Expectations about the source of a speaker's accent affect accent adaptation

Charlotte R. Vaughn

Citation: The Journal of the Acoustical Society of America **145**, 3218 (2019); doi: 10.1121/1.5108831
View online: https://doi.org/10.1121/1.5108831
View Table of Contents: https://asa.scitation.org/toc/jas/145/5
Published by the Acoustical Society of America

Expectations about the source of a speaker's accent affect accent adaptation

Charlotte R. Vaughn[a)]
Department of Linguistics, University of Oregon, 1290 University of Oregon, Eugene, Oregon 97403-1290, USA

(Received 28 December 2018; revised 28 April 2019; accepted 30 April 2019; published online 31 May 2019)

When encountering speakers whose accents differ from the listener's own, listeners initially show a processing cost, but that cost can be attenuated after short term exposure. The extent to which processing foreign accents (*L2-accents*) and within-language accents (*L1-accents*) is similar is still an open question. This study considers whether listeners' expectations about the source of a speaker's accent—whether the speaker is purported to be an L1 or an L2 speaker—affect intelligibility. Prior work has indirectly manipulated expectations about a speaker's accent through photographs, but the present study primes listeners with a description of the speaker's accent itself. In experiment 1, native English listeners transcribed Spanish-accented English sentences in noise under three different conditions (speaker's accent: monolingual L1 Latinx English, L1-Spanish/L2-English, no information given). Results indicate that, by the end of the experiment, listeners given some information about the accent outperformed listeners given no information, and listeners told the speaker was L1-accented outperformed listeners told to expect L2-accented speech. Findings are interpreted in terms of listeners' expectations about task difficulty, and a follow-up experiment (experiment 2) found that priming listeners to expect that their ability to understand L2-accented speech can improve does in fact improve intelligibility. © *2019 Acoustical Society of America*.
https://doi.org/10.1121/1.5108831

[TCB] Pages: 3218–3232

I. INTRODUCTION

When encountering speakers whose accent differs from the listener's own, listeners initially show a processing cost, but that cost can be attenuated after short term exposure (e.g., Bradlow and Bent, 2008; Floccia *et al.*, 2006). The extent to which processing a foreign accent (*L2-accent*) is similar to processing a within-language accent (*L1-accent*, e.g., a regional accent) is still an open question (Cristia *et al.*, 2012), but there is some evidence that different types of accents yield differences in listener performance (e.g., Adank *et al.*, 2009; Clarke and Garrett, 2004; Floccia *et al.*, 2006; Goslin *et al.*, 2012). One plausible source of the processing cost for accented speech[1] comes from the effort involved in resolving a potential mismatch between the incoming signal and a listener's existing exemplar category representations (e.g., Bradlow and Bent, 2008; Van Engen and Peele, 2014). Further, listeners' expectations about speakers have been known to affect speech perception in various ways (e.g., Magnuson and Nusbaum, 2007; McGowan, 2015; Niedzielski, 1999; Remez *et al.*, 1981; Rubin, 1992; Staum Casasanto, 2008; Strand, 1999). The mismatch between the signal and the activation of stereotype-based group-specific representations has been offered as one factor underlying those expectation effects (McGowan, 2015). However, prior work investigating the effect of expectations on accented speech processing has only examined listeners' expectations about speakers holistically, generated from photos or nationality labels. Whether listeners' processing of accented speech can be affected specifically by listeners' expectations about the *type of accent* a speaker has—due to being an L2 speaker, or variation in L1—is unknown. It may be that a listener's prior experiences processing L1-accented and L2-accented speech are relevant when entering into a new interaction with an accented speaker. In fact, work in social psychology has demonstrated that prior experiences with a task feed into expectations, which can affect processing, known as *processing fluency* (e.g., Lick and Johnson, 2015; Oppenheimer, 2008). Thus, listeners' previous experiences processing L1- and L2-accented speech may contribute to their expectations about a speaker, or the difficulty of the task, which may then affect processing.

This study examines the nature of the expectations listeners have about speakers, asking whether listeners' ability to transcribe an accented speaker's speech in noise is affected by their beliefs about the source of a speaker's accent. In a matched guise design, three between-subjects conditions (in experiment 1) presented different information about the speaker: describing the speaker as an L2 learner of English (*L2-accent* condition, which in fact matches the speaker's actual background), a monolingual Latinx English speaker (*L1-accent* condition, which provides a different rationale for the source of the speaker's accent), or not giving listeners any information about the source of his accent (*No guise*[2] condition). A follow up experiment (experiment 2), provides two modifications of the L2-accent condition, each priming listeners with a different factor that potentially contributes to their L2-related expectations. Taken together, the results provide insight into the mechanisms contributing to the processing of accented speech, and therefore the kinds of

[a)]Electronic mail: cvaughn@uoregon.edu

 0001-4966/2019/145(5)/3218/15/$30.00

interventions that may be most effective at reducing the adverse outcomes that accented speakers face (e.g., Kang and Rubin, 2009; Lindemann, 2002; Lippi-Green, 1997; Pantos and Perkins, 2013).

A. Processing L1- and L2-accented speech

Much previous research has been devoted to investigating how listeners process accented speech. In general, prior work has found that listeners appear to treat native and non-native speakers differently. For example, listeners have different expectations about non-native speakers' levels of syntactic competence as compared to native speakers, as indexed by event-related potentials (ERP) (Hanulikova *et al.*, 2012), and this pattern is conditioned on listeners' ability to identify the foreign accent of the speaker (Grey and Van Hell, 2017). Further, Lev-Ari (2015) argues that listeners rely more on top-down processing and less on the actual speech when listening to non-native speakers compared to native speakers. It should be noted that it is difficult to draw overall conclusions from the body of work comparing native and non-native accent processing, however, because the points of comparison are not always defined equivalently; sometimes L2-accent processing is compared to L1-accent processing on the whole with the implicit assumption that the L1 speaker in question shares an L1-accent with the listener, and other times L2-accent processing is compared explicitly to processing an L1-accent unfamiliar to the listeners. That said, studies examining L2-accented speech as compared to native (or implicitly "familiar" accent) speech (e.g., Hanulikova *et al.*, 2012; Lev-Ari, 2015; Van Engen and Peele, 2014), and studies examining L2-accented speech as compared to unfamiliar L1-accented speech (Adank *et al.*, 2009; Floccia *et al.*, 2006; Goslin *et al.*, 2012) tend to conclude that L1- and L2-accent processing are different in nature.

Thus, many studies report that listeners encounter greater processing difficulties for L2-accents than L1-accents. However, the bases for this effect are debated. Although it is tempting to imagine that the distance along a perceptual similarity continuum between a listener's own accent and a speaker's accent would completely explain how much difficulty the listener has, L1-accents are not always perceptually closer than L2-native accents to a native speaker's own accent (Cristia *et al.*, 2012). Indeed, work attempting to identify the reason for differences in processing between L1- and L2-accents has found evidence for distinct normalization mechanisms for the two types of accents, rather than solely gradient, perceptually based effects (Goslin *et al.*, 2012). Greater within-speaker acoustic variability for L2-accent than L1-accent speakers has also been presumed to factor into differences in processing (e.g., Floccia *et al.*, 2006; though see Vaughn *et al.*, 2018). In general, the extent to which the incoming accented acoustic signal matches or mismatches listeners' exemplar representations appears to play a major role in accent adaptation (e.g., Bradlow and Bent, 2008; Goslin *et al.*, 2012; Van Engen and Peele, 2014).

In addition to the signal, characteristics of listeners themselves contribute to processing both L1- and L2-accents. A range of listener characteristics has been shown to modulate a listener's adaptation to accented speech of various types, from cognitive properties like receptive vocabulary (e.g., Banks *et al.*, 2015; Bent *et al.*, 2016; Ingvalson *et al.*, 2017a) to listeners' affective properties like mood (e.g., Walker, 2016) to listeners' biases toward speakers and groups of speakers (e.g., Babel and Russell, 2015; Kang and Rubin, 2009; Ingvalson *et al.*, 2017b; Laturnus, 2018; McGowan, 2015; Yi *et al.*, 2013; though note that these results have been inconsistent across studies). The current study focuses on a different potential factor in accent adaptation: listeners' expectations about speakers' speech, specifically.

B. Expectations and accented speech processing

There is a long research tradition demonstrating that listeners' beliefs can affect speech perception and/or evaluation (e.g., Carmichael, 2016; Magnuson and Nusbaum, 2007; Niedzielski, 1999; Remez *et al.*, 1981; Staum Casasanto, 2008; Strand, 1999). For example, listeners who believed that they were listening to two different synthesized voices exhibited more of a processing slowdown in a word monitoring task than those who expected only one voice (Magnuson and Nusbaum, 2007). In terms of accented speech processing, listeners' expectations about whether they will encounter an accent change have been shown to affect the degree of processing slowdown that occurs (Floccia *et al.*, 2009). Additionally, several studies have demonstrated an expectation-based effect that Kang and Rubin (2009) call "reverse linguistic stereotyping." In this effect, activating a stereotype about the speaker (delivered, for example, through a picture purportedly of the speaker), triggers particular expectations about their speech, such as that the speech will be accented, which has impacts on the listeners' apprehension of the speaker and their speech (e.g., Babel and Russell, 2015; Gnevsheva, 2018; Hu and Su, 2015; Kang and Rubin, 2009; McGowan, 2015; Rubin, 1992; Zheng and Samuel, 2017). In a classic demonstration of this effect, Rubin (1992) played North American undergraduate students short lectures produced by a standard American-accented English speaker who was rated as more accented when the listeners were presented with a picture indicating the speaker was Asian than when presented with a Caucasian picture. Through a series of tasks tapping into a range of processing stages, Zheng and Samuel (2017) determined that the effect of faces on perceived accentedness operates post-perceptually; that is, stereotype-based expectations from faces affect how listeners make accentedness judgments not how listeners actually perceive the stimuli. Hu and Su (2015) found that Cantonese learners of English performed better on a range of comprehension tasks after listening to a native English speaker when told that the speaker was a native English speaker than when told the speaker was a learner of English. McGowan (2015) and Babel and Russell (2015) used transcription accuracy in noise as the dependent variable rather than accentedness and comprehension, finding differences in intelligibility based on the photo presented, providing precedents for the present study. Traditional interpretations of such expectation effects (e.g., Kang and

Rubin, 2009; Lippi-Green, 1997; Rubin, 1992) suggest that listeners' negative biases toward foreigners and their stereotype-based expectations that a speaker with a "foreign-looking" photo will be accented cause them to hallucinate an accent. In this account, listeners' biases cause them to bring less effort to the activity of listening, thereby reducing speech comprehension due to lack of attention to the acoustic signal, effectively upholding less of their end of the communicative burden (Lippi-Green, 1997; Rubin, 1992, in line with the account offered by Lev-Ari, 2015). More recently such results have also been explained by incongruence in listeners' stereotype-based expectations about the speaker and the incoming signal (Gnevsheva, 2018; McGowan, 2015). McGowan (2015) found that when listening to Chinese-accented English speech, native English speaking listeners shown a photo of an Asian American face transcribe the speech *better* than listeners shown a Caucasian American face, suggesting that prior results (e.g., Hu and Su, 2015; Rubin, 1992) may have been partially due to perceived incongruence between the acoustic signal and stereotype-based expectations about the face or guise. Under such an account, it is not that listeners put less attention toward processing speech they expect to be foreign accented, but rather that stereotype-based incongruence induces additional processing costs, taking up resources (see also Gnevsheva, 2018, for a discussion of these accounts).

With the exception of Hu and Su (2015), prior studies investigating listener expectations have presented either photos of faces (Rubin, 1992; Babel and Russell, 2015; McGowan, 2015; Staum Casasanto, 2008; Zheng and Samuel, 2017) or nationality labels (Niedzielski, 1999) to listeners as a way to manipulate listeners' beliefs about speakers. These methods make an assumption about the behavior of listeners (albeit a reasonable one given the prevalence of stereotype-based inferences, Fiske, 1998; and findings that perceivers have intuitions about matches between faces and voices, Campbell-Kibler and McCullough, 2015; Kamachi *et al.*, 2003), presuming that when presented with a photo of a Hispanic-looking person, for example, the listener generates an expected linguistic representation of that speaker based on the photo, such as that the person is a native speaker of Spanish whose English would therefore be Spanish-accented. This corresponds to a stereotype-based assumption about the cue level from the category level in the model of person construal by Freeman and Ambady (2011). In contrast, the present study provides a direct recommendation to the listener about *what kind of speech* to expect from the talker, priming expectations about the cue level in the model by Freeman and Ambady (2011) while holding the category level mostly constant. By explicitly orienting the listener to the type of *speech* to expect, namely L1 English or L2 English speech, the present study influences listeners' linguistic expectations more directly. This manipulation is possible by drawing on the fact that the speech of Latinx English speakers (L1 English speakers) and Spanish-accented English speakers (L2 English speakers) have been shown to overlap acoustically and perceptually (Tseng, 2015).

C. Latinx English and L1-Spanish-accented English

In parts of the United States with long-standing Spanish speaking populations, a dialect of English referred to here as *Latinx English*[3] has developed, which is influenced by contact with Spanish, but whose speakers may or may not be bilingual Spanish-English speakers (Fought, 2003, 2006; Godinez and Maddieson, 1985; Mendoza-Denton, 2008; Ornstein-Galicia, 1984; Santa Ana, 1993). That such speakers can and are often L1 English, or even nearly monolingual English speakers, is emphasized by Santa Ana (1993; here using the term Chicano English): "Chicano English is an ethnic dialect that children *acquire* as they acquire English…during their language acquisition period. Chicano English is to be distinguished from the English of second-language learners. The latter *learn* English after the so-called critical age of language acquisition. Thus defined, Chicano English is spoken only by native English speakers" (Santa Ana, 1993, p. 15, emphasis in original). That is, Latinx Englishes are dialects of native English speakers, whose initial formation is likely due to contact with Spanish and L1-Spanish/L2-English speakers (Fought, 2006). However, the maintenance of Latinx English within individuals and communities may be due to its importance for identity work, affording its speakers the linguistic repertoire to index their Latinx identity, or other identities, even if they are monolingual English speakers (Eckert, 2008; Fought, 2006; Mendoza-Denton, 2008). Although each specific variety of Latinx English is situated within its regional English variety, there are commonly described features of Latinx English that show connections to L1 Spanish (see Tseng, 2015, for a review), including properties of vowels (e.g., lack of reduction of unstressed high vowels, /ɪ/ raising, /ɑ/ fronting), and consonants (e.g., trilled /r/, stopping of fricatives, devoicing of final stops; Fought, 2003; Godinez and Maddieson, 1985; Mendoza-Denton, 2008; Santa Ana and Bayley, 2004). For example, Godinez and Maddieson (1985) found no significant vocalic differences between monolingual Latinx English and L1-Spanish/L2-English speakers in their sample. That said, the patterning of these features is not identical for Latinx English and L1-Spanish/L2-English speakers with certain L1-transfer properties such as categorical /ɪ/ raising or epenthetic /ə/ being more likely for non-native than native English speakers (e.g., Fought, 2003; Konopka and Pierrehumbert, 2008). However, of crucial importance for the purposes of this study, prior work has shown that many listeners cannot tell the L1 or L2 status of Latinx English speakers (Fought, 2003). And, Morales (2014) found nearly identical foreign accent ratings for Latinx English speakers and early bilingual Spanish-English speakers. Thus, given these results, it is expected that listeners in the present study could reasonably believe an L1-Spanish/L2-English speaker to be either L1-accented (a monolingual Latinx English speaker) or L2-accented (a L1 Spanish/L2 English speaker).

D. The present study

The present study gives different sets of listeners differing information about the source of a speaker's accent, and

examines whether that information affects how accurate listeners are at transcribing accented speech in noise. Since prior work has manipulated expectations in more indirect but perhaps more contextually rich ways (e.g., photographs), the first goal of this study (research question 1) is to test this accent source manipulation. In other words, can a written description of information about the source of the speaker's accent affect transcription accuracy at all? This first research question will be addressed by examining the performance of listeners given no information about the speaker's accent (*No guise*) compared to groups given background information (*L1-accent* and *L2-accent guises*). According to models like predictive coding (Clark, 2013), receiving any socially informative expectations may orient the listener in advance to the appropriate linguistic exemplar representations, minimizing prediction error, and thus facilitating initial transcription accuracy as well as perceptual adaptation over time (Sohoglu and Davis, 2016). Under this framework, the L1- and L2-accent groups, both given informative language models, should outperform the No guise group, given no language model.

The second research question examines the effect of presenting listeners with L1-accent vs L2-accent expectations about the speaker on transcription accuracy. Research question 2 considers: if listeners are thought to process native and non-native speech differently (e.g., Goslin *et al.*, 2012), does that difference arise solely from the linguistic signal itself, or might expectations about the type of speech one is hearing (while minimizing other inferences about a speaker) contribute to those differences? The design of the study thus attempts to isolate listeners' expectations about the source of a speaker's accent, removing potential confounds of interspeaker differences in the acoustic signal by using the same speaker, and removing the myriad inferences that photographs can trigger about a speaker beyond their native language status (e.g., D'Onofrio, 2015) by using a written description. Although presenting descriptions of a speaker's accent surely trigger some inferences about a speaker beyond just those about their accent, here, it is assumed that inferences generated in response to a description of a speaker's accent are at least more constrained and less varied than those generated in response to a photo.

If the L1- and L2-accent guise conditions do show differences in transcription accuracy, previous work suggests that listeners' prior associations with listening to native and non-native speech may guide their task performance. For example, listeners may have negative past experiences listening to L2-accented speakers (e.g., Lindemann, 2002; Pantos and Perkins, 2013), which may lead them to expect that foreign accents are difficult to process. Work in social psychology has found that expectations about task difficulty can indeed lead to differences in processing. The construct of *processing fluency*, or the "difference between expected difficulty and actual difficulty" (Oppenheimer, 2008, p. 238), proposes that tasks that are hard or harder than expected are more disfluent for listeners, and tasks that are easy or easier than expected are more fluent. If this account is at work, it predicts for research question 2 that listeners given the L2-accent guise should have lower accuracy than listeners primed with an L1-accent guise because L2-accent guise listeners enter the task primed with the expectation that the talker is non-native and thus harder to process.

Further, prior work highlights perceivers' different *naive theories* about what makes their processing more or less disfluent (Dweck, 1999; Miele and Molden, 2010). A listener may attribute their subjective experience of processing disfluency to a range of factors, from properties of the stimulus or the task to their own mental states. And, these different naive theories (that is, explanations about why a task feels difficult) can have differential impacts on task behavior (Miele and Molden, 2010). Thus, experiment 2 investigates two different naive theories that participants may have about the processing disfluency involved in listening to L2-accented speech, one based on properties of the L2-accented *speaker* and another based on the *task* of listening to L2-accented speech.

Finally, both experiments address each of the two main research questions in terms of the time course of accent adaptation. Prior findings are mixed regarding the extent to which listeners recover from initial costs when listening to accented speech, ranging from lack of clear improvement to significant gains to complete adaptation. That is, some prior work has found that listeners improve in terms of intelligibility or comprehensibility of accented speech over time (e.g., Clarke and Garrett, 2004; Bradlow and Bent, 2008; Xie *et al.*, 2018), while in other studies short term exposure has not resulted in consistent improvements (e.g., Adank and McQueen, 2007; Floccia *et al.*, 2009; Smith *et al.*, 2014). Analyses in this study incorporate the amount of exposure the listener has had to the speaker (measured by quartiles of experimental trials, following Clarke and Garrett, 2004; Bradlow and Bent, 2008; and others) in order to measure potential adaptation to the speech.[4] Examining time points at which any potential differences between guise conditions may emerge is informative: differences emerging early as opposed to later in the experiment point to distinct implications about the effects of expectations on speech processing.

II. EXPERIMENT 1

A. Materials and methodology

1. Acoustic stimuli

Forty semantically anomalous, syntactically legal sentences (from Liss *et al.*, 2009), e.g., "Secure but lease apart," "Hold a page of fortune," "Younger rusty viewers," were recorded. Semantically anomalous sentences were chosen in an effort to minimize the amount of contextual or top-down information listeners could use, especially given that studies have suggested that listeners use more top-down processing when listening to L2-accented than native speech (e.g., Lev-Ari, 2015).

The speaker was a 34-year-old male L1-Spanish/L2-English speaker, originally from Chile but living in the United States (US) for graduate school, who identified as Hispanic. He reported learning English informally as child, and taking six months of English language courses as an adult. A small group of trained phoneticians familiar with

both Spanish-accented English and Latinx English judged that this speaker was believable as either type of speaker.

Stimuli were recorded in a sound-attenuated booth using a SM93 microphone (Shure, Inc., Niles, IL) and a PMD-661 recorder (Marantz America, Inc., Mahwah, NJ). Sentences were amplitude normalized to 65 dB sound pressure level (SPL) and were mixed with speech-shaped noise at +5 signal-to-noise ratio (SNR), bounded by a leader and tail of 500 ms of noise. The SNR of +5 was chosen based on pilot testing of several ratios to select a SNR that had the desired effect of reducing overall accuracy while still preserving listeners' ability to perceive indexical characteristics of the speaker's voice.

2. Procedure

Participants in this study came from two populations, Amazon's Mechanical Turk (Amazon.com, Inc., Seattle, WA) workers participating online, and undergraduate students participating in the laboratory. Thus, this study adds to the growing body of literature conducting accent adaptation experiments online (Burchill *et al.*, 2018; Liu and Jaeger, 2018; Xie *et al.*, 2018); however, this study has the additional benefit of directly comparing performance of listeners participating on Mechanical Turk vs those in the laboratory. The experiment was conducted in a custom web browser–based presentation software, so all participants used the identical interface. Laboratory participants were seated at an individual computer workstation with HD-202 headphones (Sennheiser electronic GmbH & Co. KG, Wedemark, Germany), and Mechanical Turk participants were asked to use their own headphones.

Before the experimental phase, several important instructions were given. The primary intended manipulation of the different guise conditions was to generate different expectations about the native language of the speaker, and thus the source of his accent. A small amount of additional information about the speaker beyond his native language was provided to listeners (see Table I) for two reasons. First, it was hoped that providing a short narrative about the speaker would make it seem more natural to participants that any information was being given at all (thereby drawing less attention to the source of the accent possibly being the purpose of the experiment). Second, it was hoped that the additional information, congruent with and explanatory of each language situation, would increase the likelihood that participants retained and believed the accent information (see Sec. IV for more on this point).

Given that the guise condition manipulation is crucial to interpreting the results of this study, it was important to ensure that participants had, in fact, paid attention to the guises in order to draw valid conclusions about the guises' effects. To this end, participants were asked a question about the language background of the speaker (called "Daniel") in order to ensure that they retained that critical information. On the screen immediately following the information about Daniel displayed in Table I, participants were asked "What do you know about Daniel's knowledge of Spanish?" and were given four possible response options: (1) Daniel speaks Spanish, (2) Daniel only speaks English, and no Spanish, (3) Unknown—I was not told whether or not he speaks Spanish, and (4) I do not remember. Participants were encouraged not to guess, but rather to answer "I do not remember" if they did not know the answer. Participants' data were excluded if they did not answer in accordance with their assigned guise (see Sec. II A 3).

After the instructions, the 40 semantically anomalous sentences mixed with noise were presented in random order to each participant. On the screen during every trial were the instructions "Type what you hear. If you do not understand what Daniel is saying, please make your best guess." Intermixed with these 40 sentences were several attention checks (e.g., true/false responses to auditory prompts about basic facts, like "A quarter is worth 25 cents"). After the transcription section, participants answered a series of questions about their experience of the task and then filled out a language background questionnaire.

3. Participants

A total of 230 participants took part in experiment 1 (94 participants in the laboratory, and 136 Mechanical Turk workers), randomly assigned to the 3 guise conditions. Mechanical Turk participants, restricted to IP addresses in the U.S., were paid the equivalent of $10/h for their participation. Undergraduate participants were recruited from the University of Oregon Psychology and Linguistics Human Subjects Pool, and were given partial course credit for their time.

Eighty-six participants were removed for not responding in accordance with their guise on the question regarding Daniel's knowledge of Spanish or for failing attention checks (2 in laboratory, 84 Mechanical Turk participants removed). The large difference in exclusion rate across participant types (removal of 2.1% in laboratory but 61.8% of Mechanical Turk participants) indicates that studies involving careful attention to detailed instructions may not be particularly well-suited for a Mechanical Turk setting. Three additional participants were removed due to reporting uncorrected hearing loss (one in laboratory, two Mechanical Turk). One in laboratory participant was removed because they responded to each trial with a single keypress. Six of

TABLE I. Text of the guise descriptions given to participants in each of the three conditions in experiment 1.

No guise	"Today you will be listening to speech from Daniel."
L1-accent guise	"Meet Daniel! Daniel grew up in Southern California. His parents are from Chile and they speak Spanish and English. But, Daniel only speaks English. He never learned Spanish. Daniel's first language and only language is English."
L2-accent guise	"Meet Daniel! Daniel grew up in Southern California. His parents are from Chile, and he speaks Spanish at home. Daniel's first language was Spanish, and he only began learning English when he started school."

the remaining participants were removed because they were not native English speakers (four in laboratory, two Mechanical Turk).

Following Xie *et al.* (2018), participants' reaction times (RTs) were assessed to ensure that those who were likely multitasking or otherwise clearly distracted during the task were not included. A script was run that was designed to remove participants whose mean log RT was greater than three standard deviations (SDs) outside of their guise group's mean log RT, but no participants met that criterion so no additional participants were removed. In total, 134 native English speaking participants remained, distributed across conditions according to Table II. Of these participants, 76 identified as female, 53 as male, and 5 declined to report their gender. Fifteen participants identified as Hispanic or Latinx, 113 identified as not Hispanic or Latinx, and 6 declined to report their ethnicity. Participants fell into the following age groups: 18–24 years, 89; 25–32, 17; 33–45, 18; 46–65, 8; over 65, 2. In the language background questionnaire, all participants were asked about whether they spoke Spanish (yes, 26; no, 108); their exposure to L1-Spanish, L2-English speech ("How much experience do you have listening to people who speak Spanish-accented English?," a lot, 36; some, 88; none, 10), their exposure to Latinx English ("Have you ever met anyone who sounds like they have a Spanish accent, even if they do not know Spanish?," yes, 42; "No. I believe that is possible but do not personally know anyone fitting that description.," 81; "No. I do not believe that is possible.," 11).[5] None of these factors significantly improved any statistical models in either experiment, so they are not considered further.

B. Analysis and results

5360 data points were collected (40 sentences × 134 participants). Following Xie *et al.* (2018), individual trials (sentences) whose log reaction times were outside 3 SDs from the participant's mean log reaction time were discarded, removing 111 total trials (2.1% of the overall data; 76 trials or 2.2% of in laboratory participants' data, and 35 trials or 1.8% of Mechanical Turk participants' data), and leaving a total of 5289 trials for analysis.

To assess performance on the speech in noise task, transcription accuracy was assessed following the procedure of Babel and Russell (2015). First, responses were converted to all lowercase and punctuation was removed. Spelling errors were corrected sparingly. For example, *transend* was corrected to *transcend*, *persue* to *pursue*, and *percieve* to *perceive*. Then, a word-matching script calculated the percentage of words correct for each sentence, which took into account all words rather than only key words. All homophones were counted as correct spellings (e.g., *counsel*, *council*). Percent correct was converted to rationalized arcsine units (RAUs) after Studebaker (1985), following previous work (e.g., Babel and Russell, 2015; Bradlow and Bent, 2008). The RAU transformation converts proportional data into a normally distributed scale, which ranges from −23 to 123. Descriptive statistics and figures present accuracy in percent correct for maximum interpretability, but statistical analyses used RAU as the dependent measure. (As some work has questioned the wisdom of the RAU transform, e.g., Warton and Hui, 2011, all statistical models were also run using percent correct as the dependent variable. Results were substantively similar, and only models in RAU are presented.)

TABLE II. Final number of participants in each guise condition in experiment 1 after exclusions.

	No guise	L1-accent guise	L2-accent guise	Total
In laboratory	25	32	29	86
Mechanical Turk	18	13	17	48
Total	43	45	48	134

The overall percent correct across conditions was 58.1% (SD = 31.5%). Accuracy was lowest for the No guise condition ($M = 56.1\%$, SD = 31.8%), highest for the L1-accent guise condition ($M = 60.8\%$, SD = 30.9%), with the L2-accent guise in the middle ($M = 57.5\%$, SD = 31.7%). Figure 1 visualizes percent correct by guise condition across trial quartiles.

In order to examine the effect of guise on transcription accuracy, mixed effects linear regression modeling was conducted, using per-sentence RAU as the dependent variable. Model building followed a step-up procedure, using likelihood ratio testing to test each factor one at a time. Factors significantly improving the model were left in, building up to the most complex model. Participant type (in laboratory or Mechanical Turk) was included as a fixed effect. The model included sum contrast-coded fixed effects of guise condition (No guise, L1-accent guise, and L2-accent guise) in order to test the two research questions of interest (research question 1, No guise vs guise; research question 2, L1-accent vs L2-accent; coded as listed in Table III). Trial quartile was included as a fixed effect to model the amount of adaptation throughout the experiment [first quartile (trials 1–10), second (11–20), third (21–30), and fourth (31–40); dummy coded with the final quartile (31–40) as reference level]. All two-way interactions between fixed effects were also examined. The random effects included the maximal structure that improved the model by likelihood ratio testing, and included random intercepts for participant and sentence, and random slopes for participant type and guise condition by sentence. The best model (reported in Table III) included main effects of participant type, guise condition, and trial quartile, and included an interaction between guise condition and trial quartile.

Results confirmed significant differences in transcription accuracy according to guise condition. With respect to research question 1, by the end of the experiment (i.e., in the final quartile), the groups receiving information about the speaker (L1- and L2-accent conditions) transcribed the accented speech more accurately than the group that did not receive any information [$\beta = 6.544$, standard error (SE) = 1.700, $p < 0.001$]. Not only did the information about the speaker have an effect, but the effect was a beneficial one. With respect to research question 2, in the final quartile transcription accuracy was higher for the group receiving the L1-accent guise than the group receiving the L2-accent guise ($\beta = -4.985$, SE = 1.953,

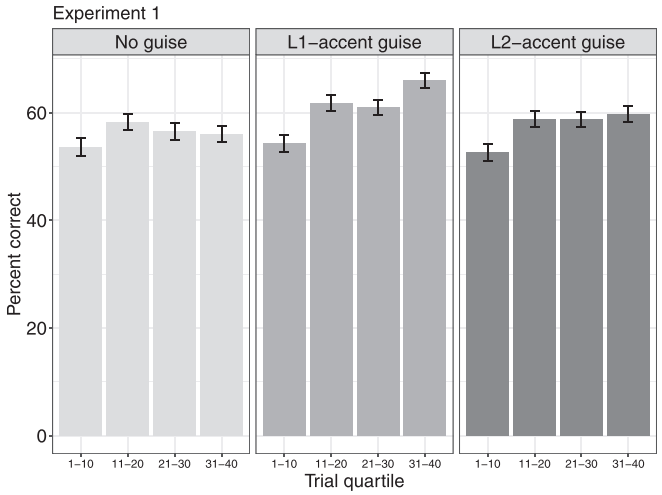

FIG. 1. Experiment 1 results by guise condition in percent correct across trial quartiles. Error bars represent ±1 standard error (SE).

$p = 0.011$). Listeners who thought the speaker was a native speaker transcribed speech more accurately than listeners who thought the speaker was non-native.

In terms of adaptation, results indicated a main effect of quartile, indicating that listeners generally improved in their ability to transcribe accented speech in noise over time; across guises, there were significant differences between the first and fourth quartiles ($\beta = -6.420$, SE = 0.804, $p < 0.001$), and between the third and fourth quartiles ($\beta = -1.611$, SE = 0.801, $p = 0.044$). That said, inspecting Fig. 1 it is likely that the significant main effect of quartile was driven by listeners in the L1- and L2-accent guise conditions, as there does not appear to be significant adaptation in the No guise condition. Relatedly, as is evident in Fig. 1, the model returned several significant interactions between guise condition and quartile, indicating that the magnitude of adaptation differed as a function of guise assignment. This result provides guidance as to where in the experiment the effect of guise was most pronounced. For research question 1: the difference between No guise and L1-/L2-accent guise conditions was significant when comparing the first quartile to the final quartile ($\beta = -6.028$, SE = 1.727, $p < 0.001$), the second to the final quartile ($\beta = -3.742$, SE = 1.701, $p = 0.029$), and the third to the final quartile ($\beta = -3.649$, SE = 1.720, $p = 0.034$). This pattern suggests that the superior performance of guise conditions compared to the No guise condition was rather consistent throughout. For research question 2, L1- and L2-accent guises were not significantly different between any quartile comparisons, but Fig. 1 indicates that most of the gains from the L1-accent guise compared to L2-accent occurred between the third and fourth quartiles, which is hinted at by the trend toward significance in Table III ($\beta = 3.447$, SE = 1.937, $p = 0.075$).

Finally, there was a significant main effect of participant type ($\beta = 4.075$, SE = 1.513, $p = 0.008$), indicating that Mechanical Turk participants transcribed words more accurately than in laboratory participants.

C. Discussion

The results of experiment 1 demonstrated that listeners receiving some information about the source of the speaker's accent (the L1- and L2-accent guise conditions) transcribed the speaker's Spanish-accented speech in noise more accurately than listeners receiving no information about the speaker's accent (No guise condition). Further, most of those gains were present from relatively early on in the experiment, as the No guise condition showed generally flat performance across quartiles (in line with prior work discussed above failing to find robust accent adaptation). Further, listeners who were told that the speaker was a native Latinx English speaker (L1-accent) transcribed more words correctly than listeners who were told the speaker was a non-native English speaker (L2-accent).

TABLE III. Statistical model results of fixed effects for experiment 1.

Coding scheme (contrast code matrix):
No guise vs L1 and L2: No guise = -1, L1-accent guise = 0.5, L2-accent guise = 0.5
L1 vs L2: No guise = 0, L1-accent guise = -1, L2-accent guise = 1

Model formula: RAU ~ participant type + guise × trial quartile 4th + (1 | participant) + (guise | sentence)

Fixed effect	Estimate	Standard error	*t* value	*p* value
(Intercept)	58.6604	3.5760	16.404	<2e-16
Participant type	4.0748	1.5125	2.694	0.008130
No guise vs L1- and L2-accent	6.5438	1.6997	3.850	0.000142
L1-accent v. L2-accent	−4.9850	1.9533	2.552	0.011169
Trial quartile 4th v. 1st	−6.4200	0.8042	−7.983	1.75e-15
Trial quartile 4th vs 2nd	−0.9982	0.8002	−1.247	0.212312
Trial quartile 4th vs 3rd	−1.6107	0.8008	−2.011	0.044345
No guise vs L1 and L2: 4th vs 1st	−6.0276	1.7269	−3.490	0.000486
L1 vs L2: 4th vs 1st	1.5051	1.9486	−0.772	0.439921
No guise vs L1 and L2: 4th vs 2nd	−3.7416	1.7099	−2.188	0.028696
L1 vs L2: 4th vs 2nd	1.0507	1.9416	−0.541	0.588432
No guise vs L1 and L2: 4th vs 3rd	−3.6492	1.7204	−2.121	0.033955
L1 vs L2: 4th vs 3rd	3.4470	1.9371	−1.779	0.075218

The first pattern of results (research question 1) demonstrates that adaptation to an accented speaker may indeed be aided by having prior information about the speaker, as predicted by predictive coding accounts (Clark, 2013; Sohoglu and Davis, 2016), and by accounts suggesting that congruent expectations about speakers can be facilitative (McGowan, 2015). Furthermore, the present results isolate the potential benefit of having a certain type of prior information, information about the source of a speaker's accent. How could such information help listeners? Predictive coding suggests that prior information about the talker could serve to minimize prediction error, and thus facilitate initial transcription accuracy as well as perceptual adaptation over time (Sohoglu and Davis, 2016). The present results are in line with prior findings in this domain (e.g., as reviewed in Koelewijn *et al.*, 2015). For example, prior knowledge about who is going to speak has been shown to enhance speech intelligibility in multi-talker contexts by way of focusing listeners' attention (e.g., Brungart *et al.*, 2001; Kitterick *et al.*, 2010). As for the present results, it may be that giving any kind of elaborated contextual information about a speaker put listeners in a processing mode where they approached the signal using deeper processing than those with no contextual information (Craik and Tulving, 1975). Or, it may be that, even though the L1- and L2-accent guises gave different information about how the speaker's accent related to Spanish, both guises pre-activated for listeners the general concept "Spanish," which was enough to guide those listeners to find the appropriate exemplar space. That said, listeners did benefit more from L1- than L2-accent guises, discussed in detail below.

Further, listeners in the No guise condition may have expended processing resources attempting to determine what type of accent the speaker had, leaving fewer overall processing resources available for adapting to the accent. Atagi and Bent (2015) documented a relationship between listeners' ability to categorize the type of non-native accent a speaker has (e.g., Japanese- vs French-accented English) and their ability to recognize non-native speech in noise. They suggest that listeners who have lexical representations that are strongly tied to representations of accents have more cognitive resources available for linguistic processing. Thus, in the present study, listeners given any believable information about the type of accent the speaker had (giving them a strong representation of the speaker's accent) may have had an easier time locating an appropriate exemplar set, and thus their transcription accuracy may be facilitated as compared to listeners not given such information. This explanation also accords with the time course of this effect first appearing rather early: having information about the speaker gives listeners an initial foothold, and also allows for increased learning to occur (Sohoglu and Davis, 2016).

The second pattern of results (research question 2) showed that listeners who expected L1-accented speech had higher transcription accuracy than another group of listeners expecting the same speech to be from an L2 speaker. This finding is consistent with Hu and Su (2015), who showed that English learners had higher scores on comprehension tests in an L1-guise than L2-guise condition when listening to a native English speaker. A possible implication of such findings is that listeners' past experiences listening to L2-accented speakers lead them to have the expectation that foreign accents are difficult to process (e.g., Lindemann, 2002; Pantos and Perkins, 2013), and thus listeners primed with the expectation that the talker is non-native may have entered the task primed with that expectation. This explanation aligns with work in social psychology on processing fluency (e.g., Lick and Johnson, 2015; Oppenheimer, 2008), though the precise mechanism of this and similar effects (e.g., stereotype threat, Pennington *et al.*, 2016; Schmader *et al.*, 2004; wishful seeing, Dunning and Balcetis, 2013; mindset, or implicit theory, Dweck, 1999) remains unsettled in the literature. Worse performance by participants primed to expect that a task will be difficult may be due to, for example, increased anxiety, which depletes processes of executive functioning like attention and working memory (Schmader *et al.*, 2004), changes in the processing strategy used (Jamieson and Harkins, 2011), lack of filtering of irrelevant or unlikely interpretations of stimuli (Dunning and Balcetis, 2013), or less error monitoring and subsequent error correction (Moser *et al.*, 2011). In general, most of these mechanisms would posit that listeners entering the task primed to use an L2-mode of processing would have used their resources less efficiently and therefore transcribed fewer words accurately. Conversely, then, listeners in the L1-accent condition would not be discouraged by expectations that the task would be difficult, freeing up their resources to continue to adapt to the accent, explaining the putative later time course of adaptation in these results.

Finally, results showed that Mechanical Turk workers performed better on the task than participants in the laboratory. That the addition of an interaction between participant type and group did not improve the model, however, suggests that guise assignment did not differentially affect the two types of participants. It is not possible to know for certain the source of the difference between the two types of participants, but it may be that Mechanical Turk workers have more experience with this type of speech transcription task. Or, it may be that middle-aged adults, who are more represented in the Mechanical Turk than in laboratory participants (62.5% Mechanical Turk vs 5.8% of in laboratory participants between 25 and 45 year old) may be better than younger adults, who are the majority of the in laboratory participants (16.6% Mechanical Turk vs 94.2% of in laboratory participants between 18 and 24 years old), in perception of non-native speech in noise, due to factors like vocabulary size or greater exposure to speech variability. Future work investigating accented speech processing in noise across these age groups could test this hypothesis. However, since a much higher percentage of Mechanical Turk participants had to be excluded for failing crucial attention checks, concluding that Mechanical Turk participants are more desirable participants in this type of task is not advisable.

On the whole, there is evidence that expectations about the source of a speaker's accent matter to listeners, and that listeners' prior experiences may help explain these results. Listeners appeared to benefit from information about the language background of the speaker in this study,

particularly when that information suggested that the listeners would hear a native speaker (which, of course, was not the actual background of the speaker; this potential incongruity is explored in more detail in Sec. IV). These results raise questions about precisely what aspects of listeners' expectations make their processing more or less disfluent, that is, the *naive theories* these listeners may have about the experience of listening to accented speech. As discussed above, it may be that listeners in the L2-accent group experience processing disfluency that influences accuracy because the L2-accent guise calls to mind listeners' models of past accented speakers, in particular, and thus calibrates expectations about the *speaker* accordingly. Or, alternatively, past experiences may be influential because the L2-accent guise information calls to mind entire prior communicative interactions with non-native speakers, and listeners' experiences of the act of listening to L2-accented speech, and thus calibrates expectations about the *task* accordingly.

One way to begin to understand the aspects of listeners' prior experience that may be important—or, which naive theories listeners may hold—is to attempt to make the experiences of listeners expecting L2-accented speech *less* disfluent in different ways, to see whether those modifications change performance as compared to listeners given the L2-accent guise from experiment 1. As a first pass at this question, experiment 2 uses two additional instructional manipulations that emphasize two different factors that may make listeners' processing of L2-accented speech easier than they expect.

III. EXPERIMENT 2

Experiment 2 presented two new groups of listeners with new guises, both elaborations of the L2-accent guise from experiment 1. In experiment 1, since listeners in the L2-accent guise had poorer performance as compared to the L1-accent guise, one possible inference is that listeners have had prior experience suggesting that L2-accented speech processing will be hard, and thus listeners in the L2-accent guise approach task with that expectation. Experiment 2 aims to determine whether ameliorating the expected difficulty (processing disfluency) for the L2-accent group via additional instructions will affect performance as compared with the L2-accent condition from experiment 1. In experiment 2, one new group of listeners was given a guise designed to decrease perceived disfluency attributed to the accented *speaker* (by framing the speaker's L2 accent more positively), and another new group was given a guise designed to decrease perceived disfluency attributed to the listening *task* (by framing listeners' experience of the task of listening to L2 speech, and their ability to succeed, more positively). Of course other potential naive theories may exist, and other potential factors may play a role, such as listeners' implicit biases and listeners' own speech patterns, but the two factors investigated here directly follow up on the L2-accent guise from experiment 1.

A. Materials and methodology

1. Acoustic stimuli

Stimuli in experiment 2 were identical to those used in experiment 1.

2. Procedure

The procedure was identical to that of experiment 1, but with different guise instructions. Table IV shows the text for the two guises used in experiment 2, which contain (in italics in Table IV) the entirety of the text from the L2-accent guise in experiment 1 (shown in Table I), plus the additional sentences unique to the new guises shown (in bold in Table IV). The guise referred to as *L2-accent guise:Speaker* emphasizes that Daniel's L2 accent is reportedly easy to understand, and the guise called *L2-accent guise:Task* emphasizes that listeners have been shown to improve their understanding of accented speech through tasks such as this one. Participants in these two conditions will be compared against the L2-accent guise participants from experiment 1 in analyses.

3. Participants

Only in laboratory participants were used in experiment 2. The participants were all undergraduates recruited from the University of Oregon Psychology and Linguistics Human Subjects Pool who were given partial course credit for their time. Thus, only the subset of participants in the L2-accent condition from experiment 1 who participated in the laboratory ($N = 29$) will be used as a baseline for comparison in experiment 2.

A total of 48 new participants took part in experiment 2 and were randomly assigned to receive one of the two new L2-accent guises. Five participants were removed for not responding in accordance with their guise on the question regarding Daniel's knowledge of Spanish or for failing attention checks. Reaction times were checked for outliers following the procedure described for experiment 1, but again no participants were removed. In total, data from 41 new

TABLE IV. Text of the guise descriptions given to participants in each of the new conditions in experiment 2. The portion of guise text in *italics* in this table was also present in the L2-accent guise from experiment 1. The portion of guise text in **bold** was specific to each respective new guise in experiment 2.

L2-accent guise:Speaker	*Meet Daniel! Daniel grew up in Southern California. His parents are from Chile, and he speaks Spanish at home. Daniel's first language was Spanish, and he only began learning English when he started school.* **He became highly proficient in English. Now, Daniel's English is very good, and his English-speaking friends say he is very easy to understand.**
L2-accent guise:Task	*Meet Daniel! Daniel grew up in Southern California. His parents are from Chile, and he speaks Spanish at home. Daniel's first language was Spanish, and he only began learning English when he started school.* **Scientific studies have shown that practice listening to accented speech helps listeners improve their comprehension of accented speakers. By the time you're done with this task, you will be much better at understanding Daniel than you were when you started!**

TABLE V. Number of participants in each guise condition in experiment 2, including in laboratory participants from the L2-accent guise in experiment 1.

	L2-accent guise:Exp 1	L2-accent guise:Speaker	L2-accent guise:Task	Total
In laboratory only	29	21	20	70

participants plus 29 participants from experiment 1 were used in the following analyses, distributed across conditions according to Table V. Of the 41 new participants, 20 identified as female and 21 as male. Two participants identified as Hispanic or Latinx and 39 identified as not Hispanic or Latinx. Participants fell into the following age groups: 18–24 years; 40; 25–32, 1. Participants responded as follows regarding whether they spoke Spanish (yes, 7; no, 34), their experience listening to L1-Spanish, L2-English speech (a lot, 7; some, 32; none, 2), their exposure to Latinx English ("Have you ever met anyone who sounds like they have a Spanish accent, even if they do not know Spanish?," yes, 12; "No. I believe that is possible but do not personally know anyone fitting that description.," 26; "No. I do not believe that is possible.," 3). Again, none of these measures significantly improved any models, so are not considered further.

B. Analysis and results

1640 new data points were collected (40 sentences × 41 new participants). Again, individual trials (sentences) whose log reaction times were outside 3 SDs from the participant's mean were discarded, removing 15 total trials (0.92% of the data), and leaving 1625 trials for analysis. Transcription accuracy was assessed using the same procedure as in experiment 1.

The overall percent correct across the three conditions in experiment 2 was 58.8% (SD = 31.1%). Accuracy was lowest for the L2-accent guise condition from experiment 1 ($M = 56.2\%$, SD = 32.0%), highest for the L2-accent guise:Task condition ($M = 61.3\%$, SD = 29.8%) with the L2-accent guise:Speaker condition intermediate ($M = 59.9\%$, SD = 31.0%). Figure 2 visualizes percent correct by guise condition across trial quartiles.

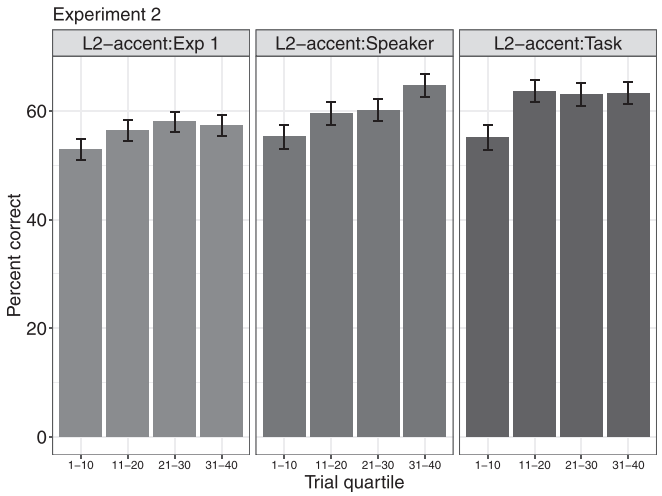

FIG. 2. Experiment 2 results by guise condition in percent correct across trial quartiles. Error bars represent ±1 SE.

Mixed effects linear regression modeling was conducted as described in experiment 1 with per-sentence RAU as the dependent variable. The model included a fixed effect of guise condition (L2-accent guise:Experiment1, L2-accent guise:Speaker, and L2-accent guise:Task), dummy coded, with L2-accent guise:Experiment1 as the reference level in order to compare each of the two new guises to the baseline. Trial quartile was included as a fixed effect (first, second, third, and fourth; dummy coded with fourth quartile again as the reference level). All two-way interactions between fixed effects were examined. Similar random effects as in experiment 1's best model were used, which included random intercepts of participant and item, and a random slope for guise condition by item. The best model (reported in Table VI) included main effects of guise condition and trial quartile (with no interaction between guise condition and trial quartile, unlike experiment 1).

The significant main effect of guise indicates that transcription performance was significantly higher for the L2-accent guise:Task condition than the L2-accent:Exp1 condition ($\beta = 4.635$, SE = 1.910, $p = 0.018$). Although transcription performance was numerically higher for the L2-accent guise:Speaker group than the L2-accent guise group from experiment 1, this comparison did not reach significance ($\beta = 3.407$, SE = 1.806, $p = 0.063$). In terms of adaptation, the only significant difference between the final quartile and another quartile was between the first and fourth quartiles ($\beta = -6.269$, SE = 1.068, $p < 0.001$).

Given listeners' good performance in the experiment 2 guises, one may wonder whether these new guises improved performance as much as being told the listener is an L1 speaker did in experiment 1. To explore this question, a new model was fit, including all five conditions across both experiments (still in laboratory participants only), using the same model structure as in experiment 1 and setting the L1-guise from experiment 1 as the reference level. The model finds no significant differences between the L1-accent condition from experiment 1 and either the L2-accent:Task condition ($\beta = 0.759$, SE = 2.265, $p = 0.738$), or the L2-accent:Speaker condition ($\beta = 0.104$, SE = 2.234, $p = 0.963$). (Similar results were obtained for these comparisons when using all participants, not just in laboratory participants only.) This finding suggests that the guises given to the listeners in these two new conditions may be equivalent to the L1-accent guise's advantage.

C. Discussion

Even though all listeners in experiment 2 were told that the speaker was an L2-accented speaker (in accordance with his actual language background), listeners who were additionally told that their performance on the task was likely to improve transcribed more words correctly than listeners only given information about the speaker's accent. Listeners told to expect a less accented L2 speaker showed numeric improvement over listeners only given accent source information, but this was not statistically significant. The model

did not support estimating different patterns of adaptation for different guise conditions, and thus across all conditions results indicated that most of the improvement occurred toward the beginning of the experiment. However, given that the total number of trials was limited to 40 sentences, future work should determine the time course of task-based gains in more detail.

In terms of interventions, these results present an intriguing suggestion that providing relatively simple information about the speaker (the source of their accent) alongside information emphasizing the non-fixed nature of listener ability (that listeners can/will improve in their accuracy) may improve intelligibility of foreign-accented speech, though perhaps not dramatically. This proposal aligns with social and educational psychological findings documenting the power of growth mindset vs fixed mindset (Dweck, 1999) and priming growth mindset (e.g., Aronson *et al.*, 2002; Lou and Noels, 2016).

IV. GENERAL DISCUSSION

Experiment 1 demonstrated that listeners are sensitive to the source of a speaker's accent. By the end of the experiment, listeners transcribed L2-accented speech better when they expected an L1-accented speaker as opposed to an L2-accented speaker. And, being given any prior expectations about the accent resulted in a benefit in intelligibility. These results were interpreted in terms of the construct of processing fluency, where being given expectations about the source of the speaker's accent reduced processing disfluency; listeners who did not entertain an L2-mode of processing (the L1-accent guise listeners) experienced the least amount of difficulty. Experiment 2 explored ways to reduce the difficulty encountered by listeners expecting an L2 accent, and found that listeners who were primed with the information that their ability to process L2 speech may improve did in fact improve more than those not given that information. That this information had an effect suggests that when listeners approach L2-accented speech *without* that information, their expectations based on prior experiences—that the task of processing L2-accented speech is difficult—may, in fact, contribute to the processing disfluency they experience in the task and thus their worse performance (though, of course, recall that the group receiving no information showed the worst performance). That is, since listeners in experiment 2 benefitted from being told that their ability to understand L2-accented speech may improve, this suggests that listeners primed to expect an L2-accent in experiment 1 may not have expected to succeed.

Taken together, the results of experiments 1 and 2 contribute to our understanding of the mechanisms involved in accented speech processing, and potential differences between processing L1 and L2 accents. First, it is noteworthy that, although prior work has found some acoustic similarities and perceptual confusability between L1 Latinx English speakers and L1 Spanish-L2 English speakers (Fought, 2003; Morales, 2014), it appears that listeners' beliefs about whether a speaker is a Latinx English vs L2 English speaker may have consequences on their processing. Next, the present results are in line with prior studies proposing differences in how L1- and L2-accented speech is processed (e.g., Adank *et al.*, 2009; Goslin *et al.*, 2012). Because intelligibility varied between conditions in this study, even though the acoustic signal remained the same, part of the difference between processing L1- and L2-accented speech may have to do with listeners' prior experiences with each type of listening, experiences that they bring with them to the task. To test this possibility further, future work could examine more detailed measures of individual listeners' reported prior experiences with various types of accented speech in relation to task performance. Of course, this general interpretation of the present results is not meant to suggest that processing disfluency or expectation-driven factors are the predominant forces in accented speech processing (differences observed between guise conditions in this study, though statistically significant, were small); certainly, lack of alignment between the acoustic signal and mental representation (Van Engen and Peelle, 2014) is a major cause. Rather, when the acoustic signal is equated, and other stereotype-based inferences are not as strongly activated (through photographs, for example), processing disfluency emerges as a potential explanation for why those expecting L1 speech outperformed those expecting L2 speech. This suggestion is in line with current models of listening effort, such as the Framework for Understanding Effortful Listening (FUEL; Pichora-Fuller *et al.*, 2016), which recognize the importance of factors beyond (mis)match to the acoustic signal to language processing, like motivation. In other words, the present results suggest that listeners' prior experiences with processing L2-accented speech are, alongside motivation, a component of the increased listening effort involved in processing L2-accented speech.

These results also augment prior work on the role of listener expectations in accented speech processing using

TABLE VI. Statistical model results of fixed effects for experiment 2.

Model formula: RAU ~ guise + trial quartile 4th + (1 | participant) + (1 + guise | sentence)				
Fixed effect	Estimate	Standard error	*t* value	*p* value
(Intercept)	58.212	3.872	15.035	<2e-16
L2-accent:Exp1 vs L2-accent:Speaker	3.407	1.806	1.887	0.0633
L2-accent:Exp1 vs L2-accent:Task	4.635	1.910	2.426	0.0178
Trial quartile 4th vs 1st	−6.269	1.068	−5.869	4.92e-09
Trial quartile 4th vs 2nd	−1.086	1.064	−1.021	0.3075
Trial quartile 4th vs 3rd	−1.299	1.061	−1.223	0.2213

photographs or nationality labels to trigger expectations (e.g., Rubin, 1992; McGowan, 2015; Niedzielski, 1999). In this study, explicitly directing listeners toward the type of speech they should expect indeed affected transcription accuracy, meaning that directly priming expectations about the source of an accent can have relevance for listeners, consistent with Hu and Su (2015) for other dependent measures. Further, based on prior work (e.g., Kleinschmidt and Jaeger, 2015; McMurray and Jongman, 2011), it is likely that listeners update their beliefs about the speaker during the course of the experiment, weighing the information in the guise presented to them against their growing experience with the speaker. Listeners' top-down expectations about the speaker may interact with their monitoring of the bottom-up signal, assessing incongruence and congruence (McGowan, 2015) over time in a dynamic process. Future work using matched guise methodologies to evaluate expectations may benefit from more explicitly assessing the integration of expectations with the bottom-up signal. Relatedly, since the present study suggests that different factors may exert effects on accented speech processing at different amounts of exposure to the speaker, alongside prior mixed findings regarding adaptation, more work addressing the time course of accented speech processing is warranted.

The integration of top-down and bottom-up information also applies to other aspects of listeners' processing strategies when listening to L1- vs L2-accented speech. The present study employed semantically anomalous sentences to impede listeners' ability to use contextual information in word recognition. An open question is how listeners expecting L1- vs L2-accented speech would behave when listening to semantically coherent sentences (as they would encounter in regular life). This is a particularly interesting question given prior proposals of different processing "modes" for different types of speech (Lev-Ari, 2015; Sumner, 2013; Sumner *et al.*, 2013), for example, favoring more top-down processing for L2- than L1-accented speech (Lev-Ari, 2015). Future work could investigate whether, separate from the speaker's actual L1- or L2-accent status, listeners *expecting* to hear an L2-accented speaker would employ more top-down strategies from the beginning than listeners expecting to hear an L1-accented speaker, or whether listeners deploy specific strategies only in response to the bottom-up signal they are actually hearing (or, whether the two factors interact over time).

To account for the differences between guises observed in this study, it has been suggested that a crucial factor is listeners' prior experiences with accented speech, which affects their level of processing fluency in the task. However, it may be that other accounts are at work, particularly the hypothesis suggesting that listeners do not uphold their end of the communicative burden when listening to accented speech (e.g., Kang and Rubin, 2009; Lippi-Green, 1997; though cf. McGowan, 2015). In such an explanation (discussed in Sec. I B), listeners are thought to attend less closely to speech they expect to be foreign-accented because of biases against accented speakers. In contrast, the processing fluency account suggests that listeners who expect foreign-accented speech do not limit the overall attention they bring to the task, but rather that their processing is not as effective because of divided attention or misallocation of resources. The present paper cannot tease apart these two accounts and, in fact, they may be linked; biases toward accented speakers could give rise to difficulty processing accented speech (as suggested by Lippi-Green, 1997), and difficulty processing accented speech could give rise to biases toward accented speakers (as suggested by Dragojevic *et al.*, 2017). However, the two accounts posit different roles of attention in processing accented speech, and would therefore make contrasting predictions about, for example, a divided attention task, which could be used in future work to attempt to disentangle these hypotheses (to the extent that they are separable).

Finally, this work has potential implications for the types of interventions that might be most effective in enhancing communication between native listeners and L2-accented speakers. On the surface, these results seem to suggest that an ideal strategy for communicative success would be for non-native speakers to untruthfully inform native listeners that they are, in fact, native speakers. However, McGowan's (2015) results show that giving *any* additional information about a speaker does not necessarily help; in his study, listeners who saw a photo of an Asian face when listening to Chinese-accented English transcribed that speech more accurately than those who saw a Caucasian face, but listeners who saw a Caucasian face (representing additional, though stereotype-incongruent information) did not improve compared to listeners who saw no guise (a silhouetted face). As discussed above, it is likely that listeners use expectations to guide their processing, but also continually integrate the incoming signal into their expectations, updating their model with additional information (Kleinschmidt and Jaeger, 2015; McMurray and Jongman, 2011). Just like McGowan's listeners had strong stereotype-driven expectations that Caucasians do not have Chinese accents, listeners in this study likely would not have believed that the Spanish-accented voice in this study had never had any contact with Spanish. Rather, the supporting information given about the speaker's use of Latinx English in the (otherwise incongruent) L1-accent guise helped listeners interpret the guise information and the voice in a congruent manner, benefiting processing. This design difference likely explains the difference between the incongruent guise results in the present study and McGowan's (2015) study, while not contradicting McGowan's incongruence-based account.

Thus, taken together with McGowan's (2015) results, these findings suggest that native listeners may benefit from a bit of additional plausible, congruent information about the language background of L2 speakers. Moreover, the results of experiment 2 augment this understanding, suggesting that the knowledge that listening to foreign accents is not a fixed skill may aid in intelligibility (potentially providing as much benefit as expecting to hear an L1-accent). The possibility that such interventions may help is supported by research documenting that stereotype threat can be reduced by drawing individuals' attention to the existence of stereotype threat (e.g., Johns *et al.*, 2005), and priming a growth mindset (e.g., Aronson *et al.*, 2002; Lou and Noels, 2016). That said, providing this additional contextual information would likely

only result in modest gains in intelligibility since in this study improvements were relatively small. Future work could determine whether priming listeners about their ability to improve in accented speech processing would also improve intelligibility of L1 accents. Future work could also explore the impact of contextual information on other outcome measures, such as comprehensibility or communicative success, since intelligibility and comprehensibility are not always correlated (Derwing and Munro, 1997), and since processing fluency has been found to affect similar measures of effort (such as perceived comprehension or judgments of learning) in other domains (e.g., Miele and Molden, 2010).

The proposed intervention puts more of the communicative burden on the shoulders of native listeners, presenting them with the knowledge that their own behavior is part of the communicative equation, and that they have the ability to improve. Again, it is not known how results from this web-based task with one speaker's semantically anomalous speech embedded in noise would translate to real-life communicative interactions (where listeners potentially have more exposure to the speaker, and certainly have more at stake in the interaction). However, these findings raise the intriguing possibility that giving native listeners a bit of situating information about the source of speakers' accents, and giving them the knowledge that their ability to understand accented speech can improve, could be potential interventions in a range of settings from classrooms to courtrooms to call centers and other increasingly common real-world settings involving interactions between native and non-native speakers.

ACKNOWLEDGMENTS

The author would like to thank Chun Liang Chan, Ann Bradlow, and the Speech Communication Research Group of the Linguistics Department at Northwestern University for use of and support with the online presentation software. Thanks to Molly Babel for a transcription scoring script and Tyler Kendall, three anonymous reviewers, and the associate editor for helpful feedback. Thanks also to Matthew Bauer, Sam Elliott, Celina Maldonado, Blair Prater, and Brandon Zuel for assistance with subject running.

[1]Of course, every speaker is, in actuality, an "accented speaker." In this paper, the term *accented speech* is used as a shorthand to represent situations where the speaker's accent is different from the listener's own.

[2]The term *guise* is used diversely across the literature, originally to represent the language variety presented in a given experimental condition (e.g., Lambert *et al.*, 1960), but also to designate the linguistic variant used in a condition (e.g., Campbell-Kibler, 2007), and the information given to listeners about a speaker in a condition (e.g., Babel and Russell, 2015; Kang and Rubin, 2009; McGowan, 2015); the present paper employs this third usage of the term.

[3]Specific varieties of Latinx English are sometimes called Latino English, Hispanic English, Chicano English, or Mexican-American English; *Latinx English* is the term adopted in this paper, based on Fought's (2003, 2006) use of the term Latino English.

[4]In the present study, it is not possible to disentangle whether any adaptation observed is best characterized as adaptation to the accent itself, adaptation to speech in noise, or some combination of the two.

[5]This question was phrased in this way because the term Latinx English is not widely used by the general public, and the question's potential answers were given in this way because participants may answer "no" for different reasons, likely due to their degree of familiarity with Latinx communities.

Adank, P., Evans, B. G., Stuart-Smith, J., and Scott, S. K. (**2009**). "Comprehension of familiar and unfamiliar native accents under adverse listening conditions," J. Exp. Psychol. Hum. Percept. Perform. **35**, 520–529.

Adank, P., and McQueen, J. M. (**2007**). "The effect of an unfamiliar regional accent on spoken-word comprehension," in *16th International Congress of Phonetic Sciences (ICPhS)*, pp. 1925–1928.

Aronson, J., Fried, C. B., and Good, C. (**2002**). "Reducing the effect of stereotype threat on African American college students by shaping theories of intelligence," J. Exp. Soc. Psychol. **38**, 113–125.

Atagi, E., and Bent, T. (**2015**). "Relationship between listeners' nonnative speech recognition and categorization abilities," J. Acoust. Soc. Am. **137**(1), EL44–EL50.

Babel, M., and Russell, J. (**2015**). "Expectations and speech intelligibility," J. Acoust. Soc. Am. **137**(5), 2823–2833.

Banks, B., Gowen, E., Munro, K. J., and Adank, P. (**2015**). "Cognitive predictors of perceptual adaptation to accented speech," J. Acoust. Soc. Am. **137**(4), 2015–2024.

Bent, T., Baese-Berk, M., Borrie, S. A., and McKee, M. (**2016**). "Individual differences in the perception of regional, nonnative, and disordered speech varieties," J. Acoust. Soc. Am. **140**(5), 3775–3786.

Bradlow, A. R., and Bent, T. (**2008**). "Perceptual adaptation to non-native speech," Cognition **106**(2), 707–729.

Brungart, D. S., Simpson, B. D., Ericson, M. A., and Scott, K. R. (**2001**). "Informational and energetic masking effects in the perception of multiple simultaneous talkers," J. Acoust. Soc. Am. **110**, 2527–2538.

Burchill, Z., Liu, L., and Jaeger, T. F. (**2018**). "Maintaining information about speech input during accent adaptation," PloS One **13**(8), e0199358.

Campbell-Kibler, K. (**2007**). "Accent, (ING), and the social logic of listener perceptions," Am. Speech **82**(1), 32–64.

Campbell-Kibler, K., and McCullough, E. (**2015**). "Perceived foreign accent as a predictor of face-voice match," in *Responses to Language Varieties: Variability, Processes and Outcomes*, edited by D. R. Preston and A. Prokhodkine (John Benjamins, Amsterdam), pp. 175–190.

Carmichael, K. (**2016**). "Place-linked expectations and listener awareness of regional accents," in *Awareness and Control in Sociolinguistic Research*, edited by A. Babel (Cambridge University Press, Cambridge, UK), pp. 152–176.

Clark, A. (**2013**). "Whatever next? Predictive brains, situated agents, and the future of cognitive science," Behav. Brain Sci. **36**(3), 181–204.

Clarke, C. M., and Garrett, M. F. (**2004**). "Rapid adaptation to foreign-accented English," J. Acoust. Soc. Am. **116**(6), 3647–3658.

Craik, F. I., and Tulving, E. (**1975**). "Depth of processing and the retention of words in episodic memory," J. Exp. Psychol.: Gen. **104**(3), 268–294.

Cristia, A., Seidl, A., Vaughn, C., Schmale, R., Bradlow, A., and Floccia, C. (**2012**). "Linguistic processing of accented speech across the lifespan," Front. Psychol. **3**, 479.

Derwing, T. M., and Munro, M. J. (**1997**). "Accent, intelligibility, and comprehensibility: Evidence from four L1s," Stud. Second Lang. Acquis. **19**(1), 1–16.

D'Onofrio, A. (**2015**). "Persona-based information shapes linguistic perception: Valley girls and California vowels," J. Sociolinguist. **19**(2), 241–256.

Dragojevic, M., Giles, H., Beck, A. C., and Tatum, N. T. (**2017**). "The fluency principle: Why foreign accent strength negatively biases language attitudes," Commun. Monogr. **84**(3), 385–405.

Dunning, D., and Balcetis, E. (**2013**). "Wishful seeing: How preferences shape visual perception," Curr. Dir. Psychol. Sci. **22**(1), 33–37.

Dweck, C. S. (**1999**). *Self-Theories: Their Role in Motivation, Personality, and Development* (Taylor and Francis, Philadelphia, PA).

Eckert, P. (**2008**). "Where do ethnolects stop?," Int. J. Bilingual. **12**(1-2), 25–42.

Fiske S. T. (**1998**). "Stereotyping, prejudice, and discrimination," in *The Handbook of Social* Psychology, edited by D. T. Gilbert, S. T. Fiske, and L. Gardner (McGraw-Hill, New York), pp. 357–411.

Floccia, C., Butler, J., Goslin, J., and Ellis, L. (**2009**). "Regional and foreign accent processing in English: Can listeners adapt?," J. Psycholinguist. Res. **38**, 379–412.

Floccia, C., Goslin, J., Girard, F., and Konopczynski, G. (**2006**). "Does a regional accent perturb speech processing?," J. Exp. Psychol. Hum. Percept. Perform. **32**, 1276–1293.

Fought, C. (**2003**). *Chicano English in Context* (Palgrave/MacMillan, New York).

Fought, C. (**2006**). *Language and Ethnicity* (Cambridge University Press, Cambridge, UK).

Freeman, J. B., and Ambady, N. (**2011**). "A dynamic interactive theory of person construal," Psychol. Rev. **118**(2), 247.
Gnevsheva, K. (**2018**). "The expectation mismatch effect in accentedness perception of Asian and Caucasian non-native speakers of English," Linguistics **56**(3), 581–598.
Godinez, M., and Maddieson, I. (**1985**). "Vowel differences between Chicano and general Californian English?," Int. J. Soc. Lang. **53**, 43–58.
Goslin, J., Duffy, H., and Floccia, C. (**2012**). "An ERP investigation of regional and foreign accent processing," Brain Lang. **122**, 92–102.
Grey, S., and van Hell, J. G. (**2017**). "Foreign-accented speaker identity affects neural correlates of language comprehension," J. Neurolinguist. **42**, 93–108.
Hanulikova, A., van Alphen, P. M., van Goch, M., and Weber, A. (**2012**). "When one person's mistake is another's standard usage: The effect of foreign accent on syntactic processing," J. Cognit. Neurosci. **24**, 878–887.
Hu, G., and Su, J. (**2015**). "The effect of native/non-native information on non-native listeners' comprehension," Lang. Awareness **24**(3), 273–281.
Ingvalson, E. M., Lansford, K. L., Fedorova, V., and Fernandez, G. (**2017a**). "Cognitive factors as predictors of accented speech perception for younger and older adults," J. Acoust. Soc. Am. **141**(6), 4652–4659.
Ingvalson, E. M., Lansford, K. L., Federova, V., and Fernandez, G. (**2017b**). "Listeners' attitudes toward accented talkers uniquely predicts accented speech perception," J. Acoust. Soc. Am. **141**(3), EL234–EL238.
Jamieson, J. P., and Harkins, S. G. (**2011**). "The intervening task method: Implications for measuring mediation," Pers. Social Psychol. Bull. **37**(5), 652–661.
Johns, M., Schmader, T., and Martens, A. (**2005**). "Knowing is half the battle: Teaching stereotype threat as a means of improving women's math performance," Psychol. Sci. **16**, 175–179.
Kamachi, M., Hill, H., Lander, K., and Vatikiotis-Bateson, E. (**2003**). "Putting the face to the voice: Matching identity across modality," Curr. Biol. **13**(19), 1709–1714.
Kang, O., and Rubin, D. L. (**2009**). "Reverse linguistic stereotyping: Measuring the effect of listener expectations on speech evaluation," J. Lang. Social Psychol. **28**, 441–456.
Kitterick, P. T., Bailey, P. J., and Summerfield, A. Q. (**2010**). "Benefits of knowing who, where, and when in multi-talker listening," J. Acoust. Soc. Am. **127**, 2498–2508.
Kleinschmidt, D. F., and Jaeger, T. F. (**2015**). "Robust speech perception: Recognize the familiar, generalize to the similar, and adapt to the novel," Psychol. Rev. **122**(2), 148.
Koelewijn, T., de Kluiver, H., Shinn-Cunningham, B. G., Zekveld, A. A., and Kramer, S. E. (**2015**). "The pupil response reveals increased listening effort when it is difficult to focus attention," Hear. Res. **323**, 81–90.
Konopka, K., and Pierrehumbert, J. (**2008**). "Vowels in contact: Mexican heritage English in Chicago," in *Symposium about Language and Society Austin (SALSA)*, Austin, TX.
Lambert, W. E., Hodgson, R., Gardner, R. C., and Fillenbaum, S. (**1960**). "Evaluational reactions to spoken languages," J. Abnorm. Social Psychol. **60**, 44–51.
Laturnus, R. (**2018**). "Perceptual adaptation to non-native speech: The effects of bias, exposure, and input variation," Doctoral dissertation, New York University.
Lev-Ari, S. (**2015**). "Comprehending non-native speakers: Theory and evidence for adjustment in manner of processing," Front. Psychol. **5**, 1546.
Lick, D. J., and Johnson, K. L. (**2015**). "The interpersonal consequences of processing ease: Fluency as a metacognitive foundation for prejudice," Curr. Direct. Psychol. Sci. **24**(2), 143–148.
Lindemann, S. (**2002**). "Listening with an attitude: A model of native-speaker comprehension of non-native speakers in the United States," Lang. Soc. **31**(03), 419–441.
Lippi-Green, R. (**1997**). *English with an Accent: Language, Ideology, and Discrimination in the United States* (Routledge, London, UK).
Liss, J. M., White, L., Mattys, S. L., Lansford, K., Lotto, A. J., Spitzer, S. M., and Caviness, J. N. (**2009**). "Quantifying speech rhythm abnormalities in the dysarthrias," J. Speech, Lang. Hear. Res. **52**(5), 1334–1352.
Liu, L., and Jaeger, T. F. (**2018**). "Inferring causes during speech perception," Cognition **174**, 55–70.
Lou, N. M., and Noels, K. A. (**2016**). "Changing language mindsets: Implications for goal orientations and responses to failure in and outside the second language classroom," Contemp. Educ. Psychol. **46**, 22–33.
Magnuson, J. S., and Nusbaum, H. C. (**2007**). "Acoustic differences, listener expectations, and the perceptual accommodation of talker variability," J. Exp. Psychol. Hum. Percept. Perform. **33**(2), 391.
McGowan, K. B. (**2015**). "Social expectation improves speech perception in noise," Lang. Speech **58**(4), 502–521.
McMurray, B., and Jongman, A. (**2011**). "What information is necessary for speech categorization? Harnessing variability in the speech signal by integrating cues computed relative to expectations," Psychol. Rev. **118**(2), 219.
Mendoza-Denton, N. (**2008**). *Homegirls: Language and Cultural Practice among Latina Youth Gangs* (Blackwell Publishing, Malden).
Miele, D. B., and Molden, D. C. (**2010**). "Naive theories of intelligence and the role of processing fluency in perceived comprehension," J. Exp. Psychol.: Gen. **139**(3), 535.
Morales, A. (**2014**). "Effects of listener characteristics on foreign-accentedness rating of a non-standard English dialect," Honors thesis, University of Arizona.
Moser, J. S., Schroder, H. S., Heeter, C., Moran, T. P., and Lee, Y. H. (**2011**). "Mind your errors: Evidence for a neural mechanism linking growth mind-set to adaptive posterior adjustments," Psychol. Sci. **22**, 1484–1489.
Niedzielski, N. (**1999**). "The effect of social information on the perception of sociolinguistic variables," J. Lang. Social Psychol. **18**(1), 62–85.
Oppenheimer, D. M. (**2008**). "The secret life of fluency," Trends Cognit. Sci. **12**(6), 237–241.
Ornstein-Galicia, J. (Ed.) (**1984**). *Form and Function in Chicano English* (Newbury House, Rowley, MA).
Pantos, A. J., and Perkins, A. W. (**2013**). "Measuring implicit and explicit attitudes toward foreign accented speech," J. Lang. Soc. Psychol. **32**, 3–20.
Pennington, C. R., Heim, D., Levy, A. R., and Larkin, D. T. (**2016**). "Twenty years of stereotype threat research: A review of psychological mediators," PLoS One **11**(1), e0146487.
Pichora-Fuller, M. K., Kramer, S. E., Eckert, M. A., Edwards, B., Hornsby, B. W., Humes, L. E., and Wingfield, A. D. (**2016**). "Hearing impairment and cognitive energy: The framework for understanding effortful listening (FUEL)," Ear Hear. **37**(Suppl. 1), 5S–27S.
Remez, R. E., Rubin, P. E., Pisoni, D. B., and Carrell, T. D. (**1981**). "Speech perception without traditional speech cues," Science **212**(4497), 947–949.
Rubin, D. L. (**1992**). "Nonlanguage factors affecting undergraduates' judgments of nonnative English-speaking teaching assistants," Res. Higher Educ. **33**(4), 511–531.
Santa Ana, O. (**1993**). "Chicano English and the Nature of the Chicano language setting," Hispanic J. Behav. Sci. **15**(1), 1–35.
Santa Ana, O., and Bayley, R. (**2004**). "Chicano English: Phonology," in *A Handbook of Varieties of English* (Mouton de Gruyter, Berlin), Vol. 1, pp. 417–434.
Schmader, T., Johns, M., and Barquissau, M. (**2004**). "The costs of accepting gender differences: The role of stereotype endorsement in women's experience in the math domain," Sex Roles **50**(11-12), 835–850.
Smith, R., Holmes-Elliott, S., Pettinato, M., and Knight, R. (**2014**). "Cross-accent intelligibility of speech in noise: Long-term familiarity and short term familiarization," Q. J. Exp. Psychol. **67**(3), 590–608.
Sohoglu, E., and Davis, M. H. (**2016**). "Perceptual learning of degraded speech by minimizing prediction error," Proc. Natl. Acad. Sci. U.S.A. **113**(12), E1747–E1756.
Staum Casasanto, L. (**2008**). "Does social information influence sentence processing?," in *Proceedings of the Annual Meeting of the Cognitive Science Society*, Vol. 30, No. 30.
Strand, E. A. (**1999**). "Uncovering the role of gender stereotypes in speech perception," J. Lang. Social Psychol. **18**(1), 86–100.
Studebaker, G. A. (**1985**). "A rationalized arcsine transform," J. Speech, Lang. Hear. Res. **28**(3), 455–462.
Sumner, M. (**2013**). "A phonetic explanation of pronunciation variant effects," J. Acoust. Soc. Am. **134**(1), EL26–EL32.
Sumner, M., Kurumada, C., Gafter, R., and Casillas, M. (**2013**). "Phonetic variation and the recognition of words with pronunciation variants," in *The Proceedings of the 35th Annual Meeting of the Cognitive Science Society*, pp. 3486–3492.
Tseng, A. (**2015**). "Vowel variation, style, and identity construction in the English of Latinos in Washington, DC," Doctoral dissertation, Georgetown University.
Van Engen, K. J., and Peelle, J. E. (**2014**). "Listening effort and accented speech," Front. Human Neurosc. **8**, 577.

Vaughn, C., Baese-Berk, M., and Idemaru, K. (**2018**). "Reexamining phonetic variability in native and non-native speech," Phonetica (published online).
Walker, A. (**2016**). "Advantage accented? Listener differences in understanding speech in noise," Univ. Penn. Work. Pap. Linguist. **22**(2), 20.
Warton, D. I., and Hui, F. K. (**2011**). "The arcsine is asinine: The analysis of proportions in ecology," Ecology **92**(1), 3–10.
Xie, X., Weatherholtz, K., Bainton, L., Rowe, E., Burchill, Z., Liu, L., and Jaeger, T. F. (**2018**). "Rapid adaptation to foreign-accented speech and its transfer to an unfamiliar talker," J. Acoust. Soc. Am. **143**(4), 2013–2031.
Yi, H. G., Phelps, J. E., Smiljanic, R., and Chandrasekaran, B. (**2013**). "Reduced efficiency of audiovisual integration for nonnative speech," J. Acoust. Soc. Am. **134**(5), EL387–EL393.
Zheng, Y., and Samuel, A. G. (**2017**). "Does seeing an Asian face make speech sound more accented?," Atten. Percept. Psychophys. **79**(6), 1841–1859.