Effects of Perceived Race on Comprehensibility, Intelligibility, and Accentedness Ratings

Bradford J. Lee
Fukui University of Technology
Justin L. Bailey
Fukui University of Technology

Reference Data:

Lee, B. J. & Bailey, J. L. (2022). Effects of perceived race on comprehensibility, intelligibility, and accentedness ratings. In P. Ferguson, & R. Derrah (Eds.), *Reflections and new perspectives*. JALT. https://doi.org/10.37546/JALTPCP2021-29

Reverse linguistic stereotyping (RLS) is the phenomenon whereby listeners' speech perception is altered by information (or assumptions) about a speaker (e.g., gender, age, social class, ethnicity, nationality). However, while listeners tend to downwardly appraise speakers' accents and comprehensibility when they perceive them to be of a different language community, past research has generally relied solely on Likert scale data. The current study suggests that qualitative investigation is also necessary to validly interpret raters' responses. After collecting ratings of accentedness, comprehensibility, language proficiency, and intelligibility from 223 Japanese university students listening to a Japanese speech under a matched-guise, students reported what factors influenced their ratings. Listeners rated the speech of the Caucasian guise as significantly more accented. A key finding was that while comprehensibility ratings were equal, they were qualitatively different. Caucasian guise members cited pronunciation significantly more than the Asian-guise group, who reported grammatical and lexical factors as informing their rating decisions.

逆言語的ステレオタイプ（RLS）とは、話者に関する情報（性別、年齢、社会階級、民族、国籍など）により、聞き手の音声知覚が変化する現象のことである。過去の研究によると、聞き手が異なる言語コミュニティの話者と認識した場合、話者のアクセントや理解力を低く評価する傾向がある。本研究では、評価者の回答を的確に解釈するためには、リッカート尺度法に加えて質的な調査も必要であると考え、matched-guise techniqueを用いて、223名の日本人大学生が日本語のスピーチを聞き、アクセント、理解度、言語能力と明瞭度の評価を集めた後、どのような要因が評価に影響を与えたかを調査した。結果からみると、白人系の話し手はとりわけアクセントが強いと評価され、アジア系の話し手は文法的、語彙的な要因で低い理解度に繋がったという質的な違いがあることが分かった。

All aspects of human language are heavily laden with information, not only the semantic message encoded in the lexis, but pragmatic messages and emotive aspects such as attitude and mood. While much of this encoding is done purposefully (though not always consciously) by a speaker, certain linguistic features, such as accents, can transmit information about the speaker unintentionally. In fact, listeners often make inferences about a speaker based solely on their speech variety, a phenomenon known in the literature as *linguistic stereotyping* (see Gluszek & Dovidio, 2010; Trofimovich et al., 2019, for social status; Russo et al., 2017, for aptitude; and Anderson et al., 2007, for intelligence).The current study sought to investigate the converse phenomenon, called *reverse linguistic stereotyping* (RLS) (Kang & Rubin, 2009), where prior information about a speaker is thought to alter a listener's perception of that person's speech.

As pronunciation is typically the first linguistic feature that listeners attend to (Carlo, 2013), many students and researchers have understandably been invested in understanding its mechanisms and effective training methods (see Saito & Plonsky, 2019, for a meta-analysis of pronunciation teaching techniques and their effects). Research on accented speech typically refers to *accentedness* (the degree to which a speech variety differs from a set point of reference), *comprehensibility* (how much effort a listener exerts to understand speech), and *intelligibility* (the degree to which speech is accurately decoded) (e.g., Derwing & Munro, 1997; Munro & Derwing, 1995). These components must be considered independently, as highly accented speech does not necessarily result in low intelligibility, particularly when the listener has a high degree of exposure to the target variety (Huang, 2013).

Review of the Literature

RLS describes a process in which listeners ascribe (often stereotypical) linguistic traits of a certain language community to an individual before hearing them speak. In a seminal report by Rubin (1992), it was shown that a person's perception of speech

could be influenced by images they are presented with visually. He used a mixed-guise test and had American undergraduate students listen to a recorded lecture while being shown a photograph depicting the lecturer as either Caucasian or Chinese. In fact, the lecturer was a Caucasian speaker of L1 American English. Despite this, the students who were shown the Chinese guise reported that the lecturer was more accented and less comprehensible than the group who were under the Caucasian guise. Other researchers using the mixed-guise test have also found evidence suggesting RLS exists (e.g., Zheng & Samuel, 2017), with Hu and Su (2015) also reporting that not only attitudes, but also listening comprehension performance was disparate between a Caucasian and Asian guise.

Evidence has also been reported that simply priming listeners by writing terms such as "Canadian" or "Australian" on materials before allowing them to listen to speech samples significantly influences perception (e.g., Niedzielski, 1999; Hay, Nolan, & Drager, 2006). Participants in these studies reported the presence of non-existent phonemes that were typical of the indicated speech communities, which formed the basis of the exemplar-model of speech perception (Hay, Nolan, & Drager, 2006). Essentially, this model proposes that the human brain creates individual files for every interlocutor we interact with and stores their unique patterns of speech. This makes it possible to identify a voice on the other end of a telephone call, or to 'hear' a loved one's voice inside our minds. Then, when we interact with these individuals again, the brain recalls these speech patterns to more quickly and effortlessly process their speech in real time. Researchers posit that "activating a social category raises the activation level of exemplars associated with that category" (Hay, Nolan, & Drager, 2006, p. 371) which makes it more likely that a listener will erroneously hear non-existent phonemes in the speech stream. Essentially, the brain's evolutionary mechanisms which helped us become more efficient at processing speech have also made us more susceptible to misdirection.

RLS has received significant attention since its inception and has been highlighted as an important factor in perceptions of gendered speech (e.g., Johnson, et al., 1999; Strand, 1999), age (Hay, Warren, & Drager, 2006), and perceived social class (Hay, Warren, & Drager, 2006). However, previous studies have thus far been limited to numerical ratings of comprehensibility based on a Likert scale, which offer little insight into why speakers were up/downgraded. The current study sought to address this gap in the literature by asking listeners to also report problem areas which they felt impacted the speaker's level of comprehensibility, allowing for a more accurate interpretation of comprehensibility scores. Therefore, the current project sought to answer the following research questions (RQs):

RQ1: To what degree, if any, does the perceived race of a speaker have on Japanese listeners' ratings of comprehensibility, accentedness, and language proficiency?

RQ2: Do the listeners cite different factors which affect comprehensibility depending on the perceived race of a speaker?

RQ3: What effect, if any, does a speaker's perceived race have on the intelligibility of the speech among Japanese listeners?

Study Design

Participants

The participants were all students at a private university in a rural area of western Japan. It should be noted that the rural location and the absence of an English major at the university place the students in an EFL context. Recruitment efforts resulted in obtaining 240 volunteers including 17 non-Japanese participants. However, their data was not included in the study and instead held back for future use. The result was a population of $N = 223$ native Japanese L1 speakers, ranging in age from 19 to 21 years old. There were $n = 192$ males (86.1%) and $n = 31$ females (13.9%), nearly equivalent to the gender ratio of the university's student body. This ratio should therefore not be considered as indicative of a sampling bias.

Materials and Methods

A recording was created of a Japanese man reciting a passage (in Japanese) from "夜中の汽笛について、あるいは物語の効用について" [Concerning the Sound of a Train Whistle in the Night or On the Efficacy of Fiction], (Murakami, 1995). This passage recounted a conversation between a young couple as the man describes (through a series of metaphors) how much he loves the woman. The passage was selected as the topic was considered relevant and engaging for the target population (e.g., university students). The recording was 2 minutes 28 seconds long, which was considered long enough for the participants to form their ratings without the influence of listener fatigue.

The reader, a man in his early 50s, was a local resident of the same city as the university; therefore, his accent was similar to the one the students (even those who came from outside of the prefecture) encountered daily in the community. This individual volunteered to be the audio source as he had both a linguistics background and extensive public speaking experience and indicated confidence in his oratory skills. The recording was made in the first author's studio using an Audio-Technica AT2035

cardioid condenser microphone with a shock mount and pop filter, connected via USB to a laptop computer running the open-source audio editor Audacity. This software was also used to normalize peak amplitude, in line with L2 speech research standards (Derwing & Munro, 1997).

Two photographic guises were prepared: Group C was shown a Caucasian male with light brown hair and hazel eyes, while Group J was shown a Japanese man with black hair and dark brown eyes. Neither photograph was of the actual speaker. Both photographs were professional portraits, with similar seated poses against solid backgrounds, looking directly at the camera and smiling. The models were of similar age (in their late 40s) and wearing similar clothing.

After participants gave informed consent to participate in the study, they were told that they would listen to a short, 2.5-minute audio passage. The researchers displayed either the Caucasian or Japanese guise photo on a computer screen for the duration of the audio passage, without any explanation. After the audio was played through once, the participants were given a four-item questionnaire which asked them to rate the speaker's comprehensibility, accentedness, and level of Japanese ability on a 5-point Likert scale (e.g., 1: completely incomprehensible; 5: completely comprehensible). Item Four was a multiple-choice item asking students to indicate any applicable factors which they felt had hindered their comprehension of the speech. Students also had the option of choosing *I had no problems at all* or writing novel responses that were not listed.

When all participants had completed the questionnaire, the audio was played a second time while participants completed a cloze test. This test consisted of eight gaps in the script (four nouns, one adjective, three verbs) and was used to quantify intelligibility scores. Only perfect responses (i.e., all phonemes correct) were scored as correct; any errors or omissions were counted as incorrect.

Results

Datasets were created in the Statistical Package for the Social Sciences (SPSS), Version 23, by which all statistical calculations were conducted. Descriptive statistics were first calculated for each variable, with means and standard deviations presented in Table 1.

Table 1

Descriptive Statistics of Results

	Group C			Group J		
	N	*M*	*SD*	*N*	*M*	*SD*
Comprehensibility	108	3.92	1.19	115	3.85	.96
Accentedness	108	4.06	.77	115	4.57	.61
Proficiency	108	4.38	.68	115	4.30	.77
Q1	106	.92	.27	114	.95	.22
Q2	106	.96	.19	114	.92	.27
Q3	106	.94	.23	114	.96	.21
Q4	106	.99	.10	114	.97	.16
Q5	106	.96	.19	114	.95	.22
Q6	106	.99	.10	114	.98	.13
Q7	106	.97	.17	114	.95	.22
Q8	106	.99	.10	114	.98	.13

Note. Group C: Caucasian male; Group J: Japanese male.
Means for Section One are based on a 5-point Likert scale (0: negative, 5: positive).
Means for Section Two are based on a binary scale (0: incorrect, 1: correct).
Discrepancies in *N* reflect three participants who did not complete Section Two

Perceived Race of Speaker

A series of *t*-tests were run to determine if there were statistical differences in the means between groups in any of the variables. As shown in Table 2, below, the only significant difference in scores was regarding the perceived accentedness of the speaker. This difference was statistically significant, $t(221) = -5.59$, $p < .001$ to a medium effect size of $d = -.75$ (for a discussion of L2 field-specific interpretations of effect size, see Plonsky & Oswald, 2014). None of the other variables revealed significant differences, with *p*-values over .05, and 95% *CI* values straddling zero.

Table 2

Inferential Statistics (independent samples t-tests) Between Group C and Group J

	95% *CI*	*t*	*p*	*d*
Comprehensibility	-.22, .35	.45	.66	.06
Accentedness	-.70, -.34	-5.59**	< .001	-.75**
Proficiency	-.12, .27	.77	.44	.10
Q1	-.09, .04	-.69	.49	-.09
Q2	-.02, .10	1.29	.20	.17
Q3	-.07, .05	-.43	.67	-.06
Q4	-.02, .05	.93	.35	.13
Q5	-.04, .07	.53	.60	.07
Q6	-.02, .04	.52	.61	.07
Q7	-.03, .08	.91	.36	.12
Q8	-.02, .04	.52	.61	.07

Note. ** indicates significance at the < .001 level

The results presented in Tables 1 and 2 address our first RQ, which asked to what degree, if any, does a speaker's perceived race impact listeners' ratings of comprehensibility, accentedness, and language proficiency. When the Japanese listeners viewed a person of a different race (the Caucasian guise, in this instance), they assigned higher ratings of accentedness to the speech compared to when viewing someone with Japanese facial features. However, ratings of comprehensibility or language proficiency were statistically equal under the same guises.

Comprehensibility Factors

Responses to Item Four of the survey are reported in Table 3, including relative frequencies of each response, chi-square values, statistical significance, and strengths of association (Cramer's *V*).

Table 3

Responses to Item Four (What problems did you have in comprehending the speaker?)

	Group C	Group J	Chi-Square	*p*	Cramer's *V*
Speed	8	12	.74	.39	.05 [-.07, .18]
Pronunciation	32	3	29.08**	< .001	.34 [.23, .45]
Grammar	8	24	8.60**	.003	.19 [.06, .30]
Vocabulary	5	15	5.09*	.02	.14 [.02, .26]
No Problems	67	70	.01	.92	.01 [-.11, .13]
Other	2	2	0	.97	< .005 [-.12, .13]
Total *N*	122	126	-	-	

Note. Participants could make multiple selections, resulting in a larger *N* than other tables.
* denotes significance at the < .05 level.
** denotes significance at the < .001 level.

As reported in Table 3, a chi-square test of independence indicated that there was significant association between group membership and responses in the categories of pronunciation, grammar, and vocabulary. Group C reported feeling that pronunciation was a hindrance to comprehension at a greater degree than Group J, X^2 (1, N = 248) = 29.08, p < .001, at a moderate strength of association, V = .34. On the other hand, Group J reported issues of grammar, X^2 (1, N = 248) = 8.60, p = .003, and vocabulary, X^2 (1, N = 248) = 5.09, p = .02, significantly more than Group C, but at a low strength of association (V = .19 and .14, respectively).

RQ2, which asked if listeners cite different factors that affect comprehensibility depending on the perceived race of a speaker, can therefore be answered in the affirmative. Despite the comprehensibility ratings for both groups being statistically equivalent (p = .66; see Table 2), qualitatively different rationales were provided to support these ratings. Group C attributed their comprehension difficulty mainly to issues relating to the speaker's pronunciation, while Group J seemed to attend more to issues of lexis and grammar. This is a salient finding, as previous research into the comprehensibility of accented speech typically reports only numerical scores and not listeners' justifications. The implications of this finding, which will be discussed further in the next section, are that study designs that do not include this reporting are unable to be accurately interpreted as they may lead to conclusions of false equivalency.

Intelligibility

RQ3 asked, to what effect, if any, does a speaker's perceived race have on intelligibility. As reported in Table 2, no significant differences were detected between the groups in terms of listening accuracy on the cloze test. We, therefore, answer this RQ by concluding that in the current study, perceived race had no significant impact on the intelligibility of the speech.

Discussion and Pedagogical Implications

In partial replication of the main findings of Rubin (1992), the current study found that under a mixed-guise test, Japanese listeners rated a speaker as being more accented when viewing a Caucasian face than when they were shown an Asian one (though Rubin's results were reversed due to his experiment having been conducted in English). These results contribute to the literature by providing support to both the construct of RLS and that of the exemplar-model of speech perception in a language other than English. They also fell largely in line with expectations. Even though Japanese people tend to have positive attitudes toward English speakers and praise their usage of Japanese (Neustupny, 1986), this does not mean that they believe them to be equal to native speakers on a phonetic level.

Even more salient are the findings related to comprehensibility. As previously detailed, though comprehensibility ratings were statistically equivalent between groups, they were shown to be derived from completely distinct linguistic factors. As the current study demonstrated that a Likert scale is insufficient for interpreting listeners' ratings of comprehensibility, we propose that a paradigm shift is needed whereby study designs elicit listeners' subjective reasons alongside their ratings. While this requires that results from past studies be reexamined, it also opens up new possible avenues of investigation by probing further into listeners' cognition. This suggested retooling of study designs is timely and comes at a moment when researchers have also advocated for the addition of *listener engagement* to the established paradigm of accentedness, comprehensibility, and intelligibility in research on accented speech (e.g., Lee, 2022). Now 30 years after Rubin's (1992) study on RLS, it is a positive sign that the field continues to develop and expand into new territories as our understanding continues to evolve.

Limitations and Future Directions

While the current study attempted to address many of the weaknesses in previous studies of accented speech thus far discussed, it is not without its own limitations. For one, we are unable to quantify the participants' attitudinal orientation towards Caucasian speakers of Japanese and their associated accents, which may have interacted with their intelligibility scores through the variable of listener engagement. In future studies, we plan to correct this by incorporating both data on attitudinal orientation and levels of listener engagement. This would allow us to get a fuller picture of the participants' cognition and potential biases, similar to how the current study examined students' cognition regarding comprehensibility. This relates closely to the limitation of physical location. As explained in the subsection on participants, the study was conducted in a rural area of Japan. This means that due to the low number of foreign residents in the area, the participants have had much less exposure to non-native Japanese speakers than students at other, more urban locations would have. This reduced exposure to foreign residents (and thus, Caucasians), has two salient implications. First, sociolinguistic theories (e.g., Tajfel & Turner, 1979; Giles & Johnson, 1987) posit that in multicultural contexts, group associations often form along language community lines, typically with one group enjoying institutionally-backed status over others. However, the power balance is somewhat unclear in the current context. Due to the generally low numbers of non-native Japanese speakers in the area of the current study, the only Caucasians the participants meet regularly are likely to be English teachers and professors. As these professions are generally regarded as having high prestige, particularly from the viewpoint of language students, this may have had an influence on their subjective ratings. This idea is supported by Reid et al. (2019), who similarly demonstrated that positive attitudes towards a particular language community typically results in more favorable ratings.

The second implication involves the participants' degree of exposure to Caucasians' Japanese speech patterns. As discussed previously, the exemplar-model of speech perception is based on the idea that people store all instances of speech as examples of a particular person or group's way of speaking. This implies that the more exposure a person has to a speech pattern, the more samples of speech they store, and the easier perception of that speech becomes (see also Huang, 2013 for empirical evidence on accent familiarity). However, as it can be postulated that the participants of the current study had far less exposure to Caucasians speaking Japanese than residents of a more urban area with a larger Caucasian population, it is not possible to accurately quantify this variable. It should also be noted that the exemplar-model of speech perception is only a theory that cannot be definitively proven.

Finally, the study was limited in the use of only two guises. In future studies, we plan to expand on these findings by including further guises to tease apart the factors

of ethnicity. As previously mentioned, Caucasians, many of whom come from native English-speaking countries, are often seen in positions of power, such as language teachers or celebrities. English is also the main foreign language in Japan, which further increases the status of many Caucasians, who are typically assumed to be American or British citizens (see also Lee & Bailey, 2020). Incorporating non-white guises, such as those from Southeast Asia, Africa, or even neighboring East Asian countries such as China and Korea, might give different results due to differences in perceived power balance and social status. In addition, the use of a control group where students are asked to rate a speech sample without any visual image, would provide a standard to which the effects of RLS could be compared.

Conclusion

Over the past 30 years, studies across multiple fields have investigated how speech perception can be influenced by assumptions about a speaker's characteristics (RLS). This has largely been explained psychologically through sociolinguistic models of intercultural communication, which essentially describe people's desire to identify with their own language community by creating distance with others. However, the actual linguistic mechanics of how speech perception can be affected has been more recently addressed by the exemplar-model, which suggests that the brain's evolutionary processes which helped speed up speech perception might also have made it more susceptible to errors or deception (Hay, Nolan et al., 2006).

The current study provided evidence of RLS in a language other than English (e.g., Japanese) in a population of 223 Japanese university students who were asked to provide ratings of accentedness, comprehensibility, language proficiency, and intelligibility of a recorded Japanese speech. The study used a mixed-methods approach and matched-guise study design, which also illustrated the weakness of the standard practice of only reporting Likert scale scores. Our findings suggested that the matched-guise test resulted in significantly different ratings of accentedness, with the Caucasian guise group scoring the speech as significantly more accented. At the same time, ratings for comprehensibility, language proficiency, and intelligibility were statistically equal.

The unique aspect of the study was the elicitation of qualitative data, whereby participants reported what factors influenced their comprehensibility ratings. Our study showed, for the first time to our knowledge, that while ratings were statistically equal, the way in which these ratings were arrived at was qualitatively very different. Factors such as accent and pronunciation played a large role in the students' decisions regarding the Caucasian-guise, whereas grammatical and lexical factors largely informed the listeners in the Asian guise group. This clearly demonstrates that interpretation of Likert scores is impossible without such explicit data, making a strong case for including this information in future studies on accented speech.

Bio Data

Bradford J. Lee (ORCID ID 0000-0001-9833-5631) holds an Ed.D. in TESOL from Anaheim University, a M.A. in Applied Linguistics from the University of Newcastle, and a B.A. in Linguistics from the University of Hawaii. He is currently Director of the International Center and an Associate Professor in the Organization for Fundamental Education at Fukui University of Technology. His main research areas include phonology/pronunciation instruction, noticing/perception-based instruction, and smartphone-based writing. <bradford-lee@fukui-ut.ac.jp>

Justin L. Bailey (ORCID ID 0000-0003-1983-821X) holds a M.A. in Applied Linguistics from the University of Leicester and a B.Sc. in Psychology from Aston University. He is currently working as an Assistant Professor in the Organization for Fundamental Education at Fukui University of Technology. His research areas include test washback and English as a world language. <j-bailey@fukui-ut.ac.jp>

References

Anderson, S., Downs, S. D., Faucette, K., Griffin, J., King, T., & Woolstenhulme, S. (2007). How accents affect perception of intelligence, physical attractiveness, and trust-worthiness of Middle-Eastern-, Latin-American-, British-, and Standard-American-English-accented speakers. *The BYU Undergraduate Journal of Psychology, 3*(1), 1-7. https://scholarsarchive.byu.edu/intuition/vol3/iss1/3

Derwing, T. M., & Munro, M. J. (1997). Accent, intelligibility, and comprehensibility: Evidence from four L1s. *Studies in Second Language Acquisition, 19*(1), 1-16. https://doi.org/10.1002/9781118540190.wbeic006

di Carlo, G. S. (2013). Lexical differences between American and British English: A survey study. *Language Design, 15*(2013), 61-75.

Giles, H., & Johnson, P. (1987). Ethnolinguistic identity theory: A social psychological approach to language maintenance. *International Journal of the Sociology of Language, 1987*(68), 69-99. https://doi.org/10.1515/ijsl.1987.68.69

Gluszek, A., & Dovidio, J. F. (2010). The way they speak: A social psychological perspective on the stigma of non-native accents in communication. *Personality and Social Psychology Review, 14*(2), 214-237. https://doi.org/10.1177%2F1088868309359288

Hay, J., Nolan, A., & Drager, K. (2006). From fush to feesh: Exemplar priming in speech perception. *The Linguistic Review, 23*(3), 351-379. https://doi.org/10.1515/TLR.2006.014

Hay, J., Warren, P., & Drager, K. (2006). Factors influencing speech perception in the context of a merger-in-progress. *Journal of Phonetics, 34*(4), 458-484. https://doi.org/10.1016/j.wocn.2005.10.001

Hu, G., & Su, J. (2015). The effect of native/non-native information on non-native listeners' comprehension. Language Awareness, 24(3), 273–281. https://doi.org/10.1080/09658416.2015.1077853

Huang, B. H. (2013). The effects of accent familiarity and language teaching experience on raters' judgments of non-native speech. *System, 41*(3), 770-785. https://doi.org/10.1016/j.system.2013.07.009

Johnson, K., Strand, E. A., & D'Imperio, M. (1999). Auditory-visual integration of talker gender in vowel perception. *Journal of Phonetics, 27*(4), 359–384. https://doi.org/10.1006/jpho.1999.0100

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), 44–456. https://doi.org/10.1177%2F0261927X09341950

Lee, B. J. (2022). Listener engagement: The missing link in research on accented speech. *International Review of Applied Linguistics in Language Teaching.* https://doi.org/10.1515/iral-2021-0177

Lee, B. J., & Bailey, J. L. (2020). Japanese tertiary-level students' cognition of World Englishes. *Memoirs of Fukui University of Technology, 50*, 254-267. http://hdl.handle.net/10461/29339

Munro, M. J., & Derwing, T. M. (1995). Foreign accent, comprehensibility, and intelligibility in the speech of second language learners. *Language Learning*, *45*(1). 73-97. https://doi.org/10.1111/j.1467-1770.1995.tb00963.x

Murakami, H. "夜中の汽笛について、あるいは物語の効用について" [Concerning the Sound of a Train Whistle in the Night or On the Efficacy of Fiction]. 夜のくもざる一村上朝日堂超短小説 *[Cloudy Night: Short Stories by Murakami Asashido]*, H. Murakami. Shincho Bunko, 1995, pp. 234-240.

Niedzielski, N. (1999). The effect of social information on the perception of sociolinguistic variables. *Journal of Language and Social Psychology, 18*(1), 62-85. https://doi.org/10.1177%2F0261927X99018001005

Neustupny, J. V. (1986). *Communication with foreigners*. Tokyo: Iwanami.

Plonsky, L., & Oswald, F. L. (2014). How big is "big"? Interpreting effect sizes in L2 research. *Language Learning, 64*(4), 878-912. https://doi.org/10.1111/lang.12079

Reid, K. T., Trofimovich, P., O'Brien, M. G. (2019). Social attitudes and speech ratings: Effects of positive and negative bias on multiage listeners' judgments of second language speech. *Studies in Second Language Acquisition*, *41*(2), 419-442. https://doi.org/10.1017/S0272263118000244

Rubin, D. L. (1992). Nonlanguage factors affecting undergraduates' judgements of nonnative English-speaking teaching assistants. *Research in Higher Education, 33*(4), 511-531. https://doi.org/10.1007/BF00973770

Russo, M., Islam, G., & Koyuncu, B. (2017). Non-native accents and stigma: How self-fulfilling prophesies can affect career outcomes. *Human Resource Management Review, 27*(3), 507-520. https://doi.org/10.1016/j.hrmr.2016.12.001

Saito, K., & Plonsky, L. (2019). Effects of second language pronunciation teaching revisited: A proposed measurement framework and meta-analysis. *Language Learning, 69*(3), 652-708. https://doi.org/10.1111/lang.12345

Strand, E.A. (1999). Uncovering the role of gender stereotypes in speech perception. *Journal of Language and Social Psychology, 18*(1), 86–99. https://doi.org/10.1177%2F0261927X99018001006

Tajfel, H., & Turner, J. C. (1979). An integrative theory of intergroup conflict. In W. G. Austin, & S. Worchel (Eds.), *The social psychology of intergroup relations* (pp. 33-47). Monterey, CA: Brooks/Cole.

Trofimovich, P., McDonough, K., Dao, P., & Abashidze, D. (2019). Attitudinal bias, individual differences, and second language speakers' interactional performance. *Applied Linguistics Review*, *13*(1). https://doi.org/10.1515/applirev-2019-0010

Zheng, Y., & Samuel, A. G. (2017). Does seeing an Asian face make speech sound more accented? *Attention, Perception, & Psychophysics*, *79*, 1841-1859. https://doi.org/10.3758/s13414-017-1329-2