

Okim Kang¹ , Kate Yaw² , and Maria Kostromitina¹ 

¹ English Department, Northern Arizona University, United States

² Department of World Languages, University of South Florida, United States

The effects of situational contexts and occupational roles on listeners' judgements on accented speech

Much language attitude research has demonstrated that people make biased judgements based on speakers' language choice and accent. However, the influence of occupational context on listeners' perceptions of accented English is largely unknown. This verbal guise study examined the extent to which academic contexts and workforce-related professional contexts affect listeners' judgements of accented speech. Results revealed that simulated contexts made a significant difference in listeners' perceptual judgements, with speakers perceived as significantly more comprehensible and acceptable in service-occupational roles than in academic contexts. These findings suggest that listeners' speech judgements can be heavily influenced by speakers' situational contexts. The study also provides evidence in support of the fluency principle, showing that listeners may evaluate accented speech more negatively if it requires more processing effort. The findings inform the domains of linguistic stereotyping and listeners' attitudes towards accented speech.

Key words: listener bias, stereotyping, situational context, occupational context, accented speech

Address for correspondence: Okim Kang

English Department, Northern Arizona University, Flagstaff, AZ, 86011, United States.

E-mail: okim.kang@nau.edu

This is an open access article licensed under the CC BY NC ND 4.0 License.

Listeners use accent as a salient feature of spoken communication to index a speaker's membership in a speech community (i.e., a group with shared linguistic norms and speech patterns, Wardhaugh & Fuller, 2015). In discussions of accent variation among first language (L1) English speakers, speech communities may be described using notions of standard (e.g., Received Pronunciation, General American) or nonstandard (e.g., Texas, Boston) accents, in which a standard accent is a way of speaking considered superior to others (Wardhaugh & Fuller, 2015). In this definition of standard, the emphasis is on speaker prestige as indexed by accent rather than uniformity imposed upon the phonological production (Milroy, 2001). Indeed, linguists caution that the concept of a uniform, monolithic accent is a myth (for a detailed discussion, see Lippi-Green, 2012). Additional variation enters when we consider the phonological features of spoken language produced by LX users (i.e., speakers who learned English as an additional language after the age of three; see Dewaele, 2018). Again, it is the listener perception of this variation that seems to matter most, with accentedness being measured as the perceived degree to which a speaker's pronunciation deviates from the listener's expected norm (Munro & Derwing, 1995). Then, in accordance with the fluency principle of language attitudes, accentedness may impact listeners' processing fluency, which in turn mediates the ensuing language attitudes (Dragojevic et al., 2017). Thus, the goal of the present study was to apply the processing fluency model to better understand why listeners might have different perceptual ratings across different academic and occupational contexts, as some contents might require more processing effort than others, which might lead to biased evaluations. The study also made use of linguistic and reverse linguistic stereotyping paradigms as well as the expectancy violations theory (EVT) and language attitude theories in order to add further interpretations of potential rating differences of LX-accented speech across occupational roles and expectations.

Processing Fluency

Extracting information from the speech stream is an active cognitive process in which listeners rely on the speech signal itself plus the context in which it occurs and their own expectations (Heald & Nusbaum, 2014). The amount of effort involved in mapping the acoustic signal onto linguistic representations and thereby gleaning meaning from speech is known as processing fluency (Dragojevic et al., 2017; Dragojevic & Giles, 2016). Due to the fluency aspects of speech processing, speech that is more difficult to process (i.e., requiring more effort) tends to cue more negative evaluations of the speaker whereas speech that is easier to process can positively bias how listeners view the speaker (Dragojevic, 2020). There are two potential means by which this operates. One is that listeners tend to shift the communicative burden to the speaker (Lippi-Green, 2012) so that when they experience processing difficulties, these are directly attributable

to the speaker's (in)ability to communicate. The second is that processing fluency cues an affective reaction, which then prompts listeners to up- or downgrade their evaluations of speakers in line with their affective response (Dragojevic, 2020).

Processing fluency research most commonly measures cognitive processing through scalar ratings of the speaker's comprehensibility, using multi-item scales with a small number of speakers (see Dragojevic & Giles, 2016; Hansen & Dovidio, 2016). This provides an in-depth look at how listeners perceive individual speakers, though generalizability to novel talkers may be hard to claim without aggregating findings from multiple studies with this design. In research on speech perception, an alternative is to present listeners with a larger number of speakers/recordings and gather quick, impressionistic ratings through single-item scales (e.g., Munro & Derwing, 1995). This approach exposes listeners to a wider range of audio stimuli with a tradeoff in how in-depth the rating process for each individual stimulus can be before listener fatigue becomes problematic. The current study took the latter approach.

Language Attitudes

Listeners' attitudes toward different accents can prompt evaluative reactions to the language varieties through a process of social categorization and stereotypic trait attribution (Dragojevic, 2018), moderated by the degree of processing fluency experienced. Indeed, the human brain relies on social categorization to facilitate processing of new voices. This becomes more problematic when listeners overgeneralize and assume that everyone who speaks a certain way automatically shares other nonlinguistic traits (Kinzler, 2020). Such evaluative reactions can be categorized into two related attitudinal phenomena: linguistic stereotyping and reverse linguistic stereotyping.

Linguistic Stereotyping

When listeners use speakers' accents and language varieties as grounds for judgements about the speakers themselves, they are engaging in linguistic stereotyping (Bradac et al., 2001; Lambert et al., 1960). Research suggests that linguistic stereotypes reflect social group associations learned from a young age rather than inherent differences in speech itself (Giles & Niedzielski, 1998; Kinzler et al., 2010). These stereotypes relate to various nonlinguistic attributes of the speaker such as ethnicity, social status, intelligence, personality, physical characteristics, and job suitability.

The situational context in which language varieties are used can play a major role in how listeners evaluate a speaker (Dragojevic, 2018). One such example is suitability for employment. On this front, speakers of nonstandard varieties are less likely to be hired, retained, and/or promoted compared to colleagues who speak standard varieties (Lippi-Green, 2012). In U.S. higher education,

international teaching assistants (ITAs) may have lower course enrollments than domestic TAs (Bresnahan & Sun Kim, 1993), face student blame for low grades (Fitch & Morgan, 2003), and be rated lower in course evaluations (Smith et al., 2005). Speakers with LX accents have been viewed as less credible than L1-accented speakers (Lev-Ari & Keysar, 2010). Attitudinal research generally shows that the perceived status of the job matters, with speakers of standard varieties of English being viewed as significantly more employable in high-status jobs than are those with nonstandard accents (see De La Zerda & Hopper, 1979; Giles et al., 1981), though Cargile (2000) found no difference in job suitability ratings between L1 and Chinese accented speech regardless of job status. In a study controlling for perceived job status by asking participants to rate each speaker guise for an entry-level position, Carlson and McHenry (2006) found that more salient nonstandard accents elicited lower employability ratings, suggesting that the situational context in which the speech occurs may interact with the degree of accent perceived by the listener.

Reverse Linguistic Stereotyping

While research has established that listeners judge a speaker's traits based on the way they talk, we also know that the reverse is true – listeners' perceptions of a speaker can be influenced by traits ascribed to the speaker (Kang & Rubin, 2009; Rubin, 2012). Nonlinguistic speaker information, such as name, job title, biography, or appearance can shape how listeners perceive speakers' accents (Lindemann, 2002). This phenomenon, known as reverse linguistic stereotyping, has been demonstrated in multiple academic and employment contexts. In higher education, such as among U.S. undergraduate students responding to ITA speech, visual guises have been used to elicit evidence of reverse linguistic stereotyping (Rubin, 1992). For instance, Kang and Rubin (2009) found that listeners rated the same L1 English speaker as more LX-accented when accompanied by a photograph of an East Asian lecturer versus a photograph of a Caucasian American lecturer. This bias in perceptual judgement has also emerged in HIV prevention counseling (Rubin et al., 1997), manager-employee relations (Rubin et al., 1999), and perception of health care aides (Rubin et al., 2016). Other nonvisual aspects of identity have also been applied as guises to test their impact on perception of speech. In a study on the impact of the ethnicity of proper names in Switzerland, Prikhodkine et al. (2016) found that multiple L1 speakers of Swiss French were perceived as more foreign-accented and less suitable for a job as a bank manager when they were presented under an ethnic minority name guise than when the same speakers were assigned a more traditional Swiss name. Therefore, perceptual judgements of speaker acceptability, or the extent to which their way of speaking is considered appropriate for a given context (Thomson, 2018), may be susceptible to both linguistic and reverse linguistic stereotyping.

Expectancy Violations Theory

In addition to the processing fluency principle, EVT offers a useful lens on the possible mechanisms underpinning linguistic and reverse linguistic stereotyping. Originally developed to explain the effects of proxemic expectations on communication, EVT has been expanded to include the role of various verbal and nonverbal behaviors in interpersonal communication (Burgoon, 1993). According to EVT, an expectancy is “an enduring cognition about the behavior anticipated of others” (Burgoon, 2015, p. 2), stemming from knowledge of social norms and any additional information available about the interlocutor. Thus, in the case of linguistic stereotyping, a speaker’s accent may cue expectations about “people who sound like that,” leading to an automatic association of stereotypes with the speaker’s accent. However, for reverse linguistic stereotyping, it is the contextual information that sets the expectations for the language and accent the speaker will produce. The perceptual ramifications of these expectations, as well as their violation, are intriguing. Babel and Russell (2015) found that providing listeners with photos of L1 English speakers of Chinese or White ethnic backgrounds impeded intelligibility for the ethnically Chinese speakers while improving accentedness ratings for the ethnically White speakers. Gnevsheva (2018) found that when listeners’ expectations were violated (i.e., by encountering an LX accent from a Caucasian speaker presumed to be an L1 English speaker), they rated the degree of LX accentedness as significantly stronger than in an audio-only condition without this contextual information about the speaker.

The Current Study

What remains unexplored in perception of speech research is the impact of the speaker’s occupational context as an independent variable on listeners’ accentedness, comprehensibility, and acceptability judgements of LX-accented speech. Through this study, we sought to investigate the extent to which situational contexts affect listeners’ judgements of LX-accented speech. We hypothesized that varying the situational contexts could yield evidence of reverse linguistic stereotyping by showing biased perceptual judgements of LX-accented speech across different contexts. The study was guided by the following three research questions:

What are the rating differences in listeners’ judgements of accentedness, acceptability, and comprehensibility of LX-accented speech

- (1) between academic and nonacademic professional contexts?
- (2) across three academic disciplines (i.e., humanities, social sciences, and physical sciences)?
- (3) across three occupations (i.e., waiter, teacher, and doctor)?

Method

To address the research questions, we employed a quasi-experimental design in which we manipulated academic specialization (i.e., humanities, physical sciences, and social sciences) and professional contexts (i.e., waiter, doctor, and teacher). We did so for three L1 accent conditions (i.e., Arabic, Mandarin, and Spanish) using an adaptation of the verbal guise technique (Garrett, 2010). This method has been successfully applied in other studies of linguistic and reverse linguistic stereotyping (e.g., Kang & Rubin, 2009; Prikhodkine et al., 2016).

Participants

Participants (listeners) were 64 undergraduate students recruited from an introductory linguistics course at a US university. In this sample, 68.75% identified as female and 31.25% identified as male. They ranged in age from 18 to 35 years old (87.50% = 18–21, 10.94% = 22–25, 1.56% = 30 or older), and were all L1 English users. Almost two thirds (63%) had already taken one or two introductory linguistics courses before they participated in this study. Listeners received extra credit for completing the study.

Procedure

Data were collected using the online platform SurveyGizmo. Participants completed a brief demographic questionnaire, then responded to 24 sound files (18 target stimuli plus six distractors). Stimuli were presented one at a time in randomized blocks, and listeners were not informed that they would hear the same speaker more than once, which was later the case. Each stimulus appeared with a written description introducing the speaker's profession and the context of the speech. All speakers were introduced as having 10 years of experience in their respective fields. Participants listened to each audio file once, then rated it for accentedness, comprehensibility, and acceptability. Participants were also asked to provide brief online comments immediately after their ratings for each speech construct. These comments were provided completely voluntarily but helped better understand reasoning behind each rating. Survey completion time was 40–50 minutes.

To achieve variation in the audio files, we used an adaptation of the verbal guise technique (Garrett, 2010). A typical verbal guise task includes audio recordings from multiple speakers. In contrast, a matched-guise approach involves one bidialectal speaker producing two or more recordings in different dialects/accents (e.g., Lambert et al., 1960). The 18 target recordings in our study were produced by three speakers: one L1 Arabic, one L1 Mandarin, and one L1 Spanish. The speakers were not instructed to vary their accent or style of speaking. Nonetheless, elements of the matched-guise technique were present, in that these three speakers produced multiple audio files, with each file containing different

content and representing a different occupational identity. Listeners heard each of the target speakers six times, each time under a different occupational guise. This adaptation addressed some of the matched-guise limitations summarized in Solís Obiols (2002), namely, concerns about the accentuation of stereotypes in the recording process (Gardner & Lambert, 1972) and the repetition of the same content, prompting listeners to focus more on linguistic features than they would with novel content (Lee, 1971). Regarding the former, speakers were asked to produce six unique recordings without explicitly varying their accent, meaning that the voice remained the same while the content of each recording differed. Thus, the speakers were less likely to (over)emphasize differences between accent varieties that could otherwise introduce more stereotyped linguistic features. For the latter, this study included scripted and semi-scripted recordings (see the Materials section). While there was repetition of the message in the scripted lecture recordings, they were interspersed with the semi-scripted recordings such that the listeners never heard the same recording twice in a row.

A follow-up online interview with several open-ended questions was presented upon the completion of the survey. The interview concerned the participants' general attitude toward the study. To check the guise manipulation, participants were asked if they had heard any speaker more than once during the listening tasks. Responses were generally "no" (35%), "it's possible, but I didn't notice it" (33%), "I didn't think about it (22%)," or "maybe, I don't know (10%)". None of the participants gave a confirmatory response ("yes") about hearing the speakers more than once.

Materials

Audio Stimuli

All audio stimuli were recorded by adult male speakers ranging in age from 25 to 35 years old. Three university instructors were selected to record the target audio stimuli. They were highly intelligible LX English speakers from three L1 backgrounds: Arabic, Chinese (Mandarin), and Spanish. These three speakers recorded six sound files each, one per situational context (see below), for a total of 18 target stimuli.

This study included two types of monologic stimuli: scripted academic lectures and semi-scripted professional contexts. Academic lecture scripts were excerpts of humanities, social sciences, and physical sciences lectures (representing different disciplines) from the Michigan Corpus of Academic Spoken English (MICASE; Simpson et al., 1999). All lectures were originally delivered by L1 speakers of American English. For the current study, the first 30 seconds (67–75 words) of each lecture were used.

The semi-scripted stimuli presented a common scenario for each target profession (waiter, doctor, and school teacher). For instance, in a scenario of a waiter telling a customer about the daily specials, the speaker was provided with five pieces of information to communicate to the customer (e.g., "soup of the

day is split pea with ham”) and three keywords that should be included when communicating this information (e.g., “soup,” “split pea,” “ham”). The school teacher context included a scenario of discussing a student’s class performance with a parent, and the doctor context was about a doctor explaining test results to a patient. The semi-scripted stimuli included numerous authentic features of real-world speech such as phonological modifications, speaker hesitations, and self-corrections (Geddes & White, 1978), making them increasingly popular as listening comprehension stimuli (e.g., Wagner, 2014).

The inclusion of both scripted and semi-scripted tasks in the study raises concerns about their comparability. In fact, the original study design was to use scripted tasks for all six audio files. However, piloting revealed that while the scripted academic lectures were perceived by listeners as authentic, scripted versions of the professional scenarios were not. Therefore, to preserve the plausibility of the identity guises, we chose to use the semi-scripted versions instead.

Audio stimuli were controlled for vocabulary range, intelligibility, and speech rate. Vocabulary range, represented by type-token ratio (TTR, i.e., the ratio of the number of unique word types to the total number of words; Nation, 2013), was measured using the LexTutor vocabulary profiler (Version 4, Cobb, 2020). The TTR values ranged from 0.28 to 0.32. Intelligibility and speech rate were evaluated by eight trained linguistics PhD students. Speakers’ intelligibility was confirmed by scores of 90 or higher on a 1–100 scale marking what percent of the speaker’s words were understood (Kang et al., 2018). Regarding the speech rate, all speakers were given a sample with an ideal speech rate after which to model themselves. They also were asked to practice their recording as many times as they needed. The speakers’ speech rate mostly fell within the range of approximately 3.2–3.7 syllables per second. If their speech rate fell outside of this range, they were asked to rerecord.

Finally, there were also six distractor recordings interspersed with the target audio stimuli. Four additional adult males recorded the distractors. Three of these speakers were international teaching assistants from other L1 backgrounds. The fourth was an L1-English university ESL instructor. Each distractor speaker recorded one or two sound files, for a total of six distractors.

Rating scales

Each audio stimulus was rated for three constructs on a nine-point slider scale. The constructs were comprehensibility (1 = *extremely difficult to understand*, 9 = *extremely easy to understand*), LX-accentedness (1 = *very strongly accented*, 9 = *not accented at all*), acceptability (1 = *not acceptable at all to fulfill the job described above*, 9 = *completely acceptable to fulfill the job described above*). Such scales are commonly employed in LX speech research and validated by other studies (see Munro, 2018; Thomson, 2018). In addition, this impressionistic approach was intentionally adopted for two reasons: (a) because the purpose of the study was to examine nontrained listeners’ instant judgements

on speech files and (b) because by completing a quick judgement scale, listeners could move on promptly through each stimulus without realizing that they had listened to the same speaker. The construct of accentedness was measured from the semantic sense of "accent standardness."; However, in the current study, the term "accentedness" was used deliberately in order to maintain the consistency of the current practice in L2 speech research (Munro, 2018). The inter-rater reliability coefficient among these listeners was above .92 and responses on all three variables were normally distributed.

Data Analyses

To answer our three research questions, we fit nine linear mixed-effects models (LMEMs) onto the data using the lme4 package (Bates et al., 2007) in R (ver. 4.0.2; R Core Team, 2020). This allowed us to examine the main effect of situational context while controlling for the random factors of listeners and speakers. Thus, the response variable in each model was one of the three speech perception constructs, while the target situational contexts were included as the fixed factor. Listeners and speakers' L1 backgrounds were entered as random factors. Before interpreting the model results, residual scatterplots were evaluated for the evidence of normality, linearity, and homoscedasticity in the data. The visual examination of the residual versus fitted plots did not reveal substantial violations of the assumptions. Finally, in running the nine LMEMs, the corrected alpha level was set at $p = .006$.

In order to determine a sufficient sample size for the study, attempts were made to conduct an a priori power analysis only for the fixed effects in the models using G*Power (Faul et al., 2007). This solution was deemed possible as the analysis of fixed effects in a LMEM resembles that of an analysis of variance (ANOVA). Specifically, we chose a repeated-measures within-factors ANOVA as a statistical test for this power analysis. Two analyses were run in *G*Power*: one for a two-level fixed factor and the other for a three-level fixed factor.

However, upon further deliberations and consultations with the support team of G*Power, it was concluded that the chosen approach was not practical for an a priori power analysis in the current study, since this tool was not intended to calculate statistical power for a LMEM. Furthermore, as noted by Kumle et al. (2021), although mixed-effects models are a powerful statistical technique, they do not offer a straightforward way of estimating the probability of correctly rejecting the null hypothesis. One way to approach power analysis for this technique would be through simulation (see e.g., Brysbaert & Stevens, 2018; Westfall et al., 2014). Then, such an analysis would require either a similar existing dataset or predetermined estimates for the parameters in the simulation. This process needs to be carefully selected on the basis of similar research, if such research is available. Due to these obstacles, an a priori power analysis was concluded not feasible. Instead, we conducted a post hoc simulation-based power

analysis by using the *simr* package in R (Green & MacLeod, 2016). The results of the analysis with 200 simulations showed that the size of the current dataset ($N = 64$ participants, 18 measurements per participant: $6 \text{ stimuli} \times 3 \text{ speech perception constructs}$) was sufficient to reject the null hypothesis at the alpha level $p < .05$ (achieved $b > .95$).

Results

The current study addressed three research questions that focused on the effect of situational contexts on listeners' judgements of LX-accented speech on accentedness, comprehensibility, and acceptability. Table 1 presents the descriptive statistics and correlations between the three listener judgement variables.

Overall, listeners rated the speakers relatively high on acceptability and comprehensibility. The accentedness ratings were somewhat lower, indicating that listeners could hear a foreign accent in the speech samples. As seen in Table 1, the three constructs were moderately correlated, but none demonstrated strong correlations, thus confirming that the measured criteria should be treated separately.

Differences in Listeners' Judgements Between Academic and Occupational Contexts

Three LMEMs were performed to examine the differences in listeners' judgements of LX-accented speech in different situational contexts. The fixed factor of situational context in these models consisted of two levels: (a) academic context that included three lecture types (humanities, social sciences, and physical sciences) and (b) occupational context that included three professions (doctor, waiter, and teacher). Figure 1 plots the means and associated 95% CIs of accentedness, comprehensibility, and acceptability ratings of the stimuli in the academic and other occupational contexts. It shows that occupational stimuli were rated higher overall, with speech rated as less accented and more comprehensible, and the speakers as more acceptable.

Table 2 presents the results of the three LMEMs, showing that the observed differences in accentedness, acceptability, and comprehensibility were statistically significant between contexts. To estimate the effect sizes, we followed Brysbaert and Stevens (2018), who suggested calculating Cohen's d for mixed-effect models in psychology. The calculated effect sizes given in Table 2 revealed medium to large effects of situational context on accentedness ratings

Table 1. *Descriptive Statistics and Correlations Among Listener Judgement Constructs*

	<i>M</i>	<i>SD</i>	<i>SE</i>	1	2	3
1. Accentedness	4.29	1.92	0.06	1	0.32	0.51
2. Comprehensibility	6.48	1.98	0.06		1	0.57
3. Acceptability	7.40	1.77	0.05			1

Figure 1. Means and 95% CIs of Accentedness, Comprehensibility, and Acceptability Ratings in Academic vs. Occupational Context (Mean of the Three Accents: Arabic, Mandarin Chinese, Spanish).

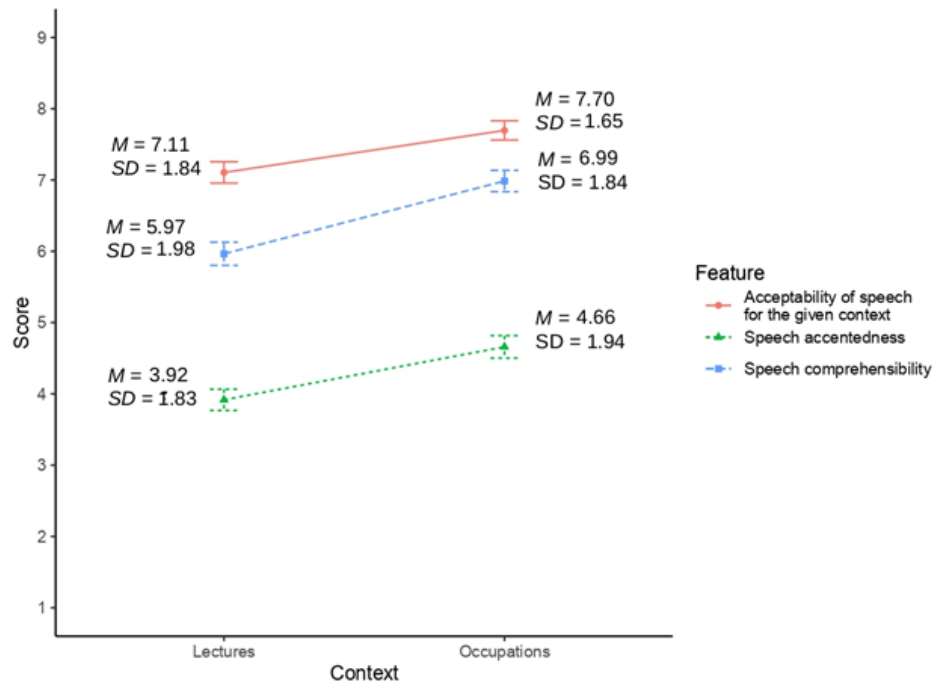


Table 2. Summary of Results of Linear Mixed-Effects Models for Accentedness, Comprehensibility, and Acceptability Rating										
Fixed effects							Random effects			
Parameters (N = 576 per context)	Est.	SE	df	t	p	d	Spk	Listn	Con. R ²	Mar. R ²
							SD	SD		
Accentedness										
Lectures	3.92	0.43	2.56	9.10			0.70	1.12	0.49	0.04
Occupations	4.66	0.08	1085	8.96	< .001 ^{e-5}	0.55				
Comprehensibility										
Lectures	5.97	0.46	2.37	13.0			0.76	0.99	0.44	0.06
Occupations	6.99	0.09	1085	11.4	0.01 ^{e-5}	0.69				
Acceptability										
Lectures	7.11	0.33	3.21	21.6			0.50	1.18	0.53	0.03
Occupations	7.70	0.07	1085	8.16	0.01 ^{e-5}	0.50				

Note. Con. R² = conditional R², Mar. R² = marginal R², adjusted alpha level .006

($d = 0.55$) and acceptability ratings ($d = 0.55$), and a medium-to-large effect on comprehensibility ratings ($d = 0.69$). Each model explained a substantial amount of variance in the responses, as indicated by the conditional R^2 values in Table 2, though the fixed factors accounted for only a small portion (4–6%) of this variance (see marginal R^2 values in Table 2). In other words, the majority of variance in the listeners' perception ratings was explained by the random factors of listeners and speakers' L1 background (38-50%).

Differences in Listeners' Judgements Across Different Academic Disciplines

In order to examine the differences in listeners' judgements across academic contexts, we performed the LMEM with a three-level fixed factor representing lectures in three academic fields (humanities, social sciences, and physical sciences). Figure 2 presents the average ratings and 95% CIs for accentedness, comprehensibility, and acceptability ratings across the three lecture types. The figure demonstrates an overall decline in listeners' ratings from humanities to social sciences lectures. The decline is even more noticeable in the physical science lecture ratings, especially with regard to comprehensibility.

Figure 2. Means and 95% CIs of Accentedness, Acceptability, and Comprehensibility Ratings in Academic Contexts

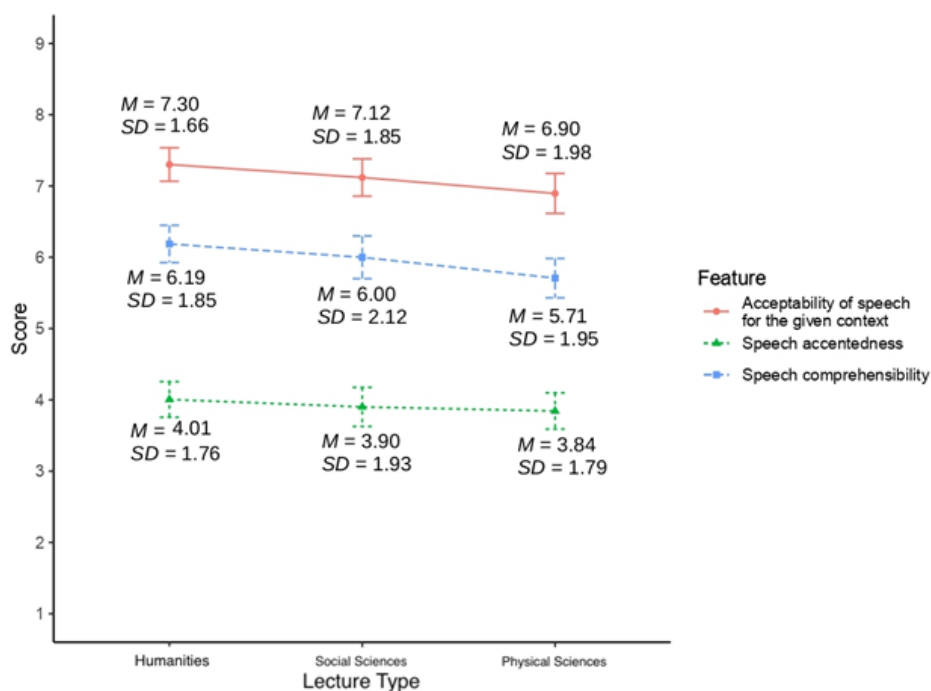


Table 3 summarizes the results of the LMEMs for the fixed and random effects as well as the variance explained by the fixed factors and the total variance explained by the models.

The results computed for the estimates for the three different lecture type conditions revealed statistically nonsignificant differences between lecture types in the ratings of accent. Further analysis of the estimates showed that acceptability and comprehensibility ratings were statistically significantly different for humanities and physical sciences lectures in particular ($p = .003$). More specifically, lectures in humanities were rated as statistically significantly more acceptable and comprehensible than lectures in physical sciences. Cohen's d effect sizes were calculated for the estimates in the LMEM. As reported in Table 3, the effect of the different types of lectures on the three ratings was generally small. However, a near- medium effect size was detected in acceptability and comprehensibility ratings between humanities and physical sciences ($d = 0.48$ and 0.44 respectively).

Similar to the first research question, the majority of variance in listeners' ratings was explained by the random factors, that is, the listeners and speakers (29%–50%). The fixed factor of lecture type explained between 1% and 9% of the variance.

Differences in Listeners' Judgements Across Different Occupations

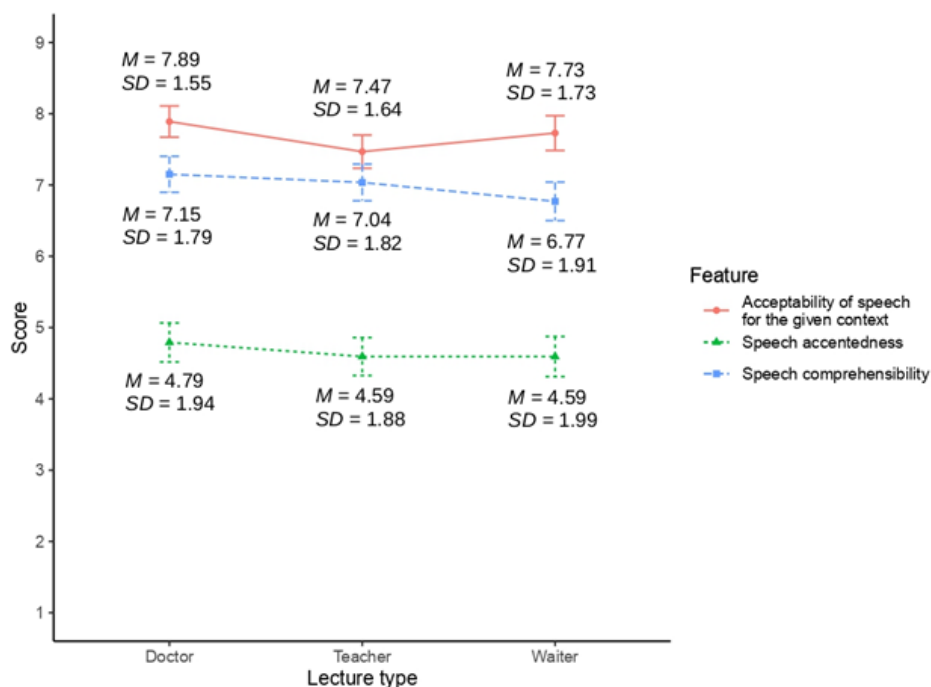
The third research question addressed the differences in listeners' rating of accented speech across professional contexts, namely, a waiter announcing the specials of the day, a teacher discussing a student's class performance with a parent, and a doctor explaining test results to a patient. Figure 3 depicts the

Table 3. *Summary of Results of Linear Mixed-Effects Models for Accentedness, Comprehensibility, and Acceptability Rating of Lectures*

Parameters ($N = 182$ per context)	Fixed effects					Random effects				
	Est.	SE	df	t	p	d	Spk SD	Listn SD	Con. R^2	Mar. R^2
Accentedness										
Humanities	4.01	0.32	3.69	12.48			0.47	1.15	0.45	0.01
Soc. Science	3.90	0.14	508	-0.74	.46	0.11				
Phys. Science	3.84	0.14	508	-1.15	.25	0.17				
Comprehensibility										
Humanities	6.19	0.37	3.07	16.47			0.57	1.09	0.38	0.09
Soc. Science	6.00	0.16	508	-1.16	.25	0.17				
Phys. Science	5.71	0.16	508	-2.96	.003	0.44				
Acceptability										
Humanities	7.30	0.35	3.57	20.68			0.52	1.30	0.57	0.07
Soc. Science	7.12	0.13	508	-1.45	.15	0.21				
Phys. Science	6.89	0.13	508	-3.23	.001	0.48				

Note. Con. R^2 = conditional R^2 , Mar. R^2 = marginal R^2 , adjusted alpha level .006

Figure 3. Means and 95% CIs of Accentedness, Comprehensibility, and Acceptability Ratings in Occupational Contexts



means and 95% CIs of accentedness, comprehensibility, and acceptability ratings for each occupational context, revealing several trends.

Whereas listeners evaluated the speech samples of waiters as the most accented and least comprehensible, the lowest acceptability rating was given to the teacher samples. As summarized in Table 4, the results of the LMEMs with the professional contexts as fixed factors in each model showed that the type of occupational context generally had a small, statistically nonsignificant effect on accentedness and comprehensibility ratings. However, there was a statistically significant difference in acceptability judgements for teachers and doctors. Specifically, the results of the LMEM in Table 4 indicated that the ratings of acceptability were statistically significantly lower for teachers than for doctors, with a medium effect of the context ($d = 0.56$). Moreover, waiters were evaluated as statistically significantly less comprehensible than doctors, with the effect of the context being near-medium ($d = 0.40$).

The R^2 values further indicated that the majority of variance in each model was explained by the random factors of listeners (22–46%) and speakers' LX background (10–28%). The fixed factor of occupational context explained only between 1% and 7% of the variance.

Table 4. *Summary of Results of Linear Mixed-Effects Models for Accentedness, Comprehensibility, And Acceptability Ratings of Speech Samples in Professional Contexts*

Fixed effects							Random effects			
Parameters (<i>N</i> = 182 per context)	Est.	SE	df	<i>t</i>	<i>p</i>	<i>d</i>	Spk	Listn	Con. <i>R</i> ²	Mar. <i>R</i> ²
							<i>SD</i>	<i>SD</i>		
Accentedness										
Doctor	4.79	0.58	2.34	8.24			0.96	1.10	0.53	0.02
Teacher	4.59	0.14	508	−1.39	.16	0.21				
Waiter	4.59	0.14	508	−1.39	.16	0.21				
Comprehensibility										
Doctor	7.15	0.61	2.23	11.80			1.02	0.90	0.50	0.07
Teacher	7.04	0.14	508	−0.82	.41	0.12				
Waiter	6.77	0.14	508	−2.72	.02	0.40				
Acceptability										
Doctor	7.89	0.35	3.13	22.6			0.53	1.14	0.57	0.01
Teacher	7.47	0.11	508	−3.75	.0002	0.56				
Waiter	7.73	0.11	508	−1.434	.15	0.21				

Note. Con. *R*² = conditional *R*², Mar. *R*² = marginal *R*², adjusted alpha level .006

Discussion

The objective of the current study was to examine the extent to which situational contexts affected listeners' judgements of LX-accented speech. Specifically, using an adaptation of a verbal guise technique, we manipulated the academic specializations and professional occupations of accented speakers. We then asked listeners to evaluate the accentedness, comprehensibility, and acceptability of the LX-accented speech. The results showed that intelligible but accented LX-speech was perceived to be significantly more acceptable, more comprehensible, and less accented in nonacademic-based contexts (i.e., school teachers) as compared to academic contexts. In other words, even though the speaker's degree of accentedness and comprehensibility was identical across the situations, listeners perceived the LX-accented speech differently depending on the situational contexts in which the speaker's speech was presented. Studies (e.g., Lev-Ari & Keysar, 2010) have shown that listeners have higher and harsher expectations towards speakers in university-level academic contexts. International instructors are known to receive low course evaluations or negative reactions from U.S. undergraduate students (Bresnahan & Sun Kim, 1993; Fitch & Morgan, 2003; Smith et al., 2005) due to their accent. The current finding suggests that listeners' higher expectations in the academic context may lead to their harsher evaluation of accented speech. When listeners harbor certain expectations about speakers, their evaluation of a speaker's accent can be distorted (Kang & Rubin, 2009; Rubin, 1992) and such biased processes can result in language-based discrimination, which can occur commonly in our daily life (Kang & Rubin, 2009).

Among various factors that could have affected listeners' speech evaluation, listeners' previous experience in interacting with professionals (e.g., doctors or teachers) could have possibly caused their perceptual differences. Listeners' backgrounds or previous experience do indeed impact their evaluation of accented speech (Kang, 2012). In addition, listeners' situational familiarity or semantic context could play a role in their perceptual judgements (Kennedy & Trofimovich, 2008). However, given that research on listeners' familiarity with particular professionals' accents is limited, their LX-accented speech-rating patterns should be further investigated in the context of their previous LX-related experience and specific communication purposes.

For listeners' perceptions of LX-accented speech in three different lecture types, the results revealed that the LX-accented physics lecture was judged as significantly less comprehensible and the lecturer as less acceptable than for humanities, but no difference was found for social science. Interestingly, but perhaps not surprisingly (as the same speakers were used across the contexts), accentedness ratings did not show any significant differences across lecture types. These findings can be supported by the processing fluency model, in which the physics lecture might have required listeners more effort to process, which could lead them to evaluate the speakers' speech more negatively (Dragojevic, 2020). Note that the majority of listener participants in this study had a nonphysics background. Listeners might have found the physics lecture itself conceptually difficult and therefore less comprehensible. As introduced earlier, the scalar rating of the speaker's comprehensibility is a common measure for processing fluency (Dragojevic & Giles, 2016). These results further indicate that the listeners' expectations might have been higher for the speakers that explained conceptually difficult science content than for those who described humanities-related content. However, the current finding should not be generalized yet, because there can be large individual listener differences to consider, as seen from the large random effect variance (29%–50%). Listeners from different academic backgrounds may provide different types of responses.

Regarding listeners' judgements within the three occupational contexts, the accented school teachers were judged as less acceptable than doctors, but waiters were judged as less comprehensible than doctors. In post-rating qualitative comments, some listeners reported having reservations against their future children having a teacher with a non-native accent, confirming educational stakeholders' negative perceptions of teachers' accents (Moran et al., 2021). Moran et al.'s study demonstrated that parents were harsh in their ratings of accented teachers' teaching ability due to the investment that parents have in their children's education. Although the current study revealed a hint of such a trend, this topic was not primarily investigated in detail in this study, and these selective comments should not be generalized in a quantitative sense.

The findings about listeners' perceptions of accented speech in work-related contexts further verified listeners' biased judgements based on speakers'

occupational status (Carlson & McHenry, 2006; Dragojevic, 2018), demonstrating a significant difference in listeners' comprehensibility ratings between accented speech of individuals presented as waiters and doctors. Professionals with authority might have received more attention and credibility, which could have made their speech sound more comprehensible. An expectancy, as defined in EVT, or anticipated behavior based on knowledge of social norms about the speaker, could affect listeners' judgements or perspectives of others (Burgoon, 2015). As an example of reverse linguistic stereotyping, nonlinguistic contexts could set the expectations for the speaker's accent and comprehensibility (Kang & Rubin, 2009; Rubin, 1992). However, research findings on listeners' attitudes toward accented speakers' job suitability are still mixed (Carlson & McHenry, 2006; Giles et al., 1981). Thus, this authority-related credibility argument is not conclusive at this point, and more research is needed on this topic to confirm the argument. Here, it is important to note that accentedness judgements did not differ across the three professional contexts. This finding consistently upholds the pattern that regardless of accents, listeners perceived speakers' speech differently across professional and situational contexts.

Conclusion

This verbal guise study attempted to examine the influence of occupational roles on listeners' perceptions of accented English and found that the simulated situational contexts made a substantial difference in listeners' perceptual judgements. Overall, our findings are exploratory and limited in scope. The sample size was relatively small and was only supported by a post hoc power analysis. In addition, only three academic and three occupational contexts were tested. Further research can consider listeners' individual characteristics along with more extended situations and contexts as well as larger participant groups. Future studies can also explore the effect of different listener backgrounds on their perception of accents in various real-life contexts. Overall, even though the effect sizes of the fixed effects in our current findings were relatively small, we believe that our study can offer an initial discussion point regarding the importance of contexts in the speech perception and evaluation process.

Acknowledgments

We would like to give special thanks to ENG 703 seminar students who helped with the data collection of the project.

Conflict of Interest Disclosure

The Authors have no conflicts of interest to disclose.

Funding

The Authors declare no sources of funding.

Research Ethics Statement

This study has been approved by the Institutional Review Board at Northern Arizona University. (Reference 1038331-2).

Authorship Details

Okim Kang: research concept and design, collection and/or assembly of data, data analysis and interpretation, writing the article, critical revision of the article, final approval of the article.

Kate Yaw: writing the article.

Maria Kostromitina: data analysis and interpretation.

References

- Babel, M., & Russell, J. (2015). Expectations and speech intelligibility. *Journal of the Acoustical Society of America*, 137(5), 2823–2833. <https://doi.org/10.1121/1.4919317>
- Bates D, Mächler M, Bolker B, & Walker S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, 67(1), 1–48. <https://doi.org/10.18637/jss.v067.i01>.
- Bradac, J.J., Cargile, A.C., & Hallett, J.S. (2001). Language attitudes: Retrospect, conspect, and prospect. In W. P. Robinson & H. Giles (Eds.), *The new handbook of language and social psychology* (pp. 137–158). Wiley.
- Bresnahan, M. I., & Sun Kim, M. (1993). The impact of positive and negative messages on change in attitude toward international teaching assistants. *Folia Linguistica*, 27(3/4), 347–363. <https://doi.org/10.1515/flin.1993.27.3-4.347>
- Brysbaert, M., & Stevens, M. (2018). Power analysis and effect size in mixed

- effects models: A tutorial. *Journal of Cognition*, 1(1), 9. <https://doi.org/10.5334/joc.10>
- Burgoon, J. K. (1993). Interpersonal expectations, expectancy violations, and emotional communication. *Journal of Language and Social Psychology*, 12(1-2), 30–48. <https://doi.org/10.1177/0261927X93121003>
- Burgoon, J. K. (2015). Expectancy violations theory. In Berger, C. R., Roloff, M. E., Wilson, S. R., Dillard, J. P., Caughlin, J., & Solomon, D. (Eds.), *The international encyclopedia of interpersonal communication* (pp. 1–9). Wiley. <https://doi.org/10.1002/9781118540190.wbeic102>
- Cargile, A. C. (2000). Evaluations of employment suitability: Does accent always matter? *Journal of Employment Counseling*, 37(3), 165–177. <https://doi.org/10.1002/j.2161-1920.2000.tb00483.x>
- Carlson, H. K., & McHenry, M. A. (2006). Effect of accent and dialect on employability. *Journal of Employment Counseling*, 43, 70–83. <https://doi.org/10.1002/j.2161-1920.2006.tb00008.x>
- Cobb, T. (2020, May). *Web VP classic v.4* [computer program]. <https://www.lexutor.ca/vp/eng/>
- De La Zerda, N., & Hopper, R. (1979). Employment interviewers' reactions to Mexican American speech. *Communication Monographs*, 46(2), 126–134. <https://doi.org/10.1080/0367757909375998>
- Dewaele, J.-M. (2018). Why the dichotomy 'L1 versus LX user' is better than 'native versus non-native speaker.' *Applied Linguistics*, 39(2), 236–240. <https://doi.org/10.1093/applin/amw055>
- Dragojevic, M. (2018). Language attitudes. In H. Giles & J. Harwood (Eds.), *Oxford research encyclopedia of intergroup communication* (Vol. 2, pp. 179–192). Oxford University Press. <https://doi.org/10.1093/acrefore/9780190228613.013.437>
- Dragojevic, M. (2020). Extending the fluency principle: Factors that increase listeners' processing fluency positively bias their language attitudes. *Communication Monographs*, 87(2), 158–178. <https://doi.org/10.1080/03637751.2019.1663543>
- Dragojevic, M., & Giles, H. (2016). I don't like you because you're hard to understand: The role of processing fluency in the language attitudes process. *Human Research Communication*, 42(3), 396–420. <https://doi.org/10.1111/here.12079>
- Dragojevic, M., Giles, H., Beck, A., & Tatum, N. T. (2017). The fluency principle: Why foreign accent strength negatively biases language attitudes. *Communication Monographs*, 84(3), 385–405. <https://doi.org/10.1111/10.1080/03637751.2017.1322213>
- Faul, F., Erdfelder, E., Lang, A.-G., & Buchner, A. (2007). G*Power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behavior Research Methods*, 39(2), 175–191. <https://doi.org/10.3758/BF03193146>

- Fitch, F., & Morgan, S. E. (2003). "Not a lick of English": Constructing the ITA identity through student narratives. *Communication Education*, 52(3/4), 297–310. <https://doi.org/10.1080/0363452032000156262>
- Gardner, R. C., & Lambert, W. E. (1972). *Attitudes and motivation in second-language learning*. Newbury House.
- Garrett, P. (2010). *Attitudes to language*. Cambridge University Press.
- Geddes, M. & White, R. (1978). The use of semi-scripted simulated authentic speech and listening comprehension. *Audio-Visual Language Journal*, 16(3), 137–145.
- Giles, H., & Niedzielski, N. (1998). Italian is beautiful, German is ugly. In L. Bauer & P. Trudgill (Eds.), *Language myths* (pp. 85–93). Penguin.
- Giles, H., Wilson, P., & Conway, A. (1981). Accent and lexical diversity as determinants of impression formation and perceived employment suitability. *Language Sciences*, 3, 91–103. [https://doi.org/10.1016/S0388-0001\(81\)80015-0](https://doi.org/10.1016/S0388-0001(81)80015-0)
- Gnevsheva, K. (2018). The expectation mismatch effect in accentedness perception of Asian and Caucasian non-native speakers of English. *Linguistics*, 56(3), 581–598. <https://doi.org/10.1515/ling-2018-0006>
- Green, P., & MacLeod, C. J. (2016). simr: An R package for power analysis of generalised linear mixed models by simulation. *Methods in Ecology and Evolution*, 7(4), 493–498. <https://doi.org/10.1111/2041-210X.12504>
- Hansen, K., & Dovidio, J. F. (2016). Social dominance orientation, nonnative accents, and hiring recommendations. *Cultural Diversity and Ethnic Minority Psychology*, 22(4), 544–551. <https://doi.org/10.1037/cdp0000101>
- Heald, S. L. M., & Nusbaum, H. C. (2014). Speech perception as an active cognitive process. *Frontiers in Systems Neuroscience*, 8, 35. <https://doi.org/10.3389/fnsys.2014.00035>
- Kang, O. (2012). Impact of rater characteristics on ratings of international teaching assistants' oral performance. *Language Assessment Quarterly*, 9(3), 249–269. <https://doi.org/10.1080/15434303.2011.642631>
- 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. <https://doi.org/10.1177/0261927X09341950>
- Kang, O., Thomson, R., & Moran, M. (2018). Empirical approaches to measuring intelligibility of different varieties of English. *Language Learning*, 68(1), 115–146. <https://doi.org/10.1111/lang.12270>
- Kennedy, S., & Trofimovich, P. (2008). Intelligibility, comprehensibility, and accentedness of L2 speech: The role of listener experience and semantic context. *The Canadian Modern Language Review*, 64(3), 459–489. <https://doi.org/10.3138/cmlr.64.3.459>
- Kinzler, K. D. (2020). *How you say it: Why you talk the way you do – and what it says about you*. Houghton Mifflin Harcourt.

- Kinzler, K. D., Shutts, K., & Correll, J. (2010). Priorities in social categories. *European Journal of Social Psychology*, 40(4), 581–592. <https://doi.org/10.1002/ejsp.739>
- Kumle, L., Vö, M.L.H. & Draschkow, D. (2021). Estimating power in (generalized) linear mixed models: An open introduction and tutorial in R. *Behavior Research Methods* 53, 2528–2543. <https://doi.org/10.3758/s13428-021-01546-0>
- Lambert, W. E., Hodgson, R. C., Gardner, R. C., & Fillenbaum, S. (1960). Evaluational reactions to spoken language. *Journal of Abnormal and Social Psychology*, 60(1), 44–51. <https://doi.org/10.1037/h0044430>
- Lee, R. R. (1971). Dialect perception: A critical review and re-evaluation. *Quarterly Journal of Speech*, 57(4), 410–417. <https://doi.org/10.1080/00335637109383086>
- Lev-Ari, S., & Keysar, B. (2010). Why don't we believe non-native speakers? The influence of accent on credibility. *Journal of Experimental Social Psychology*, 46(6), 1093–1096. <https://doi.org/10.1016/j.jesp.2010.05.025>
- Lindemann, S. (2002). Listening with an attitude: A model of native-speaker comprehension of non-native speakers in the United States. *Language in Society*, 31, 419–441. <https://doi.org/10.1017/S0047404502020286>
- Lippi-Green, R. (2012). *English with an accent: Language, ideology, and discrimination in the United States (2nd ed.)*. Routledge.
- Ludwig, A., & Mora, J. C. (2017). Processing time and comprehensibility judgements in non-native listeners' perception of L2 speech. *Journal of Second Language Pronunciation*, 3(2), 167–198. <https://doi.org/10.1075/jslp.3.2.01lud>
- Milroy, J. (2001). Language ideologies and the consequences of standardization. *Journal of Sociolinguistics*, 5(4), 530–555. <https://doi.org/10.1111/1467-9481.00163>
- Moran, M., Kang, O., & McGroarty, M. (2021). Arizona educational stakeholders' perceptions of K12 teachers' accented speech. *Journal of Multilingual and Multicultural Development*. Advance online publication. <https://doi.org/10.1080/01434632.2021.1921783>
- Munro, M. J. (2018). Dimensions of pronunciation. In O. Kang, R. I. Thomson, & J. M. Murphy (Eds.), *The Routledge handbook of contemporary English pronunciation* (pp. 413–431). Routledge.
- 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>
- Nation, I. S. P. (2013). *Learning vocabulary in another language (2nd ed.)*. Cambridge University Press.
- Prikkhodkine, A., Correia Saavedra, D., & Dos Santos Mamed, M. (2016). “Give me your name and I’ll tell you whether you speak with an accent” The effect of proper names ethnicity on listener expectations. *CALL: Irish Journal for Culture, Arts,*

- Literature and Language*, 1(1), 10. <https://doi.org/10.21427/D7D592>
- R Core Team. (2020). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing. <https://www.R-project.org/>
- Rubin, D. L. (1992). Nonlanguage factors affecting undergraduate's judgements of nonnative English-speaking teaching assistants. *Research in Higher Education*, 33(4), 511–531. <https://doi.org/10.1007/BF00973770>
- Rubin, D. (2012). The power of prejudice in accent perception: Reverse linguistic stereotyping and its impact on listener judgements and decisions. In J. Levis & K. LeVelle (Eds.). *Proceedings of the 3rd Pronunciation in Second Language Learning and Teaching Conference, Sept. 2011*. (pp. 11–17). Iowa State University.
- Rubin, D.L., Ainsworth, S., Cho, E., Turk, D., and Winn, L. (1999). Are Greek letter social organizations a factor in undergraduates' perceptions of international instructors? *International Journal of Intercultural Relations*, 23(1), 1–12. [https://doi.org/10.1016/S0147-1767\(98\)00023-6](https://doi.org/10.1016/S0147-1767(98)00023-6)
- Rubin, D. L., Coles, V. B., & Barnett, J. T. (2016). Linguistic stereotyping in older adults' perceptions of health care aides. *Health Communication*, 31(7), 911–916. <https://doi.org/10.1080/10410236.2015.1007549>
- Rubin, D.L., Healy, P., Zath, R.C., Gardiner, T.C., & Moore, C.P. (1997). Non-native physicians as message sources: Effects of accent and ethnicity on patients' responses to AIDS prevention counseling. *Health Communication*, 9(4), 351–368. https://doi.org/10.1207/s15327027hc0904_4
- Simpson, R. C., Briggs, S. L., Ovens, J., & Swales, J. M. (1999). The Michigan corpus of academic spoken English. *The Regents of the University of Michigan*. <https://doi.org/10.21415/QT9V-2J96>
- Smith, R. A., Strom, R. E., & Muthuswamy, N. (2005). Undergraduates' rating of domestic and international teaching assistants: Timing of data collection and communication intervention. *Journal of Intercultural Communication and Research*, 34(1), 3–21.
- Solís Obiols, M. (2002). The matched guise technique: A critical approximation to a classic test for formal measurement of language attitudes. *Revista de Sociolingüística*, 1, 1–6.
- Thomson, R. (2018). Measurement of accentedness, intelligibility, and comprehensibility. In O. Kang & A. Ginther (Eds.), *Assessment in second language pronunciation* (pp. 11–29). Routledge.
- Wagner, E. (2014). Using unscripted spoken texts in the teaching of second language listening. *TESOL Journal*, 5(2), 288–311. <https://doi.org/10.1002/tesj.120>
- Wardhaugh, R., & Fuller, J. M. (2015). *An introduction to sociolinguistics* (7th ed.). Wiley Blackwell.
- Westfall J., Kenny D. A., & Judd C. M. (2014). Statistical power and optimal design in experiments in which samples of participants respond to samples of stimuli. *Journal of Experimental Psychology General*, 143(5), 2020–2045. <https://doi.org/10.1037/xge0000014>