COGNITIVE SCIENCE
A Multidisciplinary Journal

Cognitive Science 45 (2021) e12986

ISSN: 1551-6709 online
DOI: 10.1111/cogs.12986

The Other Accent Effect in Talker Recognition: Now You See It, Now You Don't

Madeleine E. Yu,[a] Jessamyn Schertz,[b] Elizabeth K. Johnson[a]

[a]*Department of Psychology, University of Toronto*
[b]*Department of Language Studies, University of Toronto Mississauga*

Received 12 December 2020; received in revised form 16 April 2021; accepted 29 April 2021

Abstract

The existence of the Language Familiarity Effect (LFE), where talkers of a familiar language are easier to identify than talkers of an unfamiliar language, is well-documented and uncontroversial. However, a closely related phenomenon known as the Other Accent Effect (OAE), where accented talkers are more difficult to recognize, is less well understood. There are several possible explanations for why the OAE exists, but to date, little data exist to adjudicate differences between them. Here, we begin to address this issue by directly comparing listeners' recognition of talkers who speak in different types of accents, and by examining both the LFE and OAE in the same set of listeners. Specifically, Canadian English listeners were tested on their ability to recognize talkers within four types of voice line-ups: Canadian English talkers, Australian English talkers, Mandarin-accented English talkers, and Mandarin talkers. We predicted that the OAE would be present for talkers of Mandarin-accented English but not for talkers of Australian English—which is precisely what we observed. We also observed a disconnect between listeners' confidence and performance across different types of accents; that is, listeners performed equally poorly with Mandarin and Mandarin-accented talkers, but they were more confident with their performance with the latter group of talkers. The present findings set the stage for further investigation into the nature of the OAE by exploring a range of potential explanations for the effect, and introducing important implications for forensic scientists' evaluation of ear witness testimony.

Keywords: Talker recognition; Other Accent Effect; Language Familiarity Effect; Regional accent; Non-native accent; Accented speech; Speech processing

Correspondence should be sent to Madeleine E. Yu, Dept. of Psychology, University of Toronto Mississauga, 3359 Mississauga Road, Mississauga, Ontario L5L 1C6, Canada. E-mail: madeleine.yu@mail.utoronto.ca

1. Introduction

Listeners recognize talkers of a familiar language better than they recognize talkers of an unfamiliar language. For example, using a now-standard paradigm for testing talker recognition, German listeners were shown to be better at recalling previously heard voices from a line-up when hearing German than English, and vice versa (Goggin et al., 1991). This Language Familiarity Effect (LFE; Thompson, 1987) is well documented and often cited as evidence for the role of linguistic information in talker recognition (see Levi, 2019, for review). There is also evidence for what has been termed the Other Accent Effect (OAE; Stevenage et al., 2012). In the OAE, same-accented talkers are better recognized than other-accented talkers. For example, North American English listeners recognize talkers who speak North American English better than talkers who speak Spanish-accented English (Thompson, 1987).

Some question the robustness of the OAE, suggesting that it is simply a weaker iteration of the LFE (Goggin et al., 1991), or that it only applies to certain types of accents (Johnson et al., 2018). For example, some argue that the effect is more likely to occur with non-native accents (i.e., speech produced by second-language [L2] talkers that is influenced by elements of their first language [L1]; Johnson et al. 2018), or only in contexts were a particular accent is socially marked in a specific community (see, e.g., Braun et al., 2018). Here, we examine the OAE in different types of voice line-ups, with the aim of better understanding the factors that affect listeners' recognition of accented talkers.

Why are talkers who speak in accents other than one's own difficult to recognize? There are four likely explanations, none of which are mutually exclusive. One possibility is that the OAE is linked to abstract phonological knowledge (Johnson et al., 2018). According to this explanation, talker recognition crucially depends on the use of native language sound structure (Remez et al., 1997). When there is a substantial mismatch between the L1 phonology of the talker and the listener (e.g., their L1's have distinct segmental, phonotactic and/or prosodic structure), talker recognition is hampered. If this is the case, the OAE would increase in strength as the distance in phonological representations between the talker and listener increases. While the multidimensionality of phonological distance makes it hard to quantify with a single number, general support for this view comes from studies that suggest children's accumulating knowledge of the native language's sound structure improves talker recognition (Fecher & Johnson, 2021). Additional support comes from the finding that Canadian English listeners show no OAE for Australian English, an English variety that is immediately recognizeable as distinct from Canadian English, but shares many phonological characteristics with Canadian English (Johnson et al., 2018).

A second possible explanation for the OAE is that it is linked to language-based social biases, with talkers classified as either belonging (in-group) or not belonging (out-group) to a certain language community based on how they speak (Kozlowski, 2015). These biases are prevalent in adults (Bestelmeyer et al., 2015; Bresnahan et al., 2002; Cargile & Giles, 1997; Fuertes et al., 2012; Gluszek & Dovidio, 2010) and early emerging children (St. Pierre & Johnson, 2020; Paquette-Smith et al., 2019; Kinzler et al., 2007). Indeed, at least in young children, language-based social biases seem to outweigh race-based biases (e.g., Caucasian

American children prefer to be friends with other-race children who speak the same variety of English than same-race children who speak a non-native variety of English; Kinzler et al., 2009). If the OAE is largely due to language-based social biases, then one would predict that the OAE would be weak (or absent) with accents that are less socially marked in the community under investigation.

A third possible explanation is that the OAE, like the LFE, is linked to comprehension difficulties (see, e.g., Goggin et al., 1991; Köster & Schiller, 1997). For example, as talkers with accents distinct from one's own accent can be difficult to comprehend (see, e.g., Bradlow & Bent, 2008; Munro & Derwing, 1995), listeners may exhaust more processing resources to understand these other-accented talkers than they would with a same-accented talker (Brown et al., 2020; Dragojevic & Giles, 2016; Van Engen & Peelle, 2014) and, as a result, may be less accurate at recognizing the voices of other-accented talkers. If comprehension were the primary explanation for the OAE, the main prediction is that the strength of OAE increases as a function of accent incomprehensibility.

Lastly, a fourth possible explanation for the OAE is that it is linked to exemplar-based schemas for familiar languages (see Goggin et al., 1991, for a discussion in relation to the LFE). According to this view, listeners' exemplar-based schemas for frequently encountered languages or accents may improve their ability to distinguish members of a particular community, regardless of the relationship between the L1 phonology of the listener and the talker they are attempting to identify, and any language-based social biases the listener holds. Support for this view comes from work on the LFE, suggesting that merely over-hearing a language in daily life may lead to improved recognition of talkers of that language (Johnson et al., 2011; Orena et al., 2015). However, other studies call this prediction into question by reporting no improvements in talker recognition after two weeks of passive exposure to talkers (Fecher & Johnson, 2018b). If the OAE is modulated by the prevalence of relevant exemplars, then one might predict that the strength of the OAE would be stronger for accents to which listeners have greater exposure.

To summarize, we have laid out four possible types of explanations for the OAE, and to date, we have very little data to distinguish between these possibilities. In the current study, we take a first step towards understanding the factors that drive the OAE. Though the single experiment reported here cannot definitively identify a single cause for the OAE, our findings indicate which explanations of the OAE are more likely than others, and set the stage for additional work in this area.

2. Current study

In the current study, we adopted a voice line-up paradigm often used in the LFE literature (see, e.g., Johnson et al., 2011, 2018) in which listeners are familiarized with a single talker's voice at the beginning of each trial and then presented with a line-up of voices and asked to identify which voice they had just heard. We also asked listeners to report their confidence following the talker selection stage. Each participant completed a total of 16 voice line-ups—four different types of voice line-ups (e.g., Canadian English, Australian English, Mandarin-accented English, Mandarin) occurred four times each.

This is the first study to directly compare performance on a familiar accent to performance on two other types of accents (i.e., performance on the Canadian English line-ups versus performance on the Australian and Mandarin-accented line-ups), as well as to simultaneously compare the OAE to the LFE (i.e., we can compare performance on the Mandarin line-ups to performance on the Canadian English line-ups). Notably, by including accents that likely differ from the native accent in phonological distance, social status, comprehensibility, and familiarity, this enabled us to test different possible explanations for the OAE.

If the OAE is driven by phonological differences between accents, then a stronger OAE would be expected for more phonologically distinct accents. Given the distinctive phonology of Mandarin and English, we suggest that Mandarin-accented English is likely further from Canadian English in terms of its phonology than other closely related regional varieties such as Australian English (Johnson et al., 2018; see also Bent et al. 2016; Floccia et al., 2009; Mulak & Best, 2013 for support for the idea that non-native accents are often more phonologically distinct than regional accents). Based on this presumption, we predicted a stronger OAE for Mandarin-accented English than for Australian English. If the OAE is driven by social biases, then we again predicted that Mandarin-accented English would elicit a stronger OAE than Australian English would, given the well-established social bias against non-native accented talkers (see, e.g., Gluszek & Dovidio, 2010; Lambert et al., 1960; Lev-Ari & Keysar, 2010), particularly considering that Canadian listeners have been shown to be biased against Chinese-accented speech (Baquiran & Nicoladis, 2020). If the OAE is driven by processing limitations linked to comprehension difficulties, then we expected performance for the three types of accents to reflect Canadian listeners' relative comprehension of them (with Mandarin-accented English being the most difficult to understand compared to either Canadian or Australian English). And finally, if the OAE is due to exemplar-based schematic representations for particular accents, then we predicted that the OAE would be equally strong for all accents that are not routinely encountered (i.e., the OAE should be similar for Australian and Mandarin-accented English since our participants did not have routine exposure to either).

Following the work of prior studies (see, e.g., Yarmey, 1993), we also included a measure of listeners' confidence following talker selection and predicted that listener confidence would parallel performance accuracy. This confidence measure allowed us to assess the relationship between listener accuracy and listener confidence for each line-up type, an aspect that is highly relevant in forensic contexts when the reliability of ear-witness testimony by the witness against defendants from different linguistic backgrounds must be judged.

3. Method

3.1. Participants

Fifty-six Canadian English-speaking adults who learned English in Canada before the age of five (M_{age} = 21.9 years, SD = 3.6; 44 female) were tested. Although not explicitly asked about their race, participants were racially diverse, reflecting the make-up of the cosmopolitan community sampled. All participants reported no knowledge of Mandarin and

no routine exposure to either Australian English or Mandarin (including Mandarin-accented English). Participants received course credit or $10 compensation for participating. An additional 12 participants were excluded prior to the final analysis because they did not fit the language exposure criteria (8) or due to technical errors (4). This study was approved by the University of Toronto Research Ethics Board.

3.2. Stimuli

Auditory stimuli consisted of recordings by four female adult talkers for each of the four different voice line-up conditions in the study (M_{age} = 21.6 years, SD = 1.1). The Canadian English talkers were from the Greater Toronto Area; the Australian English talkers were from Sydney, New South Wales. The Mandarin and Mandarin-accented English talkers were native speakers of Standard Mandarin and learned English after the age of 5. Although it was not possible for us to record our Mandarin speakers from a single metropolitan area (as we did for our Canadian and Australian speakers), two native talkers of Mandarin ensured that the talkers all spoke what they recognized as a standard Mandarin variety. Mandarin-accented talkers were closely matched in perceived accent strength, with all having a highly perceptible non-native accent (see Yu, 2019). The script for English sentences was drawn from Johnson et al. (2011) and Mandarin sentences were modeled after them (see Appendix A).

Sentence length was controlled across languages and accents (range of 15–21 syllables). The relative variability of f0 (mean and standard deviation) and duration did not differ across talker sets (see Appendix B, Appendix C, and Table C1 for measurements and analyses), and care was taken to ensure that there were no obvious impressionistic differences in voice quality across talkers within a set. We also measured the relative comprehensibility of each English talker group based on speech-in-noise transcriptions by 12 English-speaking adults. As expected, the non-native Mandarin accent was substantially less intelligible than the native Australian and Canadian accents. Surprisingly, our Australian talkers were also slightly more intelligible to Canadians than our Canadian talkers (see Appendix D for additional details and analysis). The ease with which Canadians understood the Australian talkers is in line with our presumption that the phonological distance between Canadian English and Australian English is relatively small.

During the experiment, auditory stimuli were presented to participants over Sennheiser HD 280 PRO headphones at a comfortable listening level. Visual stimuli were presented over a computer monitor using Articulate Storyline 3 via an online learning management system CourseSites by Blackboard.

4. Procedure

At the start of each of 16 trials (see Fig. 1), participants were familiarized with a talker who produced a repetition of a pair of sentences (interstimulus interval [ISI] between sentences = 300 ms; ISI between the repetition of the set = 500 ms), accompanied by a visual representation of a cartoon creature. On-screen instructions directed them to remember the talker before viewing a 1-minute distractor video clip which featured instrumental music and sound

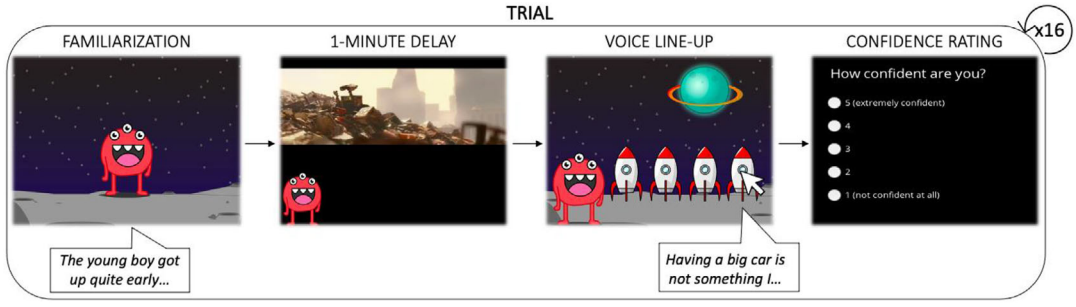

Fig. 1. Illustration of trial procedure. In each trial, participants were familiarized with a talker and, following a 1-minute delay, had to identify the same talker from a closed-set voice line-up that featured the target talker and three distractor talkers. The delay featured a short video without speech. Participants reported their confidence following each line-up. A single testing session took about 45 minutes to complete.

effects, but no speech (Fecher & Johnson, 2018a). Next, participants were presented with a four voice line-up that contained the target talker and three distractor talkers. The distractor talkers always spoke the same language/accent as the target, and each produced a unique sentence. Participants had to judge which of the four talkers was the target. They were able to replay the voices in the line-up if desired, and were instructed to listen to all voices before making a decision. Participants were told to guess if unsure. After each trial, participants reported their confidence with their talker selection (1 = not confident at all; 5 = extremely confident). A single practice trial, featuring acoustically distinct voices exclusive from the rest of the study, preceded the test trials and ensured participants understood the task. Participants received feedback on their performance following the practice trial.

Across trials, all 16 talkers were presented once as the target and three times as the distractor. Each participant heard all 40 recorded sentences. Sentences were never repeated within conditions, and familiarization sentences were unique across all conditions. Trial order was pseudo-randomized so that no language/accent condition occurred more than twice in a row. The order of language/accent conditions and the position of the target talker within line-ups were fully counterbalanced across all participants. Participants were individually tested in a quiet testing room and a 4-minute break occurred halfway through the study.

5. Results

5.1. Accuracy

Listeners' accuracy in talker recognition is shown in Fig. 2. To compare the effect of language condition on accuracy, we fit a generalized logistic mixed-effects model to our data using the *glmer* function in the *lme4* package Version 1.1-21 (Bates, Mächler, Boler, & Walker, 2015) in R. The model included the binary response variable, talker recognition Accuracy (1 = correct response). The independent variable, Language Condition, was entered as a fixed effect and was forward difference-coded to allow for adjacent comparisons:

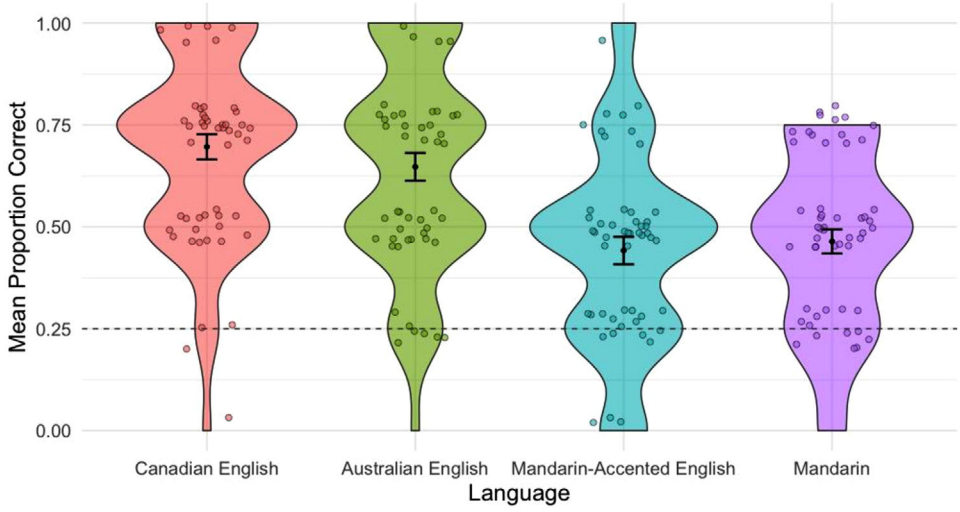

Fig. 2. Mean proportion of correct talker recognition by language condition. Error bars indicate SE of by-participant means, and the dashed line indicates chance level performance. Listeners performed significantly worse with Mandarin-accented English than with Australian English. Performance was equal with Canadian English and Australian English and equal with Mandarin and Mandarin-accented English.

(1) Canadian vs. Australian, (2) Australian vs. Mandarin-accented English, and (3) Mandarin-accented English vs. Mandarin. The maximal random effects structure that would converge was implemented, including random intercepts for Talker and a random by-participant intercept and slope for Language Condition. The β-coefficient for the intercept represents the log odds of a correct response averaged across all language conditions, and the β-coefficient for each Language Condition comparison represents the difference in log odds of a correct response between the two relevant levels.

A significant positive intercept indicated that listeners' performance, averaged across all Language Conditions, was above chance, $\beta = 0.28$, $SE = 0.10$, $z = 2.74$, $p < .01$. There was no significant difference in accuracy between Canadian and Australian accents ($M_{accuracy} = 69.6\%$ vs. 64.7%), $\beta = 0.23$, $SE = 0.28$, $z = 0.83$, $p = .41$, or between Mandarin and Mandarin-accented English ($M_{accuracy} = 46.4\%$ vs. 44.2%), $\beta = \hat{\text{R}}0.10$, $SE = 0.26$, $z = -0.39$, $p = .69$, but there was a significant difference between Australian and Mandarin-accented English ($M_{accuracy} = 64.7\%$ vs. 44.2%), $\beta = 0.90$, $SE = 0.28$, $z = 3.21$, $p < .005$. Given that there were condition-based differences, we performed follow-up tests, comparing the set of by-participant means in each language condition to chance (25%) via one-sample t-tests, to determine whether there was above-chance performance for all groups. Results showed that listeners were above-chance in all groups (all $p < .001$). Overall, the results provide evidence for the LFE as there was a significant difference in performance between Canadian English and Mandarin, and it also replicates Johnson et al.'s lack of the OAE with Canadian and Australian English. Furthermore, the difference between Australian and Mandarin-accented

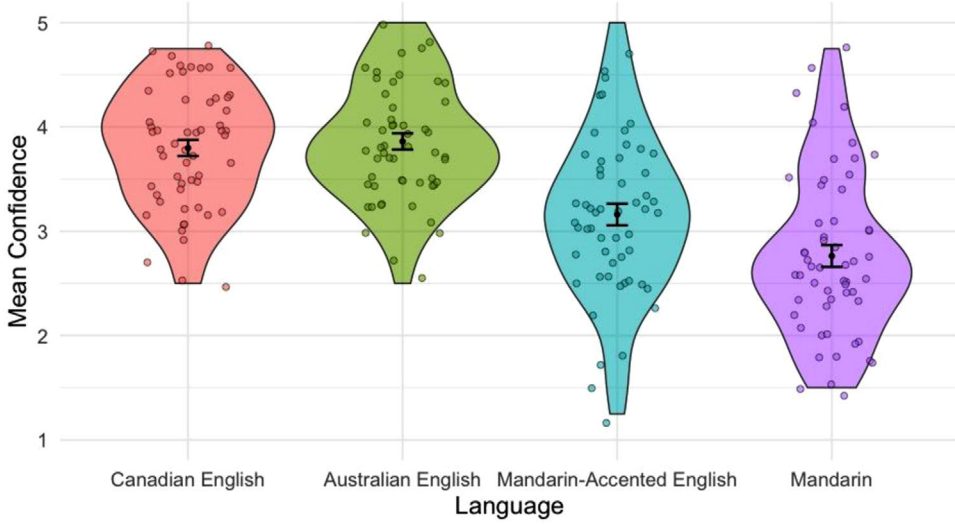

Fig. 3. Mean confidence in talker recognition by language condition. Error bars indicate SE of by-participant means. Listeners were equally confident with Canadian English and Australian English, but significantly less confident with Mandarin and Mandarin-accented English. Confidence was significantly higher with Mandarin-accented English than with Mandarin.

English supports predictions that the OAE is dependent on social and phonological factors and aligns with the notion that the LFE and OAE could be driven by similar underlying mechanisms.

5.2. Confidence

Listeners' confidence in talker recognition is shown in Fig. 3. To compare the effect of Language Condition on Confidence, we fit a linear mixed-effects regression model to our data using the *lmer* function in the *lme4* package in R. The model included Confidence Rating as the dependent variable and Language Condition as the independent variable, with the same coding scheme as above. The maximal random effects structure that would converge was implemented, including a random by-participant intercept and slope for Language. The β-coefficient corresponding to each effect represents the difference in confidence rating between the two levels of that comparison.

The model revealed that, parallel to performance, there was no significant difference in confidence between Canadian and Australian accents ($M_{confidence} = 3.80$ vs. 3.86), $\beta = -0.06$, $SE = 0.09$, $t = -0.70$, $p = .49$, and confidence was significantly different between both Australian and Mandarin-accented English ($M_{confidence} = 3.86$ vs. 3.16), $\beta = 0.70$, $SE = 0.10$, $t = 6.70$, $p < .001$. Interestingly, there was a difference in listeners' confidence between Mandarin-accented English and Mandarin ($M_{confidence} = 3.16$ vs. 2.76), $\beta = 0.40$, $SE = 0.09$, $z = 4.40$, $p < .001$, despite the fact that, as discussed above, accuracy was equivalent in the

two conditions. This disconnect between overall confidence and performance with Mandarin and Mandarin-accented English indicates that the relationship between confidence and accuracy is modulated by the type of speech heard.

Given the mismatch found above, we investigated the relationship between confidence and accuracy, and whether this relationship differed by condition. To do this, we fit a generalized logistic mixed-effects model to our data using the *glmer* function in the *lme4* package in R. The model included talker recognition Accuracy as the binary response variable, with Confidence Rating, Language Condition, and their interaction, included as independent variables. Language Condition was forward-coded as in the previous models. The maximal random effects structure that would converge was implemented, including random intercepts for Talker and a random by-participant intercept and slope for Language Condition.

There was a main effect of Confidence Rating, $\beta = 0.47$, $SE = 0.08$, $z = 6.21$, $p < .001$, indicating that the correct talker was chosen more often when higher confidence ratings were reported. However, there was no significant interaction between confidence and any of the condition comparisons; the relationship between confidence and accuracy did not differ between Canadian and Australian English, $\beta = -0.17$, $SE = 0.23$, $z = -0.75$, $p = .46$, between Australian and Mandarin-accented English, $\beta = 0.18$, $SE = 0.22$, $z = 0.86$, $p = .39$, or between Mandarin-accented English and Mandarin, $\beta = 0.18$, $SE = 0.19$, $z = 0.99$, $p = .32$. This indicates that in all line-ups, listeners selected the correct talker when they were more confident, and they selected the incorrect talker when they were less confident.

6. General discussion

The OAE has been observed for many accent varieties, including non-native accents (e.g., Spanish-accented English; Thompson, 1987) and regional accents (e.g, Standard Received Dutch vs. The Hague Dutch). But what are the underlying mechanisms driving the OAE? And what is the relationship between the OAE and the LFE? Is the OAE present for all, or only some, accents? If the latter, can we predict which accents will impact talker recognition most? In the current study, we presented listeners with four types of talkers whose speech varied in key features (e.g., phonological similiarity to listeners' native variety of English, ease of comprehension). This allowed us to explore various explanations for the OAE and to examine how accuracy in talker recognition related to confidence during performance.

We found that Canadian listeners are equally good at identifying Canadian and Australian English talkers, and equally poor at identifying Mandarin and Mandarin-accented talkers. Listeners' confidence ratings generally mirrored accuracy, but interestingly, listeners reported greater confidence with Mandarin-accented English talkers than with Mandarin talkers, despite equal performance between the two line-up types. Thus, our findings makes three key contributions to the study of talker recognition. First, we confirm that the magnitude of the OAE differs by accent, indicating that not all "other" accents are processed in the same way. Second, we demonstrate that the OAE can disrupt talker recognition as severely as the LFE. And third, while confidence is related to accuracy within a given accent, listeners' confidence ratings do not faithfully reflect accent-based differences in talker recognition performance.

Based on our findings, what can we conclude regarding the OAE's underlying mechanisms? The fact that listeners showed a robust OAE for Mandarin-accented English, but not Australian English, is arguably consistent with phonological explanations for the OAE, under the assumption that the phonology of the Mandarin-accented talkers in our study diverged more from the local variety than that of the Australian-accented talkers (e.g., Bent et al., 2016; Floccia et al., 2009; Johnson et al., 2018). Given earlier discussed tendencies for biases against non-native English talkers (see, e.g., Baquiran & Nicoladis, 2020; Gluszek & Dovidio, 2010; Lambert et al., 1960; Lev-Ari & Keysar, 2010), the superior performance we observed with Australian and Canadian English compared with Mandarin-accented English is also consistent with social explanations for the OAE. Thus, phonological and social factors may both play a role in the OAE. To begin to tease apart the contributions of these two factors to the OAE, an interesting direction for future work might be to test Mandarin-English early bilinguals who are more familiar with the phonology of Mandarin-accented English than monolingual English talkers, but who might still possess an implicit bias against Mandarin-accented English (Baquiran & Nicholadis, 2020; DeJesus et al., 2017; Hanzlíková & Skarnitzl, 2017). Looking at individual differences in the possession of accent biases and performance on talker recognition might also be particularly informative.

Our results also align fairly well with the hypothesis that the OAE is driven by comprehension difficulties. Recall that we predicted that performance with the accent varieties should reflect how accurately Canadian listeners could transcribe those recordings (i.e., the Mandarin-accented English condition should show the strongest OAE because listeners must expend the most processing effort to comprehend that speech). Indeed, our transcription data supported the idea that comprehension impacts the OAE as Mandarin-accented English talkers were both the most difficult to understand and the most difficult to identify among the accent varieties. Yet our findings do not align perfectly with the comprehension hypothesis: Although the stimuli produced by the Australian talkers were rated as more comprehensible than the Canadian English, talker identification was equal across the two line-ups.

The fourth and final explanation we discussed in the Introduction received the least support from our findings. That is, contrary to predictions of the exemplar hypothesis, we observed substantial performance differences between Australian and Mandarin-accented English—two accents that participants had no routine exposure to. Moreover, we find no difference in performance between Australian and Canadian English, even though our participants were far more familiar with the latter than the former variety of English. These findings are in stark contrast to the predictions of the exemplar hypothesis, which suggests that a listener's success in identifying a talker should be closely related to how familiar the listener is with the language or accent spoken by that talker.

A major strength of the current study is that it is the first to test the same listeners on four different types of voice line-ups. By doing so, we begin to directly test some key hypotheses regarding the mechanisms behind the OAE. But at the same time, a limitation of the current study is that we tested listeners on *only* four different types of line-ups. So although the current study represents a substantial advance in the understanding of the OAE, it leaves many questions unanswered. Future studies should incorporate a wider variety of languages

and accents, both to confirm that the current findings are generalizable and to tease apart the social and phonological explanations by selecting varieties that do not covary (i.e., a socially marked but phonologically similar accent). In addition, future research could more closely examine the role of comprehension in the OAE by systematically manipulating how easily talkers are understood.

Another aspect of this study that merits discussion is the outcome of our confidence rating measures; while higher confidence ratings generally corresponded to greater accuracy, confidence can also be modulated by the type of speech listeners heard. For instance, listeners reported an overall higher confidence with both Canadian and Australian English compared to both Mandarin-accented English and Mandarin. And, despite equal performance between the latter two types of line-ups, participants reported higher confidence on average with Mandarin-accented English than with Mandarin. This disconnect between overall confidence and accuracy across conditions suggests that the strength of confidence is dependent on the type of line-up involved.

Overall, this work offers insight into the mechanisms at work during human speech processing and lays the groundwork for future research to look at a broader range of accents and to control for factors such as implicit accent biases more carefully. On a theoretical level, we begin to explore the possible mechanisms underlying the OAE. And finally, on a practical level, this work shows that the OAE varies by accent type, an important factor for forensic scientists to consider when evaluating ear witness testimony.

Acknowledgments

We thank Changhao Gao and Priscilla Fung for their assistance with stimulus creation. This research was supported by grants awarded to E. K. J. and J. S. from SSHRC (Social Sciences and Humanities Research Council of Canada), and to E. K. J. from NSERC (Natural Sciences and Engineering Research Council of Canada) and the CRC (Canada Research Chairs) program.

Open Research Badges

This article has earned Open Data and Open Materials badges. Data and materials are available at https://osf.io/qg5kx/?view_only=7715acd7c3c046f2ae84dad5b7e9b82a.

References

Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, *67*(1), 1–48. https://doi.org/10.18637/jss.v067.i01

Baquiran, C., & Nicoladis, E. (2020). A doctor's foreign accent affects perceptions of competence. *Health Communication*, *35*(6), 726–730. https://doi.org/10.1080/10410236.2019.1584779

Bent, T., Atagi, E., Akbik, A., & Bonifield, E. (2016). Classification of regional dialects, international dialects, and nonnative accents. *Journal of Phonetics*, *58*, 104–117. https://doi.org/10.1016/j.wocn.2016.08.004

Bestelmeyer, P., Belin, P., & Ladd, D. (2015). A neural marker for social bias toward In-group accents. *Cerebral Cortex*, *25*(10), 3953–3961. https://doi.org/10.1093/cercor/bhu282

Bradlow A. R., & Bent T. (2008). Perceptual adaptation to non-native speech. *Cognition*, *106*(2), 707–729. https://doi.org/10.1016/j.cognition.2007.04.005

Braun, A., Llamas, C., Watt, D., French, P., & Robertson, D. (2018). Sub-regional "other-accent" effects on lay listeners' speaker identification abilities: a voice line-up study with speakers and listeners from the North East of England. *International Journal of Speech, Language, and the Law*, *25*(2), 231–255. https://doi.org/10.1558/ijsll.37340

Bresnahan, M. J., Ohashi, R., Nebashi, R., Liu, W. Y., & Shearman, S. M. (2002). Attitudinal and affective response toward accented English. *Language & Communication*, *22*(2), 171–185. https://doi.org/10.1016/S0271-5309(01)00025-8

Brown, V.A., McLaughlin, D.J., Strand, J.F., & Van Engen, K.J. (2020). Rapid adaptation to fully intelligible nonnative-accented speech reduces listening effort. *Quarterly Journal of Experimental Psychology*, *73*(9), 1431–1443. https://doi.org/10.1177/1747021820916726

Cargile, A. C., & Giles, H. (1997). Understanding language attitudes: Exploring listener affect and identity. *Language & Communication*, *17*(2), 195–217. https://doi.org/10.1016/S0271-5309(97)00016-5

DeJesus, J. M., Hwang, H. G., Dautel, J. B., & Kinzler, K. D. (2017). Bilingual children's social preferences hinge on accent. *Journal of Experimental Child Psychology*, *164*, 178–191. https://doi.org/10.1016/j.jecp.2017.07.005

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

Fecher, N., & Johnson, E. K. (2021). Developmental improvements in talker recognition are specific to the native language. *Journal of Experimental Child Psychology*, *202*. https://doi.org/10.1016/j.jecp.2020.104991.

Fecher, N. & Johnson, E. K. (2018a). Effects of language experience and task demands on talker recognition by children and adults. *The Journal of the Acoustical Society of America*, *143*(4), 2409–2418. https://doi.org/10.1121/1.5032199

Fecher, N. & Johnson, E. K. (2018b). The native-language benefit for talker identification is robust in 7.5-month-old infants. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *44*(12), 1911–1920. https://doi.org/10.1037/xlm0000555

Floccia, C., Butler, J., Girard, F., and Goslin, J. (2009). Categorization of regional and foreign accent in 5-to 7-year-old British children. *International Journal of Behavioral Development*, *33*(4), 366–375. https://doi.org/10.1177/2F0165025409103871

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

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

Goggin, J. P., Thompson, C. P., Strube, G., & Simental, L. R. (1991). The role of language familiarity in voice identification, *Memory and Cognition*, *19*(5), 448–458. https://doi.org/10.3758/BF03199567

Johnson, E. K., Bruggeman, L., & Cutler, A. (2018). Abstraction and the (misnamed) language familiarity effect. *Cognitive Science*, *42*(2), 633–645. https://doi.org/10.1111/cogs.12520

Johnson, E. K., Westrek, E., Nazzi, T., & Cutler, A. (2011). Infant ability to tell voices apart rests on language experience. *Developmental Science*, *14*(5), 1002–1011. https://doi.org/10.1111/j.1467-7687.2011.01052.x

Hanzlíková, D., & Skarnitzl, R. (2017). Credibility of native and non-native speakers revisited: Do non-native listeners feel the same? *Research in Language*, *15*(3), 285–298. https://doi.org/10.1515/rela-2017-0016

Kinzler, K. D., Shutts, K., Dejesus, J., & Spelke, E. S. (2009). Accent trumps race in guiding children's social preferences. *Social Cognition*, *27*(4), 623–634. https://doi.org/10.1521/soco.2009.27.4.623

Kinzler, K. D., Dupoux, E., & Spelke, E. S. (2007). The native language of social cognition. *Proceedings of the National Academy of Sciences*, *104*(30), 12577–12580. https://doi.org/10.1073/pnas.0705345104

Köster, O., & Schiller, N. O. (1997). Different influences of the native language of a listener on speaker recognition. *Forensic Linguistics*, *4*(1), 18–28.

Kozlowski, A. (2015). The influence of accents on social perception. *Inkblot*, *4*, 12–16. https://inkblotuoft.files.wordpress.com/2015/06/alexandra1.pdf

Lambert, W. E., Hodgson, R. C., Gardner, R. C., & Fillenbaum, S. (1960). Evaluational reactions to spoken languages. *Journal of Abnormal and Social Psychology*, *60*(1), 44–51. https://psycnet.apa.org/doi/10.1037/h0044430

Lev-Ari S., & Keysar B. (2010). Why don't we believe non-native speakers? The influence of accent on credibility. *Journal of Experimental Social Psychology*, *46*(6), 1093–1096. https://doi.org/10.1016/j.jesp.2010.05.025

Levi, S. V. (2019). Methodological considerations for interpreting the Language Familiarity Effect in talker processing. *WIREs Cognitive Science*, *10*(2), e1483. https://doi.org/10.1002/wcs.1483

Mulak, K. E., & Best, C. T. (2013). Development of word recognition across speakers and accents. In L. Gogate, & G. Hollich (Eds.), *Theoretical and computational models of word learning: Trends in psychology and artificial intelligence* (pp. 242–269). IGI Global. 10.4018/978-1-4666-2973-8.ch011

Munro, M. J., & Derwing, T. M. (1995). Processing time, accent, and comprehensibility in the perception of native and foreign-accented speech. *Language and Speech*, *38*(3), 289–306. https://doi.org/10.1177/2F002383099503800305

Paquette-Smith, M., Buckler, H., White, K., Choi, J. & Johnson, E. K. (2019). The effect of accent exposure on children's sociolinguistic evaluation of peers. *Developmental Psychology*, *55*(4), 809–822. https://doi.org/10.1037/dev0000659

Remez, R. E., Fellowes, J. M., & Rubin, P. E. (1997). Talker identification based on phonetic information. *Journal of Experimental Psychology: Human Perception and Performance*, *23*(3), 651–666. http://doi.org/10.1037/0096-1523.23.3.651

Stevenage, S. V., Clarke, G., & McNeill, A. (2012). The 'other-accent' effect in voice recognition. *Journal of Cognitive Psychology*, *24*(6), 647–653. https://doi.org/10.1080/20445911.2012.675321

St Pierre, T., & Johnson, E. K. (2020). The development of accent-based friendship preferences: age and language exposure matter. *Proceedings of the 42nd Annual Meeting of the Cognitive Science Society*, Toronto, ON.

Thompson, C. P. (1987). A language effect in voice identification. *Applied Cognitive Psychology*, *1*(2), 121–131. https://psycnet.apa.org/doi/10.1002/acp.2350010205

Van Engen, K. J., & Peelle, J. E. (2014). Listening effort and accented speech. *Frontiers in Human Neuroscience*, *8*, Article 577. https://doi.org/10.3389/fnhum.2014.00577

Yarmey, A. D. (1993). Stereotypes and recognition memory for faces and voices of good guys and bad guys. *Applied Cognitive Psychology*, *7*(5), 419–431. https://doi.org/10.1002/acp.2350070505

Yu, M. (2019). *Re-evaluating the other accent effect in talker recognition* (Publication No. 27541135) [Master's Thesis, University of Toronto]. ProQuest Dissertations & Theses Global.

Appendix A: Sample of sentence stimuli

English

1. The young boy got up quite early in order to watch the sun rise.
2. This supermarket had to close due to economic problems.
3. The committee will meet this afternoon for a special debate.
4. Having a big car is not something I would recommend in the city.
5. Moms often leave the maternity unit two days after giving birth.

Mandarin (same order)

1. 这个小男孩为了看日出早早的起床了.
2. 这个超市因为经济原因必须要关门了.
3. 委员会在下午的会议上进行特别辩论.
4. 在这座城市我不是很会推荐拥有一辆大车.
5. 妈妈们总是生完孩子后两天就离开妇产科.

Appendix B: Mean acoustic measures, with standard deviations in parentheses, for the 16 talkers used in the voice line-ups

Talker	Language Condition	Mean F0 (Hz)	SD F0 (Hz)	Duration (s)	Articulation Rate (syll./s)
Talker 1	Canadian English	200.0 (6.6)	30.7 (7.1)	3.5 (0.3)	4.9 (0.4)
Talker 2	Canadian English	214.7 (11.4)	48.5 (9.7)	3.1 (0.3)	5.6 (0.5)
Talker 3	Canadian English	199.4 (8.8)	40.9 (8.7)	3.1 (0.4)	5.6 (0.6)
Talker 4	Canadian English	200.0 (7.2)	37.5 (6.4)	3.5 (0.3)	4.9 (0.4)
Talker 1	Mandarin Accented English	207.0 (9.3)	31.5 (10.5)	4.3 (0.5)	4.0 (0.4)
Talker 2	Mandarin Accented English	210.1 (7.3)	21.8 (6.4)	4.5 (0.4)	3.9 (0.3)
Talker 3	Mandarin Accented English	197.4 (6.9)	30.6 (7.3)	4.2 (0.4)	4.1 (0.4)
Talker 4	Mandarin Accented English	194.0 (8.4)	30.9 (7.8)	3.9 (0.4)	4.4 (0.4)
Talker 1	Australian English	205.4 (8.4)	39.5 (5.1)	3.8 (0.4)	4.7 (0.4)
Talker 2	Australian English	200.6 (7.6)	35.0 (4.8)	3.7 (0.5)	4.8 (0.5)
Talker 3	Australian English	188.9 (7.5)	39.5 (6.7)	3.5 (0.5)	5.1 (0.5)
Talker 4	Australian English	213.0 (9.5)	47.3 (8.2)	3.3 (0.4)	5.4 (0.6)
Talker 1	Mandarin	223.9 (5.9)	30.9 (5.2)	4.0 (0.3)	4.3 (0.2)
Talker 2	Mandarin	231.2 (9.0)	42.3 (7.9)	3.6 (0.3)	4.8 (0.4)
Talker 3	Mandarin	208.8 (4.9)	30.9 (5.7)	3.2 (0.2)	5.4 (0.3)
Talker 4	Mandarin	230.4 (6.2)	42.0 (6.4)	3.6 (0.2)	4.8 (0.3)

Note. Duration = average sentence duration; Hz = Hertz; s = second; syll. = syllable.

Table C1
Relative acoustic variability of talker sets, in F-values calculated as the ratio of the two variances

Language Pair	Duration	Mean F0	SD F0
CAN-AusE	1.22	1.85	2.08
CAN-MandAccE	1.04	1.07	2.54
CAN-Mandarin	1.86	1.95	1.30
AusE-MandAccE	1.17	1.73	1.22
AusE-Mandarin	2.26	1.06	1.61
MandAccE-Mandarin	1.93	1.83	1.96

Note. Critical *F*-values for significance at $p < 0.05$ are > 9.1.

Appendix C

Following previous work in this area (Johnson et al., 2011, 2018), in order to assess relative variability across each talker set, we compared variance across Set A versus Set B for each measure and each language pair. As different talkers were used in each language condition, it was important to ensure that the acoustic parameters of talker voices were matched in variability so that performance differences could be attributed to the language condition, rather than the particular set of talkers. The ratio of these two variances allows derivation of an *F*-value with degrees of freedom (*n*-1, *n*-1); here these degrees of freedom are (3,3), yielding a critical *F*-value of 9.1 for significance at $p < 0.05$. The *F*-values for all comparisons across the language conditions are far below 9.1. Thus, the talker sets were matched in relative variability.

Appendix D

Twelve additional adults (M_{age} = 30.8 years, SD = 8.8; 9 female) who learned English before the age of 6 and did not have any routine exposure to the accents in the study transcribed accented speech stimuli in noise (0 SNR). Each transcriber was presented with recordings of 12 different sentences from each accent for a total of 36 unique sentences in randomized order. Each subset of sentences from each accent included three sentences produced by each of the four different talkers. Across all transcribers, recordings were presented in each accent type an even number of times. As expected, mean transcription accuracy was lowest for Mandarin-accented English compared to the other English variants ($M_{\text{Australian English}} = 0.80$, SD = 0.24; $M_{\text{Canadian English}} = 0.70$, SD = 0.29; $M_{\text{Mandarin-accented English}} = 0.42$, SD = 0.32). Moreover, as expected, paired t-tests revealed that transcription accuracy was significantly worse for the Mandarin-accented recordings than for either the Australian, $t(11) = 18.94$, $p < .001$, or Canadian English, $t(11) = 8.93$, $p < .001$, recordings. Surprisingly, the Canadian English transcriptions were also slightly less accurate than the Australian English transcriptions, $t(11) = 3.98$, $p < .01$.