

DECEMBER 07 2023

Exploring effects of social information on talker-independent accent adaptation ^{EP}

Drew J. McLaughlin  ; Kristin J. Van Engen



JASA Express Lett. 3, 125201 (2023)

<https://doi.org/10.1121/10.0022536>



View
Online



Export
Citation

CrossMark



LEARN MORE

Advance your science and career as a member of the
Acoustical Society of America

Exploring effects of social information on talker-independent accent adaptation

Drew J. McLaughlin^{1,2,a)}  and Kristin J. Van Engen²

¹Basque Center on Cognition, Brain and Language, Donostia-San Sebastián, Gipuzkoa 20018, Spain

²Department of Psychological & Brain Sciences, Washington University in St. Louis, St. Louis, Missouri 63130, USA

d.mclaughlin@bcb.l.eu, kvanengen@wustl.edu

Abstract: The present study examined whether race information about speakers can promote rapid and generalizable perceptual adaptation to second-language accent. First-language English listeners were presented with Cantonese-accented English sentences in speech-shaped noise during a training session with three intermixed talkers, followed by a test session with a novel (i.e., fourth) talker. Participants were assigned to view either three East Asian or three White faces during training, corresponding to each speaker. Results indicated no effect of the social priming manipulation on the training or test sessions, although both groups performed better at test than a control group. © 2023 Author(s). All article content, except where otherwise noted, is licensed under a Creative Commons Attribution (CC BY) license (<http://creativecommons.org/licenses/by/4.0/>).

[Editor: Charles C. Church]

<https://doi.org/10.1121/10.0022536>

Received: 14 June 2023 **Accepted:** 1 November 2023 **Published Online:** 7 December 2023

1. Introduction

Although speech processing often appears effortless, listeners are constantly accommodating variation in speakers' productions. Every speaker produces speech somewhat uniquely (idiosyncratic variation), and groups of speakers of the same language also develop unique phonological patterns over time due to regional separation, social groupings, and additional language backgrounds (accent variation). The wide range of variation in spoken language can pose a challenge to listeners, who have to map speaker productions onto their internal linguistic categories. The task of rapidly accommodating unique speaker productions has been shown to induce a processing cost, such that blocks with multiple talkers (changing from trial-to-trial) result in slower and less accurate perception by listeners than blocks with a single talker (Mullennix *et al.*, 1989; Magnuson and Nusbaum, 2007; Choi and Perrachione, 2019; Kapadia and Perrachione, 2020). When changing between speakers of different accents, this processing cost is even larger, suggesting that the degree of processing cost may scale with the complexity of the mapping the listener needs to compute between a speaker's productions and their own internal linguistic categories (McLaughlin *et al.*, 2023).

Of particular interest in the present study is how listeners rapidly accommodate second-language (L2) accent variation, particularly from late L2 learners. Late learners of a given language often (although not always) produce speech differently than early (first language; L1) learners. At a phonetic level, for example, L2 speakers of a language may replace unfamiliar sounds in the inventory of their L2 with similar sounds from their L1. Because Cantonese does not have interdental fricatives (/θ/ and /ð/), for example, L1 Cantonese late learners of L2 English often replace these sounds with alveolar stops (/t/ and /d/). It is worth noting that replacement of sounds in this manner is seen in dialect variation as well and likely poses a familiar challenge to the listener. In other words, although the present study will focus on how listeners accommodate variation present in L2 speech, the challenge of reconciling systematic segmental differences (as well as suprasegmental differences) between a speaker's productions and the listener's own perceptual categories is prolific in all types of multi-accent communication. When external acoustic information and internal perceptual categories are misaligned (Van Engen and Peelle, 2014), it can slow processing (Porretta *et al.*, 2020), increase processing costs (McLaughlin and Van Engen, 2020), and reduce the accuracy of speech perception (Munro and Derwing, 2020). However, listeners are also able to rapidly adapt to speaker productions (Clarke and Garrett, 2004; Brown *et al.*, 2020), thereby improving their ability to perceive speech.

1.1 Accent accommodation

Improvement of listening accuracy for L2 accent perception can occur rapidly. By examining perceptual processing speed, Clarke and Garrett (2004) demonstrated rapid adaptation by L1 English listeners to both Spanish- and Mandarin Chinese-accented English. The authors used a word identification task that probed subjects after presentation of each sentence

^{a)} Author to whom correspondence should be addressed.

regarding the final word (i.e., requiring a “match” or “no match” response). Although listeners were initially slower to respond to the L2-accented speech, after only approximately 1 min of exposure, this processing cost was significantly reduced. Similar work by [Brown et al. \(2020\)](#) showed a rapid decrease in cognitive demands for L1 listeners processing Mandarin-accented English. In a pupillometry experiment, the authors found that task-evoked pupil response (a cognitive correlate of cognitive demands) reduced across trials at a faster rate for L2 accent than L1 accent. A dual-task (reaction time) experiment using the same stimuli revealed similar results, although only in exploratory analyses of the first 20 trials.

Notably, adaptation to L2 accent can also occur while listening to multiple talkers of the same accent. A seminal study by [Bradlow and Bent \(2008\)](#) found that L1 English listeners exposed to speech from multiple L2 speakers of the same accent performed better at post-test than listeners trained with a single L2 talker (see also [Sidasar et al., 2009](#)). Additionally, performance at post-test was better not only for the L2 speakers from the training session, but also for novel L2 speakers with the same accent. Their data further indicated that multi-talker accent training may be more effective than single-talker accent training, benefiting the listener when they encounter a novel L2 accent (e.g., encountering a Slovakian-accented speaker after training with a Mandarin accent). Work by [Baese-Berk et al. \(2013\)](#) with L1 English listeners expanded on this latter finding, further demonstrating that training with multiple L2-accented talkers of different language backgrounds can also improve performance on a post-test with a novel talker with an unfamiliar/untrained L2 accent (as compared to training with a single talker).

Altogether, work examining accommodation of L2 accent indicates not only that listeners adapt to a given L2 speaker, but that they can learn the systematic accent qualities present across multiple speakers with L2 accent and generalize that learning ([Bent and Baese-Berk, 2021](#)). The mechanisms supporting the process of accent accommodation, however, remain underspecified [see discussion by [Xie et al. \(2023\)](#)]. One potential mechanism supporting accommodation is *phonemic recalibration*, in which the listener “tunes” their own perceptual categories to better match the productions of the talker. Specifically, a recalibration mechanism posits that critical linguistic category boundaries are directionally shifted; for example, voice onset time (VOT) boundaries between voiced (e.g., /b/) and voiceless (e.g., /p/) stop pairs might be adjusted by the listener to be either more positive or negative, depending on a given speaker’s productions. Indeed, work examining acoustically manipulated L1 accent provides strong evidence that listeners can adjust their phonemic category boundaries [[Norris et al., 2003](#); see review by [Samuel and Kraljic \(2009\)](#)], and this type of phonemic recalibration mechanism appears to also support adjustments to natural L2 accent ([Xie et al., 2017](#)).

An additional—not mutually exclusive—accent accommodation mechanism that has been proposed is *criteria relaxation*. Similar to phonemic recalibration, a criteria relaxation mechanism posits changes to the listener’s linguistic categories; the criteria relaxation mechanism, however, allows the listener to accommodate speaker productions by expanding the variance of a category distribution [see discussion by [Zheng and Samuel \(2020\)](#)]. Note that recent work by [Xie et al. \(2023\)](#) would refer to the former as a “category shift” and the latter as a “category expansion,” with both falling under the larger umbrella of mechanisms that promote perceptual adaptation via changes to listeners’ perceptual categories. Evidence from [Babel et al. \(2021\)](#) supports the conclusion that phonemic recalibration and criteria relaxation mechanisms may both be involved in the process of accent accommodation. The authors examined accommodation in L1 English listeners for two specific accented phonemes (/z/-devoicing or /s/-voicing) and found that, while listeners were able to accommodate both, the more common of the two accented phonemes in North American English (/z/-devoicing) resulted in a directional phonemic recalibration, and the less common of the two accented phonemes (/s/-voicing) resulted in an expansion of the category. Together, these results indicated that both of these perceptual mechanisms may support listener accommodation of familiar and unfamiliar spoken language variation. In the present study, we do not aim to disentangle which mechanisms are responsible for the observed accent accommodation effects. Our focus is instead on whether social information modulates the accommodation that does occur. We do note, however, that in naive listeners (as in the present study), it is highly likely that a criteria relaxation mechanism supports early adaptation—particularly before the listener has received enough exposure to have learned the systematic qualities of the accent.

1.2 Social priming

A substantial body of work indicates that social information can be conveyed by a person’s speech, such that listeners can identify speaker characteristics, such as age, race, gender, sexual orientation, and social class ([Strand, 1999](#); [Labov, 1986](#); [Purnell et al., 1999](#); [Munson and Babel, 2007](#)). Furthermore, social cues from outside the acoustic signal can also affect speech processing. Examining comprehension of a short lecture (as measured with a cloze test), [Rubin \(1992\)](#) found that performance was poorer for participants who listened while viewing a picture of an East Asian face than when viewing a White face, despite the speech in both conditions being from the same American-accented speaker (see also [Kang and Rubin, 2009](#)). Although Rubin attributed this to listener biases against East Asian speakers, further work on this topic has indicated that race information can also facilitate perception of speech. For example, [McGowan \(2015\)](#) examined American listeners’ recognition accuracy of Mandarin Chinese-accented speech presented in babble for three guise conditions: an East Asian face, a White face, or control silhouette (i.e., with no race information). Results indicated that

performance was better for listeners randomly assigned to view the East Asian guise compared to the White guise (differences with the control prime were non-significant).

Thus, it appears that extralinguistic information may prime listeners to expect specific accent qualities, which can either facilitate or inhibit recognition accuracy. [Babel and Russell \(2015\)](#), for example, found that recognition accuracy for L1 Canadian-accented English speech presented in pink noise was reduced for Chinese-Canadian talkers when still images of the talkers were shown on screen (as compared to a fixation cross-only). This was not the case for the White-Canadian talkers presented in the same experiment, indicating that expectations of an L2 accent for the Chinese-Canadian talkers may have affected speech perception. Similarly, work by [Kutlu et al. \(2022\)](#) in American and Canadian listeners has shown reduced recognition accuracy following presentation of minority (South Asian) faces for both L1 Indian English and L1 British English.

Examinations of these “social priming” effects on perception of L2 accent have garnered mixed outcomes. [Hanulíková \(2021\)](#) explored the effects of East Asian and White faces on speech perception in teenage, young adult, and older adult L1 German listeners. For teens and older adults, Korean-accented German speech was more accurately perceived when presented with an East Asian face than a White face [matching the results of [McGowan \(2015\)](#)], while for young adults, there was no difference between priming conditions. [Melguy and Johnson \(2021\)](#) also examined the effects of multiple racial primes on perception of L2 accent, focusing on perception of Mandarin-accented English in young adult American L1 English listeners. However, the results of two experiments showed no main effects of any of the visual primes on transcription accuracy [deviating from the outcomes of [McGowan \(2015\)](#)].

To our knowledge, the only study that has specifically examined the effect of social information on adaptation to L2 accent (i.e., improvement across a task as opposed to overall performance) is [Vaughn \(2019\)](#). In that study, Vaughn compared L1 English listeners’ adaptation to an L2 Spanish-accented speaker following presentation of a verbal guise (i.e., a description of the talker). Three verbal guises were examined between subjects: (1) No guise (no information provided); (2) L1 accent guise (e.g., “the speaker’s first and only language is English, but his parents are Spanish-English bilingual speakers from Chile”); and (3) L2 accent guise (e.g., “the speaker’s first language is Spanish, his parents are from Chile, and he only began to learn English in school”). Results indicated improved overall recognition accuracy and greater adaptation for participants provided with language background information about the speaker. Surprisingly, participants given the L1 accent guise outperformed subjects given the L2 accent guise (despite the latter being the true description of the speaker).

1.3 A theoretical framework connecting social priming and accent accommodation

One framework that may account for the effects of social information on speech processing is exemplar theory ([Johnson, 1997](#); [Goldinger, 1998](#); [Pierrehumbert, 2001](#)), which proposes that phonetically detailed episodic traces (or “exemplars”) are stored in the lexicon. On such a view, recognition of the incoming acoustic signal is assisted by activation of acoustically similar exemplars in memory. Critical to the present study, it has also been proposed in prior work [see discussion by [Melguy and Johnson \(2021\)](#)] that the phonetic patterns of exemplars are linked to social information. In other words, listeners may create larger socio-phonetic categories by extrapolating across co-occurring phonetic and social patterns encountered throughout their lives. These socio-phonetic connections, thus, provide a framework by which top-down information can influence speech perception.

Based on an exemplar framework, one might predict that only experienced listeners would show a social priming effect. The role of familiarity with a given accent, and how that may or may not impact effects of social information on speech processing, has been explored in prior work. The seminal study by [McGowan \(2015\)](#) examined two groups of listeners: one with a low amount of prior experience with Mandarin-accented English and one with a high amount of prior experience with Mandarin-accented English. However, results of [McGowan \(2015\)](#) indicated that both experienced and inexperienced listeners benefitted equally from social information. One possible explanation for this outcome is that inexperienced listeners are able to prepare for an unknown accent based on social information by engaging a criteria relaxation mechanism. In other words, listeners may not need to know what accent to expect to benefit from social information because they are expanding their perceptual categories (as opposed to directionally shifting them). Note that this account assumes that even inexperienced listeners may broadly associate some races and ethnicities with L2 accent; indeed, work by [Zou and Cheryan \(2017\)](#) examining Americans’ stereotyping of racial and ethnic groups in the United States does appear to support this assumption (with results indicating that Asian, Latinx, and Middle Eastern Americans are stereotyped as having L2 accents, while Black and White Americans are not).

In the present study, we examine social priming effects in listeners who had little experience with the target accent. We believe that this design choice was justifiable based on the outcomes of [McGowan \(2015\)](#) but also note that re-examining the role of listener experience may prove informative in future work.

1.4 The present study

We aimed to investigate whether visually presented social primes can benefit L1 listeners during perceptual adaptation to unfamiliar, L2-accented speech. We hypothesized the following:

- (1) Listeners assigned to view East Asian faces during the multi-talker training session will have better transcription accuracy than subjects assigned to view White faces, given that Cantonese-accented speakers of English are more typically East Asian than White.
 - (a) Additionally, this benefit of the East Asian faces will be immediate (significant at trial 1).
- (2) Listeners who were trained with the East Asian faces will outperform subjects who were trained with the White faces in the post-test (novel talker) session.

We also expected that the East Asian primed group would do better in the post-test session than subjects in the control group (matching prior work; Bradlow and Bent, 2008). We did not have a prediction as to whether the White primed group would do better than the control group.

2. Method

Pre-registration of the experiment design and predictions is available from <https://osf.io/dr4m5>. Data and analysis scripts are available from <https://osf.io/5xmf4/files/>.

2.1 Participants

Young adult subjects (age mean = 26.6; age range = 18–35) were recruited using the website Prolific to participate in the experiment online. Separate postings were used to recruit the control group subjects and the training group subjects because the control group's session was substantially shorter (approximately 10 min, as compared to 25 min for the training group subjects). Both groups were paid at the same rate approved by the Washington University Institutional Review Board, rounding up to the nearest 15-min increment: \$2.50 for the control group, \$5.00 for the training groups. Inclusion criteria (set via Prolific's demographic filters) selected for White¹ young adults between 18 and 35 years old reporting English as their first and dominant language, currently residing in the United States and being of United States nationality, and having normal hearing and vision (or corrected-to-normal vision). Our recruitment plan and budget allowed up to a maximum of 300 subjects (100 per group), although we did recruit three replacement subjects to even out the final number of subjects per group after exclusions (90 per group). In total, 303 subjects participated in the experiment, 33 of which were excluded for one of the following reasons: failing to meet eligibility criteria (despite Prolific's pre-screening; 14), self-reporting the use of speakers instead of headphones for any task (two), failing attention-check trials in the speech transcription task (four), performing greater than or equal to three standard deviations away from the group average in the speech transcription task (ten), self-reporting that their data should be excluded (one), or software error (two).

2.2 Materials

During the training session, both training groups were presented with the same auditory stimuli featuring three female Cantonese-accented speakers of English. In the post-test session, a fourth female Cantonese-accented speaker was presented for all random assignment groups. All targets were Hearing-In-Noise-Test sentences (HINT; Nilsson *et al.*, 1994), retrieved from the SpeechBox corpus (Bradlow, 2009). The four target speakers were selected after piloting eight speakers to determine their relative intelligibility levels; the four selected talkers were similarly intelligible when presented in speech-shaped noise at a -2 dB signal-to-noise ratio (SNR). The speech-shaped noise was created in Praat (version 6.1.10; Boersma and Weenink, 2021) using the long-term average spectrum of the full set of target speaker files. Background noise was presented before and after the target sentence for 500 ms. Three additional audio files were also recorded by a native speaker of American English for attention-check trials. These files were recordings of the sentences "please type a single G," "please type a single Q," and "please type a single X" and were presented without background noise. The first two catch trials were presented during the training session, and the third was presented during the post-test session.

For the visual stimuli, pictures of three White and three East Asian females who were similarly rated for attractiveness, neutrality of expression, and high prototypicality of race were selected from the Chicago Face Database [Ma *et al.*, 2015; Fig. 1(A)]. The specific codes associated with the images were as follows: AF-246, AF-253, AF-255, WF-011, WF-233, and WF-244. Subjects randomly assigned to the East Asian priming condition were only shown the three East Asian female faces, while those assigned to the White priming condition were only shown the three White female faces. No prime was presented during the test session.

2.3 Procedure

Subjects were redirected from Prolific to the experiment hosted on Gorilla (Anwyl-Irvine *et al.*, 2020). Subjects in the training groups completed the following tasks (in order): a headphone screening, the training session, the post-test session, and the demographic and language questionnaire. The training session was the same for both training groups, with the exception of the race of the visual stimuli. Subjects recruited for the control group completed the following tasks (in order): a headphone screening, the post-test session, and the demographic and language questionnaire.

2.3.1 Headphone screening

The headphone screening was designed by Milne *et al.* (2021). Subjects are played three bursts of noise and have to determine which contained a hidden tone. Due to some issues with the screening mistakenly flagging subjects who use Apple EarPods, we

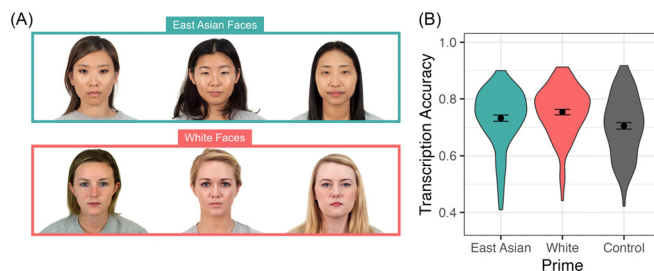


Fig. 1. (A) Images of the faces used for the East Asian and White priming conditions are shown. Images are sourced from the Chicago Face Database (Ma *et al.*, 2015). (B) By-subject transcription accuracy during the post-test session is summarized using proportion of words correctly transcribed. Violin plots show distribution of subjects' average performance and points show group means with standard error bars.

decided against using it to determine which subjects to exclude. However, many Prolific subjects drop out of the experiment at the headphone screening (presumably because they are not wearing headphones), which still made it a useful screening tool. A self-report question about use of headphones at the end of the task was used to exclude subjects from analyses instead.

2.3.2 Training session

Before beginning the training session, subjects were instructed to do the speech transcription task in a distraction-free environment, listen closely to each sentence, take their best guess if they did not understand the speaker, and use correct spellings (to the best of their abilities). Subjects were also reminded of the headphone requirement at this time. Additional instructions for the training session informed subjects that a photograph of each given speaker would appear on the screen before each trial and that they should not close their eyes during the task.

During the training session, each prime co-occurred with the same talker's voice consistently (20 trials per prime/talker), appearing on screen 2000 ms before the audio stimulus began. After the audio stimulus finished, a response box appeared on the screen for subjects to transcribe what they heard. A 2-s delay was inserted between trials. The session contained 62 trials total, including the two attention-check trials, which were set to occur approximately one-third and two-thirds of the way through the task. The presentation order of the target sentences was randomized across subjects. Halfway through the experiment, subjects were told that they could take a short break before continuing the task.

2.3.3 Post-test session

Instructions for the post-test session matched those for the training session, with the exception of the instructions about the images of the speakers. Subjects were told that during the post-test session, there would be nothing on the screen and that they would just be listening closely to the target sentences read aloud by a single speaker. The post-test session was the same for the control subjects and the trained subjects. Twenty-one trials were presented, including one attention-check trial set to occur halfway through the session. All target sentences presented during the post-test session were novel. No breaks were offered during the post-test session.

2.3.4 Demographic and language background questionnaire

After a reminder that responses would not affect pay for the study, a series of questions asked subjects to report their age, gender, race, ethnicity, hearing status, nationality, and languages. Last, subjects were asked to report if they thought that there was any reason that their data should be excluded from the study.

2.4 Data preparation

Keywords in the Hearing in Noise Test sentences were identified for each sentence. Specifically, determiners were excluded from scoring (e.g., "the," "a," "an," "his," "her"). Transcription accuracy for both the training and the post-test sessions was calculated with the R package *autoscore* (Borrie *et al.*, 2019). The scoring procedure allowed for differences in tense (i.e., "jump" vs "jumped") and plurality (i.e., "cat" vs "cats") and double-letter misspellings ("atack" would be acceptable for "attack" and "ocassion" for "occasion"). The tallied number of correct vs incorrect (missed) keywords per sentence was used for analyses. In the training session data, the effect of trial was renumbered from 0 to 59 (i.e., differences at the intercept represent differences at trial 1) and then scaled to assist with model convergence.

3. Results

Analyses of the training data are reported first, followed by analyses of the post-test data. For both datasets, we used the *glmer()* function from the *lme4* package in R to fit the data from the speech transcription task. The predicted variable, transcription accuracy, was treated as a grouped binomial. That is, each trial of the task (for a given subject) corresponded to a single row of the dataframe in R, and the predicted variable of the models was two columns: correctly identified target words and missed target words. A logit link function was specified for the models. Random effects included intercepts for

subjects and items and random slopes of trial and speaker by subject. Random slopes of condition by item were attempted but produced model singularity and, thus, were removed from all models.

3.1 Training results

Fixed effects in the model included prime (levels: East Asian, White; dummy-coded reference level: East Asian), trial, and the interaction between prime and trial. Log-likelihood model comparisons were used to determine whether each fixed effect significantly improved model fit. The effect of prime did not significantly improve model fit [$\chi^2(1) = 2.26, p = 0.13$; Fig. 2(A)]. The effect of trial did significantly improve fit [$\chi^2(1) = 45.26, p < 0.001$; Fig. 2(B)], indicating that subjects improved across the session, but the interaction between prime and trial was not significant [$\chi^2(1) = 0.80, p = 0.37$]; nor were differences between prime at the intercept.

3.2 Post-test results

For the post-test dataset, the only fixed effect examined was condition (levels: control group, East Asian primed group, White primed group). Our initial dummy-coding scheme set the control condition as the reference level, and in a separate model, we changed the reference level to the East Asian primed group (to directly compare the two training groups). A log-likelihood model comparison indicated that the effect of condition significantly improved model fit [$\chi^2(2) = 10.38, p = 0.006$]. When the control condition was set as the reference level, model estimates indicated that subjects assigned to the White prime in training performed significantly better than those in the control group ($\beta = 0.27, p = 0.001$), and those assigned to the East Asian prime training performed marginally better [$\beta = 0.15, p = 0.08$; Fig. 1(B)]. When the East Asian prime condition was set as the reference level, the model estimate for the difference between the East Asian and White prime training groups was not significant ($\beta = 0.13, p = 0.13$).

4. Discussion

The present study examined the effect of social information (i.e., perceived speaker race) on talker-independent perceptual adaptation to L2 accent. Participants who received a training session with East Asian faces vs those who received the same training session with White faces did not perform significantly differently during training or test. These results indicate that our matched-guise manipulation of speaker race did not affect perceptual adaptation to the Cantonese-accented talkers or the generalizability of perceptual learning to a novel talker of the same L2 accent.

Although the results of the present study indicated no effect of the social priming manipulation [see also results of young adult listeners in Hanulíková (2021) and Melguy and Johnson (2021)], there are notable differences between our sample and prior work that did find effects of social information on speech perception. In particular, we note that our sample differed from McGowan’s (2015) sample (i.e., focused on White listeners) and that, additionally, we focused exclusively on listeners who had little experience with the target accent. In particular—although it did not prove informative in McGowan (2015)—listener experience and/or interactions with racial/ethnic groups of interest may be a topic worth revisiting in future research. Babel and Russell (2015), for example, found that listeners’ social networks (i.e., time spent with Asian Canadians) affected their priming susceptibility. Thus, examining familiarity with the race/ethnicity used in priming manipulations, as opposed to (solely) familiarity with the target accent, may prove fruitful in future research.

In line with prior work (Bradlow and Bent, 2008), subjects who received training performed better in the post-test session than subjects given no training (i.e., the control group). This outcome indicates that listeners were accommodating the Cantonese-accented English speech during the training session and generalized this perceptual learning when they encountered a novel Cantonese-accented talker. Surprisingly, subjects assigned to the White prime training session performed significantly better than the control group, while subjects assigned to the East Asian prime training session only

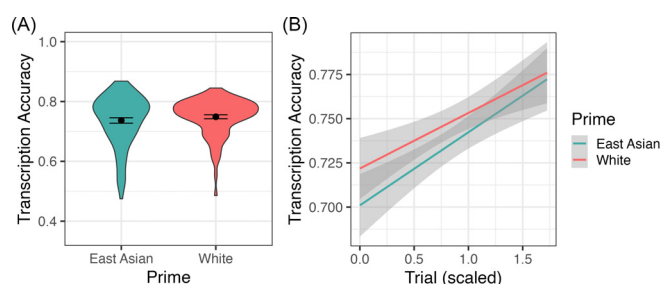


Fig. 2. Transcription accuracy during the training session is summarized using proportion of words correctly transcribed. The main effect of prime and the interaction of prime with trial were both non-significant. The main effect of trial indicated significant improvement across the training session. Note that y axis scales differ between plots. (A) Violin plots show distribution of subjects’ average performance, and points show group means with standard error bars. (B) Improvement across trials is shown for each condition with predicted fit lines and standard error ribbons.

performed marginally better than the control group. This trend is the opposite of what we predicted based on prior work that has found benefits of East Asian faces for the perception of Mandarin Chinese-accented English (McGowan, 2015).

As suggested in the Introduction, it is possible that accommodation of L2 accent is supported by phonemic recalibration (Xie *et al.*, 2017) and/or criteria relaxation (Zheng and Samuel, 2020; Babel *et al.*, 2021). Our results do not provide clear insights as to whether one or both of these mechanisms supported perceptual adaptation to Cantonese accent in the present study. However, they do indicate that adaptation was robust against an interference effect of social information. As noted above, both groups trained with the L2 accent performed better than a control group at test [as in Bradlow and Bent (2008)], which confirms that the exposure to the L2 accent promoted adaptation and generalization of perceptual learning. Thus, it does not appear to be the case that incongruent perceived speaker race inhibits the processes of adaptation or generalization. It remains to be determined whether the mechanism(s) supporting adaptation is equally robust under more naturalistic settings.

One limitation of the present study is that we focused exclusively on short-term effects of social information on perceptual adaptation. In future research, examining adaptation over longer timescales (i.e., weeks) may provide more ecological insights into how co-occurring accents and social cues are learned and leveraged during speech processing. It is possible that, particularly in listeners with limited exposure to a given accent, a prolonged training session would allow us to better examine how social priming develops in the listener. Additionally, by including evaluations of listener biases (in addition to performance measures), future work can examine whether prolonged accent trainings may afford the added benefit of reducing linguistic prejudice. Particularly in research examining accommodation of speech from L2 speakers of minority races and/or ethnicities, determining whether biases impact listener performance (e.g., via listener motivation) will be crucial to fully understanding the integration of social information during spoken language processing.

Acknowledgments

This work was supported by National Science Foundation Graduate Research Fellowship DGE-1745038 (to D.J.M.), National Science Foundation Doctoral Dissertation Research Improvement Grant No. 2116319 (to K.J.V.E. and D.J.M.), the Basque Government through the BERC 2022–2025 program, the Spanish State Research Agency through BCBL Severo Ochoa excellence accreditation CEX2020-001010-S, and the European Union's Horizon 2020 research and innovation programme under the Marie Skłodowska-Curie Grant Agreement No. 101103964 (PerMSAA).

Author Declarations

Conflict of Interest

The authors have no conflicts of interest to disclose.

Ethics Approval

The present study received ethics approval from the Washington University in St. Louis Internal Review Board.

Data Availability

Data, scripts, and materials are available from <https://osf.io/5xmf4/files/>.

References

- ¹The present study comes from a dissertation of work that examined similar social priming effects. A preceding experiment indicated that social priming effects may only be present in White listeners. Thus, in the present study White subjects were specifically recruited. We anticipated that this may increase the likelihood of finding an effect of racial primes.
- Anwyl-Irvine, A. L., Massonnié, J., Flitton, A., Kirkham, N., and Evershed, J. K. (2020). "Gorilla in our midst: An online behavioral experiment builder," *Behav. Res.* **52**, 388–407.
- Babel, M., Johnson, K. A., and Sen, C. (2021). "Asymmetries in perceptual adjustments to non-canonical pronunciations," *Lab. Phonol.* (published online).
- Babel, M., and Russell, J. (2015). "Expectations and speech intelligibility," *J. Acoust. Soc. Am.* **137**(5), 2823–2833.
- Baese-Berk, M. M., Bradlow, A. R., and Wright, B. A. (2013). "Accent-independent adaptation to foreign accented speech," *J. Acoust. Soc. Am.* **133**(3), EL174–EL180.
- Bent, T., and Baese-Berk, M. M. (2021). "Perceptual learning of accented speech," in *The Handbook of Speech Perception*, edited by J. S. Pardo, L. C. Nygaard, R. E. Remez, and D. B. Pisoni (Wiley, New York), pp. 428–464.
- Boersma, P., and Weenink, D. (2021). "Praat: Doing phonetics by computer (version 6.1.10) [computer program]," <http://www.praat.org> (Last viewed December 13, 2021).
- Borrie, S. A., Barrett, T. S., and Yoho, S. E. (2019). "Autoscore: An open-source automated tool for scoring listener perception of speech," *J. Acoust. Soc. Am.* **145**(1), 392–399.
- Bradlow, A. R. (2009). "SpeechBox," <https://speechbox.linguistics.northwestern.edu> (Last viewed September 1, 2020).
- Bradlow, A. R., and Bent, T. (2008). "Perceptual adaptation to non-native speech," *Cognition* **106**(2), 707–729.
- Brown, V. A., McLaughlin, D. J., Strand, J. F., and Van Engen, K. J. (2020). "Rapid adaptation to fully intelligible nonnative-accented speech reduces listening effort," *Q. J. Exp. Psychol.* **73**(9), 1431–1443.
- Choi, J. Y., and Perrachione, T. K. (2019). "Time and information in perceptual adaptation to speech," *Cognition* **192**, 103982.

- Clarke, C. M., and Garrett, M. F. (2004). "Rapid adaptation to foreign-accented English," *J. Acoust. Soc. Am.* **116**(6), 3647–3658.
- Goldinger, S. D. (1998). "Echoes of echoes? An episodic theory of lexical access," *Psychol. Rev.* **105**(2), 251–272.
- Hanulíková, A. (2021). "Do faces speak volumes? Social expectations in speech comprehension and evaluation across three age groups," *PLoS One* **16**(10), e0259230.
- Johnson, K. (1997). "Speech perception without speaker normalization: An exemplar model," in *Talker Variability in Speech Processing*, edited by K. Johnson and J. Mullenix (Academic, New York), pp. 145–166.
- Kang, O., and Rubin, D. L. (2009). "Reverse linguistic stereotyping: Measuring the effect of listener expectations on speech evaluation," *J. Lang. Soc. Psychol.* **28**(4), 441–456.
- Kapadia, A. M., and Perrachione, T. K. (2020). "Selecting among competing models of talker adaptation: Attention, cognition, and memory in speech processing efficiency," *Cognition* **204**, 104393.
- Kutlu, E., Tiv, M., Wulff, S., and Titone, D. (2022). "Does race impact speech perception? An account of accented speech in two different multilingual locales," *Cogn. Res.* **7**(1), 7.
- Labov, W. (1986). "The social stratification of (r) in New York City department stores," in *Dialect and Language Variation* (Academic, New York), pp. 304–329.
- Ma, D. S., Correll, J., and Wittenbrink, B. (2015). "The Chicago Face Database: A free stimulus set of faces and norming data," *Behav. Res.* **47**(4), 1122–1135.
- Magnuson, J. S., and Nusbaum, H. C. (2007). "Acoustic differences, listener expectations, and the perceptual accommodation of talker variability," *J. Exp. Psychol. Hum. Percept. Perform.* **33**(2), 391–409.
- McGowan, K. B. (2015). "Social expectation improves speech perception in noise," *Lang. Speech* **58**(Pt. 4), 502–521.
- McLaughlin, D. J., Colvett, J. S., Bugg, J. M., and Van Engen, K. J. (2023). "Sequence effects and speech processing: Cognitive load for speaker-switching within and across accents," *Psychon. Bull. Rev.* (published online).
- McLaughlin, D. J., and Van Engen, K. J. (2020). "Task-evoked pupil response for accurately recognized accented speech," *J. Acoust. Soc. Am.* **147**(2), EL151–EL156.
- Melguy, Y. V., and Johnson, K. (2021). "General adaptation to accented English: Speech intelligibility unaffected by perceived source of non-native accent," *J. Acoust. Soc. Am.* **149**(4), 2602–2614.
- Milne, A. E., Bianco, R., Poole, K. C., Zhao, S., Oxenham, A. J., Billig, A. J., and Chait, M. (2021). "An online headphone screening test based on dichotic pitch," *Behav. Res.* **53**, 1551–1562.
- Mullenix, J. W., Pisoni, D. B., and Martin, C. S. (1989). "Some effects of talker variability on spoken word recognition," *J. Acoust. Soc. Am.* **85**(1), 365–378.
- Munro, M. J., and Derwing, T. M. (2020). "Foreign accent, comprehensibility and intelligibility, redux," *J. Second Lang. Pronunciation* **6**(3), 283–309.
- Munson, B., and Babel, M. (2007). "Loose lips and silver tongues, or, projecting sexual orientation through speech," *Lang. Linguist. Compass* **1**(5), 416–449.
- Nilsson, M., Soli, S. D., and Sullivan, J. A. (1994). "Development of the Hearing in Noise Test for the measurement of speech reception thresholds in quiet and in noise," *J. Acoust. Soc. Am.* **95**(2), 1085–1099.
- Norris, D., McQueen, J. M., and Cutler, A. (2003). "Perceptual learning in speech," *Cogn. Psychol.* **47**(2), 204–238.
- Pierrehumbert, J. (2001). "Exemplar dynamics: Word frequency, lenition and contrast," in *Frequency and the Emergence of Linguistic Structure* (John Benjamins, Amsterdam), pp. 137–157.
- Porretta, V., Buchanan, L., and Järvikivi, J. (2020). "When processing costs impact predictive processing: The case of foreign-accented speech and accent experience," *Atten. Percept. Psychophys.* **82**(4), 1558–1565.
- Purnell, T., Idsardi, W., and Baugh, J. (1999). "Perceptual and phonetic experiments on American English dialect identification," *J. Lang. Soc. Psychol.* **18**(1), 10–30.
- Rubin, D. L. (1992). "Nonlanguage factors affecting undergraduates' judgments of nonnative English-speaking teaching assistants," *Res. High. Educ.* **33**(4), 511–531.
- Samuel, A. G., and Kraljic, T. (2009). "Perceptual learning for speech," *Atten. Percept. Psychophys.* **71**(6), 1207–1218.
- Sidaras, S. K., Alexander, J. E., and Nygaard, L. C. (2009). "Perceptual learning of systematic variation in Spanish-accented speech," *J. Acoust. Soc. Am.* **125**(5), 3306–3316.
- Strand, E. A. (1999). "Uncovering the role of gender stereotypes in speech perception," *J. Lang. Soc. Psychol.* **18**(1), 86–100.
- Van Engen, K. J., and Peelle, J. E. (2014). "Listening effort and accented speech," *Front. Hum. Neurosci.* **8**, 577.
- Vaughn, C. R. (2019). "Expectations about the source of a speaker's accent affect accent adaptation," *J. Acoust. Soc. Am.* **145**(5), 3218–3232.
- Xie, X., Jaeger, T. F., and Kurumada, C. (2023). "What we do (not) know about the mechanisms underlying adaptive speech perception: A computational framework and review," *Cortex* **166**, 377–424.
- Xie, X., Theodore, R. M., and Myers, E. B. (2017). "More than a boundary shift: Perceptual adaptation to foreign-accented speech reshapes the internal structure of phonetic categories," *J. Exp. Psychol. Hum. Percept. Perform.* **43**(1), 206–217.
- Zheng, Y., and Samuel, A. G. (2020). "The relationship between phonemic category boundary changes and perceptual adjustments to natural accents," *J. Exp. Psychol. Learn. Mem. Cogn.* **46**(7), 1270–1292.
- Zou, L. X., and Cheryan, S. (2017). "Two axes of subordination: A new model of racial position," *J. Pers. Soc. Psychol.* **112**(5), 696–717.