

Expecting a challenge: A behavioral and neurophysiological investigation of talker identity in cross-dialectal speech perception

Holly A. Zaharchuk, Department of Speech, Language, and Hearing Sciences, University of Connecticut

Abby Walker*, Department of English, Virginia Tech, ajwalker@vt.edu

Adriana R. Miller, Department of Psychology, Pennsylvania State University

Carla Fernandez, Universidad Politécnica Taiwan-Paraguay

Janet G. van Hell, Department of Psychology, Pennsylvania State University

*Corresponding author.

Listeners use their knowledge about a talker to guide speech perception. In the present study, we manipulated both the familiarity and predictability of this talker-specific knowledge. Twenty-two listeners from western Pennsylvania completed an audio-visual lexical decision task while EEG was recorded. Critically, listeners were introduced to the talkers beforehand, with short videos establishing the talker's U.S. English accent identity: Mainstream, more similar to the participants themselves; Southern, a relatively less familiar variety; and Unpredictable, which switched between Mainstream and Southern accents. During the task, listeners watched these talkers produce tokens with accents that either aligned with or violated their accent identity. Behaviorally, alignment between talker and token accent influenced performance on Southern-accented tokens, while accuracy on Mainstream-accented tokens was nearly at ceiling regardless of talker accent identity. Neurally, accent familiarity had the largest effect pre-speech (in response to the visual presentation of the talker), while both predictability and familiarity impacted processing of the speech signal itself. Overall, our results suggest that listeners use information about a talker's accent, even when it is unfamiliar or unpredictable, to alter their listening strategies for more successful speech recognition.



1. Introduction

There is substantial non-random variability in the speech signal. Listeners take advantage of the systematicity in variation and use context to map particular signals to phonological and lexical representations. For example, on average, men have a lower frequency boundary between /s/ and /ʃ/ than women. Previous research has shown that the same ambiguous fricative is heard as /ʃ/ when listeners think the speaker is a woman, but as /s/ when they think the speaker is a man (Strand, 1999; Strand & Johnson, 1996). Johnson et al. (1999, p. 380) attribute these findings to listeners having pre-existing expectations around gendered language that they employ in speech perception. In fact, there is general agreement that context impacts perception by prompting listeners to shift their perceptual systems, though accounts of how these expectations shape processing differ. Listeners could be activating different exemplars (Johnson et al., 1999), using different mappings (Samuel & Larraza, 2015), or assuming different distributional priors (Kleinschmidt & Jaeger, 2015).

The source of variation we are interested in comes from pronunciation differences across regional dialects. For example, speakers in the Southern United States often produce PRIZE-class words with a monophthong compared to speakers from California, who produce a diphthong. There is evidence that listeners also use contextual information to adjust to this type of variation. If a listener is led to believe a speaker is from a particular dialect region (D’Onofrio, 2015; Niedzielski, 1999), or even if a dialect region is simply primed (Hay & Drager, 2010; Hurring et al., 2022; Walker et al., 2018a; Weatherholtz et al., 2014), listeners can interpret the same auditory signal differently. Researchers generally agree that listeners use the same mechanisms to adjust to dialectal variation as they do to adjust to other sources of variation, like gender (Foulkes & Docherty, 2006). For example, Hay and Drager’s (2010, p. 883) explanation for a shift in perception by New Zealanders in response to Australian primes is that listeners activate the concept of “Australia” and, through spreading activation, speech exemplars associated with this concept.

However, one difference between regionally-based variation and other sources of structured variation is a listener’s experience. While we would expect most listeners to regularly hear speech from both men and women, for example, we would not expect all listeners to have substantial exposure to different dialects. This means that listeners may have relatively little experience with mapping a regionally-accented signal to the intended utterance. Indeed, there is substantial evidence that processing less familiar dialects is more difficult than processing familiar dialects (e.g., Clopper & Bradlow, 2008; Floccia et al., 2006; Labov & Ash, 1997), and that listeners perform better with a second dialect (D2) as a function of experience with this particular D2 (Evans & Iverson, 2004; Sumner & Samuel, 2009; Walker, 2018). In the present study, we test the hypothesis that listeners rapidly form expectations about a talker’s speech patterns and use this

information to facilitate word recognition. Specifically, we investigate how the *predictability* of a talker's speech patterns and a listener's *familiarity* with these patterns influence the perception of regional accents in U.S. English. We use both behavioral and neural methods to measure the effects of talker identity at different stages of cross-dialectal processing.

1.1 Effects of context on cross-dialectal speech perception

While context has been argued to impact speech perception because listeners shift into a dialect-specific listening mode to facilitate processing (by activating exemplars, changing priors or mappings, etc.), it is not clear if a dialect-specific listening mode is available to listeners with limited familiarity with the contextually-cued dialect. If we look to the literature, we can build three predictions for how context may impact perception in these cases. On the one hand, context may have no impact on speech perception when experience is limited; a listener's representation of the unfamiliar dialect is too weak to be activated by context, and listeners perform poorly with the less familiar accent regardless of whether there is facilitative context or not (cf. Evans & Iverson, 2004; Van der Feest & Johnson, 2016; Walker et al., 2020; Weatherholtz et al., 2014). On the other hand, context may impact the processing of familiar and less familiar accents alike by prompting listeners to activate specific mappings, distributions, or exemplars in both cases. The robustness of the dialect representation does not matter, and even a small amount of lived experience or knowledge of stereotypes can be invoked with context (McGowan, 2015; see Wade et al., 2023, and Walker, 2019, for similar arguments about speech production).

A third hypothesis is that listeners do not shift into a *dialect-specific* listening mode for less familiar dialects, but instead shift into a more *uncertain* listening mode, allowing for more flexibility in signal-to-word mappings. Work coming from the perceptual adaptation literature suggests that, in response to exposure to atypical pronunciations, listeners may relax their criteria for phonemic categorization, demonstrated by a general increase in the endorsement of pseudowords (rather than a specific increase based on exposure to the variant of interest; e.g., Babel et al., 2021; Bissell & Clopper, 2025; Weatherholtz, 2015; Zheng & Samuel, 2020).¹ Relatedly, there is evidence that listeners are more flexible in their mappings when the acoustic signal is overall less reliable (Brouwer et al., 2012; McMurray et al., 2019; Van Ooijen, 1996) or if they have reason to expect dialectal variation (Clopper & Walker, 2017). Based on these findings, listeners may respond to a less familiar dialect by shifting into an uncertain listening mode, characterized by relaxed categorization criteria or overall category broadening. In other

¹ There are analogous findings in the syntactic literature that listeners are less likely to show the expected P600 to ungrammatical (unattested) forms (Weissler & Brennan, 2020) and are more likely to associate ungrammatical (unattested) sentences (Maher et al., under review) with talkers whose accents are stigmatized or less familiar (though see Zaharchuk et al., 2021).

words, the context may make listeners aware that different acoustic-phonetic mappings are likely to occur but not increase their confidence about what these mappings should be.²

1.2 Talker identity as a contextual cue to talker dialect

In the present study, we play listeners speech samples from actresses performing two different dialect guises: Mainstream U.S. English (MUSE), a term we use here to capture something akin to a pan-regional variety associated with the middle class and regionally most similar to Midland and West Coast dialect regions (Clopper & Bradlow, 2008; cf. Lippi-Green, 1997, p. 59) and Southern U.S. English (SUSE), a highly enregistered regional variety tied to the southeast U.S. (Labov et al., 2006; Preston, 2018). The listeners in our study are from western Pennsylvania (treated as a distinct dialect region between the Midland and Inland North in Labov et al., 2006), and MUSE is expected to be more familiar to them than SUSE.³ Other studies have found that similar participants perform worse in listening tasks with SUSE compared to MUSE accents (Clopper & Bradlow, 2008; Walker et al., 2018b). We are interested in how performance with these two dialects (one more familiar, one less familiar) changes with contextual cues.

The contextual manipulation in our study is talker identity. While most studies manipulate *perceived* talker identity by pairing a target speech sample with photos or videos of people of different genders, ethnicities, ages or social classes (e.g., Drager, 2011; McGowan, 2015; Strand & Johnson, 1996), we use the term “talker identity” to refer to a talker that listeners are already familiar with. They have heard *this particular person* speak before and therefore have had a chance to form expectations about this person’s broader linguistic patterns (i.e., their accent and style), as well as their idiosyncrasies and vocal tract characteristics. Kleinschmidt and Jaeger (2015, p. 171) frame talker identity as the most valuable type of information a listener can have in making predictions about speech. There is plenty of evidence that listeners are tracking fairly fine-grained patterns in the speech of known individuals (Podesva et al., 2015; Remez et al., 1997) and that processing speech is easier if produced by a previously encountered talker (Clapp et al., 2023; Nygaard et al., 1994; Nygaard & Pisoni, 1998; Palmeri et al., 1993; Sheffert & Olson, 2004). In short, manipulating talker identity should be a strong and realistic way to alter listeners’ expectations of a talker’s regional accent.

In the present study, we use video (audio-visual presentation) to familiarize our listeners with the talkers—and, critically, with their regional accents—to arm them with talker-specific

² Maher (2023, p. 18), talking about morphosyntactic processing, and referring to Labov (1973), nicely puts it this way: “In Labov’s framework of knowledge, they are *evaluating* that a particular language variety is associated with grammatical anomalies, but not accurately *predicting* the typical limits of the anomalies.”

³ Another difference between these dialects is their societal status, with MUSE, essentially by definition, being the standardized variety and SUSE being a regionalized and often stigmatized variety (Hasty, 2018; Preston, 2018). In this way, it may be difficult to disentangle the effects of familiarity from the effects of prestige, a point that we address more directly in the Discussion.

knowledge for later speech processing. Bimodal presentation of speech has been shown to aid in speech perception (e.g., Summerfield, 1979; see also Walker et al., 2020) and the learning of specific voices (Sheffert & Olson, 2004). Of most relevance to our study, Molnar et al. (2015) showed that early bilingual listeners use audio-visual talker identity to form predictions about which language to expect from a talker, as evidenced by an increase in reaction time for incongruent talker-language pairs (see also Walker et al., 1995). Critically, they found that late bilinguals simply show a slow down for their less dominant language, regardless of talker-language (in)congruence. Taken together, these findings suggest that different levels of experience with a language variety impact the usefulness of talker identity.

In addition to manipulating talker identity, Molnar et al. (2015) also manipulated talker-language predictability. This study, and their related EEG study (Martin, Molnar et al., 2016), familiarized Spanish-Basque bilingual listeners with two types of talkers: monolingual talkers, who only used either Spanish or Basque during familiarization, and bilingual talkers, who switched between Spanish and Basque during familiarization. In other words, the language of monolingual talkers was predictable, and the language of bilingual talkers was unpredictable. Both studies found evidence that predictability impacted performance on the subsequent test phase: In terms of both accuracy and ERPs, processing the unpredictable bilingual talkers was more challenging than processing the predictable monolingual talkers. The authors argue that, in the case of the bilingual code-switching talkers, listeners could not use talker identity as a cue to predict the language of the upcoming signal. In fact, Martin, Molnar et al. (2016) even found differences between bilingual and monolingual talkers *before* the onset of audio, suggesting that listeners *prepare* to listen differently based on their linguistic expectations for a talker.

1.3 Regional dialects in online processing research

Like Molnar et al. (2015) and Martin, Molnar et al. (2016), we use talker identity as a cue to language variety, but with *dialects*, not languages. Specifically, we introduce participants to talkers who use MUSE (familiar), SUSE (less familiar), or who switch between both (unpredictable) during familiarization. We then see how listeners respond to the two dialects from these talkers in a lexical decision task. In addition to measuring lexical decision performance *behaviorally*,⁴ we also measure the impact of dialect familiarity and talker identity *neurally* with event-related potential (ERP) analysis of electroencephalographic (EEG) data. ERPs are measurements of electrical brain activity that are time-locked to an external event, such as hearing a word. ERPs provide a millisecond-by-millisecond measurement of the brain's activity during mental processing as it unfolds over time, and behavioral measures (e.g., reaction times, accuracy) reflect the outcome of this process. ERPs can thus detect early stimulus-related changes in language processing that cannot be measured via behavioral responses, which typically capture the endpoint of processing.

⁴ For a solely behavioral investigation of the same stimuli with a different population, see Walker et al. (2020).

Combining neural (ERP) and behavioral (reaction time, accuracy) measures allows us to capture the different phases of language comprehension, beginning from the very onset of the stimulus and going through the participant's manual response. In general, this combination of methods provides a full picture of the comprehension process. In our particular study, measuring both neural activity and behavior allows us to connect the sociolinguistic literature on cross-dialect communication to the neurocognitive literature on word recognition.

ERP research has found that familiarity with a particular dialect or accent impacts online processing of phonological (Brunellière & Soto-Faraco, 2013, 2015; Goslin et al., 2012), semantic (Martin, Garcia et al., 2016), and grammatical (Garcia et al., 2022; Weissler, 2021; cf. Zaharchuk et al., 2021) aspects of this variety. For example, Goslin et al. (2012) presented sentences in both the listeners' own (D1) accent and several less familiar regional (D2) accents in a passive English listening task. They found that the D2 accents elicited larger phonological mapping negativities (PMN), indexing activation of phonological candidates, than the D1 accent. On the later N400 component, indexing lexico-semantic access, D1 and D2 processing did not differ. Together, these results suggest that mapping external acoustic signals onto internal phonetic categories requires more resources for less familiar regional accents than for the listener's own accent; however, this effortful normalization process allows lexico-semantic access to proceed unimpeded. These results also demonstrate the benefit of investigating cross-dialectal communication with ERPs; observing differences on the PMN rather than the N400 component revealed that phonological rather than lexico-semantic processing was the source of comprehension difficulties in the D2. In other words, the latency of the ERP effects revealed the underlying mechanisms of cross-dialectal processing. By contrast, the behavioral measure of comprehension question accuracy only revealed that there were difficulties, but could not indicate at which stage of processing these difficulties arose.

In the present study, our participants are exposed to D1 and D2 accents through an auditory lexical decision task. In this task, listeners decide whether the token they heard was a real word or a pseudoword. ERP studies comparing real word and pseudoword processing typically observe a lexicality effect, where pseudowords elicit larger N400 amplitudes than real words (e.g., Chwilla et al., 1995; Holcomb et al., 2002; Liu & Van Hell, 2020; Martin, Molnar et al., 2016). A larger N400 is understood to reflect challenges in lexico-semantic access, with real words being easily found in the lexicon and pseudowords being searched for but never found (Federmeier, 2022; Kutas & Federmeier, 2011). The size of the N400 lexicality effect can also be influenced by talker identity; indeed, in their work with bilingual listeners, Martin, Molnar et al. (2016) found that bilingual talkers elicited smaller differences between real words and pseudowords on the N400 than monolingual talkers. In other words, lexical status was more easily established when the language was predictable. Differences in N400 amplitude have also been associated with neighborhood density, where real words and pseudowords with more

lexical neighbors elicit larger N400 amplitudes than similar items with fewer neighbors (e.g., Carrasco-Ortiz et al., 2017; Meade et al., 2019). This N400 modulation is thought to be driven by an increase in global lexico-semantic processing due to the many coactivated neighbors. The effect of lexical competitors on the N400 is especially important to consider given the potential responses to dialect context we outlined in Section 1.1. If listeners respond to an unfamiliar dialect by broadening their expectations of possible mappings, this would introduce more lexical competition, reducing the N400 effect.

While lexicality effects are typically observed on the N400, Martin, Molnar et al. (2016), among others (e.g., Signoret et al., 2013), have also observed differences between real words and pseudowords on the N1, a much earlier component. The N1 is associated with processing the acoustic features of the signal (e.g., Helenius et al., 2002; Kapnoula & McMurray, 2021; for a review, see Getz & Toscano, 2021). A related component, the P2/P200, has been associated with processing phonetic category information (Brunellière et al., 2009). Both of these components are elicited by complex auditory stimuli and can be enhanced by attention (Čeponienė et al., 2005; Picton & Hillyard, 1974). In Martin, Molnar et al., the N1 lexicality effect—a larger response to pseudowords than to real words—was present only for the (predictable) monolingual talkers and not for the (unpredictable) bilingual talkers. This suggests that talker identity can facilitate lexical access at the earliest stages of auditory processing by cueing the upcoming acoustic-phonetic information in the speech signal. Additionally, as we mentioned earlier, Martin, Molnar et al. also observed differences in brain responses to talkers *before* the auditory stimulus was even played, with the visual presentation of the talker's faces eliciting a larger P3b for bilingual talkers than for monolingual talkers. This pre-speech effect may reflect the activation of talker identity in anticipation of the upcoming auditory information. To summarize, the pre-lexical (i.e., earlier than the N400) effects of interest are the pre-speech P3b and the post-speech N1 and P2, which index different aspects of acoustic-phonetic processing. Together, these components may reflect the proactive use of talker identity for language processing.

1.4 Present study

Building on previous work, the present study investigates cross-dialectal (D1–D2) language processing both behaviorally and with EEG/ERPs. We manipulated talker identity—predictable monodialectal talkers and unpredictable bidialectal talkers—and accent familiarity—familiar Mainstream (MUSE) accents and less familiar Southern (SUSE) accents—to understand how listeners use their language experience for speech recognition. We audio-visually introduced participants to three types of talkers during an initial familiarization phase: Mainstream talkers, who used a MUSE accent throughout; Southern talkers, who used a SUSE accent throughout; and Unpredictable talkers, who switched between a MUSE and a SUSE accent. These videos established each talker's identity, as well as the predictability of their accent. During the subsequent lexical

decision task, listeners watched videos of the same talkers producing real word and pseudoword tokens. Critically, these tokens bore either a MUSE or a SUSE accent, which either aligned with or violated the listener's expectations about the talker.

Following the work by Martin, Molnar et al. (2016), our hypotheses for the ERP data were focused on the N400 effect, where we had the strongest *a priori* predictions about the effects of lexicality. These predictions are laid out below in **Table 1**. We present three ways that talker and token accent could impact the N400, assuming that our western Pennsylvania listeners will be more familiar with MUSE compared to SUSE. First, if high familiarity with a dialect is necessary to form specific expectations about a talker's pronunciation, we would expect the properties of the token to matter more than the properties of the talker. That is, we would expect pseudowords to elicit larger N400 effects than real words regardless of the talker's *expected* accent, but this effect would be mediated by token accent (with the expectation of a weaker N400 lexicality effect for the less familiar SUSE accent). Alternatively, if listeners adopt a more uncertain listening strategy when faced with a talker of a less familiar dialect, we would expect to see strong lexicality effects for Mainstream talkers that would be attenuated for Southern and Unpredictable talkers. Finally, if listeners form robust expectations regardless of the degree of their previous experience with an accent, we would expect congruency between talker and token to matter more than the token itself. That is, we would expect SUSE real words and pseudowords to elicit weaker N400 effects than MUSE real words for the Mainstream talkers, and we would expect MUSE real words and pseudowords to elicit weaker N400 effects than SUSE real words for Southern talkers. For the Unpredictable talkers, we would expect a lexicality effect regardless of token accent, since neither variety has a stronger association with the talker than the other; however, this lexicality effect would probably be attenuated compared to the monodialectal talkers, who predictably use one variety or the other.

In addition, we were motivated to investigate three other ERP components: the pre-speech P3b and the pre-lexical N1 and P2. The pre-speech analysis was based on Martin, Molnar et al. (2016), who observed differential activation of talker identity for the bilingual versus the monolingual talkers. We expect a similar difference between our Unpredictable talkers and the two predictable talkers (Mainstream and Southern), which would reflect pre-activation of talker identity to facilitate perception of the unpredictable acoustics. We also investigated the pre-lexical N1 and P2, which index different levels of processing but tend to pattern together in response to salient stimuli. Martin, Molnar et al. observed a larger N1 response to pseudowords than to real words for monolingual but not for bilingual talkers. An analogous pattern in our study would be an N1 (and potentially P2) lexicality effect for Mainstream and Southern talkers but not for Unpredictable talkers. Together, these components should help us understand how the differences between a listener's expectations, their underlying phonological representations, and the actual token accents are detected and resolved (or not) before lexico-semantic access takes place.

Hypothesis	Talker accent	MUSE real words vs. pseudowords	SUSE real words vs. pseudowords
Listeners do not adjust to less familiar accents (token accent matters most)	Mainstream talker	Lexicality effect	Weaker/no lexicality effect
	Southern talker		
	Unpredictable talker		
Listeners expect a challenge from “accented” talkers (talker accent matters most)	Mainstream talker	Lexicality effect	
	Southern talker	Weaker/no lexicality effect	
	Unpredictable talker	Weakest/no lexicality effect	
Listeners adjust to less familiar accents (congruency between token and talker accent matters most)	Mainstream talker	Lexicality effect	Weaker/no lexicality effect
	Southern talker	Weaker/no lexicality effect	Lexicality effect
	Unpredictable talker	Same lexicality effect regardless of trial dialect (weaker than effects for congruent trials for predictable talkers)	

Table 1: Competing predictions for the N400 lexicality effect.

2. Methodology

2.1 Participants

Thirty undergraduate students from Penn State’s subject pool received course credit for their participation. We recruited participants according to the following criteria: 18–35 years of age, right-handed, normal or corrected-to-normal vision, normal hearing, no history of head trauma, no history of language or neurological disorders, and monolingual English language background with limited second language experience. Two participants with bilingual language backgrounds were excluded prior to the EEG session. Five participants were unable to complete the EEG session due to technical difficulties. One additional participant was excluded after reporting abnormal hearing. Data from the remaining 22 participants were analyzed (Age: $M = 18.36$, $SD = 0.58$, $Min = 18$, $Max = 20$; Sex: 18 female, 4 male). All participants provided informed consent.

2.2 Materials

2.2.1 Talkers

We recruited six actresses to record our stimuli, all of whom identified as white and none of whom identified as native speakers of SUSE. Two actresses were from Lynchburg, Virginia, and one each from northern Virginia; Suffolk, Virginia; Maryland; and Wisconsin. The actresses produced

two different accent guises for the stimuli: one SUSE (rhotic Southern US English accent) and one MUSE (a regional Mainstream US English accent). We recruited non-Southern actresses based on previous work (Walker et al., 2018b). In that study, self-identified bidialectal Southerners were asked to produce MUSE and SUSE guises on command, but were still frequently heard as Southern in their MUSE guise (see Weissler, 2021, for related results using a bidialectal talker). In the present study, we needed the talkers to switch categorically between the two varieties, both for more interpretable effects of accent and for better comparison to Martin, Molnar et al.'s (2016) bilingual/monolingual interlocutors. **Table 2** presents mean ratings of accent strength and difficulty of understanding for each of our talkers (collected after the audio-visual lexical decision task; see Section 2.2.3 for survey details). Every talker received stronger accent ratings in their Southern talker identity than in their Mainstream talker identity (for a significant overall effect of identity on accented ratings, $t(84.83) = -6.19, p < .001$); and all but one were rated as harder to understand in their Southern guise (for a significant overall effect of identity on ease of comprehension ratings, $t(85.95) = -3.01, p = .003$). These comparisons also demonstrate that the actresses were heard as doing something different in each guise.

	<i>Mainstream talker</i>			<i>Southern talker</i>			<i>Unpredictable talker</i>		
<i>Talker</i>	Accent strength	Comprehension difficulty	<i>N</i>	Accent strength	Comprehension difficulty	<i>N</i>	Accent strength	Comprehension difficulty	<i>N</i>
Alice	4.00	2.89	9	6.60	4.00	6	6.00	3.29	7
Amber	2.57	2.14	7	4.86	3.43	7	3.38	2.25	8
Hannah	3.20	1.80	5	5.00	4.00	9	5.00	2.62	8
Kendall	3.17	2.67	6	5.43	4.43	7	5.89	3.44	9
Nora	1.88	2.00	8	3.43	1.71	7	2.71	2.29	7
Sara	2.78	1.67	9	4.62	2.50	8	4.20	4.00	5
Average	2.93	2.20	44	4.91	3.34	44	4.59	2.93	44

Table 2: Post-experiment participant evaluations of talkers.

Despite our confidence that the speakers were heard as *doing* Southern in their Southern guise, readers might reasonably be concerned about the authenticity of their performance. In Appendix A, we compare our actress' vowel productions to other recent recordings of both older and younger women from southwest Virginia. In the Southern talker accent, our actresses look like speakers from the area, though they align best with an older sample: They produce a monophthongal PRIZE, a diphthongal THOUGHT, and show the Southern Vowel Shift (high DRESS and KIT vowels and a diphthongal FACE that starts under DRESS and ends above it).

An analysis of their stop consonants published in Walker (2020) also shows that they produced significantly more negative lag Voice Onset Time in their SUSE versus MUSE tokens (reflecting documented patterns in SUSE), demonstrating that they also changed features below the level of conscious awareness.

2.2.2 Stimuli

The stimuli were recorded at a TV studio on the Virginia Tech campus. The talkers produced stimuli to camera with the aid of a teleprompter. Audio was recorded with a Sennheiser MKE 600 Shotgun Condenser Microphone (off-screen) connected to a Panasonic AJHPD2500 P2 Recorder (40KHz, 16 bit). Stimuli were recorded as MOV files and later converted to WMA files for compatibility with E-Prime.

Talkers recorded two types of stimuli: introductory monologues and single-word tokens. Each actress was assigned a character name and two unique monologues about their character's life. For example, below are the introductions to the two monologues from "Sara":

Hi, I'm Sara! Something interesting about me? Well, I don't know if it's super interesting, but I love to paint. I try to do one different painting a week, just something small...

Hi, it's Sara again! A pet peeve of mine is televisions in restaurants. I get it at sports bars, like, obviously you need TVs for the game, but I don't get why we need them anywhere else...

The talker recorded each monologue twice: once in a MUSE accent and once in a SUSE accent. Each monologue, edited tightly with jump cuts, was approximately one minute long. From the Southern and Mainstream versions of a monologue, we created a third version that switched between the two different versions at some, but not all, of the jump cuts. We created this version as an analog to the bilingual code-switching condition in Molnar et al. (2015) and Martin, Molnar et al. (2016) to make our results maximally comparable, especially since some of their most interesting results came from responses to the bilingual talkers. However, it is important to note that in the context of dialect-shifting, this is a fairly unnatural condition; bidialectal talkers rarely show such extreme shifts between styles across their whole repertoires (Hazen, 2001; Labov, 1998), let alone within a one-minute conversation with no clear change in external factors (though see Sharma, 2018). This is why we choose to call this version "Unpredictable" as opposed to "Bidialectal," to avoid the claim that this represents a naturalistic degree of style-shifting (which is not a necessary condition of our research question).

The purpose of the introductory monologues was to establish each character's *Talker Accent*—Mainstream, Southern, or Unpredictable—for the primary experimental task. This task was a lexical decision task that included 300 monosyllabic real words and 120 monosyllabic pseudowords. The variable of real word versus pseudoword is referred to as *Token Type*. The

real words were selected because they contained vowels that can strongly mark SUSE (Gunter et al., 2020): KIT (52), DRESS (52), PRIZE (55), FACE (53), STRUT (55) and THOUGHT (33). The majority of the pseudowords (69%) did not contain these vowels. The actresses recorded each token twice: once in a MUSE accent and once in a SUSE accent. We refer to accents in the lexical decision task as the *Token Accent*.

The recordings were edited such that the video always started before the audio ($M = 386\text{ms}$, $SD = 94\text{ms}$, $Min = 161\text{ms}$, $Max = 580\text{ms}$). The video ended an average of 305ms after the audio ($SD = 51\text{ms}$, $Min = 90\text{ms}$, $Max = 602\text{ms}$). Across the two token types and six actresses, the duration of SUSE words was 617ms on average ($SD = 126\text{ms}$), and the duration of MUSE words was 604ms on average ($SD = 122\text{ms}$). While this difference is consistent with both documentation (Clopper & Smiljanic, 2015; Jacewicz et al., 2007) and stereotypes (Niedzielski & Preston, 2000) that SUSE is slower than MUSE, word length did not differ significantly between the two accents (see Appendix B for details).

2.2.3 Additional materials

A language history questionnaire elicited self-reports of language acquisition, exposure, and proficiency, in addition to demographic, health, and handedness information. This was used to confirm participants' eligibility for the experiment.

A debriefing questionnaire elicited judgments about each of the six actresses from the EEG session. Participants were first asked to indicate whether there was a difference among the six talkers in terms of accent (all but one said yes). Participants were then prompted to rate the strength of each talker's accent and the difficulty with which they understood each talker on a scale from one (very little accent/very easy to understand) to seven (very strong accent/very hard to understand; see **Table 2** for mean accent strength and comprehension difficulty ratings, respectively). Participants were also asked where they thought each talker was from and whether they had additional comments about the talkers.⁵

A post-experiment questionnaire also measured each participant's accent affiliation and prescriptive language attitudes. In addition, we collected individual differences measures of working memory (Automated Operation Span Task [O-Span]; Redick et al., 2012), inhibitory control (AX-Continuous Performance Test [AX-CPT], Morales et al., 2013), and language proficiency (verbal fluency), but they will not be analyzed in the present study.

⁵ When asked if they perceived any differences in accent between the talkers, only one participant indicated that they noticed any within-talker variation: "Some had Southern accents at times." Otherwise, nine responses included a simple binary comparison between talkers with Southern or "country" accents and talkers with Mainstream accents. Another eight responses commented that the strength of the Southern accent varied between talkers. When asked if they had any other comments about the talkers, another participant noted: "Yeah, the brunette one [unclear which talker this is] kept changing it. Like it was Southern to no accent at all." Overall, participants either did not overtly detect a difference between the Southern and Unpredictable talkers or encoded the Unpredictable talkers as having a weaker Southern accent than the Southern talkers.

2.3 Design

Twelve experimental lists were created so that each word was said by all actresses in both accents across participants (6 actresses x 2 token accents) and each actress was presented with each talker identity in four lists (3 talker accents x 4). In each list, two of the actresses were presented as Southern talkers, two as Mainstream talkers, and two as Unpredictable talkers (**Figure 1**). An actress' talker accent was reinforced throughout the lexical decision task, such that 58 of the 70 tokens from each SUSE talker (>80%; 38 real words and 20 pseudowords) were presented in a SUSE accent, 58/70 tokens from each MUSE talker were presented in a MUSE accent, and 35/70 tokens from each Unpredictable talker (50%; 25 real words and 10 pseudowords) were presented in each accent. In other words, only 12 of the 70 tokens for each monodialectal talker were produced with incongruent token accents (and these incongruencies were all real words). These distributions allowed us to maintain the Talker Accent established in the monologues (Southern, Mainstream, Unpredictable) while also investigating alignment with Token Accent (MUSE, SUSE).

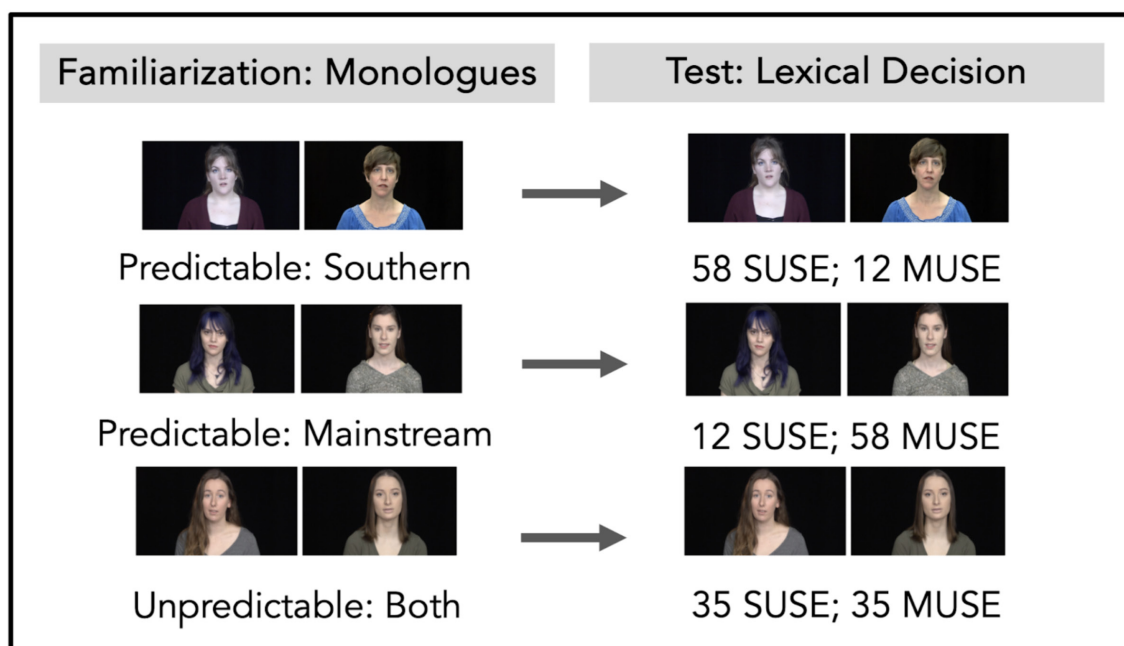


Figure 1: Design for one of the 12 experimental lists. Numbers reflect how many trials were spoken by each individual talker at test, including both real words and pseudowords. For Predictable talkers, all of the incongruent trials were real words.

2.4 Procedure

Participants completed the language history questionnaire at the beginning of the experiment. The EEG session then began with the familiarization phase, in which each character's Talker Accent was established. Recall that each actress was assigned two monologues. The monologues

were presented to participants in two runs, such that the first monologue from each of the six talkers was presented in a row before the second set of monologues. Participants answered four short comprehension questions after each monologue (48 total) to encourage attention.

Immediately following the familiarization phase, participants completed the audio-visual lexical decision task while EEG was recorded. At the beginning of each trial, a fixation cross appeared for approximately 500ms. Next, the audio-visual stimulus began playing, in which one of the actresses was depicted producing either a real word or a pseudoword with either a MUSE or a SUSE accent. After the stimulus finished playing, participants saw “Real word?” on the screen and indicated whether the word was “real” or “fake” by pressing one of two buttons on the button box (response laterality was counterbalanced across participants). A response initiated the next trial. Talker accent, token accent, and token type were pseudo-randomized across trials (420 total), such that no more than three congruent trials from the same talker (e.g., a MUSE token produced by Sara as a Mainstream talker) appeared in a row and no more than two incongruent trials from the same talker (e.g., a MUSE token produced by Sara as a Southern talker) appeared in a row.

After the EEG session, participants were seated in a separate sound-attenuated booth at a desktop computer. The debriefing questionnaire was completed first, followed by the AX-CPT, O-Span, and verbal fluency tasks. Finally, participants completed the post-experiment questionnaire.

2.5 EEG acquisition and pre-processing

During the EEG session, participants sat in a comfortable chair approximately three feet from a computer monitor in a sound-attenuated booth. An elastic cap (Brain Products ActiCap, Germany) with 31 active Ag/AgCl electrodes located along the midline (Fz, FCz, Cz, Pz, Oz) and laterally (FP1/2, F7/8, F3/4, FC5/6, FC1/2, T7/8, C3/4, CP5/6, CP1/2, P7/8, P3/4, O1/2, PO9/10) recorded EEG. Bipolar recordings above and below the left eye (VEOG) and at the outer canthi of both eyes (HEOG) monitored for vertical and horizontal eye movements, respectively. Electrode impedances were kept below 10 k Ω . Online, the EEG signal used a vertex reference (FCz), was amplified by a NeuroScan SynampsRT amplifier with a 0.05–100 Hz bandpass filter (first-order Butterworth with 6 dB/octave roll-off), and was continuously sampled at a rate of 500 Hz. The EEG sessions were programmed with E-Prime 2.0 software. Auditory stimuli were presented over insert headphones (Etymotic Research Inc., Elk Grove Village, IL). Visual stimuli were presented on the computer monitor on a black screen. Button presses were recorded using a serial response box (Psychology Software Tools, Pittsburgh, PA).

EEG pre-processing and ERP measurement were conducted with the EEGLAB and ERPLAB MATLAB toolboxes (Brunner et al., 2013; Lopez-Calderon & Luck, 2014). We used a 30 Hz low-pass filter (24 dB/octave roll-off) and re-referenced the EEG signal to the average of the

two mastoids. We conducted manual artifact rejection to remove any atypical eye or muscle activity or periods of line or channel noise. Head channels were removed if they exceeded a maximum flatline duration of five seconds, exhibited a channel correlation of lower than 0.6, or exceeded the line noise threshold of four standard deviations. The data were then submitted to independent component analysis (ICA). We removed ICA components associated with (1) eye activity over 70% and brain activity less than 25% or (2) channel noise over 90% and brain activity less than 10% (Number removed: $M = 1.09$, $SD = 0.81$, $Min = 0$, $Max = 3$). Any bad head channels that had been removed before ICA were then interpolated before we time-locked the EEG signal (Number removed: $M = 1.23$, $SD = 1.51$, $Min = 0$, $Max = 6$).

We conducted two time-locking procedures: one to the onset of the video (before the audio) to replicate Martin, Molnar et al.'s (2016) analysis of talker predictability; and one to the onset of the audio, to investigate processing of the speech signal itself. Time-locking to the onset of the word is a typical procedure for ERP research with auditory stimuli (see Kutas & Federmeier, 2011). Baseline correction (–200ms to 0ms) and artifact rejection (peak-to-peak activity exceeding ∓ 60 mV in eye channels or ∓ 100 mV in head channels) were conducted separately for the video and audio ERPs (from 0 to 600ms and from 0 to 1000ms, respectively). Overall, 3.32% of trials for the video analysis and 3.42% of trials for the audio analysis were removed during pre-processing. The data for each channel of interest (13) and trial (420 possible) were then extracted in 2ms increments. The channels of interest were frontal/fronto-central (F3, Fz, F4, FC1, FC2), central/centro-parietal (C3, Cz, C4, CP1, CP2), and parietal (P3, Pz, P4), as in Martin, Molnar et al.

2.6 Data preparation and analysis approach

Data preparation and analysis were conducted with R version 4.2.2 (R Core Team, 2022). Mixed-effects models were fitted to trial-level data with the *lme4* package (Bates et al., 2015). Generalized linear mixed-effects models with a binomial family function were fitted to the behavioral accuracy data, while linear mixed-effects models were fitted to the behavioral RT and ERP data. The fixed and random effects structure for each model is described below. To test the main effects and interactions in each model, Type-III analysis-of-deviance tables were calculated and Wald chi-square tests were conducted with the *car* package (Fox & Weisberg, 2019). To investigate simple effects, estimated marginal means were calculated and pairwise comparisons were conducted with the *emmeans* package (Lenth, 2022). Pairwise *p*-values were adjusted with the Hommel method to control the family-wise error rate (Blakesley et al., 2009).

All models included Talker Accent (Mainstream, Southern, Unpredictable), Helmert contrast-coded, as a fixed effect. The fixed effect of Token Condition (MUSE real word, SUSE real word, Pseudoword), Helmert contrast-coded, was also included in all models except for the pre-audio ERP analysis, where it was not relevant (listeners had not yet heard the actual token). Token Condition combines Token Accent (MUSE, SUSE) and Token Type (real word, pseudoword)

into one factor in order to model a full-rank interaction with Talker Accent. Recall that the monodialectal talkers only produced accent-congruent pseudowords (i.e., Mainstream talkers only produced MUSE pseudowords), while the bidialectal talkers produced pseudowords in both token accents. This approach was taken in order to balance (1) maintaining each actress' Talker Accent throughout the study while (2) achieving an equal number of MUSE and SUSE tokens across conditions; however, this design resulted in an unbalanced three-way interaction among Talker Accent, Token Accent, and Token Type. Using Token Condition allowed us to model the interaction with Talker Accent without rank deficiency.

In addition to the fixed effects described above, models included mean-centered trial as a covariate to control for global changes in response patterns over the course of the experiment. Models also included random intercepts for Item, nested under Talker, and Participant. In addition, the ERP analyses included random intercepts for Channel. Model fitting began with random by-participant slopes for Talker Accent and Token Condition (except for the pre-audio ERP analysis). In the case of non-convergence, singularity, or correlations above 0.95, the random effects structure was simplified iteratively—first by removing the correlation between the slope and intercept estimates, then by removing random slopes but including the correlation parameter—such that the final model for each analysis reflected the maximally-supported structure (Matuschek et al., 2017).

2.6.1 Behavioral data

The lexical decision task yielded both accuracy and reaction time (RT) data. RTs were measured from the onset of the response screen ("Real word?"), which appeared after the audio-visual stimulus had finished playing. This forced a delay in response for participants, which was necessary to avoid introducing ERP components associated with response preparation that overlap in time with our components of interest (Wascher et al., 1996). Any responses with RTs below 50ms or above 2000ms were removed before proceeding with either analysis (5.92% of all responses). For the RT analysis, incorrect responses were removed (9.20% of true responses), then the data were inverse-transformed ($-1000/\text{RT}$) to normalize the distribution. Outliers, defined as 2.5 standard deviations above or below each participant's mean inverse RT, were then removed (2.29% of correct true responses).

2.6.2 ERP data

Before averaging the ERPs in each time window, trials with incorrect behavioral responses were removed. Overall, 9.46% of the pre-processed video data and 9.45% of the pre-processed audio data were removed. On average, there were 364 trials per participant remaining for the video analysis and 363 trials per participant remaining for the audio analysis. For the visual ERPs (pre-speech analysis), average responses were taken in a 225–375ms time window. For the audio

ERPs, average responses were taken in three time windows: 100–200ms, 200–350ms, and 500–800ms. The time windows were chosen based on those in Martin, Molnar et al. (2016), as well as visual inspection of the mean waveforms in each channel across participants and conditions (i.e., without reference to the experimental manipulations).

2.6.3 ERP interpretation

ERP components are distinguished by both timing and polarity. For example, the N400 is named this way because it peaks approximately 400ms after the onset of the stimulus (for visual presentation; the peak is later for auditory presentation, which is why our time window is not centered at 400ms), and this peak has a negative voltage. An N400 response that has a larger negative amplitude than another (e.g., -4 mV vs. -1 mV) is thought to reflect more effortful or less efficient lexical access. In other words, a *stronger* or *larger* N400 effect means that lexical search is more difficult for one stimulus than another. The N1 has the same polarity as the N400 but emerges earlier in the course of processing speech, with a more negative-going N1 waveform indicating that one stimulus is more perceptually salient (or otherwise more attention-grabbing due to stimulus intensity, task demands, etc.) than another. The P2 differs from the N1 and N400 in both polarity and timing, with more positive-going P2 responses reflecting greater salience at more abstract levels of processing (e.g., category structure, emotional valence, etc.). Overall, the timing of differences in mean amplitude between stimuli implies that there are differences in specific linguistic processes between stimuli. Named ERP components like the N400 reflect associations between timing and processing that are well-established in the literature.

3. Results

3.1 Behavioral analysis

3.1.1 Accuracy

There was a significant main effect of Token Condition ($\chi^2(2, N = 3) = 123.26, p < .001$), qualified by a significant interaction with Talker Accent ($\chi^2(4, N = 5) = 14.12, p = .007$), on lexical decision accuracy. The main effect of Token Condition reflected a graded pattern of accuracy descending from MUSE real words ($M = 0.96$, 95% CI [0.95, 0.98]) to SUSE real words ($M = 0.93$, 95% CI [0.90, 0.95]) to pseudowords ($M = 0.87$, 95% CI [0.83, 0.90]). All three token conditions differed significantly from one another: MUSE real words versus SUSE real words ($z = 6.33, p < .001$), MUSE real words versus pseudowords ($z = 11.07, p < .001$), SUSE real words versus pseudowords ($z = 5.91, p < .001$).

To investigate the interaction between Token Condition and Talker Accent, we conducted pairwise comparisons between token conditions within each talker accent. These are shown in **Figure 2**. For Mainstream talkers, accuracy was higher on MUSE real words ($M = 0.96$, 95%

CI [0.95, 0.98]) than on SUSE real words ($M = 0.91$, 95% CI [0.86, 0.94]; $z = 4.92$, $p < .001$) or pseudowords ($M = 0.90$, 95% CI [0.85, 0.93]; $z = 6.11$, $p < .001$), which did not differ from one another. For Southern talkers, accuracy was again higher on MUSE real words ($M = 0.96$, 95% CI [0.94, 0.98]) than on SUSE real words ($M = 0.94$, 95% CI [0.91, 0.96]; $z = 1.99$, $p = .047$) or pseudowords ($M = 0.86$, 95% CI [0.80, 0.90]; $z = 5.82$, $p < .001$). In addition, accuracy on SUSE real words was higher than on pseudowords ($z = 5.95$, $p < .001$). For Unpredictable talkers, accuracy was higher on MUSE real words ($M = 0.97$, 95% CI [0.95, 0.98]) than on SUSE real words ($M = 0.93$, 95% CI [0.90, 0.95]; $z = 4.01$, $p < .001$) or pseudowords ($M = 0.85$, 95% CI [0.79, 0.89]; $z = 7.91$, $p < .001$). Accuracy was also higher on SUSE real words than on pseudowords ($z = 4.95$, $p < .001$).

To better understand the pattern of results within each talker accent, we also compared talker accents within each token condition. For pseudowords, accuracy was higher for Mainstream talkers ($M = 0.90$, 95% CI [0.85, 0.93]) than for Southern talkers ($M = 0.86$, 95% CI [0.80, 0.90]; $z = 2.44$, $p = .029$) or Unpredictable talkers ($M = 0.85$, 95% CI [0.79, 0.89]; $z = 2.47$,

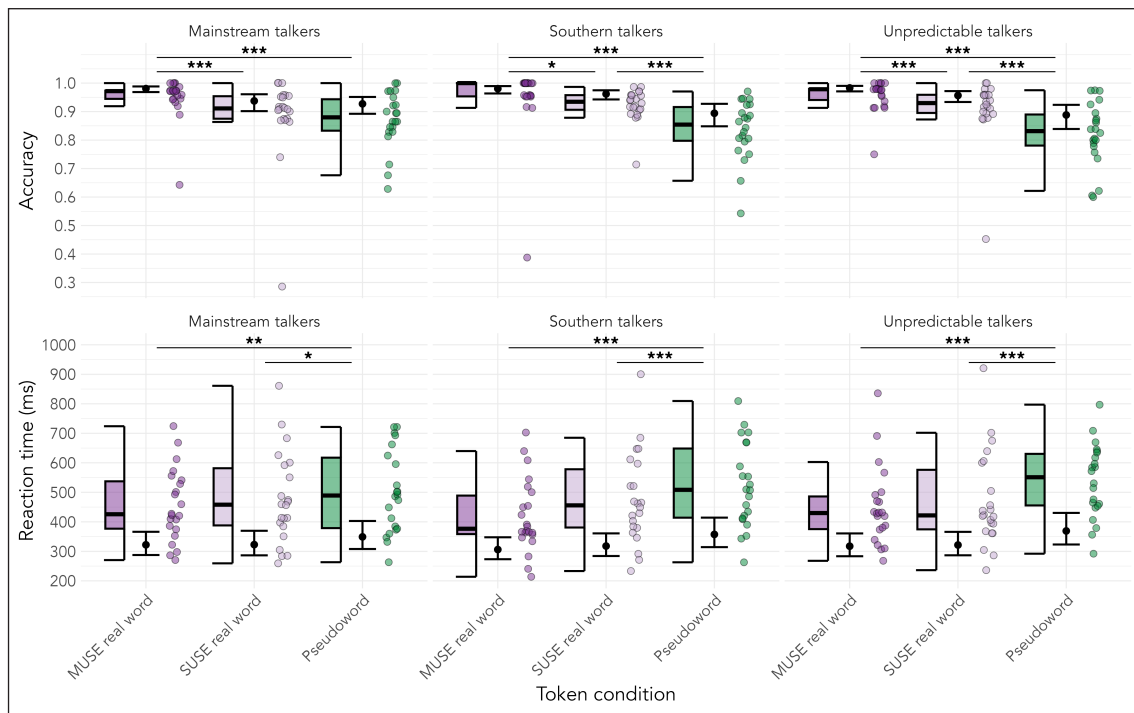


Figure 2: Test task results. Top panel: Boxplots and point plots show distributions of mean accuracy by participant for each combination of Talker Accent and Token Condition. Bottom panel: Boxplots and point plots show distributions of mean reaction times by participant for each combination of Talker Accent and Token Condition. In both panels, black dots represent estimated marginal means with 95% confidence intervals. Significant comparisons within Talker Accent are labeled with asterisks (*** < .001, ** < .01, * < .05).

$p = .027$). In addition, for SUSE real words, accuracy was higher for Southern talkers ($M = 0.94$, 95% CI [0.91, 0.96]) than for Mainstream talkers ($M = 0.91$, 95% CI [0.86, 0.94]; $z = 2.42$, $p = .047$). None of the other contrasts between talker accents within each token condition was significant ($ps > .05$).

3.1.2 Reaction time

There was a main effect of Token Condition ($\chi^2(2, N = 3) = 71.88, p < .001$) on lexical decision RTs. Responses to pseudowords ($M = 358$, 95% CI [316, 413]) were slower than responses to either MUSE real words ($M = 315$, 95% CI [282, 357]; $z = 8.02, p < .001$) or SUSE real words ($M = 321$, 95% CI [287, 364]; $z = 6.77, p < .001$). MUSE and SUSE real words did not differ from one another ($z = 1.24, p = .214$). The interaction between Token Condition and Talker Accent was not significant ($\chi^2(4, N = 5) = 5.65, p = .227$); however, we conducted pairwise comparisons between token conditions within each talker accent to align with the accuracy analysis. The results reflected the main effect of Token Condition, with slower responses to pseudowords than to MUSE or SUSE real word across all three talker accents. These comparisons can be found in Appendix C and are shown in **Figure 2**. We also conducted pairwise comparisons between talker accents within each token condition, but none of the response times differed significantly from one another (see Appendix C).

3.1.3 Behavioral data summary

In terms of accuracy, we observed a consistent MUSE accent advantage for real words that was unaffected by talker identity. By contrast, SUSE real word and pseudoword performance were both impacted by talker identity. While listeners were significantly more accurate on SUSE real words from Southern talkers than from Mainstream talkers, listeners also erroneously accepted more pseudowords from Southern talkers than from Mainstream talkers. In terms of RT, we observed an overall lexicality effect that was not affected by either token accent or talker accent.

We attribute this difference in the effect of token accent between the two behavioral measures to the time course of decision-making. Because we were recording EEG during the lexical decision task, participants were required to wait hundreds of milliseconds after the word ended before making their response. By contrast, in a previous behavioral study with the same stimuli (Walker et al., 2020), participants were able to respond as soon as they had made a decision about lexical status (which could be before the end of the stimulus). The results of that study reflected a MUSE token accent advantage, with slower responses to SUSE real words than to MUSE real words. Comparing the two studies suggests that, while the ultimate decision about lexical status (i.e., accuracy) can survive a delay from stimulus presentation to behavioral response, the subtle differences in decision-making that led to this outcome cannot. In the EEG data, we will see how auditory processing contributes to lexical decision-making.

3.2 ERP analysis

Figures 3 and 4 provide two ways of looking at ERP data. The left panel of **Figure 3** and the top panel of **Figure 4** show grand mean waveforms, or the overall trajectory of neural responses. These waveforms represent the mean voltage every 2ms across trials, electrodes, and participants. The gray boxes behind the waveforms show the time windows in which each analysis was conducted. The right panel of **Figure 3** and the bottom panel of **Figure 4** summarize the values used for each analysis. For example, consider a trial in which Participant A heard the real word *dress* spoken in a SUSE accent by a Mainstream talker. In the N400 time window, voltages were measured in 2ms increments, starting 500ms after the onset of this token and ending 300ms later, for a total of 150 data points per electrode. These data points were then averaged together to yield the N400 response for that particular stimulus. The analyses were performed on these individual values, but the figures show the values summarized by participant.

3.2.1 Video onset (pre-speech analysis)

Before the onset of the auditory token, in the 225–375ms time window post video-onset (in response to the visual presentation of the talker), there was a significant main effect of Talker Accent ($\chi^2(2, N = 3) = 36.90, p < .001$). The statistical comparisons between talker accents are shown in the right panel of **Figure 3**. The ERPs for Southern talkers ($M = -5.40$, 95% CI $[-7.19, -3.62]$) were more negative-going than the ERPs for either Mainstream talkers ($M = -4.98$, 95% CI $[-6.76, -3.19]$; $z = 5.86, p < .001$) or Unpredictable talkers ($M = -4.84$, 95% CI $[-6.65, -3.04]$; $z = 2.58, p = .020$); Mainstream and Unpredictable talkers did not differ from one another ($p > .05$).

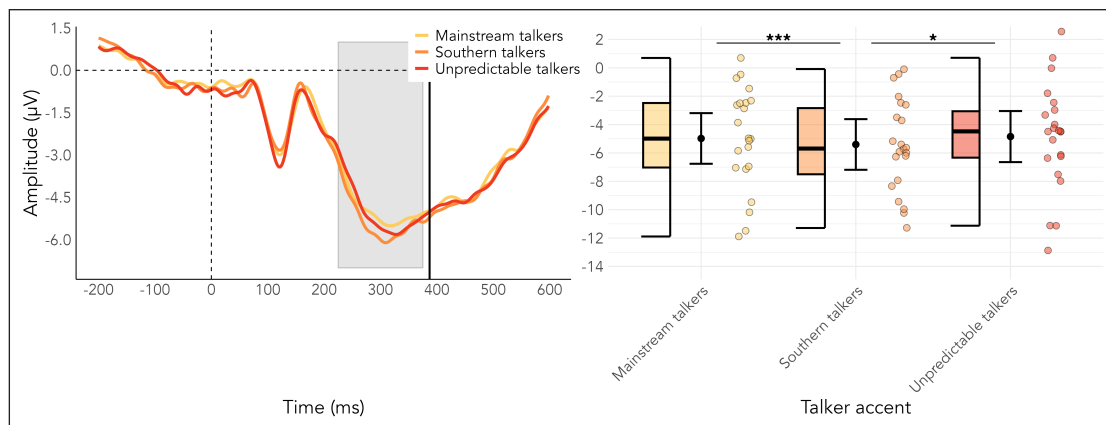


Figure 3: Test task ERPs time-locked to video onset (pre-speech analysis). Left panel: Grand mean waveforms for Talker Accent. Dotted black lines mark plot origin. Solid black line marks mean audio onset. Light grey box marks time window for analysis. Right panel: Boxplots and point plots show distributions of mean amplitude between 225 and 375ms post-video onset by participant for Talker Accent. Black dots represent estimated marginal means with 95% confidence intervals. Significant comparisons are labeled with asterisks ($*** < .001$, $** < .01$, $* < .05$).

3.2.2 Audio onset

3.2.2.1 N1 time window: 100–200ms

Following the onset of the auditory tokens in the 100–200ms time window, we observed a significant interaction between Talker Accent and Token Condition ($\chi^2(4, N = 5) = 97.51, p < .001$). Neither of the main effects was significant ($ps > .05$).

To investigate this interaction, we conducted pairwise comparisons between token conditions within each talker accent. These are shown in the bottom panel of **Figure 4** (top row). For Unpredictable talkers, SUSE real words ($M = 0.99$, 95% CI $[-0.14, 2.11]$) elicited less negative-going waveforms than either MUSE real words ($M = 0.50$, 95% CI $[-0.63, 1.63]$; $z = 4.54, p < .001$) or pseudowords ($M = -0.06$, 95% CI $[-1.29, 1.17]$; $z = 2.69, p = .014$); MUSE real words did not differ from pseudowords ($p > .05$). For Mainstream talkers, pseudowords ($M = 1.19$, 95% CI $[0.04, 2.34]$) elicited less negative-going waveforms than SUSE real words ($M = 0.24$, 95% CI $[-0.96, 1.45]$; $z = 2.43, p = .045$). The responses to MUSE real words ($M = 0.78$, 95% CI $[-0.32, 1.88]$) did not differ from either of the other two token conditions ($ps > .05$). For Southern talkers, no differences between token conditions emerged.

To better understand the relation between Talker Accent and Token Condition, we also conducted pairwise comparisons between talker accents within each token condition. We only observed differences within pseudowords, such that Mainstream talkers ($M = 1.19$, 95% CI $[0.04, 2.34]$) elicited less negative-going waveforms than either Southern talkers ($M = 0.33$, 95% CI $[-0.82, 1.48]$; $z = 6.73, p < .001$) or Unpredictable talkers ($M = -0.06$, 95% CI $[-1.29, 1.17]$; $z = 3.10, p = .004$). Pseudowords from Southern and Unpredictable talkers did not differ from one another, and none of the contrasts within MUSE or SUSE real words was significant ($ps > .05$).

3.2.2.2 P2 time window: 200–350ms

Immediately following the N1 time window, we observed a main effect of Talker Accent ($\chi^2(2, N = 3) = 6.36, p = .042$) that was qualified by an interaction with Token Condition ($\chi^2(4, N = 5) = 35.03, p < .001$). The main effect of Talker Accent reflected a graded pattern of sensitivity from Southern ($M = 4.00$, 95% CI $[2.40, 5.60]$) to Mainstream ($M = 3.82$, 95% CI $[2.22, 5.41]$) to Unpredictable ($M = 3.66$, 95% CI $[2.05, 5.27]$), though none of the pairwise comparisons was significant ($ps > .05$).

To investigate the interaction between Token Condition and Talker Accent, we conducted pairwise comparisons between token conditions within each talker accent. These are shown in the bottom panel of **Figure 4** (middle row). For the P2 time window, we observed a similar pattern of results for Unpredictable talkers as in the N1 time window, with SUSE real words ($M = 4.21$, 95% CI $[2.58, 5.84]$) eliciting more positive-going waveforms than either MUSE real words ($M = 3.63$, 95% CI $[2.00, 5.26]$; $z = 4.89, p < .001$) or pseudowords ($M = 3.15$, 95% CI $[1.44, 4.85]$, $z = 2.60, p = .019$); MUSE real words and pseudowords did not differ. None of the contrasts within Mainstream or Southern talkers was significant ($ps > .05$).

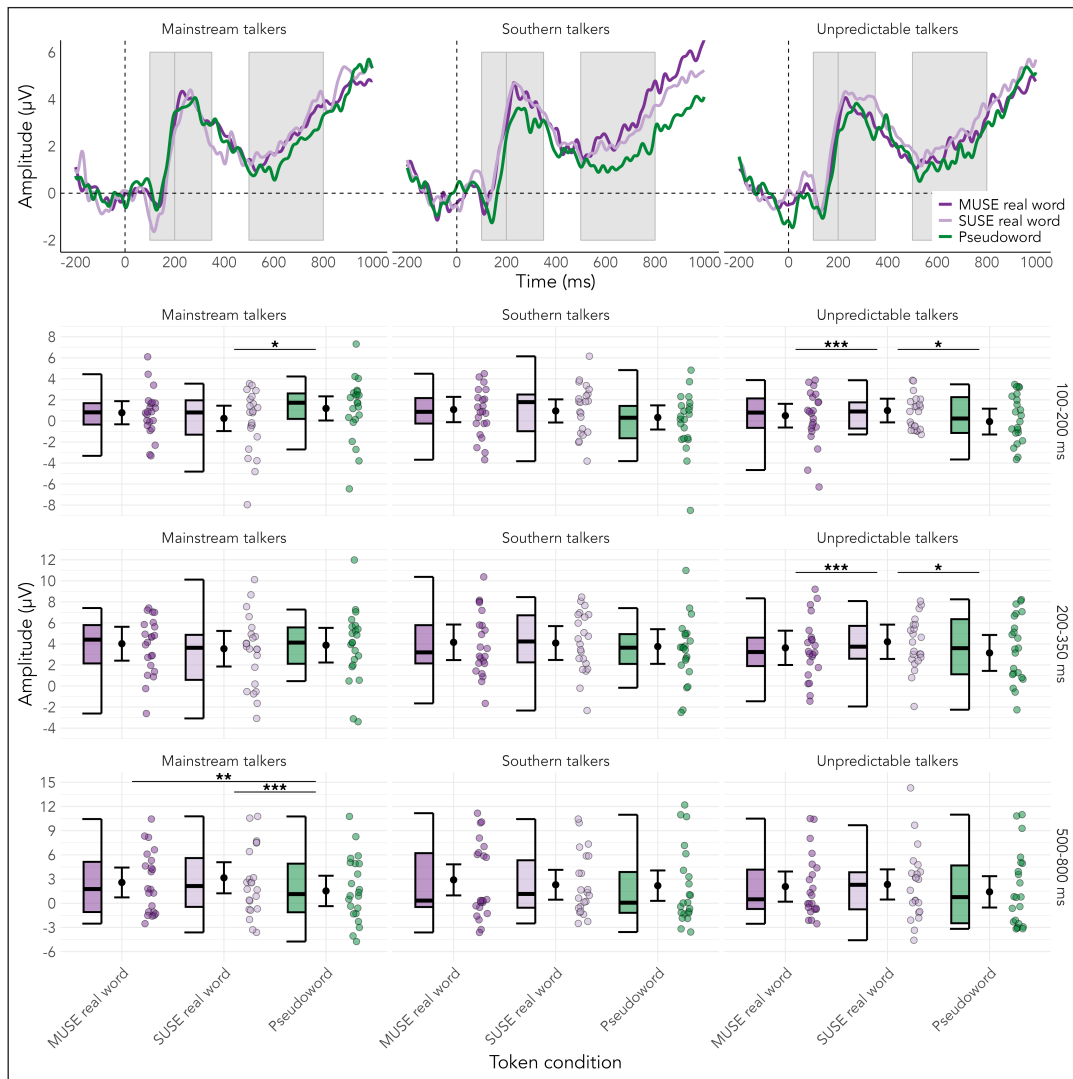


Figure 4: Test task ERPs time-locked to audio onset. Top panel: Grand mean waveforms for Token Condition within Talker Accent. Dotted black lines mark plot origin. Light grey boxes mark time windows for analysis. Bottom panel: Boxplots and point plots show distributions of mean amplitude by participant for each combination of Talker Accent and Token Condition within each time window. Black dots represent estimated marginal means with 95% confidence intervals. Significant comparisons within Talker Accent are labeled with asterisks (***) < .001, ** < .01, * < .05).

We also conducted pairwise comparisons between talker accents within each token condition but did not find any significant differences ($p > .05$).

3.2.2.3 N400 time window: 500–800ms

In this later (500–800ms) time window, there was a main effect of Token Condition ($\chi^2(2, N = 3) = 12.09, p = .002$) qualified by an interaction with Talker Accent ($\chi^2(4, N = 5) = 32.71, p < .001$). The main effect of Token Condition reflected an overall lexicality effect, with pseudowords

($M = 1.71$, 95% CI $[-0.15, 3.57]$) eliciting more negative-going waveforms than either MUSE ($M = 2.52$, 95% CI $[0.68, 4.35]$; $z = 3.16$, $p = .003$) or SUSE ($M = 2.60$, 95% CI $[0.76, 4.43]$; $z = 3.46$, $p = .002$) real words. The two real word conditions did not differ from one another ($p > .05$).

To investigate the interaction between Token Condition and Talker Accent, we conducted pairwise comparisons between token conditions within each talker accent. These are shown in the bottom panel of **Figure 4** (bottom row). For Mainstream talkers we observed a lexicality effect: Pseudowords ($M = 1.53$, 95% CI $[-0.36, 3.41]$) elicited a more negative-going waveform than either MUSE real words ($M = 2.58$, 95% CI $[0.73, 4.43]$; $z = 3.23$, $p = .002$) or SUSE real words ($M = 3.16$, 95% CI $[1.23, 5.09]$; $z = 3.81$, $p < .001$), which did not differ from one another ($p > .05$). None of the contrasts within Southern or Unpredictable talkers was significant ($ps > .05$).

To better understand the relation between Talker Accent and Token Condition, we also conducted pairwise comparisons between talker accents within each token condition. As in the 100–200ms time window, we only observed differences within pseudowords, such that Mainstream talkers ($M = 1.53$, 95% CI $[-0.36, 3.41]$) elicited a more negative-going waveform than Southern talkers ($M = 2.18$, 95% CI $[0.30, 4.07]$; $z = 4.17$, $p < .001$). Neither differed from Unpredictable talkers ($M = 1.41$, 95% CI $[-0.53, 3.36]$; $ps > .05$). None of the contrasts within MUSE or SUSE real words was significant ($ps > .05$).

3.2.3 ERP summary

The two sets of ERP analyses—one time-locked to the visual presentation of the talker (pre-speech analysis), the other to the auditory presentation of the token—illustrated how talker identity is accessed and deployed during online processing.

In response to the visual presentation of the talker (pre-speech), we observed a more negative-going deflection for Southern talkers than for Mainstream or Unpredictable talkers. This enhanced negativity for Southern talkers may reflect our listeners' relative lack of familiarity with the anticipated dialect. We relate this effect to the literature on complex visual processing in the Discussion (see Section 4.2).

In the speech-onset analysis, our strongest predictions related to the N400 effect, which reflects lexico-semantic processing. We expected to observe larger negative-going responses to pseudowords than to real words on the N400 component. In addition, we explored several hypotheses of how accent predictability and familiarity might impact the N400 lexicality effect. Based on Martin, Molnar et al. (2016) and related work (e.g., Brunellière et al., 2009), we also investigated differences in pre-lexical processing before the N400. Together, we analyzed the ERPs in three time windows: N1 (100–200ms), P2 (200–350ms), and N400 (500–800ms).

In the first time window, for Mainstream talkers, we observed stronger N1 responses to SUSE real words compared to (MUSE) pseudowords. This N1 effect may reflect early detection of the incongruity between a SUSE token accent and a Mainstream talker identity. By contrast, for

Unpredictable talkers, we observed attenuated N1 responses to SUSE real words compared to both MUSE real words and (mixed) pseudowords; however, this effect was paired with a larger P2 response. These components reflect the salience of auditory information at different levels of processing, with the N1 being sensitive to low-level acoustic information and the P2 being sensitive to higher-level phonetic information. This pattern of results may reflect the in-between status of Unpredictable talkers, with MUSE accents being more salient in the initial stages of auditory processing (N1) and SUSE accents becoming more salient downstream (P2).

In the last time window, pseudowords elicited larger N400 responses than real words for Mainstream talkers, which reflects the expected effect of lexical status. Critically, this effect was observed regardless of whether the real words were spoken with a MUSE or a SUSE accent. This suggests that when listeners *expect* to hear a MUSE accent, the ease of lexical access is not affected by the actual token accent. While the familiar, predictable Mainstream talkers elicited typical N400 lexicality effects, the Southern talkers and Unpredictable talkers did not. There were no significant differences between real word and pseudoword responses for these talkers, regardless of token accent. Overall, the N400 results suggest that lexical access is more strongly influenced by a listener's expectations about a talker than by the actual token accent. These challenges in lexical access with unfamiliar and unpredictable talkers may have resulted from greater lexical competition due to dialectal ambiguity (Carasco-Ortiz, 2017; Meade et al., 2019). We relate these online results to the offline data in the Discussion (see Section 4.1).

4. Discussion

We investigated the effects of familiarity and predictability on cross-dialectal speech recognition, combining measures of comprehension accuracy (lexical decision) and online processing as the stimulus unfolds in real time (ERPs). We introduced a group of listeners from western Pennsylvania to six talkers with three different accents: Mainstream, which was both familiar to our listeners and predictable throughout the experiment; Southern, which was less familiar but also predictable; and Unpredictable, which switched between the familiar MUSE and less familiar SUSE accents. These talkers produced MUSE-accented real words, SUSE-accented real words, and pseudowords in an audio-visual lexical decision task while EEG was recorded. This combination of neural and behavioral measures revealed different talker vs. accent familiarity effects depending on the stage of processing. Early in processing, we observed the strongest effects of talker identity, with Mainstream talkers eliciting N400 lexicality effects regardless of token accent. At the endpoint of processing, we observed the strongest effects of token accent, with the highest accuracy on MUSE real words regardless of talker identity. Our results suggest that cross-dialectal communication contexts promote an uncertain listening strategy.

4.1 The effects of talker identity on lexical processing

Previous behavioral work shows that Northern listeners have a MUSE versus SUSE advantage (e.g., Clopper & Bradlow, 2008; Walker et al., 2018b). There is also a larger body of work showing that listeners process familiar regional accents more easily than less familiar regional accents (e.g., Floccia et al., 2006; Labov & Ash, 1997). In the present study, we observed evidence for this MUSE token accent advantage only in behavioral responses. Lexical decision performance was most sensitive to token accent, with significantly higher accuracy on MUSE real words compared to SUSE real words, even for Southern talkers. By contrast, RTs were only sensitive to lexical status, with faster responses to real words than to pseudowords, regardless of talker identity or token accent. The auditory ERP analyses were also relatively insensitive to token accent. During early auditory processing (N1), SUSE real words were more salient than (MUSE) pseudowords when produced by Mainstream talkers. For Unpredictable talkers, SUSE real words were less perceptually salient than MUSE real words or (mixed) pseudowords (N1) but caught listeners' attention later during auditory processing (P2). Lexico-semantic processing was also conditioned on talker identity, with larger N400 responses to pseudowords compared to real words only for Mainstream talkers. Overall, our results only support a MUSE *token* advantage in the very latest stage of lexical processing (the lexical decision itself) and instead see evidence of a Mainstream *talker* advantage early on.

In the Introduction, we laid out three possible ways in which contextual cueing of talker identity could impact word recognition of less familiar dialects: (1) token accent may impact processing more than context; (2) context may facilitate processing by promoting dialect-specific listening strategies; (3) context may disrupt (or delay) processing by promoting an uncertain listening strategy. Our results align with the third option, with listeners adopting an uncertain listening strategy in the presence of Southern and Unpredictable talkers. We believe this was illustrated in two places. First, the absence of the N400 lexicality effect for these talkers suggests that listeners are entertaining more competitors. In **Table 1**, we said that if listeners simply maintained a MUSE advantage in the context of these talkers, we would expect to see a difference between SUSE and MUSE real words: We do not. We also said that if listeners were shifting their expectations to dialect-specific processing, we would see talker-accent congruency effects: We do not. Instead, what we observe in the N400 time window is that expectations about the talker matter, such that Southern and Unpredictable talkers broaden the possible mappings that listeners entertain.

Lexical decision accuracy further illustrates this uncertain listening strategy for unfamiliar dialects. We found that listeners were significantly more accurate on SUSE real words from Southern talkers than from Mainstream talkers (with SUSE words from Unpredictable talkers non-significantly falling in between). At first glance, we might take this to mean that our

participants *did* activate Southern-accented mappings to some degree when seeing a Southern talker. However, we also found that a) listeners performed similarly well on MUSE real words across talkers and b) listeners made more mistakes on pseudowords from Southern and Unpredictable talkers compared to Mainstream talkers. Therefore, the higher accuracy for SUSE real words spoken by Southern talkers was not simply about shifting to a more SUSE-like perceptual system. If that had been the case, listeners should have had an easier time rejecting SUSE pseudowords and struggled more with MUSE real words. Rather, these results suggest that listeners broadened the signal-to-word mappings they considered from these talkers (Babel et al., 2021; Clopper & Walker, 2017; McMurray et al., 2019; Weatherholtz, 2015; Zheng & Samuel, 2020). This strategy was evident whether that challenge came from a talker who was overtly unpredictable (Unpredictable talker), or who simply had a predictably less familiar accent (Southern talker). While this uncertain listening strategy resulted in poorer lexical decision performance, this could be an optimal strategy in real-world situations where talkers are not likely to utter pseudowords. In the face of uncertainty, hearing tokens as real words is likely to result in successful comprehension (Ganong, 1980).

The transition from Mainstream talker advantage in the ERP data to MUSE accent advantage in the accuracy data highlights how contextual information shapes the time course of comprehension. During acoustic-phonetic processing, we observed sensitivity to SUSE-accented tokens on the N1 and P2 for Mainstream and Unpredictable talkers but not for Southern talkers. These early effects of acoustic-phonetic salience only transitioned to robust N400 effects for Mainstream talkers during lexico-semantic processing, with differences between real words and pseudowords that were insensitive to token accent. Southern and Unpredictable talkers did *not* elicit strong lexicality effects, suggesting that listeners were entertaining more competitors for these talkers. This uncertain listening strategy made listeners more likely to endorse SUSE real words and pseudowords as real words when they were produced by Southern talkers than when they were produced by Mainstream talkers. However, listeners were overall more accurate on MUSE real words than on SUSE real words regardless of the talker. Together, these differences in accuracy show how relying on talker identity during online processing led to less accent-specific mappings for SUSE tokens offline. We were able to observe this trade-off not only by disrupting the relation between talker identity and token accent, which have been perfectly aligned in previous work (e.g., Goslin et al., 2012), but also by measuring both online (EEG/ERPs) and offline (reaction time, accuracy) processing. Future work on cross-dialectal communication might also consider combining neurocognitive and behavioral methods, as well as pursuing other ways of disentangling bottom-up speech perception from top-down expectations.

So far we have framed the difference between the two accents in this study—MUSE and SUSE—as primarily being about the familiarity our western Pennsylvania listeners had with these accents. However, given that SUSE is also a more stigmatized variety (Hasty, 2018; Preston 2018), the two dialects also differ in overt prestige. Sumner and colleagues have argued that

listeners encode prestigious dialects with higher social weight than less prestigious dialects, and that the benefit of familiarity cannot outweigh prestige (Clapp et al., 2023; Sumner & Kataoka, 2013; Sumner et al., 2014; see also Clopper & Bradlow, 2008; Evans & Iverson, 2007; Maher et al., under review; Zaharchuk et al., 2021). This means that the mental representations of SUSE and MUSE *before* the experiment and the responses to these accents *during* the experiment may have been shaped primarily by their relative prestige in the United States, rather than their relative familiarity to participants in the study. We cannot disentangle the issue of social status in this particular study; minimally, participants who frequently hear or use SUSE would be needed, which we are pursuing in other work. For now, however, we note that regardless of *why* the representations of SUSE and MUSE differed for our listeners, both prestige- and exposure-centric accounts would fundamentally predict the same thing—less robust SUSE representations—and what we ultimately argue here is that a weaker representation appears to lead to a less committed listening style.

4.2 The effects of talker identity on pre-speech processing

We also observed differences in talker identity in the pre-speech ERP analysis. The visual presentation of the Southern talkers elicited a negative-going deflection compared to the visual presentation of the Mainstream and Unpredictable talkers. Our enhanced negativity in the 225–375ms window to the visual presentation of unfamiliar Southern talkers is reminiscent of the enhanced negativity, peaking around 300ms after the onset of a visual stimulus, reported in the literature on processing complex visual stimuli (termed N300; e.g., Kumar et al., 2021). Taken to index predictive coding of complex visual objects and scenes, in particular the final stages of processing complex visual stimuli, this component has been shown to be sensitive to how well an exemplar fits into a category. Extending this reasoning to our data, the enhanced negativity for the presentation of the Southern talkers suggests that, to our listeners, the talkers who are designated as “Southern” are not as easily assigned to a talker identity category. Regardless of the specific interpretation, this effect suggests that talker identity can impact speech processing before speech even begins, and could be related to the well-established impacts of perceived talker ethnicity on speech perception performance (e.g., Kutlu et al., 2022; McGowan, 2015).

4.3 Comparison with Martin, Molnar, and Carreiras (2016)

Our study applied the experimental design from Martin, Molnar et al. (2016) to a bidialectal language context. Their study investigated the impact of talker identity on *bilingual* language processing by presenting bilingual Spanish-Basque listeners with monolingual Spanish talkers, monolingual Basque talkers, and bilingual Spanish-Basque talkers, whose bilingual identity was signaled by code-switching between Spanish and Basque. Our study investigated the impact of talker identity on *cross-dialectal* language processing by presenting our listeners with monodialectal MUSE talkers, monodialectal SUSE talkers, and bidialectal talkers. These two

changes—from bilingual to bidialectal talkers and from bilingual to monodialectal listeners—resulted in different patterns of neural responses. In the pre-audio analysis, we both observed effects of talker identity; however, they were differentially impacted by predictability and familiarity. Specifically, Martin, Molnar et al. observed a positive-going deflection for the bilingual talker, which was interpreted as a marker of uncertainty in language prediction, while we observed a negative-going deflection for the Southern talkers. Together, these patterns of sensitivity to talker identity appear to be driven by a lack of predictability; in our case, from a lack of familiarity with Southern varieties.

The post-speech analyses diverged even further. On the N1, Martin, Molnar et al. (2016) observed an interaction between lexical status and talker identity, with pseudowords eliciting larger effects than real words for monolingual talkers but not for bilingual talkers. We did not observe such a lexicality effect on either the N1 or P2; rather, our early auditory effects reflected the perceptual salience of particular talker-token accent pairs. However, we did observe differences in real word and pseudoword processing on the N400 for the Mainstream talkers. Our N400 lexicality effect for Mainstream talkers is analogous to their N1 lexicality effect for monolingual talkers, in that lexical status was more easily detected when produced by predictable talkers; however, the difference in the timing of these effects shows that our listeners were unable to predict lexical status as early as their listeners.

Martin, Molnar et al. (2016) also observed typical N400 lexicality effects for both monolingual and bilingual talkers. While the lexicality effects were attenuated for the bilingual talkers, we failed to observe any lexicality effects at all for either our Southern or Unpredictable talkers. This comparison further highlights the uncertain listening strategy adopted by our participants. While their listeners were able to engage typical lexico-semantic processing mechanisms for both predictable (monolingual) and unpredictable (bilingual) talkers, our listeners engaged relatively shallow processing for unfamiliar or otherwise unpredictable talkers. These differences in lexical processing between their study and ours may reflect differences in processing languages versus dialects, the high level of proficiency and familiarity of their bilingual listeners with both varieties (versus the relatively low level of familiarity of our listeners with Southern accents), or the relative naturalness of switching between two different languages in a bilingual community versus categorical shifts between two different dialects (see below).

4.4 Limitations and future directions

We would like to highlight two aspects of our experimental design that could be the focus of future work. First, as mentioned in Section 2.2.2, our Unpredictable talkers exhibited extreme shifts between dialects in their monologues that are atypical for bidialectal speakers. We operationalized dialect switching this way as a direct parallel of the language switching condition in Martin, Molnar et al. (2016). The goal was to manipulate the predictability of the upcoming signal through talker identity; however, the social strangeness of the Unpredictable

talkers' patterns of speech may have exaggerated the effects of uncertainty in cross-dialectal communication. It is an open question whether listeners see more naturally occurring patterns of style-shifting as equally unpredictable; in other words, future studies may investigate to what degree, and on what dimensions, variation from a single talker is expected.

Second, our study implemented a mixed rather than blocked design, such that, trial-to-trial, participants did not know which talker or accent they would hear next. Apart from following Martin, Molnar et al. (2016), this decision allowed us to investigate the immediate and non-habitualized effect of talker type and token accent on speech perception. It possibly also pushed listeners into more categorical versus talker-specific approaches to processing (cf. Clapp et al., 2023, p. 13), which would mean our listeners were labelling the talkers as “Southern” and “non-Southern,” and using more general-Southern vs. talker-specific knowledge in processing (which is ultimately how we have interpreted our results). However, since previous work has shown that language processing can be different in experiments which are blocked by talker or by dialect and those that are not (Clapp et al., 2023), future work could investigate cross-dialectal processing when blocked by talker or by talker accent. We might expect overall accuracy to improve by facilitating talker-specific adaptation (see Luthra, 2024, for a review of talker variability); in addition, we may see evidence that western Pennsylvania listeners *can* form more dialect-specific expectations for Southern talkers when trial-to-trial uncertainty is reduced. Such a finding would suggest that the uncertain listening strategy we see in the current study is related to the high uncertainty of each trial rather than the predictability of the talker's accent.

5. Conclusion

In this study, we combined behavioral and neurocognitive measures to investigate the effects of familiarity and predictability on cross-dialectal speech perception. In behavior, we found that non-Southern U.S. listeners performed worse on SUSE-accented real words than on MUSE-accented real words in a lexical decision task, even when the talker had a Southern accent identity. While expectations of a Southern accent did increase accuracy on SUSE-accented real words compared to expectations of a Mainstream accent, this pattern was reversed for pseudowords. In our ERP analyses, an enhanced negativity for Southern talkers pre-audio was not followed by any lexicality effects in response to the real words and pseudowords; by contrast, Mainstream talkers elicited canonical N400 effects irrespective of token accent. Together, these results suggest that listeners considered a broader range of acceptable mappings from Southern talkers than for Mainstream talkers. While we find some similarities between listener responses to a predictable talker with a less familiar accent and an unpredictable-accented talker, there is evidence from early brain responses that listeners engaged different auditory processing mechanisms for these two types of talkers. This work shows the value of pairing behavioral measures with neurophysiological approaches to better understand the mechanisms underlying (cross-)dialectal processing.

Appendix

Appendix A. Acoustic analysis of actresses' vowels

In Figure A.1, we present average trajectories of six vowels (DRESS was not used in the present task, but we include it as an anchor for interpreting FACE) for the actors in their MUSE and SUSE performance (top row). The actresses make substantial changes across guises: in their SUSE guise, FACE and THOUGHT become more diphthongal, PRIZE becomes more monophthongal, and DRESS and KIT raise. In the second row, we show comparative vowel plots for five older women from southwest Virginia who demonstrate the Southern Vowel Shift (collected by Caleigh Hampton, taken from a mix of interview and citation speech), and four self-identified bidialectal younger southwest Virginia talkers, producing single words in a “Southern” guise (cf. Walker et al. 2018; Walker 2020). The actors have larger vowel spaces than the other two groups in both their guises. Apart from that, they look most similar to the older talkers, reflecting a more conservative Southern accent target.

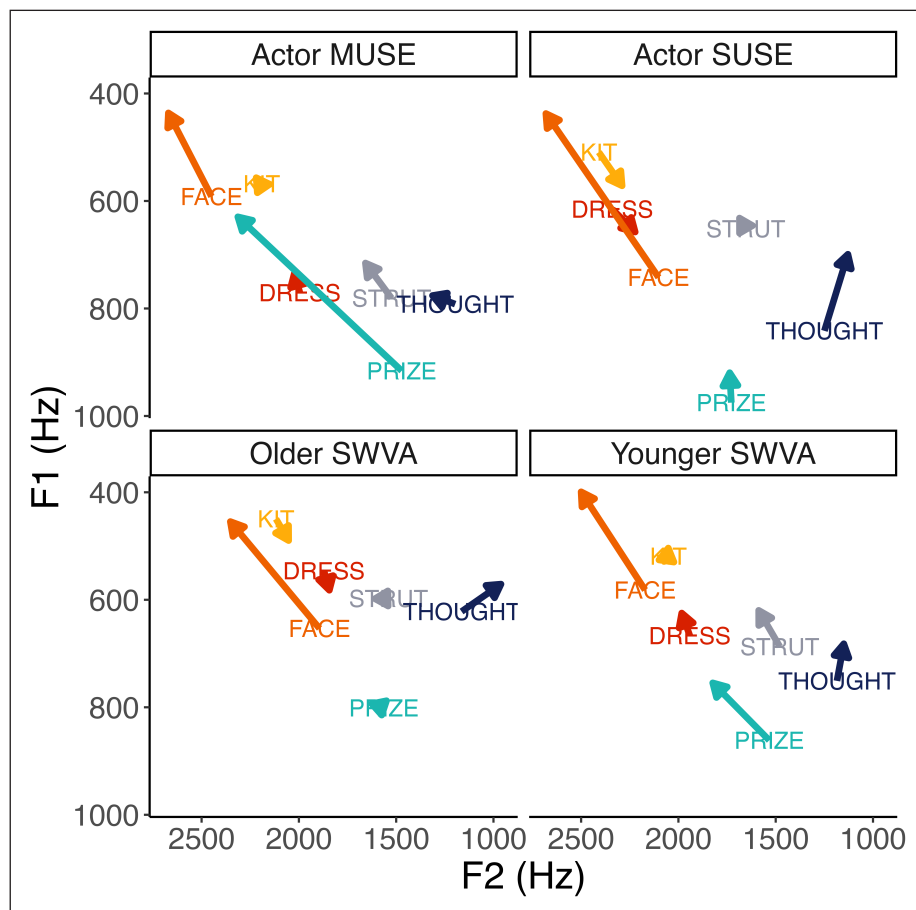


Figure A.1. Vowel plots of actors across the two guises (top row), compared to women actually from SWVA (bottom row).

Appendix B. Analysis of word length

We compared the duration of each item in the lexical decision task by Token Accent with the following linear mixed-effects model (the default treatment contrast codes for Token Accent were used, with Southern as the reference level): Token duration $\sim$ Token Accent + (Token Accent | Talker) + (Token Accent | Item). As Table B.1 shows, the difference in token duration between Token Accents was not significant.

Fixed effect	Estimate	SE	df	<i>t</i>	<i>p</i>
(Intercept)	615.57	29.93	5.21	20.57	< .001
Token Accent	-10.03	7.78	5.10	-1.32	.241

Table B.1. Model coefficients for the fixed effect of Token Accent on token duration (Satterthwaite approximation used to calculate df and *p* values). N = 5040.

Appendix C. Analysis of reaction time data

Factor	Level	Contrast level 1	Contrast level 2	Mean 1 [95% CI]	Mean 2 [95% CI]	Contrast
Talker accent	MUSE	Standard real word	Southern real word	322.12 [287.49, 366.22]	322.75 [286.41, 369.64]	$z = -0.07$, $p = .943$
		Standard real word	Pseudoword	322.12 [287.49, 366.22]	349.03 [307.90, 402.84]	$z = -3.36$, $p = .002$
		Southern real word	Pseudoword	322.75 [286.41, 369.64]	349.03 [307.90, 402.84]	$z = -2.47$, $p = .027$
	SUSE	Standard real word	Southern real word	306.01 [273.29, 347.64]	317.91 [284.11, 360.83]	$z = -1.48$, $p = .140$
		Standard real word	Pseudoword	306.01 [273.29, 347.64]	357.37 [314.24, 414.22]	$z = -5.00$, $p < .001$
		Southern real word	Pseudoword	317.91 [284.11, 360.83]	357.37 [314.24, 414.22]	$z = -4.72$, $p < .001$
	Unpredictable	Standard real word	Southern real word	317.39 [283.36, 360.70]	321.44 [286.57, 365.98]	$z = -0.57$, $p = .570$

(Contd.)

Factor	Level	Contrast level 1	Contrast level 2	Mean 1 [95% CI]	Mean 2 [95% CI]	Contrast
		Standard real word	Pseudoword	317.39 [283.36, 360.70]	369.09 [323.16, 430.24]	$z = -5.51$, $p < .001$
		Southern real word	Pseudoword	321.44 [286.57, 365.98]	369.09 [323.16, 430.24]	$z = -4.99$, $p < .001$
Token condition	Standard real word	MUSE	SUSE	322.12 [287.49, 366.22]	306.01 [273.29, 347.64]	$z = 1.98$, $p = .143$
		MUSE	Unpredictable	322.12 [287.49, 366.22]	317.39 [283.36, 360.70]	$z = 0.71$, $p = .475$
		SUSE	Unpredictable	306.01 [273.29, 347.64]	317.39 [283.36, 360.70]	$z = -1.34$, $p = .362$
	Southern real word	MUSE	SUSE	322.75 [286.41, 369.64]	317.91 [284.11, 360.83]	$z = 0.55$, $p = .889$
		MUSE	Unpredictable	322.75 [286.41, 369.64]	321.44 [286.57, 365.98]	$z = 0.14$, $p = .889$
		SUSE	Unpredictable	317.91 [284.11, 360.83]	321.44 [286.57, 365.98]	$z = -0.53$, $p = .889$
	Pseudoword	MUSE	SUSE	349.03 [307.90, 402.84]	357.37 [314.24, 414.22]	$z = -0.81$, $p = .418$
		MUSE	Unpredictable	349.03 [307.90, 402.84]	369.09 [323.16, 430.24]	$z = -1.82$, $p = .205$
		SUSE	Unpredictable	357.37 [314.24, 414.22]	369.09 [323.16, 430.24]	$z = -1.02$, $p = .418$

Table C.1. Reaction time means, 95% confidence intervals, z statistics, and p values for comparisons between each level of Token Condition within each level of Talker Accent (values plotted in bottom panel of **Figure 2**) and between each level of Talker Accent within each level of Token Condition.

Acknowledgements

This project started with a grant from Virginia Tech's Institute for Society, Culture and Environment (ISCE) to the second author and Dr. Mike Bowers, who passed away in 2021. We are grateful to Mike for his enthusiasm, encouragement and insights. The project was later supported by an NSF collaborative grant to Walker and Van Hell (BCS-2041264 and BCS-2041081) and BCS-2234907 to Zaharchuk and Van Hell. Thanks to research assistants Sherree Ann Shuler, Adam Bowen, Jessie Yu and Paloma Barongan, and to our actors and participants. This paper is part of a special issue based on LabPhon 19, and we are grateful to the guest editors Taehong Cho, Jeff Holliday and especially Sang-Im Lee-Kim, and to our reviewers.

Competing interests

The authors have no competing interests to declare.

Authors' contributions

H.Z. – data collection; data analysis (lead); data processing; writing (co-lead); review and editing

A.W. – funding; conceptualization (co-lead); design; experiment implementation; writing (co-lead); data analysis; review and editing

A.M. – data analysis; data processing; review and editing

C.F. – design; experiment implementation (lead); data collection;

J.VH. – funding; conceptualization (co-lead); design; writing; data analysis; review and editing; supervision

References

Babel, M., Johnson, K., & Sen, C. (2021). Asymmetries in perceptual adjustments to non-canonical pronunciations. *Laboratory Phonology*, 12(1), 1–43. <https://doi.org/10.16995/labphon.6442>

Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, 67(1), 1–48. <https://doi.org/10.18637/jss.v067.i01>

Bissell, M., & Clopper, C. (2025). The effect of listener dialect experience on perceptual adaptation to and generalization of a novel vowel shift. *Laboratory Phonology*, 16(1), 1–24. <https://doi.org/10.16995/labphon.11588>

Blakesley, R. E., Mazumdar, S., Dew, M. A., Houck, P. R., Tang, G., Reynolds III, C. F., & Butters, M. A. (2009). Comparisons of methods for multiple hypothesis testing in neuropsychological research. *Neuropsychology*, 23(2), 255–264. <https://doi.org/10.1037/a0012850>

Brouwer, S., Mitterer, H., & Huettig, F. (2012). Speech reductions change the dynamics of competition during spoken word recognition. *Language and Cognitive Processes*, 27(4), 539–571. <https://doi.org/10.1080/01690965.2011.555268>

- Brunellière, A., Dufour, S., Nguyen, N., & Frauenfelder, U. H. (2009). Behavioral and electrophysiological evidence for the impact of regional variation on phoneme perception. *Cognition*, 111(3), 390–396.
- Brunellière, A., & Soto-Faraco, S. (2013). The speakers' accent shapes the listeners' phonological predictions during speech perception. *Brain and Language*, 125(1), 82–93. <https://doi.org/10.1016/j.bandl.2013.01.007>
- Brunellière, A., & Soto-Faraco, S. (2015). The interplay between semantic and phonological constraints during spoken-word comprehension. *Psychophysiology*, 52(1), 46–58. <https://doi.org/10.1111/psyp.12285>
- Brunner, C., Delorme, A., & Makeig, S. (2013). EEGLAB—An open source MATLAB toolbox for electrophysiological research. *Biomedical Engineering/Biomedizinische Technik*, 58(SI-1-Track-G), 000010151520134182. <https://doi.org/10.1515/bmt-2013-4182>
- Carrasco-Ortiz, H., Midgley, K. J., Grainger, J., & Holcomb, P. J. (2017). Interactions in the neighborhood: Effects of orthographic and phonological neighbors on N400 amplitude. *Journal of Neurolinguistics*, 41, 1–10. <https://doi.org/10.1016/j.jneuroling.2016.06.007>
- Čeponienė, R., Alku, P., Westerfield, M., Torki, M., & Townsend, J. (2005). ERPs differentiate syllable and nonphonetic sound processing in children and adults. *Psychophysiology*, 42(4), 391–406. <https://doi.org/10.1111/j.1469-8986.2005.00305.x>
- Chwilla, D. J., Brown, C. M., & Hagoort, P. (1995). The N400 as a function of the level of processing. *Psychophysiology*, 32(3), 274–285. <https://doi.org/10.1111/j.1469-8986.1995.tb02956.x>
- Clapp, W., Vaughn, C., & Sumner, M. (2023). The episodic encoding of talker voice attributes across diverse voices. *Journal of Memory and Language*, 128, 104376. <https://doi.org/10.1016/j.jml.2022.104376>
- Clopper, C. G., & Bradlow, A. R. (2008). Perception of dialect variation in noise: Intelligibility and classification. *Language and Speech*, 51(3), 175–198. <https://doi.org/10.1177/0023830908098539>
- Clopper, C. G., & Smiljanic, R. (2015). Regional variation in temporal organization in American English. *Journal of Phonetics*, 49, 1–15. <https://doi.org/10.1016/j.wocn.2014.10.002>
- Clopper, C. G., & Walker, A. (2017). Effects of lexical competition and dialect exposure on phonological priming. *Language and Speech*, 60(1), 85–109. <https://doi.org/10.1177/0023830916643737>
- D'Onofrio, A. (2015). Persona-based information shapes linguistic perception: Valley Girls and California vowels. *Journal of Sociolinguistics*, 19, 241–256. <https://doi.org/10.1111/josl.12115>
- Evans, B. G., & Iverson, P., 2007. Plasticity in vowel perception and production: A study of accent change in young adults. *Journal of the Acoustical Society of America*, 121(6), 3814–3826.
- Floccia, C., Goslin, J., Girard, F., & Konopczynski, G. (2006). Does a regional accent perturb speech processing? *Journal of Experimental Psychology: Human Perception and Performance*, 32(5), 1276–1293. <https://doi.org/10.1037/0096-1523.32.5.1276>
- Fox, J., & Weisberg, S., (2019). *An R companion to applied regression*, 3rd ed. Sage. <https://socialsciences.mcmaster.ca/jfox/Books/Companion/>

- Ganong, W. F. (1980). Phonetic categorization in auditory word perception. *Journal of Experimental Psychology: Human Perception and Performance*, 6(1), 110–125. <https://doi.org/10.1037/0096-1523.6.1.110>
- Garcia, F. M., Shen, G., Avery, T., Green, H. L., Godoy, P., Khamis, R., & Froud, K. (2022). Bidialectal and monodialectal speech processing: Neurophysiological evidence from mismatch negativity. *Journal of Communication Disorders*, 100, 106267. <https://doi.org/10.1016/j.jcomdis.2022.106267>
- Getz, L. M., & Toscano, J. C. (2021). The time-course of speech perception revealed by temporally-sensitive neural measures. *Wiley Interdisciplinary Reviews: CognitiveScience*, 12(2), e1541. <https://doi.org/10.1002/wcs.1541>
- Goslin, J., Duffy, H., & Floccia, C. (2012). An ERP investigation of regional and foreign accent processing. *Brain and Language*, 122(2), 92–102. <https://doi.org/10.1016/j.bandl.2012.04.017>
- Gunter, K., Vaughn, C., & Kendall, T. (2020). Perceiving Southernness: Vowel categories and acoustic cues in Southernness ratings. *Journal of the Acoustical Society of America*, 147(1), 643–656. <https://doi.org/10.1121/10.0000550>
- Hasty, J. D. (2018). They sound better than we do: Language attitudes in Alabama. In T. E. Nunnally (Ed.), *Speaking of Alabama: The history, diversity, function, and change of language* (pp. 192–200). University of Alabama Press.
- Hay, J., & Drager, K. (2010). Stuffed toys and speech perception. *Linguistics*, 48(4), 865–892. <https://doi.org/10.1515/ling.2010.027>
- Hazen, K. (2001). An introductory investigation into bidialectalism. *Penn Working Papers in Linguistics: Selected Papers from NWAV*, 29, 85–99.
- Helenius, P., Salmelin, R., Richardson, U., Leinonen, S., & Lyytinen, H. (2002). Abnormal auditory cortical activation in dyslexia 100 msec after speech onset. *Journal of Cognitive Neuroscience*, 14(4), 603–617. <https://doi.org/10.1162/08989290260045846>
- Holcomb, P. J., Grainger, J., & O'Rourke, T. (2002). An electrophysiological study of the effects of orthographic neighborhood size on printed word perception. *Journal of Cognitive Neuroscience*, 14(6), 938–950. <https://doi.org/10.1162/089892902760191153>
- Hurring, G., Hay, J., Drager, K., Podlubny, R., Manhire, L., & Ellis, A. (2022). Social priming in speech perception: Revisiting kangaroo/kiwi priming in New Zealand English. *Brain Sciences*, 12(6), 684. <https://doi.org/10.3390/brainsci12060684>
- Jacewicz, E., Fox, R. A., & Salmons, J. (2007). Vowel duration in three American English dialects. *American Speech*, 82(4), 367–385. <https://doi.org/10.1215/00031283-2007-024>
- Johnson, K., Strand, E. A., & D'Imperio, M. (1999). Auditory–visual integration of talker gender in vowel perception. *Journal of Phonetics*, 27(4), 359–384. <https://doi.org/10.1006/jpho.1999.0100>
- Kapnoula, E. C., & McMurray, B. (2021). Idiosyncratic use of bottom-up and top-down information leads to differences in speech perception flexibility: Converging evidence from ERPs and eye-tracking. *Brain and Language*, 223, 105031. <https://doi.org/10.1016/j.bandl.2021.105031>
- Kleinschmidt, D. F., & Jaeger, T. F. (2015). Robust speech perception: Recognize the familiar, generalize to the similar, and adapt to the novel. *Psychological Review*, 122(2), 148–203. <https://doi.org/10.1037/a0038695>

- Kumar, M., Federmeier, K. D., Beck, D. M. (2021). The N300: An index for predictive coding of complex visual objects and scenes. *Cerebral Cortex Communications*, 2(2), tgab030. <https://doi.org/10.1093/texcom/tgab030>
- Kutas, M., Federmeier, K. (2011) Thirty years and counting: Finding meaning in the N400 component of the event-related brain potential (ERP). *Annual Review of Psychology*, 62, 621–647. <https://doi.org/10.1146/annurev.psych.093008.131123>
- Kutlu, E., Tiv, M., Wulff, S., & Titone, D. (2022). Does race impact speech perception? An account of accented speech in two different multilingual locales. *Cognitive Research: Principles and Implications*, 7(1), 7.
- Labov, W. (1973). Where do grammars stop? In R. Shuy (Ed.), *Report of the Twenty-Third Annual Round Table Meeting on Linguistics and Language Studies* (pp. 43–88). Georgetown University Press.
- Labov, W. (1998). Co-existent systems in African-American English. In S. Mufwene, J. Rickford, G. Bailey, & J. Baugh (Eds.), *African American English* (pp. 110–153). Routledge.
- Labov, W., & Ash, S. (1997). Understanding Birmingham. In C. Bernstein, T. Nunnally, & R. Sabino (Eds.), *Language variety in the South revisited* (pp. 508–573). Alabama University Press.
- Labov, W., Ash, S., & Boberg, C. (2006). *The atlas of North American English: Phonetics, phonology and sound change*. Mouton de Gruyter.
- Lenth, R. V. (2022). emmeans: Estimated Marginal Means, aka Least-Squares Means. URL: <https://CRAN.R-project.org/package=emmeans>. R package version 1.8.3.
- Liu, Y., & Van Hell, J. G. (2020). Learning novel word meanings: An ERP study on lexical consolidation in monolingual, inexperienced foreign language learners. *Language Learning*, 70(S2), 45–74. <https://doi.org/10.1111/lang.12403>
- Lopez-Calderon, J., & Luck, S. J. (2014). ERPLAB: An open-source toolbox for the analysis of event-related potentials. *Frontiers in Human Neuroscience*, 8, 213. <https://doi.org/10.3389/fnhum.2014.00213>
- Luthra, S. (2024). Why are listeners hindered by talker variability? *Psychonomic Bulletin & Review*, 31(1), 104–121.
- Maher, Z. (2023). *Knowledge and processing of morphosyntactic variation in African American language and mainstream American English*. [Doctoral Dissertation, University of Maryland].
- Maher, Z., Vaughn, C., & Novick, J. (Under review). Listeners expect grammatical variation across dialects. https://osf.io/preprints/psyarxiv/eqvm7_v3
- Martin, C. D., Garcia, X., Potter, D., Melinger, A., & Costa, A. (2016). Holiday or vacation? The processing of variation in vocabulary across dialects. *Language, Cognition and Neuroscience*, 31(3), 375–390. <https://doi.org/10.1080/23273798.2015.1100750>
- Martin, C. D., Molnar, M., & Carreiras, M. (2016). The proactive bilingual brain: Using interlocutor identity to generate predictions of language processing. *Scientific Reports*, 6, 26171. <https://doi.org/10.1038/srep26171>
- Matuschek, H., Kliegl, R., Vasishth, S., Baayen, H., & Bates, D. (2017). Balancing Type I error and power in linear mixed models. *Journal of Memory and Language*, 94, 305–315.

- McGowan, K. B. (2015). Social expectation improves speech perception in noise. *Language and Speech*, 58(4), 502–521. <https://doi.org/10.1177/0023830914565191>
- McMurray, B., Ellis, T. P., & Apfelbaum, K. S. (2019). How do you deal with uncertainty? Cochlear implant users differ in the dynamics of lexical processing of noncanonical inputs. *Ear and Hearing*, 40(4), 961–980. <https://doi.org/10.1097/AUD.0000000000000681>
- Meade, G., Grainger, J., & Holcomb, P. J. (2019). Task modulates ERP effects of orthographic neighborhood for pseudowords but not words. *Neuropsychologia*, 129, 385–396. <https://doi.org/10.1016/j.neuropsychologia.2019.02.014>
- Molnar, M., Carreiras, M., & Ibáñez-Molina, A. J. (2015). Interlocutor identity affects language activation in bilinguals. *Journal of Memory and Language*, 81, 91–104. <https://doi.org/10.1016/j.jml.2015.01.002>
- Morales, J., Gómez-Ariza, C. J., & Bajo, M. T. (2013). Dual mechanisms of cognitive control in bilinguals and monolinguals. *Journal of Cognitive Psychology*, 25(5), 531–546. <https://doi.org/10.1080/20445911.2013.807812>
- Niedzielski, N. (1999). The effect of social information on the perception of sociolinguistic variables. *Journal of Language and Social Psychology*, 18(1), 62–85. <https://doi.org/10.1177/0261927X99018001005>
- Niedzielski, N. A., & Preston, D. R. (2000). *Folk linguistics*. Mouton de Gruyter. <https://doi.org/10.1515/9783110803389>
- Nygaard, L. C., & Pisoni, D. B. (1998). Talker-specific learning in speech perception. *Perception & Psychophysics*, 60(3), 355–376. <https://doi.org/10.3758/BF03206860>
- Nygaard, L. C., Sommers, M. S., & Pisoni, D. B. (1994). Speech perception as a talker-contingent process. *Psychological Science*, 5(1), 42–46. <https://doi.org/10.1111/j.1467-9280.1994.tb00612.x>
- Palmeri, T. J., Goldinger, S. D., & Pisoni, D. B. (1993). Episodic encoding of voice attributes and recognition memory for spoken words. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 19(2), 309–328. <https://doi.org/10.1037/0278-7393.19.2.309>
- Podesva, R. J., Reynolds, J., Callier, P., & Baptiste, J. (2015). Constraints on the social meaning of released /t/: A production and perception study of US politicians. *Language Variation and Change*, 27(1), 59–87.
- Preston, D. R. (2018). Changing research on the changing perceptions of Southern US English. *American Speech*, 93(3–4), 471–496. <https://doi.org/10.1215/00031283-7271283>
- R Core Team. (2022). R: A language and environment for statistical computing. R Foundation for Statistical Computing. <https://www.R-project.org/>
- Redick, T. S., Broadway, J. M., Meier, M. E., Kuriakose, P. S., Unsworth, N., Kane, M. J., & Engle, R. W. (2012). Measuring working memory capacity with automated complex span tasks. *European Journal of Psychological Assessment*, 28(3), 164–171. <https://doi.org/10.1027/1015-5759/a000123>
- Remez, R. E., Fellowes, J. M., & Rubin, P. E. (1997). Talker identification based on phonetic information. *Journal of Experimental Psychology: Human Perception and Performance* 23(3), 651–666.

- Samuel, A. G., & Larraza, S. (2015). Does listening to non-native speech impair speech perception? *Journal of Memory and Language*, 81, 51–71. <https://doi.org/10.1016/j.jml.2015.01.003>
- Sheffert, S. M., & Olson, E. (2004). Audiovisual speech facilitates voice learning. *Perception & Psychophysics*, 66, 352–362. <https://doi.org/10.3758/bf03194884>
- Signoret, C., Gaudrain, E., & Perrin, F. (2013). Similarities in the neural signature for the processing of behaviorally categorized and uncategorized speech sounds. *European Journal of Neuroscience*, 37, 777–785. <https://doi.org/10.1111/ejn.12097>
- Strand, E. A. (1999). Uncovering the role of gender stereotypes in speech perception. *Journal of Language and Social Psychology*, 18(1), 86–99. <https://doi.org/10.1177/0261927X99018001006>
- Strand, E. A., & Johnson, K. (1996). Gradient and visual speaker normalization in the perception of fricatives. In D. Gibbon (Ed.), *Natural language processing and speech technology: Results of the 3rd KONVENS Conference, Bielefeld, October 1996* (pp. 14–26). De Gruyter Mouton. <https://doi.org/10.1515/9783110821895-003>
- Summerfield, Q. (1979). Use of visual information for phonetic perception. *Phonetica*, 36(4–5), 314–331. <https://doi.org/10.1159/000259969>
- Sumner, M., & Kataoka, R. (2013). Effects of phonetically-cued talker variation on semantic encoding. *Journal of the Acoustical Society of America*, 134(6), EL485–EL491.
- Sumner, M., Kim, S. K., King, E., & McGowan, K. B. (2014). The socially weighted encoding of spoken words: A dual-route approach to speech perception. *Frontiers in Psychology*, 4, 1015.
- Sumner, M., & Samuel, A. G. (2009). The effect of experience on the perception and representation of dialect variants. *Journal of Memory and Language*, 60(4), 487–501. <https://doi.org/10.1016/j.jml.2009.01.001>
- Van der Feest, S. V., & Johnson, E. K. (2016). Input-driven differences in toddlers' perception of a disappearing phonological contrast. *Language Acquisition*, 23(2), 89–111. <https://doi.org/10.1080/10489223.2015.1047096>
- Van Ooijen, B. (1996). Vowel mutability and lexical selection in English: Evidence from a word reconstruction task. *Memory & Cognition*, 24, 573–583. <https://doi.org/10.3758/BF03201084>
- Wade, L. R., Embick, D., & Tamminga, M. (2023). Dialect experience modulates cue reliance in sociolinguistic convergence. *Glossa: A Journal of General Linguistics*, 2(1), 19. <https://doi.org/10.5070/G6011187>
- Walker, A. (2018). The effect of long-term second dialect exposure on sentence transcription in noise. *Journal of Phonetics*, 71, 162–176. <https://doi.org/10.1016/j.wocn.2018.08.001>
- Walker, A. (2019). The role of dialect experience in topic-based shifts in speech production. *Language Variation and Change*, 31(2), 135–163. <https://doi.org/10.1017/S0954394519000152>
- Walker, A. (2020). Voiced stops in the command performance of Southern US English. In Irina Shport & Wendy Herd (Eds.), *The Southern United States: Social factors and language variation*, a special issue of the *Journal of the Acoustical Society of America*, 147(1), 606–615.
- Walker, A., Hay, J., Drager, K., & Sanchez, K. (2018a). Divergence in speech perception. *Linguistics*, 56(1), 257–278.

- Walker, A., Fernandez, C., & Van Hell, J. G. (2020). The effect of talker identity on dialect processing. *University of Pennsylvania Working Papers in Linguistics*, 26(2), 1–12.
- Walker, A., Van Hell, J. G., & Bowers, M. (2018b). The effect of style-shifting on speech perception. Paper presented at *New Ways of Analyzing Variation (NWAV)*. New York, NY, October 18–21.
- Walker, S., Bruce, V., & O'Malley, C. (1995). Facial identity and facial speech processing: Familiar faces and voices in the McGurk effect. *Perception & Psychophysics*, 57(8), 1124–1133. <https://doi.org/10.3758/bf03208369>
- Wascher, E., Verleger, R., Jaskowski, P., & Wauschkuhn, B. (1996). Preparation for action: An ERP study about two tasks provoking variability in response speed. *Psychophysiology*, 33(3), 262–272.
- Weatherholtz, K. (2015). *Perceptual learning of systemic cross-category vowel variation*. [Doctoral Dissertation, The Ohio State University].
- Weatherholtz, K., Walker, A., Melvin, S., Royer, A., & Clopper, C. G. (2014). Effects of experience and expectations on adaptation to dialect variation in noise. Poster presented at the *27th Annual CUNY Conference on Human Sentence Processing*. Columbus, Ohio, March 14–15.
- Weissler, R. E. (2021). *Leveraging African American English knowledge: Cognition and multidialectal processing*. [Doctoral Dissertation, University of Michigan].
- Weissler, R. E., & Brennan, J. R. (2020). How do listeners form grammatical expectations to African American Language? *Penn Working Papers in Linguistics*, 25(2), 135–141.
- Zaharchuk, H. A., Shevlin, A., & Van Hell, J. G. (2021). Are our brains more prescriptive than our mouths? Experience with dialectal variation in syntax differentially impacts ERPs and behavior. *Brain and Language*, 218, 104949. <https://doi.org/10.1016/j.bandl.2021.104949>
- Zheng, Y., & Samuel, A. (2020). The relationship between phonemic category boundary changes and perceptual adjustments to natural accents. *Journal of Experimental Psychology: Learning, Memory & Cognition*, 46(7), 1270–1292. <https://doi.org/10.1037/xlm0000788>

