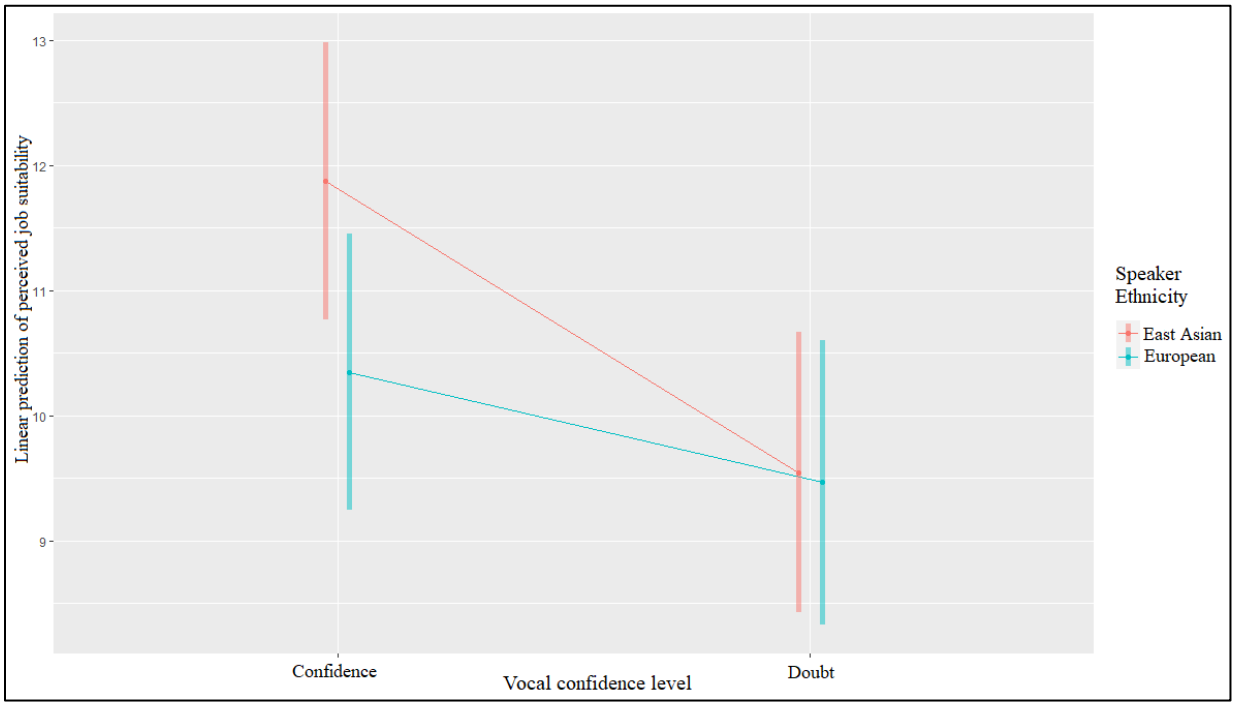

Figure 3.1. Interaction between vocal confidence and speaker ethnicity for average job suitability (out of 18) based on model prediction. Red indicates speakers of East Asian heritage, and blue represents speakers of European heritage. The vertical bars indicate a 95% confidence interval.

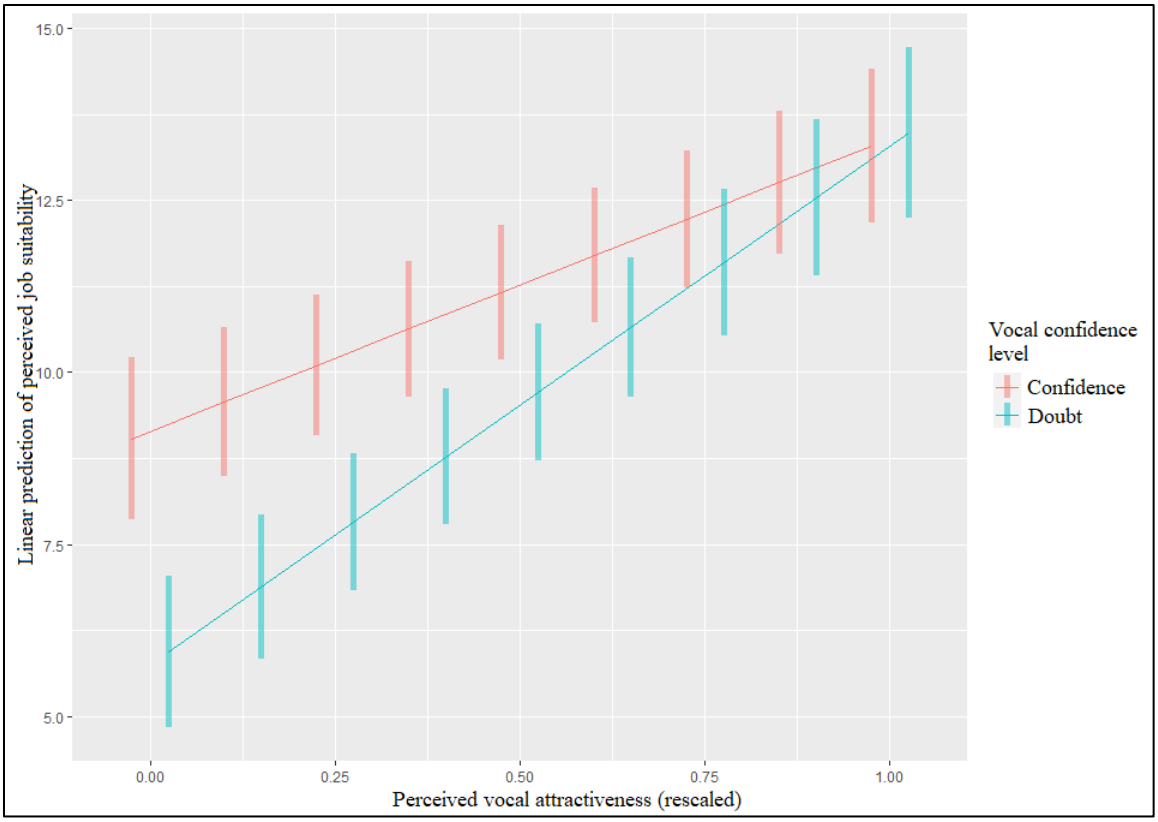

Figure 3.2. Interaction between speaker vocal confidence level and vocal attractiveness for average job suitability (out of 18) based on model prediction. The vertical bars indicate a 95% confidence interval.

Job candidate rankings

Participants' rankings of candidates as "most likely to hire" were analyzed. To determine each rank position (reported in X and Y pixel coordinates of the top of each image), images had to be separated by at least 80 pixels since the height of each image was 100 pixels. Data from six participants were excluded who either did not move the images at all (n = 2) or overlapped two or more of the images (n = 4) so the rank positions were not distinct. Overall, participants were more likely to rank East Asian candidates as their first choice (67.6%). There was no significant association between the top candidate's ethnicity and vocal confidence level, $\chi 2$ (1, N = 145) = 0.40, p = 0.53, or the candidate's ethnicity and sex, $\chi 2$ (1, N = 145) = 1.22, p = 0.27.

3.2.4 Discussion

This experiment aimed to better understand if the standard accent advantage is specific to speakers who are also part of a dominant racial/ethnic group or if it also applies to speakers who are visible ethnic minorities, as they vocally convey high or low competence. Specifically, we examined the interactive effects of speaker's ethnic appearance (European, East Asian) and vocal confidence level (confidence, doubt) on their perceived job suitability when all speakers have an (in-group) standard English accent.

Processing a speaker's ethnic appearance

When perceivers only saw a speaker's ethnic appearance, speakers of East Asian heritage received higher job suitability ratings, potentially suggesting that participants activated associated concepts or cultural stereotypes about people of East Asian heritage. This result differs from Hansen, Rakić, & Steffens (2018) where Turkish and German males were rated as comparably competent when participants only saw their faces, despite people of Turkish heritage being stereotyped as low in competence (Hansen et al., 2018). In contrast, people of East Asian heritage, specifically Chinese, are stereotyped as high in competence compared to other immigrant populations or visible ethnic minority groups (e.g., Cargile, 1997; 2000; Fiske et al., 2002; Kil et al., 2019). Although, participants' scores on the SAAAS did not significantly predict candidate's job suitability. It measured their explicit cultural attitudes towards Asian Canadians including competence-based stereotypes. This may have occurred because participants provided socially desirable answers, and/or this scale does not activate the same concepts about speakers of East Asian heritage as when viewing their face. Thus, for visible ethnic minorities who may be associated with high competence stereotypes, their ethnic appearance alone may be used as a diagnostic cue of their job suitability.

At the same time, other qualities of a job candidate's facial appearance more strongly affected their perceived job suitability. The more physically attractive a job candidate was perceived to be, the higher their job suitability ratings. This heuristic is reported to affect social judgments in a job interview (Dion et al., 1972; Eagly et al., 1991; Hosoda, Stone-Romero, & Coats, 2003; Nisbett & Wilson, 1977). Previous studies examining the effect of job candidate physical attractiveness have largely involved images of Caucasian men or women, or candidate ethnicity was not disclosed (Eagly et al., 1991; Hosoda, Stone-Romero, & Coats, 2003). Our result suggests this effect of physical attractiveness extends to an interethnic hiring context involving visible ethnic minorities.

Processing a speaker's ethnic appearance and vocal confidence level

We then examined how a speaker's vocal confidence level and standard accent interacts with their ethnic appearance. Of interest was the observed interaction between speaker vocal confidence level and ethnic appearance. When speakers expressed doubt, there was no difference in job suitability ratings between speakers of East Asian and European heritage. This result provides evidence that a doubtful tone of voice is a more diagnostic cue for assessing a candidate's job suitability amidst ethnic appearance cues. Conversely, when speakers expressed vocal confidence, speakers of East Asian heritage received higher job suitability ratings than speakers of European heritage and were more often ranked as the candidate most likely to be hired. This result demonstrates that when candidates vocally express high competence, ethnicity cues carry more perceptual weight than having a standard accent or expressing confidence.

The higher ratings for speakers of East Asian heritage may be due to a few reasons. One, participants experienced a positive violation of their expectations by speakers of East Asian heritage having a Canadian English accent. This result is in line with findings by Hansen and

colleagues (2017; 2018) where speakers who were visible ethnic minorities with standard accents were perceived as more competent than speakers who were part of the dominant racial/ethnic group with standard accents. Relatedly, speakers of Chinese heritage depicted as having a Canadian English accent may be highly regarded because their standard accent demonstrates they have acculturated into Canadian society or more generally the host culture (Baquiran & Nicoladis, 2020). This perceived level of acculturation can positively impact a job candidate's perceived level of cultural fit and hirability (Bye et al., 2014; Hofhuis et al., 2016; Horverak et al., 2013). On the other hand, in Canada, speakers of various visible ethnic minority groups can have standard English accents. Thus, participants may not have been surprised or impressed that speakers of East Asian heritage had standard accents. Instead, participants may have strongly activated stereotypes of high competence for people of East Asian heritage (e.g., Cargile, 1997; 2000; Fiske et al., 2002; Kil et al., 2019). Alternatively, participants may have responded in a socially desirable way by rating members of a visible ethnic minority more favourably than members of a dominant racial/ethnic group. This has been observed when race/ethnicity is a highly salient cue (e.g., Mullins, 1982) and when participants evaluate job candidates that differ in race from themselves (Lewis & Sherman, 2003). However, if participants were responding in a socially desirable way, candidates of East Asian heritage should have also received higher ratings in the vocal doubt condition, which was not the case. Then again, speaker ethnicity may have been more salient to participants when job candidates were expressing vocal confidence, compared to doubt. More research is needed to better control for participant's tendency to respond in a socially desirable way and to understand the extent to which job interviewers' judgments in an interethnic hiring context are affected by perceived social pressures.

Between Block 1 and 2, with the addition of vocal cues, all speakers regardless of vocal confidence level, received lower job suitability ratings. This result is unlike Hansen and colleagues (2017; 2018) where job candidates with a standard accent, regardless of ethnicity, experienced an increase in competence and hirability ratings compared to when participants were only presented their face. Also, contrasting Hansen and colleagues, we observed a greater mean difference in job suitability ratings between Block 1 and 2. This decrease in perceived job suitability ratings may be due to a couple of reasons. One, participants' expectations for the speakers' voices were not met after viewing their appearance (e.g., a speaker sounded less attractive than expected). Two, participants evaluated speakers differently due to the pragmatic function of the speech stimuli used. In Hansen and colleagues (2017; 2018), speakers uttered a greeting ("Good morning, nice to meet you" in German), which can allow a speaker to socially connect with a listener and may be produced with other non-verbal cues of positive affect (Bach & Harnish, 1979; Duranti, 1997). Conversely, in our study speakers produced an assertion ("I'll be promoted very soon"), which allows a speaker to inform a listener of information or express beliefs (e.g., Williamson, 2002). Thus, participants may have been less probed to form a positive impression of speakers based on the linguistic content of their speech. In an interview, job candidates may produce both of these phrases at different stages (e.g., greetings at the beginning versus asserting their competence in the middle). Future research could explore how a job interviewer's impressions of a job candidate changes over time based on the type of speech act produced.

Social processing of a speaker's voice

In Block 2, speaker's job suitability was most strongly explained by speaker's perceived vocal attractiveness. This effect may demonstrate that participants continued to draw inferences about the speaker's personality and other social attributes as they integrated the speaker's vocal

competence level and ethnic appearance. Specifically, a speaker's perceived vocal attractiveness relates to their perceived likability and an interviewer's affect towards a job candidate (Gallois et al., 1992; Howard & Ferris, 1996). It is found to be positively associated with perceived extraversion, warmth, and low neuroticism (i.e., being calm or at ease) (Zuckerman et al., 1995). Previous studies also report a positive association between a speaker's vocal attractiveness with their job interview ratings, job performance ratings (DeGroot & Kluemper, 2007) and perceived achievement (i.e., competent, lazy, industrious) (Zuckerman & Driver, 1989). Thus, our result may suggest that participants' inferences about a candidate's social competence had a greater impact than their vocal confidence level. This impact of vocal attractiveness was not moderated by speaker ethnicity suggesting that in an interethnic hiring context, vocal qualities related to a job interviewer's social connection to a candidate carries greater weight on their impression than ethnicity cues. Our results also suggest that having an attractive sounding voice may counteract the negative effects of expressing doubt or uncertainty.

Limitations and Conclusion

The ecological validity of our speech stimuli could be improved by using speech from speakers with standard English accents who are visible ethnic minority members. Our speakers were all Caucasian. There is some research suggesting that people can accurately detect the ethnicity of native English speakers who are of East Asian, Southeast Asian, or South Asian heritage (Hanna, 1997; Newman & Wu, 2011; Wong & Babel, 2017), though there is a lot of variability across speakers. Moreover, some research has found ethnic differences in social anxiety between Americans of European versus Asian heritage (Krieg & Xu, 2015), which may contribute to potential differences in their vocal expression of low competence. Also, our study used faces of average physical attractiveness, with females not wearing any makeup, and all faces having the

same relative hair colour. In a real hiring context, other aspects of a job candidates' physical appearance including their body weight/type, height, clothing, and hair may contribute to job interviewers' impression of their competence (Lowman, Harms, & Mills, 2019). More research is needed regarding how these factors may interact with a speaker's ethnic appearance and vocal cues to affect their perceived job suitability.

Also, the study methodology of presenting faces only in Block 1 and then combining this information with the speaker's voices in Block 2, may have constrained participants' job suitability ratings. Although we were interested in the integration of visual ethnicity cues and vocal competence cues, we also created a context where participants may have used the speaker's physical appearance as a relevant criterion for the human resources manager position. This context may have encouraged participants to commit the speaker's physical appearance as the basis of their impression (e.g., Uhlmann & Cohen, 2005), potentially limiting the change in job suitability ratings we observed in Block 2. Other methodologies are needed that can allow us to understand the individual and combined weight of visual ethnicity cues and vocal competence cues as they may affect impressions in real-world contexts.

Overall, our findings reveal that perceivers' integration of a job candidate's standard English accent, vocal confidence level, and ethnic appearance affects their job suitability. The standard accent advantage may apply to more than just speakers of a dominant race/ethnic group. When job candidates all have a standard accent and are conveying high vocal competence, ethnicity cues, may impact an interviewer's competence impression. In contrast, perceivers may shift their processing of a candidates' speech when candidates convey low vocal competence. This may occur because sounding doubtful carries more diagnostic value for assessing job suitability than a candidate's standard accent or ethnic appearance. Above all, the social connection formed

with a candidate's voice via their perceived social competence may carry the greatest weight on their perceived suitability for a job.

3.3 Experiment 2 Perceptual study (In-group and Out-group Accent)

3.3.1. Objective

This experiment investigated the effect of a speaker's ethnic appearance and accent on their perceived job suitability (i.e., competence and hirability). Although having a non-standard (out-group) accent can lead to lower competence ratings, less is understood regarding the impact of a speaker's ethnic appearance on their perceived job suitability, especially as this cue is combined with a speaker's accent and vocal competence level. We examined the perceived job suitability of speakers with standard (American English) accents who were of European heritage and speakers with non-standard (Singaporean English) accents who were of East Asian heritage as they conveyed vocal pride or shame (vocal affective presentation style).

We predicted that when participants evaluate speaker's competence and hirability via ethnic appearance only (Block 1), like Experiment 1, if people are strongly activating cultural stereotypes related to competence or expectations that people of East Asian heritage have a non-standard accent, then there will be a difference in competence and hirability ratings between speakers of East Asian heritage versus European heritage. Alternatively, if a speaker's ethnic appearance is not a reliable diagnostic cue for assessing competence, then speakers of East Asian or European heritage may receive comparable competence and hirability ratings.

Then when participants assess competence and hirability via speaker's ethnic appearance, vocal affective presentation style, and standard or non-standard accent (Block 2), we predicted that speakers expressing vocal pride will be perceived as more competent than those expressing vocal shame. This result would indicate that perceivers are accurately perceiving a speaker's vocal affective state, impacting the perceived competence rating.

Of interest is the potential interaction between speaker vocal affective presentation style and speaker accent. We predict that speakers with American English accents and European heritage expressing pride will be perceived as more competent than speakers with Singaporean English accents if their East Asian appearance increases the salience of their out-group accent. When speakers are expressing vocal shame, we have similar predictions to Study 1 perceptual experiment with potentially amplified effects. There may be no difference in the perceived competence and hirability ratings of speakers with Singaporean or American English accents if the vocal expression of shame is a stronger more diagnostic cue of a speaker's competence, compared to their accent or ethnic appearance. Alternatively, speakers with Singaporean English accents with East Asian appearances may receive higher competence ratings if listeners engage in reduced or different inferential processing of speech in a non-standard accent, similar to Study 1. Otherwise, speakers with Singaporean English accents may receive lower competence ratings than speakers with American English accents if the salient cue of vocal shame, a non-standard accent, and possible negative attitudes towards East Asians are compounded. Like the findings from Study 1 and Study 2 Experiment 1, we predict that a speaker's perceived vocal attractiveness and/or friendliness may further moderate the impact of a speaker's vocal affective presentation style, accent, and ethnic appearance.

Lastly, analyses were conducted between experiments and studies to explore the impact of a job candidate's ethnic appearance on their subsequent job suitability. First, we compared the perceived job suitability of East Asian job candidates with standard accents (Experiment 1 Block 2) and non-standard accents (Experiment 2 Block 2). If a speaker's ethnic appearance greatly shapes their perceived job suitability, there may be no difference in perceptual ratings between these experiments. Conversely, if a speaker's accent carries more perceptual weight on the job

suitability impression, East Asian speakers with standard accents will receive higher ratings, when expressing high vocal competence. They may receive lower ratings when expressing vocal shame, similar to the results found in Study 2 Experiment 1. Second, we compared the perceived job suitability of job candidates based on their vocal cues only (Study 1 experiment) and job candidates based on their vocal cues and ethnic appearance (Experiment 2 Block 2) to determine the additive effect of speaker's ethnic appearance. For speakers with Singaporean English accents, if their non-standard accent greatly shapes their perceived job suitability, there may be no difference in ratings between these communicative contexts, regardless of vocal competence level. This may occur because hearing a speaker's non-standard accent, activates concepts for a typical ethnic appearance. Alternatively, if a speaker's ethnic appearance makes their non-standard accent more salient, then we may observe lower job suitability ratings compared to Study 1.

3.3.2 Methods

Participants

Participants were 102 native Canadian English speakers (47 males, 55 females, Mean Age = 29.97 years, SD = 5.32 years, Range = 24-40 years) recruited through the online platform Prolific Academic (www.prolific.co) [accessed April – June 2022]. Sample size was determined a priori from a power analysis in G*Power (Faul, Erdfelder, Lang & Buchner, 2007) based on a repeated measures ANOVA for the between-subjects factor of vocal affective presentation style. This sample size allows us to detect medium effects ($\eta_p^2 = 0.06$ or $d = 0.50$) with $\alpha = 0.05$, an assumed correlation of 0.50 between repeated measures, and a statistical power of 0.80, $F(3, 92) = 2.70$. Two participants were excluded from the analyses because they did not play all the audio stimuli. The following analyses are based on 100 participants (45 males, 55 females). None of the participants had taken part in Study 1 or Study 2 Experiment 1.

All participants were born in and currently reside in Canada and identified their nationality as Canadian[3]. Approximately half of the participants were monolingual English speakers (n = 48). The remaining participants were bilingual (n = 42) or multilingual (n = 10). Some of these additional languages included French, Spanish, Cantonese, Polish, Tagalog, Tamil, and German. Most participants began learning only English at birth (n = 93). The remaining participants were simultaneous bilinguals (n = 2) or sequential bilinguals (n = 5), who began learning English between the ages of 1-4 years old. Participants attended school in English (excluding 3 participants with measurement error in their responses) and were interested in working in English (1 participant indicated an interest in working in English and French).

Most participants identified their race/ethnicity as White/European (n = 65). The remaining participants identified as Asian (n = 19, East Asian: n = 10, South Asian = 6, Southeast Asian = 1, a combination of two Asian races/ethnicities: n =2), Black/African/Caribbean (n = 3), Middle Eastern/Arab (n = 1), Indigenous (n =1), two races/ethnicities (one of which included White/European) (n = 6) or two or more races/ethnicities (not including White/European) (n = 3). Two participants responded 'Other' or did not indicate any race/ethnicity. Also, approximately half of the participants were third or more generation Canadian (n = 52). The remaining participants were second generation Canadian (n = 46), or unspecified (n = 2). Participants had a wide range of occupations including accountant, data analyst, teacher, nurse, and graphic designer. 51% of participants had experience helping with the interviewing and hiring process (Range = 1 days to 10 years).

[3] Participants reported their current province of residence as follows: Ontario (57%), Alberta (17%), British Columbia (10%), Nova Scotia (4%), New Brunswick (4%), Manitoba (3%), Saskatchewan (3%), Quebec (1%), and Newfoundland (1%).

Participants reported normal or corrected-to-normal vision and hearing, and no speech, language or hearing disorders, or neurological and psychiatric disorders. Participants provided informed consent prior to the experiment and were compensated 7.50 CAD for their participation. This study was approved by the McGill Faculty of Medicine Institutional Review Board.

Stimuli/materials

Visual stimuli

The same images as Study 2 Experiment 1 were used.

Speech stimuli

The same vocal pride and shame utterances by speakers with American English or Singaporean English accents were used, as described in Study 1 perceptual experiment. Utterances from two male and two female speakers from each accent group were selected (8 speakers total). Selected utterances had the largest mean difference in perceived confidence ratings between the vocal pride and shame expressions, and the lowest variability via standard deviation (see Appendix F). This selection process ensured that the vocal affective states can be differentiated, especially before combining them with speaker ethnic appearance.

Due to the differences in vocal intention of Experiment 1 and 2, we validated that there was no significant difference in the perceived confidence of vocal confidence and pride utterances, or between the vocal doubt and shame utterances. We used the validation data for the vocal confidence and doubt utterances by speakers with Canadian English accents (Jiang & Pell, 2017), and the perceived confidence ratings collected during Study 1 for the vocal pride and shame utterances by speakers with American English accents. Since the perceived confidence ratings were on different Likert scales (5-point and 9-point scales, respectively), we first scaled the mean perceived confidence ratings for each speaker's vocal expression to be a value between 0 and 1.

This allowed us to compare the perceived confidence ratings from the two experiments. Then we examined the difference in perceived confidence ratings based on speaker accent (American English versus Canadian English accent), and vocal expression (pride, shame, confidence, or doubt) (Figure 3.3). There was no effect of speaker accent ($F(1, 12) = 0.41$, $p = 0.54$). In other words, speech in an American English or Canadian English accent received comparable confidence ratings. There was an effect of vocal expression ($F(2, 12) = 40.78$, $p < .001$, $\eta_p^2 = 0.87$, 95% CI for η_p^2 [0.66, 0.93] (Table 3.3). No other effects were significant. The Tukey pairwise comparisons revealed that there were comparable confidence ratings between the vocal confidence and pride utterances (Mean difference = 0.01, $p = 1.00$), and between the vocal doubt and shame utterances (Mean difference = -0.01, $p = 1.00$) (Figure 3.3). Also, the vocal pride utterances were rated as sounding more confident than the vocal doubt utterances (Mean difference = 0.70, $p < .001$), and the vocal confidence utterances were rated as sounding more confident than the vocal shame utterances (Mean difference = 0.70, $p < .001$). Due to these comparable differences, we compared the vocal expressions between Experiment 1 and 2 as referring to high and low vocal competence levels.

Table 3.3. *Perceived confidence ratings for utterances conveying a vocal competence level. Perceived confidence ratings are scaled to be between 0 (low confidence) and 1 (high confidence).*

		Vocal confidence	
Speaker accent	Vocal competence expression	*M*	*SD*
Canadian English	confidence	0.92	0.13
	doubt	0.23	0.16
American English	pride	0.88	0.13
	shame	0.17	0.19

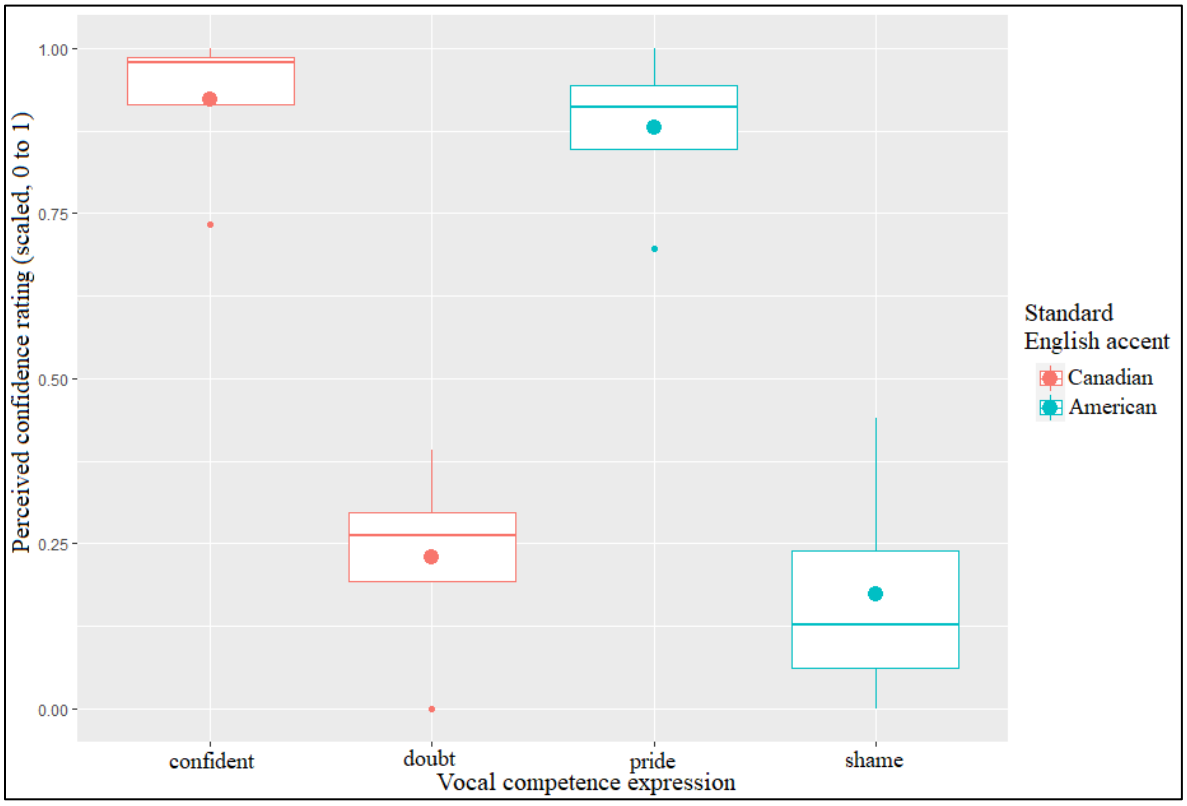

Figure 3.3. Scaled perceived confidence ratings (range from 0 to 1) between the utterances conveying vocal pride and shame (vocal affective presentation style), and vocal confidence and doubt (vocal mental state) by speakers with Canadian or American English accents.

Design and Procedure

This online study used the same mixed-factorial design as Study 2 Experiment 1. All participants completed competence and hirability ratings at two time points (Block 1: Visual Only, Block 2: Audio and Visual) and were presented speakers (or job candidates) of East Asian and European heritage (within-subject factors). All speakers of East Asian heritage had a non-standard (Singaporean English) accent, and all speakers of European heritage had a standard (American English) accent. In Block1, participants were presented an image of each job candidate's face. In Block 2, participants were again presented an image of each job candidate's face paired with a

voice. Participants were randomly assigned to hear all job candidates produce either vocal pride or shame (between-subjects factor). Participants were randomly assigned to one of four versions of the experiment, created by pairing each speaker's vocal pride and shame utterances with the two groups of faces (i.e., Group 1 faces-pride utterances, Group 1 faces–shame utterances, Group 2 faces–pride utterances, Group 2 faces–shame utterances).

The experiment was presented online using jsPsych (de Leeuw, 2015) and Just Another Tool for Online Studies (JATOS) (Lange, Kühn, Filevich, 2015) via the BRAMS Online Testing Platform (OTP)(https://brams.org/category/online-testing-platform/), where the instructions and task were written in English. The same procedure as Study 2 Experiment 1 was used, except participants evaluated eight, rather than four, job candidates. Similar to Study 1, we measured participants' explicit attitudes towards people with non-standard accents via the Accent Belief Scale (Hansen, 2020), and participants' tendency to minimize their general explicit negative bias via the Social Desirability Scale (He et al., 2014) in addition to the Tolerance of Ambiguity scale (Herman et al., 2010). Then, in the final blocks, participants rated the job candidate's vocal attractiveness and vocal friendliness, similar to Study 1, and rated the job candidate's physical attractiveness, like in Experiment 1.

Statistical analysis

Like Study 2 Experiment 1, conditions were combined based on vocal affective presentation style (i.e., pride conditions: Group 1 faces-pride utterances, Group 2 faces-pride utterances; shame conditions: Group 1 faces-shame utterances, Group 2 faces-shame utterances). The following analyses involved 54 participants in the vocal pride condition (26 males, 28 females) and 46 participants in the vocal shame condition (19 males, 27 females). Also, the same statistical analysis approach was used as Study 2 Experiment 1, with the analyses performed and figures

created in RStudio (version 2022.7.1.554) (R Studio, 2022) (R Version 4.2.1, https://www.R-project.org) (R Core Team, 2022). For the linear mixed effects model, speaker ethnicity and vocal affective presentation style were effects (or deviation) coded, coding each level as 1 (pride and East Asian) and -1 (shame and European). Other numerical predictors were rescaled (between 0 and 1) and centered, so the intercepts in the presented models can be interpreted in reference to the lowest rating for that variable (e.g., a rating of not at all attractive for vocal attractiveness).

Cronbach's alpha (α) was calculated for each of the perceptual rating scales to assess their level of internal consistency. High internal consistency means that the average inter-item correlation between the visual or speech stimuli for a given measure is high.

For the visual stimuli, there was high internal consistency on the competence scale (pride conditions: α = 0.946, 95% CI [0.891, 0.975]; α= 0.941, 95% CI [0.884, 0.968], shame conditions: α = 0.961, 95% CI [0.907, 0.986]; α= 0.954, 95% CI [0.836, 0.984]) and hirability scale (pride conditions: α = 0.935, 95% CI [0.884, 0.965];α = 0.925, 95% CI [0.851, 0.959], shame condition: α = 0.963, 95% CI [0.912, 0.986]; α= 0.957, 95% CI [0.856, 0.984]), and acceptable internal consistency for the physical attractiveness scale (group 1 faces: α = 0.768, 95% CI [0.515, 0.881], group 2 faces: α = 0.807, 95% CI [0.595, 0.895]). For the speech stimuli, the internal consistency was high for the competence scale (pride conditions: α = 0.868, 95% CI [0.695, 0.928];α = 0.798, 95% CI[0.619, 0.882] , shame conditions: α = 0.961, 95% CI [0.898, 0.987]; α = 0.881, 95% CI [0.749, 0.937]) and high to acceptable for the hirability scale (pride conditions: α = 0.828, 95% CI [0.652, 0.901]; α = 0.756, 95% CI [0.548, 0.847], shame conditions: α = 0.963, 95% CI [0.904, 0.988]; α = 0.887, 95% CI [0.810, 0.932]). It was acceptable for the accent similarity scale (pride conditions: α = 0.759, 95% CI [0.512, 0.854]; α = 0.761, 95% CI [0.523, 0.858], shame conditions: α = 0.812, 95% CI [0.503, 0.905]; α = 0.753, 95% CI [0.575, 0.831]), the vocal attractiveness scale

(pride conditions: α = 0.642, 95% CI [0.163, 0.804]; α = 0.742, 95% CI [0.39, 0.87], shame conditions: α = 0.862, 95% CI [0.706, 0.920]; α = 0.776, 95% CI [0.610, 0.861]), and low to acceptable for the vocal friendliness scale (pride conditions: α = 0.44, 95% CI [-0.10, 0.69]; α = 0.464, 95% CI [-0.003, 0.722], shame conditions: α = 0.828, 95% CI [0.64, 0.90], α = 0.518, 95% CI [-0.002, 0.739]. Thus, the ratings for each measure were averaged across participants and used as dependent measures or predictors in the following analyses.

3.3.3 Results

Initial analyses

Competence and hirability ratings in Block 1 and 2

In Block 1, across conditions, faces of East Asian heritage (M = 6.06, SD = 1.47) were rated as marginally more competent than faces of European heritage (M = 5.85, SD = 1.49), t (797.99) = 1.94, 95% CI [-0.002, 0.41], p = 0.05. Also, faces of East Asian heritage (M = 5.98, SD = 1.50) received higher hirability ratings than faces of European heritage (M = 5.72, SD = 1.53), t (797.6) = 2.43, p = 0.02, 95% CI [0.05, 0.47]. Thus, all speakers were perceived to be moderately competent and hirable based only on their face, with East Asian candidates receiving higher ratings.

In Block 2, speakers of East Asian heritage with Singaporean English accents were rated as comparably competent (F (1,796) = 0.32, p = 0.57, and comparably hirable (F (1,796) = 0.74, p = 0.39, to speakers of European heritage with American English accents. Speakers expressing vocal pride were rated as more competent (F (1, 796) = 46.76, p < .001, η_p^2 = 0.06, 95% CI for η_p^2 [0.03, 0.09]), and more hirable (F (1,796) = 38.73, p < .001, η_p^2 = 0.05, 95% CI for η_p^2 [0.02, 0.08]), than those expressing vocal shame. These latter results demonstrate that our manipulation of vocal affective presentation style affected the competence and hirability ratings as predicted. There was a significant interaction between speaker vocal affective presentation style and ethnicity

for both perceived competence and hirability. When expressing vocal pride, speakers of European heritage with American English accents and East Asian heritage with Singaporean English accents were rated comparably competent (mean difference = 0.43, p = 0.06, 95% CI [-0.01, 0.86]) and comparably hirable (mean difference = 0.41, p = 0.09, 95% CI [-0.05, 0.87]). In contrast when speakers expressed vocal shame, speakers of East Asian heritage with Singaporean English accents were rated as more competent (mean difference = 0.65, p = 0.002, 95% CI [1.12, 0.18]) and more hirable (mean difference = 0.73, p = 0.001, 95% CI [1.22, 0.23] than speakers of European heritage with American English accents.

The competence and hirability ratings were highly correlated across both blocks, $\rho = 0.94$, $p < .001$, and within each block (Block 1: $\rho = 0.94$, $p < .001$; Block 2: $\rho = 0.94$, $p < .001$). In other words, the more competent speakers were rated for the human resources manager position, the more likely they were rated to be hired for the position. Thus, we summed the competence and hirability ratings (now out of 18) and subsequently refer to this composite rating as, "*Job suitability*", our primary outcome variable. See Table 3.4 for these descriptive statistics.

Table 3.4. *Job suitability ratings based on speaker accent and vocal affective presentation style*

Speaker accent	Job suitability			
	Block 1: Visual Only		Block 2: Audio and Visual	
	M	*SD*	*M*	*SD*
	Pride			
American English	11.55	2.97	11.56	3.67
Singaporean English	12.02	2.93	10.72	3.47
	Shame			
American English	11.60	2.99	8.79	3.81
Singaporean English	12.05	3.00	10.17	3.11

Social perception of speaker's voices

For perceived accent similarity, there was a marginal effect of vocal affective presentation style, F (1, 796) = 4.34, p = 0.04, η_p^2 = 0, 95% CI for η_p^2 [0.00, 0.02]), with speakers expressing vocal pride receiving marginally higher accent similarity ratings than speakers expressing vocal shame. There was a main effect of speaker accent, (F (1,796) = 1692.96, p < .001, η_p^2 = 0.68, 95% CI for η_p^2 [0.58, 0.71]), speakers with American English accents were rated as having more similar accents to participants than speakers with Singaporean English accents. There was also a significant interaction between speaker vocal affective presentation style and accent, (F (1,796) = 8.54, p = 0.004, η_p^2 = 0.01, 95% CI for η_p^2 [0.00, 0.03]). Speakers with American English accents received similar accent similarity ratings when they expressed vocal pride or shame (mean difference = 0.10, p = 0.93). In contrast, speakers with Singaporean English accents were rated as having more similar accents to participants when they expressed vocal pride compared to shame (mean difference = -0.60, p = 0.002). This result served as a manipulation check that participants perceived the American English accent as more of an in-group accent compared to the Singaporean English accent.

For perceived vocal attractiveness, there was a significant effect of speaker accent (F (1,796) = 18.77, p < .001, η_p^2 = 0.02, 95% CI for η_p^2 [0.01, 0.05]), speakers with American English accents were rated as having more attractive voices than speakers with Singaporean English accents. There was a significant effect of vocal affective presentation style, (F (1,796) = 37.07, p < .001, η_p^2 = 0.04, 95% CI for η_p^2 [0.02, 0.08]), with speakers expressing vocal pride rated as sounding more attractive than those expressing vocal shame. There was a significant interaction between speaker vocal affective presentation style and speaker accent, (F (1,796) = 31.37, p < .001, η_p^2 = 0.04, 95% CI for η_p^2 [0.02, 0.07]). When speakers expressed vocal pride, speakers with

American English accents were perceived as more vocally attractive than speakers with Singaporean English accents (mean difference = 1.34, $p < .001$). When speakers expressed vocal shame, speakers with American English accents were perceived as sounding comparably attractive to speakers with Singaporean English accents (mean difference = -0.24, $p = 0.64$).

For perceived vocal friendliness, there was a significant effect of speaker accent, ($F(1,796) = 19.77$, $p < .001$, $\eta_p^2 = 0.02$, 95% CI for η_p^2 [0.01, 0.05]), speakers with American English accents were rated as having less friendly voices than speakers with Singaporean English accents. There was a significant effect of vocal affective presentation style, ($F(1,796) = 147.19$, $p < .001$, $\eta_p^2 = 0.16$, 95% CI for η_p^2 [0.11, 0.20]), with speakers expressing vocal pride rated as sounding more friendly than those expressing vocal shame. There was a significant interaction between speaker vocal affective presentation style and accent/ethnicity, ($F(1,796) = 15.82$, $p < .001$, $\eta_p^2 = 0.02$, 95% CI for η_p^2 [0.01, 0.04]). When speakers expressed vocal pride, speakers with American English accents were perceived as sounding comparably friendly to speakers with Singaporean English accents, mean difference = -0.10, $p = 0.94$. When speakers expressed shame, speakers with Singaporean English accents were perceived as sounding more friendly than speakers with American English accents (mean difference = -1.10, $p < .001$).

Then for perceived physical attractiveness, speakers of East Asian ($M = 5.13$, $SD = 1.66$) and European ($M = 5.15$, $SD = 1.67$) heritage were comparable in physical attractiveness, $t(797.91) = -0.15$, $p = 0.88$, 95% CI [-0.25, 0.21]. Based on these mean ratings, participants perceived speakers' faces to be moderately attractive.

In summary, when speakers had an American English accent, participants were more likely to perceive their accent as similar to their own, when they were expressing vocal pride or shame, compared to when speakers had a Singaporean English accent. Then in terms of perceived social

qualities, when speakers expressed vocal pride speakers with American English accents were perceived to sound more attractive and comparably friendly to speakers with Singaporean English accents. In contrast, when speakers expressed vocal shame participants perceived speakers with American English accents were perceived to sound comparably attractive and less friendly than speakers with Singaporean English accents. Also, speakers of European heritage and East Asian heritage received comparable physical attractiveness ratings. See Table 3.5 for descriptive statistics of the social perceptual ratings.

Table 3.5. *Social perceptual ratings based on speaker accent and vocal affective presentation style*

Speaker accent	Accent similarity		Vocal attractiveness		Vocal friendliness	
	M	*SD*	*M*	*SD*	*M*	*SD*
			Pride			
American English	6.89	2.14	5.88	2.02	6.22	1.83
Singaporean English	2.31	1.47	4.54	1.93	6.32	1.56
			Shame			
American English	6.99	1.76	4.23	2.23	4.20	2.08
Singaporean English	1.72	1.13	4.47	1.75	5.29	1.60

Job suitability ratings at Block 1 – Visual only

A linear mixed effects model was used to determine how speakers' perceived job suitability varied based on their ethnicity. The fixed effects included speaker ethnicity (European, East Asian). Speaker sex, participant sex, physical attractiveness ratings, the two-way interaction between physical attractiveness ratings and speaker sex and participant tolerance of ambiguity total scores were included as control variables. The model also included a random intercept by participant with a correlated random slope to control for variability among participants as a

function of the effect of physical attractiveness ratings. Vocal affective presentation style was not included as a fixed effect because at Block 1 all participants were not exposed to speech cues.

The model had an adjusted R^2 of 0.76 in accounting for variation in job suitability ratings (0.63 from random effects, 0.13 from fixed effects). The model returned a significant intercept (B = 11.45, SE = 0.73, t (136.18) = 15.70, p < .001, 95% CI [10.02, 12.90], with the by-participant random intercept contributing SD = 3.11. There was an effect of speaker ethnicity, B = 0.23, SE = 0.05, t (700.39) = 4.15, p < .001, 95% CI [0.12, 0.33], with speakers of East Asian heritage receiving higher job suitability ratings. Participants gave higher job suitability ratings for speakers they perceived to look more attractive, B = 3.24, SE = 0.51, t (88.12) = 6.40, p < .001, 95% CI [2.23, 4.24], controlling for the above-mentioned fixed and random effects, (by-subject random slope of physical attractiveness contributing SD = 3.41). There was no effect of speaker sex, B = 0.23, SE = 0.15, t (691.41) = 1.50, p = 0.14, 95% CI [-0.07, 0.53]. There was also a significant effect of participants' tolerance of ambiguity scores, B = -3.01, SE = 1.38, t (100.98) = -2.17, p = 0.03, 95% CI [-5.75, -0.26], where the higher participants' scores (i.e., the more tolerant they are of ambiguity and uncertainty), the lower job suitability ratings given. No other effects were significant.

Change in job suitability ratings

To investigate the extent to which hearing a speaker's vocal affective presentation style changed the job suitability ratings for speakers with Singaporean English accents/East Asian heritage or with American English accents/European heritage, we calculated the difference in job suitability ratings for each speaker's face between Block 1 (Visual only) and Block 2 (Audio and Visual), the same as Study 2 Experiment 1. As seen from Table 3.5, on average speakers experienced a decrease in job suitability ratings in Block 2 compared to Block 1.

The difference in job suitability ratings was not affected by speaker accent/ethnicity, F (1, 796) = 1.35, p = 0.25. The difference in job suitability ratings was significantly affected by speaker vocal affective presentation style, F (1, 796) = 48.79, p = <.001, η_p^2 = 0.06, 95% CI for η_p^2 [0.03, 0.09]), with speakers expressing vocal shame experiencing a greater decrease in job suitability ratings. There was a significant interaction between speaker accent/ethnicity and vocal affective presentation style, F (1, 796) = 21.08, p < .001, η_p^2 = 0.03, 95% CI for η_p^2 [0.01, 0.05]. When speakers expressed vocal pride, speakers with Singaporean English accents/East Asian heritage showed a greater decrease in job suitability ratings than speakers with American English accents/European heritage (mean difference = 1.31, p < .001). When speakers expressed vocal shame, speakers with American English accents/European heritage showed a marginally greater decrease in job suitability ratings than speakers with Singaporean English accents/East Asian heritage (mean difference = -0.93, p = 0.048).

Job suitability ratings at Block 2 – Audio and visual

A linear mixed effects model was used to determine how speakers' perceived job suitability varied based on their accent/ethnicity and vocal affective presentation style. The fixed effects included speaker accent/ethnicity (American English/European, Singaporean English/East Asian), vocal affective presentation style (pride, shame) and their interaction. Vocal attractiveness ratings, vocal friendliness ratings and their two-way and three-way interactions with speaker accent/ethnicity and vocal affective presentation style were included as control variables. Also, speaker sex, participant sex, physical attractiveness ratings, and participants' scores on the accent belief scale, and tolerance of ambiguity scale, were included as control variables. The model also included a random intercept by-participant and by-speaker.

The model had an adjusted R^2 of 0.66 in accounting for variation in job suitability ratings (0.31 from random effects, 0.35 from fixed effects). The model returned a significant intercept (B = 6.43, SE = 1.03, t (136.04) = 6.66, p < .001, 95% CI [4.40, 8.47], with the by-participant random intercept contributing SD = 1.94 and the by-speaker random intercept contributing SD = 0.62. Participants gave higher job suitability ratings if the speaker expressed vocal pride compared to shame, B = 0.68, SE = 0.32, t (386.40) = 2.16, p = 0.03, 95% CI [0.06, 1.30]. Like Block 1, participants gave higher job suitability ratings for candidates they perceived to be more physically attractive, B = 0.95, SE = 0.47, t (765.65) = 2.05, p = 0.04, 95% CI [0.04, 1.87]. There was a significant effect of speaker accent/ethnicity, B = 1.09, SE = 0.33, t (30.81) = 3.29, p = 0.003, 95% CI [0.41, 1.76]. Participants gave higher job suitability ratings for speakers they perceived to sound more attractive, B = 4.13, SE = 0.48, t (765.47) = 8.52, p < .001, 95% CI [3.18, 5.08], and more friendly, B = 3.56, SE = 0.52, t (743.41) = 6.78, p < .001, 95% CI [2.53, 4.59]. There was also a significant effect of participants' accent belief diagnosticity scores, B = -2.12, SE = 0.98, t (99.39) = -2.15, p = 0.03, 95% CI [-4.07, -0.17], meaning the more participants believe that they can infer other traits about a speaker based on their accent, the lower their job suitability ratings. Unlike Experiment 1, there was no effect of speaker sex, B = -0.30, SE = 0.24, t (8.15) = -1.29, p = 0.23, 95% CI [-0.82, 0.22], and no significant interaction between speaker accent/ethnicity and vocal affective presentation style, B = 0.08, SE = 0.24, t (727.18) = 0.32, p = 0.75, 95% CI [-0.39, 0.54].

Like Experiment 1, there was a significant interaction between speaker vocal affective presentation style and vocal attractiveness, B = 1.60, SE = 0.49, t (772.78) = 3.26, p = 0.001, 95% CI [0.64, 2.57] (Figure 3.4). If speakers expressed vocal pride, there was a steeper positive slope between vocal attractiveness and perceived job suitability. In contrast, when speakers expressed

vocal shame, the slope was shallower indicating that sounding more attractive did not result in as great of a change in job suitability ratings compared to expressing vocal pride.

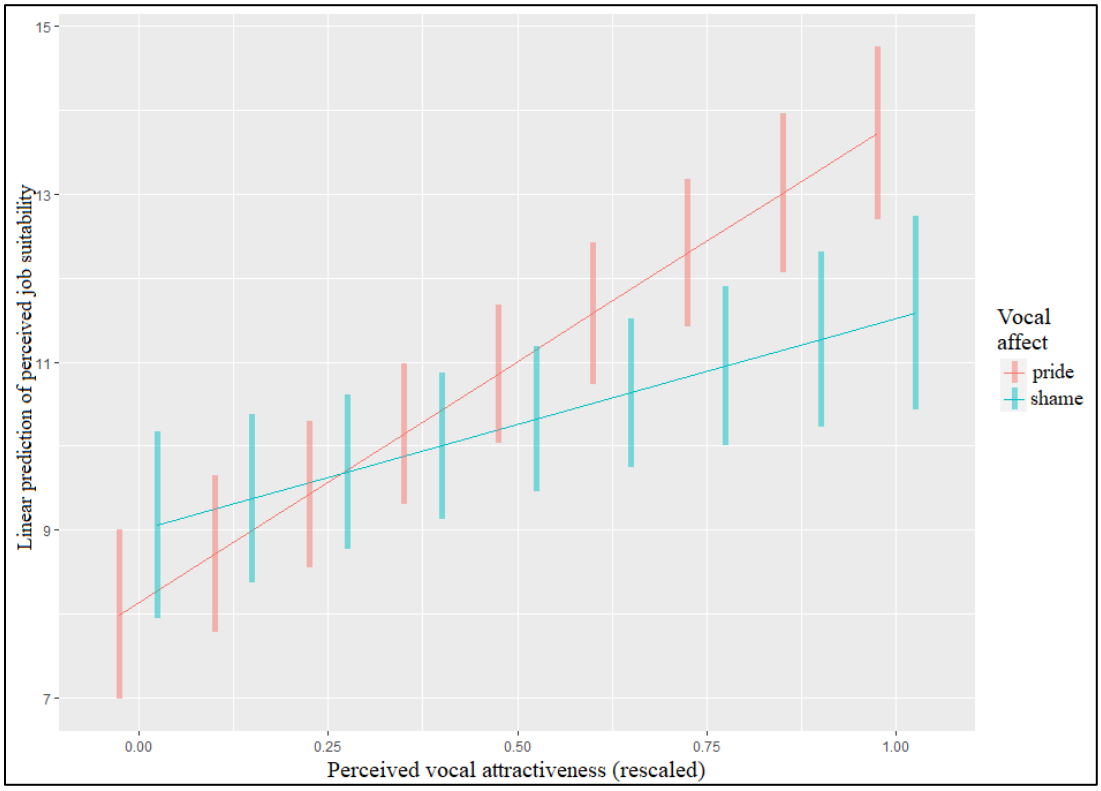

Figure 3.4. Interaction between speaker vocal affective presentation style and vocal attractiveness for average job suitability (out of 18) based on model prediction. The vertical bars indicate a 95% confidence interval.

There was a significant interaction between speaker accent/ethnicity and vocal friendliness, B = -1.32, SE = 0.51, t (737.84) = -2.61, p = 0.01, 95% CI [-2.32, -0.32] (Figure 3.5). If speakers had an American English accent/European heritage, there was a steeper slope, meaning the more friendly they were perceived to sound, the higher their job suitability ratings. In contrast, if speakers had a Singaporean English accent/East Asian heritage, the slope was shallower, meaning that with increased vocal friendliness, their perceived job suitability did not proportionally increase as much as speakers with American English accents/European heritage.

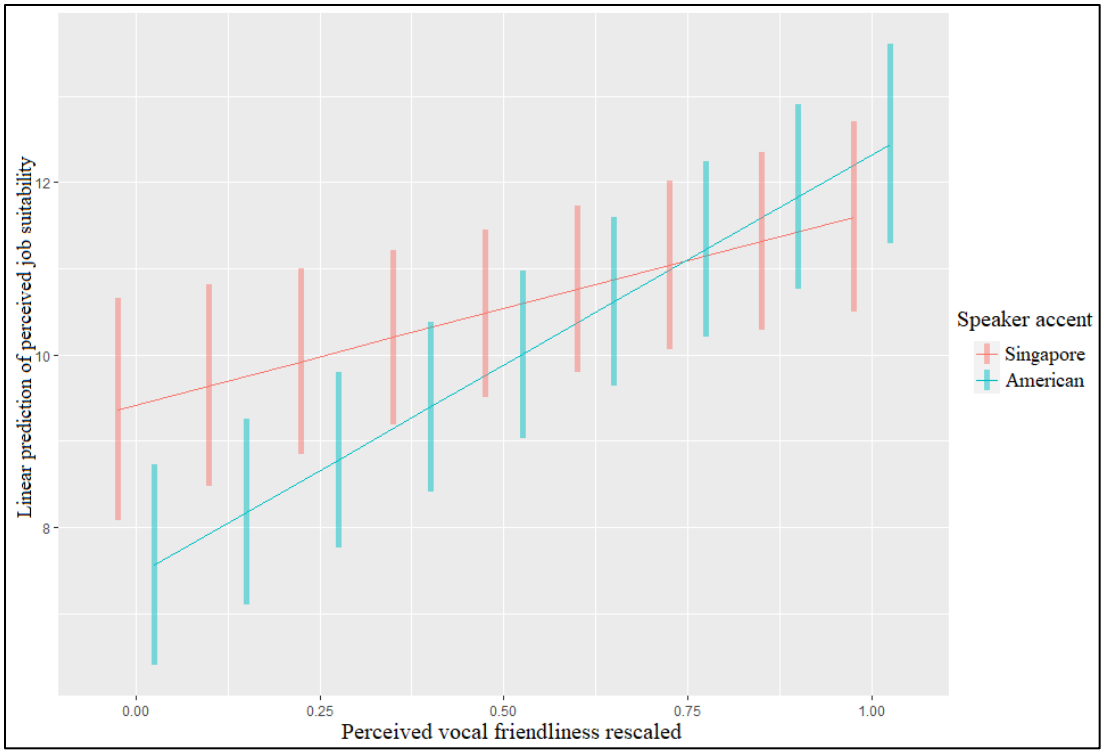

Figure 3.5. Interaction between speaker accent and perceived vocal friendliness average job suitability (out of 18) based on model prediction. The vertical bars indicate a 95% confidence interval.

There was also a significant interaction between speaker vocal affective presentation style and vocal friendliness, B = -2.12, SE = 0.50, t (742.91) = -4.22, p < .001, 95% CI [-3.11, -1.14]. The pairwise comparisons revealed that the least friendly speakers' voices were perceived to sound, speakers expressing vocal pride received higher job suitability ratings than speakers expressing vocal shame (rating of 1: t (533) = 4.01, SE = 0.72, p = 0.01, rating of 2: t (384) = 3.77, SE = 0.63, p = 0.02). There was no significant difference in speakers' job suitability ratings with higher vocal friendliness ratings (rating of 3: t (249) = 3.37, SE = 0.54, p = 0.08, rating of 4: t (161) = 2.70, SE = 0.48, p = 0.39, rating of 5: t (120) = 1.72, SE = 0.45, p = 0.96, rating of 6: t (120) = 0.54, SE = 0.45, p = 1.00, rating of 7: t (161) = -0.61, SE = 0.48, p = 1.00, rating of 8: t (249) = -

1.51, SE = 0.54, p = 0.99, rating of 9: t (384) = -2.16, SE = 0.63, p = 0.78). No other effects were significant.

Selecting the best job candidate

Participants selected one candidate for the human resources manager position. Participants most often selected European candidates with American English accents as the best candidate (62%). There was no significant association between the selected candidate's accent and vocal affective presentation style, $\chi 2$ (1, N = 100) = 0.70, p = 0.40, or the candidate's sex, $\chi 2$ (1, N = 100) = 0.61, p = 0.43.

Effect of East Asian candidates' accent: Experiment 1 and Experiment 2 Analyses

Analyses were conducted between Experiment 1 (in-group accent only) and Experiment 2 (in-group and out-group accent) Block 2 job suitability ratings to determine the effect of East Asian job candidates' accent in a hiring context. The job suitability ratings of East Asian speakers paired with a Canadian English accent (Experiment 1) versus a Singaporean English accent (Experiment 2) were compared, for each vocal competence level (high: confidence, pride; low: doubt shame). As previously found, speakers who conveyed high vocal competence received higher job suitability ratings than those conveying low vocal competence, F (1, 698) = 35.52, p < .001, η_p^2 = 0.05, 95% CI for η_p^2 [0.02, 0.08]) (Figure 3.6). This effect was further moderated by speaker's accent, F (1, 698) = 18.98, p < .001, η_p^2 = 0.03, 95% CI for η_p^2 [0.01, 0.05]). When East Asian job candidates conveyed high vocal competence, speakers with Canadian English accents received higher job suitability ratings than speakers with Singaporean English accents (mean difference = 1.30, p = 0.001, 95% CI [0.40, 2.20]). When East Asian candidates conveyed low vocal competence, speakers with Singaporean English accents received marginally higher job suitability ratings (mean difference = 0.94, p = 0.06, 95% CI [-0.03, 1.91]). Also, when East Asian job

candidates had a Canadian English accent, they received higher job suitability ratings when they expressed high compared to low vocal competence (mean difference = 2.80, $p < .001$, 95% CI [1.80, 3.80]. When East Asian job candidates had a Singaporean English accent, they received comparable job suitability ratings across the vocal competence levels (mean difference = 0.55, $p = 0.36$).

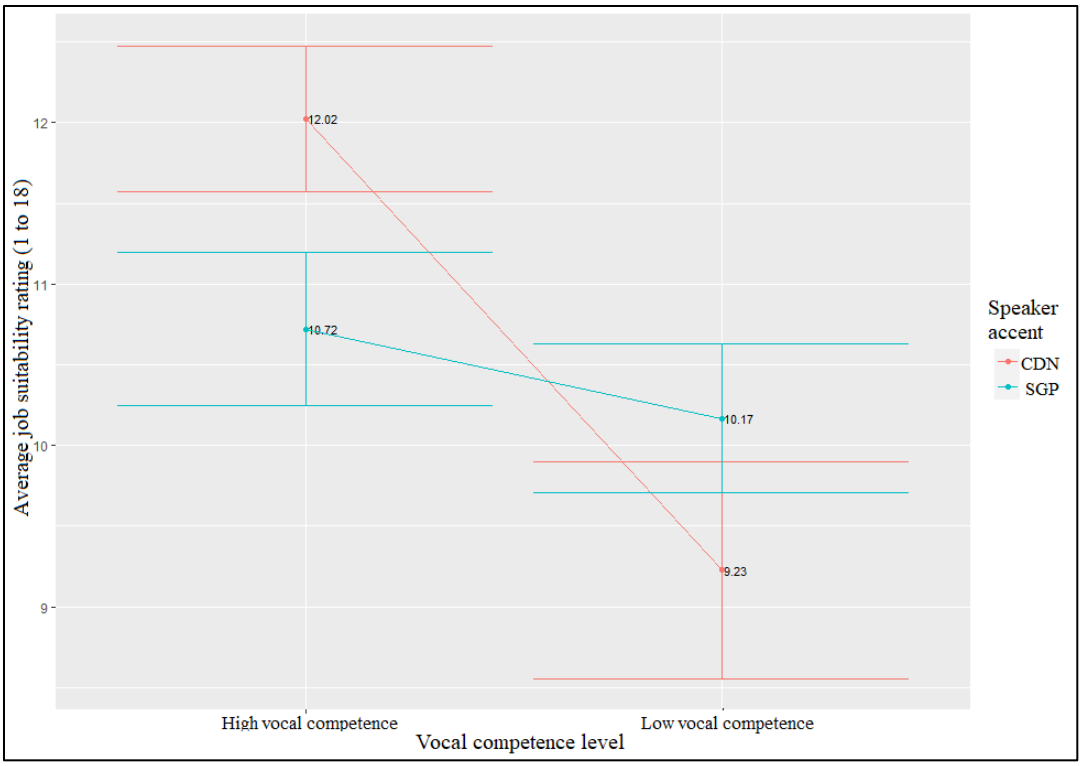

Figure 3.6. Mean job suitability ratings for East Asian job candidates in Experiment 1 and 2 expressing high vocal competence (pride, confidence) or low vocal competence (shame, doubt)

Ethnic appearance effect for candidates with non-standard accents: Study 1 and 2 analyses

Analyses were conducted between the job suitability ratings from Study 1 (audio only) and Block 2 of this experiment (audio and visual) to determine the effect of ethnic appearance for speakers with standard and non-standard accents. Analyses only included the speech stimuli used in both experiments to streamline results. Speakers received higher ratings when participants were exposed to their ethnic appearance and accent, compared to only their accent, $F(1, 1752) = 21.92$,

$p < .001$, $\eta_p^2 = 0.01$, 95% CI for η_p^2 [0.00, 0.02]) (Figure 3.7). As reported in other models, speakers received higher ratings if they expressed vocal pride compared to shame, $F (1, 1752) = 272.32$, $p < .001$, $\eta_p^2 = 0.13$, 95% CI for η_p^2 [0.11, 0.16]), which was further moderated by speakers' accent, $F (1, 1752) = 59.61$, $p < .001$, $\eta_p^2 = 0.03$, 95% CI for η_p^2 [0.02, 0.05]). Of interest, there was no significant effect of speaker accent, $F (1, 1752) = 1.09$, $p = 0.30$, or an interaction between speaker accent and communicative context, $F (1, 1752) = 0.001$, $p = 0.98$. In other words, the perceived job suitability of speakers with standard or non-standard accents did not differ when participants only heard their voice or were presented with their voice with their ethnic appearance. There was also a significant interaction between communicative context and vocal affective presentation style, regardless of speaker accent, $F (1, 1752) = 37.41$, $p < .001$, $\eta_p^2 = 0.02$, 95% CI for η_p^2 [0.01, 0.04]). When speakers expressed vocal pride, they received comparable job suitability ratings across both communicative contexts (mean difference = -0.31, $p = 0.55$). When speakers expressed vocal shame, they received higher job suitability ratings in the audio and visual context compared to the audio only context (mean difference = 1.78, $p < .001$). Also, the mean difference in job suitability ratings between vocal pride and shame was over two times greater in the audio only condition (mean difference = 3.75, $p < .001$) compared to the audio and visual condition (mean difference = 1.66, $p < .001$).

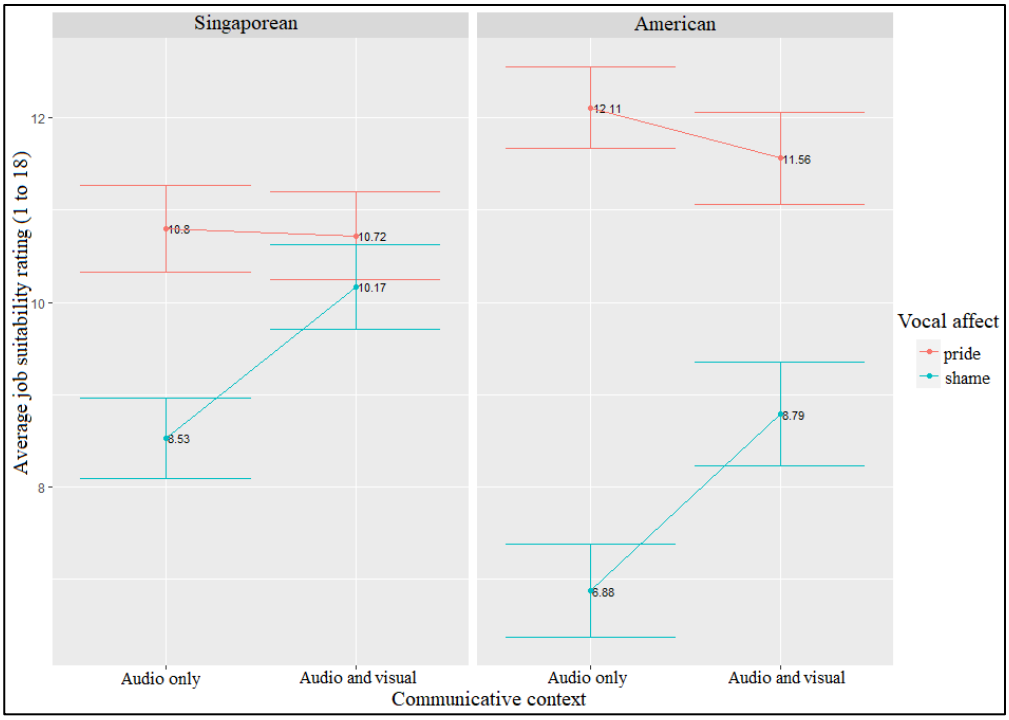

Figure 3.7. Mean job suitability ratings for job candidates with standard (American English) and non-standard (Singaporean English) accents in Study 1 (audio only) and Study 2 Experiment 2 Block 2 (audio and visual), expressing vocal pride or shame.

3.3.4 Discussion

This experiment aimed to better understand the impact of a speaker's ethnic appearance on the job suitability impressions of speakers with standard and non-standard accents expressing a vocal competence level. Specifically, we examined the interactive effects of a speaker's accent/ ethnic appearance (American English accent/European, Singaporean English accent/East Asian) and vocal affective presentation style (pride, shame) on their perceived job suitability.

Processing a speaker's ethnic appearance

When perceivers only saw a speaker's ethnic appearance, there was an effect of speaker ethnicity on their perceived job suitability, as seen in Experiment 1. Job candidates of East Asian heritage were rated as more suitable for the human resources manager position based on their appearance only. This result may suggest that participants activated associated concepts or cultural

stereotypes about people of East Asian heritage. Also, like Experiment 1, job candidates perceived physical attractiveness had the largest impact on their job suitability ratings, providing further support for the effect of physical attractiveness in a hiring context (Dion et al., 1972; Eagly et al., 1991; Hosoda, Stone-Romero, & Coats, 2003), and specifically an interethnic hiring context involving visible ethnic minorities. Unlike Experiment 1, we also measured participants' social desirability (He et al., 2014) to potentially explain their job suitability ratings; however, it was not a significant predictor. The questionnaire may not have been sensitive enough for our experimental context, as it measures participants' tendency to minimize their general explicit negative bias (He et al., 2014), not specifically their (negative) bias towards speakers with non-standard accents or people who are visible ethnic minorities. Another possibility is that participants' implicit perceptions of speakers of East Asian heritage would also need to be measured to further understand participants' impressions when only exposed to a speaker's ethnic appearance. There is evidence that participants implicitly perceive members of their own racial group more positively than other racial groups based on their ethnic appearance, with people of Asian heritage perceived less positively compared to White people (Axt, Ebersole, & Nosek, 2014).

Processing a speaker's ethnic appearance, accent, and vocal affective presentation style

Compared to when participants evaluated speaker's job suitability based on their face only, the initial positive impression of speakers of East Asian heritage was reversed in Block 2. Hearing speakers of East Asian heritage with Singaporean English accents resulted in lower job suitability ratings than speakers of European heritage with American English accents. Additionally, when speakers expressed vocal pride, speakers of East Asian heritage with Singaporean English accents experienced a greater change in their perceived job suitability from Block 1 compared to speakers of European heritage (with American English accents). These results demonstrate that integrating

a speaker's vocal cues with their physical appearance can greatly shift their perceived job suitability. The inferences drawn from a speaker's voice, particularly their accent, carries a lot of social weight on competence impressions. This may occur when speakers express high vocal competence because perceivers can process the speech with greater fluency (Alter & Oppenheimer, 2009), allowing them to draw more social inferences from the voice. These changes in job suitability ratings also suggests that the impressions we form of speakers depends on the type of ethnicity or cultural cues presented. For example, while people of Chinese heritage may be associated with high competence stereotypes (e.g., Fiske et al., 2002; Kil, Noels, Lascano, & Schweickart, 2019; Leong & Hayes, 1990), these findings are based on people being prompted by words. When participants are exposed to a speaker's voice and ethnic appearance, speakers with Chinese English accents are perceived as having low competence (Baquiran & Nicoladis, 2020; Houshmand, Spanierman, & Tafarodi, 2014). Thus, it may be beneficial for interviewers to assess job candidates through various communication channels to form a comprehensive impression of their competence.

We then examined how a speaker's vocal affective presentation style and accent interacted with their ethnic appearance to affect their perceived job suitability. There was an effect of speaker vocal affective presentation style, supporting our hypothesis. Speakers expressing vocal pride received higher job suitability ratings than those expressing vocal shame, indicating that to some extent, participants accurately perceived speaker's vocal competence level, regardless of a speaker's accent. Of interest was the potential interaction between speaker vocal affective presentation style and accent/ethnic appearance. Unlike Experiment 1 and Study 1, this interaction was not observed. This result contrasts with Bradac and Wisegarver (1984), Jiang, Sanford, and Pell (2018), and Jiang, Keenan-Gossack, and Pell (2020), where a speaker's vocal competence

level and accent moderated their perceived competence. Then again, in Experiment 1 and Study 1, this interaction had one of the smallest effects on changes to speaker's perceived job suitability, so it may not be as critical to explaining participants' Block 2 ratings.

Rather, speaker's perceived vocal attractiveness followed by their perceived vocal friendliness had the largest impact on their perceived job suitability, as previously reported. The effect of these social attributes was further moderated by speaker's vocal affective presentation style or accent. These findings provide more evidence for the impact of job interviewers' affective reactions of job candidates (e.g., Howard & Ferris, 1996; Rivera, 2015; Young & Kacmar, 1998). It also provides evidence that a job candidate's vocal competence level can affect how friendly and attractive their voice sounds (e.g., Feiler & Powell, 2016), which can subsequently affect job interviewers social interest in them. We also found that a job candidate's accent/ethnic appearance and perceived vocal friendliness impacted their job suitability. Speakers with American English accents and European appearances showed a steeper slope, meaning the more friendly they were perceived to sound the higher their job suitability. In contrast, speakers with Singaporean English accents showed a shallower slope, meaning that sounding more friendly did not result in similar increases in job suitability. This different inferential processing of speech in a non-standard accent is in line with the findings in Study 1, where participants had greater difficulty decoding the vocal cues of speakers with non-standard accents, impacting their social inferences. Moreover, participants with higher accent diagnosticity scores (i.e., the more they believe that they can infer other traits about a speaker based on their accent) gave lower job suitability ratings. Taken together, sounding friendly and attractive can benefit the perceived job suitability of all job candidates. However, for speakers with non-standard accents, this benefit can be determined by interviewer's attitude towards/experience with their accent and vocal competence.

Impact of East Asian candidates' accent and vocal competence level

We explored the effect of a candidate's ethnic appearance as a visible ethnic minority in combination with their accent and vocal competence level by examining the difference in job suitability ratings for East Asian candidates in Experiment 1 and this experiment. Again, a speaker's accent played a prominent role on their perceived job suitability. When speakers conveyed high vocal competence, East Asian candidates with standard accents received higher job suitability ratings than East Asian candidates with non-standard accents. When speakers conveyed low vocal competence, East Asian candidates with standard accents received lower job suitability ratings than East Asian candidates with non-standard accents. These results are consistent with Study 1 and Jiang and colleagues (2018; 2020), where participants were only exposed to speech, demonstrating that a speaker's vocal cues (accent and vocal competence level) carry greater social weight on competence-based impressions than their ethnic appearance. Similar results have also been observed in Baquiran and Nicoladis (2020), where participants (Caucasian and Chinese Canadians) were saw an image of a Chinese doctor paired with a voice saying good or bad medical news in a Canadian English or Chinese English accent. The Chinese doctor with a Canadian English accent was rated as more competent than with a Chinese English accent regardless of participant ethnicity. Thus, a speaker's accent is a stronger cue for activating in-group membership compared to ethnic appearance. However, more research is needed to understand the impact of a candidate's ethnic appearance in an interethnic hiring context involving East Asian candidates with standard and non-standard accents as well job candidates of other races/ethnicities.

Additive effect of candidate's ethnic appearance

The effect of communicative context was explored to determine the additive effect of job candidate's ethnic appearance, with their accent and vocal competence level. Job candidates,

regardless of speaker accent, received higher job suitability ratings in Block 2 of this experiment, compared to when it was assessed via speaker's accent only (Study 1). This may suggest that a speaker's physical appearance (ethnic appearance and attractiveness) positively contributes to their perceived job suitability, as it provides an additional source of cues to draw social inferences from (Todorov, Olivola, Dotsch, & Mende-Siedlecki, 2015). In our case, a speaker's physical appearance was particularly beneficial to speaker's perceived job suitability if they expressed low vocal competence. When speakers expressed vocal shame, they received higher job suitability ratings in the audio and visual context, compared to the audio only context, while they received comparable ratings in these contexts when they expressed vocal pride. This may have occurred because when speakers expressed low vocal competence participants engaged in mentalizing processes to try to understand speaker's communicative intention (e.g., Jiang & Pell, 2015; Kuhlen et al., 2015), and their physical appearance provided positive cues related to their job suitability. However, when speakers expressed high vocal competence, it is unclear why their vocal cues carried more weight than their ethnic appearance. More research is needed to understand the cognitive and social processes involved when a speaker conveys high vocal competence in combination with other vocal and visual cues. Also, results may have differed if participants were presented videos of job candidate's vocally expressing low competence. By participants viewing static images of faces, dynamic visual cues (e.g., changes in eye gaze, a thinking face, shifts in posture) of low competence were absent, which can amplify participants' potentially negative judgments (DeGroot & Motowidlo, 1999; Gallois, Callan & Palmer, 1992). At the same time, this methodological decision allowed us to focus on a speaker's vocal cues, and their interaction with static ethnic appearance features.

Unlike Hansen and colleagues (2018) we did not observe an interaction between speaker accent and communicative context on their perceived competence. They found that Turkish looking candidates with Turkish German accents were perceived as comparably competent to when participants only saw their face. Conversely, German looking candidates with native German accents were perceived as more competent in the audio and visual context (Hansen et al., 2018). Instead, we observed that communicative context similarly affected speakers with standard accents and dominant ethnic appearances and speakers with non-standard accents who were visible ethnic minorities. This difference may indicate that the effect of the communicative context on competence-based impressions may differ with the accent and ethnic groups being compared.

Limitations and Conclusion

One limitation to this experiment is that participants selected the best candidate for the human resource manager position by only seeing the speaker's faces again, due to experimental constraints. As a result, this decision may have been greatly shaped by speaker's ethnic appearance and perceived physical attractiveness, rather than their visual and vocal cues. At the same time, speaker's ethnic appearance was tied to their accent, so in selecting a speaker of European heritage as the top candidate, they may have indicated their preference for a speaker with a standard accent.

Overall, our findings reveal that for speakers with non-standard accents, their ethnic appearance does not amplify negative perceptions related to their accent. Compared to only hearing their accent and vocal competence level, both speakers with standard and non-standard accents were perceived as more suitable for the job with the addition of their ethnic appearance. Although, the social attraction inferences drawn from a speaker's vocal cues more greatly shapes their job suitability, with speaker's accent carrying more social weight than their ethnic appearance. However, the impact of these inferences varies across participants. For job

interviewers assessing job candidates with non-standard accents, improving their experience with/attitude towards these speakers may aid their ability to decode their vocal cues, and potentially fostering their social connections.

Chapter 4

General Discussion

Summary of results

This study aimed to better understand why speakers with non-standard accents are often perceived as less competent than speakers with standard accents, particularly in job interviews. To do this, we investigated the effect of a speaker's tone of voice, specifically their vocal competence level, as it interacts with their accent and other ethnicity-related information to contribute to their perceived competence and hirability (i.e., job suitability). The impact of a speaker's accent and vocal cues is not well understood because previous studies have predominantly elicited speech by having speakers read standard narrative passages or scripted sentences with a constant speech rate and volume (e.g., Dragojevic & Giles, 2014; Fuse, Navichkova, & Alloggio, 2018; Hansen, Rakić, & Steffens, 2017; 2018, Hosoda & Stone-Romero, 2010; Wong & Babel, 2017), resulting in speech with a neutral or flat affect. Although this elicitation procedure has allowed researchers to control the linguistic content between speakers and the amount of variability in their manner of speaking, it has not allowed us to accurately and wholly understand the competence impressions formed of speakers with non-standard accents in real-world contexts. During an interview, job candidates meaningfully produce changes in their prosody to indicate their varying mental and/or affective states. These vocal expressions can subsequently influence the social inferential processing that listeners engage in (Jiang & Pell, 2015; Jiang, Sanford, & Pell, 2018; Jiang, Gossack-Keenan, & Pell, 2020). Thus, this study examined how processing a speaker's tone of voice, in addition to their accent, can shape their perceived competence in a job interview.

In Study 1, we examined the perceived job suitability of speakers with American English and Singaporean English accents expressing a vocal affective presentation style (vocal pride and shame), while controlling for speaker's perceived social attraction (vocal friendliness and

attractiveness). We found that listeners accurately decoded the vocal competence level of speakers with non-standard accents, though with reduced accuracy compared to speech in a standard accent, supporting previous trends (Bradac & Wisegarver, 1984; Jiang, Sanford, & Pell, 2018; Jiang, Keenan-Gossack, & Pell, 2020). This reduced decoding resulted in speakers with non-standard accents showing a smaller difference in their perceived job suitability when they expressed high versus low vocal competence. Though sounding very friendly led to comparable job suitability for speakers with standard and non-standard accents, vocal friendliness was a stronger mediator between speaker vocal affective presentation style and perceived job suitability, for speakers with standard compared to non-standard accents. This impact of perceived social attraction may indicate native listener's greater experience with decoding the vocal cues of speakers with standard (compared to non-standard) accents. Listeners may find it easier to accurately decode vocal friendliness in a standard accent, because of its greater social relevance for their interactions (Bestelmeyer et al., 2015), or their greater perceived similarity with, and subsequent attraction to speakers with standard accents (Byrne, 1971).

We also acoustically analyzed the vocal competence utterances to characterize the cues that may contribute to a speaker's perceived job suitability. Contrary to our hypotheses, we did not find differences between the vocal pride and shame utterances in frequency, amplitude, and spectral measures for speakers with standard and non-standard accents. A couple temporal differences were found when speakers expressed vocal shame, where speakers with American English accents produced longer pauses, and had greater variation in their pause duration compared to speakers with Singaporean English accents.

Then in Study 2, we examined the effect of speaker's ethnic appearance, as it interacts with their accent and vocal competence level to affect their perceived job suitability. Again, we found

that speaker's perceived job suitability was greatly impacted by the perceived social attraction of their voice. Similar to Study 1, the effect of speaker's perceived vocal friendliness was moderated by speaker's accent/ethnic appearance, where the benefit of sounding more friendly on job suitability ratings was smaller for speakers with non-standard accents. Contrary to our predictions, the ethnic appearance of speakers with non-standard accents did not amplify potential negative perceptions related to their accent. For East Asian candidates with standard or non-standard accents, we observed the same interaction between vocal affective presentation style and accent as in Study 1 where participants were only exposed to speaker's voices. This result suggests that a speaker's accent may carry more perceptual weight on their job suitability than their ethnic appearance. Moreover, speakers, regardless of their accent, received higher job suitability ratings when participants were presented with their accent, vocal competence level, and ethnic appearance, compared to only their vocal cues. Thus, a speaker's physical appearance, not just as a marker of their ethnicity, can contribute to their perceived job suitability.

Discussion

Vocal competence and warmth

This research contributes to an ongoing discussion regarding the major dimensions that shape our social impressions of others. Various fields within psychology have proposed that our impressions of others can be characterized by competence (e.g., intelligent, confident, competent) and warmth (e.g., warm, good natured, sincere, friendliness) features (e.g., Fiske, Cuddy, Glick, & Xu, 2002; Fiske et al., 2007). In the accent literature, these dimensions are roughly equivalent to perceptions of status (e.g., intelligence, competence, social class, education) and solidarity (e.g., similarity, attractiveness, trustworthiness) features (Giles & Billings, 2004; Zahn & Hopper, 1985); and research on interpersonal relationships describe these dimensions in terms of agency (e.g., submissive, dominant) and communion (e.g., hostile, friendly) features (Horowitz et al.,

2006; Wiggins, 1979). Across these theories, a person's competence relates to their perceived capability to pursue their goals as an individual and complete tasks (Abele & Wojciszke, 2007; Fiske et al., 2002). In contrast, a person's warmth relates to their perceived intentions (positive or negative) which affect their social interactions with others (Abele & Wojciszke, 2007; Fiske et al., 2002). For example, being friendly affects the interests of others more than oneself. There is evidence that a person's perceived warmth has a stronger effect on our impression of them compared to their perceived competence (e.g., Abele & Wojciszke, 2007; Wojciszke & Abele, 2008; Wojciszke, Bazinska, & Jaworski, 1998), particularly when evaluating strangers (e.g., Wojciszke & Abele, 2008), like job candidates. This dominance of warmth judgments is hypothesized to occur because it is more crucial for us to assess a person's intentions compared to their ability to act on these intentions (Fiske et al., 2007). For job candidates with non-standard accents who may be visible ethnic minorities, this theory may explain why being of high versus low cultural fit (perceived similarity, shared cultural beliefs and values) results in comparable perceived skills and knowledge (Bye et al., 2014; Horverak et al., 2013). Yet, their perceived hirability is associated with having high perceived cultural fit and similarity to a job interviewer (Bye et al., 2014; Deprez-Sims & Morris, 2010; Horverak et al., 2013).

Our results add two insights to this discussion. Firstly, a speaker's perceived vocal warmth can mediate the relationship between their vocal competence level and their perceived job suitability. In other words, interviewers may judge a candidate's capability for a job based on their perceived intention, regardless of their accent. Though studies have found that warmth and competence judgments can be positively correlated (e.g., Rosenberg et al., 1968), when multiple traits/behaviors are combined and people are compared against each other, these dimensions can be negatively correlated (e.g., Judd, Hawkins, Yzerbyt, & Kashima, 2005; Kervyn, Yzerbyt, Judd,

& Nunes, 2009). Our results suggest that our impression of a speaker's vocal warmth can greatly colour our judgment of their job-related capabilities based on a speaker's vocal cues, and this provides evidence for the dominance of vocal warmth perceptions in a hiring context.

Secondly, a speaker's vocal competence level in combination with their accent, can contribute to the perceived vocal warmth of speakers. When speakers produce speech with a neutral tone of voice (or no intended mental or affective state), there are mixed results on the perceived warmth/solidarity of speakers with non-standard accents. Some have found that speakers with non-standard accents are rated as lower in solidarity than speakers with standard accents (Foucart et al., 2020; Fuertes et al., 2012, Dragojevic & Giles, 2016). This may occur because it is more cognitively taxing to process a speaker's non-standard accent, so listeners generate negative affect towards the speaker (Foucart et al., 2020). Conversely, others have found no difference in the perceived solidarity of speakers with non-standard or standard accents (e.g., Callan, Gallois, & Forbes, 1983; Dragojevic et al., 2017; Lindemann, 2003), or that speakers with non-standard accents are higher in solidarity (e.g., Acheme & Cionea, 2022). Researchers suggest these differences may be due to the accents compared, listener's cultural identity, and the type of speech that speakers produce (Bradac & Wisegarver, 1984). Our results demonstrate that when speakers with non-standard accents convey a vocal competence level, it can differentially affect their perceived vocal friendliness/attractiveness. However, more research is needed to understand the relationship between a speaker's perceived competence and warmth when they are conveying a vocal intention.

Decoding vocal competence in a non-standard accent

This study also provided evidence that we are less proficient at decoding the vocal competence cues of speakers with non-standard accents compared to speakers with standard

accents. The decoding process may also be more effortful and take longer for listeners compared to speech in a standard accent. The cognitive model of vocal expression processing proposes that the vocal cues of speakers with unfamiliar out-group accents are analyzed via an "indirect route" that involves iterative cognitive analysis, more so than speech in an unfamiliar in-group accent (Jiang, Keenan-Gossack, & Pell, 2020). This cognitive analysis involves three components: specifying the meaning of the vocal cue, integrating the speech into the communicative context, and forming pragmatic inferences (Jiang et al., 2020). For job interviewers, the additional pressure to make quick, accurate judgments of job candidates (Derous et al., 2016) likely does not aid this extensive processing of speech in a non-standard accent. Our results also demonstrate that the in-group advantage for recognizing vocal emotions (Elfenbein & Ambady, 2002; 2003) applies to the perception of more complex emotions/mental states by speakers with in-group accents and the social inferences drawn from their speech. Thus, by job interviewers being patient and potentially reducing time pressures, they may accurately and thoroughly evaluate the competence of job candidates with non-standard accents.

Other broader recommendations for improving listeners' processing of speech in a non-standard accent may also benefit their decoding ability. These recommendations promote increased long-term exposure to speech in a non-standard accent to improve listener's perceptual adaptation and/or attitude towards speakers with non-standard accents. One, engaging in media (e.g., movies, TV shows) that involves speech in a non-standard (or out-group) accent can provide a low-stakes, passive way for listeners to increase their experience with processing speech in a non-standard accent. For example, watching TV in a non-standard accent can improve listener's acceptability judgments and associate warmth-related attributes to speakers with non-standard accents (Peng, 2020). Two, listeners can increase the language and cultural diversity in their social

network. This can be done actively by making more social connections with people who have variable language and cultural backgrounds. For example, listeners with more racially diverse social networks are more likely to perceive standard and non-standard accents as less accented (Kutlu, Tiv, Wulff, & Titone, 2021), potentially suggesting a wider acceptability of language varieties. Workplaces can also increase cross-cultural/linguistic interactions by having a culturally and language diverse group of employees. This can involve companies hiring people with non-standard accents at all levels (e.g., entry-level to managerial) (Kim, Roberson, Russo, & Briganti, 2019). By people working together over an extended period they may be encouraged to form deeper social connections (Kim et al., 2019), including understanding each other's vocal mental and affective states.

Lastly, listeners can expand their language background. Bi-/multi-lingual speakers are hypothesized to have greater perceptual flexibility in spoken word recognition as they have greater familiarity with variable pronunciations both within and between their languages (Weber, Di Betta, & McQueen, 2014). Moreover, speaking a second language can allow listeners to demonstrate more empathy and perspective taking of speakers with non-standard accents, reducing their potential negative bias towards speakers with non-standard accents (Hansen, Rakić & Steffens, 2014). However, the effectiveness of these recommendations may vary across individuals (including their extant cultural/language attitudes), the non-standard accent being processed (including its perceived strength), and potentially the complexity of a speaker's vocal mental/affective state. More research is needed to investigate how listeners can better decode the vocal cues of speakers with non-standard accents.

Limitations

The main limitation of this study is that we used short speech samples and few samples from each speaker. This allowed us to capture people's first impressions of speakers from a thin slice

(e.g., Ambady & Rosenthal, 1992; Ambady, Bernieri, & Richeson, 2000). Increasing this quantity of speech stimuli and variation in displays of vocal competence may have improved and/or changed the results of our acoustic analysis. At the same time, we detected medium-sized effects across all three perceptual experiments, and the experiments produced relatively similar results. Thus, despite using a small quantity of speech stimuli expressing a vocal competence level, the stimuli still meaningfully conveyed these vocal mental/affective states which were accurately decoded by participants. More research is needed on how a speaker's perceived job suitability is affected by the vocal expression of competence using more speech samples with greater variability, longer speech samples, and speech that provides more context. For example, job candidates are often asked different types of questions to assess their competence, including behavioral questions (e.g., candidate's personal values and beliefs, solving hypothetical problems) and technical questions (e.g., assess level of expertise for job-related tasks) (Brosy, Bangerter, & Ribeiro, 2020). Job interviewers may evaluate the answers to some questions with greater weight compared to others, and they may be exposed to a greater wealth of vocal and linguistic cues. Understanding how interviewers' initial competence impressions change over time using speech that is comparable in quantity and richness to a real interview, will provide a comprehensive understanding of how a job candidate's voice shapes their perceived competence.

Also, we specifically examined explicit attitudes and biases that may occur in a job interview. The impressions of a job candidate's competence may differ if implicit attitudes and bias were investigated. Additionally, in Study 2, the vocal competence level was a between-subjects factor, so we created a context where, for example, all job candidates expressed high vocal competence. While a context where all candidates sound confident is likely, a context where they all sound uncertain may be less likely. This methodological decision supported a simpler experimental

design requiring less factorial conditions. Future research could explore the use of a full factorial design as well as the presentation of audio only in Block 1, and audio and visual stimuli in Block 2. It is also unclear if we would observe different results if candidates were evaluated for a different job position. For example, jobs involving less spoken communication (e.g., data analyst, software developer) may not have resulted in differences between speakers with standard and non-standard accents (Hosoda & Stone-Romero, 2010), and occupational stereotypes related to speaker's accent/ethnicity may have been activated (Leong & Hayes, 1990; He, Kang, Tse, & Toh, 2019).

Lastly, we studied a narrow range of accents and ethnicities and focused on one visible ethnic minority group. This allowed us to build on extant literature surrounding the perception of East Asians' speech and competence (Cargile, 1997; 2000; Fiske et al., 2002; Gnevsheva, 2018; Kil et al., 2019; Wong & Phooi-Ching, 2000), while conducting a novel investigation surrounding the effect of speaker's accent, vocal competence level, and ethnic appearance. However, the results may change if a different comparison of accents or ethnicities was studied. Also, participants were largely Caucasian (the dominant racial group in Canada). Although this allowed us to obtain perceptions from the people that job candidates with non-standard accents may likely encounter in job interviews, other perceptions are possible. More research is needed to understand the impact of listener's accent and race/ethnicity on their impression of job candidates with non-standard accents who are visible ethnic minorities.

Future directions

With constant technological advancements and the pandemic pushing people to work from home, work environments have changed drastically in the last thirty years. There are now more jobs that involve exclusively working remotely (Adrjan et al., 2021), and organizations greatly rely on technology to communicate remotely such as videoconferencing meetings, messaging platforms and online collaborative tools (Rimol, 2021; Turner et al., 2010). Also, job interviews

often consist of several stages, primarily involving phone calls and video calls, with face-to-face in-person job interviews being a less frequent occurrence. Due to this dominance in online communication, interviewers must rely on cues from less communication channels compared to in-person interactions to foster social connections with prospective employees and co-workers. Research has found that compared to in-person interactions, communicating via video platforms can increase the cognitive workload for participants, encouraging a reliance on heuristics in their processing of information and judgment of others (Ferran & Watts, 2008, Fiechter, Fealing, Gerrard, & Kornell, 2018). This can potentially make it harder for job candidates to connect with an interviewer (McIlvaine, 2019). However, it is not clear how this increased workload, due to technological restraints (e.g., quality of audio and video) and limited access to a candidate's non-verbal cues, may potentially affect job interviewers' decoding of job candidate's vocal mental/affective states during online interviews. Moreover, for job candidates with non-standard accents, it is unclear if meeting interviewers for the first time via online platforms prompts interviewers to greatly rely on heuristics to shape their judgment, combined with the cognitive challenges of processing speech in a non-standard accent. More research is needed to better understand the experiences of people with non-standard accents in online job interviews and the impact of online versus in-person contexts on our socio-emotional understanding of others.

Outside of a job interview context, this work may also inform the speech of computer interfaces and personal assistant technology (e.g., Apple Siri, Amazon Alexa, Microsoft Cortana). Building on the similarity-attraction paradigm (Byrne, 1971) and the mediating effect of cognitive evaluations (Montoya & Horton, 2004), research has found that people are more socially attracted to interfaces that share their accent. When using human voices for a computer interface, participants perceive voices with an in-group accent as more socially rich and likable (Dahlbäck

et al., 2001), and are more trusting of the content of their speech (Dahlbäck et al., 2007) compared to interfaces with out-group accents. Similar results are also found when using computer synthesized speech, where a speaker's accent affects their perceived credibility (Cowan et al., 2016; Feijóo-García et al., 2021). Since the prosody of this interface speech is never mentioned, it is unclear how using speech that conveys warmth could enhance a user's experience with the technology (Sutton et al., 2019), such as navigating a product with greater ease and speed.

Conclusion

In conclusion, this thesis contributes significant findings to the experiences of speakers with non-standard accents who may be immigrants to Canada. Despite beliefs that their non-standard accent determines their job interview outcome, other cues conveyed and inferred from their voice and physical appearance more strongly contribute to their perceived suitability for a job. Specifically, the perceived warmth of a speaker's voice greatly predicted speaker's competence and hirability for a managerial job involving a lot of spoken communication. This finding, across three perceptual experiments, indicates the significance of interviewers' interest in socially connecting with job candidates, regardless of their accent. Native listeners may be better at differentiating the vocal competence level of speakers with standard (compared to non-standard) accents. This may relate to the potential miscommunications that occur in job interviews involving job candidates with non-standard accents, and the related frustrations or negative affect that interviewers may form towards candidates with non-standard accents. However, with increased experience processing speech in a non-standard accent, particularly in a workplace, this decoding ability may improve. We hope that this work inspires more research on the real-world experiences of speakers with non-standard accents by investigating the intersectionality of various speaker identity variables with speech cues (e.g., gender, accent, vocal mental/affective state, and race/ethnicity). This would allow us to better understand the true weight of these various social and speech cues on the social impressions formed in various communicative contexts.

References

Abele, A. E., & Wojciszke, B. (2007). Agency and communion from the perspective of self versus others. *Journal of personality and social psychology*, *93*(5), 751-763.

Acheme, D. E., & Cionea, I. A. (2022). "Oh, I like Your Accent": Perceptions and Evaluations of Standard and Non-standard Accented English Speakers. *Communication Reports*, *35*(2), 92-105.

Adolphs, R. (2002). Recognizing emotion from facial expressions: Psychological and neurological mechanisms. *Behavioral and Cognitive Neuroscience Reviews*, *1*, 21–62. doi:10.1177/153458 2302001001003.

Alter, A. L., & Oppenheimer, D. M. (2009). Uniting the Tribes of Fluency to Form a Metacognitive Nation. *Personality and Social Psychology Review*, *13*(3), 219–235. https://doi.org/10.1177/1088868309341564

Ambady, N., & Rosenthal, R. (1992). Thin slices of expressive behavior as predictors of interpersonal consequences: A meta-analysis. *Psychological bulletin*, *111*(2), 256-274.

Ambady, N., Bernieri, F. J., & Richeson, J. A. (2000). Toward a histology of social behavior: Judgmental accuracy from thin slices of the behavioral stream. In *Advances in experimental social psychology* (Vol. 32, pp. 201-271). Academic Press.

Adrjan, P., Ciminelli, G., Judes, A., Koelle, M., Schwellnus, C., & Sinclair, T., (2021). Will it stay or will it go? Analysing developments in telework during COVID-19 using online job postings data, OECD Productivity Working Papers, 2021-30, OECD Publishing.

Asbrock, F. (2010). Stereotypes of social groups in Germany in terms of warmth and competence. *Social Psychology*, *41*(2), 76–81.

Baquiran, C. L. C., & Nicoladis, E. (2020). A doctor's foreign accent affects perceptions of competence. *Health Communication*, *35*(6), 726-730.

Bencharit, L. Z., Ho, Y. W., Fung, H. H., Yeung, D. Y., Stephens, N. M., Romero-Canyas, R., & Tsai, J. L. (2019). Should job applicants be excited or calm? The role of culture and ideal affect in employment settings. *Emotion*, *19*(3), 377-401.

Ben-Shachar, M., Lüdecke, D., & Makowski, D. (2020). effectsize: Estimation of Effect Size Indices and Standardized Parameters. *Journal of Open Source Software*, *5*(56), 2815. doi: 10.21105/joss.02815

Berger, C. R., & Calabrese, R. J. (1975). Some explorations in initial interaction and beyond: Toward a developmental theory of interpersonal communication. *Human Communication Research*, *1*,99–112.

Bergeron, A., & Trofimovich, P. (2017). Linguistic Dimensions of Accentedness and Comprehensibility: Exploring Task and Listener Effects in Second Language French. *Foreign Language Annals*, *50*(3), 547–566. https://doi.org/10.1111/flan.12285

Bestelmeyer, P. E., Belin, P., & Ladd, D. R. (2015). A neural marker for social bias toward in-group accents. *Cerebral Cortex*, *25*(10), 3953-3961.

Bhanji, J. P., & Beer, J. S. (2013). Dissociable neural modulation underlying lasting first impressions, changing your mind for the better, and changing it for the worse. *Journal of Neuroscience*, *33*(22), 9337-9344.

Billig, M., & Tajfel, H. (1973). Social categorization and similarity in intergroup behaviour. *European Journal of Social Psychology*, *3*(1), 27–52.

Birman, D., & Simon, C. D. (2014). Acculturation research: Challenges, complexities, and possibilities. In F. T. L. Leong, L. Comas-Díaz, G. C. Nagayama Hall, V. C. McLoyd, &

J. E. Trimble (Eds.), *APA handbook of multicultural psychology, Vol. 1. Theory and research* (pp. 207–230). American Psychological Association. https://doi.org/10.1037/14189-011

Boersma, P., & Weenink, D. (2020). Praat: Doing phonetics by computer v 6.1. 21.

Bradac, J. J., Konsky, C. W., & Davies, R. A. (1976). Two studies of the effects of linguistic diversity upon judgments of communicator attributes and message effectiveness. *Communications Monographs*, *43*(1), 70-79.

Bradac, J. J., Konsky, C. W., & Elliott, N. D. (1976). Verbal behavior of interviewees: The effects of several situational variables on verbal productivity, disfluency, and lexical diversity. *Journal of communication disorders*, *9*(3), 211-225.

Bradac, J., & Wisegarver, R. (1984). Ascribed status, lexical diversity, and accent: Determinants of perceived status, solidarity, and control of speech style. *Journal of Language and Social Psychology*, *3*(4), 239–255.

Bradac, J. J., & Street Jr, R. L. (1989). Powerful and powerless styles of talk: A theoretical analysis of language and impression formation. *Research on Language & Social Interaction*, *23*(1-4), 195-241.

Bradlow, A. R., Nygaard, L. C., & Pisoni, D. B. (1999). Effects of talker, rate, and amplitude variation on recognition memory for spoken words. *Perception & Psychophysics, 61,* 206–219.

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

Brambilla, M., Carraro, L., Castelli, L., & Sacchi, S. (2019). Changing impressions: Moral character dominates impression updating. *Journal of Experimental Social Psychology*, *82*, 64-73.

Branker, R. R. (2017). Labour market discrimination: The lived experiences of English-speaking Caribbean immigrants in Toronto. *Journal of International Migration and Integration*, *18*(1), 203-222.

Braun, B., Dainora, A., & Ernestus, M. (2011). An unfamiliar intonation contour slows down online speech comprehension. *Language and Cognitive Processes*, *26*(3), 350-375.

Brosy, J., Bangerter, A., & Ribeiro, S. (2020). Encouraging the production of narrative responses to past-behaviour interview questions: effects of probing and information. *European Journal of Work and Organizational Psychology*, *29*(3), 330-343.

Bye, H. H., Horverak, J. G., Sandal, G. M., Sam, D. L., & Van de Vijver, F. J. (2014). Cultural fit and ethnic background in the job interview. *International Journal of Cross Cultural Management*, *14*(1), 7-26.

Byrne, D. (1971). *The attraction paradigm*. Academic Press.

Callan, V. J., Gallois, C., & Forbes, P. A. (1983). Evaluative reactions to accented English: Ethnicity, sex role, and context. *Journal of Cross-Cultural Psychology*, *14*(4), 407-426.

Campbell, S., & Roberts, C. (2007). Migration, ethnicity and competing discourses in the job interview: Synthesizing the institutional and personal. *Discourse & Society*, *18*(3), 243-271.

Cargile, A. C. (2000). Evaluations of Employment Suitability: Does Accent Always Matter? *Journal of Employment Counseling*, *37*(3), 165–177. https://doi.org/10.1002/j.2161-1920.2000.tb00483.x

Cheung, C., & Sung, M. (2016). Does accent matter? Investigating the relationship between accent and identity in English as a lingua franca communication. *System*, *60*, 55-65.

Clarke, C. M., & Garrett, M. F. (2004). Rapid adaptation to foreign-accented English. The *Journal of the Acoustical Society of America*, *116*(6), 3647–3658. https://doi.org/10.1121/1.1815131

Cowan, B. R., Gannon, D., Walsh, J., Kinneen, J., O'Keefe, E., & Xie, L. (2016, May). Towards understanding how speech output affects navigation system credibility. In *Proceedings of the 2016 CHI Conference Extended Abstracts on Human Factors in Computing Systems* (pp. 2805-2812).

Cowen, A. S., Laukka, P., Elfenbein, H. A., Liu, R., & Keltner, D. (2019). The primacy of categories in the recognition of 12 emotions in speech prosody across two cultures. *Nature human behaviour*, *3*(4), 369-382.

Creese, G., & Kambere, E. N. (2003). What colour is your English?. *Canadian Review of Sociology/Revue canadienne de sociologie*, *40*(5), 565-573.

Creese, G. (2010). Erasing English language competency: African migrants in Vancouver, Canada. *Journal of International Migration and Integration/Revue de l'integration et de la migration internationale*, *11*(3), 295-313.

Cristia, A., Seidl, A., Vaughn, C., Schmale, R., Bradlow, A., & Floccia, C. (2012). Linguistic processing of accented speech across the lifespan. *Frontiers in Psychology*, *3*, 479. https://doi.org/10.3389/fpsyg.2012.00479

Dahlbäck, N., Swamy, S., Nass, C., Arvidsson, F., & Skågeby, J. (2001). Spoken interaction with computers in a native or non-native language-same or different. In *Proceedings of INTERACT* (pp. 294-301).

Dahlbäck, N., Wang, Q., Nass, C., & Alwin, J. (2007, April). Similarity is more important than expertise: Accent effects in speech interfaces. In *Proceedings of the SIGCHI conference on Human factors in computing systems* (pp. 1553-1556).

DeGroot, T., & Motowidlo, S. J. (1999). Why visual and vocal interview cues can affect interviewers' judgments and predict job performance. *Journal of Applied Psychology*, *84*(6), 986.

De Leeuw, J. R. (2015). jsPsych: A JavaScript library for creating behavioral experiments in a Web browser. *Behavior research methods*, *47*(1), 1-12.

Demo, D.A. (2006), 'Interactional Competence in Gatekeeping Encounters: A Discourse Analysis of Cross-Cultural Employment Interviews,' unpublished Ph.D. thesis, Georgetown University, Washington, DC.

Deprez-Sims, A. S., & Morris, S. B. (2010). Accents in the workplace: Their effects during a job interview. *International Journal of Psychology*, *45*(6), 417–426. https://doi.org/10.1080/00207594.2010.499950

Deprez-Sims, A. S., & Morris, S. B. (2013). The Effect of Non-native Accents on the Evaluation of Applicants during an Employment Interview: The development of a path model. *International Journal of Selection and Assessment*, *21*(4), 355–367. https://doi.org/10.1111/ijsa.12045

Derous, E., Buijsrogge, A., Roulin, N., & Duyck, W. (2016). Why your stigma isn't hired: A dual-process framework of interview bias. *Human Resource Management Review*, *26*(2), 90-111.

Derwing, T. M., Fraser, H., Kang, O., & Thomson, R. I. (2014). L2 Accent and Ethics: Issues that Merit Attention. In A. Mahboob & L. Barratt (Eds.), *English in a Multilingual Context* (pp. 63–80). New York: Springer.

Derwing, T. M., & Munro, M. J. (1997). Accent, intelligibility, and comprehensibility: Evidence from four L1s. *Studies in Second Language Acquisition*, *20*, 1–16.

Derwing, T. M., & Munro, M. J. (2009). Comprehensibility as a Factor in Listener Interaction Preferences: Implications for the Workplace. *Canadian Modern Language Review*, *66*(2), 181–202. https://doi.org/10.3138/cmlr.66.2.181

Dion, K., Berscheid, E., & Walster, E. (1972). What is beautiful is good. *Journal of personality and social psychology*, *24*(3), 285-290.

Doucerain, M. M., Varnaamkhaasti, R. S., Segalowitz, N., & Ryder, A. G. (2015). Second language social networks and communication-related acculturative stress: The role of interconnectedness. *Frontiers in psychology*, *6*, 1111.

Dragojevic, M., & Giles, H. (2014). The reference frame effect: An intergroup perspective on language attitudes. *Human Communication Research*, *40*(1), 91-111.

Dragojevic, M., & Giles, H. (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, 396–420. https://doi.org/10.1111/hcre.12079

Dragojevic, M., Giles, H., Beck, A.C., & Tatum, N. (2017). The fluency principle: Why foreign accent strength negatively biases language attitudes. *Communication Monographs*, *84*(3), 385-405. https://doi.org/10.1080/03637751.2017.1322213

Eagly, A. H., Ashmore, R. D., Makhijani, M. G., & Longo, L. C. (1991). What is beautiful is good, but…: A meta-analytic review of research on the physical attractiveness stereotype. *Psychological bulletin*, *110*(1), 109-128.

Egan, J. P. (1948). Articulation testing methods. *Laryngoscope*, *58*, 955–991

Elfenbein, H. A., & Ambady, N. (2002). On the universality and cultural specificity of emotion recognition: a meta-analysis. *Psychological bulletin*, *128*(2), 203-235.

Elfenbein, H. A., & Ambady, N. (2003). Universals and cultural differences in recognizing emotions. *Current directions in psychological science*, *12*(5), 159-164.

Elfenbein, H. A., Laukka, P., Althoff, J., Chui, W., Iraki, F. K., Rockstuhl, T., & Thingujam, N. S. (2022). What Do We Hear in the Voice? An Open-Ended Judgment Study of Emotional Speech Prosody. *Personality and Social Psychology Bulletin*, *48*(7), 1087-1104.

Elrick, J. (2016). Screening, skills and cultural fit: Theorizing immigrant skill utilization from an organizational perspective. *Journal of International Migration and Integration*, *17*(3), 801-817.

Eyben, F., Scherer, K. R., Schuller, B. W., Sundberg, J., André, E., Busso, C., ... & Truong, K. P. (2015). The Geneva minimalistic acoustic parameter set (GeMAPS) for voice research and affective computing. *IEEE transactions on affective computing*, *7*(2), 190-202.

Fairbanks, G. (1960). The rainbow passage. *Voice and articulation drillbook*, (2nd ed.). Harper & Row.

Faul, F., Erdfelder, E., Lang, A. G., & Buchner, A. (2007). G* Power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behavior research methods*, *39*(2), 175-191.

Fiechter, J. L., Fealing, C., Gerrard, R., & Kornell, N. (2018). Audiovisual quality impacts assessments of job candidates in video interviews: Evidence for an AV quality bias. *Cognitive research: principles and implications*, *3*(1), 1-5.

Feijóo-García, P. G., Zalake, M., de Siqueira, A. G., Lok, B., & Hamza-Lup, F. (2021, September). Effects of Virtual Humans' Gender and Spoken Accent on Users' Perceptions of Expertise in Mental Wellness Conversations. In *Proceedings of the 21st ACM International Conference on Intelligent Virtual Agents* (pp. 68-75).

Feiler, A. R., & Powell, D. M. (2016). Behavioral expression of job interview anxiety. *Journal of Business and Psychology*, *31*(1), 155-171.

Ferran, C., & Watts, S. (2008). Videoconferencing in the field: A heuristic processing model. *Management science*, *54*(9), 1565-1578.

Fiske, S., Cuddy, A., Glick, P., & Xu, J. (2002). A model of (often mixed) stereotype content: Competence and warmth respectively follow from perceived status and competition. *Journal of Personality and Social Psychology*, *82*(6), 878–902.

Fiske, S. T., Cuddy, A. J. C., & Glick, P. (2007). Universal dimensions of social cognition: Warmth, then competence. *Trends in Cognitive Sciences*, *11*, 77–83.

Floccia, C., Goslin, J., Girard, F., & Konopczynski, G. (2006). Does a regional accent perturb speech processing? *Journal of Experimental Psychology: Human Perception and Performance*, *32*(5), 1276–1293. https://doi.org/10.1037/0096-1523.32.5.1276

Foon, A. E. (1986). A social structural approach to speech evaluation. *The Journal of Social Psychology*, *126*(4), 521-530.

Freeman, J. B., & Ambady, N. (2011). A dynamic interactive theory of person construal. *Psychological review*, *118*(2), 247-279.

Fuertes, J. N., Gottdiener, W. H., Martin, H., Gilbert, T. C., & Giles, H. (2012). A meta-analysis of the effects of speakers' accents on interpersonal evaluations. *European Journal of Social Psychology*, *42*(1), 120–133. https://doi.org/10.1002/ejsp.862

Fuse, A., Navichkova, Y., & Alloggio, K. (2018). Perception of intelligibility and qualities of non-native accented speakers. *Journal of Communication Disorders*, *71*, 37–51. https://doi.org/10.1016/j.jcomdis.2017.12.006

Gallois, C., Callan, V. J., & Palmer, J. A. M. (1992). The influence of applicant communication style and interviewer characteristics on hiring decisions. *Journal of Applied Social Psychology*, *22*(13), 1041-1060.

Gass, S., & Varonis, E. M. (1984). The effect of familiarity on the comprehensibility of nonnative speech. *Language learning*, *34*(1), 65-87.

Giles, H. (1970). Evaluative reactions to accents. *Educational Review*, *22*(3), 211–227. https://doi.org/10.1080/0013191700220301

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

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

Goldberg, C. B. (2005). Relational Demography and Similarity-Attraction in Interview Assessments and Subsequent Offer Decisions: Are we missing something? *Group & Organization Management*, *30*(6), 597–624. https://doi.org/10.1177/1059601104267661

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

Goslin, J., Duffy, H., & Floccia, C. (2012). An ERP investigation of regional and foreign accent processing. *Brain and Language*, *122*(2), 92–102. https://doi.org/10.1016/j.bandl.2012.04.017

Gudykunst, W. B. (1985). A model of uncertainty reduction in intercultural encounters. *Journal of language and Social Psychology*, *4*(2), 79-98.

Guerrero, L., & Rothstein, M. G. (2012). Antecedents of underemployment: Job search of skilled immigrants in Canada. *Applied Psychology*, *61*(2), 323-346.

Guyer, J. J., Fabrigar, L. R., & Vaughan-Johnston, T. I. (2019). Speech rate, intonation, and pitch: Investigating the bias and cue effects of vocal confidence on persuasion. *Personality and Social Psychology Bulletin*, *45*(3), 389-405.

Hansen, K., & Dovidio, J. F. (2016). Social dominance orientation, non-native accents, and hiring recommendations. *Cultural Diversity and Ethnic Minority Psychology*, *22*(4), 544–551. https://doi.org/10.1037/cdp0000101

Hansen, K., Rakić, T., & Steffens, M. C. (2014). When actions speak louder than words: Preventing discrimination of nonstandard speakers. *Journal of Language and Social Psychology*, *33*(1), 68-77.

Hansen, K., Rakić, T., & Steffens, M. C. (2017). Competent and Warm? How Mismatching Appearance and Accent Influence First Impressions. *Experimental Psychology*, *64*(1), 27–36. https://doi.org/10.1027/1618-3169/a000348

Hansen, K., Rakić, T., & Steffens, M. C. (2018). Foreign-looking native-accented people: More competent when first seen rather than heard. *Social Psychological and Personality Science*, *9*(8), 1001-1009.

Hansen, K. (2020). Accent Beliefs Scale (ABS): Scale Development and Validation. *Journal of Language and Social Psychology*, *39*(1), 148-171.

He, J. C., Kang, S. K., Tse, K., & Toh, S. M. (2019). Stereotypes at work: Occupational stereotypes predict race and gender segregation in the workforce. *Journal of Vocational Behavior*, *115*, 103318.

He, J., Van de Vijver, F. J., Dominguez Espinosa, A., Abubakar, A., Dimitrova, R., Adams, B. G., ... & Villieux, A. (2015). Socially desirable responding: Enhancement and denial in 20 countries. *Cross-Cultural Research*, *49*(3), 227-249.

Hebbani, A., & Colic-Peisker, V. (2012). Communicating one's way to employment: A case study of African settlers in Brisbane, Australia. *Journal of Intercultural Studies*, *33*(5), 529-547.

Herman, J. L., Stevens, M. J., Bird, A., Mendenhall, M., & Oddou, G. (2010). The tolerance for ambiguity scale: Towards a more refined measure for international management research. *International Journal of Intercultural Relations*, *34*(1), 58-65.

Hofhuis, J., van der Zee, K. I., & Otten, S. (2016). Dealing with differences: the impact of perceived diversity outcomes on selection and assessment of minority candidates. The *International Journal of Human Resource Management, 27(*12), 1319–1339. https://doi.org/10.1080/09585192.2015.1072100

Hofstede, G. (1980). Culture and organizations. *International studies of management & organization*, *10*(4), 15-41.

Hofstede, G. (1984). *Culture's consequences: International differences in work-related values* (Vol. 5). Sage.

Hofstede, G. (2001). *Culture's consequences: Comparing values, behaviors, institutions, and organizations across nations* (2nd ed.). Sage.

Hollandsworth, J. G., Kazelskis, R., Stevens, J., & Dressel, M. E. (1979). Relative contributions of verbal, articulative, and nonverbal communication to employment decisions in the job interview setting. *Personnel Psychology*, *32*, 359–367.

Honorof, D. N., McCullough, J., & Somerville, B. (2000) Comma gets a cure: a diagnostic passage for accent study. https://www.dialectsarchive.com/wp-content/uploads/2012/10/COMMA-GETS-A-CURE.pdf

Horowitz, L. M., Wilson, K. R., Turan, B., Zolotsev, P., Constantino, M. J., & Henderson, L. (2006). How interpersonal motives clarify the meaning of interpersonal behavior: A revised circumplex model. *Personality and social psychology review*, *10*(1), 67-86.

Horverak, J. G., Bye, H. H., Sandal, G. M., & Pallesen, S. (2013). Managers' evaluations of immigrant job applicants: The influence of acculturation strategy on perceived person-organization fit (PO Fit) and hiring outcome. *Journal of Cross-Cultural Psychology*, *44*(1), 46-60.

Hosoda, M., Stone-Romero, E. F., & Walter, J. N. (2007). Listeners' cognitive and affective reactions to English speakers with standard American English and Asian accents. *Perceptual and Motor Skills*, *104*(1), 307-326.

Hosoda, M., & Stone-Romero, E. (2010). The effects of foreign accents on employment-related decisions. *Journal of Managerial Psychology*, *25*(2), 113–132. https://doi.org/http://dx.doi.org.ezproxy.rice.edu/10.1108/02683941011019339

Houshmand, S., Spanierman, L. B., & Tafarodi, R. W. (2014). Excluded and avoided: Racial microaggressions targeting Asian international students in Canada. *Cultural Diversity and Ethnic Minority Psychology*, *20*(3), 377-388.

Howard, J. L., & Ferris, G. R. (1996). The employment interview context: Social and situational influences on interviewer decisions 1. *Journal of applied social psychology*, *26*(2), 112-136.

Hyndman, J., & Hynie, M. (2016) "From newcomer to Canadian: Making refugee integration work", Policy Options, Institute for Research on Public Policy. https://policyoptions.irpp.org/magazines/may-2016/from-newcomer-to-canadian-making-refugee-integration-work/

IEEE. (1969). IEEE recommended practice for speech quality measurements. *IEEE Transactions on Audio and Electroacoustics*, *17*, 225–246.

International Phonetic Association (1949). *The principles of the international phonetic association*. University College.

Jansen, A., König, C. J., Stadelmann, E. H., & Kleinmann, M. (2012). Applicants' self-presentational behavior: What do recruiters expect and what do they get?. *Journal of Personnel Psychology*, *11*(2), 77-85.

Jiang, X., & Pell, M. D. (2015). On how the brain decodes vocal cues about speaker confidence. *Cortex*, *66*, 9-34.

Jiang, X., & Pell, M. D. (2017). The sound of confidence and doubt. *Speech Communication*, *88*, 106–126. https://doi.org/10.1016/j.specom.2017.01.011

Jiang, X., Sanford, R., & Pell, M. D. (2018). Neural architecture underlying person perception from in-group and out-group voices. *NeuroImage*, *181*, 582-597.

Jiang, X., Gossack-Keenan, K., & Pell, M. D. (2020). To believe or not to believe? How voice and accent information in speech alter listener impressions of trust. *Quarterly Journal of Experimental Psychology*, *73*(1), 55-79.

Johnson, K. (1997). Speech perception without speaker normalization: An exemplar model. In K. Johnson & J. Mullenix (Eds.), *Talker variability in speech processing* (pp. 145–166). Academic Press.

Judd, C. M., James-Hawkins, L., Yzerbyt, V., & Kashima, Y. (2005). Fundamental dimensions of social judgment: understanding the relations between judgments of competence and warmth. *Journal of personality and social psychology*, *89*(6), 899-913.

Kalin, R., & Rayko, D. S. (1978). Discrimination in evaluative judgments against foreign-accented job candidates. *Psychological Reports*, *43*, 1203–1207.

Keltner, D. (1995). Signs of appeasement: Evidence for the distinct displays of embarrassment, amusement, and shame. *Journal of personality and social psychology*, *68*(3), 441-454.

Kerekes, J. (2007). The co-construction of a gatekeeping encounter: An inventory of verbal actions. *Journal of Pragmatics*, *39*(11), 1942-1973.

Kervyn, N., Yzerbyt, V. Y., Judd, C. M., & Nunes, A. (2009). A question of compensation: the social life of the fundamental dimensions of social perception. *Journal of Personality and Social Psychology*, *96*(4), 828-842.

Kil, H., Noels, K. A., Lascano, D. I. V., & Schweickart, O. (2019). English Canadians' cultural stereotypes of ethnic minority groups: Implications of stereotype content for acculturation ideologies and immigration attitudes. *International Journal of Intercultural Relations*, *70*, 104-118.

Kim, R., Roberson, L., Russo, M., & Briganti, P. (2019). Language diversity, nonnative accents, and their consequences at the workplace: Recommendations for individuals, teams, and organizations. *The Journal of Applied Behavioral Science*, *55*(1), 73-95.

König, C. J., Hafsteinsson, L. G., Jansen, A., & Stadelmann, E. H. (2011). Applicants' Self-presentational Behavior across Cultures: Less self-presentation in Switzerland and Iceland than in the United States. *International Journal of Selection and Assessment*, *19*(4), 331-339.

Krieg, A., & Xu, Y. (2015). Ethnic differences in social anxiety between individuals of Asian heritage and European heritage: A meta-analytic review. *Asian American Journal of Psychology, 6*(1), 66–80. https://doi.org/10.1037/a0036993

Kristof-Brown, A. L. (2000). Perceived applicant fit: Distinguishing between recruiters' perceptions of person-job and person-organization fit. *Personnel Psychology*, *53*(3), 643-671.

Kuhlen, A. K., Bogler, C., Swerts, M., & Haynes, J. D. (2015). Neural coding of assessing another person's knowledge based on nonverbal cues. *Social Cognitive and Affective Neuroscience*, *10*(5), 729-734.

Kutlu, E., Tiv, M., Wulff, S., & Titone, D. (2021). The impact of race on speech perception and accentedness judgements in racially diverse and non-diverse groups. *Applied Linguistics*. https://doi.org/10.1093/applin/amab072

Kuznetsova, A., Brockhoff, P.B., & Christensen, R.H.B. (2017). "lmerTest Package: Tests in Linear Mixed Effects Models." *Journal of Statistical Software*, *82*(13), 1-26. doi:10.18637/jss.v082.i13.

Lambert, W. (1967). A Social Psychology of Bilingualism. *Journal of Social Issues*, *23*(2), 91–109.

Lange, K., Kühn, S., & Filevich, E. (2015). "Just Another Tool for Online Studies" (JATOS): An easy solution for setup and management of web servers supporting online studies. *PloS one*, *10*(6), e0130834.

Laukka, P., Neiberg, D., & Elfenbein, H. A. (2014). Evidence for cultural dialects in vocal emotion expression: Acoustic classification within and across five nations. *Emotion*, *14*(3), 445-449.

Laukka, P., Elfenbein, H. A., Thingujam, N. S., Rockstuhl, T., Iraki, F. K., Chui, W., & Althoff, J. (2016). The expression and recognition of emotions in the voice across five nations: A lens model analysis based on acoustic features. *Journal of personality and social psychology*, *111*(5), 686-705.

Laukka, P., & Elfenbein, H. A. (2021). Cross-cultural emotion recognition and in-group advantage in vocal expression: A meta-analysis. *Emotion Review*, *13*(1), 3-11.

Lee, Y. T., Ottati, V. C., Lin, C., & Chan, S. X. (2014). How are Asian Americans seen and evaluated? Examining ethnic stereotypes and their cultural complexity. *Cross-Cultural Communication*, *10*, 98 –111.

Leong, F. T., & Hayes, T. J. (1990). Occupational stereotyping of asian americans. *The Career Development Quarterly*, *39*(2), 143-154.

Lenth, R. (2022). emmeans: Estimated Marginal Means, aka Least-Squares Means. R package version 1.8.0, https://CRAN.R-project.org/package=emmeans.

Lev-Ari, S., & Keysar, B. (2012). Less-Detailed Representation of Non-Native Language: Why Non-Native Speakers' Stories Seem More Vague. *Discourse Processes*, *49*(7), 523–538. https://doi.org/10.1080/0163853X.2012.698493

Lewis. M. (1993). Self-conscious emotions: Embarrassment, pride, shame, and guilt. In M. Lewis &J. M. Havtiand (Eds.), *Handbook of emotions* (pp. 353-364). Guilford.

Lindemann, S. (2003). Koreans, Chinese or Indians? Attitudes and ideologies about non-native English speakers in the United States. *Journal of Sociolinguistics, 7*(3), 348–364. https://doi.org/10.1111/1467-9481.00228

Lippi-Green, R. (1997). *English with an accent: Language, ideology, and discrimination in the United States*. Routledge.

Macan, T. (2009). The employment interview: A review of current studies and directions for future research. *Human Resource Management Review*, *19*(3), 203-218.

Madziva, R., McGrath, S., & Thondhlana, J. (2016). Communicating employability: The role of communicative competence for Zimbabwean highly skilled migrants in the UK. *Journal of International Migration and Integration*, *17*(1), 235-252.

Mai, R., & Hoffmann, S. (2014). Accents in Business Communication: An integrative model and propositions for future research. *Journal of Consumer Psychology*, *24*(1), 137–158. https://doi.org/10.1016/j.jcps.2013.09.004

Manroop, L., Boekhorst, J. A., & Harrison, J. A. (2013). The influence of cross-cultural differences on job interview selection decisions. *The International Journal of Human Resource Management*, *24*(18), 3512-3533.

McIlvaine, A.R. (2019, September 5). Why do most Americans prefer in-person job interviews? *Human Resource Executive*. Retrieved from https://hrexecutive.com/why-do-most-americans-prefer-in-person-job-interviews/

McGovern, T. V., & Tinsley, H. E. (1978). Interviewer evaluations of interviewee nonverbal behavior. *Journal of Vocational Behavior*, *13*(2), 163-171.

Miller, R. O., Gayfer, B. L., & Powell, D. M. (2018). Influence of vocal and verbal cues on ratings of interview anxiety and interview performance. *Personnel Assessment and Decisions*, *4*(2), 26-41.

Miller, R. S., & Leary, M. R. (1992). Social sources and interactive functions of embarrassment. In M. Clark (Ed.), *Emotion and social behavior* (pp. 202-221). Sage.

Montoya, R. M., & Horton, R. S. (2004). On the importance of cognitive evaluation as a determinant of interpersonal attraction. *Journal of personality and social psychology*, *86*(5), 696-712.

Montoya, R. M., Horton, R. S., & Kirchner, J. (2008). Is actual similarity necessary for attraction? A meta-analysis of actual and perceived similarity. *Journal of Social and Personal Relationships*, *25*(6), 889-922.

Montoya, R. M., & Horton, R. S. (2012). The reciprocity of liking effect. In M. Paludi (Ed.), *The psychology of love* (pp. 39–57). Praeger.

Munro, M. J. (2003). A primer on accent discrimination in the Canadian context. *TESL Canada Journal*, *20*(2), 38-51.

Munro, M., & Derwing, T. (1994). Evaluations of foreign accent in extemporaneous and read material. *Language Testing, 11*(3), 253–266.

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

Munro, M., & Derwing, T. (1995b). Processing Time, Accent, and Comprehensibility in the Perception of Native and Foreign-Accented Speech. *Language and Speech*, *38*, 2–9. https://doi.org/10.1177/002383099503800305

Munro, M. J., & Derwing, T. M. (1999). Foreign accent, comprehensibility, and intelligibility in the speech of second language learners. *Language learning*, *49*, 285-310.

Nisbett, R. E., & Wilson, T. D. (1977). The halo effect: Evidence for unconscious alteration of judgments. *Journal of personality and social psychology*, *35*(4), 250-256.

Nordstrom, C.R., Hall, R.J., & Bartels, L.K. (1998). First impressions versus good impressions: The effect of self-regulation on interview evaluations. *The Journal of Psychology*, *132*,477–491. http://dx.doi.org/10.1080/00223989809599281.

Nygaard, L. C., & Queen, J. S. (2008). Communicating emotion: linking affective prosody and word meaning. *Journal of Experimental Psychology: Human Perception and Performance*, *34*(4), 1017-1030.

Padgett, J. K., Lou, E., Lalonde, R. N., & Sasaki, J. Y. (2020). Too Asian? The model minority stereotype in a Canadian context. *Asian American Journal of Psychology*, *11*(4), 223-232.

Paladino, M. P., & Mazzurega, M. (2020). One of us: On the role of accent and race in real-time in-group categorization. *Journal of Language and Social Psychology*, *39*(1), 22-39.

Pantos, A. J. (2012). Defining the cognitive mechanisms underlying reactions to foreign accented speech: An experimental approach. *Review of Cognitive Linguistics*, *10*(2), 427–453. https://doi.org/10.1075/rcl.10.2.08pan

Parton, S. R., Siltanen, S. A., Hosman, L. A., & Langenderfer, J. (2002). Employment interview outcomes and speech style effects. *Journal of Language and Social Psychology*, *21*(2), 144-161.

Pell, M. D., Jaywant, A., Monetta, L., & Kotz, S. A. (2011). Emotional speech processing: Disentangling the effects of prosody and semantic cues. *Cognition & Emotion*, *25*(5), 834-853.

Peng, C. Y. (2020). The effects of media exposure and language attitudes on grammaticality judgments. *Global Chinese*, *6*(1), 69-95.

Peng, Y., Zebrowitz, L. A., & Lee, H. K. (1993). The impact of cultural background and cross-cultural experience on impressions of American and Korean male speakers. *Journal of Cross-Cultural Psychology*, *24*(2), 203-220.

Pisoni, D. (1997). Some thoughts on "normalization" in speech perception. In K. Johnson & J. Mullennix (Eds.), *Talker variability in speech processing* (pp. 9–32). Academic Press.

Porretta, V., Tucker, B. V., & Järvikivi, J. (2016). The influence of gradient foreign accentedness and listener experience on word recognition. *Journal of Phonetics*, *58*, 1-21.

Price, P. C., & Stone, E. R. (2004). Intuitive evaluation of likelihood judgment producers: Evidence for a confidence heuristic. *Journal of Behavioral Decision Making*, *17*(1), 39-57.

Puhacheuskaya, V., & Järvikivi, J. (2022). I was being sarcastic!: The effect of foreign accent and political ideology on irony (mis) understanding. *Acta Psychologica*, *222*, 103479, 1-13.

Raghuram, S. (2013). Identities on call: Impact of impression management on Indian call center agents. *Human Relations*, *66*(11), 1471-1496.

R Core Team (2021). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. https://www.R-project.org/.

Reysen, S. (2005). Construction of a new scale: The Reysen likability scale. *Social Behavior and Personality: an international journal*, *33*(2), 201-208.

Rimol, M. (2019). Gartner Survey Reveals a 44% Rise in Workers' Use of Collaboration Tools Since 2019. *Gartner*. Retrieved from https://www.gartner.com/en/newsroom/press-releases/2021-08-23-gartner-survey-reveals-44-percent-rise-in-workers-use-of-collaboration-tools-since-2019.

Rivera, L. A. (2015). Go with your gut: Emotion and evaluation in job interviews. *American journal of sociology*, *120*(5), 1339-1389.

Roessel, J., Schoel, C., Zimmermann, R., & Stahlberg, D. (2019). Shedding new light on the evaluation of accented speakers: Basic mechanisms behind non-native listeners' evaluations of non-native accented job candidates. *Journal of Language and Social Psychology*, *38*(1), 3–32. https://doi.org/10.1177/0261927X17747904

Rosenberg, S., Nelson, C., & Vivekananthan, P. S. (1968). A multidimensional approach to the structure of personality impressions. *Journal of Personality and Social Psychology*, *9*(4), 283–294. https://doi.org/10.1037/h0026086

Ryan, E. B., Hewstone, M., & Giles, H. (1984). Language and intergroup attitudes. In *Attitudinal judgment* (pp. 135-158). Springer.

Sakamoto, I., Ku, J., & Wei, Y. (2009). The deep plunge: Luocha and the experiences of earlier skilled immigrants from mainland China in Toronto. *Qualitative Social Work*, *8*(4), 427-447.

Sakamoto, I., Chin, M., & Young, M. (2010). Canadian Experience," Employment Challenges, and Skilled Immigrants: A Close Look Through "Tacit Knowledge". *Canadian Social Work*, *12* (1), 145-151.

Sam, D. L., & Berry, J. W. (2010). Acculturation: When individuals and groups of different cultural backgrounds meet. *Perspectives on psychological science*, *5*(4), 472-481.

Sandal, G. M., van de Vijver, F., Bye, H. H., Sam, D. L., Amponsah, B., Cakar, N., ... & Tien-Lun Sun, C. (2014). Intended self-presentation tactics in job interviews: A 10-country study. *Journal of Cross-Cultural Psychology*, *45*(6), 939-958.

Saslow, L. R., McCoy, S., van der Löwe, I., Cosley, B., Vartan, A., Oveis, C., ... & Epel, E. S. (2014). Speaking under pressure: Low linguistic complexity is linked to high physiological and emotional stress reactivity. *Psychophysiology*, *51*(3), 257-266.

Sauter, D. A., Eisner, F., Ekman, P., & Scott, S. K. (2010). Cross-cultural recognition of basic emotions through nonverbal emotional vocalizations. *Proceedings of the National Academy of Sciences*, *107*(6), 2408-2412.

Scherer, K. R., London, H., & Wolf, J. J. (1973). The voice of confidence: Paralinguistic cues and audience evaluation. *Journal of Research in Personality*, *7*(1), 31–44. https://doi.org/10.1016/0092-6566(73)90030-5

Schmid, M. S., & Hopp, H. (2014). Comparing foreign accent in L1 attrition and L2 acquisition: Range and rater effects. *Language Testing, 31* (3), 367-388. https://doi.org/10.1177/0265532214526175

Schmid Mast, M., Frauendorfer, D., & Popovic, L. (2011). Self-promoting and modest job applicants in different cultures. *Journal of Personnel Psychology*, *10*(2), 70-77.

Schroeder, J., & Epley, N. (2015). The sound of intellect: Speech reveals a thoughtful mind, increasing a job candidate's appeal. *Psychological science*, *26*(6), 877-891.

Schwarz, N. (2004). Metacognitive experiences in consumer judgment and decision making. *Journal of Consumer Psychology*, *14*(4), 332-348.

Semnani-Azad, Z., & Adair, W. L. (2013). Watch your tone… relational paralinguistic messages in negotiation: The Case of East and West. *International Studies of Management & Organization*, *43*(4), 64-89.

Skowronski, J. J., & Carlston, D. E. (1987). Social judgment and social memory: The role of cue diagnosticity in negativity, positivity, and extremity biases. *Journal of personality and social psychology*, *52*(4), 689-699.

Smith, V. L., & Clark, H. H. (1993). On the course of answering questions. *Journal of memory and language*, *32*(1), 25-38.

Statistics Canada (2016). Ethnic and cultural origins of Canadians: Portrait of a rich heritage. https://www12.statcan.gc.ca/census-recensement/2016/as-sa/98-200-x/2016016/98-200-x2016016-eng.cfm

Statistics Canada (2018, July 25). Generation status: Canadian-born children of immigrants. https://www12.statcan.gc.ca/nhs-enm/2011/as-sa/99-010-x/99-010-x2011003_2-eng.cfm

Statistics Canada (2021). Increasing diversity of languages, other than English or French, spoken at home [Infographic]. https://www150.statcan.gc.ca/n1/pub/11-627-m/11-627-m2022051-eng.htm

Sumner, M., Kim, S. K., King, E., & McGowan, K. B. (2014). The socially weighted encoding of spoken words: A dual-route approach to speech perception. *Frontiers in psychology*, *4*, 1-13.

Sutton, S. J., Foulkes, P., Kirk, D., & Lawson, S. (2019, May). Voice as a design material: Sociophonetic inspired design strategies in human-computer interaction. In *Proceedings of the 2019 CHI conference on human factors in computing systems* (pp. 1-14).

Tajfel, H., & Turner, J. (1979). An integrative theory of intergroup conflict. In W. G. Austin, & S. Worchel (Eds.), *The social psychology of intergroup relations* (pp. 33–47). Brooks/Cole.

Thai, M., Szeszeran, N. A., Hornsey, M. J., & Barlow, F. K. (2020). Forever foreign? Asian Australians assimilating to Australian culture are still perceived as less Australian than White Australians. *Social Psychological and Personality Science*, *11*(6), 812-820.

Timming, A. R. (2017). The effect of foreign accent on employability: a study of the aural dimensions of aesthetic labour in customer-facing and non-customer-facing jobs. *Work, Employment and Society*, *31*(3), 409–428. https://doi.org/10.1177/0950017016630260

Tingley, D., Yamamoto, T., Hirose, K., Keele, L., & Imai, K. (2014). mediation: R Package for Causal Mediation Analysis. *Journal of Statistical Software*, *59*(5), 1-38. http://www.jstatsoft.org/v59/i05/.

Todorov, A., Olivola, C. Y., Dotsch, R., & Mende-Siedlecki, P. (2015). Social attributions from faces: Determinants, consequences, accuracy, and functional significance. *Annual review of psychology*, *66*(1), 519-545.

Tracy, J. L., & Robins, R. W. (2004). Putting the self into self-conscious emotions: A theoretical model. *Psychological inquiry*, *15*(2), 103-125.

Tracy, J. L., & Robins, R. W. (2004). Show your pride: Evidence for a discrete emotion expression. *Psychological science*, *15*(3), 194-197.

Tracy, J. L., & Matsumoto, D. (2008). The spontaneous expression of pride and shame: Evidence for biologically innate nonverbal displays. *Proceedings of the National Academy of Sciences*, *105*(33), 11655-11660.

Tsai, J. L., Knutson, B., & Fung, H. H. (2006). Cultural variation in affect valuation. *Journal of personality and social psychology*, *90*(2), 288-307.

Tsai, J. L. (2007). Ideal affect: Cultural causes and behavioral consequences. *Perspectives on Psychological Science*, *2*, 242–259. http://dx.doi.org/10.1111/j.1745-6916.2007.00043.x

Tsurutani, C. (2012). Evaluation of speakers with foreign-accented speech in Japan: The effect of accent produced by English native speakers. *Journal of multilingual and multicultural development*, *33*(6), 589-603.

Turner, T., Qvarfordt, P., Biehl, J. T., Golovchinsky, G., & Back, M. (2010, April). Exploring the workplace communication ecology. In *Proceedings of the SIGCHI conference on human factors in computing systems* (pp. 841-850).

Uhlmann, E. L., & Cohen, G. L. (2005). Constructed criteria: Redefining merit to justify discrimination. *Psychological Science*, *16*(6), 474-480.

Van Engen, K. J., Baese-Berk, M., Baker, R. E., Choi, A., Kim, M., & Bradlow, A. R. (2010). The Wildcat Corpus of native-and foreign-accented English: Communicative efficiency across conversational dyads with varying language alignment profiles. *Language and speech*, *53*(4), 510-540.

Van Engen, K. J., & Peelle, J. E. (2014). Listening effort and accented speech. *Frontiers in Human Neuroscience*, *8*, 1-4. doi: 10.3389/fnhum.2014.00577

Van Kleef, G. A. (2009). How emotions regulate social life: The emotions as social information (EASI) model. *Current directions in psychological science*, *18*(3), 184-188.

Waaramaa, T., Laukkanen, A. M., Airas, M., & Alku, P. (2010). Perception of emotional valences and activity levels from vowel segments of continuous speech. *Journal of voice*, *24*(1), 30-38.

Wang, J., Cheng, G. H. L., Chen, T., & Leung, K. (2019). Team creativity/innovation in culturally diverse teams: A meta-analysis. *Journal of Organizational Behavior*, *40*(6), 693-708.

Weber, A., Di Betta, A. M., & McQueen, J. M. (2014). Treack or trit: Adaptation to genuine and arbitrary foreign accents by monolingual and bilingual listeners. *Journal of phonetics*, *46*, 34-51.

Wells, J. C. (1982). *Accents of English*. Cambridge University Press.

Wickham, H. (2016). *ggplot2: Elegant Graphics for Data Analysis*. Springer-Verlag.

Wiggins, J. S. (1979). A psychological taxonomy of trait-descriptive terms: The interpersonal domain. *Journal of Personality and Social Psychology*, *37*, 395–412.

Winkielman, P., Schwarz, N., Fazendeiro, T, & Reber, R. (2003). The hedonic marking of processing fluency: Implications for evaluative judgment. In J. Much & K. C. Klauer (Eds.), *The psychology of evaluation: Affective processes in cognition and emotion* (pp. 189-217). Lawrence Erlbaum Associates, Inc.

Wojciszke, B., Bazinska, R., & Jaworski, M. (1998). On the dominance of moral categories in impression formation. *Personality and Social Psychology Bulletin*, *24*(12), 1251-1263.

Wojciszke, B., & Abele, A. E. (2008). The primacy of communion over agency and its reversals in evaluations. *European Journal of Social Psychology*, *38*(7), 1139-1147.

Wolgast, S., Björklund, F., & Bäckström, M. (2018). Applicant ethnicity affects which questions are asked in a job interview: The role of expected fit. *Journal of Personnel Psychology*, *17*(2), 66-74.

Wong, P., & Babel, M. (2017). Perceptual identification of talker ethnicity in Vancouver English. *Journal of Sociolinguistics*, *21*(5), 603-628.

Wong, I. F., & Phooi-Ching, L. (2000). Chinese cultural values and performance at job interviews: A Singapore perspective. *Business communication quarterly*, *63*(1), 9-22.

Yang, Z., Huynh, J., Tabata, R., Cestero, N., Aharoni, T., & Hirschberg, J. (2020). What makes a speaker charismatic? producing and perceiving charismatic speech. In *Proc. 10th International Conference on Speech Prosody* (Vol. 2020, pp. 685-689).

Young, A. M., & Kacmar, K. M. (1998). ABCs of the interview: The role of affective, behavioral, and cognitive responses by applicants in the employment interview. *International Journal of Selection and Assessment*, *6*(4), 211-221.

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

Zhang, D. (2022). rsq: R-Squared and Related Measures. R package version 2.5, https://CRAN.R-project.org/package=rsq.

Appendices

Appendix A

Job description for human resources manager (adapted from Deprez-Sims & Morris, 2010)

A human resources manager is expected to complete the following tasks:

Training and Development

- Plans policies relating to personnel training and development
- Conducts employee orientation with human resources assistants to ensure adherence to company goals and policies

Recruitment

- Recruits, interviews, and selects employees to fill vacant positions, with supervisors and managers

Employee and Labour Relations

- Mediates conflicts between employees or between employees and supervisors/managers
- Maintains company record of insurance coverage, pension plans, and personnel transactions (e.g., hires, promotions, and terminations)
- Investigates on-the-job accidents and prepares report for insurance providers

Compensation Management

- Conducts research on labour market to determine competitive salaries for employees
- Prepares budget of personnel operations

Appendix B

The North Wind and Sun (International Phonetic Association, 1949)

The North Wind and the Sun were disputing which was the stronger when a traveller came along wrapped in a warm cloak. They agreed that the one who first succeeded in making the traveller take his cloak off should be considered stronger than the other. Then the North Wind blew as hard as he could, but the more he blew, the more closely did the traveller fold his cloak around him; and at last the North Wind gave up the attempt. Then the Sun shined out warmly and immediately the traveller took his cloak off. And so the North wind was obliged to confess that the Sun was the stronger of the two.

Appendix C

Table C1. Descriptive characteristics of the two groups of European and East Asian faces for each gender. Images of faces are from the MR2 Database (Strohminger et al., 2016)

Face Group	Face Sex	Face Ethnicity	Face ID	Perceived race accuracy (%)	Perceived as multi-racial (%)	Mean Perceived physical attractiveness (out of 7)	Mean Perceived age (years)
Group 1	Female	East Asian	AF01	97.62	2.38	3.43 (1.50)	23.10 (3.83)
	Female	European	WF09	100	0.00	3.41 (1.29)	24.12 (4.07)
	Male	East Asian	AM01	95.83	4.17	3.00 (1.19)	25.15 (6.01)
	Male	European	WM08	97.92	0.00	3.58 (1.58)	25.81 (3.93)
Group 2	Female	East Asian	AF11	95.74	6.38	3.62 (1.53)	28.66 (5.65)
	Female	European	WF06	97.96	0.00	3.94 (1.31)	28.02 (5.44)
	Male	East Asian	AM03	98.18	5.45	2.67 (1.44)	29.20 (5.19)
	Male	European	WM10	100	2.33	3.47 (1.59)	29.93 (4.94)

Appendix D

Study 1 Supplementary Material

Attention question accuracy attempts

We examined the number of times participants answered the attention questions about the job description to receive 100% accuracy. Most participants answered all the questions correctly within two attempts (M = 2.1, SD = 1.49, Median = 2, Mode = 1). Four participants answered the questions within 5-9 attempts, potentially suggesting that they were not attentively reading the job description or the question options. There was a significant correlation between the job suitability ratings and the number of attention question accuracy attempts when these participants were included, r = -0.08, t (1198) = -2.90, p = 0.004. This correlation was not significant when these participants were excluded. During exploratory analyses, the exclusion of these 4 participants from the main model of interest: Job suitability ~ vocal affective presentation style * speaker accent + vocal attractiveness + vocal friendliness, did not result in any difference to the model outcome. The model had an adjusted R^2 of 0.51 in accounting for variation in job suitability ratings, 0.17 from random effects, 0.34 from fixed effects, with a residual error variance of 7.88 (versus model 1 which has a residual error variance of 7.86). The model returned a significant intercept (B = 7.13, SE = 0.60, t (33.06) = 11.82, p < .001, 95% CI [5.91, 8.36]) with the by-participant random intercept contributing SD = 1.36, and the by-speaker random intercept contributing SD = 0.95. Participants gave higher job suitability ratings for speakers expressing pride compared to shame, B = -1.30, SE = 0.25, t (1067.03) = -5.20, p < .001, 95% CI [-1.79, -0.81], controlling for the above-mentioned fixed and random effects. There was no main effect of speaker accent. Participants gave higher job suitability ratings for speakers they perceived to sound more attractive, B = 0.48, SE = 0.06, t (1116.94) = 8.37, p < .001, 95% CI [0.36, 0.59], or friendly, B = 0.27, SE = 0.06, t (1113.37) = 4.43, p < .001, 95% CI [0.15, 0.40], controlling for the above-mentioned fixed and random effects. There was a significant interaction of Speaker accent and Vocal affective presentation style, B = -1.74, SE = 0.35, t (1072.09) = -4.94, p < .001, 95% CI [-2.43, -1.05].

Cronbach's alpha

The following values are from all the 60 participants. The Cronbach's alpha represents the level of internal consistency between the various speech stimuli in being perceived for each dimension. High internal consistency means that the average inter-item correlation between the speech stimuli is high. The Cronbach's alpha for the competence ratings for the 20 speech stimuli was 0.854, 95% CI [0.789, 0.894]. This is considered good internal consistency. The Cronbach's alpha for the hirability ratings was 0.826, 95% CI [0.729, 0.881], which is considered good internal consistency. The Cronbach's alpha for the vocal attractiveness ratings was 0.891, 95% CI [0.815, 0.929]. The Cronbach's alpha for the vocal friendliness ratings was 0.853, 95% CI [0.731, 0.908]. The Cronbach's alpha for the accent strength ratings for the 10 speech stimuli (from the North Wind and Sun) was 0.712, 95% CI [0.557, 0.802] which is considered an acceptable internal consistency. The Cronbach's alpha for the comprehensibility ratings was 0.812, 95% CI [0.735, 0.865].

Appendix E

Study 2 Experiment 1 Supplementary Material

Randomization check

The age of participants in the doubtful condition (M = 30.99, SD = 4.71) was comparable to the age of participants in the confident condition (M = 29.98, SD = 4.33), t (141.49) = -1.36, p = 0.17, 95% CI [-2.47, 0.45]. Also, the confident and doubtful conditions did not significantly differ in terms of the proportions of males and females, χ^2 (1, N = 151) = 0.01, p = 0.93.

Attention question accuracy attempts

We examined the number of times participants answered the attention questions about the job description to receive 100% accuracy. Most participants answered all the questions correctly within two attempts (M = 2.12, SD = 1.91, Median = 2, Mode = 1). Thirteen participants answered the questions within 5-13 attempts, potentially suggesting that they were not attentively reading the job description or the question options. There was not a significant correlation between the competence or hirability ratings and the number of attention question accuracy attempts (competence: r = 0.01, t (1206) = 0.46, p = 0.65; hirability, r = 0.02, t (1206) = 0.83, p = 0.41).

Appendix F

Study 2 Experiment 2 Supplementary Material

Table F1. *Difference in mean perceived confidence ratings (out of 5) by speaker ID*

Speaker accent	Speaker sex	Speaker ID	Mean difference rating (SD)
American English	M	6.USA	2.67 (2.46)
	M	10.USA	4.67 (2.00)
	M	17.USA	3.93 (2.25)
	F	1.USA	2.13 (2.70)
	F	15.USA	2.45 (2.40)
Singaporean English	M	2.SGP	0.20 (2.21)
	M	5.SGP	0.75 (2.17)
	M	9.SGP	0.72 (1.96)
	F	1.SGP	2.33 (1.84)
	F	3.SGP	4.25 (2.65)

Randomization check

The age of participants in the pride condition (M = 30.59, SD = 5.55) was comparable to the age of participants in the shame condition (M = 29.24, SD = 4.98), t (97.73) = 1.28, p = 0.20, 95% CI [-0.74, 3.44]. Also, the confident and doubtful conditions did not significantly differ in terms of the proportions of males and females, χ^2 (1, N = 100) = 0.23, p = 0.63.

Attention question accuracy attempts

We examined the number of times participants answered the attention questions about the job description to receive 100% accuracy. Most participants answered all the questions correctly within two attempts (M = 2.11, SD = 1.64, Median = 1.50, Mode = 1). Eight participants answered the questions within 5-9 attempts, potentially suggesting that they were not attentively reading the job description or the question options. There was not significant correlation between the competence or hirability ratings and the number of attention question accuracy attempts (competence: r = 0.03, t (2398) = 1.48, p = 0.13, 95% CI [-0.01, 0.07]; hirability, r = 0.03, t (2398) = 1.50, p = 0.13, 95% CI [-0.01, 0.07]).

We also examined participants' responses to the two attention questions in the post-experiment questionnaires where they were asked to select a rating along the Likert scale. Their responses were coded as correct (for selecting the appropriate rating) or incorrect. Most participants answered both questions correctly (Question 1: 96% accuracy, Question 2: 100% accuracy).