Dialect Classification and Speech Intelligibility in Noise

Undergraduate Research Thesis

Presented in Partial Fulfillment for the Requirements for graduation with "Honors Research Distinction in Linguistics" in the undergraduate colleges of The Ohio State University

by

Megan Dailey

The Ohio State University

April 2018

Project Advisor: Professor Cynthia Clopper, Department of Linguistics

Abstract

Listeners have good intuitions about regional differences between speakers. Less studied is the relationship between dialect classification and speech intelligibility. Regional dialect is known to affect speech intelligibility: in some cases, familiar dialects can facilitate speech processing, but in other cases, familiar dialects inhibit speech processing. Many other lexical, discourse, and social factors also affect speech intelligibility. The present study explores predictors of dialect classification accuracy and intelligibility accuracy for the Northern and Midland dialects of American English. In previous studies, Midland speech in noise was found to be more intelligible than Northern speech in noise for both Midland and Northern listeners. Given that listeners are sensitive to differences between Northern and Midland speech in terms of intelligibility, we might expect that listeners use differences in intelligibility to identify where talkers are from. To explore this possibility, participants completed a speech intelligibility in noise task followed by a dialect classification task with Northern and Midland speech. Stimulus materials were balanced for lexical, discourse, and social factors that affect speech intelligibility. A linear regression model revealed that mean token intelligibility did not predict token classification accuracy. The results reveal that easily classifiable tokens are not always the least intelligible. These results suggest that processing mechanisms adapt for regional variation, but dialect-specific cues are not always available to the listener in classification. Predictive models of intelligibility and classification revealed that social, lexical, and discourse factors interact to affect accuracy on both tasks. Ongoing analysis will more precisely determine the role of each predictor in each task, leading towards a better understanding of the relationship between dialect perception and processing mechanisms.

Introduction

Listeners are sensitive to regional dialect variation in the speech signal. On the one hand, dialect classification tasks have shown that listeners can use dialect-specific acoustic cues to classify talkers by dialect (Clopper & Pisoni, 2004, 2006, 2007; Williams, Garrett, & Coupland, 1999). On the other hand, speech intelligibility in noise tasks have revealed that regional variation can facilitate or inhibit speech processing (Mason, 1946; Labov & Ash, 1997; Clopper & Bradlow, 2008). Less is known about the relationship between the perception and processing of regional variation. Because dialect-specific cues used for dialect classification are perceived in the context [of], a relationship between perception and processing is hypothesized such that easy-to-identify dialect-specific forms are harder to process, and vice versa. The hypothesis is tested in a two-part experiment consisting of a dialect classification task and a speech intelligibility in noise task of speech stimuli from two Midwestern dialects of American English. Further, acoustic reduction has been found to mediate both the production of dialect-specific forms and listener accuracy on speech intelligibility in noise tasks (Luce & Pisoni, 1997; Picheny, Braida, & Durlach, 1985; Kalikow, Stevens, & Elliott, 1977; Bradlow, Torretta, & Pisoni, 1996; Labov & Ash, 1997; Clopper & Bradlow, 2008). To gain a clearer understanding of the relationship between the two tasks, lexical, discourse, and social factors known to influence acoustic reduction are operationalized as predictor variables in two linear regression models predicting accuracy on both classification and intelligibility. The present study uses multiple-factor models to explore interactions of as many predictors of classification and intelligibility as possible to identify which factors are most important to perceptual and processing mechanisms in online processing. Finally, the analysis initiates a comparison of the models that serves to motivate future hypotheses about the relationship between dialect perception and speech processing.

Regional and social dialect variation has been documented for decades, if not centuries, and can take many forms (Fees, 1991). Much early documentation of dialect variation in American English focused on lexical differences between groups of speakers (Atwood, 1953; Allen, 1954; Kurath, 1967), but phonological variation has also been widely documented (Kurath, 1961; Labov, 1972). Labov's (1972) influential study on the use of /ɹ/ in New York City department stores inspired interest in the study of systematic phonological variation across social groups. Phonological variation is most commonly used to delineate regional dialect boundaries in the recent literature, at least in American English (Labov, Ash, & Boberg, 2005). Labov et al.'s (2005) linguistic atlas of American English is constructed from an analysis of acoustic-phonetic differences observed in telephone interview recordings of more than 600 speakers across the country. The atlas serves as an important reference point from which commonly-accepted dialect regions grounded in phonological variation are derived. These dialect boundaries can be used to observe listeners' sensitivity to regional acoustic-phonetic differences and to explore their ability to exploit these differences to classify speakers by dialect.

There is overwhelming evidence from the perceptual dialect classification literature that listeners are sensitive to regional and social differences between speakers. While listeners could plausibly use lexical variation to identify the dialect of a speaker, more recent work has sought to determine listeners' sensitivity to acoustic-phonetic variation between speakers that stems from regional background differences (Clopper & Pisoni, 2004, 2006, 2007; Williams, Garrett, & Coupland, 1999). One of the first dialect classification studies that presented listeners with recorded speech stimuli of speakers from different regional backgrounds was Williams et al. (1999). The researchers presented casual spontaneous stories of speakers from six regional dialects of English in Wales and speakers of standard British English to adults and teenagers in all six corresponding regions. In a forced-choice task, participants identified whether the speaker was from one of the six regions, from England, or if their region of origin was unknown. The teenagers broadly identified three dialect regions of English in Wales, and their systematic confusion patterns suggested more fine-grained mental representations of the six dialects.

In American English (AE), Clopper and Pisoni (2004) conducted a similar task in which participants from Indiana listened to speakers from six distinct dialect regions (New England, North, North Midland, South Midland, South, and West) reading the same two sentences. Participants chose which of the six regions each speaker was from. Results showed that listeners were able to categorize speakers into three broad dialect regions (New England, South, North/West) with accuracy above chance. The researchers conducted a similar study, this time a free-classification task, in which listeners organized controlled speech stimuli from speakers of six regional dialects into their own groups (Clopper & Pisoni, 2007). Results showed that listeners made more groups than predicted that varied along dimensions of gender, linguistic markedness, and geography, revealing that listeners perceive more-detailed regional dialect differences than were defined for them in the groups they were given in a related forced-choice classification task (Clopper & Pisoni, 2006). Taken together, the results from these three studies provide strong evidence that listeners can use phonological and acoustic-phonetic variation to determine the regional background of a speaker.

However, listeners are not equally sensitive to all phonological variation. Several recent studies demonstrate cases in which acoustic-phonetic information in the speech signal is not available for or not used by listeners to classify speakers by dialect. Niedzielski (1999) observed that listeners from Detroit, Michigan did not perceive vowels that are characteristic of Northern Midwestern speech, such as that spoken in Detroit, in a fellow speaker from Detroit. But when listeners heard the same speaker and were told that she was Canadian, they were able to perceive the shifted vowels. Similarly, in their free-classification task described above, Clopper and Pisoni (2007) found that listeners from the Northern dialect region categorized speakers of their own dialect more similarly to Midland speakers than Midland listeners did, providing further evidence that Northern listeners in particular, and more generally Midwestern listeners, are not sensitive to the phonological differences between Northern vowels and Midland vowels. This phenomenon has been confirmed in follow-up investigations (Clopper & Bradlow, 2008; Jones et al., 2017) and is puzzling, particularly because the acoustic-phonetic differences between the sets of vowels are not so small as to be undetectable. In particular, Northern vowels take part in a characteristic chain shift, known as the Northern Cities Vowel Shift (Labov, 1998), in which the raising of the low front vowel /æ/ is hypothesized to have initiated a sequential shift of vowels that maintains perceptual and acoustic space between vowel phonemes. Midland vowels do not exhibit a similar shift, but instead are characterized by back-vowel fronting and a low-back vowel merger known as the cot/caught merger. Vowel formant spaces of the Northern and Midland dialects marked for dialect-specific vowel phoneme variation are shown in Figure 1. The goal of the current project was to provide a new analysis of this mismatch between perception and production by comparing predictors of Midwesterners' performance on a dialect classification task to predictors of their performance on a speech intelligibility in noise task with Northern and Midland speech. Because Midwestern listeners do not seem to have access to Midland and Northern regional differences in dialect classification, it is of interest to examine whether they are sensitive to this same regional variation during speech intelligibility in noise.

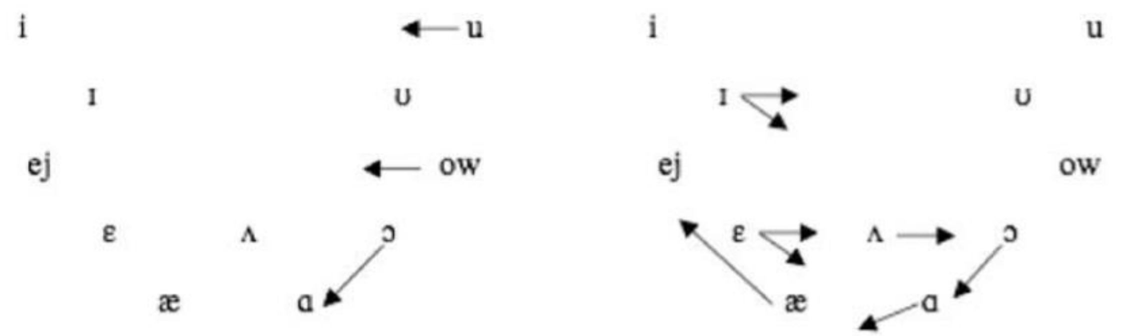

Figure 1. Vowel systems of the Midland (left) and Northern (right) dialects. From Clopper et al. (2010).

The role of regional dialect in speech intelligibility in noise is similarly variable. For example, local dialects can facilitate speech processing (Mason, 1946; Labov & Ash, 1997). In a transcription task, Labov and Ash (1997) found that listeners from Birmingham, Alabama could more quickly and accurately identify the speech of speakers from Alabama than listeners from Chicago or Philadelphia. Familiar, non-local dialects can also facilitate speech processing (Clopper & Bradlow, 2008). In a speech intelligibility in noise task, Clopper and Bradlow (2008) observed that a collection of perceptually similar and relatively unmarked dialects they labelled General American English (i.e., non-rhotic New England, Midland, and Western dialects) were more intelligible than Southern and Mid-Atlantic dialects in noise for listeners from various regions in the U.S. Specifically, Northern listeners found Midland speech easier to process in noise than the speech of their own dialect.

Clopper and Bradlow (2008) acknowledged the potential problem that these two separate sets of anomalies observed in Northern listeners pose. In speech intelligibility in noise, Northern listeners have a harder time processing Northern speech in noise than they do Midland speech. Therefore, processing mechanisms are sensitive to Northern/Midland acoustic-phonetic differences because they have a harder time recovering Northern forms, at least from a degraded signal. But in a dialect classification in noise task, Northern listeners did not perceive differences between Northern and Midland talkers, and could not distinguish between the Northern Cities Shifted-vowels that they themselves presumably produce and Midland vowels. These results suggest that in dialect classification tasks, listeners don't have access to the same dialect-specific information in the speech signal that they do in speech processing. This mismatch between processing and perception merits further investigation if a relationship between the two processes is to be identified.

The mismatch between processing and perception motivates a hypothesis about the relationship between the two that draws upon the findings summarized above. In speech intelligibility in noise tasks, unfamiliar dialects and more linguistically marked dialects inhibit speech processing (Labov & Ash, 1997; Clopper & Bradlow, 2008). That is, forms that deviate from an expected norm or stored representation are harder to process than forms that do not. In dialect free-classification tasks, listeners categorize speakers along axes of linguistic markedness, or how many different phonological variants are present in a dialect (Clopper & Pisoni, 2007). That is, forms that deviate from an expected norm or stored representation are easier to identify than forms that are not. Such observations suggest that a linear relationship between the processing of regional variants and dialect classification exists: forms that are easy to process are hard to identify, and v-ice versa. If such a relationship exists, it would suggest that stored lexical representations do not match unfamiliar or linguistically marked dialect forms, so mapping these forms is harder than mapping familiar or linguistically unmarked forms to stored representations. If a linear relationship does not exist between the two tasks, i.e. if easy-to-identify, linguistically marked forms are not always the hardest to process, that would suggest some adaptation for dialect-specific forms in processing, which could be specified in the lexical representation.

This hypothesis is what the present study seeks to test. Participants completed both a speech intelligibility in noise task and a forced-choice dialect classification task in which they transcribed and classified Northern and Midland American English speech. Crucially, stimulus materials were identical across tasks for all participants so that direct, token by token comparisons could be made to better understand the relationship between the two tasks. If a linear relationship of listener performance on the two tasks can be observed, like one described above, this result would account for Midwestern listeners' difficulty in perceiving Northern vowels. However, in the event of poor performance on the dialect classification task, as has been previously observed for Midland and Northern speech (Clopper & Bradlow, 2008), an analysis of the correlation between intelligibility and classification still may not resolve the source of the perceptual difficulty that Midwesterners encounter in explicit classification of the two dialects. For this reason, a secondary analysis of predictors of performance on both tasks could reveal additional similarities between the factors that influence classification and intelligibility, from which more can be understood about the mismatch between dialect perception and speech processing.

To examine factors that affect listener performance on a dialect classification task, it is necessary to understand what is required of listeners to successfully identify the dialect of a speaker. First, listeners must have adequate exposure to regional dialect variation to be able to distinguish what marks regional differences between talkers. Clopper and Pisoni (2006) explored listener experience with regional variation in a forced-choice classification task of American English dialects by balancing listener groups for region of origin and geographic mobility. They found that listeners confused dialects that were closer to their region of origin more often than dialects that are not, but listeners who had lived in multiple different regions did so less often. It seems, then, that the level of listener experience matters for successful classification, as listeners who have more experience with the dialects being presented to them were more successful in classification.

In addition to having experience with regional variation, for successful dialect classification performance listeners must be able to identify dialect-specific acoustic cues in the speech signal during processing. The most studied dialect-specific acoustic cues in American English are differences in the realization of vowel phonemes (Labov et al., 2005), with much less known about regional prosodic and other acoustic cues (Clopper & Smiljanic, 2011). While listeners could be relying on acoustic cues not previously identified by researchers, the present study controls for lexical variation, but all other sources of acoustic-phonetic regional variationare presumed to be available and relevant to listeners in both tasks.

Not only do listeners have to be sensitive to dialect-specific cues, but for successful dialect classification, they must be able to attribute these cues to regional background differences between speakers, and subsequently map the correct cues onto the correct commonly-accepted dialect groups. Williams et al. (1999) pointed out that a knowledge of the extent to which listeners are aware that regional variation is in fact regionally-motivated is a key component to understanding both their performance in dialect classification and their mental representations of regional variation. But acoustic cues in the speech signal are not the only factors that affect how listeners classify speakers by dialect, which may explain why explicit, forced-choice classification can be hard for listeners.

Dialect labels can also affect dialect classification performance by provoking listener bias. In a dialect free-classification task using the same stimulus materials as Clopper and Pisoni’s (2006) forced-choice task (Clopper & Pisoni, 2007), listeners had different patterns of confusion: New England speech was confused with other relatively unmarked dialects. This finding led the researchers to the conclusion that the New England label elicited a response bias from the participants in the forced-choice task. Niedzielski's (1999) results cannot be attributed to anything but listener bias, as the label of the speaker was the only factor manipulated in the task. However, there aren't strong identities associated with being "Northern" vs. being "Midland", and neither dialect is collectively considered more prestigious than the other (Campbell-Kibler, 2012). Therefore, a listener bias is not expected in the classification task in the present study. Nevertheless, the results from the two studies together suggest that mapping their own representations of regional variation onto pre-existing labels is harder for listeners than using acoustic cues in the speech signal to define their own perceptually salient groups.

Williams et al. (1999) preferred an affective account to explain listener bias and resolve some of the perceptual confusions in their data. They observed that listeners that received higher "likeability" ratings were the most claimed, i.e., the most likely to be classified to the listener's own dialect. They proposed that group-level processes in which listeners attribute in-group or out-group characteristics to speakers significantly influenced how the speakers were classified by dialect. In a similar way, a lack of affective in-group/out-group attribution could mask acoustic differences between Northern vowels and vowels of other dialects of American English for Northern listeners.

It is equally important to consider what is required of the listener in a speech intelligibility in noise task to determine which factors affect performance on the task. The primary focus of the listener in a speech intelligibility in noise task is to identify words in noise, or to correctly map a degraded signal onto stored lexical representations. It follows that any sources of variation in the speech signal that affect how degraded the signal is or that would affect a word's deviation from a stored representation would influence listener performance in speech intelligibility in noise. So, both acoustic reduction, or signal degradation that is measured both temporally and in terms of dispersion and intensity, and regional and social variation, which affect a word's deviation from its stored representation, affect performance on speech intelligibility in noise. As noted above, acoustic reduction also affects the realization of dialect-specific cues in the speech signal. In this way, factors that mediate acoustic reduction could affect both speech intelligibility in noise and dialect classification. Previous work on factors that mediate acoustic reduction, how acoustic reduction affects the realization of regional variation, and how both acoustic reduction and regional variation affect speech intelligibility in noise are considered below.

Researchers have found many lexical and discourse factors that mediate acoustic reduction. Word frequency and phonological neighborhood density have been found to affect acoustic reduction: words that are highly frequent are more reduced than words that are less frequent, while words that have few phonological neighbors, or low phonological density, are reduced relative to high density words (i.e. Munson & Solomon, 2004). Discourse factors that affect reduction include style, order of mention in a conversation, and predictability. Words in a plain speech style are reduced relative to words in a clearer speech style (Picheny, Durlach, & Braida, 1986). Words that are mentioned for the second time in a discourse are reduced relative to words mentioned for the first time (Fowler & Housum, 1987). Words that are highly predictable in their semantic context are reduced relative to words that are not predictable (Lieberman, 1963).

Further, acoustic reduction affects the realization of dialect-specific forms, and this effect can vary across dialects, resulting in complex interactions in production (Clopper &

Pierrehumbert, 2008; Clopper, Mitsch, & Tamati, 2017). Clopper and Pierrehumbert (2008) found that in words that were highly predictable (more acoustic reduction), Southern vowels were produced with less dispersion, Northern vowels were produced more dispersion, or more extreme Northern Cities-Shifting, and Midland vowels were unaffected by the predictability of their carrier words in context. Therefore, factors such as semantic predictability that mediate acoustic reduction influence the realization of the dialect-specific cues that listeners need to identify speakers in a dialect classification task. Discourse factors tied to the degree of reduction produced by a speaker also affect the realization of dialect-specific cues during speech production. For example, speech context can affect how often dialect-specific forms are produced (Labov, 1972). Labov (1972) found that in more casual contexts, more dialect-specific forms are produced, and in more formal contexts fewer dialect-specific forms are produced. Taken together, the observations of both studies suggest that dialect-specific cues may not always be available to the same degree for the listener to identify the dialect of a speaker.

Lexical and discourse factors that mediate acoustic reduction have been shown to influence speech intelligibility in noise. For example, even though high frequency and low density words are more reduced, they are more intelligible in noise than low frequency and high density words (Luce & Pisoni, 1997), suggesting that processing mechanisms can overcome the difficulties presented by acoustic reduction. Similarly, predictable words are more intelligible than unpredictable words in noise, even if they are more reduced, when they are presented in context (Kalikow, Stevens, & Elliott, 1977). Picheny, Braida, and Durlach (1985) found that conversational speech was less intelligible for hearing-impaired listeners than speech spoken in a clear style, suggesting that the degree of reduction can affect speech processing.

Finally, social variation affects speech intelligibility in noise, either facilitating or inhibiting speech processing. Female speech has been found to be more intelligible in noise than male speech (Bradlow, Torretta, & Pisoni, 1996), suggesting that female speech is easier to process than male speech. As has been previously discussed, regional dialect is another social-indexical source of variation that affects speech intelligibility in noise in complex ways. Local dialects can facilitate speech processing, but not always (Labov & Ash, 1997; Clopper & Bradlow, 2008). It is again hypothesized that dialect-specific forms that are less intelligible in noise, and thus harder to process, deviate from an expected norm or stored representation.

These lexical, discourse, and social factors interact to mediate both perception and production. For example, Clopper and Bradlow (2008) found that female speakers of less linguistically-marked dialects were more intelligible than male speakers of the same dialects, but female speakers of more marked dialects, such as the Northern and Southern dialects, were less intelligible than their male counterparts. Further, Burdin, Turnbull, and Clopper (2014) found that lexical frequency, phonological density, speech style, discourse mention, and semantic predictability interacted in the production of acoustic reduction. These observations of interactions between the relevant lexical, discourse, and social factors are what drive the motivation in the current study to include all of them in a predictive model of performance on both intelligibility and classification.

To model the predictors of performance on intelligibility and classification in the present study, the stimuli were balanced for the linguistic, discourse, and stimulus factors known to mediate acoustic reduction and influence speech intelligibility in noise. Because acoustic reduction affects both the processing of dialect-specific forms and the retrieval of dialect-specific cues for dialect identification, the factors responsible for mediating acoustic reduction were expected to affect both dialect classification and speech intelligibility. Balancing for these factors allowed for an analysis of intelligibility and classification performance that investigates which predictors they share and which they do not. Because the goal of the listener is different across the two tasks, the predictors of intelligibility performance are not expected to identically map onto classification performance. The ensemble of analyses in the present study, including examinations of intelligibility performance, classification performance, the relationship between the tasks and their predictors, will allow for a more in-depth exploration of the previously observed mismatch between intelligibility in noise and dialect classification of the Northern and Midland dialects. Moreover, the project will lead to a better understanding of both the processing of regional variation and dialect perception, as well as the relationship between them.

Methods

Materials

The stimuli in both tasks were 117 English phrases of three to seven words in length extracted from short stories from the Ohio State Stories Corpus (Burdin, Turnbull, & Clopper, 2014). The selected stimuli were extracted from passages read by eight talkers (four male, four female) each from the Northern and Midland dialect regions, for a total of 16 talkers. The dialect regions are shown on the map in Figure 2. Four stimulus lists were created of the 117 English phrases for a total of 468 stimulus tokens. The stimuli were balanced for social, linguistic, and discourse factors that are known to affect speech processing. In particular, phrases within each list were balanced by talker gender, talker dialect, semantic predictability, and speech style (clear or plain). In the clear speech style, talkers were instructed to read the passages as if they were talking to a hearing-impaired or non-native listener, whereas in the plain speech style, talkers were instructed to read the passages as if they were talking to a close friend or family member. Additionally, each phrase contained one target word. Target words in the phrases were balanced within each list for word frequency, phonological neighborhood density, mention in the passage (first or second), and by vowel (/æ/, /ɑ/, /ɔ/, /ɛ/, /i/, and /u/). The same stimulus list was used across tasks for each participant, except that in the intelligibility task, the phrases were mixed with speech-shaped noise at a 0dB signal-to-noise ratio. These intelligibility stimuli had a 500-millisecond noise lead and a 500-millisecond noise tail. The intelligibility phrases were mixed with noise to avoid a ceiling effect in participant accuracy, so that differences in dialect intelligibility could be revealed. Classification phrases were not mixed with noise because performance on these tasks is already poor (Clopper & Pisoni, 2004; Clopper & Bradlow, 2008). The lack of noise in the classification stimuli ensured that participants could rely on non-degraded acoustic cues in the speech signal to determine the dialect of the talker. At least 24 participants, balanced for listener dialect (12 Midland listeners, 12 Northern listeners), heard each stimulus list.

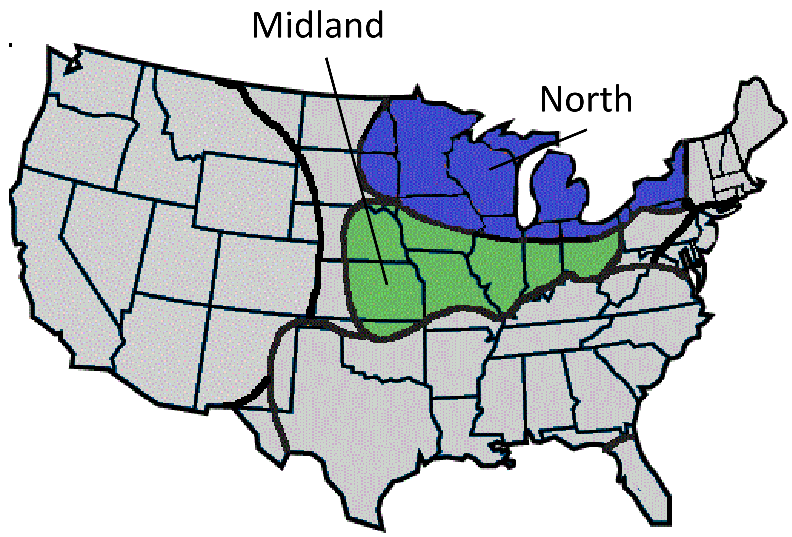

Figure 2. Map of regional dialects of American English, featuring the Northern and Midland dialects.

Participants

Participants were 108 monolingual American English-speaking undergraduates recruited from introductory linguistics courses at The Ohio State University. The median age of the participants was 21.4 years (range: 18-50 years). The participants were balanced by dialect: 54 participants had only lived in the Northern dialect region before the age of 18 and are considered Northern listeners, and the remaining 54 participants had only lived in the Midland dialect region before the age of 18 and are considered Midland listeners.

Procedure

All participants completed a speech intelligibility in noise task followed by a forced-choice dialect classification task. Because phrases across the two tasks were identical for each listener, participants always completed the intelligibility in noise task before the classification task so that the latter did not affect performance on the former. Participants were seated in a quiet room at a computer with a keyboard, button box, and headphones.

In the intelligibility task, participants were presented with the phrases mixed with noise one at a time over headphones, in randomized order, at a comfortable listening level. Participants were instructed to type what they heard into a box on the computer screen. Responses were scored for target word accuracy.

In the classification task, participants were presented with the same set of phrases without noise one at a time over headphones, in a different randomized order, at a comfortable listening level. While each phrase played, they were presented with a map similar to that in Figure 2 on the computer screen and were instructed to select whether the talker was from the Northern or the Midland dialect region. Participants made their responses using colored buttons on a button box (blue for North and green for Midland). Responses were scored for accuracy.

Results

Effects of talker and listener dialect on intelligibility

Table 1 shows the mean accuracy scores for the intelligibility task by talker dialect and listener dialect, aggregated over subject means. To determine whether there was an effect of talker dialect, listener dialect, or an interaction between talker and listener dialect on intelligibility accuracy aggregated over subject means, a repeated-measures ANOVA on target word accuracy was conducted with a within-subjects factor of talker dialect and a between-subjects factor of listener dialect. The ANOVA confirmed that participant accuracy was significantly higher for Midland target words than for Northern target words ($F(1, 105) = 112.79$, $p < 0.001$), suggesting a processing benefit for Midland phrases over Northern phrases. The main effect of listener dialect ($p = 0.080$) and the interaction ($p = 0.498$) were not significant.

To examine whether the effect of talker dialect and the lack of effect of listener dialect and their interaction would remain consistent across samples of phrases, a by-items analysis was

Table 1. Mean intelligibility accuracy scores by talker and listener dialect, aggregated by subject means. Standard deviations are shown in parentheses.

	Talker dialect	
Listener dialect	Midland	Northern
Midland	0.70 (*0.11*)	0.62 (*0.08*)
Northern	0.72 (*0.08*)	0.64 (*0.08*)

performed on the intelligibility accuracy data. Means were aggregated over phrase tokens. A repeated-measures ANOVA with a within-items factor of listener dialect and a between-items factor of talker dialect revealed main effects of talker dialect ($F(1, 466) = 8.94$, $p < 0.01$) and listener dialect ($F(1, 466) = 11.60$, $p < 0.001$) on intelligibility accuracy. The by-items analysis supported the talker dialect effect found in the by-subjects analysis: Midland phrases were more intelligible than Northern phrases. The listener dialect effect that emerged in the by-items ANOVA reveals that Northern listeners performed significantly better on the phrase intelligibility task than Midland listeners. As shown in Table 1, the effect is small, consistent with the lack of a significant listener dialect effect in the by-subject analysis. It is possible that the Northern listeners' greater exposure to Northern speech and current residence in the Midland dialect region gave them a small advantage over the Midland listeners in the intelligibility task.

Effects of talker and listener dialect on classification

Table 2 shows mean classification accuracy scores by talker and listener dialect, aggregated over subject means. To determine whether there was an effect of talker dialect, listener dialect, or an interaction between talker and listener dialect on classification accuracy, a repeated-measures ANOVA on dialect classification accuracy was conducted with a within-subjects factor of talker dialect and a between-subjects factor of listener dialect. The ANOVA revealed that participant accuracy scores were significantly higher for Midland phrases than for Northern phrases ($F(1, 105) = 42.96$, $p < 0.001$). There was no main effect of listener dialect on classification accuracy. There was a marginal interaction between talker dialect and listener dialect on classification accuracy ($F(105) = 3.80$, $p = 0.054$). To explore this interaction further, two independent-sample t-tests were conducted in which classification accuracy for each talker dialect was compared between listener groups. Results showed that the Midland and Northern listener groups did not perform differently from one another in classification of either Midland or Northern phrases, ($t(105)=1.68$, $p = 0.095$ and $t(105)=-1.58$, $p = 0.118$, respectively).

Table 2. Mean classification accuracy scores by talker and listener dialect, aggregated by subject means. Standard deviations are shown in parentheses.

	Talker dialect	
Listener dialect	Midland	Northern
Midland	0.57	0.44
	(*0.10*)	(*0.10*)
Northern	0.54	0.47
	(*0.09*)	(*0.08*)

Given the relatively poor performance in the classification task overall, a series of one-sample t-tests was conducted to determine whether the mean classification accuracy scores by talker dialect and listener dialect were significantly above chance. Results showed that both Midland and Northern listeners classified Midland phrases significantly above chance ($t(52) = 5.40$, $p < 0.001$ and $t(52) = 3.50$, $p < 0.001$, respectively). Additionally, both Midland and

Northern listeners classified Northern phrases significantly below chance ($t(52) = -4.10$, $p < 0.001$ and $t(52) = -2.30$, $p = 0.026$, respectively). Because Midland phrases were classified significantly above chance but Northern phrases significantly below chance, it is likely that there was a response bias for Midland phrases in the classification task. The results from the independent-sample t-tests support that both Midland and Northern listeners had a Midland response bias, as both listener groups performed similarly in classification of both Midland and Northern phrases. The response bias for Midland phrases and the poor mean classification accuracy scores overall further suggest that participants from both dialect regions found the task difficult and could not accurately classify phrases by talker dialect.

A by-items analysis of the classification data supported the by-subject analyses. A repeated-measures ANOVA with a within-items factor of listener dialect and a between-items factor of talker dialect conducted on the by-items data revealed a main effect of talker dialect on classification accuracy ($F(1, 466) = 77.6$, $p < 0.001$). No listener effect was observed. An interaction of talker dialect and listener dialect ($F(1, 466) = 6.30$, $p < 0.05$) was observed, and paired-sample t-tests were conducted to explore the interaction further. The tests revealed Midland listeners classified Midland phrases marginally better than Northern listeners ($t(232) = 1.96$, $p = 0.051$). No significant differences were observed between the two groups for the Northern phrases ($t(234) = -1.61$, $p = 0.109$).

To parallel the by-subjects analysis, a series of one-sample t-tests was conducted to determine whether the performance of Midland and Northern listeners was significantly above chance for either Midland or Northern phrases. Results revealed that Midland phrase tokens were classified significantly above chance by both Midland and Northern listeners ($t(232) = 6.79$, $p < 0.001$ and $t(232) = 4.64$, $p < 0.001$, respectively), and Northern phrase tokens were classified significantly worse than chance for both Midland and Northern listeners ($t(234) = -4.20$, $p < 0.001$ and $t(234) = -2.74$, $p < 0.01$, respectively). These results parallel the one-sample t-tests on the by-subjects classification means, and further support a response bias for Midland phrases.

The results of the by-items and by-subjects analyses of the effect of talker and listener dialect on intelligibility and classification are conclusive: Midland phrases are more intelligible than Northern phrases for both Midland and Northern listeners. However, while Midland and Northern listeners classified Midland phrases significantly above chance, they classified Northern phrases significantly below chance, pointing to a response bias for Midland phrases. Poor overall performance on classification suggests that listeners were not able to successfully classify talkers by dialect. The results from the two tasks are puzzling for a few reasons. Northern listeners had lower intelligibility accuracy on Northern phrases, which would not be observed if familiar dialects always facilitated speech processing, but is consistent with previous results (i.e. Clopper & Bradlow, 2008). Further, not only did Northern listeners fall short of successfully classifying speakers by dialect, but they showed a response bias for Midland speakers. If Northern listeners were simply unaware that their own dialect is distinct from the Midland dialect, a Northern response bias might be expected. The results are consistent with the paradox suggested by previous work (Clopper & Bradlow, 2008): while regional dialect variation can pose processing difficulties for listeners, listeners are not able to use those processing differences to successfully distinguish between the dialects in the classification task. These results therefore merit an examination of the correlation between the two tasks to explore any observable relationship between speech processing and dialect identification at the level of individual subjects and/or items.

Relationship between intelligibility and classification

Two linear regression analyses were conducted to examine the relationship between the mean accuracy scores in the two tasks, talker dialect, and listener dialect, separately for subjects and items. In particular, I wanted to examine whether performance on the intelligibility task could predict performance on the classification task to determine whether the ability to process the speech of a speaker from a specific dialect region corresponds with the ability to identify where they are from. It might be expected that higher intelligibility accuracy for Northern phrases would lead to higher classification accuracy of Northern talkers, assuming that listeners with higher intelligibility accuracy for Northern phrases have more experience processing Northern speech. Further, I wanted to examine any interaction between talker and listener dialect to explore processing and perception differences between the two dialects across listener groups.

Results from the linear regression of by-subjects means are shown in Tables 3 and 4, and a scatterplot of the relationship is shown in Figure 3. The dependent variable in the analysis was classification accuracy, and the predictor variables were intelligibility target word accuracy, talker dialect, and listener dialect. Intelligibility means were centered around zero, but are plotted as raw values in Figure 3. Table 3 shows the model output with the Northern dialect as the reference level, or default, for both talker and listener dialect. Thus, the Intercept estimate in Table 3 is Northern listeners' mean Northern phrase classification accuracy, or 0.47. Each subsequent estimate is interpreted with respect to the Intercept. For example, in the third row of Table 3, the estimate for Midland listeners' mean Northern phrase classification accuracy is 0.06 greater than the Intercept, or 0.53. When Listener and/or Talker dialect are not specified, the effect is for Northern listeners and/or talkers, and when Listener and/or Talker dialect are specified, the effect is for Midland listeners and/or talkers. Estimates involving intelligibility are slopes, and those without are intercepts. Table 4 shows the model output with the Midland dialect as the reference level for both talker and listener dialect.

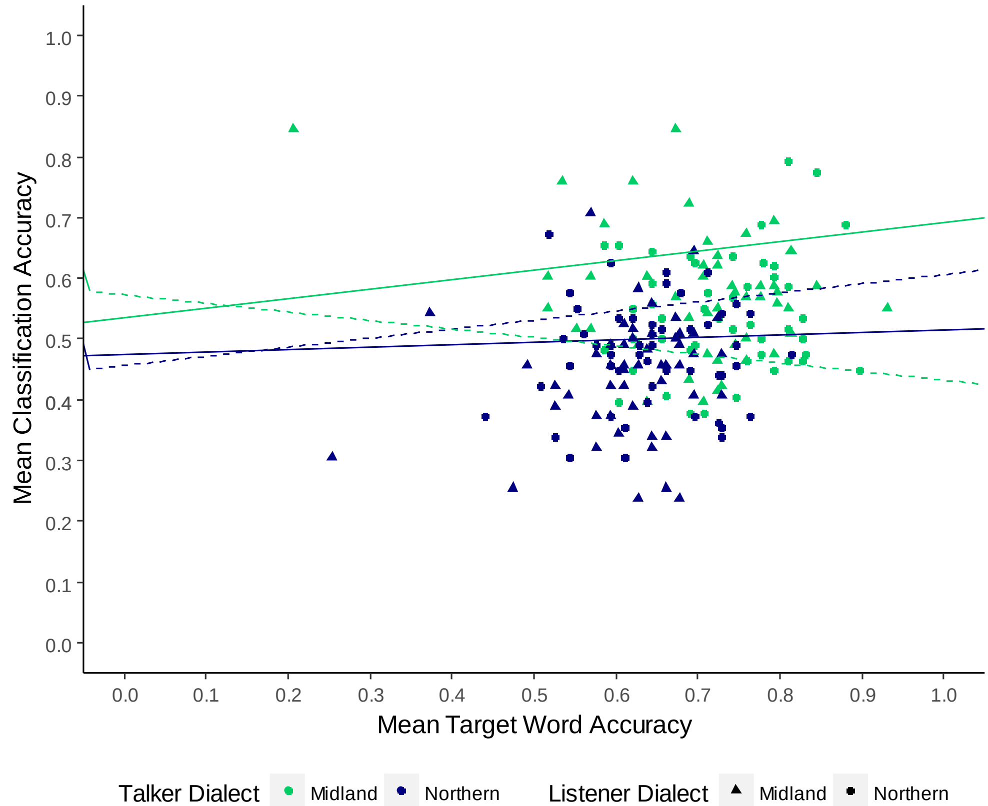

Figure 3. Scatterplot of relationship between intelligibility and classification tasks with regression lines for talker and listener dialect groups. Each data point is one participant's aggregated mean accuracy scores for phrases from one dialect (each participant has a mean Midland and mean Northern data point). Solid lines are Northern listeners, dashed lines are Midland listeners.

The results from the by-subjects linear regression reestablished the previous talker dialect effect on classification accuracy: Midland talker classification accuracy was greater than Northern talker classification accuracy for both Midland and Northern listeners. Midland listeners' classification accuracy of Midland phrases was significantly greater than Northern listeners', signaling a stronger Midland response bias for Midland listeners than for Northern listeners. This effect is consistent with the interaction of talker and listener dialect in the classification accuracy ANOVA and subsequent results of the paired-sample t-test on Midland phrases. Interestingly, the three-way interaction between talker dialect, listener dialect, and intelligibility accuracy was also significant. Lower intelligibility accuracy on the Midland phrases corresponded with higher classification accuracy of Midland phrases for Midland listeners. This result is in the opposite direction of the expected positive correlation between the tasks, which would lead to the conclusion that better processing leads to better perception. Because the analysis of the classification data revealed a response bias for Midland phrases, the Intelligibility effect can be interpreted in such a way that Midland listeners who had a greater Midland response bias in the classification task also had lower intelligibility accuracy on Midland phrases. This effect could be related to participant motivation in the experiment: the same Midland participants who were less focused in the intelligibility task were also more likely to only hit the Midland button in the classification task. No other slopes involving intelligibility accuracy were different from zero, indicating that intelligibility did not predict classification accuracy for any other combinations of listener and talker dialect groups.

Table 3. Linear model of classification accuracy predicted by intelligibility accuracy, talker dialect, and listener dialect aggregated by subject means, with the Northern dialect as the reference level for both talker and listener dialect.

Effect	Estimate	*t*-value	*p*-value
Intercept	0.47	35.24	< 0.001
Intelligibility accuracy	0.04	0.25	0.802
Talker dialect (Midland)	0.06	2.90	0.004
Listener dialect (Midland)	-0.02	-0.82	0.416
Talker dialect x Intelligibility accuracy	0.12	0.51	0.610
Listener dialect x Intelligibility accuracy	0.20	0.89	0.377
Talker dialect x Listener dialect	0.06	2.25	0.025

Talker dialect x Listener dialect x Intelligibility accuracy	-0.65	-2.21	0.029

Table 4. Linear model of classification accuracy predicted by intelligibility accuracy, talker dialect, and listener dialect aggregated by subject means, with the Midland dialect as the reference level for both talker and listener dialect.

Effect	Estimate	*t*-value	*p*-value
Intercept	0.58	44.93	< 0.001
Intelligibility accuracy	-0.30	-2.60	0.010
Talker dialect (Northern)	-0.12	-6.25	< 0.001
Listener dialect (Northern)	-0.05	-2.37	0.019
Talker dialect x Intelligibility accuracy	0.54	2.82	0.005
Listener dialect x Intelligibility accuracy	0.46	2.33	0.021
Talker dialect x Listener dialect	0.06	2.25	0.025
Talker dialect x Listener dialect x Intelligibility accuracy	-0.65	-2.21	0.029

A second linear regression analysis was conducted on the data aggregated by item means, shown in Tables 5 and 6 and visualized in Figure 4. Dependent and predictor variables remained the same, and Intelligibility means were centered around zero. The talker dialect effect held in the by-items data, but the effect of intelligibility accuracy on classification accuracy of Midland phrases by Midland listeners was not observed. Because individual differences in listener performance are ignored in the by items analysis, and individual listener differences in response bias are thought to have driven the intelligibility effect in the by-subject analysis, the by-item collapsing across individual listeners could explain why the same intelligibility effect is not present in the by-items linear model. However, there was a three-way interaction of intelligibility accuracy, talker dialect, and listener dialect on classification accuracy. Specifically, the Northern listener Midland phrase slope was significantly less than the Midland listener Midland phrase slope. After rerunning the model with reference levels of Northern listener dialect and Midland talker dialect, neither slope was found to be significantly different from zero. No other significant slopes involving intelligibility accuracy were observed in either the by- items or by-subjects data, suggesting that performance on intelligibility was not a predictor of performance on classification.

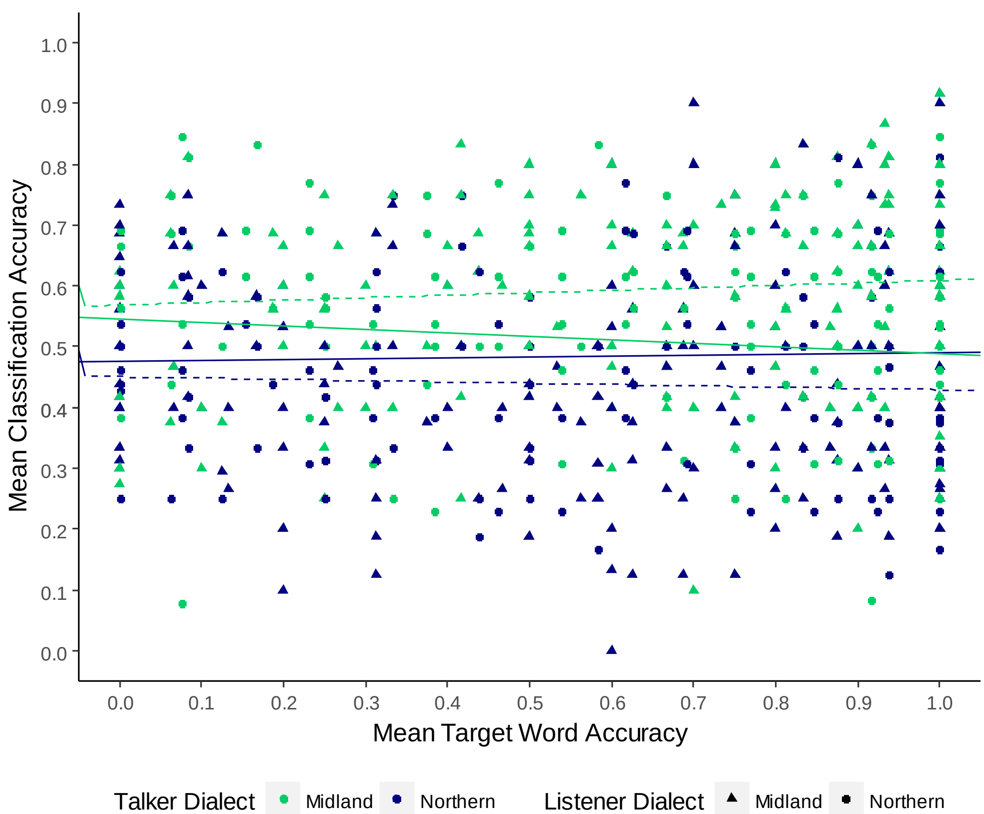

Figure 4. Scatterplot with linear regression of relationship between intelligibility and classification tasks, with regression lines for talker and listener dialect groups. Each data point is one listener groups averaged means of one token. Solid lines are Northern listeners, dashed lines are Midland listeners.

Table 5. Linear model of classification accuracy predicted by intelligibility accuracy, talker dialect, and listener dialect aggregated by item means, with the Northern dialect as the reference level for both talker and listener dialect.

Effect	Estimate	*t*-value	*p*-value
Intercept	0.47	46.91	< 0.001
Intelligibility accuracy	0.01	0.46	0.643
Talker dialect (Midland)	0.07	4.95	< 0.001
Listener dialect (Midland)	-0.02	-1.70	0.089
Talker dialect x Intelligibility accuracy	-0.07	-1.58	0.114
Listener dialect x Intelligibility accuracy	-0.04	-0.83	0.409
Talker dialect x Listener dialect	0.05	2.24	0.026
Talker dialect x Listener dialect x Intelligibility accuracy	0.13	2.07	0.038

Table 6. Linear model of classification accuracy predicted by intelligibility accuracy, talker dialect, and listener dialect aggregated by item means, with the Midland dialect as the reference level for both talker and listener dialect.

Effect	Estimate	*t*-value	*p*-value
Intercept	0.57	55.83	< 0.001
Intelligibility accuracy	0.04	1.18	0.239
Talker dialect (Northern)	-0.12	-8.13	< 0.001
Listener dialect (Northern)	-0.02	-1.46	0.145
Talker dialect x Intelligibility accuracy	-0.06	-1.35	0.177
Listener dialect x Intelligibility accuracy	-0.10	-2.05	0.041
Talker dialect x Listener dialect	0.05	2.24	0.026
Talker dialect x Listener dialect x Intelligibility accuracy	0.13	2.7	0.038

The results could suggest that a listener's ability to classify a speaker by dialect and their ability to understand a speaker are not strongly correlated. This observation runs contrary to the hypothesis that talkers or phrases that are less intelligible are easier to classify by dialect. The results instead suggest that performance on intelligibility and classification are mitigated by separate mechanisms. The clustering of data points in Figure 3 revealed that participants fared similarly to one another on both tasks. However, the dispersion of data points in Figure 4 revealed that there was a considerable amount of variation in performance on individual target words and phrases in both intelligibility and classification. It is useful to investigate what could be causing such considerable variation. In the next sections, I use linear regression models to identify predictors of both intelligibility and classification accuracy, with the goal of investigating the linguistic, sociolinguistic, and discourse factors that interact to affect listeners' processing and perception of the phrase stimuli in these two tasks.

Predicting intelligibility accuracy

A linear regression predicting intelligibility accuracy was conducted to explore the interactions of factors that have been previously shown to affect acoustic reduction and speech intelligibility in noise. Both lexical and discourse factors that mediate acoustic reduction and social factors previously found to affect speech intelligibility in noise were examined in the model. A model that includes as many factors as possible is of particular interest because assessing multiple factors simultaneously could reveal which factors are most important to the listener during speech processing. Intelligibility accuracy was the dependent variable in the analysis, and predictor variables were talker dialect, talker gender, listener dialect, frequency, density, predictability, speech style, and mention. Continuous variables, including frequency, density, and predictability, were centered around zero. The significant effects and interactions of two model outputs are shown in Figures 5 and 6. In Figure 5, the Northern dialect is the reference level for both talker and listener dialect. In Figure 6, the Midland dialect is the reference level for both talker and listener dialect.

Figure 5. Linear model of intelligibility accuracy predicted by eight factors aggregated by items, with the Northern dialect as the reference level for both talker and listener dialect. Only significant main effects and interactions are shown.

Effect	Estimate	Std. Error	t value	Pr(>|t|)	
(Intercept)	0.8001125	0.0701958	11.398	< 2e-16	***
Style (plain)	-0.2865897	0.0891499	-3.215	0.00137	**
Talker dialect (Midland) x Listener dialect (Midland) x Frequency x Predictability x Style (plain)	-2.0034627	0.9586602	-2.090	0.03699	*
Talker dialect x Talker gender (male) x Frequency x Style x Mention	0.5485951	0.2629976	2.086	0.03735	*
Talker dialect x Listener dialect x Frequency x Predictability x Style x Mention	2.5810692	1.1627781	2.220	0.02676	*
Talker dialect x Listener dialect x Talker gender x Frequency x Style x Density	0.0835378	0.0414462	2.016	0.04423	*
Talker dialect x Listener dialect x Frequency x Predictability x Style x Mention x Density	0.4357239	0.1837720	2.371	0.01801	*

Figure 6. Linear model of intelligibility accuracy predicted by eight factors aggregated by items, with the Midland dialect as the reference level for both talker and listener dialect. Only significant main effects and interactions are shown.

Effect	Estimate	Std. Error	t value	Pr(>|t|)	
(Intercept)	0.7284298	0.0587891	12.391	< 2e-16	***
Predictability	-0.5145464	0.2442859	-2.106	0.035533	*
Frequency x Predictability	0.8972491	0.4198554	2.137	0.032942	*
Talker dialect (Northern) x Style (plain)	-0.4799561	0.1277392	-3.757	0.000186	***
Talker dialect x Mention (second)	-0.3031797	0.1261369	-2.404	0.016496	*
Predictability x Mention	0.6117037	0.3099667	1.973	0.048839	*
Frequency x Density	0.0280999	0.0113055	2.486	0.013170	*
Predictability x Density	-0.0582127	0.0278607	-2.089	0.037032	*
Listener dialect (Northern) x Frequency x Predictability	-1.2923022	0.5817153	-2.222	0.026635	*
Talker dialect x Style x Mention	0.6597688	0.1815287	3.635	0.000299	***
Talker dialect x Frequency x Density	-0.0374082	0.0155501	-2.406	0.016403	*
Talker gender (male) x Frequency x Density	-0.0379954	0.0164409	-2.311	0.021122	*
Talker dialect x Predictability x Density	0.0917823	0.0358356	2.561	0.010640	*
Frequency x Style x Density	-0.0302522	0.0136236	-2.221	0.026701	*
Listener dialect x Frequency x Predictability x Style	1.5097589	0.7300874	2.068	0.039016	*
Talker dialect x Talker gender x Frequency x Density	0.0525381	0.0212988	2.467	0.013875	*
Talker dialect x Talker gender x Predictability x Density	-0.1152932	0.0544740	-2.116	0.034658	*
Talker dialect x Frequency x Style x Density	0.0410189	0.0200270	2.048	0.040917	*
Talker gender x Frequency x Style x Density	0.0504733	0.0228536	2.209	0.027531	*
Talker dialect x Frequency x Mention x Density	0.0508729	0.0196591	2.588	0.009862	**
Talker dialect x Predictability x Mention x Density	-0.1372898	0.0596694	-2.301	0.021695	*
Talker dialect x Listener dialect x Frequency x Predictability x Style	-2.0034627	0.9586602	-2.090	0.036993	*
Talker dialect x Talker gender x Frequency x Style x Mention	-0.5485951	0.2629976	-2.086	0.037348	*
Talker dialect x Talker gender x Frequency x Style x Density	-0.0698470	0.0260366	-2.683	0.007477	**
Talker dialect x Talker gender x Predictability x Style x Density	0.1786137	0.0761973	2.344	0.019353	*
Talker dialect x Frequency x Style x Mention x Density	-0.0617602	0.0278757	-2.216	0.027044	*
Frequency x Predictability x Style x Mention x Density	0.1762340	0.0827461	2.130	0.033537	*
Talker dialect x Listener dialect x Frequency x Predictability x Style x Mention	2.5810692	1.1627781	2.220	0.026757	*
Talker dialect x Listener dialect x Talker gender x Frequency x Style x Density	0.0835378	0.0414462	2.016	0.044228	*
Talker dialect x Frequency x Predictability x Style x Mention x Density	-0.2883357	0.1263182	-2.283	0.022753	*
Talker dialect x Listener dialect x Frequency x Predictability x Style x Mention x Density	0.4357239	0.1837720	2.371	0.018011	*

The results returned no main effect for talker dialect, listener dialect, or gender, but all three interacted with the lexical and discourse factors to affect speech intelligibility in noise. Talker dialect, in particular, was involved in many lower- and higher-order interactions (for the purposes of this paper, I will consider two- and three-way interactions as lower-order interactions and any interaction with four or more factors as higher-order interactions). For example, Northern target words were less intelligible in noise when they were in plain speech style, and Northern second mention words were less intelligible than Midland second mention words, which suggests that reduced Northern phrases were more difficult for listeners of both dialects. Because Clopper & Pierrehumbert (2008) found that Northern talkers have more dispersed (and, by extension, more dialect-specific) vowel spaces in reduced speech, it could be that listeners had a harder time accommodating for this source of regional variation in noise. Further, there was a main effect of predictability on Midland target words such that more predictable Midland words were less intelligible to listeners. Since highly predictable words are more reduced than less predictable words (Lieberman, 1963), it could be that listeners had a hard time accommodating for more reduced Midland words, even though semantic predictability has been shown facilitate speech processing (Kalikow, Stevens, & Elliott, 1977). However, it is possible that the speech signal in the intelligibility task was degraded to the point that the predictable context that high predictability target words were situated in was not able to be recovered by listeners.

There were eight higher-order interactions with talker and listener dialect mediated by the lexical and discourse factors. Seven of these interactions in both model outputs were in a positive direction. The reference levels of the model outputs in Figures 5 and 6 are such that both listener and talker dialect are the same dialect in each. Thus, if most of the higher-order interactions involving listener and talker dialect are in a positive direction, it would follow that in certain contexts involving multiple factors known to affect speech intelligibility in noise, there is an intelligibility benefit for Midland listeners processing Midland words and Northern listeners processing Northern words. There has been little previous work that includes eight social, lexical, and discourse factors as predictors of intelligibility, so this finding, if confirmed with further examination and visualization of the model, is novel.

Talker dialect and talker gender were also involved in eight higher-order interactions, but the directions of the effects were variable. However, six of these effects were in Figure 6, and so involved Northern male talkers. Additionally, in Figure 6, there were five lower-order interactions involving only the lexical and discourse factors, suggesting that Midland listener intelligibility accuracy was particularly affected by lexical and discourse variation in Midland phrases. However, every predictor variable was involved in some lower- or higher-order interaction in the intelligibility model. Further analysis and visualization of the model is necessary to fully understand how each predictor affects intelligibility in noise, and which effects matter the most. Next, a similar linear model on classification accuracy was built to explore similarities and differences between factors that affect performance on each task. The lack of relationship found between the two tasks in the previous analysis foreshadows differences between the Intelligibility and classification models, but the hypothesis that performance on the two tasks is mediated by factors affecting acoustic reduction could lead to several similarities.

Predicting classification accuracy

A linear regression analysis predicting classification accuracy was run to explore predictors of classification accuracy and to compare them with the predictors of intelligibility accuracy. Since acoustic reduction is hypothesized to affect dialect perception during processing, and for consistency between intelligibility and classification models, the same predictor variables used in the intelligibility model were used in the classification model. The dependent variable in the analysis was classification accuracy, and the continuous variables frequency, density, and predictability were again centered around zero. Figure 7 shows the model output with the Northern dialect as the reference level for both talker and listener dialect. Figure 8 shows the model output with the Midland dialect as the reference level for both talker and listener dialect.

Figure 7. Linear model of classification accuracy predicted by eight factors aggregated by items, with the Northern dialect as the reference level for both talker and listener dialect. Only significant main effects and interactions are shown.

Effect	Estimate	Std. Error	t value	Pr(>|t|)
(Intercept)	0.5166535	0.0375867	13.746	< 2e-16 ***
Talker dialect (Midland) x Listener dialect (Midland)	0.1553796	0.0716819	2.168	0.03053 *
Listener dialect x Predictability	-0.3911450	0.1831349	-2.136	0.03305 *
Frequency x Predictability	-0.4555360	0.2157443	-2.111	0.03509 *
Listener dialect x Mention (second)	0.1425226	0.0698400	2.041	0.04166 *
Listener dialect x Talker gender (male) x Predictability	0.7551620	0.2621231	2.881	0.00409 **
Listener dialect x Frequency x Predictability	0.6646076	0.2999233	2.216	0.02702 *
Frequency x Predictability x Style (plain)	0.6398223	0.2812344	2.275	0.02321 *
Listener dialect x Predictability x Mention	0.7470773	0.2732016	2.735	0.00641 **
Listener dialect x Frequency x Density	-0.0305750	0.0151316	-2.021	0.04371 *
Frequency x Predictability x Density	-0.1407215	0.0531819	-2.646	0.00833 **
Listener dialect x Talker gender x Frequency x Predictability	-0.7871696	0.3921478	-2.007	0.04511 *
Listener dialect x Frequency x Predictability x Style	-0.8086702	0.3901581	-2.073	0.03858 *
Listener dialect x Talker gender x Predictability x Mention	-1.0073500	0.4313332	-2.335	0.01981 *
Listener dialect x Predictability x Style x Mention	-0.8403239	0.4221267	-1.991	0.04691 *
Frequency x Predictability x Style x Mention	-0.7825697	0.3558866	-2.199	0.02822 *
Listener dialect x Talker gender x Frequency x Density	0.0366226	0.0162838	2.249	0.02483 *
Talker dialect x Frequency x Predictability x Density	0.1381787	0.0566709	2.438	0.01501 *
Listener dialect x Frequency x Predictability x Density	0.1214078	0.0555423	2.186	0.02916 *
Talker gender x Frequency x Predictability x Density	0.1551220	0.0550043	2.820	0.00494 **
Frequency x Predictability x Style x Density	0.1297873	0.0590107	2.199	0.02818 *
Talker gender x Predictability x Mention x Density	0.0942259	0.0408824	2.305	0.02148 *
Frequency x Predictability x Mention x Density	0.1214763	0.0575084	2.112	0.03502 *
Predictability x Style x Mention x Density	0.0940894	0.0429920	2.189	0.02897 *
Talker dialect x Listener dialect x Frequency x Predictability x Density	-0.1202733	0.0586381	-2.051	0.04064 *
Talker dialect x Talker gender x Frequency x Predictability x Density	-0.1801313	0.0585879	-3.075	0.00219 **
Listener dialect x Talker gender x Frequency x Style x Density	-0.0411135	0.0169593	-2.424	0.01560 *
Talker dialect x Frequency x Predictability x Style x Density	-0.1468852	0.0626427	-2.345	0.01932 *
Talker gender x Frequency x Predictability x Style x Density	-0.1601944	0.0612610	-2.615	0.00912 **
Talker dialect x Talker gender x Predictability x Mention x Density	-0.1267443	0.0598319	-2.118	0.03451 *
Talker dialect x Frequency x Predictability x Mention x Density	-0.1477725	0.0653115	-2.263	0.02397 *
Talker gender x Predictability x Style x Mention x Density	-0.1551257	0.0532254	-2.915	0.00368 **
Talker dialect x Listener dialect x Talker gender x Frequency x Predictability x Density	0.1130229	0.0521520	2.167	0.03056 *
Talker dialect x Listener dialect x Talker gender x Predictability x Style x Density	0.2057933	0.0837987	2.456	0.01430 *
Talker dialect x Talker gender x Frequency x Predictability x Style x Density	0.1480736	0.0635864	2.329	0.02017 *
Talker dialect x Talker gender x Frequency x Predictability x Mention x Density	0.1650999	0.0755780	2.184	0.02926 *
Talker dialect x Talker gender x Predictability x Style x Mention x Density	0.2138694	0.0893124	2.395	0.01691 *
Listener dialect x Talker gender x Predictability x Style x Mention x Density	0.1761994	0.0784774	2.245	0.02507 *

Figure 8. Linear model of classification accuracy predicted by eight factors aggregated by items, with the Midland dialect as the reference level for both talker and listener dialect. Only significant main effects and interactions are shown.

Effect	Estimate	Std. Error	t value	Pr(>|t|)
(Intercept)	0.6100669	0.0309059	19.740	< 2e-16 ***
Talker dialect (Northern)	-0.1710565	0.0502165	-3.406	0.000697 ***
Talker gender (male)	-0.0958059	0.0484045	-1.979	0.048185 *
Talker dialect x Listener dialect (Northern)	0.1553796	0.0716819	2.168	0.030531 *
Talker dialect x Talker gender	0.1362381	0.0690715	1.972	0.048963 *
Talker gender x Style (plain) x Mention (second)	0.2769225	0.1087054	2.547	0.011069 *
Talker gender x Style x Density	-0.0232827	0.0108636	-2.143	0.032449 *
Listener dialect x Talker gender x Style x Mention	-0.3336196	0.1547377	-2.156	0.031429 *
Talker dialect x Talker gender x Frequency x Density	0.0276974	0.0119318	2.321	0.020563 *
Listener dialect x Talker gender x Frequency x Density	-0.0076285	0.0129922	-0.587	0.557287
Talker dialect x Talker gender x Style x Density	0.0314168	0.0137841	2.279	0.022962 *
Talker dialect x Predictability x Mention x Density	-0.0760713	0.0367072	-2.072	0.038604 *
Talker gender x Predictability x Mention x Density	-0.1303582	0.0529428	-2.462	0.014052 *
Talker gender x Style x Mention x Density	0.0358198	0.0161637	2.216	0.027014 *
Listener dialect x Frequency x Predictability x Style x Mention	-1.3139818	0.5570549	-2.359	0.018614 *
Talker gender x Frequency x Predictability x Style x Mention	-1.3108311	0.6425512	-2.040	0.041729 *
Talker dialect x Listener dialect x Frequency x Predictability x Density	-0.1202733	0.0586381	-2.051	0.040636 *
Listener dialect x Talker gender x Frequency x Style x Density	0.0411135	0.0169593	2.424	0.015598 *
Talker dialect x Talker gender x Frequency x Mention x Density	-0.0342730	0.0170376	-2.012	0.044652 *
Talker dialect x Talker gender x Predictability x Mention x Density	0.1940806	0.0635948	3.052	0.002362 **
Listener dialect x Talker gender x Frequency x Predictability x Style x Mention	2.4352371	0.9202852	2.646	0.008328 **
Talker dialect x Listener dialect x Talker gender x Frequency x Predictability x Density	0.1130229	0.0521520	2.167	0.030565 *
Talker dialect x Listener dialect x Talker gender x Predictability x Style x Density	0.2057933	0.0837987	2.456	0.014304 *
Talker dialect x Talker gender x Frequency x Predictability x Style x Density	-0.1480736	0.0635864	-2.329	0.020165 *
Talker dialect x Talker gender x Frequency x Predictability x Mention x Density	-0.1650999	0.0755780	-2.184	0.029264 *
Talker gender x Frequency x Predictability x Style x Mention x Density	-0.2139089	0.1002449	-2.134	0.033209 *

Unlike in the intelligibility model, main effects were observed for talker dialect and talker gender in the classification model. Consistent with previous analyses of the classification data, the main effect of talker dialect in Figure 8 confirms that Northern talkers were classified significantly worse than Midland talkers for Midland listeners, which was interpreted as a response bias for the Midland dialect. The interaction between talker dialect and listener dialect in Figure 8 shows that Northern listeners classified Northern phrases better than Midland listeners overall, again suggesting less of a Midland response bias on the part of Northern listeners. There was a main of effect of talker gender on classification of Midland phrases (Figure 8), such that classification accuracy of Midland male talkers was significantly worse than Midland female talkers for Midland listeners. The two-way interaction between talker dialect and talker gender indicates that Midland listeners are better at classification of Northern male talkers. Because there was already a response bias towards Midland phrases, the main effect of talker gender and the two-way interaction could suggest that Northern male talkers were more accurately or easily classified by Midland listeners as Northern talkers than Northern female talkers. It could be that Northern dialect-specific cues are easier to recognize in males. Alternatively, it could be that the four Northern male talkers in this experiment were better representatives of the Northern dialect than their four female counterparts.

An examination of the interactions involving listener dialect in Figure 7 extend this possibility. There were seven lower-order interactions and nine higher-order interactions that involved listener dialect in Figure 7 alone. Of the six lower-order interactions involving Midland listeners and Northern phrases in Figure 7, four were in a positive direction, which raises an interesting possibility. Effects in the positive direction suggest that Midland listeners were better than Northern listeners at classifying Northern phrases in certain contexts (certain combinations of lexical and discourse factors). If this result holds, it could mean that Midland listeners are more sensitive to Northern dialect-specific cues in certain contexts than Northern listeners. However, the directions of the higher-order interactions involving Midland listeners and Northern phrases were variable, suggesting that Midland listeners are not always more sensitive to Northern cues.

There were many lower and higher-order interactions involving both social factors and lexical and discourse factors in this analysis. This pattern of results suggests that acoustic reduction, social-indexical variation and listener experience all play important roles in listener performance on the classification task. As s00hown in Figure 7, there were seven interactions on Northern phrases that only involved lexical and discourse factors, i.e. not mediated by social-indexical variation, and none for Midland phrases. Clopper & Pierrehumbert (2008)'s finding that Northern vowels are more dispersed when reduced could again offer an explanation for the appearance of these interactions. It is possible that acoustic reduction affects the acoustic cues in

Northern speech more than in Midland speech. Additionally, because of the Midland response bias in the classification task for listeners of both dialects, it seems that listeners might have to have stronger or more perceptually salient Northern cues to motivate classifying a talker as Northern. In this way, more reduced, dispersed vowels could be necessary for Northern talker classification, but not Midland talker classification.

Notably, the opposite pattern occurred in the intelligibility model: lexical and discourse-only effects and interactions were observed for Midland phrases, but only one main effect of speech style was observed for Northern phrases. This is puzzling, because no overall intelligibility benefit was observed for Midland phrases in the intelligibility model. Further investigation of the interactions will help with an interpretation of why lexical and discourse factors interacted on their own to affect Midland phrases in processing but Northern phrases in classification. In any case, the large number of listener dialect interactions in the classification model are consistent with the finding that listener experience affects dialect classification performance (Clopper & Pisoni, 2006). The very low number of interactions involving listener dialect as the only relevant social factor in the intelligibility model suggests that listener regional background does not affect speech processing as much as dialect classification of the Northern and Midland dialects for Northern and Midland listeners.

Another notable similarity between the low-level lexical and discourse-only interactions for Midland intelligibility phrases and Northern classification phrases was that frequency and predictability were present in almost all of the interactions. It may be that for both predictability and frequency, the signal was degraded in the intelligibility task so that listeners could not use context to facilitate processing, but frequency and predictability did facilitate processing in the nondegraded signal in the classification task, so listeners could pay more attention to acoustic cues for classification. These preliminary interpretations of the model output are cautious, and a more thorough examination of both models is needed to better understand effects and interactions of lexical, discourse and social factors.

Discussion

The mismatch between the processing and perception of Midland and Northern speech was reconfirmed in the results of the intelligibility and classification tasks. In the intelligibility task, both Midland and Northern listeners had higher overall accuracy on Midland target words, confirming a processing benefit for Midland speech. For successful performance on the task, listeners must be able to transcribe short phrases in a degraded speech signal. Additionally, listeners must be able to successfully normalize for the regional and social variation they encounter and overcome processing difficulties posed by acoustic reduction in the degraded signal. Thus, a possible explanation for the Midland intelligibility benefit could be that listeners had to rely more on stored lexical representations during the processing of the degraded signal, and Midland productions match these stored representations more than Northern productions. There could be many possible reasons for why Midland forms more closely resemble the stored representations of listeners from the Northern and Midland dialects, and a discussion of this anomaly is revisited below.

The results of the classification task, on the other hand, reveal that listeners could not classify Northern and Midland speakers above chance. Listeners' poor overall performance could be explained by the complex nature of dialect classification. As described above, listeners must accomplish several steps in order to successfully classify a speaker by dialect. First, it is helpful for listeners to have had adequate experience with the regional dialects being presented to them.

It is likely that listeners in this particular task did have adequate experience with the Northern and Midland dialects, even though the Midland listeners had never before lived in the Northern dialect region. Ohio State's student body draws largely from Ohio's major metropolitan centers, including Cleveland and Toledo (Northern dialect), and Columbus and Cincinnati (Midland dialect). Listeners from both dialect regions interact with speakers from both dialect regions every day on campus and so are exposed to the speech of both dialect regions. Next, listeners must be able to adequately identify dialect-specific cues in the speech signal during processing. It is possible the experiment design could have negatively impacted listener performance. Because phrases were only three to seven words long, there may not have been adequate dialect-specific information in each phrase presented to the listener. However, it is presumed that a large number of phrases that were balanced for target vowels gave listeners ample opportunity to at least identify the dialect of the speaker in many of the phrases.

Critically, for successful classification the listener must be able to link anomalies identified in the signal during speech processing to regional background differences between speakers. In the context of a dialect classification task, (the experimenter hopes that) the listener's attentional resources are directed towards linking cues in the speech signal to regional dialect differences, which is rarely the case in everyday speech processing. But even when this was the case, the Northern and Midland listeners in this study still could not classify Northern and Midland speakers by dialect, which is a further attestation to how difficult listeners found the task. It could be that listeners simply could not identify abnormalities in the speech signal. However, it is possible that listeners perceive acoustic differences between speakers and do not attribute them to regional background differences (i.e. Campbell-Kibler, 2012).

Even if listeners can successfully perceive acoustic cues and recognize that they are regionally-motivated, for successful classification they must be able to map the correct acoustic cues to the correct commonly-accepted dialect groups. It is likely that this is where the Northern and Midland listeners in the present study had the most difficulty. It is not likely that “Northern” and “Midland” are highly salient social groups for Ohio residents. Moreover, Campbell-Kibler’s (2012) survey of listener’s perceptions of Ohio dialects reveal that attitudes towards Northern speech in Ohio are relatively neutral, and is not considered more or less prestigious than Midland speech. These results suggest that being able to distinguish between the two regional dialects may not be very socially useful. Previous dialect classification research points to affective, group-level processes that influenced how listeners classified speakers by dialect (Williams et al., 1999). Group-level identity could be necessary for the attribution of acoustic cues to a group of speakers. Conversely, a lack of “Northern” and “Midland” identity, or even a perceived, unified “Midwestern” or “Ohioan” identity could mask acoustic differences between the two dialects.

In this way, it could be the lack of the attribution of Northern Cities Shifted vowels to the Northern dialect that drives the mismatch between perception and production that Niedzielski (1999) observed. It could be that the listeners in Niedzielski’s (1999) task had the ability to accurately perceive the relevant acoustic cues, but the label of “Detroit” or “Canadian” directed their attentional resources towards certain cues collectively attributed to Detroit vs. Canadian speech. Further, if listeners attribute Northern cues to individual idiosyncrasies, or even if they cannot perceive them, Northern tokens of a word may not be encoded or may not be strongly encoded into the lexical representation. If this is the case, Midland forms would be more often encoded into the lexical representation for speakers of both dialects, and so listeners’ representations match Midland forms more than Northern forms, which could possibly explain the Midland processing benefit in the intelligibility task.

Moreover, both Midland and Northern listeners had a response bias towards Midland phrases. As mentioned above, such a bias from Midland listeners might be expected: if Midland listeners have a hard time identifying regional background differences between the two dialects, all of the speakers might sound “normal” to them, and since they are from the Midland dialect they will classify speakers that sound like them as Midland speakers. But if the same is true of Northern listeners, a Northern response bias might be expected from Northern listeners. Although results showed that Midland listeners had a stronger Midland response bias, Northern listeners still showed a Midland response bias on the classification task. A Northern response bias for Midland phrases is puzzling, and points to outside processes affecting performance on the dialect classification task. It is possible that as Columbus is in the Midland dialect region, all listeners considered the Midland dialect as the “default”, and needed specific acoustic cues to draw them away from the conclusion that a speaker is from the Midland dialect region. Further, the majority of the Northern listener group’s only experience living in the Midland dialect region is when they moved to Columbus to attend college, and the speakers in the corpus from which the speech stimuli were drawn are mostly college-aged. It could be that age-related acoustic cues (and those not necessarily phonological) make the speakers easily identifiable as college-aged, so Northern listeners who attend college in the Midland dialect region are more likely to classify these speakers as Midland. Nevertheless, the Midland response bias on the part of the Northern listeners is puzzling and warrants future investigation.

The results from the linear regression predicting classification accuracy from intelligibility accuracy and talker and listener dialect did not yield a strong correlation between the classification and intelligibility tasks. The results primarily suggest a separation of processing and perceptual mechanisms. A lack of correlation between these two tasks may be expected, however, given poor overall performance on the classification task. It may be that the difficulty of this particular classification task, where listeners listened to a short phrase one time and had to choose a dialect label to assign it, masked any correlation between the tasks. Nevertheless, the results did not confirm the hypothesis that easier-to-identify tokens would be harder to process, and vice versa. The dispersion of the data in Figure 4, along with the individual by items analyses of both tasks, revealed that even though classification of Northern tokens was significantly worse than Midland tokens, Northern tokens were not always harder to process than Midland tokens. It is likely that individual items variation is mediated by the complex interactions explored in the predictive models. If Northern tokens are not always harder to process, the results could provide some evidence that the processing system adapts for regional variation. However, this interpretation is cautious and would need further evidence from the data to support. In other words, the current analyses do not look at the dialect-specificity of each token, so there is no evidence to support that very marked Northern tokens were also easy to process.

The only significant interaction involving intelligibility in the regression was for Midland listeners and Midland phrases: Midland listeners who had more of a Midland response bias in the classification task also had lower Midland target word accuracy. This might have to do with the motivation of individual subjects in completing the experiment. Participants who did poorly on the intelligibility task may have been less focused or motivated. The same participants might have been less likely to try to distinguish acoustic cues in the classification task, and decided to classify more speakers as Midland by default. The same pattern did not show up for Northern listeners on Midland phrases. Northern listeners did have less of a Midland response bias in classification, but it also may be that they were inherently more interested in the task because they had to distinguish speech that is similar to their own from speech that is not.

Interpretation of the predictive models of intelligibility and classification is less refined, but opens many possibilities as to how social, lexical, and discourse factors affect speech processing and dialect perception. In the intelligibility model, interactions were observed in both social factors and lexical and discourse factors that mediate acoustic reduction, which suggests that both of these sources of variation in the speech signal are important for a better understanding of how listeners deal with variation in speech processing. Moreover, there were several interactions between the social, lexical, and discourse factors, a few of which are briefly summarized here. Northern, reduced words seemed to be less intelligible for listeners overall, which supports the idea that listener lexical representations more closely match Midland productions. Reduced, more dispersed Northern words may have been harder to recover from the degraded signal, because listeners had to rely on their stored representation in a degraded signal relative to a less degraded signal. The many interactions of talker and listener dialect point to an advantage for processing the speech of familiar dialects in reduced contexts. These possible findings will be clarified with future analysis of the models.

In the classification model, like in the intelligibility model, all examined social, discourse, and lexical factors interacted to affect classification accuracy, suggesting that all are relevant for an understanding of how listeners classify talker by dialect. Further, the observance of lexical and discourse effects and interactions on classification accuracy confirm that acoustic reduction influences the ability of the listener to identify regional dialect cues in the speech signal during processing. A few specific interactions are presented and interpreted here. The same talker dialect effect that was observed in the previous statistical analyses remained in the classification model with all of the predictors, again confirming a response bias for Midland phrases. In comparison with the intelligibility model, there were many more interactions with listener dialect as the only social factor in the classification model. In line with previous dialect classification research (Clopper & Pisoni, 2006), this finding suggests that listener experience or regional background significantly influences listener performance on classification, and more broadly how listeners perceive the speech of speakers of different dialects. The fact that main effects for talker gender and talker dialect were observed in the classification model, along with many interactions involving listener dialect, suggests that social information alone may be more relevant to listeners in classification. Finally, the finding that Midland listeners were significantly better at classifying Northern male talkers further suggests that less exposure to Northern acoustic cues facilitates identification of these cues as regionally motivated, at least in the context of a forced-choice task.

Further modeling of the data will also lead to better interpretation of the intelligibility and classification accuracy data. Linear regression models on accuracy aggregated by item means were used here to try to account for the variance observed in the items analysis of the relationship between the two tasks. However, listener dialect was a within-item factor in the classification regression model, and so two observations of each item were included in the model data, which violates the assumption of independence of the items in the linear model. Instead, it would be useful to run a model on the unaggregated data that can account for variation across subjects and across items.

In conclusion, even though all lexical, discourse, and social factors examined interacted to affect both intelligibility and classification accuracy in the predictive models, this is not to say that processing and perceptual mechanisms are identical. In fact, the linear regressions modeling the relationship between the two provide strong evidence for a separation of processing and perceptual mechanisms. However, it seems that both social variation and factors that mediate acoustic reduction affect intelligibility and classification, but in different ways. The processing system may be more sensitive to abnormalities in the signal during an intelligibility in noise task because the goal in the intelligibility task is to map the speech signal onto existing representations. Listeners have to be able to normalize for variation in the signal, but they do not have to identify its source. In classification, attributing variation to regional background differences is the goal of the task. Making this connection seems to be more difficult for listeners, and could explain the mismatch between processing and perception considered in the present study.

References

Allen, H. B. (1954). The primary dialect areas of the Upper Midwest. In Linn, M. D. *Handbook of Dialects and Language variation,* 123-134. San Diego, CA: Academic Press.

Atwood, E. B. (1953). A survey of verb forms in the Eastern United States. Ann Arbor: University of Michigan Press.

Bradlow, A. R., Torretta, G. M. & Pisoni, D. B. (1996). Intelligibility of normal speech I: Global and fine-grained acoustic-phonetic talker characteristics. *Speech Communication, 20,* 255-272.

Burdin, Turnbull, & Clopper, C. G. (2014). Interactions among lexical and discourse characteristics in vowel productions. *Proceedings of Meetings on Acoustics, 22.*

Campbell-Kibler, K. (2012). Contestation and enregisterment of Ohio's imagined dialects.

Journal of English Linguistics, 48(1), 281-305.

Clopper, C. G., Mitsch, J. F., & Tamati, T. N. (2017). Effects of phonetic reduction and regional dialect on vowel production. *Journal of Phonetics, 60,* 38-59.

Clopper, C. G. & Smiljanic, R. (2011). Effects of gender and regional dialect on prosodic patterns in American English. *Journal of Phonetics, 9,* 237-249.

Clopper, C. G., Pierrehumbert, J. B., & Tamati, T. N. (2010). Lexical neighborhoods and phonological confusability in cross-dialect word recognition in noise. *Laboratory Phonology, 1,* 65-92.

Clopper, C. G. & Bradlow, A. R. (2008). Perception of dialect variation in noise: Intelligibility and classification. *Language and Speech, 51(3),* 175-198.

Clopper, C. G. & Pierrehumbert, J. B. (2008). Effects of semantic predictability and regional dialect on vowel space reduction. *The Journal of the Acoustical Society of America, 124,* 1682.

Clopper, C. G. & Pisoni, D. B. (2007). Free classification of regional dialects in American English. *Journal of Phonetics, 35,* 421-438.

Clopper, C. G. & Pisoni, D. B. (2006). Effects of region of origin and geographic mobility on perceptual dialect categorization. *Language Variation and Change, 18,* 193-221.

Clopper, C. G. & Pisoni, D. B. (2004). Some acoustic cues for the perceptual categorization of American English dialects. *Journal of Phonetics, 32,* 111-140.

Fees, C. (1991). The historiography of dialectology. *Lore and Language,* 10/2, 67-74.

Fowler, C. A. & Housum, J. (1987). Talkers signaling of “new” and “old” words in speech and listeners’ perception and use of the distinction. *Journal of Memory and Language, 26(5),* 489-504.

Jones, Z., Yan, Q., Wagner, L., & Clopper, C. G. (2017). The development of dialect classification across the lifespan. *Journal of Phonetics, 60,* 20-37.

Kalikow, D. N. Stevens, K. N. & Elliott, L. L. (1977). Development of a test of speech intelligibility in noise using sentence materials with controlled word predictability. *Journal of the Acoustical Society of America, 61,* 1337-1351.

Kurath, H. (1967). A word geography of the Eastern United States. Ann Arbor: University of Michigan Press.

Kurath, H. (1961). The pronunciation of English in the Atlantic states. Ann Arbor: University of Michigan Press.

Niedzielski, N. (1999). The effect of social information on the perception of sociolinguistic variables. *Journal of Language and Social Psychology, 18,* 62-85.

Labov, W., Ash, S. & Boberg, C. (2005). The atlas of North American English: Phonetics, phonology, and sound change. New York: Mouton de Gruyter.

Labov, W. (1998). The three dialects of English. In Linn, M. D. *Handbook of Dialects and Language Variation*, 39-81. San Diego, CA: Academic Press.

Labov, W. & Ash, S. (1997). Understanding Birmingham. In Bernstein, C., Nunnally, T., & Sabino, R. *Language Variety in the South Revisited.* 508-573. Tuscaloosa, AL: Alabama University Press.

Labov, W. (1972). The social stratification of /r/ in New York City department stores. In Linn, M. D. *Handbook of Dialects and Language Variation*, 221-244. San Diego, CA: Academic Press.

Lieberman, P. (1963). Some effects of semantic and grammatical context on the production and perception of speech. *Language and Speech, 6,* 675-678.

Luce, P. A. & Pisoni, D. B. (1998). Recognizing spoken words: the neighborhood activation model. *Ear and Hearing, 19(1)*, 1-36.

Mason, H. M. (1946). Understandability of speech in noise as affected by region of origin of speaker and listener. *Speech Monographs, 13(2)*, 54-68.

Munson, B. & Solomon, N. P. (2004). The effect of phonological neighborhood density on vowel articulation. *Journal of Speech, Language, and Hearing Research, 47*, 1048-1058.

Picheny, M. A., Durlach, N. I. & Braida, L. D. (1986). Speaking clearly for the hard of hearing II: Acoustic characteristics of clear and conversational speech. *Journal of Speech, Language, and Hearing Research, 29*, 434-446.

Picheny, M. A., Durlach, N. I. & Braida, L. D. (1985). Speaking clearly for the hard of hearing I: Intelligibility differences between clear and conversational speech. *Journal of Speech, Language, and Hearing Research, 28*, 96-103.

Williams, A., Garrett, P. & Coupland, N. (1999). Dialect recognition. In Preston, D. R. *Handbook of Perceptual Dialectology,* 345-358. Philadelphia, PA; John Benjamins Publishing Company.