1 Truthfulness judgments are negligibly affected by talker traits

2 Khia A. Johnson, Gloria M. Mellesmoen, and Molly Babel

3 Department of Linguistics, University of British Columbia

4 Author Note

5 Khia A. Johnson, <https://orcid.org/0000-0001-6225-9446>

6 Gloria M. Mellesmoen, <https://orcid.org/0000-0002-3165-2234>

7 Molly Babel, <https://orcid.org/0000-0003-1153-9557>

8 Funding: Canadian Social Sciences and Humanities Research Council Grant to Molly
9 Babel (435-2017-0136)

10 Corresponding author: Molly Babel, Department of Linguistics, University of British
11 Columbia, 2613 West Mall, Vancouver, BC, Canada V6T 1Z4. Phone: +1 (604) 822-6322. Email:
12 molly.babel@ubc.ca.

Abstract

13
14 This study examines whether social evaluation and intelligibility affect judgments of
15 perceived credulity. Canadian English listeners (i) completed a speech-in-noise task, (ii) judged if
16 sentences were true, and (iii) rated voices on social dimensions. A Bayesian analysis showed that
17 neither social evaluation nor intelligibility substantially influenced perceived truthfulness. While
18 recent work suggests talker accent may impact perceived credulity, the question is far from settled
19 in the literature. We conclude that perceived credulity is negligibly affected by assessments of
20 natively-accented voices.

Truthfulness judgments are negligibly affected by talker traits

Introduction

Spoken language incorporates linguistic, social, and affective information in a single acoustic signal. While older models of spoken language processing hypothesized robust normalization that dissociates linguistic information from social and talker-specific indices (Norris, 1994), contemporary models posit an integrated processing of social cues and linguistic content (Johnson & Mullennix, 1997; Kapnoula & Samuel, 2019; Pierrehumbert, 2016; Sumner, Kim, King, & McGowan, 2014). Listeners are sensitive to social dimensions in online processing, and are biased by learned social associations at various levels of linguistic structure (Hanulíková, Van Alphen, Van Goch, & Weber, 2012; Johnson, Strand, & D’Imperio, 1999; Van Berkum, Van den Brink, Tesink, Kos, & Hagoort, 2008). Listeners learn to parse linguistic meaning in the context of the social characteristics carried by a talker’s voice.

Social psychologists have long-studied how a voice’s social characteristics affect the assessment of speech. Disfluencies, for example, are known to negatively affect perceived credulity on competence and dynamism metrics, but less so for trustworthiness (Miller & Hewgill, 1964), with these assessments being disconnected from the ultimate persuasiveness of an individual’s speech (Serenio & Hawkins, 1967). Non-standard or unfamiliar accents can provide challenges to intelligibility or comprehension, but not always in a way that is well-correlated with perceived accentedness (Derwing & Munro, 1997; Major, Fitzmaurice, Bunta, & Balasubramanian, 2005; Munro & Derwing, 1999), highlighting the unique contributions of different voice assessments language comprehension and how impressions are formed. Intelligibility is also affected by paralinguistic social context for a voice (Babel & Russell, 2015; Kang & Rubin, 2009; McGowan, 2015). A meta-analysis by Fuertes, Gottdiener, Martin, Gilbert, and Giles (2012) confirms that accents elicit evaluative judgments, particularly in the American context. Social evaluations of voices and individuals have real world implications (see Craft, Wright, Weissler, & Queen, 2020).

With the goal of better understanding evaluative voice assessment within theories of speech

and language processing—such that explanatory mechanisms can be provided—we evaluate the extent to which voice evaluations of a group of same-accent talkers impact listener judgments regarding the perceived credulity or truthfulness of trivia statements in a Canadian English context. The exact mechanism or process by which social and affective dimensions interact with linguistic processing is debated. Accents associated with low prestige have been argued to be encoded less robustly (Sumner & Kataoka, 2013), but it is unclear whether familiarity or prestige is the more relevant factor (Babel, Senior, & Bishop, 2019) or whether methodological choices invite particular listener strategies (Zheng & Samuel, 2017). Talkers considered to be heavily accented are not always less intelligible (Munro & Derwing, 1999), suggesting that non-native speech does not *necessarily* impede linguistic processing.

If accentedness alone is not a barrier to intelligibility, a logical follow-up question is to ask *how* social attitudes come to affect speech comprehension. Dragojevic and Giles (2016) propose and test a *processing fluency hypothesis*, which provides an explanatory mechanism for how attitudes towards a voice impact linguistic processing. Dragojevic and Giles (2016) found that when speech is challenging to understand—due to noise or an unfamiliar accent—a listener’s experience can be negative. Altogether, these results indicate that both adverse listening conditions and accent stereotypes can exert independent negative impacts on affect and attitudes. Dragojevic and Giles (2016) suggest that when listeners encounter processing difficulty with natively-accented talkers, they more readily attribute that difficulty to ephemeral or incidental reasons, making it less likely to influence affect and attitudes about the speaker.

Accent stereotypes for the non-native accents of L2 speakers are often pejorative (Bouchard Ryan, Carranza, & Moffie, 1977; Lippi-Green, 2012). Despite this, researchers often use non-native accents as stimuli without fully considering the implications of social evaluation. For example, Lev-Ari and Keysar (2010) used non-native accents to investigate perceptual fluency and truthfulness judgments of trivia statements. They found that listeners were less likely to judge sentences as true when produced by individuals with non-native accents. Lev-Ari and Keysar (2010) interpret their results as increased processing difficulty causing a decrease in perceived

truthfulness. Their argument states that non-native accents disrupt listeners mapping of the speech signal to linguistic categories, which has cascading effects on their ability to assess the truthfulness of utterances. One caveat is that Lev-Ari and Keysar (2010) did not include any social evaluation of the speakers, so it could be that individual voices used in their study elicited negative evaluations related to talker-specific idiosyncrasies or accent-associations. It is unclear whether Lev-Ari and Keysar's (2010) findings apply to all non-native accents (or dialects), as might be expected if they are indeed solely caused by challenges in mapping speech to linguistic categories, or whether the results hinge on stereotypes about particular accents (non-native or not).

The speech stream inherently provides a multidimensional array of paralinguistic information that prompts implicit social evaluations (Bourhis & Giles, 1977; Giles & Trudgill, 1983; Kinzler & DeJesus, 2013), and these dimensions may impact listeners' assessment of whether a statement is truthful. These social evaluations can contribute to damaging social categorizations, leading to linguistic profiling (Baugh, 2000, 2003). Rickford and King (2016) describe the language patterns of Rachel Jeantel, the key witness in *State of Florida v. Zimmerman* (2013), and how jurors difficulty in understanding her speech—resulting from a combination of accent differences and racism—discredited her as a witness in the eyes of the jury.

The multiple degrees of freedom presented in using non-native accents as stimuli could translate into replication issues. Indeed, Souza and Markman (2013) struggled to replicate Lev-Ari and Keysar (2010)'s work with Brazilian- and Korean-accented English. They also presented a native English voice in noise at three different signal-to-noise ratios (SNRs); the lower SNRs reduce intelligibility, which following Lev-Ari and Keysar (2010), should negatively impact the perceived credibility of the trivia statements—following the proposed mechanism that perceptual disfluency interrupts mapping to linguistically meaningful categories—but this was not the case. Souza and Markman's (2013) results provided no evidence that speech rendered less intelligible—via noise or accent—leads to lower truthfulness ratings. Note, however, that others have replicated Lev-Ari and Keysar's (2010) basic findings (e.g., Hanzlíková & Skarnitzl, 2017).

All together, the research reviewed above often compares native and non-native talkers,

sampling a small number from any accent group (often just one voice), thus inhibiting broad generalizations about accent groups as a whole. Individual voices differ not only in their social indexing, but also in how pronunciation and talker-specific differences affect intelligibility, even within uniform dialects (McCloy, Wright, & Souza, 2015). Subtle pronunciation differences within a dialect contribute to variable social characterizations and can affect word recognition (e.g., Vaughn & Kendall, 2018). A path towards developing an understanding of how social evaluation and perceptual fluency affect higher-level assessments, such as perceived truthfulness, is to use native-accented talkers, which reduces the amount of variability and the number of potential social dimensions relevant to the interpretation of results.

This study assesses whether social evaluation of a talker affects listeners' truthfulness judgments, using utterances produced by ten native Canadian English talkers. As a way to quantify the role of perceptual fluency in truthfulness judgments, the proportion of keywords correctly identified in a speech-in-noise transcription task is used as a measure of intelligibility. The overarching goal of this work is to improve our understanding of how social evaluation and intelligibility conspire together in ways that have consequential effects on how an individual's voice is handled by listeners in terms of credulity. Bayesian statistical analyses are deployed to better understand effect sizes and the relative importance of different social evaluation and intelligibility measures, as opposed to the significant-or-not frequentist approach. A Bayesian analysis also allows for supporting a null result—here, that such measures don't impact credibility—which is crucial given the replication issues in this body of research.

Methods

In this study, participants completed a series of three tasks in which they made a variety of decisions after listening to sentences produced by native Standard Canadian English speakers. The tasks included (1) speech-in-noise transcription, (2) truthfulness judgments, and (3) social evaluation of the talkers on a visual-analogue scale.

Participants

The tasks were completed by 82 self-identified native speakers of English with no self-identified speech disorders or hearing impairments. A total of 62 participants self-identified as female and 20 as male. They were university students ranging in age from 18 to 45 ($\mu = 21$, $\sigma = 3.5$), and they were compensated with partial course credit for their participation.

Materials

Ten native speakers of Standard Canadian English (five female, five male) were recorded at a 44.1 kHz sampling rate using a head-mounted microphone and pre-amplifier connected to a PC. Each speaker produced 88 target sentences. The sentences were short true and false trivia statements (e.g., “Snails can sleep for a decade”) adapted from Lev-Ari and Keysar (2010) and online trivia sources.¹ Half of the sentences were objectively true and half were false, and they were selected for their obscurity.

All audio files were root-mean-square amplitude normalized to a level that ensured no samples were clipped (≈ 65 dB). Additionally, a separate version was produced for the intelligibility block, where sentences were embedded in speech-shaped noise at 0 dB SNR.

Procedure

Participants completed three tasks in blocks with a fixed order, the order of which was determined by task demands. The transcription block necessarily comes first to control for familiarity with the voices and sentences. The social evaluation of voices must come last so that the overt assessment of these voices does not effect performance on the other tasks.

The first task was a speech-in-noise transcription task, and was included to provide a measure of intelligibility. In this block, listeners were presented with a random sample of 20 unique true sentences in 0 dB SNR speech-shaped noise and asked to type what they heard as accurately as possible, with the caveat that they were instructed to not worry too much about

¹ The full list of stimuli are available at (link to be provided post-review).

151 minor spelling and punctuation mistakes. To alleviate memory load, participants were able to
152 begin typing before the sentence was over.

153 In the second block, listeners made categorical true/false judgments. This block consisted
154 of 40 sentences (20 true, 20 false) that were not used during the first block. These sentences were
155 presented without background noise and fully randomized by sentence type and talker. Listeners
156 categorized each sentence as true or false using a button-box. Participants had up to five seconds
157 to make a response.

158 In the third block, listeners evaluated the voices on a Visual Analogue Scale (VAS; Massaro
159 & Cohen, 1983). The evaluations were along seven social dimensions: (i) Unattractive/Attractive;
160 (ii) Low/High Social Status; (iii) Unintelligent/Intelligent; (iv) Uninteresting/Interesting; (v)
161 Strong/No Foreign Accent; (vi) Unpleasant/Pleasant; and (vii) Untrustworthy/Trustworthy. All
162 sentences evaluated using the VAS were randomly selected from the pool of true sentences
163 presented in the T/F task. The order of VAS evaluation metrics and the talker order was
164 randomized for each listener, and all speakers were evaluated on a single dimension before
165 moving to the next VAS scale. A single randomly selected sentence was used for all voices for
166 each social dimension. Thus, each listener was presented with 7 unique sentences during this
167 block (one for each VAS scale) and was presented with each of those sentences 10 times (once for
168 each speaker).

169 Upon completion of the perception experiment, listeners were presented with a survey that
170 included all of the sentences presented in the task. For each sentence, participants identified
171 whether they “knew” the sentence was true, false, or did not know—this “knowledge” does not
172 necessarily coincide with whether a given statement was actually true or false. Participants also
173 completed a language background questionnaire.

174 Results

175 The goal of this analysis is to understand how social evaluations and intelligibility impact
176 the judgement of an utterance as truthful, where utterances comprise obscure true and false trivia

statements. This section first summarizes the social evaluations and intelligibility scores, and then reports the results of a Bayesian multilevel logistic regression model.

Intelligibility and social evaluations

Transcription responses were corrected for obvious misspellings and scored as the proportion of key content words in each utterance transcribed correctly. Mean transcription accuracy values were calculated for each talker-participant pair—this is considered a by-talker measure of intelligibility. Figure 1 shows the participants' distribution of intelligibility scores for each talker. Some talkers are consistently rated as highly intelligible (e.g., 106), while others are rated more variably (e.g., 107).

Figure 1. Density plots depict the distribution of intelligibility by talker. Solid lines indicate female talkers. Dashed lines indicate male talkers.

Raw VAS ratings in pixel counts for the social evaluations ranged from 392 to 1432. As the center value was visually marked on the screen, it is used as a meaningful landmark. Participants

identified the central tick mark ($\mu = 946, \sigma = 10$) in the VAS familiarization phase, and all ratings were centered around this meaningful value and scaled (i.e., divided by the overall standard deviation). Figure 2 shows the distribution of centered and scaled evaluation ratings for each social dimension and talker. While some talkers are consistently rated above the center point (e.g., having no foreign accent, sounds intelligent), ratings are highly variable across participants and dimensions.

Figure 2. Density plots show the distribution of centered and scaled VAS ratings for each perceptual dimension by talker. Distributions plotted with solid lines indicate female talkers.

Globally, the social dimensions and intelligibility are not significantly correlated with one another (Pearson's r , all $p > 0.05$). Figure 3 depicts the pairwise correlations, drawing on all listener ratings across talkers. While this does not the importance of these measures for perceived

197 truthfulness, it suggests they are orthogonal to one another.

Figure 3. Pairwise correlations for the social dimensions and intelligibility. A correlation's strength is indicated in each box—none were significant.

198 Bayesian model

199 The data were analyzed with a Bayesian multilevel logistic regression model implemented
 200 with the *brms* R package (Burkner, 2017). The model predicts judgments of perceived
 201 truthfulness, and estimates group-level intercepts for each Participant, Talker, and Sentence.
 202 Population-level effects included the seven social evaluation measures (numeric; Accented,
 203 Attractive, Intelligent, Interesting, Pleasant, Status, and Trustworthy), Intelligibility (numeric),
 204 and Knowledge of the trivia statement (treatment coded; Unsure: *reference level*, True, or False).
 205 Knowledge refers to whether participant indicated that they “knew” a sentence to be true or false,
 206 or were unsure, regardless of whether or not the sentence was actually true or false. Recall that all
 207 numeric population-level effects were centered around a meaningful landmark and scaled. There
 208 were no interaction terms. Specifically, the *brms* formula was: *Response (True) ~ Accented +*
 209 *Attractive + Intelligent + Interesting + Pleasant + Status + Trustworthy + Intelligibility +*
 210 *Knowledge + (1 | Participant) + (1 | Talker) + (1 | Sentence).*

Priors for all population-level effects were weakly informative, regularizing, Student's t distributions ($\nu = 3, \mu = 0, \sigma = 2.5$), following recommendations from Gelman (2020). The default weakly informative *brms* priors for the intercept and group-level parameters were used—Student's t ($\nu = 3, \mu = 0, \sigma = 2.5$) and half Student's t ($\nu = 3, \mu = 0, \sigma = 2.5$) distributions, respectively. The model was run with six chains. Each chain had 5,000 iterations (including 2,500 warm-up iterations), resulting in 15,000 total post-warm-up samples. Chains were well-mixed, and the effective sample size for all population-level effects was sufficiently large (for additional information about Bayesian statistical modeling, see Vasisshth, Nicenboim, Beckman, Li, & Kong, 2018).

The posterior mean estimate and 95% highest density credible intervals (henceforth, CrI) for the population-level parameters are summarized in Table 1. While there is a high probability that some of the social evaluations exert a nonzero effect on perceived truthfulness (e.g., $P(\beta_{\text{intelligent}} < 0 = 1)$), the mean effect size for each is extremely small (e.g., $\beta_{\text{intelligent}} = -0.16$), especially when compared to the much larger effect of Knowledge ($\beta_{\text{Knowledge}(\text{True})} = 2.56, \beta_{\text{Knowledge}(\text{False})} = -2.12$). The parameter estimates suggest that these perceptual evaluation measures and intelligibility do not substantially contribute to perceived truthfulness judgments.

One of the primary advantages of using Bayesian models is the estimation of credible intervals for all parameters, as opposed to only the main effects (here, the population-level effects). The group-level parameters in this model were Talker, Subject, and Sentence. While each of these accounted for some of the variation in the model, the group with the highest intercept standard deviation was Sentence—values are summarized in Table 2. Together with the population-level effect for Knowledge, this suggests that content and preexisting beliefs were larger drivers of perceived truthfulness. Additionally, the greater degree of variation across listeners compared to talkers suggests that who you are has a greater impact on your decisions related to perceived truthfulness than the voice on which you are making such evaluations.

Table 1

Mean estimates and 95% CrIs for the intercept and population-level parameters.

Parameter	Posterior Mean	95% CrI
Intercept	0.17	[-0.12, 0.46]
Knowledge (True)	2.56	[2.17, 2.97]
Knowledge (False)	-2.12	[-2.60, -1.68]
Intelligibility	-0.02	[-0.12, 0.07]
High/low social status	0.02	[-0.09, 0.13]
Strong/no foreign accent	0.07	[-0.05, 0.18]
(Un)attractive	0.00	[-0.11, 0.12]
(Un)intelligent	-0.16	[-0.27, -0.05]
(Un)interesting	0.07	[-0.05, 0.18]
(Un)pleasant	-0.02	[-0.14, 0.09]
(Un)trustworthy	0.03	[-0.08, 0.14]

Table 2

Group-level parameter standard deviation posterior mean estimates and 95% CrIs.

Parameter	Posterior Mean	95% CrI
Sentence intercept s.d.	1.09	[0.88, 1.33]
Subject intercept s.d.	0.28	[0.12, 0.43]
Talker intercept s.d.	0.09	[0.00, 0.25]

Discussion and Conclusion

The goal of this paper is to assess whether intelligibility and social evaluations predict truthfulness judgments in a set of ten Canadian English talkers. The intelligibility of each talker was objectively quantified as the proportion of keywords correctly identified in a speech-in-noise transcription task. Social evaluations were elicited along visual analogue scales for seven social dimensions. Ten talkers of a single variety were used, compared to the smaller samples in related work, which allows talker-specific differences to emerge. Additionally, the talkers were all natively-accented to mitigate the impacts of social stereotypes about non-native speech.

This study presents a null result—intelligibility and social dimensions did not impact whether a listener evaluated a sentence as true, especially when compared to the large effect of listener knowledge (i.e., believing a statement to be true or false). This meaningful null result is an important contribution to a mixed body of literature on this topic. By using Bayesian statistics, we can express confidence in zero as a parameter estimate, as well as confidence that our null result is both interesting and informative. Unlike frequentist analyses, this provides a more nuanced understanding of how variability is structured, as credible intervals are provided for group-level parameters (the “random effects”), in addition to the population-level parameters (the “fixed effects”). In this light, we have identified sentences and the listeners as greater sources of variability compared to talkers. While the disparity between the three group-level parameter estimates could be partially due to there being more sentences ($n = 88$) and listeners ($n = 82$) than talkers ($n = 10$), the magnitude of the differences suggests that these are dimensions worthy of attention. In sum, we present a null result in which any talker-related parameter—the talker, intelligibility, and social evaluations—exhibits a negligible impact on how the truthfulness of an obscure trivia statement is judged. Put simply, the content of the sentence and who the listener is matter far more than who is talking.

These null results echo earlier and contemporary findings (e.g., Miller & Hewgill, 1964; Sereno & Hawkins, 1967; Souza & Markman, 2013), but diverge from other work (e.g., Hanzlíková & Skarnitzl, 2017; Lev-Ari & Keysar, 2010). While we present evidence to the

contrary, we do not claim that social evaluations of a voice *never* have down-stream implications for credibility or intelligibility. Jiang, Gossack-Keenan, and Pell (2020), for example, demonstrate that confident speech styles elicit higher ratings of credulity than doubtful speech styles for familiar native accents, unfamiliar native accents, and non-native accents. They also find that familiar native accents (termed “in-group talkers”) are rated as more believable than non-native accents, supporting the results of Lev-Ari and Keysar (2010). Their interpretation for the lack of an accent effect in Souza and Markman (2013) is that confidence in a talker crucially impacts listener processing. In the current results, variability across sentences may be due (in part) to whether the talker produced a sentence in a confident or doubtful tone.

The proposed mechanism for why non-native speakers are judged to be less truthful-sounding is rooted in the variability introduced by non-native speech, and how it purportedly inhibits effective mapping to linguistic categories (Dragojevic & Giles, 2016; Lev-Ari & Keysar, 2010). It is possible that in this study—as all talkers were natively-accented and truth judgments were elicited in the clear—the talkers were all highly intelligible, and that a ceiling effect prevented any relationship between credulity and intelligibility to surface. The demonstration, however, that the talkers varied considerably in the measured dimensions, but that these had negligible impact on the truthfulness judgments warrants highlighting. The talkers elicited distinct evaluations from listeners and these did not impact credibility. Put simply, feelings only take you so far. This underscores the need to differentiate how social evaluations and intelligibility may or may not affect perception and decision-level processes in spoken language (Babel et al., 2019; Zheng & Samuel, 2017).

Acknowledgments

To be added post-review.

References

- Babel, M., & Russell, J. (2015). Expectations and speech intelligibility. *The Journal of the Acoustical Society of America*, 137(5), 2823–2833. doi: 10.1121/1.4919317
- Babel, M., Senior, B., & Bishop, S. (2019). Do social preferences matter in lexical retuning? *Laboratory Phonology*, 10(1). doi: 10.5334/labphon.133
- Baugh, J. (2000). Racial identification by speech. *American Speech*, 75(4), 362–364. doi: 10.1215/00031283-75-4-362
- Baugh, J. (2003). Linguistic profiling. In J. P. Hutchison (Ed.), *Black linguistics: Language, society, and politics in Africa and the Americas*.
- Bouchard Ryan, E., Carranza, M. A., & Moffie, R. W. (1977). Reactions toward varying degrees of accentedness in the speech of Spanish-English bilinguals. *Language and Speech*, 20(3), 267–273. doi: 10.1177/002383097702000308
- Bourhis, R. Y., & Giles, H. (1977). The language of intergroup distinctiveness. *Language, Ethnicity and Intergroup Relations*, 13, 119–135.
- Burkner, P.-C. (2017). brms: An R package for Bayesian multilevel models using stan. *Journal of Statistical Software*, 80(1), 1–28. doi: 10.18637/jss.v080.i01
- Craft, J. T., Wright, K. E., Weissler, R. E., & Queen, R. M. (2020). Language and discrimination: Generating meaning, perceiving identities, and discriminating outcomes. *Annual Review of Linguistics*, 6(1), 389–407. doi: 10.1146/annurev-linguistics-011718-011659
- Derwing, T. M., & Munro, M. J. (1997). Accent, intelligibility, and comprehensibility: Evidence from four L1s. *Studies in Second Language Acquisition*, 1–16. doi: 10.1017/S0272263197001010
- 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 Communication Research*, 42(3), 396–420. doi: 10.1111/hcre.12079
- Fuertes, J. N., Gottdiener, W. H., Martin, H., Gilbert, T. C., & Giles, H. (2012). A meta-analysis of the effects of speakers' accents on interpersonal evaluations. *European Journal of Social*

- Psychology, 42(1), 120–133. doi: 10.1002/ejsp.862
- Gelman, A. (2020). *Prior choice recommendations*. Retrieved 2021-05-21, from <https://github.com/stan-dev/stan/wiki/Prior-Choice-Recommendations>
- Giles, H., & Trudgill, P. (1983). Sociolinguistics and linguistic value judgements. In *On dialect*. Oxford: Basil Blackwell.
- Hanulíková, A., Van Alphen, P. M., Van Goch, M. M., & Weber, A. (2012). When one person's mistake is another's standard usage: The effect of foreign accent on syntactic processing. *Journal of Cognitive Neuroscience*, 24(4), 878–887. doi: 10.1162/jocn_a_00103
- Hanzlíková, D., & Skarnitzl, R. (2017). Credibility of native and non-native speakers of English revisited: Do non-native listeners feel the same? *Research in Language*, 15(3), 285–298. doi: 10.1515/rela-2017-0016
- Jiang, X., Gossack-Keenan, K., & Pell, M. D. (2020). To believe or not to believe? How voice and accent information in speech alter listener impressions of trust. *Quarterly Journal of Experimental Psychology*, 73(1), 55–79. doi: 10.1177/1747021819865833
- Johnson, K., & Mullennix, J. W. (1997). *Talker variability in speech processing*. Morgan Kaufmann Publishers Inc.
- Johnson, K., Strand, E. A., & D'Imperio, M. (1999). Auditory-visual integration of talker gender in vowel perception. *Journal of Phonetics*, 27(4), 359–384. doi: 10.1006/jpho.1999.0100
- Kang, O., & Rubin, D. L. (2009). Reverse linguistic stereotyping: Measuring the effect of listener expectations on speech evaluation. *Journal of Language and Social Psychology*, 28(4), 441–456. doi: 10.1177/0261927X09341950
- Kapnoula, E. C., & Samuel, A. G. (2019). Voices in the mental lexicon: Words carry indexical information that can affect access to their meaning. *Journal of Memory and Language*, 107, 111–127. doi: 10.1016/j.jml.2019.05.001
- Kinzler, K. D., & DeJesus, J. M. (2013). Northern = smart and Southern = nice: The development of accent attitudes in the United States. *Quarterly Journal of Experimental Psychology*, 66(6), 1146–1158. doi: 10.1080/17470218.2012.731695

- 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. doi: 10.1016/j.jesp.2010.05.025
- Lippi-Green, R. (2012). *English with an accent: Language, ideology and discrimination in the United States*. Routledge. doi: 10.4324/9780203348802
- Major, R. C., Fitzmaurice, S. M., Bunta, F., & Balasubramanian, C. (2005). Testing the effects of regional, ethnic, and international dialects of English on listening comprehension. *Language Learning*, 55(1), 37–69. doi: 10.1111/j.0023-8333.2005.00289.x
- Massaro, D. W., & Cohen, M. M. (1983). Evaluation and integration of visual and auditory information in speech perception. *Journal of Experimental Psychology: Human Perception and Performance*, 9(5), 753. doi: 10.1037/0096-1523.9.5.753
- McCloy, D. R., Wright, R. A., & Souza, P. E. (2015). Talker versus dialect effects on speech intelligibility: A symmetrical study. *Language and Speech*, 58(3), 371–386. doi: 10.1177/0023830914559234
- McGowan, K. B. (2015). Social expectation improves speech perception in noise. *Language and Speech*, 58(4), 502–521. doi: 10.1177/0023830914565191
- Miller, G. R., & Hewgill, M. A. (1964). The effect of variations in nonfluency on audience ratings of source credibility. *Quarterly Journal of Speech*, 50(1), 36–44. doi: 10.1080/00335636409382644
- Munro, M. J., & Derwing, T. M. (1999). Foreign accent, comprehensibility, and intelligibility in the speech of second language learners. *Language Learning*, 49, 285–310. doi: 10.1111/0023-8333.49.s1.8
- Norris, D. (1994). Shortlist: A connectionist model of continuous speech recognition. *Cognition*, 52(3), 189–234. doi: 10.1016/0010-0277(94)90043-4
- Pierrehumbert, J. B. (2016, jan). Phonological representation: Beyond abstract versus episodic. *Annual Review of Linguistics*, 2(1), 33–52. doi: 10.1146/annurev-linguistics-030514-125050

- Rickford, J. R., & King, S. (2016). Language and linguistics on trial: Hearing Rachel Jeantel (and other vernacular speakers) in the courtroom and beyond. *Language*, 92(4), 948–988. doi: 10.1353/lan.2016.0078
- Sereno, K. K., & Hawkins, G. J. (1967). The effects of variations in speakers' nonfluency upon audience ratings of attitude toward the speech topic and speakers' credibility. *Communications Monographs*, 34(1), 58–64. doi: 10.1080/03637756709375520
- Souza, A. L., & Markman, A. B. (2013). Foreign accent does not influence cognitive judgments. In *Proceedings of the annual meeting of the cognitive science society* (Vol. 35). Retrieved from <https://escholarship.org/uc/item/8hd9v4ff>
- Sumner, M., & Kataoka, R. (2013). Effects of phonetically-cued talker variation on semantic encoding. *The Journal of the Acoustical Society of America*, 134(6), EL485–EL491. doi: 10.1121/1.4826151
- Sumner, M., Kim, S. K., King, E., & McGowan, K. B. (2014). The socially weighted encoding of spoken words: A dual-route approach to speech perception. *Frontiers in Psychology*, 4, 1015. doi: 10.3389/fpsyg.2013.01015
- Van Berkum, J. J., Van den Brink, D., Tesink, C. M., Kos, M., & Hagoort, P. (2008). The neural integration of speaker and message. *Journal of Cognitive Neuroscience*, 20(4), 580–591. doi: 10.1162/jocn.2008.20054
- Vasishth, S., Nicenboim, B., Beckman, M. E., Li, F., & Kong, E. J. (2018, November). Bayesian data analysis in the phonetic sciences: A tutorial introduction. *Journal of Phonetics*, 71, 147–161. doi: 10.1016/j.wocn.2018.07.008
- Vaughn, C., & Kendall, T. (2018). Listener sensitivity to probabilistic conditioning of sociolinguistic variables: The case of (ing). *Journal of Memory and Language*, 103, 58–73. doi: 10.1016/j.jml.2018.07.006
- Zheng, Y., & Samuel, A. G. (2017). Does seeing an Asian face make speech sound more accented? *Attention, Perception, & Psychophysics*, 79(6), 1841–1859. doi: 10.3758/s13414-017-1329-2