Do I need to repeat myself? Getting to the root of the Other Accent Effect

Madeleine E. Yu (madeleine.yu@mail.utoronto.ca)
Department of Psychology, University of Toronto
Mississauga, ON, L5L 1C6, Canada

Jessamyn Schertz (jessamyn.schertz@utoronto.ca)
Department of Language Studies, University of Toronto
Mississauga, ON, L5L 1C6, Canada

Elizabeth K. Johnson (elizabeth.johnson@utoronto.ca)
Department of Psychology, University of Toronto
Mississauga, ON, L5L 1C6, Canada

Abstract

Listeners struggle to identify talkers with a different accent than their own, a phenomenon known as the Other Accent Effect (OAE). But for reasons that are not well understood, the OAE is not consistently observed in all studies. Comprehension-related processing demands offer one explanation, such that other-accented talkers who are more easily understood are also easier to recognize. Here, we test this hypothesis using a forensic-style voice line-up. We examine native English-speaking adults' ability to recognize talkers from four accent groups, manipulating comprehension-related processing demands by presenting listeners with predictable sentences and subtitles (low-demand condition), or variable sentences without subtitles (high-demand condition). As predicted, the OAE was only observed for talkers with non-native accents. But crucially, our comprehension manipulation had no impact on talker recognition accuracy of any accent type. We conclude that comprehension ease is likely not a key factor driving the OAE. Other possible explanations are discussed.

Keywords: talker recognition; Other Accent Effect; accented speech processing; speech perception; forensic ear witness

Introduction

Research shows that listeners sometimes find it very difficult to correctly identify talkers who speak in a different accent (e.g., Stevenage et al., 2012, Johnson et al., 2018; Yu et al., 2021). This Other Accent Effect (OAE) however, is neither consistent across studies nor well-understood. Why is it that the OAE is not always observed?

One possible explanation is that the more challenging it is to understand an accent, the more difficult it is to identify the talkers of that accent. For instance, under the assumption that humans have a finite capacity for processing information, the simultaneous engagement of multiple tasks (i.e., attempting to comprehend speech while also attempting to remember the talker's voice) can put stress on the system (Kahneman, 1973). It is possible then that the more challenging it is to comprehend the speech a talker, the more processing resources are exhausted, leaving listeners with fewer resources to process other information available in the environment at the same time, such as voice information (Brown et al., 2020; Dragojevic & Giles, 2016). Indeed, evidence indicates that when greater attention is allocated to the lexical/semantic component of speech, this interferes with a listener's ability to detect a change in talkers (e.g., Mullennix & Pisoni, 1990; Neuhoff et al., 2014). It has been argued that the processing of indexical and linguistic information may be asymmetrical, with a tendency for attention to be automatically directed to lexical/semantic speech information (Connolly et al., 1990; Parmentier, 2009). This could provide an explanation for the results of Yu et al. (2021), where Canadian English-speaking adults were the least accurate at transcribing Mandarin-accented speech, indicating that they had greater difficulty understanding the lexical/semantic content of this type of speech. In order to contend with this challenge, listeners may have directed more attentional resources towards processing the linguistic information in this condition, resulting in worse recognition of Mandarin-accented talkers compared to Canadian and Australian talkers.

However, accents differ in many ways other than in relative comprehensibility, so the fact that ease of transcription aligns with more accurate talker identification does not provide solid evidence that comprehension-based processing demands is the primary driver of the OAE. For example, while some accents are a result of producing speech in a native language differently from others based on where one lives, or the social groups one belongs to (e.g., regional accents), other accents are a result of porting elements of one's native language into one's second language (e.g., non-native accents). As such, different accents contain a significant amount of information relating to a talker's identity (e.g., nationality, social status, regional membership) which greatly influences how individuals are perceived (Kinzler et al., 2009). For instance, non-native accented talkers tend to be perceived as less competent and less socially attractive than native-accented talkers (Adank et al., 2013). Indeed, utterances that are perceived as more pleasant, truthful, familiar, and positive tend to also be more easily processed (e.g., Boduch-Grabka & Lev-Ari, 2021; Lev-Ari & Keysar, 2010; Oppenheimer, 2008). In addition to differences in social information, regional and non-native accents may

In J. Culbertson, A. Perfors, H. Rabagliati & V. Ramenzoni (Eds.), *Proceedings of the 44th Annual Conference of the Cognitive Science Society*.

also differ in phonological structure. As non-native accents port elements of a separate language's sound structure into another, these accents likely share fewer phonological characteristics with the native accent than regional accents do (e.g., see Bent et al., 2021; Cristia et al., 2012, for discussion). In fact, some evidence supports the idea that children's accumulating knowledge of their native language's sound structure may improve talker recognition (Fecher & Johnson, 2021). Thus, it remains unclear how much the comprehensibility of accented talkers directly influences listeners' ability to identify them.

The current study is the first to directly investigate the role of comprehension within the OAE by manipulating comprehension-related processing demands. We used a forensic-style voice line-up paradigm typically used in the talker recognition literature (see, e.g., Johnson et al., 2011; 2018; Yu et al., 2021) in which listeners are first familiarized with a talker's voice at the beginning of each trial, and then following a brief delay, are asked to identify the same talker from a voice line-up that contains the target voice as well as two other distractor voices. As in Yu et al. (2021), we included line-ups for a variety of accents to directly compare talker recognition between different accent types: our participants' native accent (Canadian English), a regional accent (Australian English or Southern U.S. English), and a non-native accent (Mandarin-accented English). We also took a measure of listeners' confidence in their selections. As comprehension is closely related to intelligibility, or the amount of speech that is actually understood by the listener (e.g., Munro & Derwing, 1995; 2009), a separate group of listeners transcribed the speech stimuli to provide a baseline measure of intelligibility for each accent. We predicted that compared to native and regional accented speech, non-native accented speech would be most difficult to understand.

In the main task, we manipulated comprehension demands by implementing two conditions that differed in predictability. In the high-demand condition, there was extensive variability in the speech stimuli and listeners were not provided with any prior knowledge regarding the linguistic content of the speech stimuli. In comparison, the low-demand condition featured reduced variability in speech stimuli so that linguistic content was largely identical between talkers (i.e., speech comprehensibility was increased by making speech content more predictable). Listeners were also presented with a transcription of the speech content in advance to prime lexical expectations of speech stimuli during talker familiarization. This was done to further ease processing of the non-native accent (see Baese-Berk et al., 2021; Bradlow & Alexander, 2007; Davis et al., 2005, for similar manipulations of comprehension load).

As such, we predicted that if comprehension-related processing demands are a major factor behind the OAE, then reducing demands by providing listeners with explicit knowledge of speech content prior to talker familiarization, and by reducing sentence variability in general, should improve talker recognition performance, perhaps most noticeably with non-native accented talkers. However, if other factors besides comprehensibility of speech are more important in facilitating the presence of the OAE, then talker recognition performance might not differ between comprehension demand manipulations.

Method

Participants

A total of 120 native English-speaking adults were tested (target N = 144; M_{age} = 19.6 years; 86 female; 28 male; 6 non-binary). Thus far, 62 participants (target N = 72) have completed the high-demand condition while 58 participants (target N = 72) have completed the low-demand condition. Participants in both conditions were asked to identify talkers from each of the three accent types: native (Canadian English), regional (half heard Australian English and half heard Southern U.S. English), and non-native (Mandarin-accented English). All participants learned English before the age of five in Canada, or in another country where English was the dominant language spoken. All participants reported using English at least 80% of the time and did not have routine exposure (in media or from a particular individual) to either Australian English, Southern U.S. English, or Mandarin-accented English. No hearing or vision impairments were reported at the time of testing.

Participants received course credit for participating. An additional 27 participants were excluded prior to the final analysis because they did not pass the pre-test practice trial (16), did not follow instructions (6), or experienced technical issues (5). This drop-out rate is typical for online experiments of this type. All participants used a computer and took about 30 minutes to complete the task.

Materials

Auditory stimuli consisted of recordings by four female adult talkers for each of the four accent varieties in the study (M_{age} = 21.7 years) and the script for sentence recordings was drawn from Johnson et al. (2011). Including two different regional accents (each presented to half of the participants) allowed us to explore the OAE with a wider variety of accents and to test whether prior observations of the effect generalize to other regional accents (i.e., the OAE has been found to occur with non-native accents, but not with regional accents; Yu et al., 2021). We did not expect any differences between the two regional accents. Indeed, a direct comparison showed accuracy and confidence ratings did not differ between the two regional accents. Therefore, we collapsed them here for the purposes of our analyses.

The Canadian English talkers were from the Greater Toronto Area; the Australian English talkers were from Sydney, New South Wales; the Southern U.S. English talkers were from Knoxville, Tennessee. The Mandarin-accented English talkers all learned English after the age of 5, and while not from a single metropolitan area, were confirmed by two native speakers of Mandarin to be native speakers of standard Mandarin. Mandarin-accented talkers were closely matched in perceived accent strength, and all had a highly perceptible Mandarin accent when speaking English (see Yu et al., 2019).

Sentence length was controlled across accented speech (range of 15-21 syllables). There were no obvious impressionistic differences in voice quality across talkers in each set and talker sets did not differ in relative variability of f0 (mean and standard deviation) and duration (see Table 1; following prior work of Johnson et al., 2011; 2018; Yu et al., 2021, F-values were calculated as the ratio of the two variances, and all comparisons were below the critical F-value of 9.1).

Table 1: Relative Acoustic Variability of Talker Sets in F-Values

Accent Pair	Duration	Mean F0	SD F0
Can-Aus	1.22	0.54	2.08
Aus-MandAcc	1.17	1.73	1.22
Can-MandAcc	1.04	1.07	2.54
Can-South	1.01	0.48	3.77
South-MandAcc	1.02	1.94	0.67

To confirm our expectations regarding how our accent speech samples would differ in their relative intelligibility to listeners in the main task, we collected speech-in-noise transcriptions by an additional 16 native English-speaking adults (M_{age} = 19.1 years; 14 female) who learned English before the age of five and who did not have routine exposure to any of the accents in the study. The listeners for this transcription study were drawn from the same population as the listeners for the main task. Each transcriber was presented with recordings of eight different sentences from each accent embedded in noise (0 SNR) for a total of 32 unique sentences in randomized order. Two sentences within each of the four accents were produced by the same talker, so that all 16 talkers were heard by each transcriber. Across all transcribers, recordings were presented in each accent an even number of times. As expected, mean transcription accuracy was lowest for Mandarin-accented English compared to the other English variants ($M_{Southern}$ = 0.761, SD = 0.27; $M_{Australian}$ = 0.768, SD = 0.26; $M_{Canadian}$ = 0.688, SD = 0.29; $M_{Mandarin}$ = 0.430, SD = 0.30). Moreover, paired *t*-tests revealed that transcription accuracy was significantly worse for Mandarin-accented speech than for Southern U.S., $t(15) = 9.15$, $p < .001$, Australian, $t(15) = 9.53$, $p < .001$, and Canadian, $t(15) = 7.48$, $p < .001$, speech. Surprisingly, transcription accuracy was slightly higher for Australian and Southern U.S. speech than Canadian speech (Australian vs. Canadian: $t(15) = 3.30$, $p < .01$; Southern U.S. vs. Canadian: $t(15) = 2.84$, $p < .05$), but transcription accuracy did not differ between Southern U.S. and Australian speech, $t(15) = -0.46$, $p = .65$. Overall, the significant differences in performance observed between Mandarin-accented speech and all other types of speech support our presumption that non-native-accented speech is less intelligible than native- or regional-accented speech.

Procedure

The task was created and hosted via the Gorilla Experiment Builder (www.gorilla.sc; Anwyl-Irvine et al., 2020). Prior to testing, participants were instructed to complete the task in a quiet room free of distractions and to use headphones. A headphone screening test based on an antiphase discrimination task (Woods et al., 2017) was administered to ensure that participants were wearing headphones before they were allowed to progress to the main task.

Following the design of Yu et al. (2021), participants were asked to identify talkers from a forensic-style voice identification task. All participants heard four talkers from each of the three accent types: native, regional, and non-native. All participants heard Canadian English talkers for the native condition and Mandarin-accented English talkers for the non-native condition. For the regional accent condition, half of all the participants heard Australian English talkers while the other half heard Southern U.S. English talkers.

At the start of each of 12 trials, participants were familiarized with a talker who produced a pair of sentences (interstimulus interval [ISI] between sentences = 300ms) upon a participant-initiated mouse click. Following 500ms of silence, a one-minute distractor video clip which featured instrumental music and sound effects, but no speech (Fecher & Johnson, 2018; Yu et al., 2021) was displayed. Next, participants were presented with a three-voice line-up that contained the target talker and two distractor talkers in pseudo-randomized order. The distractor talkers always spoke the same accent as the target and produced a single sentence each. Following a participant-initiated mouse click and 500ms of silence, the sentence recording by the left-most talker played and automatically advanced to the recording by the next talker to the right following a 1000ms ISI. After the presentation of all three talkers, participants were asked to judge which of the three talkers was the target. They were able to replay the voices in the line-up as many times as they desired and were told to guess if unsure. Upon making a talker selection, they were also given the option to change their selection before making their final confirmation. After each selection, participants reported their confidence with their decision using a continuous sliding scale between 0 (low confidence) and 100 (high confidence). A single practice trial, featuring acoustically distinct voices exclusive from the rest of the task, preceded the test trials and ensured that participants understood the task. No feedback on correctness was provided to participants on any of the trials. A four-minute break occurred halfway through the task.

In the high-demand version of the task, participants heard 36 different sentences throughout the entirety of the task; sentences were never repeated within an accent and familiarization sentences were unique across all conditions. Each line-up featured a different sentence by each talker. In contrast, in the low-demand version of the task, each participant heard just three different sentences throughout the entirety of the task; all talkers in all accent conditions produced the same two familiarization sentences, and line-ups featured a third sentence that was produced by all talkers in all accent conditions. Prior to the start of the main task, participants in the low-demand version of the task were also presented with the transcription of the two familiarization sentences. For all participants, the position of the target talker was fully counterbalanced such that it was equally likely for the target to appear in any of the three positions. Trial order was also pseudo-randomized so that no accent condition occurred more than twice-in-a-row. The order of accent conditions was fully counterbalanced across all participants.

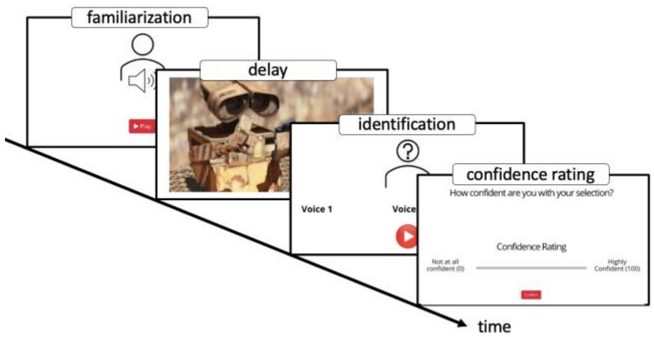

Figure 1: An example trial, in order of presentation

Results

Accuracy

Listeners' average talker recognition by comprehension demand, collapsed across accent type is shown in Figure 2.

To compare the effect of accent type and comprehension demand on talker recognition accuracy we fit a generalized logistic mixed-effects regression model to the data using the *glmer* function in the *lme4* package Version 1.1-21 (Bates et al., 2015) in R. The model included the binary response variable, talker recognition accuracy (1 = correct response). Accent type, comprehension demand, and their interaction were included as the independent variables. Accent type was forward difference-coded to allow for adjacent comparisons: (1) native vs. regional, and (2) regional vs. non-native. Comprehension demand was simple-coded (with the low-demand condition as the reference level). The maximal random effects structure that would converge was implemented and included a random intercept for talker and a random by-participant intercept and slope for accent condition.

A significant positive intercept indicated that listeners' performance, averaged across all accent and comprehension demand conditions, was above chance, $\beta = 0.87$, SE = 0.12, $z = 7.04$, $p < .001$. There was no significant difference in talker recognition between native and regional accents ($M_{accuracy} = 0.765$ vs. 0.766), $\beta = -0.02$, SE = 0.28, $z = -0.08$, $p = 0.93$, but there was a significant difference between regional and non-native accents ($M_{accuracy} = 0.766$ vs. 0.506), $\beta = 1.28$, SE = 0.27, $z = 4.72$, $p < .001$. There was no effect of comprehension demand nor any interactional effects between comprehension demand and accent type (*p*s > .05).

Since there were condition-based differences, we performed follow-up tests, comparing the set of by-participant means in each accent condition to chance (.33) via one-sample *t*-tests, to determine whether there was above-chance performance for all groups. Results showed that listeners in all groups did perform above chance (all *p*s < .001).

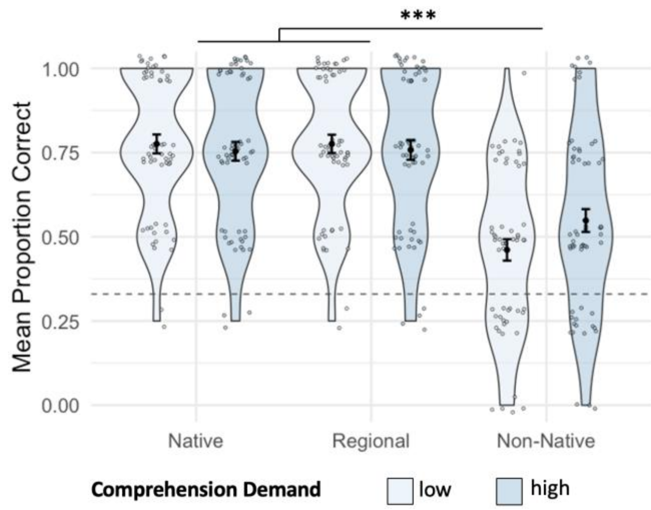

Figure 2: Mean proportion of correct talker recognition by accent type and comprehension demand. Error bars indicate SE of by-participant means, and the dashed line indicates chance level performance. Native accents include Canadian talkers, regional accents include Australian and Southern US talkers, and non-native accents include Mandarin-accented talkers.

Confidence

Listeners' confidence in talker recognition by comprehension demand, collapsed across accent type is shown in Figure 3.

To compare the effect of accent type and comprehension demand on confidence, we fit a linear mixed-effects regression model to the data using the *lmer* function in the *lmerTest* package Version 3.1-1 (Kuznetsova et al., 2017) in R. The model included confidence rating as the continuous response variable and accent condition, comprehension demand, and their interaction as the independent variables, with the same coding scheme as the previous model. The maximal random effects structure that would converge was implemented and included random intercepts for talker, accent condition, and participant.

Parallel to talker recognition performance, there was no significant difference in confidence for talker recognition in native and regional accents ($M_{confidence} = 76.6$ vs. 76.6), $\beta = 0.001$, SE = 0.03, $t = 0.04$, $p = 0.97$, but there was a significant difference between regional and non-native accents, ($M_{confidence} = 76.6$ vs. 63.0), $\beta = 0.14$, SE = 0.03, $t = 4.99$, $p < .001$. There was no effect of comprehension demand on confidence, nor any interactional effects between comprehension demand and accent type (*p*s > .05).

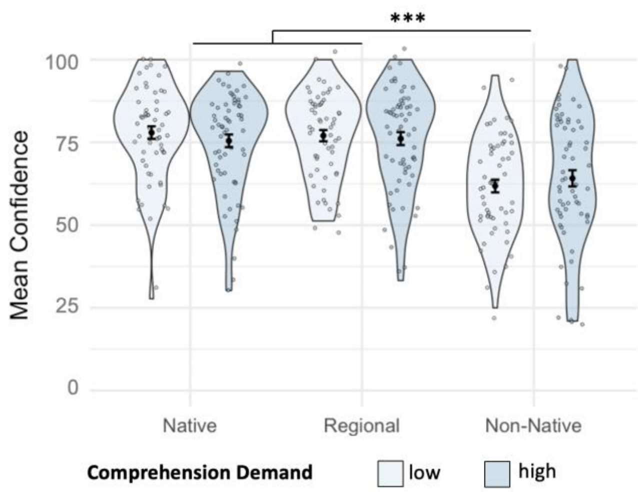

Figure 3: Mean confidence in talker recognition by accent type and comprehension demand. Error bars indicate SE of by-participant means.

Discussion

Past work indicates that the OAE may not apply to all 'other' accents equally. Why is it that talkers of certain accents are better recognized than others? We investigated whether the difference is based in relative comprehension-related demands required to process a talker's accented speech. We tested this hypothesis by comparing listener performance with a forensic-style voice identification task when the comprehension demand for speech was low or high. Listeners in both conditions recognized regional-accented (Australian or Southern U.S.) talkers equally as well as native-accented (Canadian) talkers, and better than non-native-accented (Mandarin-accented) talkers. These results replicate prior work with the OAE (Johnson et al., 2018; Yu et al., 2021) and extend it with the inclusion of another regional accent, Southern U.S. English. By doing so, this work strengthens existing evidence that the OAE does not apply to regional accents (and perhaps just to non-native accents). Contrary to our predictions however, reducing comprehension-related processing demands by manipulating speech content predictability did not affect listeners' performance. Listener confidence in talker selections, which also paralleled performance, was not modulated by easier comprehension of speech either. This suggests that at least in forensic-style voice identification tasks like the one used in the current study, speech comprehensibility is not a major factor in facilitating accented talker recognition.

Given this finding, how do we explain the observance that the accented talkers whose speech is more accurately transcribed tend to also be the ones that are more accurately identified? As previously mentioned, accents differ in many ways beyond relative comprehensibility. For instance, listeners may hold social biases against certain accents than others. Other-accented talkers tend to be perceived as less socially attractive than native-accented talkers (Adank et al., 2013) and this appears to be especially true when talkers have a non-native accent (e.g., Baquiran & Nicoladis, 2020; DeJesus et al., 2017). These negative social evaluations of non-native accents persist even when comprehension-related processing demands are eased (Vaughn & Whitty, 2020). Indeed, negative social evaluations have been shown to impede recognition memory (Philippon et al., 2008; Yarmey, 1993). Thus, negative social biases could provide an alternative explanation for why the OAE exists for some but not all regional and non-native accents (e.g., Yu et al., 2021).

But other explanations for why the OAE is only seen with some accent types remain. For example, accents differ in the extent that they share similar phonological structure to the listeners' native language (e.g., Bent et al., 2021; Cristia et al., 2012). Compared to the other accent varieties included in the current study, Mandarin-accented English is likely the most phonologically distinct from the listeners' native accent, Canadian English, given talkers' cross-language transfer from Mandarin. In line with this, some research supports this explanation as a key mechanism underlying native-language talker recognition (Fleming et al., 2014; Johnson et al., 2018; McLaughlin et al., 2019; Orena et al., 2015). Thus, future work is necessary to disentangle these two competing possibilities.

We also found that listeners' confidence in their talker selections reflected their overall performance and was similarly unaffected when comprehension-related processing demands were reduced. This observation is particularly important as current forensic voice line-up procedures typically ask earwitnesses to identify the perpetrator of a crime from a set of talkers who produce the same speech sample for comparison.

As our present observations are constrained to the accents included in the task, additional investigation should necessarily involve a greater variety of accents to build upon current understanding of the mechanisms behind the OAE. Indeed, it is not the case that non-native accents are always more distinct than regional accents. Indian English, for instance, features distinctive phonological differences from other English varieties, despite its status as a regional accent (e.g., see Wagner et al., 2014; Van Engen et al., 2010, for discussion). The current set of accents may also be potentially confounded with race biases, such that different accents may have associations with different ethnicities and by extension, different stereotypes. Such stereotypes may have in turn, elicited more positive or negative social evaluations of the talkers. Note, however, that accent-related social biases may outweigh race-related social biases in some cases (Kinzler et al., 2009). Thus, examining the presence of the OAE with a much wider variety of accents might help reveal the specific features of accented speech that generate the OAE.

In sum, this work makes two key contributions that further our understanding of the mechanisms underlying the OAE. First, we've provided the clearest support to date that the effect is likely to be observed with non-native (but not regional) accents. Second, we've shown that increasing the predictability of the stimuli that listeners encounter from talkers (thereby reducing comprehension-related processing demands), does not improve listeners' ability to process and recognize talker identity information. Moving forward, future work is needed to determine what other factors — perhaps differences in listeners' social biases and/or abstract phonological knowledge — are more instrumental in driving the OAE in talker recognition.

Acknowledgments

We would like to thank Priyanka Raj, Thomas St. Pierre, and the Child Language and Speech Studies Lab for their support. This work was supported by grants from the Social Sciences and Humanities Research Council and the Natural Sciences and Engineering Research Council.

References

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

Anwyl-Irvine, A. L., Massonnié, J., Flitton, A., Kirkham, N., & Evershed, J. K. (2020). Gorilla in our midst: An online behavioral experiment builder. *Behavior Research Methods*, *52*, 388–407. https://doi.org/10.3758/s13428-019-01237-x

Baese-Berk, M. M., Bent, T., & Walker, K. (2021). Semantic predictability and adaptation to nonnative speech. *JASA Express Letters*, *1*(1), 015207.

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., Holt, R. F., Van Engen, K. J., Jamsek, I. A., Arzbecker, L. J., Liang, L., & Brown, E. (2021). How pronunciation distance impacts word recognition in children and adults. *The Journal of the Acoustical Society of America, 150*(6), 4103-4117.

Boduch-Grabka, K., & Lev-Ari, S. (2021). Exposing Individuals to Foreign Accent Increases their Trust in What Nonnative Speakers Say. *Cognitive Science, 45*(11), e13064.

Bradlow, A. R., & Alexander, J. A. (2007). Semantic and phonetic enhancements for speech-in-noise recognition by native and non-native listeners. *The Journal of the Acoustical Society of America, 121*(4), 2339–2349. https://doi.org/10.1121/1.2642103

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

Connolly, J. F., Stewart, S. H., Phillips, N. A. (1990). The effects of processing requirements on neurophysiological responses to spoken sentences. *Brain and Language, 39*, 302–318.

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.

Davis, M. H., Johnsrude, I. S., Hervais-Adelman, A., Taylor, K., & McGettigan, C. (2005). Lexical information drives perceptual learning of distorted speech: evidence from the comprehension of noise-vocoded sentences. *Journal of Experimental Psychology: General, 134*(2), 222.

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

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

Derwing, T. M., & Munro, M. J. (1997). Accent, intelligibility, and comprehensibility: Evidence from four L1s. *Studies in second language acquisition*, *19*(1), 1-16.

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. (2018). 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

Fleming D, Giordano BL, Caldara R, & Belin P (2014). A language-familiarity effect for speaker discrimination without comprehension. *Proceedings of the National Academy of Sciences*, *11*(38), 13795–13798.

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

Kahneman, D. (1973). Attention and effort (Vol. 1063, pp. 218-226). Englewood Cliffs, NJ: Prentice-Hall.

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

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

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

McLaughlin, D. E., Carter, Y. D., Cheng, C. C., & Perrachione, T. K. (2019). Hierarchical contributions of linguistic knowledge to talker identification: Phonological versus lexical familiarity. *Attention, perception & psychophysics, 81*(4), 1088–1107. https://doi.org/10.3758/s13414-019-01778-5

Mullennix, J. W., & Pisoni, D. B. (1990). Stimulus variability and processing dependencies in speech perception. *Perception & psychophysics, 47*(4), 379-390.

Neuhoff, J. G., Schott, S. A., Kropf, A. J., & Neuhoff, E. M. (2014). Familiarity, Expertise, and Change Detection: Change Deafness is Worse in Your Native Language. *Perception, 43*(2–3), 219–222. https://doi.org/10.1068/p7665

Oppenheimer, D. M. (2008). The secret life of fluency. *Trends in Cognitive Science, 12*, 237–241.

Orena, A. J., Theodore, R. M., & Polka, L. (2015). Language exposure facilitates talker learning prior to language comprehension, even in adults. *Cognition, 143*, 36-40.

Parmentier, F. B. R. (2008). Towards a cognitive model of distraction by auditory novelty: The role of involuntary attention capture and semantic processing. *Cognition, 109*, 345–362.

Philippon, A. C., Cherryman, J., Vrij, A., & Bull, R. (2008). Why is my voice so easily recognized in identity parades? Influence of first impressions on voice identification. Psychiatry, *Psychology and Law, 15*(1), 70-77.

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*, 651–666. http://dx.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

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., 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.

Vaughn, C., & Whitty, A. (2020). Investigating the relationship between comprehensibility and social evaluation. *Journal of Second Language Pronunciation, 6*(3), 483-504.

Wagner, L., Clopper, C. G., & Pate, J. K. (2014). Children's perception of dialect variation. *Journal of child language, 41*(5), 1062-1084.

Woods, K. J. P., Siegel, M. H., Traer, J., & McDermott, J. H. (2017). Headphone screening to facilitate web-based auditory experiments. Attention, Perception, and *Psychophysics, 79*(7), 2064–2072. https://doi.org/10.3758/s13414-017-1361-2

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.E., Schertz, J., & Johnson, E.K. (2021). The other accent effect in talker recognition: now you see it, now you don't. *Cognitive Science, 45*(6). doi.org/10.1111/cogs.12986

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.