HOW ACCENTED DO CAUCASIAN-LOOKING VS. ASIAN-LOOKING NATIVE ENGLISH SPEAKERS SOUND TO A JAPANESE LISTENER?

Marzena Karpinska

The University of Tokyo
marzena.karpinska@gmail.com

ABSTRACT

Speech perception involves the integration of acoustic features with socioindexical information (e.g., speaker's ethnicity). English speakers, for instance, may rate English utterances produced by a native speaker as more accented if they believe the speaker to be Asian than if they believe the speaker to be Caucasian. This study investigates whether Japanese speakers will similarly rate English utterances as more accented when made to believe that the speaker was Asian-looking. Forty-eight participants rated the accentedness (9-point Likert-scale) of a set of sentences spoken by native English speakers. Sixteen listeners were made to believe the speakers were Caucasian, sixteen that the speakers were Asian, and sixteen were assigned to the audio-only condition. The results indicate no effect of ethnicity on accent rating by the non-native raters. Hence, non-native (Japanese) speakers may be less influenced by the speaker's perceived ethnicity when evaluating the level of English accent as spoken by native speakers.

Keywords: Speech perception, accentedness, socioindexical cues, ethnicity, exemplar theory

1. INTRODUCTION

Speech perception can be affected by the perceived age [9], gender [17], sexual orientation [11] or ethnicity [13] of the speaker. For instance, native American English listeners who listened to an audio recording of an American English female, while simultaneously being presented with a picture of an Asian female, rated her speech as more accented than those who listened to the same recording while being presented with a picture of a Caucasian female [16]. Similarly, native English listeners rated non-native English speakers from Korea as more accented when they were listening to an audio accompanied by a freeze frame featuring the Korean speaker than when they were only listening to the audio-only recording of the same speaker [19]. These findings suggest that the level of accentedness as rated by native English listeners may be affected by speaker's perceived ethnicity cued by a picture of either Asian or Caucasian face [1]. However, these studies focused only on native English listeners. Hence, it is unclear whether non-native listeners will be similarly affected by the speaker's perceived ethnicity when presented with native English utterances. This study aims to investigate whether native Japanese listeners would rate native English utterances as more accented when presented with a video of an Asian-looking speaker than when presented with a video of a Caucasian-looking speaker.

2. MODELS

Rubin [16] explains the effect of perceived ethnicity on speech perception in terms of a *negative bias*, renamed later as *reverse linguistic stereotyping* (RLS) [7]. He suggests that listeners hold negative social bias against, for instance, Asian-looking English speakers, and that this bias leads to a negative social evaluation and perception of a foreign accent even though it is not present.

Zheng and Samuel [20], on the other hand, argue that this effect may take place not on the *perception* level but on the *interpretation* level. They demonstrated that simply presenting pictures of Asian or Caucasian faces introduces demand characteristics, that is participants believing that they know the purpose of the experiment may alter their accentedness ratings. This effect, however, is mostly not present when the static picture is changed to a dubbed video.

Native and non-native listeners agree on accentedness ratings when evaluating audio-only English stimuli [3], that is when the ethnicity of the speaker is unknown. This remains true even for non-native listeners who have no familiarity with the rated language [12]. Therefore, if

[1]The term "ethnicity" is used here to indicate that a group of people has in common some certain racial traits and it is a common way of referring to "Asian" or "Caucasian" groups in sociolinguistics.

Japanese participants also demonstrate a bias towards "American = Caucasian" association then, according to the RLS model, we would expect them to act similarly to native English speaker, that is to rate English utterances as *more* accented when presented with a an Asian-looking speaker than when presented with a Caucasian-looking speaker. If, on the other hand, this effect of ethnicity was achieved due to the demand characteristics and the decision process really takes place on the *interpretation* rather than *perception* level, then we should see results similar to [20], that is no effect of speaker's ethnicity when the audio is being presented with a dubbed video stimuli.

3. METHOD

3.1. Design

Japanese listeners were divided into 3 different groups (between subject design) where they completed two rating tasks (within subject design). The first task, which will be referred to here as the *baseline* condition, consisted of audio-only stimuli and was the same for all groups in order to ensure that all the participants were initially performing the rating task in a comparable way. The second task, which will be referred to as the *experimental* condition, used the same audio files for each group, however, these audio files were either combined with different visual cues (video of an Asian speaker or video of a Caucasian speaker) or presented as audio-only stimuli. This is a variation of the matched guise technique [8] in the way that it presents the same auditory stimuli with different visual cues (the *guise*). Apart from the perception task participants were asked to complete an Implicit Association Test [5] in order to measure their bias towards "American=Caucasian" association.

3.2. Participants

Forty-eight Japanese native speakers (24 males and 24 females) were recruited in the Tokyo area. They were aged between 18 and 35 years old (mean = 22.5, SD = 4.5) with no reported history of visual or hearing impairment. All participants assessed their overall English level as pre-intermediate or above, or reported having passed an English certification exam on an intermediate level (CEFR B1). None of the participants stayed or lived abroad longer than 1 year, with a mean of 2.6 months (SD = 3.6 months).

3.3. Stimuli

Audio stimuli. Ten native English speakers from North America (5 males and 5 females) were recorded, in a soundproof booth, telling a short picture story similar to the "Suitcase story" [4]. In addition, they were asked to introduce themselves and to talk about their day. Ten sentences or phrases were extracted from each recording and the intensity was adjusted to 70 dB using Praat 6.0.16 [2]. An echo effect was added with Adobe Premiere Pro CC 7.0 in order to make the samples sound more natural when matched with the videos. The 100 audio files were randomly assigned to either *baseline* (40 utterances, 2 male and 2 female voices) or *experimental* (60 utterances, 3 male and 3 female voices) condition.

Video stimuli. The 60 utterances chosen for the *experimental* condition were combined with videos to prepare the stimuli for each of the experimental groups:

Asian group Six Asian-looking speakers (3 males and 3 females, 4 native English speakers and 2 highly proficient non-native speakers) were asked to repeat the sentences recorded for the experimental stimuli (10 sentences each) while these sentences were played in the background. A special attention was given to the lip movement. The best attempt was then chosen and the original audio was replaced with the experimental audio stimuli.

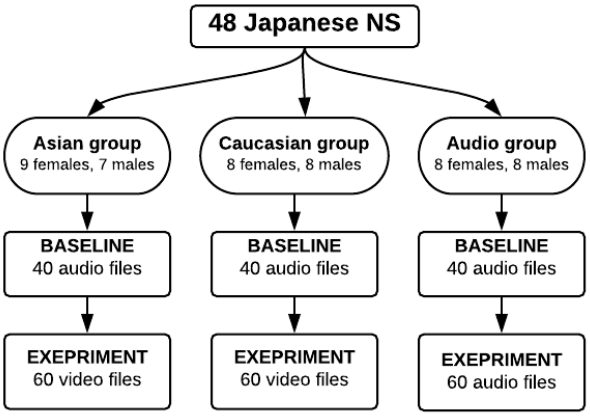
Caucasian group The same procedure as for the Asian group was followed to record Caucasian-looking native English speakers (3 males and 3 females).

Audio group In this group the experimental audio files were used without any visual cues or without any other editing.

All video editing was done using Adobe Premiere Pro CC 7.0. Speakers for the video recordings were all dressed in white t-shirts and stood in front of a white wall. The quality of the video files was evaluated by native and non-native English speakers both of whom failed to notice any mismatch between the audio and video.

IAT stimuli. Ten pictures of Asian (5 females) and Caucasian (5 females) black and white faces were used to represent the Asian and Caucasian categories. Similarly, eight places and symbols were chosen to represent the American and Japanese attributes. The stimuli was similar to the one used in [5, 19], however, the places and symbols were presented as words rather than pictures to match the requirements of the FreeIAT 1.3.3 software on which the experiemnt was administrated [14].

Figure 1: The general design of the experiment. Forty-eight Japanese native speakers was divided into 3 groups and completed two rating tasks, here *baseline* and *experiment*.



3.4. Procedure

Speech Perception. Participants were randomly assigned to 1 of 3 different groups: Asian group, Caucasian group, or Audio group. They were instructed to look at the screen and to listen to each utterance which was proceeded by a fixation cross and a beep sound in order to draw the listener's attention. After listening to each utterance they rated its' accentedness on a 9-point Likert scale (1 - non-native speaker, 9 - native speaker) and proceeded to the next item. A 9-point Likert scale was employed since it was demonstrated to be the most appropriate for accentedness judgments [3].

Participants in each group rated the stimuli in two different conditions separated by a break: (1) *baseline* (40 utterances) and (2) *experimental* (60 utterances). Half of the participants were presented with the *experimental* condition prior to the *baseline* condition. Figure 1 shows the general flow of the experiment.

Implicit Association Test. Upon completing the perception experiment participants were given an Implicit Association Test (IAT). They were asked to classify as fast as possible a set of pictures of Asian and Caucasian faces along with a set of words representing the concept of American and Japanese [6, 5] by pressing a key on the keyboard. In the congruent category Caucasian face and the concept of American (place or symbol) shared the same response key, while in the incongruent category Asian face was paired with the concept of American (place or symbol).

4. RESULTS AND DISCUSSION

The research question in this study was whether non-native listeners would perceive native English utterances as more accented (less native-like) when they believe that the speaker is Asian than when they believe that the speaker is Caucasian. Moreover, the participants were given an Implicit Association Test to measure the strength of their "American=Caucasian" association.

IAT. One participant was excluded from the IAT analysis due to having latency lower than 300 ms for more than 10% of the trails [6]. The IAT scores of Japanese participants were, on average, higher than zero (M=0.68, SD=0.20, $t(46)=23.57$, $p<0.001$) indicating overall preferences towards the "American=Caucasian" pairing.

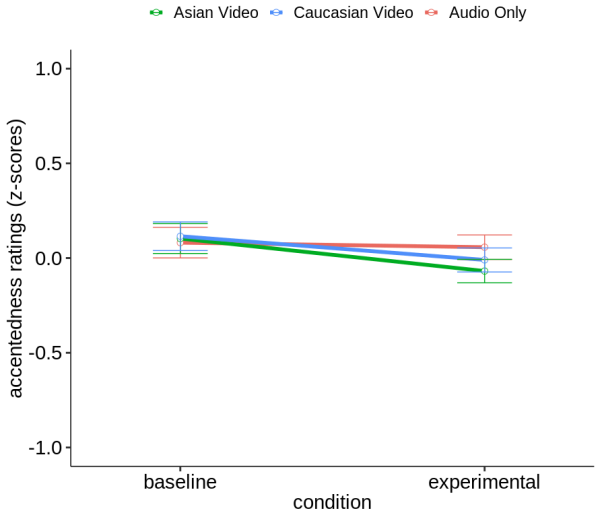
In order to explore, whether this bias did, as would the RLS model predict, lead to lower (more accented) accentedness ratings the raw accentedness ratings were converted into z-scores as advised in [18]. The plotted data in Fig. 2 indicate that there is presumably no difference between the groups in the *baseline* condition (about 0.01 to 0.03 point difference between the means). This suggests that participants in this study were rating the level of accentedness in English utterances similarly when presented with the exact same audio-only stimuli in the *baseline* condition.

The ratings in the *experimental* condition, though arguably further apart, also seem to be comparable (about 0.07 to 0.13 point difference between the means). Overall, listeners appear to be rating the utterances in the *experimental* condition as slightly more accented than the utterances in the *baseline* condition. This effect is visible across all groups regardless of the speaker's ethnicity and could be due to some acoustic features of individual speakers in the *experimental* condition, which made them "sound" less native-like when compared to the other speakers recorded for this study.

In order to determinate whether there is, indeed, no effect of the speaker's perceived ethnicity (i.e., all groups were assigning comparable accentedness ratings in the *experimental* condition) the data were analyzed with R language [15] using the linear mixed effects model implemented with the *lme4* function [1]. Group (Asian video, Caucasian video, and audio-only), Condition (*baseline* and *experimental*) and their interaction were all modeled as fixed effects. Moreover, the self-reported English level was also added to the model as a fixed effect in order to investigate whether the English proficiency had any effect on the accentedness ratings. Random

intercepts for participant and item embedded in the speaker were included along with by-participant random slopes for the effect of Condition. Using z-scores instead of the raw data as the response variable allowed also to avoid potential issues with non-normally distributed residuals. All p-values were obtained using the *anova* function [10] which provides the Satterthwaite's approximation of the degrees of freedom.

Figure 2: The mean ratings for each group in the *baseline* and *experimental* condition converted into z-scores with 95% confidence intervals. Lower z-score indicates that the utterances were rated as more accented.



The self-reported English level was not significant ($F(6, 39) = 0.35$, $p = 0.91$) indicating that participants in this study were rating the perceived accentedness level of the native English utterances similarly regardless of their English proficiency. Moreover, the Group x Condition interaction was also *not* significant ($F(2, 45) = 0.32$, $p = 0.72$) suggesting that the participants were assigning comparable accentedness ratings not only in the *baseline* condition but also in the *experimental* condition. All other effects were also not significant (all p's > .05).

Native and non-native listeners, even those non-native listeners who do not know the given language, generally agree when evaluating the level of foreign accent in audio-only stimuli [3, 12]. This claim was also confirmed by the lack of significance of the self-reported English level in the current study. It seems, therefore, that non-native listeners are capable of rating native utterances similarly as native listeners, even though they may not be proficient in the given language. Moreover, Japanese listeners demonstrated an "American=Caucasian" bias. Hence, base on the predictions of the RLS model, one might have expected that the Japanese listeners would act like the native English listeners in [16], that is they would rate videos with Asian faces as more accented than videos with Caucasian faces. However, that was not the case. Japanese participants in the current study did not rate the dubbed videos with an Asian face as being more accented than the dubbed videos with the Caucasian face. This finding is consistent with [20], where merely changing the type of stimuli from pictures to videos eliminated the initial face effect showing that native English listeners were performing the accentedness ratings based on a later *interpretation* rather than the actual *perception*. In this sense, it seems that native Japanese listeners were rating the utterance based only on their actual *perception*.

5. CONCLUSION

This study evaluated whether the believed ethnicity of the speaker would affect the accentedness ratings of native English utterances judged by non-native listeners from Japan. The results of the current experiment suggest that Japanese participants were *not* affected by the ethnicity of the speaker presented in a dubbed video when assigning the accentedness ratings to native English utterances on a 9-point Likert scale. These results differ substantially from the earliest studies where native English listeners rated native English utterances as more accented when presented with a static Asian face than when presented with a static Caucasian face. However, these results are consistent with a more recent study where the native English listeners were presented with dubbed videos of Asian and Caucasian speakers.

Non-native listeners rated the level of accentedness similarly to native listeners when presented with audio-only utterances. Moreover, non-native listeners showed a strong preference towards "American=Caucasian" pairing. However, they did not rate a video of an Asian speaker as more accented than a video of a Caucasian speaker. Therefore, the results do not support the RLS model, but they are consistent with the theory presented in [20]. Native English listeners in Rubin's study might have altered their ratings due to a later *interpretation* rather then *perception*. In contrast, non-native listeners in the current study seem to be performing the rating task based only on their *perception* without further *interpretation*.

6. REFERENCES

[1] Bates, D., Mächler, M., Bolker, B., Walker, S. 2015. Fitting linear mixed-effects models using lme4. *Journal of Statistical Software* 67(1), 1–48.

[2] Boersma, P., Weenink, D. 2010. Praat: Doing phonetics by computer [computer program]. version 6.0.16, retrieved 6 April 2016 from http://www.praat.org/.

[3] Derwing, T. M., Munro, M. J. 2009. Putting accent in its place: Rethinking obstacles to communication. *Language Teaching* 42(04), 476.

[4] Derwing, T. M., Rossiter, M. J., Munro, M. J., Thomson, R. I. 2004. Second Language Fluency: Judgments on Different Tasks. *Language Learning* 54(4), 655–679.

[5] Devos, T., Banaji, R. M. 2005. American = White? *Journal of Personality and Social Psychology* 88(3), 447–466.

[6] Greenwald, A. G., Nosek, B. A., Banaji, M. R. 2003. Understanding and using the Implicit Association Test: I. An improved scoring algorithm. *Journal of Personality and Social Psychology* 85(2), 197–216.

[7] Kang, O., Rubin, D. L. 2009. Reverse linguistic stereotyping: Measuring the effect of listener expectations on speech evaluation. *Journal of Language and Social Psychology* 28(4), 441–456.

[8] Kircher, R. 2015. The matched-guise technique. In: Hua, Z., (ed), *Research Methods in Intercultural Communication*. John Wiley & Sons, Inc. 196–211.

[9] Koops, C., Gentry, E., Pantos, A. 2008. The effect of perceived speaker age on the perception of PIN and PEN vowels in Houston, Texas. 14(2), 12.

[10] Kuznetsova, A., Brockhoff, P. B., Christensen, R. H. B. 2017. lmerTest package: Tests in linear mixed effects models. *Journal of Statistical Software* 82(13), 1–26.

[11] Levon, E. 2007. Sexuality in context: Variation and the sociolinguistic perception of identity. *Language in Society* 36(04), 533–554.

[12] Major, R. C. oct 2007. Identifying a foreign accent in an unfamiliar language. *Studies in Second Language Acquisition* 29(04).

[13] McGowan, K. B. 2015. Social Expectation Improves Speech Perception in Noise. *Language and Speech* 58(4), 502–521.

[14] Meade, A. W. Nov. 2009. FreeIAT: An Open-Source Program to Administer the Implicit Association Test. *Applied Psychological Measurement* 33(8), 643–643.

[15] R Core Team, 2018. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing Vienna, Austria.

[16] Rubin, D. L. 1992. Nonlanguage factors affecting undergraduates' judgments of nonnative English-speaking teaching assistants. *Research in Higher Education* 33(4), 511–531.

[17] Strand, E. A. Mar. 1999. Uncovering the Role of Gender Stereotypes in Speech Perception. *Journal of Language and Social Psychology* 18(1), 86–100.

[18] Yarnold, P. R., Soltysik, R. C. 2013. Ipsative Standardization is *Essential* in the Analysis of Serial Data. *Obtimal Data Analysis* 2(2), 94–97.

[19] Yi, H.-G., Phelps, J. E. B., Smiljanic, R., Chandrasekaran, B. 2013. Reduced efficiency of audiovisual integration for nonnative speech. *The Journal of the Acoustical Society of America* 134(5), EL387–EL393.

[20] Zheng, Y., Samuel, A. G. 2017. Does seeing an Asian face make speech sound more accented? *Attention, Perception, & Psychophysics* 79(6), 1841–1859.