Influences of Speaker-Listener Similarity on Shadowing and Comprehension

Lynn K. Perry (lkperry@miami.edu)
Department of Psychology, 5665 Ponce De Leon Blvd
Coral Gables, FL 33124 USA

Emily N. Mech (emech@wisc.edu)
Maryellen C. MacDonald (mcmacdonald@wisc.edu)
Mark S. Seidenberg (seidenberg@wisc.edu)
Department of Psychology, 1202 W. Johnson Street
Madison, WI 53706 USA

Abstract

We routinely encounter speakers with different accents and speaking styles. The speech perception literature offers examples of disruption of comprehension for unfamiliar speech and also of listeners' rapid accommodation to unfamiliar accents. Much of this research uses a single measure and/or focuses on isolated word perception. We investigated listeners' abilities to comprehend and shadow connected speech spoken in a familiar or unfamiliar accent. We found increases in shadowing latencies and comprehension errors in the Dissimilar Speech relative to Similar Speech conditions—especially for relatively informal rather than more academic style speech. Additionally, there was less accommodation over time to Dissimilar than Similar Speech. These results suggest that there are costs both in the immediate timescale of processing speech (necessary for shadowing) and in the longer time scale of listening comprehension when accent and other speech quality is very different from one's own speech.

Keywords: speech perception; accented speech; speech shadowing; listening comprehension

Introduction

In every perceptual domain, familiarity improves perceptual speed and accuracy. In speech, familiarity can take many forms: we can be familiar with the words spoken, with a particular speaker's voice quality, or with more general characteristics of a speech community such as accent and typical speaking rate. An important question to the study of speech perception, then, is how do listeners make sense of unfamiliar speech in order to communicate? In the current experiment, we examined the effect of familiarity of a talker's speaking style on listeners' ability to comprehend rapidly and accurately. We manipulated familiarity with recordings of two speakers reading aloud, one from the same speech community as the participants and one from a different speech community. We will refer to this manipulation as familiarity with the talker's speech style or "accent", recognizing that many factors affect listeners' perception of an individual's speaking style and its similarity to their own speech.

Previous research has demonstrated that non-native accents disrupt intelligibility of speech to native speakers, as do differences in regional accents (Adank, Evans, Stuart-Smith, & Scott, 2009; Floccia, Goslin, Girard, & Konopczynski, 2006; Major, Fitzmaurice, Bunta, & Balasubramanian, 2005). However, it is also clear that both adults and children can accommodate unfamiliar accents with sufficient exposure (Maye, Aslin, & Tanenhaus, 2008; Sumner, 2011; Sumner & Samuel, 2009; White & Aslin, 2011). The ease with which listeners can accommodate to unfamiliar speech patterns is often held as an example of how impressive human language abilities are. This accommodation is often measured via recognition of isolated words (e.g., Sumner, 2011) rather than sentences or longer utterances, and the degree of accommodation to continuous speech is less well understood.

Studies that have investigated perception of connected speech have often used speech shadowing tasks. Shadowing requires participants to repeat aloud the speech they hear as quickly and accurately as possible (e.g., Marslen-Wilson, 1973). Sabatini (2000) examined native Italian speakers trained as professional English-language interpreters' listening comprehension, shadowing, and simultaneous interpretation of nonstandard American English and Indian nonnative English. Consistent with previous findings, the unfamiliar accents were less intelligible to participants in all tasks (Sabatini, 2000). However, listening comprehension was easier than shadowing and simultaneous interpretation. The longer time scale of comprehension tasks is helpful because background knowledge and downstream speech information may clarify the interpretation of earlier input. Although top-down influences occur even on the shorter timescale in which shadowing occurs (e.g., Marslen-Wilson, 1985; Nye & Fowler, 2003), shadowing provides an index of the difficulties listeners encounter in early stages of processing unfamiliar speech. The Sabatini (2000) study clearly revealed that non-native listeners have difficulties processing unfamiliar accents. However, it is unclear whether such effects also occur for native speakers listening to unfamiliar speech, such as markedly different regional accent or voice quality.

We compared shadowing and offline listening comprehension measures to assess effects of speech familiarity at two timescales. Two native speakers of American English from different speech communities each recorded four text passages for use in shadowing and comprehension tasks. Two passages were more academic in character and two were in a more informal narrative style. This manipulation of Passage Type was included to determine whether the impact of speaker variability depended on type of material. The academic passages contained more low frequency words, which could increase the difficulty of speech comprehension for these passages compared to more informal narratives.

Methods

Participants

59 (36 female) native English speakers from the Midwest United States participated. All participants identified themselves as white native English Speakers. 29 participants were assigned to the Similar Speech Condition (16 female), and 30 to the Dissimilar Speech Condition (20 female). Participants were recruited from introductory psychology classes at the University of Wisconsin-Madison and received course credit for their participation.

Stimuli

Four text passages, two academic and two informal, were used as stimuli. The academic passages were drawn from reading comprehension portions of the Test of English as a Foreign Language (TOEFL) exam. The informal passages were written transcriptions of two stories from the radio program "This American Life." Passages varied between 308 and 350 words in length.

Each passage was recorded by two female native English speakers, both graduate students in their mid-20s. The speaker in the Similar Speech Condition is a white woman from the Midwest, the same region as the participants, while the speaker in the Dissimilar Speech Condition is an African American woman from the southeastern US and identifies herself as a speaker of African American English as well as Mainstream American English. Speakers were informed that the recordings would be used for a study of speech perception and accents. They read through the passages before making the recordings and were instructed to speak as naturally as possible, which led to slight differences in length of each speaker's recordings (Similar Speech: M=134 seconds; Dissimilar Speech: M=133 seconds).

Each speaker's recordings were rated for familiarity by fourteen white participants from the upper Midwest who did not participate in the main experiment. Similar Speech Condition recordings were rated more similar to their own speech, $t(13)=4.00$, $p=.002$; compared to the versions for the Dissimilar Speech Condition. To be sure that the Dissimilar Speech recordings were not rated lower in intelligibility and familiarity for speaker specific—rather than dialect specific—reasons, an additional eleven African American participants, primarily from the Southern U.S., rated speaker familiarity. Critically, these participants rated the Dissimilar (i.e. African American English) Speech recordings as marginally *more similar* to their own speech than the white Midwestern participants had rated these recordings, $t(23)=1.83$, $p=.081$. The African American raters judged the recordings of the white speaker as significantly less similar to their own speech than the white Midwestern participants rated these recordings, $t(18)=3.78$, $p=.001$.

Procedure

Speech similarity (Similar, Dissimilar) was manipulated between participants. Task (shadowing, comprehension), and Passage Type (Academic, Informal) were manipulated within participants, with the order of conditions randomized for each participant. Passages were presented over headphones in a quiet lab room. Each participant heard all four passages once, with assignment of conditions counterbalanced across participants. All experimental tasks were run using E-prime 2.0 software. The experiment took about 20 minutes to complete.

Shadowing Task Participants were instructed to repeat the words in the passages as they were heard as quickly and accurately as possible, speaking into a microphone directly in front of them. One Academic passage and one Informal passage were presented.

Listening Comprehension Task Participants were instructed to listen to each passage and were told they would answer true/false questions afterwards to test their comprehension. After they heard each passage, they answered six true/false questions (presented one at a time) by pressing the T and F keys on the keyboard. One Academic passage and one Informal passage were presented.

Coding

Shadowing Task Two trained research assistants, blind to the experimental hypotheses, coded each participant's speech shadowing for errors and latency.

Errors: Any errors or deviations from the original transcript were coded as omissions, constructive errors, or delivery errors. Omissions were whole words that participants omitted in shadowing. Constructive errors included any added words or changes to words that resulted in a different word or a nonword. Delivery errors included slurred hesitations, stuttering, and unintelligible responses.

Latency: Every tenth word of participants' shadowing was coded for latency relative to the original transcript. Latency was measured at the word onset, which was determined by analysis of the speech spectrogram on Praat software.

Listening Comprehension Listening comprehension accuracy was measured for each participant as the total number of true/false questions answered correctly for each passage.

Analyses

All analyses were conducted using mixed effects regression models. Latency analyses were conducted using linear regression and accuracy analyses were conducted using logistic regression. To determine the best-fit model, we used chi-square tests comparing models with and without the factor of interest. For interactions, we report coefficients and confidence intervals from the full model, and the chi-square test of model fit from the comparison to a model with the interaction removed. For main effects, we report coefficients and confidence intervals from the full model, and the chi-square test of model fit from the comparison to a model with the predictor main effect removed. To determine appropriate random effects, we began with completely specified random effects structures including random slopes for all variables in a given model. Using model comparison, we systematically removed uninformative random effects (Jaeger, 2009). All final models included random intercepts for subjects and items.

Results

We first report performance in each task separately and then examine the relationship between them. Our primary question for each task is whether speech style influenced participants' ability to shadow and comprehend that speech.

Shadowing task

The principal measures concerned the effects of speech condition on speed and accuracy of shadowing responses. We first analyzed the quantity and types of errors that were made. We next examined the speed of participants' speech shadowing by measuring the latency of their productions relative to the stimulus onset.

Shadowing errors Participants' shadowing was highly accurate overall but they made a larger proportion of errors in the Dissimilar Speech Condition, M=.09, than in the Similar Speech Condition, M=.06. Although participants were more likely to make errors for academic speech passages, M=.10, than informal speech passages, M=.06, this difference was larger for those in the Similar Speech Condition (academic: M=.08 of words; informal: M=.04 of words) than in the Dissimilar Speech Condition (academic: M=.10 of words; informal: M=.08 of words). Model comparisons revealed a main effect of speech condition, b=.34, 95% CI[.02, .67]; $X^2(1)=11.46$, $p<.001$, such that participants in the Similar Speech Condition were more accurate than participants in the Dissimilar Speech Condition, and a main effect of passage type, b=-.91, 95% CI[-1.14, -.68]; $X^2(1)=8.49$, p=.004, such that participants were more accurate at shadowing academic than informal speech. The interaction between passage type (academic v. informal) and speech condition (similar v. dissimilar) was also significant, b=.59, 95% CI[.43, .75]; $X^2(1)=51.54$, $p<.00001$.

Planned follow-up comparisons revealed that the interaction was driven primarily by participants performance on informal passages, with a highly significant effect of speech condition on informal passages, b=.90, 95% CI[.56, 1.24]; $X^2(1)=23.35$, $p<.001$, but only a marginally significant effect of speech condition on academic passages, b=.34, 95% CI[-.03, .71]; $X^2(1)=3.18$, p=.07. Additional follow-up comparisons revealed that both the participants in the Similar Speech Condition, b=-.34, 95% CI[-.46, -.22]; $X^2(1)=6.22$, p=.02, and those in the Dissimilar Speech Condition, b=-.92, 95% CI[-1.21, -.62]; $X^2(1)=8.39$, p=.004, demonstrated a significant effect of passage type, such that they made fewer errors on academic than informal passages.

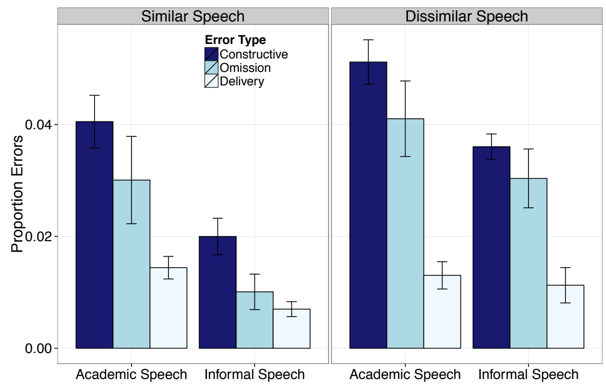

Figure 1. Proportion of errors of each type made in the shadowing task by participants in each speech condition for each passage type. Error bars depict standard error of mean.

We next examined the types of errors that were made. As seen in Figure 1, constructive errors were most frequent, M=.04 of words, followed by omissions, M=.03, and then delivery errors, M=.01. Model comparison revealed a three-way interaction between speech condition, passage type, and error type (constructive, omission, delivery) on participants' accuracy, $X^2(2)=6.15$, p=.05. Follow-up comparisons revealed that this interaction was carried by the interaction between speech condition and passage type, as there was a significant interaction between speech condition and passage type on the number of constructive, $X^2(1)=10.63$, p=.001, omission, $X^2(1)=37.36$, $p<.001$, and delivery errors, $X^2(1)=10.19$, p=.001. This result suggests that participants show a somewhat similar pattern of errors for each speech condition and passage type.

Finally, we examined whether accuracy changed over time because participants were able to adapt to the speakers' voice and improve in their shadowing performance. Model comparisons revealed that the interaction between speech condition and block was not significant, $X^2(1)=2.13$, p=.14, nor was the three-way interaction between speech condition, block, and passage type, $X^2(1)=.40$, p=.53, suggesting little effect of speech style on accommodation.

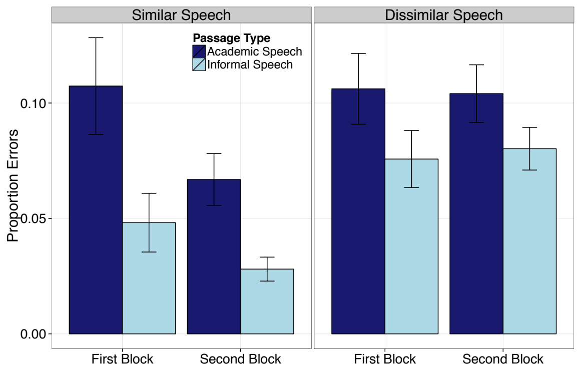

Figure 2. Proportion of errors participants made in each speech condition during shadowing of each passage type depending on whether they did the shadowing task first or second. Error bars depict standard error of the mean.

Shadowing latency In addition to making more errors, participants in the Dissimilar Speech Condition also shadowed more slowly, *M*=1108ms, *SD*=.513; than participants in the Similar Speech Condition, *M*=1032ms, *SD*=480ms (see Figure 3). Similarly, although participants were slower to shadow academic, *M*=1094ms, *SD*=514ms; than informal speech, *M*=1048ms, *SD*=.483; this difference was larger in the Similar Speech Condition (Academic Speech: *M*=1069ms, *SD*=509ms; Informal Speech: *M*=1001ms, *SD*=452ms) than in the Dissimilar Speech Condition (Academic Speech: *M*=1118ms, *SD*=518ms; Informal Speech: *M*=1098ms, *SD*=509ms). Indeed, model comparisons revealed a marginally significant interaction between speech condition and passage type, *b*=.05, 95% CI[-.005, .10]; X^2(1)=3.27, *p*=.07. Planned follow-up comparison revealed that this interaction was primarily driven by participants in the Similar Speech Condition being faster to shadow Informal Speech than Academic Speech, *b*=-.07, 95% CI[-.14, -.01]; X^2(1)=4.34, *p*=.037. There were no other significant effects.

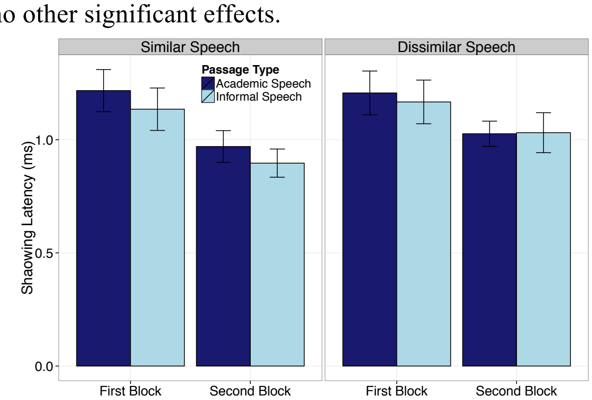

Figure 3. Latency of participants' shadowing of speech of each passage type depending on whether they did the shadowing task first or second. Error bars depict standard error of mean.

We next examined how latencies changed over the course of the task. As can be seen in Figure 3, participants had longer latencies if the shadowing task occurred in the first block, *M*=1169ms, *SD*=557ms; than in the second block, *M*=971ms, *SD*=409ms. This difference was larger for those in the Similar Speech Condition (First Block: *M*=1161ms, *SD*=558ms; Second Block: *M*=930ms, *SD*=378ms) than those in the Dissimilar Speech Condition (First Block: *M*=1176ms, *SD*=556ms; Second Block: *M*=1023ms, *SD*=440ms). However, although there was a main effect of block, *b*=-.24, 95% CI[-.46, -.02]; X^2(1)=6.43, *p*=.011, there was no interaction between speech condition and block, X^2(1)=.30, *p*=.584. Nevertheless, planned follow-up comparisons revealed that the effect of block was driven by the participants in the Similar Speech Condition, as they showed a significant effect of block, *b*=-.24, 95% CI[-.46, -.03]; X^2(1)=4.86, *p*=.028, but those in the Dissimilar Speech Condition did not X^2(1)=2.19, *p*=.14. Thus, participants shadowing speech similar to their own showed better accommodation to the speaker's voice than those shadowing dissimilar speech, as evidenced by shorter latencies as the experiment progressed.

Listening Comprehension task

Listening comprehension was higher in the Similar Speech Condition, *M*=.92, than Dissimilar Speech Condition, *M*=.88 (Fig. 4). Comprehension was higher for informal passages, *M*=.92, than academic passages, *M*=.88. As seen in the figure, participants in both speech conditions were more accurate on the Informal passages compared to the Academic ones: Similar Speech Condition (Academic Speech: *M*=.90; Informal Speech: *M*=.95); Dissimilar Speech Condition (Academic Speech: *M*=.87; Informal Speech: *M*=.89). Model comparisons revealed both marginal effects of speech condition, *b*=-.49, 95% CI[-1.04, .06]; X^2(1)=3.02, *p*=.082, and of passage type, *b*=-.24, 95% CI[-.05, .96]; X^2(1)=3.01, *p*=.083, but no significant interaction between speech condition and passage type, X^2(1)=.84, *p*=.36, on participants' listening comprehension.

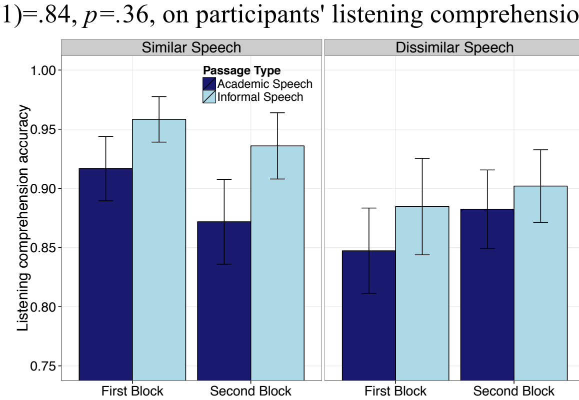

Figure 4. Accuracy in the listening comprehension task for each passage type depending on whether they completed the listening comprehension task first or second. Error bars depict standard error of mean.

We also examined whether listening comprehension changed over the course of the experiment. As can be seen in Figure 4, participants were similar in accuracy regardless of whether they completed the listening comprehension task first or second. Model comparison revealed that there was not a significant interaction between speech condition and block, $X^2(1)=1.62$, $p=.20$, nor was there a main effect of block on listening comprehension, $X^2(1)=.03$, $p=.86$. Thus, participants' listening comprehension abilities did not improve if they had already heard the speaker in the shadowing task first.

Between task comparisons

Finally, we examined the relationships between our three measures—shadowing accuracy, shadowing latency and listening comprehension. First, as can be seen in Figure 5, participants who shadowed more slowly also tended to make more shadowing errors. Comparison of linear regression models revealed that latency was a significant predictor of shadowing errors, $b=.10$, 95% CI[.07, .14]; $X^2(1)=39.76$, $p<.001$. However, there was no interaction between latency and speech condition in predicting the number of speech errors participants made, $X^2(1)=2.04$, $p=.16$. Importantly, as can be seen in Figure 5, this relationship was not significantly different for participants in the two speech conditions, demonstrating that there was not a speed/accuracy tradeoff in shadowing. Rather, speech latency and errors are highly similar measures of shadowing ability.

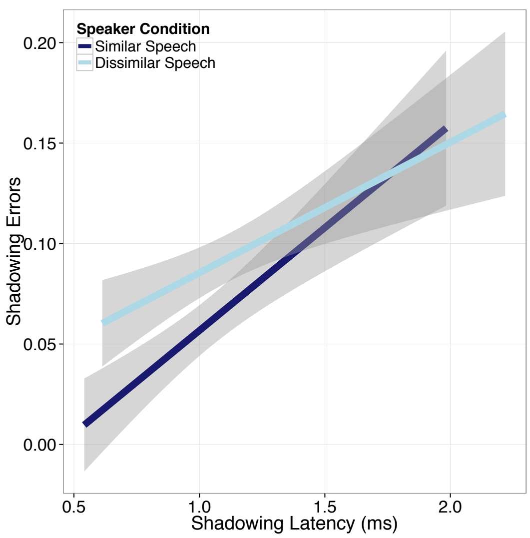

Figure 5. Relationship between shadowing errors and shadowing latency for participants in each speech condition. Error bands depict standard error of mean.

Similarly, as can be seen in Figure 6, participants with better listening comprehension tended to make fewer speech errors in shadowing. Comparison of linear regression models revealed that listening comprehension was a marginally significant predictor of speech errors, $b=-.08$, 95% CI[-.16, .003]; $X^2(1)=3.57$, $p=.06$, but listening comprehension and speech condition did not interact, $X^2(1)=.40$, $p=.53$.

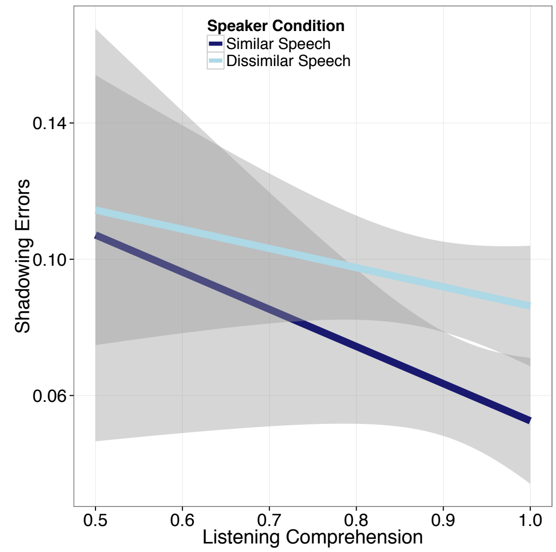

Figure 6. Relationship between shadowing errors and listening comprehension for participants in each speech condition. Error bands depict standard error of mean.

Importantly, there was a main effect of speech condition on shadowing errors, even for when both shadowing latency and listening comprehension were included as covariates, $b=.02$, 95% CI[.007, .04]; $X^2(1)=7.39$, $p=.008$, suggesting that differences in the shadowing task are not driven by individual differences in language ability. Instead, a critical predictor of participants' ability to closely shadow another person's speech, is whether or not that speech is similar to their own speech.

Discussion

In this study we asked whether the familiarity of a talker's speaking style affected listeners' ability to comprehend and closely shadow it. We found that participants were more likely to make more errors and lag further behind during shadowing and to make more comprehension errors for Dissimilar Speech than Similar Speech. The impact was larger for more informal speech than academic speech as informal speech allows for greater variation in speaking style than the more constrained academic prose. Participants make more errors and had longer shadowing latencies for Dissimilar Speech than Similar Speech, and also demonstrated less adaptation over the course of the task. This low level of adaptation is somewhat surprising given that listeners do accommodate slightly to speakers with unfamiliar accents in other studies (e.g., Maye et al., 2008). These effects are consistent with findings regarding another form of accommodation, syntactic alignment, namely the degree to which a listener subsequently uses the same sentence structures as a speaker. Weatherholtz, Campbell-Kibler & Jaeger (2014) found that participants' degree of syntactic alignment to recorded speech varied with the perceived "standardness" of the speaker's accent and perceived similarity to the participant's own speech. In our own task, speech in the Academic passage is arguably more standard than the Informal passage regardless of speech similarity. This increased standardness could explain why participants' shadowing of Academic speech was relatively unaffected by speech similarity.

Our study is the first to our knowledge to compare listeners' ability to closely shadow and comprehend speech of speakers who both were native speakers of the listeners' language but varied in regional accent and perceived similarity. By comparing listening in these two contexts we are taking an important first step in understanding how listeners process speech that is different from their own.

In the current study, our goal was to examine differences in speech style holding the content of the material constant. The overall differences between the speakers were due to a variety of individual and group factors that were not the focus of the experiment. One area for future research will be to replicate these effects with recordings from additional speakers, in order to distinguish speaker-specific and more regional factors in speech familiarity. Another important direction for future research will be to examine the impact of differences in the speech properties we examined when other factors such as dialect also vary. The comprehension of dialects may also depend on their similarity to the speaker's own. The combination of less familiar dialect and speaking style may create greater difficulties.

Future research should also consider how social information modulates participants' processing of similar and dissimilar speech. Previous research has noted a role for social influence on speech perception (Babel, 2010; Casasanto, 2008; Kinzler, Dupoux, & Spelke, 2007; Weatherholtz et al., 2014). It will be interesting to further explore how social attitudes towards people from different racial groups and regional groups influence individuals' abilities to perceive and comprehend their speech. Given that race and region were both manipulated and correlated with each other in the current study, it is possible that participant attitudes may have been contributing to processing difficulties.

Conclusions

Together our results demonstrate that even when two speakers speak the same language, differences in speaking style can create difficulties in processing at multiple timescales.

Acknowledgments

This study was supported by NICHD R24D075454. We thank Sarah Isaak for help in data collection and coding.

References

Adank, P., Evans, B. G., Stuart-Smith, J., & Scott, S. K. (2009). Comprehension of familiar and unfamiliar native accents under adverse listening conditions. *Journal of Experimental Psychology: Human Perception and Performance*, *35*(2), 520–529.

Babel, M. (2010). Dialect divergence and convergence in New Zealand English. *Language in Society*, *39*(04), 437–456.

Casasanto, L. S. (2008). Does social information influence sentence processing? *Proceedings of the 30th Annual Conference of the Cognitive Science Society*, *BC Love, K McRae, & VM Sloutsky*, 799–804.

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.

Jaeger, T. F. (2009). Random effect: Should I stay or should I go? Retrieved February 8, 2013, from http://hlplab.wordpress.com/2009/05/14/random-effect-structure/

Kinzler, K. D., Dupoux, E., & Spelke, E. S. (2007). The native language of social cognition. *Proceedings of the National Academy of Sciences*, *104*(30), 12577–12580.

Major, R. C., Fitzmaurice, S. M., Bunta, F., & Balasubramanian, C. (2005). Testing the Effects of Regional, Ethnic, and International Dialects of English on Listening Comprehension. *Language Learning*, *55*(1), 37–69.

Marslen-Wilson, W. D. (1973). Linguistic structure and speech shadowing at very short latencies. *Nature*, *244*(5417), 522–523.

Marslen-Wilson, W. D. (1985). Speech shadowing and speech comprehension. *Speech Communication*, *4*(1), 55–73.

Maye, J., Aslin, R. N., & Tanenhaus, M. K. (2008). The Weckud Wetch of the Wast: Lexical Adaptation to a Novel Accent. *Cognitive Science*, *32*(3), 543–562.

Nye, P. W., & Fowler, C. A. (2003). Shadowing latency and imitation: the effect of familiarity with the phonetic patterning of English. *Journal of Phonetics*, *31*(1), 63–79.

Sabatini, E. (2000). Listening comprehension, shadowing and simultaneous interpretation of two "non-standard" English speeches. *Interpreting*, *5*(1), 25–48.

Sumner, M. (2011). The role of variation in the perception of accented speech. *Cognition*, *119*(1), 131–136.

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.

Weatherholtz, K., Campbell-Kibler, K., & Jaeger, T. F. (2014). Socially-mediated syntactic alignment. *Language Variation and Change*, *26*(03), 387–420.
White, K. S., & Aslin, R. N. (2011). Adaptation to novel accents by toddlers. *Developmental Science*, *14*(2), 372–384.