3 Evaluations of Foreign Accented Speech: Listener Bias and Speech Signal Characteristics

Ron I. Thomson and Talia Isaacs

Abstract It is often reported that native speakers negatively evaluate personality characteristics of second language (L2) speakers on the basis of their accent. Researchers have frequently concluded that such judgements result from implicit bias on the part of listeners, for whom an L2 accent triggers stereotypes about the ethnic or racial origin of the speaker. Within this research paradigm, little attention has been paid to quantifiable features of the speech signal that may also contribute to negative evaluations of accented speakers, independent of their ethnic or racial origins. A more detailed understanding of listener reactions to foreign accented speech, and the underlying linguistic variables that influence those reactions, can reveal complementary contributions of speakers and listeners to personal interactions. In this chapter, we report results from an empirical study in which 24 native-speaking undergraduate students rated monologic speech samples, produced by 36 L2 English speakers (18 Mandarin, 18 Slavic), for perceived fluency, comprehensibility, friendliness, intelligence, and listeners' comfort interacting with each speaker. Relationships between listener ratings, and quantitative speech measures were examined across speakers, and as a function of speakers' language background and speaking task. Results indicate that quantifiable features of L2 speech and task type influence reactions to foreign-accented speech.

Keywords accent discrimination; L2 speech; foreign accent; comprehensibility; social dimensions of accent; stereotyping

1 Introduction

Experimental evidence has shown that listeners can detect a foreign accent after hearing as little as 30 milliseconds of second language (L2) speech (Flege 1984). Listeners can also recognize an L2 accent in speech played backwards (Munro et al. 2010). While some have described this skill as an accidental artifact of general speech processing mechanisms (Munro 2021), others contend that it is an evolutionary trait originally intended to thwart unintentional mating with those outside of one's own community (Scovel 1988). Recent research suggests that speakers of a majority language are less likely to form domestic partnerships with L2-accented speakers than they are with members of their own language community, and only slightly more likely to form friendships (Kogan et al. 2021). However, this appears to be a cultural rather than evolutionary phenomenon, since there is far less hesitation to marry someone of a different ethnicity if that person shares the same accent in some communities. At the furthest extreme, some argue that aversion to a foreign accent is triggered by racism (Ennser-Kananen et al. 2021).

While listener reactions to L2 speech can in part be influenced by cultural stereotypes formed in response to particular L2 accents, we take a more nuanced position. These reactions are also affected by speech signal characteristics. Non-native features either transferred from a speaker's first language (L1), or developmental in nature, may affect listener reactions when they do not match listeners' previous linguistic experience. Even if an L2-accented utterance is intelligible, it may still place increased processing demands on listeners, potentially leading to their frustration

Thomson, R. I., & Isaacs, T. (in press). Evaluations of foreign accented speech: Listener bias and speech signal characteristics. In A. Jarosz, & V. Sardegna (Eds.), *Theoretical and practical perspectives on English pronunciation teaching and research*. Springer.

and annoyance (Tulaja 2020). These reactions should be seen as independent of cultural, ethnic and racial triggers. In this chapter, we present a study which demonstrates that factors under the control of L2 speakers can contribute to listener judgments. While not wanting to minimize the potentially deleterious effect of listeners' implicit or overt biases, identifying speaker-dependent variables can inform L2 pronunciation instruction that empowers learners to take greater control over the speech they produce, in view of promoting more positive listener reactions.

2 Literature Review

2.1 Foreign accent and listener bias research

While listener reactions to L2 accents can be both positive and negative, research focussing on negative reactions is largely predicated on a belief that humans are fundamentally prejudiced against those outside of their own speech community. Within this attitudinal research paradigm, L2 accent is seen as a salient feature which allows listeners to automatically activate personal biases, which then inform implicit evaluations of a speaker's social identity (e.g., Dewaele & McCloskey 2015; Gluszek & Dovidio 2010; Lev-Ari & Keysar 2010; Lippi-Green 2012; Shah 2019).

Much of the research in this area relies on listener ratings of foreign-accented speech. Lambert et al. (1960) matched-guise technique has proven to be a popular procedure for obtaining these ratings. In this approach, listeners are asked to evaluate speech samples produced by bilingual or multilingual speakers of a target language who have recorded two or more versions of the same spoken text with different L1 and/or L2 accents. Recordings of multiple speakers are then randomized and played for listeners, who provide scalar judgments concerning inferred personal attributes and/or physical characteristics of the speakers. These attributes typically relate to speakers' status and/or solidarity, for example their social status or how friendly they are (Dragojevic & Goatley-Soan 2020). While the matched-guise technique controls for voice quality by including paired utterances produced by the same speaker, eliciting samples in this way is not always practical or possible. Verbal-guise techniques with single L2-accented speech samples from each speaker have also been used (Dragojevic & Goatley-Soan 2020; Garrett 2010). In these studies, the impact of any individual speaker's voice quality is assumed to average out over a large enough sample of L2 speakers.

Other techniques have also been used to elicit evidence of implicit bias towards L2-accented speakers. For example, even when a recorded speech sample is produced with a native accent, listeners can be prompted into perceiving an L2 accent if the recording is attributed to a picture of a foreign-looking speaker, a phenomenon Rubin (1992) termed Reverse Linguistic Stereotyping (RLS). Subsequent RLS research has demonstrated that a belief that a speaker is foreign can also impact native (Kang & Rubin 2009; Rubin 1992) and non-native (Ghanem & Kang 2021) listeners' comprehension of native speaker productions. To find evidence of bias, Lindemann (2005) avoided using aural stimuli altogether, and instead had listeners describe accents associated with countries she presented via a map. This only required that listeners refer to personal recollections of L2 accents in order to make judgments about the personal attributes of speakers with those accents.

Regardless of the precise technique used, research investigating attitudinal reactions to foreign-accented speech consistently arrive at similar conclusions. Speaking with a foreign accent results in listeners downgrading L2 speakers' perceived social status and/or the degree of solidarity with

the host or target language community (Dragojevic & Goatley-Soan 2020; Ryan 1983). Further, the extent to which an L2 speaker is downgraded reportedly depends on their ethnic origin. Lippi-Green (2012) and Lindemann (2005) argue that accents linked to non-Caucasian speakers are more likely to evoke negative evaluations than L2 accents associated with Caucasian speakers. Gilchrist and Chevrot (2017) demonstrate that explicit ethnic attribution, in which listeners are made aware of the ethnic background of the speaker, impacts assessment of speakers' global L2 proficiency. Specifically, judges assigned Arabic-accented English lower proficiency scores than Portuguese- and Chinese-accented English speakers on the basis of speech samples that contained exactly the same content.

While attitudinal research demonstrates that an L2 accent can trigger biases, which can then unduly affect listener reactions to the speakers, the strength of association may be exaggerated. First, these studies tend to be highly controlled and therefore lack ecological validity. In the real world, listeners' attention may not be as explicitly oriented towards a speaker's ethnicity or social identity as laboratory findings suggest. In fact, many listeners are not even able to accurately identify L2 accented speakers' ethnicity in the lab (Dragojevic & Goatley-Soan 2020; Gilchrist & Chevrot 2017; Lindemann 2003). Second, these studies rarely discuss individual differences across listeners, but instead focus on differences in group means. Not all listeners within a population sample respond in the same way (DeWaele & McCloskey 2015; Kang & Yaw 2021). Third, there is little overall focus on linguistic features stemming from the speech signal that may trigger positive or negative reactions quite apart from biases related to presumed group identity of L2-accented speakers. Thus, by convention, attitudinal research treats accent as a unidimensional, global phenomenon.

2.2 Impact of L2 accents on speech processing

While attitudinal research has largely focussed on listener bias, examining ways in which particular features of L2 accents impinge upon speech processing by listeners can provide a complementary and richer account for negative reactions (e.g., Kang 2012). Among researchers interested in L2 pronunciation learning and teaching, there is a widely-established literature evidencing multiple partially independent dimensions along which listeners respond to L2 speech (Derwing & Munro 2015; Thomson, 2018). Munro and Derwing (1995a, 1995b) draw a distinction between foreign accent (in relation to a target norm), comprehensibility (listeners' perception of effort in processing speech), and intelligibility (how much listeners actually understand). Foreign accent on its own may evoke listener bias, but poor comprehensibility and/or intelligibility can also trigger negative reactions independent of any assumptions about a speaker's ethnicity.

Numerous L2 pronunciation studies have demonstrated that distinctive phonological features of L2 speech affect listener perceptions of accent, comprehensibility and intelligibility in different ways (Derwing & Munro 2015; Isaacs & Thomson 2020; Kang et al. 2010; Kang et al. 2020; Levis 2018; Munro & Derwing 2006). While most L2 phonological features comprise negative transfer from the learners' L1s, some are developmental in nature, reflecting interlanguage patterns which may impact speakers across a variety of L1 backgrounds. Segmental, prosodic and temporal features of an L2 accent which are incongruent with a given listener's experience can cause processing difficulties. Derwing et al. (2009), for example, found that Mandarin L2 English speakers transferred L1 vowel length patterns to L2 English, negatively impacting their speech rate, which correlates with listeners' perception of fluency. Because the nature and extent of incongruence varies depending on each learner's L1, some L2 accents are more challenging for

listeners to process than are others. While the attitudinal research described earlier reports that non-Caucasian L2 English accents are downgraded more than Caucasian L2 English accents, is this solely the result of group bias, or does a greater phonological distance between English and specific non-Caucasian languages (e.g., Mandarin) contribute to this effect? The effect of L1-L2 phonological distance and strength of L2 accent has been considered in numerous pronunciation studies (Bongaerts et al. 2000; Bradlow et al. 2010; Cristia et al. 2012; Isaacs & Thomson 2020), but not to our knowledge in attitudinal research.

In the temporal domain, perceived oral fluency of L2 speakers has also been shown to interact with listener perceptions of accentedness, comprehensibility and intelligibility (Derwing & Rossiter 2003; Derwing et al. 2004; Thomson 2015). Attitudinal researchers have recognized the impact of perceived fluency (they call it 'processing fluency') on listener reactions to L2-accented speech (Alter & Oppenheimer 2009; Dovidio & Gluszek 2012). However, they seem to construe it as a shortcoming on the part of listeners, rather than something for which speakers bear some responsibility. In an effort to demonstrate a triggering effect of L2 identities on perceived fluency, Dragojevic and Goatley-Soan (2020) did not account for measurable differences in the L2 speakers' speech rate or proficiency. It is quite possible that those whose scores were downgraded by listeners were both less fluent and less proficient.

Other lines of research have similarly failed to consider the contribution of L2 speech processing difficulties on listener judgments. Pantos and Perkins (2013) used an implicit association test to demonstrate that response latencies to negative words were faster when associated with a foreign accent than with a native accent. However, is it the foreignness of the accent that is the issue, or might processing difficulty trigger negative emotions, which then drive listeners' association with negative words? Romero-Rivas et al. (2016) argue that difficulties listeners experience in anticipating upcoming words during sentence processing of L2 speech is caused by listeners activating negative affect in response to a speaker's accent. Could it not also be the case that properties of a particular accent are challenging, which would lead to identical results? Even evidence from neurolinguistics is used to support implicit biases. Foucart and Hartsuiker (2021) found differences in neurological activity during a sentence processing task when listeners were asked to judge the truth-value of true/false statements produced by native versus non-native speakers. They took these differences be an indicator of negative bias. Might these differences not simply be related to the greater effort that is sometimes required to process L2-accented speech? Others have found that familiarity with an L2 accent mitigates adverse reactions (Dewaele & McCloskey 2015; Kang & Yaw 2021). While they conclude that familiarity decreases bias, the fact that greater familiarity also leads to faster processing suggests that the appearance of bias might not be caused by the identity of the speaker but by psycholinguistic limitations on the part of the listener.

2.3 Task type

The nature of speaking tasks used in most attitudinal research makes determining the underlying causes of negative reactions to L2 accents difficult. Typically, this research relies upon highly controlled tasks to elicit predictable speaker output. While control is important for making comparisons across speakers, controlled tasks may not reflect reactions in real world communication. For example, a highly decontextualized reading task (e.g., Dragojevic and Goatley-Soan 2020) might interact with particular accents to induce a larger negative affect than is otherwise warranted. Heaton and Nygaard (2011) found that the specific content of a passage

can affect listener attitudes. This negative effect might not be replicated in response to potentially more engaging spontaneous speech. Another concern is that once a controlled passage is known to listeners, they will be better able to attend to pronunciation features of the speech sample, potentially increasing the saliency of accents. Listening to repetitive speaking tasks may also contribute to rater fatigue, which might be amplified in the case of foreign-accented samples.

3 Present study

While we do not dispute the contribution of linguistic stereotyping and discrimination as factors in how listeners evaluate L2-accented speech, we do not believe that it provides a complete account. In this exploratory investigation, we set out to determine if other factors might account for much of the variance in negative evaluations of L2-accented speech.

3.1 Research Questions

In this study, we examine reactions to L2 English speech samples produced by Mandarin and Slavic-accented speakers, performing two speaking tasks. We consider the influence of quantifiable features of the speech signal (i.e., speech rate and pitch) as they relate to listeners' perceptions of comprehensibility and fluency, their inferences about speakers' personality characteristics. Specifically, we asked:

1. Is there evidence of bias in listener reactions to Mandarin vs Slavic-accented English speakers?
2. Are listeners' judgments influenced by temporal and prosodic properties of L2 speakers' productions?
3. Do differences in L2 speaking performance across tasks contribute to different listener judgment patterns?

3.2 Method

3.2.1 L2 English Speakers

The L2 speech data were elicited from 36 adult newcomers to Canada who were enrolled in a government-funded ESL program. Half were Chinese (Mandarin) L1 speakers and half were Slavic L1 speakers (mostly Russians, but also two Ukrainians, two Serbians, a Pole and a Serbo Croatian). Apart from L1 differences, the groups comprised similar demographics. All were identified as beginners according to the Canadian Language Benchmarks 3-4 (CLB levels 1–4 of the instrument; Pawlikowska-Smith, 2000). The Mandarin group included 14 females and 4 males (*M* age 40.2 years; range 29-49). The Slavic group included 12 females and 6 males (*M* age 38.8 years; range 27-47). The mean age of first English exposure was 14 years (range 12-36) for the Mandarin group and 16 years (range 1-33) for the Slavic group, with most close to the mean. The L2 speakers' self-reported use of English outside of the classroom was also comparable, with approximately 1/3 of their daily communication in English, and more time spent watching English TV/Video than anything in their L1. One notable difference between the groups was in their estimated daily interactions with NSs outside of the classroom. The Mandarin group averaged less than one hour per day, while the Slavic groups averaged nearly two hours per day.

3.2.2 L2 Speaking Task

We used an eight-frame picture description task as well as a personal narrative task to elicit L2 speech. The picture sequence illustrates a humorous event in which a man and a woman mix up their identical suitcases after bumping into each other on a city street. This story has been widely used in previous research (Derwing et al. 2004; Isaacs & Thomson 2013, 2020). In the personal narrative task, participants were asked to describe their experiences during the first two weeks after their arrival to Canada. Recordings of all speaking tasks were made in a quiet room using a digital recorder, paired with a high quality unidirectional Sennheiser microphone. While recording length varied across participants, the picture description task was usually completed in less than 2 minutes, while the personal narrative tasks typically lasted between 2-3 minutes. Following Isaacs and Thomson (2013, 2020), we only used the first 20 seconds of each speaking task, after removing any initial false starts or other initial dysfluencies. This resulted in 144 items (36 speakers x 4 tasks). We created three randomizations of these items for presentation to raters. In each, we interspersed three recordings of native speakers completing the first picture description task. The native speaker items were used to ensure that the raters were scoring the correct speech sample, since we anticipated that the native speakers would receive high scores.

3.2.3 NS English Listeners

Twenty-four native English speaker listeners (21 female, 3 male) provided ratings of the L2 speech samples. All were undergraduate social science students at a mid-sized English-medium Canadian university (*M* age 22.8, range 19-49). Most were monolingual, although seven self-reported being fluent in a second language. All had spent the majority of their lives in Ontario, six were from the Toronto metropolis, and 20 from smaller cities. All reported normal hearing. None of the listeners had any previous formal experience rating L2-accented speech.

3.2.4 Rating Task

Three rating sessions were conducted in a quiet room, each with a group of eight raters. They were presented with one of three randomizations of the L2 recordings via loudspeaker and were asked to rate each speaker for fluency and comprehensibility during a first session, and friendliness, intelligence and how comfortable they were interacting with the speaker in a second session. They recorded their assessments on printed paper using 9-point Likert-type scales. Scales and their endpoints were as follows:

Fluency: Very dysfluent - very fluent
Comprehensibility: Very hard to understand to very easy to understand
Friendliness: Not very friendly - very friendly
Intelligence: Not very intelligent - very intelligent
Interactional comfort: Not very comfortable - very comfortable

We provided raters with very brief instructions at the beginning of the rating sessions, explaining that comprehensibility refers to how easy it is to understand a speaker, while fluency refers to how smooth the speaker's oral delivery is based on their use of pauses, hesitations, fillers, etc. We gave no guidance on how to interpret friendliness and intelligence as we took these to be subjective constructs. For the final category, we simply asked, 'How comfortable would you feel interacting with this person?' After giving these instructions, we had the group of raters listen to two examples and discuss together how they might rate the samples on the relevant scales. We also told the listeners that all L2 speech was produced by speakers of Chinese or Slavic origin.

Other than this brief introduction, raters were encouraged to indicate responses based on their subjective assessments of each speech sample. At the end of the second session, raters completed a short questionnaire in which they were asked to elaborate on factors that influenced their evaluation of each speech construct on each task, through both fixed and open-ended question types.

3.2.5 Discrete Measures of Speech

In addition to obtaining listener ratings, we extracted quantitative measures of the samples' temporal and acoustic characteristics. We used pruned syllables per second as a measure of speech rate. This was operationalized as the total speaking time divided by the number of fluent syllables produced (i.e., we did not count syllables comprising self-corrections, self-repetitions and nonlexical fillers such as 'um'). Among a wide variety of common speech rate measures, Derwing et al. (2004) found the pruned syllable measure to be the most strongly correlated with listener judgments of fluency. Total speaking time was measured using Sound Studio 3 and pruned syllables were calculated with reference to transcripts that had been created by a research assistant and verified by the first author.

We also calculated each speaker's minimum and maximum pitch (in Hz) and pitch range over the duration of each speaking task as a marker of affect (Ohala, 1983). Ohala found that higher pitch is associated with friendliness and politeness, while lower pitch is associated with confidence and dominance. Pitch measures were extracted using Praat Version 6. Automatic pitch tracks were used in a first pass, and manually corrected in some instances where the pitch tracker failed. After extracting pitch values for each speaker, male values were normalized to the female mean in order to combine data across all speakers.

3.3 Results

3.3.1 Interrater Reliability

We calculated interrater reliability for each rating scale using Cronbach's Alpha. Scores evidenced high overall consistency across raters as follows: Fluency (.94), Comprehensibility (.93), Friendliness (.88), Intelligence (.91), Interactional comfort (.90).

To examine L1 and task differences we computed a series of five partially repeated-measures ANOVAs, one for each speech/personality construct. Speaking task (2 levels) served as a within-subject factor, while L1 was a between-subject factor. Results (see Table 3.1) indicate a significant difference in ratings when comparing performance on the picture description versus the personal narrative task. Across all scales, speaker performance on the personal narrative task was always rated more favorably than on the picture description task (see Figure 3.1). For fluency, comprehensibility, intelligence and interactional comfort scales, Slavic-accented speakers were rated more positively than Mandarin-accented speakers, with small to medium effect sizes. For the friendliness scale, however, there was no significant difference between Slavic and Mandarin-accented speakers.

Table 3.1 Results of partially repeated measures ANOVAs comparing mean ratings for each task by L1group

	Task	L1

	F(1,34)	*p*	η2	*F*(1,34)	*p*	η2
Fluency	10.020	.003	.228	11.064	.002	.246
Comprehensibility	11.017	.002	.245	19.050	<.001	.359
Friendliness	11.939	.001	.260	.089	.768	.003
Intelligence	22.455	<.001	.398	9.703	.004	.222
Interaction comfort	14.617	<.001	.301	10.399	.003	.234

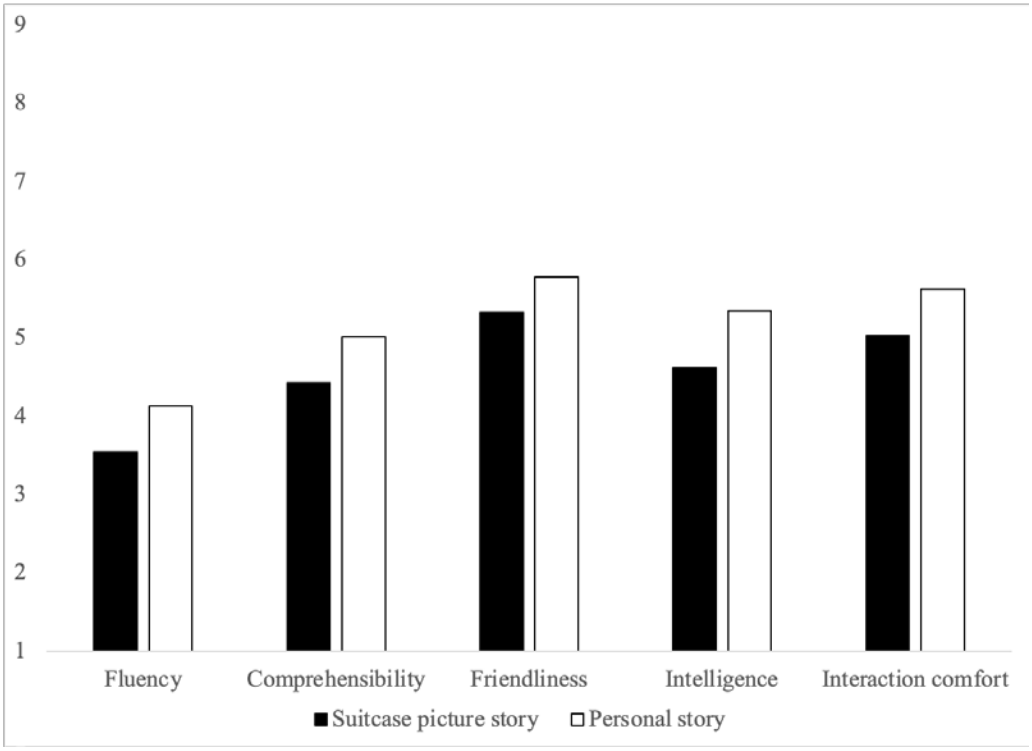

Fig. 3.1 Mean speech and personality scale ratings by task

To examine L1 and task differences for speech rate (pruned syllables/sec) and pitch range measures, we computed two partially repeated measures ANOVAs, one for each quantitative measure. Speaking task (2 levels) served as a within-subject factor, while L1 was a between-subject factor. Results (see Table 3.2) indicate a significant difference in speech rate, with a faster rate on the personal narrative task compared to the picture description task. No L1 effect for speech rate was detected, however. For maximum pitch a significant effect was found for Task, with the picture description having higher maximum pitch (*M* = 348 Hz) than the personal narrative (*M* = 321 Hz). No significant difference was found for L1. Speakers used a significantly larger pitch range on the picture description task (*M* = 271 Hz) relative to the personal narrative task (*M* = 243 Hz). Further, Slavic speakers used a significantly larger pitch range (*M* = 274 Hz) than Mandarin speakers (*M* = 239 Hz).

Table 3.2 Results of partially repeated measures ANOVA comparing speech rate and pitch measures X L1group

	Task	L1

Thomson, R. I., & Isaacs, T. (in press). Evaluations of foreign accented speech: Listener bias and speech signal characteristics. In A. Jarosz, & V. Sardegna (Eds.), *Theoretical and practical perspectives on English pronunciation teaching and research*. Springer.

	F(1,34)	*p*	η2	*F*(1,34)	*p*	η2
Speech rate (pruned syllables/sec)	10.047	.003	.228	.682	.415	.020
Max pitch	4.869	.034	.125	3.367	.075	.090
Mean pitch range	5.515	.025	.140	4.357	.044	.114

3.3.2 Multiple Regression Analyses

Stepwise linear multiple regression analyses were conducted to examine to what extent temporal characteristics of the produced speech (pruned syllables/sec) and pitch measures (maximum, minimum and range) predicted ratings on each task. On the suitcase picture description task (see Table 3.3), pruned syllables/sec were strong predictors of fluency and comprehensibility. A combination of pruned syllables/second and maximum pitch range strongly predicted intelligence and interactional comfort ratings, but only weakly predicted friendliness ratings. On the personal narrative task (See Table 3.4), pruned syllables were weaker predictors of fluency comprehensibility, intelligence and interactional comfort. Maximum pitch along with pruned syllables/sec combined to weakly predict friendliness ratings on the personal narrative task.

Table 3.3 Multiple regression of variables contributing to listener reactions to the picture description task

		Standardized Coefficients			
	Predictors Stepwise (R^2)	Standardized coefficients (β)	*t* value	*p* value	Partial correlation
Fluency	Pruned syllables (.677)	.823	8.439	<.001	.823
Comprehensibility	Pruned syllables (.810)	.900	12.052	<.001	.900
Friendliness	Max pitch (.147)	.449	3.084	.004	.473
	Pruned syllables (.317)	.418	2.871	.007	.447
Intelligence	Pruned syllables (.644)	.851	9.392	<.001	.853
	Max pitch (.736)	.307	3.392	.002	.508
Interaction comfort	Pruned syllables (.520)	.783	7.729	<.001	.721
	Max pitch (.670)	.392	3.873	<.001	.269

In addition to linear regression analyses, descriptive Pearson correlation coefficients also revealed significant relationships between pitch range and friendliness ratings on the picture description task (.373) and personal narrative task (.403). No significant correlations were found with minimum or maximum pitch, however.

Table 3.4 Multiple regression of variables contributing to listener reactions to personal narrative task

		Standardized Coefficients			
	Predictors Stepwise (R^2)	Standardized coefficients (β)	*t* value	*p* value	Partial correlations
Fluency	Pruned syllables (.436)	.660	5.124	<.001	.660
Comprehensibility	Pruned syllables (.241)	.491	3.287	.002	.491
Friendliness	Max pitch (.187)	.454	3.228	.003	.490
	Pruned syllables (.349)	.403	2.863	.007	.446
Intelligence	Pruned syllables (.463)	.680	5.415	<.001	.680
Interaction comfort	Pruned syllables (.337)	.580	4.154	<.001	.580

Looking at simple correlation plotlines, we observed a few outliers that did not fit the overall patterns. Thus, we examined the transcripts of four samples comprising the greatest mismatch between attitudinal ratings and related quantitative measures. Following Heaton and Nygaard (2011) we found some evidence that content may have played a role in the evaluation of speakers' friendliness and intelligence. For example, Sp45 was the 36th most monotone speaker, yet was rated as the sixth friendliest on the personal narrative task. At the beginning of his story, this speaker said, "Canada uh knocked off my socks. Haha. When I arrived in Canada people was very friendly." Such content may have caused positive affect among raters. Sp38 was the 35th most monotone, yet was rated as the 11th friendliest on the personal narrative. In his story he states "Our family immigrated to Canada…I met the some people. They are very nice and kindness." We found similar mismatches between intelligence ratings and quantitative measures. Sp21 was rated as the 9th most intelligent, despite being ranked 27th in terms of pruned syllables/sec. The speaker said, "when I came here, I didn't know nobody. I, I must rent apartment, and meet new people and find, must find work." The content seems to describe a confident and/or independent person. Conversely, Speaker 37 was rated as 27th in terms of intelligence, despite producing the 12th highest pruned syllables/second. She related, "at first I was very sad and nobody helped me and where the going. I no idea."

3.3.3 Questionnaire Data

Responses to the fixed-choice portion of our post-rating questionnaire also revealed how particular features of the oral texts may have contributed to listener reactions. Most raters (88%) indicated that even if individual words were clearly pronounced, a lack of coherence in the stories affected their ratings. Nearly two-thirds (63%) indicated being impacted by how fluently a speaker proceeded through the story, while almost all (92%) indicated positive affect when a speaker spent time developing details of a story (e.g., the "beautiful city" or "tall buildings"). The same number (92%) were negatively impacted by incorrect word choice for important words (e.g., "bit each other" instead of "bumped into each other").

Responses to an open-ended question asking for raters' top two influences on their ratings also revealed a diversity of influences. Some (38%) explicitly referenced fluency as a determinant in their ratings. One noted that "fluency was very important to delivery" while another stated that fluency might reflect it being "harder to tell the [picture story] than to tell their own personal

experience." Many raters (42%) commented on how easy or difficult it was to understand the speech samples. One indicated that "speaking clearly and pronouncing words correctly made the story easier to understand" while another stated, "more cohesion means it's easier to communicate using Standard English." Another rater said, "if I couldn't understand them I had a more difficult time listening to their story." Only a few raters made comments related to influences on friendliness ratings (12.5%) and intelligence ratings (12.5%). One suggested that how much the speaker "enjoyed telling [the story]" influenced how friendly they sounded, while another pointed out that monotone speech "didn't seem as friendly." One rater stated that "the story sounded better if [speakers] knew what they were talking about," while another said "when they sound intelligent the story is easier to listen to."

4 Discussion

Traditional attitudinal researchers might interpret our listeners' scalar ratings as providing evidence that listeners activate implicit (or explicit) biases that associate Caucasian, Slavic-accented speakers with greater intelligence, interactional comfort, fluency and comprehensibility relative to non-Caucasian, Mandarin-accented speakers (Lindemann 2005; Lippi-Green 2012). Differences in quantifiable features of the speech samples provide important insight, however. While we found no measurable L1-based differences in speech rate, Slavic speakers had a significantly wider pitch range than the Mandarin speakers. Interestingly, this itself is contrary to stereotypes that Slavic speakers are monotone (Crosby 2013; Svetozarova 1998) and to evidence that pitch range in Mandarin is much wider than that of English (Chen 1974). Multiple regression analyses revealed that a combination of pruned syllables and speakers' maximum pitch combined to strongly predict intelligence and interactional comfort ratings, and that pruned syllables account for much of the variance in fluency and comprehensibility ratings. This suggests that much of the difference in attitudinal reactions favouring Slavic accented speakers in our study are attributable to differences in how each L1 group controls L2 English speech rate and pitch. While there were no overall differences in how friendly Slavic and Mandarin speakers were perceived to be, maximum pitch and pitch range influenced friendliness ratings across both groups. Finally, we found that ratings were higher on all scales for the personal narrative task than for the picture description task. Examination of the content of L2 oral productions and rater questionnaire data added further nuance, suggesting that the nature of the speaking task plays a crucial role in ratings and that some tasks allow speakers to create solidarity with raters by the things that they say during the task (e.g., that they are happy in Canada). This suggests that the content of an L2 utterance or their overall message could affect listeners judgments of L2 speech, unduly resulting in higher or lower ratings when this is extraneous to the L2 speaking construct being measured (e.g., fluency, comprehensibility).

5 Implications

Like Derwing and Munro (2015), we see communication as a two-way street. In their primary research context, involving immigrant language learning, they are rightly concerned that not enough emphasis is given to listeners' responsibility in accommodating L2 accents. In attitudinal research, however, the opposite seems to be the case. Little attention is paid to what speakers can do to make themselves more intelligible. In our exploratory study, we found evidence of listener bias in reactions to L2-accented speech, but also contributions from quantifiable features of the speech signal produced by learners. We only examined two features of speech, temporal fluency

and pitch, however. There are many more segmental and prosodic features that might influence listener reactions (Derwing & Munro 2015; Kang et al. 2010; Kang et al. 2020). To the extent to which L2 learners want to (e.g., McCrocklin & Link 2016) and are able to change features of their pronunciation that induce negative reactions by listeners, they should be encouraged to do so, just as listeners should be encouraged to become more tolerant with L2 accented speakers. The good news is that L2 pronunciation instruction can be quite effective (Thomson & Derwing 2015), and if aimed at improving speakers' intelligibility or comprehensibility (rather than global accent), it is often worth the effort. One means to help listeners is to encourage them to have more interaction with L2-accented speakers, since there is evidence that familiarity will make processing more efficient (Porretta et al. 2017). Familiarity can also lead to measurably less bias in listener responses to L2-accented speakers (Dewaele & McCloskey 2015; Kang & Yaw 2021).

6 Conclusion and limitations

The results of this study suggest that there is a need for more research to tease apart the relative impacts of attitudinal bias versus quantifiable L2 speech characteristics that influence reactions to foreign-accented speakers. Listeners' reactions to positive versus negative framing of the host community's culture in an L2 speech sample is one area that would benefit from further examination in both controlled experiments, and in more in-depth qualitative data. One limitation in the current study is its lack of a native accented comparison group. Further, we did not consider listener familiarity with Mandarin and Slavic accents.

References

Alter, A. L., & Oppenheimer, D. M. (2009). Uniting the tribes of fluency to form a metacognitive nation. *Personality and Social Psychology Review*, 13(3), 219–235.

Bongaerts, T., Mennen, S., & van der Slik, F. (2000). Authenticity of pronunciation in naturalistic second language acquisition: The case of very advanced late learners of Dutch as a second language. *Studia Linguistica*, 54(2), 298–308. https://doi.org/10.1111/1467-9582.00069

Bradlow, A., Clopper, C., Smiljanic, R., & Walter, M. A. (2010). A perceptual phonetic similarity space for languages: Evidence from five native language listener groups. *Speech Communication*, 52(11–12), 930–942. https://doi.org/10.1016/j.specom.2010.06.003

Chen, G. T. (1974). The pitch range of English and Chinese speakers. *Journal of Chinese Linguistics*, 159–171.

Cristia, A., Seidl, A., Vaughn, C., Schmale, R., Bradlow, A., & Floccia. C. (2012). Linguistic processing of accented speech across the lifespan. *Frontiers in Psychology*, 3, 1–15.

Crosby, C. F. (2013). *L1 influence on L2 intonation in Russian speakers of English.* [MA Thesis]. Portland State University.

Derwing, T. M., & Munro, M. J. (2015). *Pronunciation fundamentals: Evidence-based perspectives for L2 teaching and research*. Amsterdam: John Benjamins Publishing Company.

Derwing, T. M., Munro, M. J., Thomson, R. I., & Rossiter, M. J. (2009). The relationship between L1 fluency and L2 fluency development. *Studies in Second Language Acquisition*, 31(4), 533–557.

Derwing, T. M., & Rossiter, M. J. (2003). The effects of pronunciation instruction on the accuracy, fluency, and complexity of L2 accented speech. *Applied Language Learning*, 13(1), 1–17.

Thomson, R. I., & Isaacs, T. (in press). Evaluations of foreign accented speech: Listener bias and speech signal characteristics. In A. Jarosz, & V. Sardegna (Eds.), *Theoretical and practical perspectives on English pronunciation teaching and research*. Springer.

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.

Dewaele, J. M., & McCloskey, J. (2015). Attitudes towards foreign accents among adult multilingual language users. *Journal of Multilingual and Multicultural Development*, 36(3), 221–238.

Dovidio, J. F., & Gluszek, A. (2012). Accents, nonverbal behavior, and intergroup bias. In H. Giles (Ed.), *The handbook of intergroup communication* (pp. 109–121). New York: Routledge.

Dragojevic, M., & Goatley-Soan, S. (2020). Americans' attitudes toward foreign accents: evaluative hierarchies and underlying processes. *Journal of Multilingual and Multicultural Development*, DOI: 10.1080/01434632.2020.1735402

Ennser-Kananen, J., Halonen, M., & Saarinen, T. (2021). "Come join us and lose your accent!": Accent modification courses as hierarchization of international students. *Journal of International Students*, 11, 322–340.

Flege, J. E. (1984). The detection of French accent by American listeners. *Journal of the Acoustical Society of America*, 76(3), 692–707.

Foucart, A., & Hartsuiker, R. J. (2021). Are foreign-accented speakers that 'incredible'? The impact of the speaker's indexical properties on sentence processing. *Neuropsychologia*, 107902.

Garrett, P. (2010). *Attitudes to Language*. New York, NY: Cambridge University Press.

Ghanem, R., & Kang, O. (2021). ESL students' reverse linguistic stereotyping of English teachers. *ELT Journal*. Advanced Online Access. https://doi.org/10.1093/elt/ccab011

Gilchrist, C., & Chevrot, J. P. (2017). Snap judgment: Influences of ethnicity on evaluations of foreign language speaking proficiency. *Corela. Cognition, Représentation, Langage*, 15, (1). https://doi.org/10.4000/corela.4920

Gluszek, A., & Dovidio, J. F. (2010). The way they speak: A social psychological perspective on the stigma of nonnative accents in communication. *Personality and Social Psychology Review*, 14, 214–237.

Heaton, H., & Nygaard, L. C. (2011). Charm or harm: Effect of passage content on listener attitudes toward American English accents. *Journal of language and social psychology*, 30(2), 202–211.

Isaacs, T., & Thomson, R. I. (2013). Rater experience, rating scale length, and judgments of L2 pronunciation: Revisiting research conventions. *Language Assessment Quarterly*, 10(2), 135–159.

Isaacs, T., & Thomson, R. I., (2020). Reactions to second language speech: Influences of discrete speech characteristics, rater experience, and speaker first language background. *Journal of Second Language Pronunciation*, 6(3), 402–429.

Kang, O. (2012). Impact of rater characteristics and prosodic features of speaker accentedness on ratings of international teaching assistants' oral performance. *Language Assessment Quarterly*, 9(3), 249–269.

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.

Kang, O., Rubin, D., & Pickering, L. (2010). Suprasegmental measures of accentedness and judgments of language learner proficiency in oral English. *The Modern Language Journal*, 94(4),

554–566.

Kang, O., Thomson, R. I., & Moran, M. (2020). Which features of accent affect understanding?: Exploring the intelligibility threshold of diverse accent varieties. *Applied Linguistics*, 41(4), 453–480.

Kang, O., & Yaw, K. (2021). Social judgement of L2 accented speech stereotyping and its influential factors. *Journal of Multilingual and Multicultural Development*, 1–16. https://doi.org/10.1080/01434632.2021.1931247

Kogan, I., Dollmann, J., & Weißmann, M. (2021). In the ear of the listener: The role of foreign accent in interethnic friendships and partnerships. *International Migration Review*, 55(3), 746–784.

Lambert, W. E., Hodgson, R., Gardner, R. C., & Fillenbaum, S. (1960). Evaluational reactions to spoken languages. *Journal of Abnormal and Social Psychology*, 60, 44–51.

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, 1093–1096.

Levis, J. M. (2018). *Intelligibility, oral communication, and the teaching of pronunciation*. Cambridge: Cambridge University Press.

Lindemann, S. (2003). Koreans, Chinese, or Indians? Attitudes and ideologies about non-native English speakers in the United States. *Journal of Sociolinguistics*, 7, 348–364.

Lindemann, S. (2005). Who speaks "Broken English"? US undergraduates' perceptions of non-native English. *International Journal of Applied Linguistics* 15, 187–212.

Lippi-Green, R. (2012). *English with an accent: Language, ideology and discrimination in the United States*. Second Edition. New York: Routledge.

McCrocklin, S., & Link, S. (2016). Accent, identity, and a fear of loss? ESL students' perspectives. *Canadian Modern Language Review*, 72(1), 122–148.

Munro, M. J. (2021). *Applying phonetics: Speech science in everyday life*. Hoboken, NJ: Wiley-Blackwell.

Munro, M. J., & Derwing, T. M. (1995a). Foreign accent, comprehensibility, and intelligibility in the speech of second language learners. *Language Learning*, 45(1), 73–97.

Munro, M. J., & Derwing, T. M. (1995b). Processing time, accent, and comprehensibility in the perception of native and foreign-accented speech. *Language and Speech*, 38, 289–306.

Munro, M. J., & Derwing, T. M. (2006). The functional load principle in ESL pronunciation instruction: An exploratory study. *System*, 34(4), 520–531.

Munro, M. J., Derwing, T. M., & Burgess, C. S. (2010). Detection of nonnative speaker status from content-masked speech. *Speech Communication*, 52, 626–637.

Ohala, J. J. (1983). Cross-language use of pitch: an ethological view. *Phonetica* 40(1), 1–18.

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

Pantos, A. J., & Perkins, A. W. (2013). Measuring implicit and explicit attitudes toward foreign accented speech. *Journal of Language and Social Psychology*, 32(1), 3–20.

Pawlikowska-Smith, G. (2000). *Canadian Language Benchmarks 2000: Theoretical framework*.

Ottawa, ON: Centre for Canadian Language Benchmarks.

Porretta, V., Tremblay, A., & Bolger, P. (2017). Got experience? PMN amplitudes to foreign-accented speech modulated by listener experience. *Journal of Neurolinguistics*, 44, 54–67.

Romero-Rivas, C., Martin, C. D., & Costa, A. (2016). Foreign-accented speech modulates linguistic anticipatory processes. *Neuropsychologia*, 85, 245–255.

Ryan, E. B. (1983). Social psychological mechanisms underlying native speaker evaluations of non-native speech. *Studies in Second Language Acquisition,* 5, 148–159.

Scovel, T. (1988). *A time to speak: A psycholinguistic inquiry into the critical period for human speech*. New York: Newbury House.

Shah, A. P. (2019). Why are certain accents judged the way they are? Decoding qualitative patterns of accent bias. *Advances in Language and Literary Studies*, 10(3), 128–139.

Svetozarova, N. (1998). Intonation in Russian. In D. Hirst & A. di Christo (Eds.), *Intonation systems: A survey of twenty languages* (pp. 261–274). Cambridge, England: Cambridge University Press.

Thomson, R. I. (2015). Fluency. In M. Reed & J. Levis (Eds.), *The Handbook of Pronunciation.* (pp. 209–226). Hoboken, New Jersey: Wiley.

Thomson, R. I. (2018). Measurement of accentedness, intelligibility and comprehensibility. In O. Kang& A. Ginther. (Eds.), *Assessment in second language pronunciation* (pp. 11–28). New York: Routledge.

Thomson, R. I., & Derwing, T. M. (2015). The effectiveness of L2 pronunciation instruction: A narrative review. *Applied Linguistics*, 36(3), 326–344.

Tulaja, L. (2020). Exploring acceptability: L1 judgements of L2 Danish learners' errors. In O. Kang, S. Staples, K. Yaw, & K. Hirschi (Eds.), *Proceedings of the 11th Pronunciation in Second Language Learning and Teaching conference*, ISSN 2380-9566, Northern Arizona University, September 2019 (pp. 197–206). Ames, IA: Iowa State University.