The Effect of Talker Identity on Dialect Processing

Abby Walker
Virginia Polytechnic Institute and State University

Carla B. Fernandez
Duke University

Janet D. van Hell
Pennsylvania State University

The Effect of Talker Identity on Dialect Processing

Abstract

Recent work has suggested that bilingual listeners use the visual identity of the talker to form expectations about the language the talker will use, which then facilitates lexical processing. In the current study, we extend this work to see if there are analogous effects of talker identity on dialect processing, and whether the impact of talker identity depends on the regional background of the listener. Six actresses recorded stimuli in two dialectal guises, performing Southern US accents and standardized, regionally nonspecific US accents. Participants were introduced to the actresses via video as having one particular dialect type (familiarization), and then later did an audio-visual lexical decision task (test) where some trials would be dialectally congruent and some trials would be dialectally incongruent with their earlier experience of that talker. US English listeners from both Southern and non-Southern dialect regions participated. Listeners who self-reported having (Southern) accents were impacted by talker dialect congruence, performing best with a given dialect when it matched their experience of that talker. However, other listeners were not impacted by congruency, performing better with standardized tokens regardless. This mirrors findings in bilingualism research that early bilinguals are more sensitive to talker language pairing than monolinguals or late bilinguals. We ran three additional conditions without video and/or without a familiarization stage to confirm the importance of each component to observing the effect. Generally, without familiarization, Southern US listeners performed worse with Southern vs. standardized tokens, suggesting that without strong contextual cues indicating otherwise, these listeners may expect standardized tokens in experimental settings. There is some evidence that all listeners were somewhat sensitive to talker identity even from voice alone.

The Effect of Talker Identity on Dialect Processing

Abby Walker, Carla B. Fernandez, and Janet G. van Hell*

1. Introduction

There are myriad factors that impact the ease with which we process auditory speech. One well-established and intuitively important factor is the dialect of the talker. Generally, listeners show easier processing with dialects that are more familiar to them, reflecting how perception is shaped by experience (e.g., Floccia et al. 2006, Labov and Ash 1997, Sumner and Samuel 2009, Walker 2018). However, there is also an apparently contradictory finding in the literature: listeners from regionally marked dialect regions actually appear to perform worse with their "own", presumably highly familiar dialect compared to a regionally unmarked (i.e., more standardized) dialect (Clopper and Bradlow 2008, Evans and Iverson 2007, Floccia et al. 2012, Walker et al. 2018).

The discrepancy between these two findings might be explained by another factor that impacts speech perception: context. Listeners are sensitive to contextual information when processing speech, and perform better with signal–context combinations that are congruent with their experience, or at least with their stereotype-based expectations. For example, words are recognized faster when presented in the same type of voice that listeners frequently hear say that word (Hay et al. 2019, Kim 2016). Listeners also form stereotypical expectations about a talker's L1 based on their perceived ethnicity, which impacts comprehension (Kang and Rubin 2009, McGowan 2015). Given that most experiments are run on university campuses, which are normatively standardized language institutions, the fact that speakers from marked dialect regions peform better with more standardized varieties may reflect their strong expectations for standardized language in these environments.

One important but understudied type of context is talker identity, not as a general property (i.e., this talker is old, this talker is White), but as a specific property (i.e., this talker is *Jordan*). It seems obviously true that listeners are creating strong expectations about specific talkers and dialects based on previous interactions with those talkers; for example, people who know the first author will expect her to speak with a New Zealand English accent, and would likely incur a processing cost if they heard her use a different accent. However, to our knowledge, studies have not explicitly tested this in regards to dialectal expectations using visual cues to talker identity. An exception is a study by Trude and Brown-Schmidt (2013), who used photos to show that perceptual adaptation appears to be talker-specific: participants shifted categorical boundaries for the specific talker that they had heard shifted primes from, but did not apply the learning to a different talker.

Recent work has investigated the role of talker identity in the language processing of bilinguals. Molnar, Ibáñez-Molina and Carreiras (2015) introduced bilingual Spanish–Basque listeners to talkers using videos. Each talker either spoke only in Basque, only in Spanish, or codeswitched between the languages. Listeners then did an audio–video lexical decision task where the same talkers all said some words in Spanish and Basque. The researchers found that participants were faster with experience-congruent talker–language pairings; i.e, faster with Basque than Spanish if they had been introduced to the person as a Basque talker. In addition, listeners were slower overall with the codeswitching talkers, apparently reflecting their inability to form a solid expectation about that talker's language variety.

In the study we describe here, we built on the design of Molnar et al. (2015) to explore similar effects of talker identity on dialect expectations. Specifically, we presented listeners with talkers producing speech in a performed Standardized US English dialect, or a performed Southern US English dialect. We were interested not only in whether talker identity impacted dialect

*Thanks to participants, actresses, and to research assistants Nicole DeFoor, Lily Carroll, Jessie Yu, Hayley Wood, Noel Boyle and especially Whitney Saunders. This project was financially supported by a grant awarded to Abby Walker and Mike Bowers by the Virginia Tech Institute for Society, Culture and Environment.

processing, but also in whether Southern listeners would still perform better with a Standardized US dialect vs. a Southern US dialect (cf. Walker et al. 2018).

2 Methods

2.1 Stimuli

The stimuli for this study consisted of audio–video recordings made by six white actresses who did not identify as being native speakers of Southern US English[1]. They were audio and video recorded at a TV studio at Virginia Tech, talking to camera with the aid of a teleprompter. Audio was recorded using a Sennheiser MKE 600 Shotgun Condenser Microphone (off-screen) into a Panasonic AJHPD2500 P2 Recorder (40KHz, 16 bit), into a .mov file.

Each actress was assigned a unique character name and two unique monologues that they produced to camera about their character's lives. Every monologue began with them saying their character name as a further anchor to talker identity. The actresses all also recorded the same 300 real and 120 nonsense monosyllabic words[2]. The real words contained vowels that can strongly mark Southern US English (Gunter et al. forthcoming): KIT (52), DRESS (52), PRIZE[3] (55), FACE (53), STRUT (55) and THOUGHT (33). These materials were recorded in two guises, with the actors instructed to sound clearly and authentically like a Southern US English speaker in one guise, and then like an aregional US speaker in the other guise (i.e., Standardized).

We chose to use actresses because in a pilot study (Walker et al. 2018) we had asked native, self-identified codeswitching Southerners to produce tokens in a Southern and Standardized ("aregional") accent, and found that while their speech did reliably change, they were still heard as Southern and still drew slower response times, relative to speakers who were from Northern Virginia. Given that this was our first use of this paradigm, we wanted to optimize our chances of observing talker–dialect congruency effects by investigating responses to categorical dialect shifts, akin to the language-switching seen in Molnar et al. (2015). Additionally, the task of creating the stimuli required comfort in front of cameras, bodily awareness, and the ability to produce engaging monologues. Using nonnative Southern talkers does raise concerns about the accuracy of their Southern accents, and introduces the possibility that the responses to our "Southern" tokens do not reflect how listeners would respond to authentic tokens of Southern speech. We cannot rule out this possibility. However, in optional comments left by participants, almost all of them mentioned noticing that the speakers had Southern accents (only 1/84 said the accents sounded fake), suggesting that regardless of authenticity, it appears that everyone understood the actresses as "doing" Southern.

The monologues were tightly edited with jump cuts, so that each monologue was ~1 minute, with minimal pauses. There were three types of monologues: Southern, Standardized, and Unpredictable. The Unpredictable monologues were made by switching between an actor's two dialect guises at some, but not all, of the jump cuts. We created this monologue to match the codeswitching speakers of Molnar et al. (2015), but we avoid calling this "codeswitching" or "bidialectalism" to reflect that this sort of extreme switching between dialects in a singular context may not reflect typical bidialectal/codeswitching behavior.

2.2 Experiment Design

The experiment was run using E-Prime 2.0, and for compatibility reasons videos were converted to .wma. 12 different experimental lists were created. In a given list, two of the actresses were presented in the monologues as Southern, two as Standardized, and two were presented as Unpredictable (Fig. 1). This basic pattern carried through the lexical decision task, such that 75%

[1]They came from Northern Virginia, Suffolk, VA, Lynchburg, VA (2), Maryland, and Wisconsin

[2]Files were cut so that the video started an average of 387ms before audio onset (minimum=160ms, maximum=585ms). Standardized words have an average duration of 595ms (*SD*=125), and Southern words are on average 604ms long (*SD*=127). This 9ms difference is significant in a paired t test, but in the response time models that follow that we measure RT from word end, and the effect sizes are much bigger than 9ms.

[3]We use PRIZE instead of PRICE to indicate no /ai/ vowels were before voiceless consonants.

of tokens from Southern speakers were Southern, 75% of tokens from Standardized speakers were predominantly Standardized, and tokens from Unpredictable speakers were an even mix of both dialects. These distributions allowed us to maintain the Speaker Type established in the monologues (Southern, Standardized, Unpredictable) while also testing for the impact of incongruencies. The specific actor associated with a Speaker Type differed across lists.

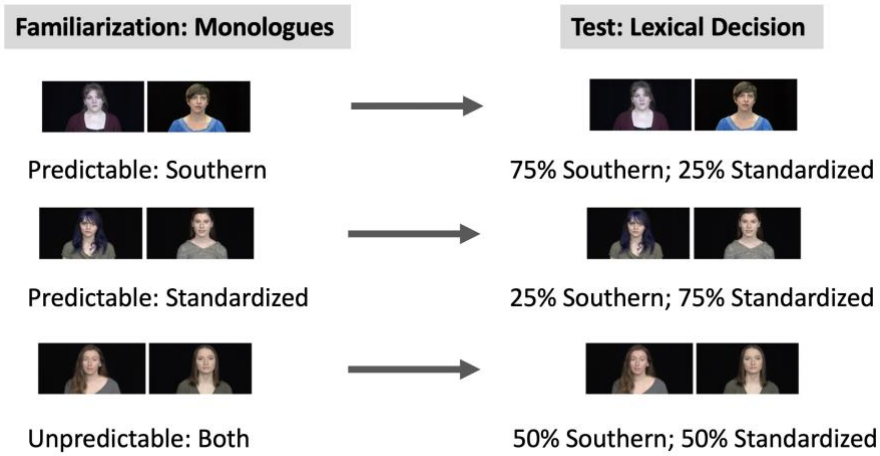

Figure 1. Experimental design for one list in the Video–Monologue condition. Percentages reflect how many real words were presented in each dialect guise.

Participants first watched and listened to two monologues from each speaker (across two cycles through the speakers), answering short comprehension questions after each monologue to encourage them to pay attention. Then participants did a speeded auditory–visual lexical decision task, where they were asked to sort the word they heard into the category of "real" or "fake" by pressing a button as quickly as possible.

The condition we have described so far is the critical *Video–Monologue* condition, which matches the experimental design described in Molnar et al. (2015). To ensure that any effects we saw relied on a) the use of video, and b) a familiarization period, we created three additional experimental conditions: a *Video* condition, where participants did the audio–visual lexical decision task but did not see the familiarization monologues; an *Audio–Monologue* condition that had the exact same audio, including the monologues, as the Video–Monologue condition, but no video; and an *Audio* condition, with neither a familiarization stage nor video.

After completing the lexical decision task, all participants completed a background questionnaire, and answered 43 8-point Likert scale questions, that covered topics related to their attitudes toward accents (including Southern accents), their own experience being heard as (un)accented, their prescriptive attitudes, and their mood.

2.3 Participants

184 people participated in the study, recruited through the Virginia Tech Psychology Department Subject Pool. 40 participants were removed for: not finishing the study (15); not providing sufficient background information (9); not meeting inclusion criteria (13); having error rates 2.5 *SD* below the mean (3). The remaining 144 participants were all native speakers of US English and had not spent more than 5 years outside of the US. We divided these speakers into two regional groups: *Southern* if they had lived in a Southern region of the US prior to age 18, and *NotSouthern* if they had never lived in the South prior to age 18. The latter group mostly, but not exclusively, consisted of participants from Northern Virginia and nearby eastern states (NJ, NY, MD). Table 1 shows the distribution of participants by regional background, across conditions.

To explore differences in speakers based on their responses to the 43 Likert scale questions in the final questionnaire, we performed factor analysis on these questions and identified three latent variables, including a "Southern Accent Score" (loading above .7 on *People tease my accent, I have an accent* and *I have a Southern accent*). Figure 2 shows the distribution of

Southern Accent Scores by Participant Region. Participants who had lived in the US South had higher scores (mean = 0.44, *SD*=1) than those who had not (mean = -0.33, *SD*=.76), and this was significant in a two-tailed, Welch's t-test, $t(109.33) = 4.91$, $p < .001$. However, substantial variation in both groups suggests that the score reflects something more than the broader dialect regions.

Condition	Not Southern	Southern	Total by Condition
Audio	16	15	31
Audio–Monologue	13	12	25
Video	31	20	51
Video–Monologue	21	16	37
Total by Dialect	81	63	144

Table 1. Distribution of participants across conditions, by regional background.

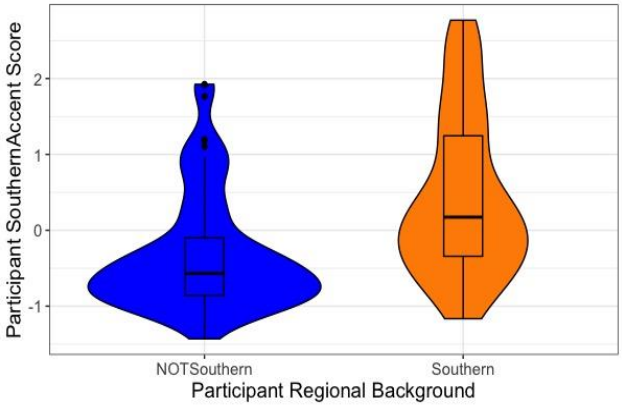

Figure 2. Southern Accent Score by Participant Region

3 Results

Prior to analysis, all responses where a trial timed out before the answer were removed, as well as responses before the start of the word (<1%). For each participant, we removed tokens with response times greater than 2 *SD* outside their response time mean for each word type (real and nonsense), excluding 2.5% of the remaining data, and leaving a total of 58,488 responses across the 4 conditions (41,800 for real words).

The means in Table 2 suggest that participants in the video conditions were more accurate, but slower, than participants in the audio-only conditions. The accuracy differences were driven largely by responses to nonsense words; performance across video and audio were similar for real words. This could reflect the fact that speech processing is easier with visual cues (Remez 2005), assuming that the real word accuracy rates here reflect a ceiling. Regardless of global differences between the two formats, in both cases participants are less accurate and slower with tokens produced in the Southern vs. Standardized guises.

Modality	Token Dialect	Real Accuracy	Real Word RT	Nonsense Accuracy	Nonsense RT
Audio	Standardized	96	403 (208)	79	528 (276)
	Southern	92	434 (211)	77	525 (272)
Video	Standardized	96	457 (260)	85	604 (336)
	Southern	91	500 (273)	83	608 (343)

Table 2. Accuracy and mean response times (*SD)* by condition modality and token dialect

3.1 Audio-only Conditions

3.1.1 Accuracy

The best fit[4] mixed effects logistic regression model for real word accuracy in the two auditory conditions is shown in Table 3. There is a three-way interaction between Token Dialect, Experimental Order, and Monologue: overall, people were less accurate with Southern than with Standardized tokens, but did better, especially with Southern tokens, as the experiment continued, and at a faster rate of improvement after having listened to the monologues. There was also an interaction between Speaker Type and Trial Number, such that people did worse with the tokens from the Unpredictable speakers as the experiment progressed.

	Estimate	Std. Error	z value	p value
(Intercept)	3.368	0.17063	19.736	<0.001
Monologue=Yes	-0.122	0.19874	-0.613	0.54
TrialNumber (log, centered)	-0.179	0.09859	-1.816	0.069
TokenDialect=Southern	-0.784	0.10312	-7.607	<0.001
SpeakerType=Unpredictable	0.108	0.08956	1.207	0.227
SpeakerType=Southern	0.202	0.09138	2.197	0.028
Monologue=Yes*TrialNumber	0.027	0.12535	0.212	0.832
Monologue=Yes*TokenDialect=Southern	0.013	0.14195	0.09	0.928
TrialNumber*TokenDialect=Southern	0.501	0.10432	4.807	<0.001
TrialNumber*SpeakerType=Unpredictable	0.256	0.08649	2.955	0.003
TrialNumber*SpeakerType=Southern	0.018	0.09291	0.188	0.851
Monologue=Yes*TrialNumber*TokenDialect=Southern	-0.357	0.14813	-2.408	0.016

Table 3. Model coefficients for best fit accuracy model (audio-only conditions). *N* items=16296, *N* Participant=56.

	Estimate	Std. Error	t value
(Intercept)	372.978	18.114	20.59
Monologue=Yes	37.151	21.661	1.715
TrialNumber (log, centered)	-7.654	2.199	-3.481
TokenDialect=Southern	26.13	3.31	7.894
SouthernAccent (centered)	-10.595	10.971	-0.966
SpeakerType=Unpredictable	26.751	3.822	7
SpeakerType=Southern	6.708	4.089	1.64
Monologue=Yes*TrialNumber	-12.969	3.247	-3.994
SouthernAccent*SpeakerType=Unpredictable	-10.559	3.663	-2.883
SouthernAccent*SpeakerType=Southern	-6.87	3.711	-1.851

Table 4. Model coefficients for best fit RT (word end) model (audio-only conditions). *N* items=15331, *N* Participants=56.

3.1.2 Response Times

The best fit linear regression mixed effects model for response times[5] (to correct answers) for the two audio conditions is shown in Table 4. Similar to the accuracy model, we find that listeners do worse (are slower) with Southern vs. Standardized tokens, and that they generally get better/faster

[4]The model selection process started with full models and model reduction was done through model comparison using a log likelihood test. All models included a random intercept for Actress and Participant.

[5]In this paper we measure response times from the end of the word. Findings are similar when we measure response time from the start of the word. with word duration included in the model as a fixed effect.

over time, this time as a general effect (not in interaction with Token Dialect). Participants who heard the monologues also have a faster slope of improvement (in part because they start off slower). There is an interaction between Speaker Type and Southern Accent Score: participants with lower scores do relatively worse/are slower with Southern and especially Unpredictable Speaker Types compared to listeners with higher scores[6]. Note that this effect is visible with both Southern and Standardized *tokens* (the interaction here is not with Token Dialect).

3.2 Video Conditions

3.2.1 Accuracy

The best fit model for real word accuracy in the video conditions (Table 5) includes a three-way interaction of Monologue, Token Dialect and Southern Accent Score: people who did not see the monologues do worse with Southern tokens independent of their own Southern Accent Score, but listeners who saw the monologues do better with Southern tokens if they have a higher score (and tend to do worse with Standardized tokens (Fig. 3)). There is also an interaction with Token Dialect and Trial Number, such that participants generally get better with tokens produced in a Southern dialect over the course of the experiment.

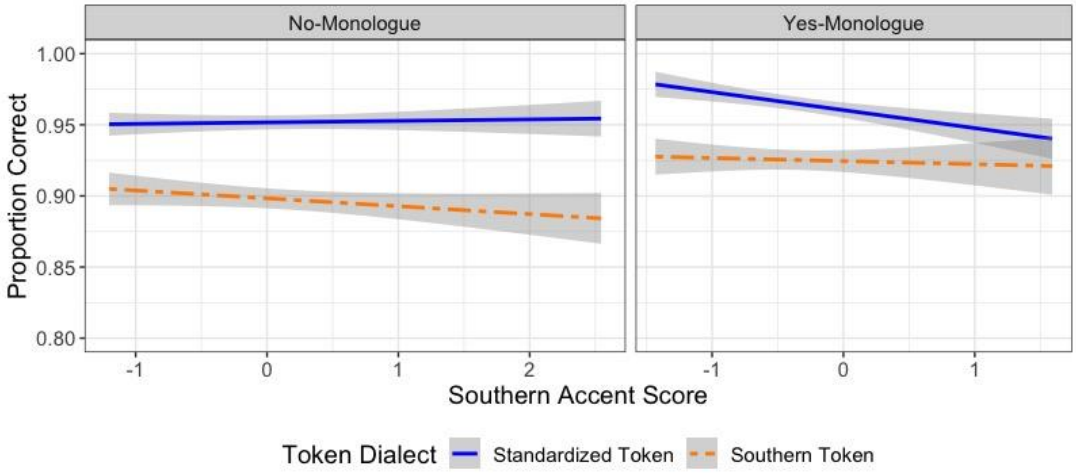

Figure 3. Effect of Southern Accent score, Monologue, and Speaker Type on real word accuracy (video conditions)

	Estimate	Std. Error	z value	p value
(Intercept)	3.206	0.13674	23.444	<0.001
TokenDialect=Southern	-0.79	0.06857	-11.515	<0.001
SouthernAccent (centered)	0.011	0.10433	0.108	0.9143
Monologue=Yes	0.148	0.17479	0.847	0.397
TrialNumber (log, centered)	0.027	0.04543	0.605	0.545
SouthernAccent*TokenDialect=Southern	-0.084	0.06564	-1.284	0.199
Monologue=Yes*TokenDialect=Southern	0.099	0.11688	0.849	0.396
SouthernAccent*Monologue=Yes	-0.357	0.2051	-1.739	0.082
TrialNumber*TokenDialect=Southern	0.262	0.05389	4.86	<0.001
SouthernAccent*TokenDialect=Southern*Monologue=Yes	0.396	0.14305	2.767	0.0057

Table 5. Model coefficients for best fit accuracy model (video conditions). *N* items=25504, *N* participants= 88.

[6]The results are similar if we substitute Southern Accent Score with Participant Region, but in later models, only Southern Accent Score and *not* Participant Region significantly impacts responses.

3.2.2 Response Times

The best fit model for response times to correct, real words in the combined video conditions (Table 6) included significant effects for Monologue, in interaction with Speaker Type * Trial Number, and in interaction with Token Dialect * Speaker Type * Southern Accent Score. The first interaction shows that while there is a general trend for participants in both conditions to get faster in later trials, participants start off slower overall when they have seen the monologue, and have a steeper negative slope (get faster more quickly). This interaction is carried by a difference between the two conditions in how Trial Number interacts with responses to the Standardized speakers, since participants without the monologue start off responding very quickly to this Speaker Type, and do not get any faster over time with them.

The second interaction is visible in Figure 4. In the condition without a monologue, there were main effects of Token Dialect (participants were slower with Southern compared to Standardized tokens), and Speaker Type (participants were slower with Unpredictable speakers). The listeners' Southern Accent Score did not matter. However, for participants who did the lexical decision task after watching and listening to the monologues, there is a three-way interaction between Token Dialect, Speaker Type, and the listener's Southern Accent Score: participants with higher scores show a cost for incongruent trials, such that they are fastest on Standardized dialect trials from Standardized Speaker Types, but slower for Standardized tokens from Unpredictable or Southern Speaker Types. There is a smaller but still significant converse effect for Southern tokens, which signifies that they are faster with the Southern dialect when produced by a Southern Speaker Type, compared to Unpredictable and Standard Speaker Types.

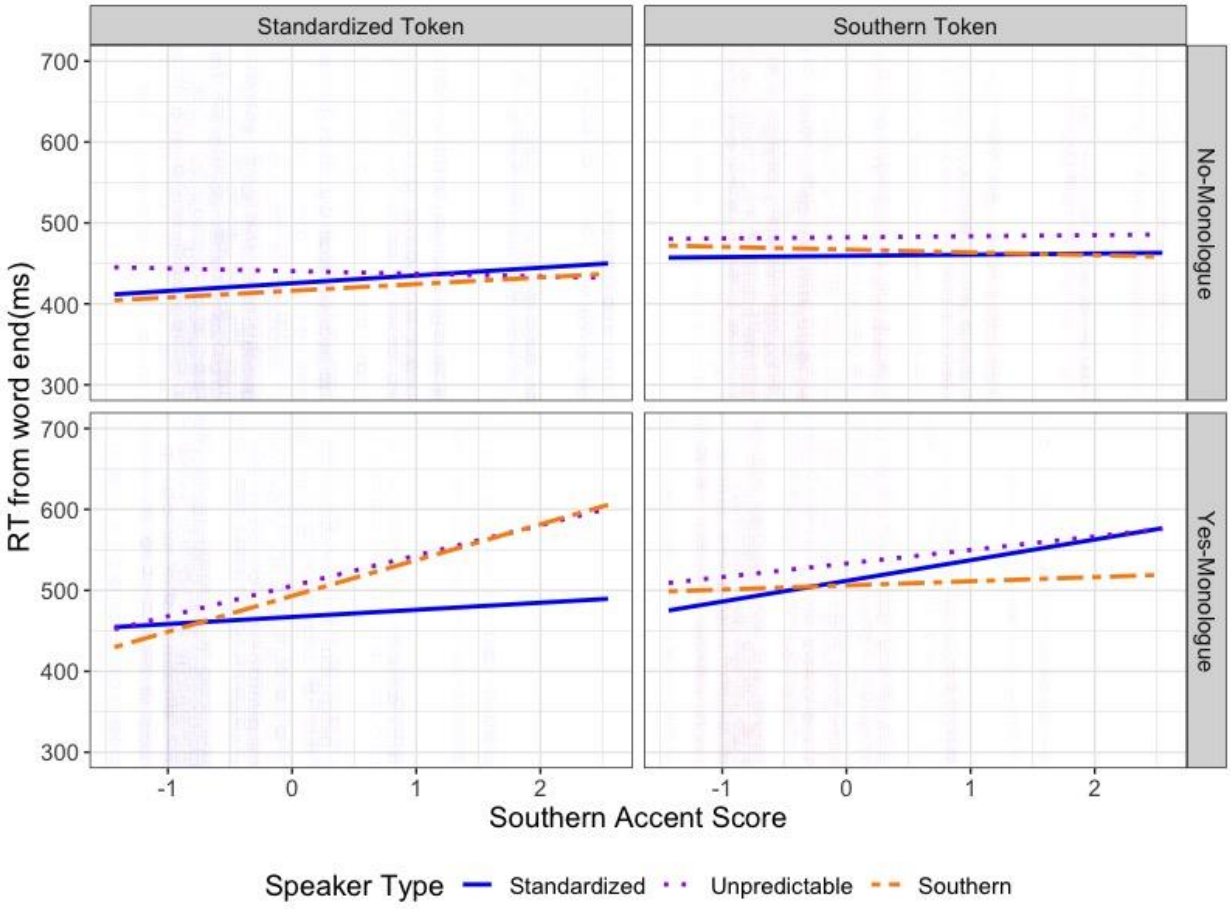

Figure 4. The interaction between Southern Accent Score, Token Dialect, Speaker Type, and Monologue on response times (video conditions only). NB: This figure only shows a range of 300-700ms to better show effect, but actual RT values range from -680 to 2580).

	Estimate	Std. Error.	t value
(Intercept)	424.653	20.54075	20.674
TokenDialect=Southern	36.223	8.01708	4.518
SpeakerType=Unpredictable	16.548	6.10957	2.709
SpeakerType=Southern	-7.317	7.84989	-0.932
SouthernAccent (centered)	7.515	17.4503	0.431
Monologue=Yes	42.149	29.02434	1.452
TrialNumber (log, centered)	-7.315	3.66967	-1.993
TokenDialect=Southern*DialectType=Unpredictable	5.67	10.51948	0.539
TokenDialect=Southern*Dialect Type=Southern	13.006	11.2655	1.154
TokenDialect=Southern*SouthernAccent	-8.126	7.76116	-1.047
DialectType=Unpredictable*SouthernAccent	-8.891	5.94259	-1.496
DialectType=Southern*SouthernAccent	0.036	7.60562	0.005
TokenDialect=Southern*Monologue=Yes	9.656	12.87622	0.75
DialectType=Unpredictable*Monologue=Yes	22.67	9.87097	2.297
DialectType=Southern*Monologue=Yes	30.6	12.70089	2.409
TokenDialect=Southern*Monologue=Yes	1.113	33.54779	0.033
DialectType=Unpredictable*TrialNumber	-7.643	5.11916	-1.493
DialectType=Southern*TrialNumber	-2.395	5.39638	-0.444
TrialNumber*Monologue=Yes	-22.736	5.59226	-4.066
TokenDialect=Southern*DialectType=Unpredictable*SouthernAccent	15.137	10.20119	1.484
TokenDialect=Southern*DialectType=Southern*SouthernAccent	-0.571	10.92271	-0.052
TokenDialect=Southern*DialectType=Unpredictable*Monologue=Yes	-22.278	16.90825	-1.318
TokenDialect=Southern*DialectType=Southern*Monologue=Yes	-43.668	18.11947	-2.41
TokenDialect=Southern*SouthernAccent*Monologue=Yes	24.194	14.77891	1.637
DialectType=Unpredictable*SouthernAccent*Monologue=Yes	36.609	11.47049	3.192
DialectType=Southern*SouthernAccent*Monologue=Yes	30.44	14.7777	2.06
DialectType=Unpredictable*Monologue=Yes*TrialNumber	19.371	7.71313	2.511
DialectType=Southern*Monologue=Yes*TrialNumber	13.51	8.06657	1.675
TokenDialect=Southern*DialectType=Unpredictable*SouthernAccent *Monologue=Yes	-50.852	19.43346	-2.617
TokenDialect=Southern*DialectType=Southern*SouthernAccent *Monologue=Yes	-50.95	20.90027	-2.438

Table 6. Model coefficients for best fit RT (word end) model (video conditions). *N* items=23809, *N* participants= 88.

4. Discussion

In this lexical decision task, we find that, generally, native US English-speaking listeners do worse (more mistakes, slower responses) with real word tokens presented in a performed Southern US dialect, as opposed to a performed Standardized US dialect. In all but the Video–Monologue condition, this effect is largely independent of whether speakers grew up in the US South or self-report having a (Southern) accent. We note that these results are similar to those of an auditory lexical decision task we ran using authentic speakers from Southwest (SWVA) and Northern Virginia (NOVA): people were slower with the SWVA speakers compared to the NOVA speakers, even if they were from SWVA (Walker et al. 2018). It also matches the results of other studies that have found that listeners perform better with a standardized dialect relative to a regionally marked dialect, even if they are from the marked dialect region (e.g., Clopper and Bradlow 2008).

However, the Video–Monologue condition yielded a different pattern. In this condition, participants are introduced audio–visually to the 6 talkers before they do the audio–visual lexical decision task. Critically then, they are able to form strong talker-specific dialect expectations, and

we found that Southern listeners stop performing better with Standardized vs. Southern tokens. In terms of accuracy, they perform equally well with Standardized and Southern tokens, and in response times, they do better with speaker-congruent tokens – i.e., Southern tokens from Southern speakers, and Standardized tokens from standardized speakers – and do worse with speaker-incongruent tokens.

Therefore, a key finding of this study is that if speakers of a regionally marked dialect have strong reason to expect their regional dialect from a speaker, they no longer do better with a comparatively standardized dialect. In fact, in terms of response times we can see them doing considerably worse with standardized, but unexpected tokens. This suggests that earlier findings (e.g., Evans and Iverson 2007, Walker et al. 2018) may also reflect participants' strong expectations for standardized language forms in university and/or experimental settings.

Another important factor may be how we classify participants as "having" a regionally marked dialect. In the Video–Monologue conditions it was critically a listener's Southern Accent Score that predicted their behavior in the task, and not where they had lived. The strength of a person's regional accent has long been tied to socioeconomic status (Trudgill 2000), and enrollment at universities is dominated by students from higher socioeconomic backgrounds (Kena et al. 2015). Moreover, young people who intend to go to university tend to use fewer local dialect features than peers who do not (e.g., Eckert 2000). Importantly, when different methods are used to classify speaker dialect, differences do emerge in listening tasks between speakers of standardized and marked dialects (Evans & Iverson 2007; Sumner & Samuel 2009; Walker 2016).

It is interesting that listeners with low Southern Accent Scores did not appear to be affected by incongruencies: they did better with Standardized tokens than Southern tokens, even if the Standardized token came from a Southern talker. This parallels Molnar et al.'s (2015) finding that it was only early, but not late, bilinguals who showed sensitivity to talker–language mismatches, and Fecher and Johnson's (2018) findings that bilingual, but not monolingual, infants appeared to be tracking the language a speaker spoke. One explanation entertained by these bilingualism researchers is that early bilinguals learn to pay more attention to talker–language pairings. The same reasoning could explain our results: Southern-accented listeners constantly hear both Southern and Standardized dialects (i.e., they are (receptively) bidialectal), and similarly deal with this variation by forming stronger talker-specific dialect expectations.

Another interpretation is that non-Southern listeners were paying similar amounts of attention to talker–dialect pairings, but that it did not help them because they have insufficient experience with Southern US English on which to draw. One reason to favor this account over the previous one is that there was evidence for general difficulty with the Unpredictable speakers across participants. If listeners were not paying attention to speaker identity, then it is hard to explain why they found the Unpredictable speakers the most difficult to process[7].

As a final note, while we only found effects of dialect–speaker incongruency explicitly in the Video–Monologue condition, there is evidence that listeners are forming some sense of talker–dialect expectations even in the audio-only conditions. For example, in terms of accuracy, listeners in the audio conditions do significantly worse with Unpredictable speakers as the experiment progresses (independent of the particular Token Dialect). The idea that listeners are forming voice-specific expectations complements other audio-only work suggesting that perceptual learning is talker-specific (e.g., Creel et al. 2008, Nygaard and Pisoni 1998).

References

Clopper, Cynthia, and Ann Bradlow. 2008. Perception of dialect variation in noise: Intelligibility and classification. *Language and Speech* 51:175–198

Creel, Sarah. C., Richard N. Aslin and Michael K. Tanenhaus. 2008. Heeding the voice of experience: The role of talker variation in lexical access. *Cognition* 106(2):633–664.

Eckert, Penelope. 2000. *Linguistic Variation as Social Practice*. Oxford: Blackwell.

Evans, Bronwen and Paul Iverson. 2007. Plasticity in vowel perception and production: A study of accent change in young adults. *Journal of the Acoustical Society of America* 121:3814–3826.

[7]Molnar and colleagues also found that their participants performed worst with the codeswitching participants.

Fecher, Natalie and Elizabeth Johnson. 2018. Bilingual infants excel at foreign-language talker recognition. *Developmental Science Short Report* e12778.
Floccia, Caroline, Jeremy Goslin, Frédérique Girard and Gabrielle Konopczynski. 2006. Does a regional accent perturb speech processing? A lexical decision study in French listeners. *Journal of Experimental Psychology: Human Perception and Performance* 32:1276–1293
Floccia, Caroline, Claire Delle Luche, Samantha Durrant, Joseph Butler and Jeremy Goslin. 2012. Parent or community: Where do 20-month-olds exposed to two accents acquire their representation of words? *Cognition* 124: 95–100.
Gunter, Kaylynn, Charlotte Vaughn and Tyler Kendall. Forthcoming. Perceiving Southernness: Vowel categories and acoustic cues in Southernness ratings. To appear in the *Journal of the Acoustical Society of America*.
Hay, Jennifer, Abby Walker, Kauyumari Sanchez and Kirsty Thompson. 2019. Abstract social categories facilitate access to socially skewed words. *PLOS ONE*, pone.0210793: 1–29.
Kang, Okim, and Donald Rubin. 2009. Reverse linguistic stereotyping: Measuring the effect of listener expectations on speech evaluation. *Journal of Language & Social Psychology* 28:441–456.
Kena, Grace, Lauren Musu-Gillette, Jennifer Robinson, Xiaolei Wang, Amy Rathbun, Jijun Zhang, Sidney Wilkinson-Flicker, Amy Barmer and Erin Dunlop Velez. 2015. The Condition of Education 2015. *National Center for Education Statistics* 144.
Kim, Jonny. 2016. Perceptual associations between words and speaker age. *Journal of the Association for Laboratory Phonology* 7(1): 1–22.
Labov, William and Sharon Ash. 1997. Understanding Birmingham. In *Language variety in the South Revisited*, ed. C. Bernstein, T. Nunnally and R. Sabino, 508–573. Tuscaloosa, AL: Alabama University Press.
McGowan, Kevin. 2015. Social expectation improves speech perception in noise. *Language and Speech* 58:502–521.
Molnar, Monica, Manuel Carreiras and A.J. Ibañez-Molina. 2015. Interlocutor identity affects language activation in bilinguals. *Journal of Memory and Language* 81: 91–104.
Nygaard, Lynne C. and David B. Pisoni. 1998. Talker-specific learning in speech perception. *Attention, Perception, & Psychophysics* 60(3), 355–376.
Remez Robert E. 2005. Three puzzles of multimodal speech perception. In *Audiovisual Speech*, ed. E. Vatikiotis-Bateson, G. Bailly and P. Perrier, 12–19. MIT Press; Cambridge, MA.
Sumner, Meghan and Arthur G. Samuel. 2009. The role of experience in the processing of cross dialectal variation. *Journal of Memory and Language* 60:487–501.
Trude, Alison and Sarah Brown-Schmidt. 2013. Talker-specific perceptual adaptation during online speech perception. *Language and Cognitive Processes* 27 (7-8): 979–1001
Trudgill, Peter. 2000. *An Introduction to Language and Society.* 4th Edition. London: Penguin Books.
Walker, Abby. 2016. Advantage Accented? Listener differences in understanding speech in noise. *University of Pennsylvania Working Papers in Linguistics* 21(2): 181–189
Walker, Abby. 2018. The effect of long-term second dialect exposure on sentence transcription in noise. *Journal of Phonetics* 71: 162–176.
Walker, Abby, Janet van Hell and Mike Bowers. 2018.. The effect of style-shifting on speech perception. Paper presented at *New Ways of Analyzing Variation (NWAV)*, New York, NY, October 18–21.

Abby Walker
Department of English
Virginia Polytechnic Institute and State University
Blacksburg, VA 24061-0203
ajwalker@vt.edu

Carla Fernandez
Department of Psychology and Neuroscience
Duke University
Durham, NC, 27708
carla.fernandez@duke.edu

Janet van Hell
Department of Psychology
Pennsylvania State University
University Park, PA, 16802
jgv3@psu.edu