AMERICAN LISTENERS' RECOGNITION OF SENTENCES UNAFFECTED BY RACIAL AND ETHNIC PRIMES

Drew J. McLaughlin[1] & Kristin J. Van Engen[2]

[1]Basque Center on Cognition, Brain and Language; [2]Washington University in St. Louis
[1]d.mclaughlin@bcbl.eu; [2]kvanengen@wustl.edu

ABSTRACT

A listener's ability to accurately understand speech can be affected by racial and ethnic information about the speaker. For example, the presentation of an East Asian versus a White face has been shown to lead to better understanding of Mandarin-accented English and poorer understanding of American- or Canadian-accented English. This social priming effect has yet to be examined with images of Latinx or Black faces for Standard American English (SAE) in American listeners. We used a matched-guise paradigm to present listeners with Black, East Asian, Latinx, and White primes (images of faces), as well as a control prime (a blurred silhouette), during transcription of SAE-accented sentences.

Results indicated no effect of the priming manipulation. In particular, the lack of an East Asian prime effect on recognition accuracy differs from prior work. We discuss the possibility that these effects may be dependent on characteristics of specific social contexts.

Keywords: speech perception; social priming; race and ethnicity

1. INTRODUCTION

Spoken language carries paralinguistic information such as cues to the speaker's race, gender, and social class [1]. For example, American listeners can identify a non-standard American English dialect from a snippet of speech as short as "hello" [2]. It is important to note that the "standard" dialect of a region is typically determined by which group(s) of speakers hold the most power [3]. As such, in the United States, the Standard American English (SAE) dialect primarily reflects the dialect(s) of White (majority race) speakers. There are also multiple other dialects spoken by communities of minority racial and ethnic groups in the United States. The best documented of these is African American Language (AAL), which differs from SAE grammatically [4] and phonologically [5]. Other dialects appear to be borne from second language (L2; "foreign") accents, such as Asian-influenced English and Spanish-influenced English. Most notable for the present study, these dialects differ from AAL because they are less homogeneous, representing influences from many different L2 accents and languages.

Capturing this difference among AAL, Spanish-influenced, and Asian-influenced dialects, a recent racial position model of Americans [6] indicates that Asian and Latinx Americans are typically stereotyped as more *culturally foreign* (a term encompassing multiple attitudes and stereotypes about immigration, as well as having a foreign accent) than White and Black Americans. Thus, while some Americans hold negative stereotypes toward Black Americans, they do not appear to be associated with foreign accents in the same way that Asian and Latinx Americans are.

1.1. Social Priming

Prior work indicates that a listener's ability to understand speech may be affected by racial and ethnic information about the speaker (here forward referred to as *social priming*). One study, for example, found reduced recognition accuracy for L1 Canadian-accented English speech (presented in pink noise) when still images of Chinese-Canadian talkers were presented (as compared to presenting the audio with a fixation cross only). This same reduction in accuracy did not occur when images of White-Canadian talkers were presented [7]. However, complementing this work, [8] examined American listeners' recognition accuracy for Mandarin Chinese-accented speech (presented in babble). Listeners were assigned to images of an East Asian, White, or control (i.e., a silhouette) face, and results indicated that the East Asian prime *facilitated* speech recognition compared to a White prime (differences with the control prime were non-significant). Most notably, when examined in conjunction with the results of [7], it appears that minority race/ethnicity primes do not always negatively affect speech perception. Instead, such cues may prime listeners to expect specific dialect/accent qualities, which can either facilitate or inhibit perception.

This interpretation aligns with an exemplar model of speech recognition [9, 10, 11], in which phonetically-detailed episodic traces ("exemplars") are stored in the lexicon and linked to social information. Across a listener's life, patterns are extrapolated across to create larger social categories that are linked to phonetic (and higher-order linguistic) patterns. These socio-phonetic connections thus provide a framework by which top-down information can influence speech perception.

However, while the results of [7] and [8] complement each other, recent work has not consistently revealed "negative" social priming effects for standard, L1-accented speech. In L1 German listeners of various ages, the effects of East Asian and White visual primes on recognition of standard L1-, non-standard (regional) L1-, and L2- (Korean-) accented speech were mixed [12]. Korean-accented German presented with an East Asian prime was more accurately perceived by teens and older adults (not younger adults), but there were no effects of priming for L1-accented speech.

1.2. Study Aims and Hypotheses

In the present study, we examined potential effects of racial and ethnic primes on perception of SAE speech for White L1 American listeners. We thus attempt to conceptually replicate the work of [7] in American listeners and extend it by examining additional racial and ethnic primes (Black and Latinx, in addition to East Asian and White). Our modified experiment includes a full matched-guise design between visual primes and SAE voices.

Based on the results of [7], we predicted that listener recognition accuracy for the East Asian priming condition would be poorer than the White priming condition. For Latinx priming condition, we predicted a similar outcome, given that Latinx and Asian Americans are similarly stereotyped as having foreign accents [6] and Spanish-influenced accents are prominent in the United States.

For the Black priming condition, however, we did not have a predicted outcome. Based on an exemplar model of speech perception, one could expect L1 American listeners to associate Black faces with AAL. This top-down effect on speech processing could result in poorer speech recognition accuracy given the present study's use of SAE stimuli. However, it is also possible that social priming effects are primarily driven by associations between racial/ethnic groups and L2 accents. In this case, one may expect no negative priming effect for the Black condition, because Black Americans are not associated with cultural foreignness in the same way as Asian and Latinx Americans [6].

2. METHODS

Data collection occurred online during Fall 2020. We acknowledge that the generalizability of our findings may be limited by the extenuating circumstances of the COVID-19 pandemic. Approval for the present study was acquired from the Washington University in St. Louis Institutional Review Board.

2.1. Participants

White young adult participants (*N* = 126) with normal hearing were recruited to participate online from the Washington University in St. Louis Psychology Participants Pool using SONA Systems. From this sample, 25 participants were excluded from analyses for one or more of the following reasons: Reporting that English was not a language learned from birth, reporting that they did not use headphones during the task, reporting that their data should be excluded (e.g., "I was too distracted"), and data collection error. This resulted in a final sample of 101 participants (age mean = 19.11; age *SD* = 1.03). All participants received course credit as compensation. Seventy-two participants reported that they were female and 29 reported that they were male.

2.2. Stimuli

Auditory stimuli included 60 unique sentence-length items collected from six separate corpora stored in the Northwestern SpeechBox [14] database. All recordings were from L1 speakers of Standard American English. One additional male was recorded in-lab to bring the numbers to an even 30 male voices and 30 female voices. The sentences were all semantically normal and contained 3-7 (mean = 4.62) common keywords and 4-10 (mean = 7.08) total syllables each. All words were included in the keywords count with the exception of the determiners "a" and "the".

All auditory stimuli were levelled to 65 dB sound pressure level using the UCLA Phonetics Lab's intensity scaling Praat script [15]. Speech-shaped noise was created using the average long-term spectrum of the entire batch of target files, and then combined with each file at a -8 dB signal-to-noise ratio (SNR). During the experiment, speech-shaped noise began two seconds before sentence onset and continued two seconds after sentence offset.

In order confirm that the speakers' accents were perceived to be SAE, we conducted an online pilot. Fourteen young adult subjects (none overlapping with the present study) rated the stimuli on a scale from 0 to 10, where 0 indicated "not at all standard" and 10 indicated "extremely standard". Stimuli were presented in speech-shaped noise at -8 dB SNR as described above. Overall, the mean rating of the stimuli was 6.56 (*SD* = 1.44). We explore whether the variation in these accent ratings affects social priming in the analyses of the main dataset.

Two additional recordings were created for attention-check trials. These recordings were of the same female voice saying either “Please type a single G” or “Please type a single Q.” These catch trials were presented without background noise in combination with a female control prime.

For the visual stimuli, images of faces were selected from the Chicago Face Database [16]. Six female and six male faces were selected for each race and ethnicity (White, Black, East Asian, and Latinx; 48 faces total) based on norming data. The faces were all rated highly for prototypicality of race/ethnicity and approximately matched for age, attractiveness, and happiness. Using these images, two control images were also created. Webmorph [17] was used to combine all of the female faces and (separately) all of the male faces into two morphed images. Next, Adobe Photoshop [18] was used to blur the images. First the face was selected (i.e., from hairline to chin, and from ear to ear) with a lasso tool and blurred with a 120 pixel-radius Gaussian blur, and then the entire image was blurred with a 40 pixel-radius Gaussian blur. This resulted in a silhouette image that was ambiguous in terms of race and ethnicity.

2.3. Procedure

The experiment was hosted on Gorilla [19], an online experiment-building platform.

The task began with a consent document and instructions. Participants were asked to complete the experiment in one sitting with their full attention. For the speech transcription task, they were told to look at the picture on the screen while listening, and type the full sentence when prompted. Instructions indicated that headphones were required for the task. Before beginning, an example file was presented for participants to adjust their volume as needed. The file could be replayed multiple times. The voice in the example did not occur in the experiment.

The transcription task contained 62 trials (60 test items and two attention checks). The attention-check stimuli occurred at fixed points one-quarter and three-quarters through the task. Halfway through the task, a self-timed break was offered to participants. The sentences-in-noise and image primes were randomly paired across subjects, with the exception that female faces were always assigned to female voices and male faces were always assigned to male voices. With the exception of the male and female control images, none of the face images repeated during the task. Audio stimuli also did not appear more than once. The order of presentation of each audio file and prime was also randomized across subjects. Thus, if any files were inadvertently harder than others, this challenge was equally represented for all priming conditions across subjects. Within each trial, the image was on screen for the full length of the audio file. Thus, the priming image appeared when the speech-shaped noise began, allowing two seconds of exposure before sentence onset.

After the transcription task, subjects completed two additional unrelated pilot tasks. Demographic and language information was collected at the end of the experiment. Additionally, after a reminder that their responses would not affect their compensation, subjects were prompted to report if they used headphones (as instructed) and if their data should be excluded for any reason (e.g., “I was not paying enough attention”).

2.4. Analyses

Transcription accuracy was determined using the R package Autoscore [20]. Common misspellings predetermined by the Autoscore package were scored as correct, and differences in tense (*kick* versus *kicked*) and plurality (*cat* versus *cats*) were allowed. Generalized linear mixed-effects regression (GLMER) was used to model the data. The predicted variable, transcription accuracy, was treated as a grouped binomial of correctly identified keywords versus incorrectly identified (or missed) keywords. For example, if for the target sentence *“the **gray mouse ate** the **cheese**”* a subject transcribed *“the **gray** house had **cheese**”*, this would be scored as two keywords correct and two keywords incorrect/missing. Although there are multiple keywords per sentence/trial, the GLMER model is nonetheless able to predict the counts of the correct versus incorrect/missing groups of data using a binomial regression.

Random intercepts by subject and item (i.e., audio file) were specified, and random slopes of prime were included by subject.

3. RESULTS

The fixed effect of prime was dummy-coded with the Control condition as the reference level. The log-likelihood comparison of GLMER models with and without the effect of prime indicated that it did not significantly improve model fit ($\chi^2 = 3.06$, $df = 4$, $p = .55$). As shown in Figure 1, accuracy did not differ across priming conditions.

We also examined the effect of sex (i.e., male versus female speaker/prime), which significantly improved model fit ($\chi^2 = 5.89$, $df = 1$, $p = .02$). Participants had better transcription accuracy for the female voices/primes than the male voices/primes ($\beta = -1.30$, $p = .01$). The effects of prime and sex did not significantly interact ($\chi^2 = 0.71$, $df = 4$, $p = .95$).

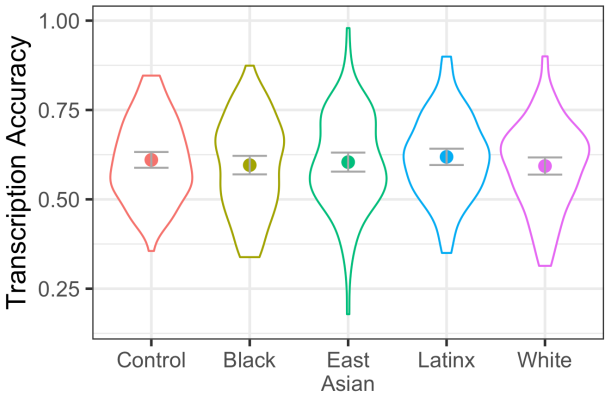

Figure 1: The non-significant effect of prime on transcription accuracy is visualized with violin distributions, mean points, and 95% confidence intervals.

Lastly, we explored whether individual differences among the speakers' voices may influence social priming. Mean standardness ratings of each stimulus (where 0 indicated "not at all standard" and 10 indicated "extremely standard") from the pilot data were added to the GLMER model with the effect of prime. Model fit was significantly improved by the effect of ratings ($\chi^2 = 37.61$, $df = 1$, $p < .001$) but not the interaction between ratings and prime ($\chi^2 = 3.37$, $df = 4$, $p = .50$). The direction of the effect of ratings indicated that higher (more standard) rated stimuli were also more intelligible ($\beta = 1.01$, $p < .001$).

4. DISCUSSION

Speakers with non-standard accents can face prejudice and stigmatization that impacts their everyday lives, including their ability to secure housing and employment [21, 22]. For second language- (L2-) accented speakers these negative effects of listener attitudes are also present, such that reduced comprehensibility of their speech affects how listeners perceive their intelligence, among other traits [23].

Prior work indicates that another challenge to successful communication may be perceived "incongruencies" between a speaker's race or ethnicity and their accent. However, these negative social priming effects have not been observed consistently across listener groups [12]. The present study focused on the effects of visually-presented racial and ethnic primes on the perception of SAE speech. Our stimuli included Black, East Asian, Latinx, and White primes, as well as a control image.

To our surprise, no significant effects of the priming images emerged. Presentation of different racial and ethnic primes did not affect listeners' recognition accuracy for SAE-accented sentences presented in noise. One methodological explanation may be the effect of the noise on listeners' ability to identify each speaker's accent. Indeed, our exploration of accent ratings indicated that more intelligible stimuli were also rated as sounding more standard, suggesting that listeners may be less able to identify a speaker's accent under adverse listening conditions. Similar results were found in [24], which showed that adding noise to the speech signal affected judgments of L1 speakers, such that listeners rated them as sounding more L2-accented. In other words, it is possible that by degrading the speech signal with noise, the present study reduced the possibility of observing a negative social priming effect. However, the initial study to observe this effect [7] presented sentences in pink noise at an exceptionally difficult signal-to-noise ratio (speech was only approximately 20% intelligible). It is possible that the findings of [7] are specific to the area in which the research was conducted (Vancouver, B.C., Canada). A key future direction for work examining social priming effects will be the influence of listeners' attitudes and implicit expectations of foreign and non-standard accents. The present work examined listeners who were White, young adult, Americans. Although the participants were all recruited from a university in the Midwestern United States, it is likely that their individual linguistic experiences vary. By examining individual differences in social networks as well as exposure to dialect and accent variation, future work may be able to determine the role of previous experience on social priming.

To the best of our knowledge, the present study is the first to examine Black and Latinx primes for American listeners and SAE accent. We hypothesized that, based on an exemplar model, the White listeners in our sample may associate Black faces with AAL and Latinx faces with Spanish-influenced English. If this were the case, a "negative" social priming effect could occur, given the use of SAE stimuli. Our results did not reveal this to be the case. It remains unclear whether social priming effects for standard L1 accent, as documented in [7], occur for additional minority race and/or ethnicity primes for American listeners.

5. SUMMARY

A listener's ability to understand speech can be affected by racial and ethnic information about the speaker. We examined this social priming effect using images of Black, East Asian, Latinx, and White faces (as well as a control silhouette) and SAE accent in White American listeners. Results indicated no effect of the priming manipulation. We suggest that social priming effects may only occur in specific populations of listeners and recommend examination of individual listener differences in future work.

6. ACKNOWLEDGMENTS

This research was supported by Doctoral Dissertation Research Improvement Grant 2116319 from the National Science Foundation.

7. REFERENCES

[1] Kraus, M. W., Torrez, B., Park, J. W., & Ghayebi, F. (2019). Evidence for the reproduction of social class in brief speech. *Proceedings of the National Academy of Sciences, 116*(46), 22998-23003.

[2] Purnell, T., Idsardi, W., & Baugh, J. (1999). Perceptual and phonetic experiments on American English dialect identification. *Journal of language and social psychology, 18*(1), 10-30.

[3] Lippi-Green, R. (2012). English with an accent: Language, ideology, and discrimination in the United States. *Routledge.*

[4] Cukor-Avila, P., & Bailey, G. (1995, June). Grammaticalization in AAVE. *In Annual Meeting of the Berkeley Linguistics Society* (Vol. 21, No. 1, pp. 401-413).

[5] Edwards, W. F. (2008). African American Vernacular English: phonology. *Varieties of English, 2,* 181-191.

[6] Zou, L. X., & Cheryan, S. (2017). Two axes of subordination: A new model of racial position. *Journal of personality and social psychology, 112*(5), 696.

[7] Babel, M., & Russell, J. (2015). Expectations and speech intelligibility. *The Journal of the Acoustical Society of America, 137*(5), 2823-2833.

[8] McGowan, K. B. (2015). Social expectation improves speech perception in noise. *Language and Speech, 58*(4), 502-521.

[9] Johnson, K. (1997). The auditory/perceptual basis for speech segmentation. *Working papers in linguistics-Ohio State University Department of Linguistics*, 101-113.

[10] Goldinger, S. D. (1998). Echoes of echoes? An episodic theory of lexical access. *Psychological review, 105*(2), 251.

[11] Pierrehumbert, J. B. (2001). Exemplar dynamics: Word frequency. *Frequency and the emergence of linguistic structure, 45*(137), 10-1075.

[12] Hanulíková, A. (2021). Do faces speak volumes? Social expectations in speech comprehension and evaluation across three age groups. *PloS one, 16*(10), e0259230.

[13] McLaughlin, D. J., & Van Engen, K. (under review). Social priming: Exploring the effects of speaker race and ethnicity on perception of nonnative accents.

[14] Bradlow, A. R. (n.d.) *SpeechBox.* Retrieved from https://speechbox.linguistics.northwestern.edu

[15] Vicenik, C. (2022, October 1). *Noise and Speech Manipulation*. Praat Script Resources. http://phonetics.linguistics.ucla.edu/facilities/acoustic/praat.html

[16] Ma, D. S., Correll, J., & Wittenbrink, B. (2015). The Chicago face database: A free stimulus set of faces and norming data. *Behavior research methods, 47*(4), 1122-1135.

[17] Lisa DeBruine. (2017, December 4). *Webmorph* (Version v0.0.0.9001). Zenodo. doi: 10.5281/zenodo.1073696

[18] Adobe Systems. (2002). *Adobe Photoshop 7.0*. Adobe Press.

[19] 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*(1), 388-407.

[20] Borrie, S. A., Barrett, T. S., & Yoho, S. E. (2019). Autoscore: An open-source automated tool for scoring listener perception of speech. *The Journal of the Acoustical Society of America, 145*(1), 392-399.

[21] Carlson, H. K., & McHenry, M. A. (2006). Effect of accent and dialect on employability. *Journal of employment counseling, 43*(2), 70-83.

[22] Purnell, T., Idsardi, W., & Baugh, J. (1999). Perceptual and phonetic experiments on American English dialect identification*. Journal of language and social psychology, 18*(1), 10-30.

[23] 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.

[24] Gittleman, S., & Van Engen, K. J. (2018). Effects of noise and talker intelligibility on judgments of accentedness. *The Journal of the Acoustical Society of America, 143*(5), 3138-3145.