Dialect Interference in Lexical Processing: Effects of Familiarity and Social Stereotypes

Cynthia G. Clopper
Department of Linguistics, Ohio State University, Columbus, Ohio, USA

Abstract

Background/Aims: The current study explored the roles of dialect familiarity and social stereotypes in dialect interference effects in a speeded lexical classification task. Methods: Listeners classified the words bad and bed or had and head produced by local Midland and non-local Northern talkers and the words sod and side or rod and ride produced by non-local, non-stereotyped Northern and non-local, stereotyped Southern talkers in single- and mixed-talker blocks. Results: Lexical classification was better for the local dialect than for the non-local dialects, and for the stereotyped non-local dialect than for the non-stereotyped non-local dialect. Dialect interference effects were observed for all three dialects, although the patterns of interference differed. For the local dialect, dialect interference was observed for response times, whereas for the non-local dialects, dialect interference was observed primarily for accuracy. Conclusion: These findings reveal complex interactions between indexical and lexical information in speech processing.

1. Introduction

Variability in the speech signal due to indexical properties of the talker can interfere with lexical processing, especially when the listener must process speech from multiple talkers, unfamiliar talkers, or talkers with unfamiliar regional or non-native accents. The classic demonstration of this interference effect is that word recognition in noise is less accurate when the talker changes from trial to trial than when all of the stimulus materials are produced by the same talker (Creeelman, 1957; Mullennix et al., 1989). Responses are also slower in speeded forced-choice lexical classification tasks when the stimulus materials are produced by multiple different talkers than when they are produced by a single talker (Mullennix and Pisoni, 1990). The observed differences in accuracy and response time between the single-talker and mixed-talker conditions in these tasks is attributed to a processing cost associated with normalization for talker variability. The mapping between the acoustic signal and the relevant phonological categories is ‘many to one’, and trial-to-trial talker variability limits the listener’s ability
to adapt to the idiosyncratic signal-to-category mappings for each of the individual talkers. This cost of normalization for talker variability is further evident in the observation that as the number of different talkers in a mixed-talker condition increases, the magnitude of the interference effect also increases. For example, Mullennix and Pisoni (1990) observed response time increases of ~60 ms from single-talker to mixed-talker conditions with two talkers, but response time increases of ~160 ms from single-talker to mixed-talker conditions with 16 talkers.

Regional dialect variation adds further complexity to this normalization process. In American English, regional dialects are characterized predominantly in terms of variation in the vowel system (Labov et al., 2006). Several ongoing vowel shifts have created acoustically overlapping, but phonologically distinct, vowel categories across dialects. That is, across regional dialects, phonological confusions result from contact between shifted and unshifted variants of different vowel categories (Clopper et al., 2010; Jacewicz and Fox, 2012; Wright and Souza, 2012). For example, the lowered and backed /ɛ/ in the Northern dialect of American English is perceptually confusable with /æ/ (Clopper et al., 2010) and the monophthongal /ɑ/ in Southern American English is perceptually confusable with /æ, a/ (Jacewicz and Fox, 2012). These phonologically ambiguous forms have the potential to lead to substantial processing costs beyond those needed to normalize for within-dialect talker variability. The first goal of the current study was therefore to determine if a dialect interference effect could be observed above and beyond a talker interference effect for these cross-dialect vowel contrasts in a speeded lexical classification task.

Previous research on cross-dialect speech processing has demonstrated that both the listener’s experience and the social stereotypes associated with specific dialect categories have a substantial impact on how these potentially phonologically ambiguous variants are processed. In particular, local listeners tend to perceive local forms more accurately than non-local listeners (Clopper et al., 2010; Jacewicz and Fox, 2012; Wright and Souza, 2012). More generally, speech intelligibility and word and sentence comprehension are more accurate for familiar dialects than for unfamiliar dialects (Adank et al., 2009; Adank and McQueen, 2007; Clopper and Bradlow, 2008; Labov and Ash, 1997; Mason, 1946), and lexical decision performance is faster and more accurate for familiar dialects than for unfamiliar dialects (Floccia et al., 2006; Impe et al., 2008; Sumner and Samuel, 2009). The familiar dialects in these studies include both local varieties and regional standard varieties, but recent research also suggests that socially salient non-local varieties, such as British English for American listeners, may also exhibit familiar-dialect effects (Sumner and Kataoka, 2013).

The external activation of stereotyped dialect categories, through either explicit place labels or implicit cultural associations, has similarly been shown to affect performance in vowel perception tasks (Hay and Drager, 2010; Hay et al., 2006; Niedzielski, 1999). For example, Niedzielski (1999) found that listeners in Detroit, Michigan, were more likely to correctly match target vowels when they believed the talker was from Canada (a locally stereotyped region) than when they believed the talker was from Michigan (the local dialect area). When they believed the talker was from Michigan, their responses corresponded to standard American English vowels rather than to the local forms, because the target variants are recognized as stereotypically Canadian forms rather than as local forms. The role of social stereotypes in cross-dialect perception is particularly relevant for the Northern dialect of American English (spoken in the northern Midwestern USA) because although this dialect is characterized by the
Northern Cities vowel shift (an acoustically distinct system of vowel variants), it is not socially stereotyped as distinct from the more standard Midland dialect, which is spoken in the southern Midwestern USA (Clopper et al., 2012; Niedzielski, 1999). Campbell-Kibler (2012) has argued that this mismatch between production and perception of the Northern dialect reflects the lack of social enregisterment or social stereotyping of the Northern dialect. The Northern dialect thus contrasts sharply with the Southern dialect of American English, which is highly enregistered and associated with non-standard, but pleasant, stereotypes (Niedzielski and Preston, 2000). The second goal of the current study was to explore these potential effects of familiarity through experience and social stereotypes/enregisterment on dialect interference effects in lexical processing.

1.1. The Current Study

The current study used a series of speeded lexical classification tasks to determine if a dialect interference effect could be observed above and beyond a talker interference effect. In particular, in experiments 1A and 2A, the processing costs associated with within-dialect talker interference were assessed by comparing performance in single-talker and mixed-talker blocks, in which the talkers shared a regional dialect. In experiments 1B and 2B, the processing costs associated with cross-dialect talker interference were assessed by comparing performance in single-talker and mixed-talker blocks, in which the talkers represented different dialects. The magnitude of the talker interference effect in experiments 1A and 2A was then compared to the magnitude of the dialect interference effect in experiments 1B and 2B, respectively. A larger dialect interference effect than talker interference effect provides evidence for cross-dialect interference above and beyond within-dialect talker interference.

To maximize the likelihood of observing a dialect interference effect, the mixed-talker conditions were limited to two talkers so that the talker interference effect would be relatively small (see Mullennix and Pisoni, 1990) and an additional effect of dialect interference could potentially be observed. Further, to maximize the dialect differences between the mixed-talker and mixed-dialect conditions across experiments, the target words were selected to contain phonologically ambiguous variants across the target dialects. The target talkers and stimulus tokens were then carefully selected to be good representatives of their respective dialect regions, based on extensive previous descriptive research on the target varieties (Clopper et al., 2005; Fox and Jacewicz, 2009; Labov et al., 2006). Given the previous research demonstrating cross-dialect lexical processing costs with small numbers of representative talkers (e.g., Clopper et al., 2010; Floccia et al., 2006; Sumner and Samuel, 2009) and the potential for phonological confusions due to dialect-specific vowel variation (e.g., Jacewicz and Fox, 2012), processing interference due to the mixed-dialect condition was expected to be larger than interference due to the within-dialect mixed-talker condition. That is, within each of experiments 1 and 2, an experiment (A vs. B) × block (single-talker vs. mixed-talker) interaction was expected, reflecting a larger block effect in the dialect interference experiment (B) than in the within-dialect talker interference experiment (A).

In experiment 1, the local Midland dialect was compared with the non-local Northern dialect of American English. The listeners were from the Midland dialect region, and were therefore expected to be highly familiar with their native Midland dialect but less familiar with the non-local Northern dialect. The Northern dialect is characterized by the Northern Cities vowel shift, which includes raised and fronted /æ/ and lowered and/or backed /ɛ/ relative to the Midland dialect (Labov et al., 2006). The
Clopper raised and fronted Northern /æ/ is confusable with the Midland /ɛ/ and the lowered and backed Northern /ɛ/ is confusable with the Midland /æ/ (Clopper et al., 2010; Jacewicz and Fox, 2012). Target words with /æ/ (bad) and /ɛ/ (bed) were therefore selected as the stimulus materials because they were predicted to be somewhat confusable across the two dialects. The processing costs associated with normalizing for this dialect-specific vowel variation in the mixed-dialect condition in experiment 1B should be greater than the costs associated with normalizing for within-dialect talker variation in experiment 1A, resulting in greater interference effects on the speed and/or accuracy of the lexical classification responses in experiment 1B than in experiment 1A. If performance accuracy is near ceiling, this predicted experiment × block interaction may be stronger for response times than for accuracy, due to the limited variation that is typically observed in near-ceiling accuracy scores (Studebaker, 1985). By contrast, if performance accuracy is not near ceiling, the interaction may be weaker for response times than accuracy, given that response times are analyzed only for correct trials. Thus, lower accuracy means that fewer trials are included in the response time analysis, leading to less robust estimates of by-subject response time means.

Given that the experimental design was limited to just two target words and two talkers in each of the mixed-talker conditions, the generalizability of the results is potentially limited (Clark, 1973). Experiments 1C and 1D therefore served as a direct replication of experiments 1A and 1B, respectively, with different target words (had, head), two separate mixed-dialect talker pairs in the mixed-dialect condition in experiment 1D, and a more heterogeneous Midland listener population, recruited from a local science museum instead of a university campus. If the results are replicated across the two experiment pairs (i.e., 1A and 1B vs. 1C and 1D), these differences in stimulus materials, listener population, and laboratory environment provide some assurance that the results do not primarily reflect sampling error. A successful replication therefore increases the generalizability of the results beyond the specific talkers, words, listeners, and laboratory environments that were used in the current study (Larzelere et al., 2015; Monin and Oppenheimer, 2014; Simons, 2014; Westfall et al., 2015).

In experiment 2, the role of social stereotypes in dialect interference effects was examined through a comparison of the non-local, non-enregistered Northern dialect and the non-local, enregistered Southern dialect of American English. As in the first set of experiments, the listeners were from the Midland dialect region and therefore did not have native familiarity with either the Northern or the Southern dialect. In this set of experiments, the target words contained the vowels /ɑː/ (side) and /ɑ/ (sod) because monophthongal /ɑː/ is a highly stereotyped feature of Southern American English (Plichta and Preston, 2005) and is also acoustically very similar to the non-stereotyped, fronted Northern /ɑ/ (see fig. 3, 4). Thus, as in the first set of experiments, the target words were potentially confusable across the two dialects in the lexical classification task to maximize the possibility of observing a dialect interference effect in experiment 2B relative to experiment 2A. However, unlike in experiment 1, the dialects in experiment 2 were both non-local varieties, but differed in their social enregisterment. The Southern dialect in general and /ɑː/ monophthongization in particular are both highly stereotyped, which allows for an examination of the effect of social enregisterment on the magnitude of dialect interference effects for two non-local dialects. Experiment 2 therefore provided the opportunity to conceptually replicate the effects of experiment 1 with different dialects and different vowel variables. Further, as in the first set of experiments, experiments 2C and 2D served as a direct replication of experiments 2A.
2. Experiment 1: Local Midland versus Non-Local Northern American English

2.1. Experiment 1A: Within-Dialect Talker Interference

2.1.1. Listeners

Fifty-seven undergraduates at Ohio State University participated as listeners in experiment 1A. They received partial course credit in an introductory linguistics course for participating. Data from listeners who were bilingual (n = 12), who reported a history of a speech or hearing disorder (n = 2), or who did not meet the residential history requirements (n = 21) were excluded prior to the analysis. Data from 1 participant were also excluded due to equipment failure. The remaining 21 listeners (15 female, 6 male) were monolingual native speakers of American English who were lifetime residents of the Midland dialect region and ranged in age from 18 to 31 years (mean = 20, SD = 3).

2.1.2. Stimulus Materials

The stimulus materials were produced by two female talkers from the Midland dialect region and two female talkers from the Northern dialect region from the Indiana Speech Project corpus (Clopper et al., 2002). The Midland talkers had lived in the Indianapolis, Indiana, metropolitan area until the age of 18 years, and their parents were also raised in the same area. The Northern talkers had lived in the Indiana suburbs of Chicago, Illinois, until the age of 18 years, and their parents were also raised in the same area. The talkers were 19–21 years old at the time of recording.

The stimulus materials consisted of one token of each of the words bad and bed produced in isolation by each of the four talkers, for a total of eight stimulus tokens. These naturally produced tokens exhibited variation in numerous dimensions, including consonant, vowel, and prosodic properties, but given that regional vowel variation in American English is well documented (e.g., Labov et al., 2006), only the differences in vowel quality and duration between the eight stimulus tokens are highlighted here. The formant trajectories for each of the two vowels are illustrated in figure 1 for the Midland (darker symbols) and Northern (lighter symbols) talkers for bad and bed. For each vowel, the first and second formant frequencies were estimated at 35%, 50%, and 65% of the vowel duration. Each trajectory is labeled in figure 1 at the beginning of the trajectory (i.e., at the 35% point) with the talker identifier (MiA, MiB, NoA, and NoB for Midland talker A, Midland talker B, Northern talker A, and Northern talker B, respectively) and the target word (bad or bed). The total duration of each vowel token, the total trajectory length of each vowel token between the 35% and 65% points (calculated as the summed Euclidean distance in the F1 × F2 plane between the 35% and 50% points and the 50% and 65% points), the average acoustic distance between each vowel pair (calculated as the mean Euclidean distance in the F1 × F2 plane at the 35%, 50%, and 65% points), and the total duration of each word token are provided in table 1.

For the Midland talkers, the vowel in bad is lower than the vowel in bed and for Midland talker B, the vowel in bad is more front than the vowel in bed. As shown...
in table 1, the vowel in *bad* is also longer in duration than the vowel in *bed* for both Midland talkers. The overall formant trajectory lengths are comparable across vowels within each talker, although Midland talker B exhibits longer trajectory lengths overall than Midland talker A. For the Northern talkers, the vowel in *bad* is more front than the vowel in *bed*, and for Northern talker B, the vowel in *bad* is lower than the vowel in *bed*. As shown in table 1, the vowel in *bad* is also longer in duration than the vowel in *bed* for the Northern talkers, but unlike for the Midland talkers, the overall formant trajectory is also longer for *bad* than for *bed* for both talkers, suggesting diphthongization of *bad* relative to *bed*. Thus, the Northern talkers exhibit lowering of /ɛ/ and fronting and/or diphthongization of /æ/, consistent with the Northern Cities vowel shift. Despite the considerable variation in the acoustic characteristics of the vowels across talkers, for all four talkers, the vowels in *bad* and *bed* are well separated in the acoustic space, as shown by the mean vowel distance measures in table 1. Further, although the primary dimension of contrast (F1 vs. F2) and the relative degree of diphthongization as captured by the trajectory length differ across the two dialects, relative duration is a robust cue for vowel category across dialects.

Table 1. The total duration of each vowel token, the total trajectory length of each vowel token between the 35% and 65% points (calculated as the summed Euclidean distance in the F1 × F2 plane between the 35% and 50% points and the 50% and 65% points), the average acoustic distance between each vowel pair (calculated as the mean Euclidean distance in the F1 × F2 plane at the 35%, 50%, and 65% points), and the total duration of each word token in experiments 1A and 1B.

<table>
<thead>
<tr>
<th>Target</th>
<th>Midland A</th>
<th>Midland B</th>
<th>Northern A</th>
<th>Northern B</th>
</tr>
</thead>
<tbody>
<tr>
<td>Vowel</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>duration, ms</td>
<td>bad</td>
<td>271</td>
<td>322</td>
<td>380</td>
</tr>
<tr>
<td></td>
<td>bed</td>
<td>207</td>
<td>194</td>
<td>237</td>
</tr>
<tr>
<td>Vowel</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>trajectory, Hz</td>
<td>bad</td>
<td>43</td>
<td>196</td>
<td>459</td>
</tr>
<tr>
<td></td>
<td>bed</td>
<td>21</td>
<td>205</td>
<td>242</td>
</tr>
<tr>
<td>Vowel</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>distance, Hz</td>
<td>bad/bed</td>
<td>261</td>
<td>372</td>
<td>285</td>
</tr>
<tr>
<td>Word</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>duration, ms</td>
<td>bad</td>
<td>396</td>
<td>411</td>
<td>518</td>
</tr>
<tr>
<td></td>
<td>bed</td>
<td>345</td>
<td>331</td>
<td>438</td>
</tr>
</tbody>
</table>

Fig. 1. First and second formant frequencies of the two Midland (MiA, MiB) and two Northern (NoA, NoB) *bad* and *bed* tokens at 35%, 50%, and 65% of the vowel duration. Trajectories are labeled at the 35% point. The Midland A and Northern A tokens were also used in experiment 1B. Black: MiA; dark gray: MiB; light gray: NoB; white: NoA. Circles denote *bed* tokens and triangles denote *bad* tokens.
The two target words are well matched for log frequency (bad: 3.15; bed: 3.10) to avoid potential item-specific biases in lexical classification. Each token was stored in an individual digital audio file with a sampling rate of 22,050 Hz and 16-bit resolution.

2.1.3. Procedure
Participants were seated at individual computers equipped with headphones and a five-button button box. Each participant completed a speeded lexical classification task, in which they were asked to identify each word token as quickly and as accurately as possible. Two of the conditions involved the Midland and Northern tokens of bad and bed. The remaining two conditions are discussed as experiment 2A below. The order of the four conditions was balanced across participants. The listeners were presented with two single-talker blocks and one mixed-talker block of trials in each condition. In the Midland condition, one single-talker block of trials included words produced by Midland talker A, one single-talker block of trials included words produced by Midland talker B, and the mixed-talker block of trials included words produced by both of the Midland talkers. Similarly, in the Northern condition, one single-talker block of trials included words produced by Northern talker A, one single-talker block of trials included words produced by Northern talker B, and the mixed-talker block of trials included words produced by both of the Northern talkers. Each block consisted of eight repetitions of each stimulus token, for a total of 16 trials in each of the single-talker blocks and 32 trials in each of the mixed-talker blocks. The order of the blocks within each condition and the order of the trials within each block were randomized separately for each listener.

The response alternatives were presented on the computer screen in the same locations for all of the participants: bad was presented on the left and bed was presented on the right. On each trial, participants listened to one token over headphones and made their response by pressing the left-most button on the button box for bad responses or the right-most button on the button box for bed responses. The listeners were instructed to hold the button box with both hands and to use their left hand to press the left-most button and their right hand to press the right-most button. All of the analyses were collapsed across left and right responses, so any differences in response time due to hand dominance are reflected equally across conditions. For each trial, the stimulus token was presented, and accuracy and response time from the onset of the auditory stimulus were recorded. The analysis of the response time data included only response times for correct trials. In addition, response times more than 2 SDs from the mean in a given block for a given participant were excluded. Prior to analysis, the accuracy data were arcsine transformed (Studebaker, 1985), and the response time data were log transformed to improve the fit of the data to a normal distribution. In the summary tables, all data are reported as raw accuracy (percent correct) and response times (ms) to facilitate interpretation.

2.1.4. Results and Discussion
Performance in the lexical classification task was excellent and overall accuracy was 96%. A summary of the accuracy and response times in the lexical classification task is shown in table 2 for each talker dialect in each block. A repeated-measures ANOVA of mean lexical classification accuracy with block (single-talker vs. mixed-talker) and talker dialect (Midland vs. North) as within-subject variables revealed significant main effects of block [F(1, 20) = 11.00, p = 0.003] and talker dialect [F(1, 20) = 6.19, p = 0.022]. The block × talker dialect interaction was not significant. Performance
was more accurate for Midland talkers than for Northern talkers, suggesting an overall benefit for the local variants relative to the non-local variants. Performance was also more accurate in the single-talker blocks than in the mixed-talker blocks, suggesting an overall talker interference effect.

A parallel repeated-measures ANOVA of lexical classification response times revealed significant main effects of block \([F(1, 20) = 25.51, p < 0.001]\) and talker dialect \([F(1, 20) = 95.78, p < 0.001]\). The block × talker dialect interaction was not significant. Performance was faster for the Midland talkers than for the Northern talkers and in the single-talker blocks than in the mixed-talker blocks, consistent with the accuracy results.

The observed near-ceiling accuracy is consistent with previous studies using the speeded lexical classification paradigm, which typically reveals performance better than 90% correct (e.g., Ben-Artzi and Marks, 1999; Mullennix and Pisoni, 1990). The near-ceiling performance further suggests that this task is very easy and that listeners are able to accurately identify isolated words without previous exposure to other vowels produced by these talkers (see also Verbrugge et al., 1976). Given the spectral variation observed in these tokens (see fig. 1), duration may have been the primary cue for lexical category in this task.

The faster and more accurate performance for the Midland talkers than for the Northern talkers is consistent with previous research demonstrating a processing benefit for local varieties relative to non-local varieties (e.g., Floccia et al., 2006). However, an inspection of table 1 reveals that the Northern target words were longer in duration than the Midland target words, suggesting that the effect of talker dialect on response times may simply reflect dialect differences in stimulus duration. Given this collinearity between talker dialect and word duration in the stimulus materials, it is impossible to definitively determine the source of the apparent talker dialect effect. However, a linear regression model predicting the mean response time for each target stimulus from talker dialect accounted for more variance \(r^2 = 0.92, \text{d.f.} = 6, p < 0.001\) than a comparable model with word duration as a predictor variable \(r^2 = 0.65, \text{d.f.} = 6, p = 0.016\), suggesting that at least some of the variance in the response times reflects an effect of talker dialect, beyond the simple effect of stimulus duration.

The faster and more accurate performance in the single-talker blocks than in the mixed-talker blocks suggests a significant talker interference effect, consistent with the previous literature demonstrating the integral nature of linguistic and indexical information in speech processing (Creelman, 1957; Mullennix et al., 1989; Nygaard et al., 1994). The magnitude of the talker interference effect (65 ms) is comparable to the results of previous studies examining two-talker interference using the speeded lexical classification paradigm (Mullennix and Pisoni, 1990).

### Table 2. Mean (SD) lexical classification accuracy and response times for each talker dialect in each block in experiment 1A

<table>
<thead>
<tr>
<th>Talker dialect</th>
<th>Accuracy</th>
<th>Response time, ms</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>single-talker</td>
<td>mixed-talker</td>
</tr>
<tr>
<td>Midland</td>
<td>0.98 (0.02)</td>
<td>0.96 (0.04)</td>
</tr>
<tr>
<td>North</td>
<td>0.97 (0.05)</td>
<td>0.93 (0.07)</td>
</tr>
</tbody>
</table>
The talker dialect × block interaction was not significant for either accuracy or response times, suggesting that the talker interference effect was similar in magnitude for both talker dialects, regardless of familiarity. The effect of dialect familiarity on within-dialect talker interference may be underestimated in this task, given the near-ceiling accuracy, and a task involving noise may be more sensitive to dialect familiarity effects in this domain. However, the response time data were not subject to floor or ceiling effects and therefore provide more compelling evidence for the lack of an overall effect of dialect familiarity on the magnitude of the talker interference effect in the within-dialect condition examined in this experiment. Experiment 1C (see section 2.3) provides a replication of this null result.

Taken together, these results demonstrate that within-dialect talker variability reduces the accuracy and speed of lexical classification performance. However, mixing two non-local Northern talkers does not create significantly more interference than mixing two local Midland talkers. The effects of cross-dialect interference on lexical classification were explored in experiment 1B, in which the mixed-talker block included both a Midland and a Northern talker.

2.2. Experiment 1B: Cross-Dialect Interference

2.2.1. Listeners

Forty-nine undergraduates at Ohio State University participated as listeners in experiment 1B. They received partial course credit in an introductory linguistics course for participating. Data from listeners who were bilingual (n = 5), who used only one hand to make their responses (n = 1), who were considerably older than the other participants (n = 1), who did not meet the residential history requirements (n = 21), or whose mean response times were more than 2 SDs slower than the mean across all participants (n = 1) were excluded prior to the analysis. The remaining 20 listeners (13 female, 7 male) were monolingual native speakers of American English who were lifetime residents of the Midland dialect region and ranged in age from 18 to 25 years (mean = 21, SD = 2).

2.2.2. Stimulus Materials

The stimulus materials consisted of the tokens of the words *bad* and *bed* produced by Midland talker A and Northern talker A from experiment 1A, for a total of four stimulus tokens (see fig. 1).

2.2.3. Procedure

The procedure was similar to experiment 1A, except that each participant completed a speeded lexical classification task and a speeded dialect classification task. Performance in the dialect classification task was not significantly different from chance and will therefore not be discussed further (cf. Clopper and Pate, 2008). Task order was counterbalanced across listeners.

As in experiment 1A, in the lexical classification task, participants were asked to identify each token as *bad* or *bed* as quickly and as accurately as possible. They were presented with two single-talker blocks and one mixed-talker block of trials. One single-talker block of trials included words produced only by the Midland talker, one single-talker block of trials included words produced only by the Northern talker, and the mixed-talker block of trials included words produced by both talkers. Stimulus presentation, data collection, and preprocessing of the accuracy and response time data were identical to experiment 1A.
2.2.4. Results and Discussion

As in experiment 1A, performance in the lexical classification task was excellent; overall accuracy was 90%. A summary of the accuracy and response times in the lexical classification task for each talker dialect in each block is shown in table 3. The results for the mixed-talker block have been split by talker dialect to illustrate the 2 (block condition) × 2 (talker dialect) design. A repeated-measures ANOVA of mean lexical classification accuracy with block (single-talker vs. mixed-talker) and talker dialect (Midland vs. North) as within-subject variables revealed significant main effects of block \( [F(1, 19) = 10.84, p = 0.004] \) and talker dialect \( [F(1, 19) = 28.51, p < 0.001] \). The block × talker dialect interaction was not significant. As in experiment 1A, performance was more accurate for the local Midland talkers than the non-local Northern talkers and in the single-talker blocks than in the mixed-talker block.

A parallel repeated-measures ANOVA of lexical classification response times revealed significant main effects of block \( [F(1, 19) = 30.80, p < 0.001] \) and talker dialect \( [F(1, 19) = 60.86, p < 0.001] \), as well as a significant block × talker dialect interaction \( [F(1, 19) = 6.28, p = 0.021] \). As in experiment 1A, responses were faster for the Midland talkers than for the Northern talkers and in the single-talker blocks than in the mixed-talker block. However, the block effect was significant for the Midland talkers \( [t(19) = 6.95, p < 0.001] \), but not for the Northern talkers \( [t(19) = 1.80, n.s.] \). Thus, a competing Northern talker significantly slowed responses for the Midland talker, but the already slower responses to the Northern talker were not significantly affected by the addition of the Midland talker.

Taken together, the results of this experiment reveal near-ceiling speeded lexical classification accuracy, faster and more accurate performance for the local Midland talker than for the non-local Northern talker, and faster and more accurate performance in the single-talker blocks than in the mixed-talker block, as in experiment 1A. Unlike in experiment 1A, however, a block × talker dialect interaction was observed for response times. In particular, a larger block effect on response times was observed for the Midland talker than for the Northern talker.

2.2.5. Talker versus Dialect Interference (Experiments 1A and 1B)

To further explore the differences in the block and talker dialect effects between the two experiments, a repeated-measures ANOVA was conducted on lexical classification accuracy with block (single-talker vs. mixed-talker) and talker dialect (Midland vs. North) as within-subject variables and experiment (1A vs. 1B) as a between-subject variable. The analysis revealed significant main effects of block \( [F(1, 39) = 21.62, p < 0.001] \), talker dialect \( [F(1, 39) = 34.63, p < 0.001] \), and experiment \( [F(1, 39) = 4.90, p =

Table 3. Mean (SD) lexical classification accuracy and response times for each talker dialect in each block in experiment 1B

<table>
<thead>
<tr>
<th>Talker dialect</th>
<th>Accuracy</th>
<th>Response time, ms</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>single-talker</td>
<td>mixed-talker</td>
</tr>
<tr>
<td>Midland</td>
<td>0.98 (0.03)</td>
<td>0.96 (0.06)</td>
</tr>
<tr>
<td>North</td>
<td>0.88 (0.15)</td>
<td>0.79 (0.17)</td>
</tr>
</tbody>
</table>
0.033], as well as a significant experiment × talker dialect interaction [F(1, 39) = 13.67, p = 0.001]. The other interactions were not significant. The main effects of block and talker dialect are consistent with the patterns observed in the individual experiments. The main effect of experiment reflects the overall higher accuracy in experiment 1A (96%) than in experiment 1B (90%). The experiment × talker dialect interaction reflects the lower overall performance for the Northern talkers in experiment 1B (84%) than in experiment 1A (95%), especially in the mixed-talker block (see tables 2, 3). An inspection of the individual participant data suggests that the lower overall performance on the Northern talkers in experiment 1B than in experiment 1A was not due to a single low-scoring outlier in experiment 1B. Thus, the difference in accuracy across experiments likely reflects the additional processing costs associated with the mixed-dialect condition in experiment 1B relative to the within-dialect condition in experiment 1A. That is, the experiment × talker dialect interaction provides indirect evidence for a dialect interference effect on accuracy for Northern talkers that exceeds the within-dialect talker interference effect.

A parallel repeated-measures ANOVA of lexical classification response times revealed significant main effects of block [F(1, 39) = 56.31, p < 0.001] and talker dialect [F(1, 39) = 152.18, p < 0.001], as well as a significant experiment × block × talker dialect interaction [F(1, 39) = 6.60, p = 0.014]. The main effect of experiment and the other interactions were not significant. The main effects of block and talker dialect are consistent with the patterns observed in the individual experiments. The significant three-way interaction reflects a significant experiment × block interaction for the Midland talkers [F(1, 39) = 6.71, p = 0.013], but not for the Northern talkers [F(1, 39) = 0.97, n.s.]. Specifically, for local Midland talkers, the dialect interference effect in experiment 1B (117 ms) was greater than the talker interference effect in experiment 1A (53 ms), whereas for non-local Northern talkers, the dialect interference effect in experiment 1B did not differ in magnitude from the talker interference effect in experiment 1A. This interaction therefore confirms a dialect interference effect on response times for Midland talkers that exceeds the simple talker interference effect.

Taken together, these findings are consistent with the hypothesis that cross-dialect talker interference produces larger effects on lexical processing than within-dialect talker interference. Mixing Midland and Northern talkers in a single experimental block results in significant dialect interference effects beyond simple talker interference effects. For the local Midland variety, this dialect interference effect was realized by greater slowing in response times in the mixed-talker block relative to the single-talker block in experiment 1B than in experiment 1A. For the less familiar Northern variety, responses were already slower overall in the single-talker block in experiment 1B, and only a modest, non-significant increase in response times was observed for the Northern talkers in the mixed-dialect block. This effect of block on response times was comparable in magnitude to the block effect for the Northern talkers in experiment 1A. However, the non-significant increase in response times for the Northern talkers in experiment 1B was accompanied by a non-significant decrease in accuracy (see table 3), which together provide a hint of a speed-accuracy trade-off for the Northern talkers in the mixed-dialect block. This speed-accuracy trade-off leads to overall lower accuracy for the Northern talkers in experiment 1B than in experiment 1A, as captured by the significant experiment × talker dialect interaction. As suggested above, this interaction is also consistent with a dialect interference effect beyond a simple talker
interference effect. Experiment 1D (see section 2.4) provides additional evidence for these two patterns of dialect interference.

The results of experiments 1A and 1B are based on a single pair of target words produced by two pairs of talkers in experiment 1A and by a single pair of talkers in experiment 1B. Replication of the results with additional materials is therefore critical for ensuring that the results do not reflect mere idiosyncrasies of the stimulus tokens (Clark, 1973; Larzelere et al., 2015; Monin and Oppenheimer, 2014; Westfall et al., 2015). In experiments 1C and 1D, the effects of talker and dialect interference were confirmed for Midland and Northern talkers with a different pair of target words (had, head), two pairs of talkers in both the talker interference (experiment 1C) and dialect interference (experiment 1D) conditions, and a more heterogeneous Midland listener population, recruited from a local science museum instead of a university campus.

2.3. Experiment 1C: Within-Dialect Talker Interference Replication

2.3.1. Listeners
Forty-five adult visitors to the Center of Science and Industry (COSI) museum in Columbus, Ohio, participated as listeners in experiment 1C. Data from listeners who used only one hand to make their responses (n = 1), who reported a history of a hearing impairment (n = 1), who did not meet the residential history requirements (n = 20), or whose mean response times were more than 2 SDs slower than the mean across all participants (n = 1) were excluded prior to the analysis. The remaining 22 listeners (14 female, 8 male) were monolingual native speakers of American English who had lived exclusively in the Midland dialect region before the age of 18 years and ranged in age from 18 to 53 years (mean = 34, SD = 12).1

2.3.2. Stimulus Materials
The same four talkers produced the stimulus materials in this experiment as in experiment 1A. The stimulus materials consisted of one token of each of the words had and head produced in isolation by each of the four talkers, for a total of eight stimulus tokens. The formant trajectories for each of the two vowels are illustrated in figure 2 for the Midland (darker symbols) and Northern (lighter symbols) talkers for had and head. For each vowel, the first and second formant frequencies were estimated at 35%, 50%, and 65% of the vowel duration. Each trajectory is labeled in figure 2 at the beginning of the trajectory with the talker identifier (MiA, MiB, NoA, or NoB) and the target word (had or head). The total duration of each vowel token, the total trajectory length of each vowel token, the average acoustic distance between each vowel pair, and the total duration of each word token are provided in table 4.

For the Midland talkers, the vowel in had is lower than the vowel in head, and for Midland talker B, the vowel in had is more front than the vowel in head. As shown in table 4, the vowel in had is also longer in duration and trajectory length than the vowel in head for both Midland talkers. For the Northern talkers, the vowel in had is more front than the vowel in head, and for Northern talker B, the vowel in had is

1 Given the wider age range of participants in this experiment, the residential history criteria were relaxed relative to those used in experiments 1A and 1B. Of the 22 participants in this experiment, 16 were lifetime residents of the Midland dialect region. Of the 6 who had lived outside this region, only 1 had lived in the Northern region.
lower than the vowel in head. As shown in table 4, the vowel in had is also longer in duration and trajectory length than the vowel in head for the Northern talkers. Thus, the Northern talkers exhibit lowering and/or backing of /ɛ/ and fronting and/or diphthongization of /æ/, consistent with the Northern Cities vowel shift. Despite the considerable variation in the acoustic characteristics of the vowels across talkers, for all four talkers, the vowels in had and head are well separated in the acoustic space, as shown by the mean vowel distance measure in table 4. Further, although the primary dimension of contrast (F1 vs. F2) differs across the two dialects, relative duration and trajectory length are robust cues for vowel category across dialects.

The two target words are reasonably well matched for log frequency (had: 4.71; head: 3.63) to avoid potential item-specific biases in lexical classification. Each token was stored in an individual digital audio file with a sampling rate of 22,050 Hz and 16-bit resolution.

**Table 4.** The total duration of each vowel token, the total trajectory length of each vowel token between the 35% and 65% points (calculated as the summed Euclidean distance in the F1 × F2 plane between the 35% and 50% points and the 50% and 65% points), the average acoustic distance between each vowel pair (calculated as the mean Euclidean distance in the F1 × F2 plane at the 35%, 50%, and 65% points), and the total duration of each word token in experiments 1C and 1D.

<table>
<thead>
<tr>
<th>Target</th>
<th>Midland A</th>
<th>Midland B</th>
<th>Northern A</th>
<th>Northern B</th>
</tr>
</thead>
<tbody>
<tr>
<td>Vowel duration, ms</td>
<td>had</td>
<td>212</td>
<td>223</td>
<td>290</td>
</tr>
<tr>
<td></td>
<td>head</td>
<td>207</td>
<td>147</td>
<td>190</td>
</tr>
<tr>
<td>Vowel trajectory, Hz</td>
<td>had</td>
<td>239</td>
<td>238</td>
<td>292</td>
</tr>
<tr>
<td></td>
<td>head</td>
<td>69</td>
<td>114</td>
<td>200</td>
</tr>
<tr>
<td>Vowel distance, Hz</td>
<td>had/head</td>
<td>319</td>
<td>308</td>
<td>327</td>
</tr>
<tr>
<td>Word duration, ms</td>
<td>had</td>
<td>438</td>
<td>423</td>
<td>549</td>
</tr>
<tr>
<td></td>
<td>head</td>
<td>434</td>
<td>330</td>
<td>540</td>
</tr>
</tbody>
</table>
2.3.3. Procedure

The procedure was identical to experiment 1A, except that two of the conditions involved the Midland and Northern tokens of had and head instead of bad and bed. The remaining two conditions are discussed as experiment 2C below. The order of the four conditions was balanced across participants.

2.3.4. Results and Discussion

As in experiment 1A, performance in the lexical classification task was excellent; overall accuracy was 91%. A summary of the accuracy and response times in the lexical classification task for each talker dialect in each block is shown in table 5. A repeated-measures ANOVA of mean lexical classification accuracy with block (single-talker vs. mixed-talker) and talker dialect (Midland vs. North) as within-subject variables revealed significant main effects of block \([F(1, 21) = 7.78, p = 0.011]\) and talker dialect \([F(1, 21) = 41.73, p < 0.001]\). The block × talker dialect interaction was not significant. As in experiment 1A, performance was more accurate for the local Midland talkers than for the non-local Northern talkers and in the single-talker blocks than in the mixed-talker blocks.

A parallel repeated-measures ANOVA of lexical classification response times revealed significant main effects of block \([F(1, 21) = 17.80, p < 0.001]\) and talker dialect \([F(1, 21) = 110.32, p < 0.001]\). The block × talker dialect interaction was not significant. As in experiment 1A, responses were faster for the Midland talkers than for the Northern talkers and in the single-talker blocks than in the mixed-talker blocks.

Taken together, the results of this experiment replicate the findings from experiment 1A with different stimulus materials and a more heterogeneous Midland listener population with respect to age and residential history: overall performance was faster and more accurate for the local Midland talkers than for the non-local Northern talkers and in the single-talker blocks than in the mixed-talker blocks. As in experiment 1A, an inspection of table 4 reveals that the Northern target words were longer in duration than the Midland target words, but a linear regression model predicting the mean response time for each target stimulus from talker dialect accounted for more variance \((r^2 = 0.70, \text{d.f.} = 6, p = 0.010)\) than a comparable model with word duration as a predictor variable \((r^2 = 0.58, \text{d.f.} = 6, p = 0.028)\), suggesting that at least some of the variance in the response times reflects an effect of talker dialect, beyond the simple effect of stimulus duration. Further, as in experiment 1A, the interaction between talker dialect and block was not significant for either accuracy or response times, suggesting that the talker interference effect was similar in magnitude across the two dialects. Experiment 1D was designed as a replication of experiment 1B with different target words (had and head), two mixed-dialect talker pairs, and the more

<table>
<thead>
<tr>
<th>Talker dialect</th>
<th>Accuracy</th>
<th>Response time, ms</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>single-talker</td>
<td>mixed-talker</td>
</tr>
<tr>
<td>Midland</td>
<td>0.97 (0.03)</td>
<td>0.96 (0.04)</td>
</tr>
<tr>
<td>North</td>
<td>0.88 (0.11)</td>
<td>0.85 (0.09)</td>
</tr>
</tbody>
</table>

**Table 5.** Mean (SD) lexical classification accuracy and response times for each talker dialect in each block in experiment 1C
heterogeneous Midland listener population to further confirm the effect of dialect interference above and beyond the talker interference effect for these two dialects for Midland listeners.

2.4. Experiment 1D: Cross-Dialect Interference Replication

2.4.1. Listeners

Fifty adult COSI visitors participated as listeners in experiment 1D. Data from listeners who reported a history of a speech or hearing impairment (n = 3), who reported a native language other than English (n = 1), who did not meet the residential history requirements (n = 23), or whose mean response times were more than 2 SDs slower than the mean across all participants (n = 2) were excluded prior to the analysis. The remaining 21 listeners (10 female, 11 male) were monolingual native speakers of American English who had lived exclusively in the Midland dialect region before the age of 18 years and ranged in age from 18 to 64 years (mean = 28, SD = 13).2

2.4.2. Stimulus Materials

The same eight stimulus tokens produced by the same four talkers were used in this experiment as in experiment 1C.

2.4.3. Procedure

The procedure was similar to experiment 1B, except that participants completed four conditions of the speeded lexical classification task. Two of the conditions involved the Midland and Northern tokens of had and head. In one of these conditions, Midland talker A and Northern talker A were presented in separate single-talker blocks and together in a mixed-talker block, as in experiment 1B. In the other condition, Midland talker B and Northern talker B were presented in separate single-talker blocks and together in a mixed-talker block. Thus, this experiment involved two mixed-dialect talker pairs, whereas experiment 1B involved only a single mixed-dialect talker pair. The remaining two conditions are discussed as experiment 2D below. The order of the four conditions was balanced across participants.

2.4.4. Results and Discussion

As in experiment 1B, performance in the lexical classification task was excellent; overall accuracy was 92%. A summary of the accuracy and response times in the lexical classification task for each talker dialect in each block is shown in Table 6. A repeated-measures ANOVA of mean lexical classification accuracy with block

Table 6. Mean (SD) lexical classification accuracy and response times for each talker dialect in each block in experiment 1D

<table>
<thead>
<tr>
<th>Talker dialect</th>
<th>Accuracy</th>
<th>Response time, ms</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Midland</td>
<td>0.98 (0.03)</td>
<td>653 (98)</td>
</tr>
<tr>
<td>North</td>
<td>0.90 (0.10)</td>
<td>778 (106)</td>
</tr>
</tbody>
</table>

2 Of the 21 participants in this experiment, 14 were lifetime residents of the Midland dialect region. Of the 7 who had lived outside this region, 2 had lived in the Northern region.
(single-talker vs. mixed-talker) and talker dialect (Midland vs. North) as within-subject variables revealed significant main effects of block \( F(1, 20) = 4.92, p = 0.038 \) and talker dialect \( F(1, 20) = 42.40, p < 0.001 \), as well as a significant block × talker dialect interaction \( F(1, 20) = 6.63, p = 0.018 \). As in experiment 1B, performance was more accurate for the local Midland talkers than for the non-local Northern talkers and in the single-talker blocks than in the mixed-talker blocks. Further, the block × talker dialect interaction reflects a significant block effect for the Northern talkers \( t(20) = 3.68, p = 0.001 \), but not for the Midland talkers \( t(20) = 0.04, \text{n.s.} \), suggesting a larger dialect interference effect on accuracy for the non-local Northern talkers than for the local Midland talkers.

A parallel repeated-measures ANOVA of lexical classification response times revealed significant main effects of block \( F(1, 20) = 30.26, p < 0.001 \) and talker dialect \( F(1, 20) = 160.43, p < 0.001 \), as well as a significant block × talker dialect interaction \( F(1, 20) = 6.38, p = 0.020 \). As in experiment 1B, responses were faster for the Midland talkers than for the Northern talkers and in the single-talker blocks than in the mixed-talker blocks. Further, the magnitude of the block effect was larger for the Midland talkers \( t(20) = 5.91, p < 0.001 \) than for the Northern talkers \( t(20) = 2.76, p = 0.012 \), consistent with the larger dialect interference effect on response times observed for the local Midland talkers than for the non-local Northern talkers in experiment 1B.

Taken together, the results of this experiment replicate the findings from experiment 1B: speeded lexical classification accuracy was near ceiling, performance was faster and more accurate for the local Midland talkers than for the non-local Northern talkers, and performance was faster and more accurate in the single-talker blocks than in the mixed-talker blocks. Further, as in experiment 1B, the interaction between block and talker dialect for response times revealed a greater dialect interference effect for the Midland talkers than for the Northern talkers. The results of experiment 1D further revealed a significant interaction between block and talker dialect for accuracy, which revealed a greater dialect interference effect for the Northern talkers than for the Midland talkers.

### 2.4.5. Talker versus Dialect Interference (Experiments 1C and 1D)

To further explore the differences in the block and talker dialect effects between experiments 1C and 1D, a repeated-measures ANOVA was conducted on lexical classification accuracy with block (single-talker vs. mixed-talker) and talker dialect (Midland vs. North) as within-subject variables and experiment (1C vs. 1D) as a between-subject variable. The analysis revealed significant main effects of block \( F(1, 41) = 11.35, p = 0.002 \) and talker dialect \( F(1, 41) = 84.07, p < 0.001 \), as well as a significant block × talker dialect interaction \( F(1, 41) = 4.35, p = 0.043 \). The main effect of experiment and the other interactions were not significant. The significant main effects of block and talker dialect are consistent with the effects observed in each of the experiments separately. The significant block × talker dialect interaction is driven primarily by the significant interaction observed in experiment 1D (see tables 5, 6), suggesting greater dialect interference for Northern talkers than Midland talkers for accuracy, although the three-way experiment × block × talker dialect interaction does not reach significance.

A parallel repeated-measures ANOVA of lexical classification response times revealed significant main effects of block \( F(1, 41) = 47.01, p < 0.001 \) and talker dialect \( F(1, 41) = 219.22, p < 0.001 \), as well as significant experiment × talker dialect \( F(1, 41) = 6.26, p = 0.016 \) and experiment × block × talker dialect \( F(1, 41) =
interactions. The main effect of experiment and the other interactions were not significant. The significant main effects of block and talker dialect are consistent with the effects observed in each of the experiments separately. The significant experiment × talker dialect interaction reflects a larger overall effect of talker dialect on response times in experiment 1C than in experiment 1D. The significant three-way interaction reflects a significant experiment × block interaction for the Midland talkers \(F(1, 41) = 4.70, p = 0.036\), but not for the Northern talkers \(F(1, 41) = 0.50, \text{n.s.}\). As in the comparison of experiments 1A and 1B, this three-way interaction confirms a dialect interference effect on response times for the Midland talkers that exceeds the simple talker interference effect.

These results also provide additional evidence for dialect interference on accuracy for the Northern talkers, reflecting a speed-accuracy trade-off. In particular, for the local Midland variety, overall accuracy was near ceiling in experiment 1D in both the single- and mixed-talker blocks (98 and 97%, respectively), and the dialect interference effect was realized by greater slowing in response times in the mixed-talker block relative to the single-talker block in experiment 1D than in experiment 1C. For the less familiar Northern variety, responses were already slower overall in the single-talker block in experiment 1D. To maintain similar response times across talker dialects in the mixed-talker block, the listeners sacrificed accuracy for response time for the Northern talkers, resulting in a significant decrease in accuracy from the single-talker blocks (90%) to the mixed-talker blocks (84%) for the Northern talkers only, as revealed by the significant block × talker dialect interaction for accuracy in experiment 1D but not in experiment 1C. Thus, the dialect interference effect for the Northern talkers was realized by less accurate lexical classification performance, whereas the dialect interference effect for the Midland talkers was realized by slower lexical classification performance. This asymmetry in the dialect interference effect for the Midland and Northern talkers reflects a classic speed-accuracy trade-off resulting from the mismatch in processing difficulty between the easy (fast and accurate) local Midland dialect and the harder (slower and less accurate) non-local Northern dialect. The experiment × talker dialect interaction for response times further captures this speed-accuracy trade-off: responses were slower overall for Midland talkers in experiment 1D than in experiment 1C, consistent with the effect of dialect interference on response times, but faster overall for Northern talkers in experiment 1D than in experiment 1C, consistent with a speed-accuracy trade-off in the more difficult mixed-dialect condition.

The results of experiment 1D therefore replicate the effects observed in experiment 1B with different target words with different acoustic properties (see fig. 1, 2; tables 1, 4), with a second mixed-dialect pair of talkers, and with a more heterogeneous Midland listener population with respect to age and residential history. Together, this set of four experiments provide strong evidence for dialect interference effects on speeded lexical processing for Midland listeners for the local Midland dialect and the non-local Northern dialect. These interference effects emerge in the accuracy domain for talkers of the more difficult, non-local Northern dialect and in the response time domain for talkers of the easier, local Midland dialect.

In experiment 2, two non-local dialects were used to explore how the dialect interference effect is realized when the local dialect is not present and the mismatch in overall processing difficulty between the two dialects is reduced. In particular, the highly enregistered Southern dialect was compared to the non-enregistered Northern dialect for the variables /aj/ and /a/. Monophthongization of /aj/ is a stereotypical
feature of Southern American English, whereas lowering and fronting of /a/ is a non-stereotyped feature of the Northern Cities vowel shift. Given that the Southern dialect is likely to be familiar through the media and shared social stereotypes (Clopper et al., 2012; Plichta and Preston, 2005), the Southern monophthongal /aj/ should be salient to the Midland listeners. By contrast, given that the Northern dialect is non-enregistered (Campbell-Kibler, 2012), the fronted Northern /a/, like the Northern /e, æ/ in experiment 1, is not expected to be salient to Midland listeners. If the dialect interference effects observed in experiment 1 critically require competition between an easy, familiar dialect and a harder, less familiar dialect, a different pattern of interference effects may emerge when Northern and Southern dialects are mixed. Although the Southern dialect is socially enregistered and may be easier to process than the non-enregistered Northern dialect for Midland listeners, the Southern dialect should still be more difficult to process than the local Midland variety (Clopper and Bradlow, 2008; Labov and Ash, 1997).

3. Experiment 2: Non-Local Northern versus Non-Local Southern American English

3.1. Experiment 2A: Within-Dialect Talker Interference

3.1.1. Listeners

The same listeners from experiment 1A also completed experiment 2A in a single experimental session.

3.1.2. Stimulus Materials

Both Northern talkers from experiment 1A were used in experiment 2A. In addition, two female talkers of the Southern dialect of American English were selected from the Indiana Speech Project corpus (Clopper et al., 2002). The Southern talkers had lived exclusively in the Indiana suburbs of Louisville, Kentucky, before the age of 18 years, and their parents were also from the same region. The Southern talkers were 22 years old at the time of recording.

The stimulus materials consisted of one token of each of the words sod and side produced by each of the four talkers, for a total of eight stimulus tokens. The formant trajectories for each of the two vowels are illustrated in figure 3 for the Southern (darker symbols) and Northern (lighter symbols) talkers for sod and side. For each vowel, the first and second formant frequencies were estimated at 35%, 50%, and 65% of the vowel duration. Each trajectory is labeled in figure 3 at the beginning of the trajectory with the talker identifier (NoA, NoB, SoA, and SoB) and the target word (sod or side). The total duration of each vowel token, the total trajectory length of each vowel token, the average acoustic distance between each vowel pair, and the total duration of each word token are provided in table 7.

For the Southern talkers, the vowel in side is more front and longer in duration than the vowel in sod. For Southern talker A, the trajectories of sod and side are similar, but for Southern talker B, the trajectory of side is longer than the trajectory of sod. However, both of the Southern talkers exhibit some monophthongization of /aj/, as expected. For Northern talker A, the vowel in side is higher than the vowel in sod, and for Northern talker B, the vowel in side is more front than the vowel in sod. All of the Northern vowels are more front than the Southern vowels. Thus, the Northern talkers...
exhibit fronting and lowering of /a/, consistent with the Northern Cities vowel shift. The Northern vowels are not clearly distinguished by duration, but the overall formant trajectory is longer for side than for sod for both talkers. Despite the considerable variation in the acoustic characteristics of the vowels across talkers, for all four talkers, the vowels in sod and side are well separated in the acoustic space, as shown by the vowel distance measures in table 7. Further, although the overall F2 of the two vowels and the relative degree of diphthongization as captured by the trajectory length differ across the two dialects, spectral differences remain a robust cue for vowel category across dialects.

The target words are somewhat less well matched for log frequency than in experiment 1 (sod: 1.48; side: 3.58), given the limited number of minimal pairs in English containing the target contrast. An inspection of the responses confirmed that side was selected as the response on 53% of the trials, suggesting a modest frequency-based response bias. Each token was stored in an individual digital audio file with a sampling rate of 22,050 Hz and 16-bit resolution.

Table 7. The total duration of each vowel token, the total trajectory length of each vowel token between the 35% and 65% points (calculated as the summed Euclidean distance in the F1 × F2 plane between the 35% and 50% points and the 50% and 65% points), the average acoustic distance between each vowel pair (calculated as the mean Euclidean distance in the F1 × F2 plane at the 35%, 50%, and 65% points), and the total duration of each word token in experiments 2A and 2B

<table>
<thead>
<tr>
<th>Target</th>
<th>Southern A</th>
<th>Southern B</th>
<th>Northern A</th>
<th>Northern B</th>
</tr>
</thead>
<tbody>
<tr>
<td>Vowel duration, ms sod</td>
<td>286</td>
<td>326</td>
<td>278</td>
<td>273</td>
</tr>
<tr>
<td>side</td>
<td>352</td>
<td>388</td>
<td>379</td>
<td>273</td>
</tr>
<tr>
<td>Vowel trajectory, Hz sod</td>
<td>149</td>
<td>99</td>
<td>141</td>
<td>70</td>
</tr>
<tr>
<td>side</td>
<td>297</td>
<td>430</td>
<td>646</td>
<td>376</td>
</tr>
<tr>
<td>Vowel distance, Hz sod/side</td>
<td>332</td>
<td>420</td>
<td>461</td>
<td>414</td>
</tr>
<tr>
<td>Word duration, ms sod</td>
<td>578</td>
<td>723</td>
<td>584</td>
<td>659</td>
</tr>
<tr>
<td>side</td>
<td>548</td>
<td>712</td>
<td>810</td>
<td>619</td>
</tr>
</tbody>
</table>
3.1.3. Procedures

The procedures were identical to those used in experiment 1A, except that the response alternatives were *sod* (left button) and *side* (right button).

3.1.4. Results and Discussion

As in experiment 1, performance in the lexical classification task was excellent; overall accuracy was 91%. A summary of the accuracy and response times in the lexical classification task for each talker dialect in each block is shown in Table 8. A repeated-measures ANOVA of lexical classification accuracy with block (single-talker vs. mixed-talker) and talker dialect (North vs. South) as within-subject variables revealed no significant main effects or interactions.

A parallel repeated-measures ANOVA of lexical classification response times revealed significant main effects of block \( [F(1, 20) = 22.76, p < 0.001] \) and talker dialect \( [F(1, 20) = 11.61, p = 0.003] \). The block × talker dialect interaction was not significant. Responses were faster for the Southern talkers than for the Northern talkers and in the single-talker blocks than in the mixed-talker blocks.

Taken together, the results of this experiment reveal that lexical classification responses were faster in the single-talker blocks than in the mixed-talker blocks, consistent with the talker interference effects observed in experiment 1. The block effects in experiments 1A and 2A were comparable in magnitude (65 and 67 ms, respectively), providing further evidence for the robustness of the within-dialect talker interference effect across dialects and vowel categories. Lexical classification performance was also faster for the Southern talkers than for the Northern talkers. However, the Southern stimulus tokens were also shorter in duration than the Northern stimulus tokens. Unlike in experiment 1, a linear regression model predicting mean response time for each target stimulus from talker dialect accounted for less variance \( (r^2 = 0.25, \text{d.f.} = 6, \text{n.s.}) \) than a comparable model with word duration as a predictor \( (r^2 = 0.55, \text{d.f.} = 6, p = 0.035) \). Thus, the faster performance on trials with Southern talkers than Northern talkers by the Midland listeners may reflect properties of the stimulus materials rather than suggesting that stereotyped non-local variants (i.e., Southern /ɑj/) are easier to process than non-stereotyped non-local variants (i.e., Northern /a/).

As in experiment 1, the block × talker dialect interactions were not significant, suggesting that although within-dialect talker variability reduced the speed of lexical classification performance, mixing two talkers of a stereotyped variety did not create significantly more or less interference than mixing two talkers of a non-stereotyped variety (see also experiment 2C in section 3.3). The lexical classification task

<table>
<thead>
<tr>
<th>Talker dialect</th>
<th>Accuracy</th>
<th>Response time, ms</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>single-talker</td>
<td>mixed-talker</td>
</tr>
<tr>
<td>South</td>
<td>0.94 (0.06)</td>
<td>0.93 (0.08)</td>
</tr>
<tr>
<td>North</td>
<td>0.90 (0.12)</td>
<td>0.89 (0.09)</td>
</tr>
</tbody>
</table>

**Table 8.** Mean (SD) lexical classification accuracy and response times for each talker dialect in each block in experiment 2A

Phonetica 2017;74:25–59
DOI: 10.1159/000446809
Clopper
3.2. Experiment 2B: Cross-Dialect Interference

3.2.1. Listeners

Sixty-four undergraduates at Ohio State University participated as listeners in experiment 2B. They received partial course credit in an introductory linguistics course for participating. Data from listeners who were bilingual (n = 9), who reported a history of a speech or hearing disorder (n = 5), who used only one hand to make their responses (n = 1), who did not meet the residential history requirements (n = 27), or whose mean response times were more than 2 SDs slower than the mean across all participants (n = 2) were excluded prior to the analysis. The remaining 20 listeners (16 female, 4 male) were monolingual native speakers of American English who were lifetime residents of the Midland dialect region and ranged in age from 18 to 28 years (mean = 20, SD = 2).

3.2.2. Stimulus Materials

The stimulus materials consisted of one token of the words sod and side produced by Northern talker A and Southern talker A from experiment 2A, for a total of four stimulus tokens (see fig. 3).

3.2.3. Procedures

The procedures were identical to those used in experiment 1B, except that the response alternatives were sod and side in the lexical classification task. As in experiment 1B, performance on the dialect classification task was not significantly different from chance and will not be discussed further.

3.2.4. Results and Discussion

Overall accuracy in the lexical classification task was excellent (81%), although somewhat lower than the accuracy observed in experiment 2A. A summary of the accuracy and response times in the lexical classification task is shown in table 9 for each talker dialect in each block. A repeated-measures ANOVA of mean lexical classification accuracy with block (single-talker vs. mixed-talker) and talker dialect (North vs. South) as within-subject variables revealed a significant main effect of block \([F(1, 19) = 8.94, p = 0.008]\). Performance was more accurate in the single-talker blocks than in the mixed-talker block. The main effect of talker dialect and the block \times talker dialect interaction were not significant.

<table>
<thead>
<tr>
<th>Talker dialect</th>
<th>Accuracy</th>
<th>Response time, ms</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>single-talker</td>
<td>mixed-talker</td>
</tr>
<tr>
<td>South</td>
<td>0.81 (0.25)</td>
<td>0.79 (0.24)</td>
</tr>
<tr>
<td>North</td>
<td>0.88 (0.17)</td>
<td>0.78 (0.16)</td>
</tr>
</tbody>
</table>

Table 9. Mean (SD) lexical classification accuracy and response times for each talker dialect in each block in experiment 2B
A parallel repeated-measures ANOVA of lexical classification response times revealed significant main effects of block \( [F(1, 19) = 12.75, p = 0.002] \) and talker dialect \( [F(1, 19) = 54.55, p < 0.001] \). Responses were faster in the single-talker blocks than in the mixed-talker block and for the Southern talker than for the Northern talker. The block × talker dialect interaction was not significant.

Taken together, the results of this experiment revealed slower and less accurate performance in the mixed-talker block than in the single-talker blocks, consistent with the results of experiment 1, in which significant talker interference effects were also observed. The analysis of the response times also revealed significantly slower responses for the Northern talker than for the Southern talker. Unlike in experiment 2A, however, the effect of block was significant for accuracy in this experiment.

3.2.5. Talker versus Dialect Interference (Experiments 2A and 2B)

To further explore the differences in the block and talker dialect effects between experiments 2A and 2B, a repeated-measures ANOVA was conducted on lexical classification accuracy with block (single-talker vs. mixed-talker) and talker dialect (North vs. South) as within-subject variables and experiment (2A vs. 2B) as a between-subject variable. The analysis revealed significant main effects of block \( [F(1, 39) = 9.90, p = 0.003] \) and experiment \( [F(1, 39) = 4.53, p = 0.040] \), as well as a significant experiment × block interaction \( [F(1, 39) = 4.68, p = 0.037] \). Performance was more accurate in the single-talker blocks than the mixed-talker blocks and in experiment 2A than in experiment 2B. The significant experiment × block interaction reflects a larger block effect in experiment 2B, in which the mixed-talker block included talkers from different dialects, than in experiment 2A, in which the mixed-talker blocks included talkers from the same dialect. This result confirms a dialect interference effect on accuracy that exceeds a simple talker interference effect. Unlike in experiment 1, however, this dialect interference effect was consistent across the two talker dialects and the three-way experiment × block × talker dialect interaction was not significant.

A parallel repeated-measures ANOVA of lexical classification response times revealed significant main effects of experiment \( [F(1, 39) = 12.14, p = 0.001] \), block \( [F(1, 39) = 31.39, p < 0.001] \), and talker dialect \( [F(1, 39) = 56.16, p < 0.001] \), as well as a significant experiment × talker dialect interaction \( [F(1, 39) = 6.99, p = 0.012] \). The other interactions were not significant. Although responses were slower overall in experiment 2B than in experiment 2A, the slowing across experiments was greater for the Northern talker (139 ms) than for the Southern talker (62 ms), especially in the mixed-talker block (see tables 8, 9). This interaction suggests that although the overall effect of talker dialect on response times in experiment 2A may reflect a simple effect of stimulus duration, the larger effect of talker dialect in experiment 2B likely reflects dialect interference for the Northern talkers, above and beyond the stimulus duration effect. That is, the experiment × talker dialect interaction provides indirect evidence for a dialect interference effect on response times for Northern talkers that exceeds within-dialect talker interference. This result parallels the comparable effect on accuracy in experiment 1, in which the effect of experiment (i.e., within- vs. cross-dialect mixed-talker blocks) was greater for the Northern talker than for the Midland talker.

This comparison between experiments 2A and 2B thus reveals further evidence for dialect interference effects beyond simple talker interference effects and suggests
that dialect interference effects can be observed with two non-local varieties. For Midland listeners, mixing Northern and Southern talkers in a single experimental block results in significant dialect interference effects on accuracy for both talker dialects and on response times for Northern talkers only. This pattern differs from the dialect interference effects observed in experiment 1, in which accuracy effects were observed primarily for Northern talkers and response time effects were observed primarily for Midland talkers. However, the combined effects of dialect interference on accuracy and response times for the Northern talkers may also reflect a speed-accuracy trade-off, as in experiment 1. In particular, the Midland listeners may sacrifice accuracy for both Northern and Southern talkers, but also sacrifice response times for the Northern talkers to avoid an even larger reduction in accuracy.

As in experiment 1, the results of experiments 2A and 2B are based on a single pair of target words produced by two pairs of talkers in experiment 2A and by a single pair of talkers in experiment 2B. Replication of the results with additional materials is therefore critical for ensuring that the results do not reflect mere idiosyncrasies of the stimulus tokens (Clark, 1973; Larzelere et al., 2015; Monin and Oppenheimer, 2014; Westfall et al., 2015). In experiments 2C and 2D, the effects of talker and dialect interference were confirmed for Northern and Southern talkers with a different pair of target words (rod, ride), two pairs of talkers in both the talker interference (experiment 2C) and dialect interference (experiment 2D) conditions, and a more heterogeneous Midland listener population, recruited from the local science museum instead of the university campus.

3.3. Experiment 2C: Within-Dialect Talker Interference Replication

3.3.1. Listeners

The same listeners from experiment 1C also completed experiment 2C in a single experimental session.3

3.3.2. Stimulus Materials

The same four talkers were used in this experiment as in experiment 2A. The stimulus materials consisted of one token of each of the words rod and ride produced by each of the four talkers, for a total of eight stimulus tokens. The formant trajectories for each of the two vowels are illustrated in figure 4 for the Southern (darker symbols) and Northern (lighter symbols) talkers for rod and ride. For each vowel, the first and second formant frequencies were estimated at 35%, 50%, and 65% of the vowel duration. Each trajectory is labeled in figure 4 at the beginning of the trajectory with the talker identifier (NoA, NoB, SoA, or SoB) and the target word (rod or ride). The total duration of each vowel token, the total trajectory length of each vowel token, the average acoustic distance between each vowel pair, and the total duration of each word token are provided in table 10.

For the Southern talkers, the vowel in ride is more front and longer in duration than the vowel in rod. The trajectory of ride is also longer than the trajectory of rod for both Southern talkers, although the relatively short trajectories of ride confirm some monophthongization of /æj/ for both talkers, as expected. For Northern talker A, the vowel in ride is more front than the vowel in rod, and for Northern talker B, the vowel

3 Of the 6 participants who had lived outside the Midland dialect region after the age of 18 years, 1 had lived in the Northern region and 2 had lived in the Southern region.
in *ride* is higher than the vowel in *rod*. The Northern *rod* vowels are more front than the Southern *rod* vowels. Thus, the Northern talkers exhibit fronting and lowering of /ɑ/, consistent with the Northern Cities vowel shift. The Northern vowels are less clearly distinguished by duration, but the overall formant trajectory is longer for *ride* than for *rod* for both talkers. Despite the considerable variation in the acoustic characteristics of the vowels across talkers, for all four talkers, the vowels in *rod* and *ride* are well separated in the acoustic space, as shown by the vowel distance measure in table 10. Further, although the overall F2 of *rod* and the relative degree of diphthongization of *ride* as captured by the trajectory length differ across the two dialects, spectral differences remain a robust cue for vowel category across dialects.

The target words are reasonably well matched for log frequency (*rod*: 2.26; *ride*: 2.69) to avoid potential item-specific biases in lexical classification. Each token was stored in an individual digital audio file with a sampling rate of 22,050 Hz and 16-bit resolution.

### Table 10. The total duration of each vowel token, the total trajectory length of each vowel token between the 35% and 65% points (calculated as the summed Euclidean distance in the F1 × F2 plane between the 35% and 50% points and the 50% and 65% points), the average acoustic distance between each vowel pair (calculated as the mean Euclidean distance in the F1 × F2 plane at the 35%, 50%, and 65% points), and the total duration of each word token in experiments 2C and 2D

<table>
<thead>
<tr>
<th>Target</th>
<th>Southern A</th>
<th>Southern B</th>
<th>Northern A</th>
<th>Northern B</th>
</tr>
</thead>
<tbody>
<tr>
<td>Vowel duration, ms</td>
<td>rod</td>
<td>277</td>
<td>248</td>
<td>276</td>
</tr>
<tr>
<td></td>
<td>ride</td>
<td>309</td>
<td>322</td>
<td>321</td>
</tr>
<tr>
<td>Vowel trajectory, Hz</td>
<td>rod</td>
<td>39</td>
<td>154</td>
<td>153</td>
</tr>
<tr>
<td></td>
<td>ride</td>
<td>347</td>
<td>363</td>
<td>657</td>
</tr>
<tr>
<td>Vowel distance, Hz</td>
<td>rod/ride</td>
<td>443</td>
<td>361</td>
<td>371</td>
</tr>
<tr>
<td>Word duration, ms</td>
<td>rod</td>
<td>432</td>
<td>484</td>
<td>592</td>
</tr>
<tr>
<td></td>
<td>ride</td>
<td>477</td>
<td>608</td>
<td>540</td>
</tr>
</tbody>
</table>

![Fig. 4. First and second formant frequencies of the two Southern (SoA, SoB) and two Northern (NoA, NoB) *rod* and *ride* tokens at 35%, 50%, and 65% of the vowel duration. Trajectories are labeled at the 35% point. All tokens were also used in experiment 2D. Black: SoA; dark gray: SoB; light gray: NoB; white: NoA. Circles denote *ride* tokens and triangles denote *rod* tokens.](image-url)
3.3.3. Procedure
The procedures were identical to those used in experiment 1C, except that the response alternatives were rod (left button) and ride (right button).

3.3.4. Results and Discussion
As in experiment 2A, performance in the lexical classification task was excellent; overall accuracy was 93%. A summary of the accuracy and response times in the lexical classification task for each talker dialect in each block is shown in table 11. A repeated-measures ANOVA of mean lexical classification accuracy with block (single-talker vs. mixed-talker) and talker dialect (North vs. South) as within-subject variables revealed no significant main effects or interactions, replicating the results of experiment 2A.

A parallel repeated-measures ANOVA of lexical classification response times revealed a significant main effect of block \[F(1, 21) = 5.31, p = 0.032\]. The main effect of talker dialect and the block × talker dialect interaction were not significant. As in experiment 2A, responses were faster in the single-talker block than in the mixed-talker block.

The results of this experiment partially replicate the results of experiment 2A with different target words, with different stimulus materials, and with a more heterogeneous Midland listener population with respect to age and residential history: speeded lexical classification accuracy was near ceiling, no significant effects of block or talker dialect were observed for accuracy, and the effect of block was significant for response times, consistent with a talker interference effect. However, unlike in experiment 2A, no effect of talker dialect was observed for response times. Consistent with this lack of a talker dialect effect on response times, a linear regression model predicting the mean response time for each target stimulus from talker dialect was also not significant \((r^2 = 0.27, \text{d.f.} = 6, \text{n.s.})\). Although an inspection of table 10 reveals that the Northern target words were longer in duration than the Southern target words, a linear regression model predicting the mean response time for each target stimulus from word duration was also not significant \((r^2 = 0.21, \text{d.f.} = 6, \text{n.s.})\), suggesting that response times were not a simple reflection of variation in stimulus duration in this experiment. Further, as in experiment 2A, the interaction between talker dialect and block was not significant for either accuracy or response times, suggesting comparable talker interference effects within each dialect. Experiment 2D was designed as a replication of experiment 2B with different target words (rod and ride), two mixed-dialect talker pairs, and the more heterogeneous Midland listener population to further confirm the effect of dialect interference above and beyond the talker interference effect for these two dialects for Midland listeners.

Table 11. Mean (SD) lexical classification accuracy and response times for each talker dialect in each block in experiment 2C

<table>
<thead>
<tr>
<th>Talker dialect</th>
<th>Accuracy</th>
<th>Response time, ms</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>single-talker</td>
<td>mixed-talker</td>
</tr>
<tr>
<td>South</td>
<td>0.93 (0.09)</td>
<td>0.93 (0.09)</td>
</tr>
<tr>
<td>North</td>
<td>0.94 (0.09)</td>
<td>0.93 (0.10)</td>
</tr>
</tbody>
</table>
3.4. Experiment 2D: Cross-Dialect Interference Replication

3.4.1. Listeners

The same listeners from experiment 1D also completed experiment 2D in a single experimental session.4

3.4.2. Stimulus Materials

The same eight stimulus tokens produced by the same four talkers were used in this experiment as in experiment 2C.

3.4.3. Procedure

The procedures were identical to those used in experiment 1D, except that the response alternatives were rod (left button) and ride (right button).

3.4.4. Results and Discussion

As in the previous experiments, performance in the lexical classification task was excellent; overall accuracy was 92%. A summary of the accuracy and response times in the lexical classification task for each talker dialect in each block is shown in table 12. A repeated-measures ANOVA of mean lexical classification accuracy with block (single-talker vs. mixed-talker) and talker dialect (North vs. South) as within-subject variables revealed no significant main effects or interactions, unlike in experiment 2B, in which performance was less accurate overall and a main effect of block was observed for accuracy.

A parallel repeated-measures ANOVA of lexical classification response times revealed a marginal main effect of block \([F(1, 20) = 3.97, p = 0.060]\) and a significant main effect of talker dialect \([F(1, 20) = 14.92, p < 0.001]\). The block × talker dialect interaction was not significant. As in experiment 2B, responses were faster for the Southern talkers than for the Northern talkers and in the single-talker blocks than in the mixed-talker blocks.

The results of this experiment therefore partially replicate the results of experiment 2B: a significant effect of talker dialect was observed for response times. However, unlike in experiment 2B, the effect of block was only marginal for response times and was not significant for accuracy. These results also contrast with the results of experiment 2C, in which a more robust effect of block was observed for response times, but the effect of talker dialect was not significant.

4 Of the 7 participants who had lived outside the Midland dialect region after the age of 18 years, 2 had lived in the Northern region and 1 had lived in the Southern region.

<table>
<thead>
<tr>
<th>Talker dialect</th>
<th>Accuracy</th>
<th>Response time, ms</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>single-talker</td>
<td>mixed-talker</td>
</tr>
<tr>
<td>South</td>
<td>0.92 (0.06)</td>
<td>0.93 (0.08)</td>
</tr>
<tr>
<td>North</td>
<td>0.93 (0.08)</td>
<td>0.92 (0.09)</td>
</tr>
</tbody>
</table>
3.4.5. Talker versus Dialect Interference (Experiments 2C and 2D)

To further explore the differences in the block and talker dialect effects between experiments 2C and 2D, a repeated-measures ANOVA was conducted on lexical classification accuracy with block (single-talker vs. mixed-talker) and talker dialect (North vs. South) as within-subject variables and experiment (2C vs. 2D) as a between-subject variable. The analysis revealed a marginal block × talker dialect interaction \([F(1, 41) = 3.97, p = 0.053]\). The main effects of experiment, block, and talker dialect, as well as the other interactions, were not significant. The marginal block × talker dialect interaction is driven primarily by the accuracy pattern observed in experiment 2D (see tables 11, 12), in which performance was slightly more accurate for the Southern talkers in the mixed-talker block than in the single-talker block, although neither the block × talker dialect interaction in experiment 2D nor the three-way experiment × block × talker dialect interaction in this analysis reach significance.

A parallel repeated-measures ANOVA of lexical classification response times revealed significant main effects of block \([F(1, 41) = 9.27, p = 0.004]\) and talker dialect \([F(1, 41) = 8.16, p = 0.007]\). The main effect of experiment and the interactions were not significant. These results reflect slower overall performance in mixed-talker blocks than in single-talker blocks and for Southern talkers than for Northern talkers. Unlike in the comparison of experiments 2A and 2B, no significant effects or interactions involving experiment were observed in the comparison of experiments 2C and 2D. Thus, although the talker dialect effect was significant for response times in experiment 2D, but not in experiment 2C, comparable to the pattern of dialect interference on response times for Northern talkers observed for experiments 2A and 2B, the experiment × talker dialect interaction was not significant in this analysis.

Together, this set of four experiments provided some additional evidence for dialect interference effects on speeded lexical classification accuracy and response times for two non-local dialects for Midland listeners, although the effects were less consistent than those observed in experiment 1. In particular, more robust interference was observed overall in experiment 2B than in experiment 2D. In experiment 2B, dialect interference was observed for both talker dialects for accuracy and for Northern talkers for response times, but in experiment 2D, no dialect interference was observed for accuracy and only weak evidence was obtained for a dialect interference effect on response times for the Northern talkers.5

4. General Discussion

4.1. Dialect Interference

The significant effects and interactions obtained from the eight experiments are summarized in table 13. Interference effects in the mixed-talker blocks were observed for response times in all eight experiments (although the effect was marginal in experiment 2D), and in five of the eight experiments for accuracy, consistent with previous research demonstrating indexical interference effects in speech processing (Creelman, 1957; Mullennix and Pisoni, 1990). The effects may have been more robust across

5 An examination of the data for the subset of lifetime Midland participants in experiments 2C and 2D did not provide any additional evidence for robust dialect interference effects for the museum visitor population.
In addition, evidence for dialect interference effects beyond talker interference effects was obtained in both experiments 1 and 2. In experiment 1, direct evidence for dialect interference for Midland talkers was obtained in the experiment × block × talker dialect interactions for response times, and additional indirect evidence for dialect interference for Northern talkers was provided by the experiment × talker dialect interactions for accuracy and response times. Dialect interference was more robustly observed in experiment 1 in response times than in accuracy, consistent with the near-ceiling performance in this relatively easy task.

In experiments 2A and 2B, direct evidence for dialect interference was obtained for both Northern and Southern talkers in the experiment × block interaction for accuracy, and additional indirect evidence for dialect interference for Northern talkers was provided by the experiment × talker dialect interaction for response times. In experiment 2B, dialect interference was more robustly observed in accuracy than in response times, consistent with the lower overall performance observed in this experiment than in any of the other experiments. Although experiment 2D provided modest evidence for the dialect interference effect on response times for the Northern talkers, the dialect interference effect on accuracy for both talker dialects was not observed in experiment 2D. The reason for the lack of a replication of the accuracy results is not clear, but it may reflect either the more variable experiences of the museum visitor population due to their larger age range or more diverse residential histories, or a ceiling effect on accuracy that masks any effects of dialect familiarity or enregisterment on performance. Thus, the evidence for dialect interference was stronger in experiment 1 than in experiment 2.

Together, the results of this study provide evidence for the hypothesis that trial-to-trial normalization for dialect variation involving potential phonological ambiguity is more costly than trial-to-trial normalization for within-dialect talker variability. Crucially, dialect interference effects were observed for all three dialects (Midland, Northern, and Southern) in both of the combinations that were examined (Midland/Northern and Southern/Northern), suggesting that even the local, native variety is not immune to processing costs associated with trial-to-trial cross-dialect normalization processes.

### Table 13. Summary of the significant effects in each of the eight experiments

<table>
<thead>
<tr>
<th>Experiment</th>
<th>Interference type</th>
<th>Accuracy</th>
<th>RTs</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>block</td>
<td>dialect</td>
</tr>
<tr>
<td>1A</td>
<td>within-dialect</td>
<td>*</td>
<td>*</td>
</tr>
<tr>
<td>1B</td>
<td>between-dialect</td>
<td>*</td>
<td>*</td>
</tr>
<tr>
<td>1C</td>
<td>within-dialect</td>
<td>*</td>
<td>*</td>
</tr>
<tr>
<td>1D</td>
<td>between-dialect</td>
<td>*</td>
<td>*</td>
</tr>
<tr>
<td>2A</td>
<td>within-dialect</td>
<td>*</td>
<td>*</td>
</tr>
<tr>
<td>2B</td>
<td>between-dialect</td>
<td>*</td>
<td>*</td>
</tr>
<tr>
<td>2C</td>
<td>within-dialect</td>
<td></td>
<td>*</td>
</tr>
<tr>
<td>2D</td>
<td>between-dialect</td>
<td>(*)</td>
<td>*</td>
</tr>
</tbody>
</table>

* Significant main effect or interaction (p < 0.05). (*) Marginal effect (p = 0.06). Shaded cells: significant cross-experiment interactions (i.e., experiment × block, experiment × talker dialect, and experiment × block × dialect, respectively, at p < 0.05).
Although dialect interference was observed for all three dialects, the pattern of interference varied across dialects and dialect pairs. In experiment 1, dialect interference effects were directly observed for response times for the Midland talkers and were indirectly observed for accuracy for the Northern talkers. In experiment 2B, dialect interference effects were directly observed for accuracy for both Southern and Northern talkers and were indirectly observed for response times for the Northern talkers. Modest indirect evidence for a dialect interference effect on response times for the Northern talkers was also obtained in experiment 2D. Thus, when the local Midland dialect was paired with the non-local Northern dialect, interference effects were observed in different domains for the two dialects, but when two non-local dialects were paired, interference effects on accuracy were observed for both dialects, and interference was observed on response times only for the non-enregistered Northern dialect. This pattern of results suggests that dialect interference effects emerge most robustly in accuracy when performance is not at ceiling (i.e., for non-local talkers), whereas interference in response times can emerge for both easier (i.e., local) and more difficult (i.e., non-local) dialects, depending on other task constraints, such as overall speed and accuracy.

In particular, in experiment 1, responses to the local Midland talkers were faster and more accurate than responses to the non-local Northern talkers overall. In the mixed-dialect blocks, responses to the Midland talkers were slowed, but responses to the Northern talkers were not. Instead, the accuracy of the responses to the Northern talkers was reduced. Thus, performance was slowed for talkers of the easy, local Midland dialect, but accuracy was reduced for talkers of the more difficult, Northern dialect. In experiment 2B, responses to the non-local Southern talkers were faster, but not more accurate, than responses to the non-local Northern talkers. In the mixed-dialect blocks, the accuracy of the responses to both dialects was reduced, consistent with the realization of the dialect interference effect in accuracy for non-local dialects. However, responses to the non-stereotyped Northern dialect were also slowed, whereas responses to the stereotyped Southern dialect were not, suggesting an effect of social stereotyping on the magnitude of the dialect interference effect in response times (see also section 4.2).

These effects of talker and dialect variability on lexical processing can be interpreted as interference due to simultaneous processing of linguistic and indexical information (Mullennix and Pisoni, 1990). In a mixed-talker block, normalization for talker variability is more difficult than in a single-talker block, and the normalization process therefore interferes with lexical access and word recognition (Mullennix et al., 1989). Crucially, the block × talker dialect interaction was not significant in any of the within-dialect talker interference experiments, confirming that this within-dialect talker variability effect is comparable in magnitude across local and non-local dialects. This pattern of results suggests that listeners can quickly adapt to dialect differences, and then normalize equally well for talker variability within the given dialect space, regardless of dialect familiarity. This adaptation critically does not require explicit recognition of the talker dialects, given that performance in the dialect classification tasks in experiments 1B and 2B was at chance level.

In a mixed-dialect block, normalization for dialect variability is more difficult than normalization for within-dialect talker variability, and a larger interference effect of normalization on lexical processing is observed. This dialect interference effect was obtained both for local/non-local (experiment 1) and, somewhat less robustly, for
non-local/non-local (experiment 2) dialect pairs. The difference in the interference effect across dialects suggests that the listener’s ability to quickly map the signal onto the correct linguistic category is a function of his or her familiarity with the variety. In particular, speed is sacrificed when a non-local dialect interferes with a local dialect, but accuracy is sacrificed when non-local dialects compete. In the latter case, speed may also be sacrificed for a socially non-enregistered dialect.

Thus, the results of this study reveal effects of dialect familiarity on the magnitude of the dialect interference effect, but no effects of dialect familiarity on the within-dialect talker interference effect. Exploring the effects of talker variability both within and across dialects on lexical classification performance therefore allows us to observe gradience in the costs associated with the normalization process: despite overall differences in lexical classification speed and accuracy across dialects, the costs associated with within-dialect talker normalization do not vary across dialects. However, the costs associated with normalizing across dialects do vary as a function of dialect familiarity, which can be realized through both native dialect experience and shared social enregisterment (see section 4.2). This gradience in the cost of normalization may reflect a hierarchy of normalization corresponding to the hierarchy of indexical categories to which talkers belong, with individual talker identity embedded within regional dialect. If the regional dialect is constant, talker normalization has a consistent cost, whereas if the regional dialect varies, the cost of normalization also varies. Alternatively, the gradience in the cost of normalization may reflect gradience in talker similarity, with talkers sharing a dialect being perceived as intrinsically more similar than talkers with different dialects. If the regional dialect is constant, talker similarity is constrained in particular dimensions and talker normalization has a consistent cost, whereas if the regional dialect varies, the degree of talker similarity is more variable, leading to variability in the cost of normalization. More research that explicitly examines the complex network of socioindexical categories to which any individual talker belongs is necessary to determine the structure of these categories and how the lexical processing system handles the variability associated with them.

An alternative interpretation of the significant block effect observed in all eight experiments in the current study is that it reflects facilitation of linguistic processing in the single-talker block relative to the mixed-talker or mixed-dialect block. Normalization for a single talker is easier than for multiple talkers or dialects, and therefore interferes less with lexical access and word recognition. In an exemplar-based account (e.g., Johnson, 1997), the recently encountered exemplars from the talker in a single-talker block facilitate talker normalization by raising the activation level of relevant talker-specific exemplars. Similarly, the recently encountered exemplars from the talkers in the single-dialect block facilitate talker normalization relative to a mixed-dialect block by raising the activation level of relevant dialect-specific exemplars. Although the interference and facilitation interpretations of the talker variability effects implicitly suggest different processing baselines, the two accounts are equivalent in that they both describe the significant difference observed between the various processing conditions. Given that overall performance in the single-talker blocks in the current set of experiments was near ceiling, the results are discussed with respect to interference, but the theoretical arguments are also compatible with a facilitation interpretation.

The dialect interference effects in the current study were elicited using two stimulus tokens from each of two talkers in each experiment. This design was selected...
to maximize the likelihood of observing significant dialect interference effects (cf. Mullennix and Pisoni, 1990). If more talker and/or token variability were introduced in the experiment, the response times would most likely be slower and more variable overall, leading to smaller differences within and across experiments. That is, variation in irrelevant dimensions (i.e., word and voice) would potentially obscure variation in the dimension of theoretical interest (i.e., talker dialect). Research examining group-level indexical variation, such as gender or regional dialect, is always faced with this trade-off between ideal random sampling of stimulus materials (Baayen et al., 2008; Clark, 1973; Judd et al., 2012) and the reality that larger samples of stimulus materials contribute to noisier data. One common solution to this problem is to take the approach used in the current study and select a small number of stimulus items that are representative of the target categories, either based on perceptual judgments from naïve listeners (e.g., Johnson et al., 1999) or based on expert assessment by the researcher (e.g., Floccia et al., 2006; Sumner and Samuel, 2009). This solution increases the generalizability of the results obtained with small stimulus sets, especially when the results can be replicated with different representative materials, as in the current study (Wells and Windschitl, 1999).

Thus, although generalization from this small number of tokens should proceed cautiously, the internal direct replication in experiment 1 provides some evidence for the generalizability of the results. The internal direct replication in experiment 2 was less robust, but several additional findings suggest that the primary findings will generalize to other stimulus materials produced by other talkers from these dialects. First, the acoustic analysis confirmed that these tokens are representative of their respective dialects (see fig. 1–4), suggesting that other, equally representative stimulus materials would produce similar results. Second, dialect interference effects were observed in both experiments 1 and 2, which involved different dialect pairs and different vowel variables, confirming that the dialect interference effect is replicable across different levels of these two factors. Thus, the results from experiment 2 provide an internal conceptual replication of the results obtained from experiment 1.

The primary caveat that remains with regard to generalization from these data relates to the selected target words. In particular, the minimal pairs in each experiment were selected to exploit the phonological ambiguity of the target vowel variants across dialects. The dialect interference effects observed in this study may therefore reflect this critical aspect of the stimulus materials. Dialect interference is therefore predicted to emerge in similar situations involving potentially phonologically ambiguous segments across dialects, but dialect interference may not emerge in other situations in which dialect-specific variants are not phonologically ambiguous. The full extent of dialect interference effects therefore remains to be determined.

4.2. Familiarity, Localness, and Enregisterment

The variation in overall accuracy for the three dialects provides some insight into the roles of familiarity and social enregisterment in cross-dialect lexical processing, independent of the dialect interference effects that were the central focus of this study. Performance on the local Midland talkers was consistently at 97% accuracy throughout experiment 1, but performance on the non-local Northern talkers ranged from 83 to 96% accuracy across experiments, and performance on the non-local Southern talkers ranged from 80 to 93% accuracy in experiment 2. These results are consistent with previous research suggesting that familiarity with a specific dialect can facilitate speech
processing (Clopper, 2014; Floccia et al., 2006; Sumner and Samuel, 2009; see also Nygaard and Pisoni, 1998; Souza et al., 2013, on comparable talker familiarity effects).

However, performance was also slower for the Northern talkers than for both the Midland and Southern talkers overall. Although the Northern stimulus tokens were longer in duration than the Midland and Southern stimulus tokens, regression analyses revealed that stimulus duration alone was not sufficient to account for the response times, suggesting that word recognition may be more difficult for non-local, non-enregistered variants than for local variants and non-local stereotyped variants. That is, explicit awareness of the social significance of non-local variants, such as Southern monophthongal /ɑ/, may speed up lexical processing relative to non-local variants that are not explicitly socially marked, although it does not contribute to overall accuracy. This result suggests that the processing costs associated with non-local, enregistered dialects, such as the Southern dialect for Midland listeners, are intermediate between the minimal processing costs associated with local dialects and the more extreme processing costs associated with non-local, non-enregistered dialects, such as the Northern dialect for Midland listeners. Crucially, accurately identifying the dialects from their variant forms is not essential for observing these effects, because dialect classification performance was not different from chance in either experiment 1B or 2B. Thus, the explicit activation of the relevant social categories is not a necessary precondition for observing differential effects of dialect on lexical classification accuracy.

The observed variation in processing costs across dialects for the Midland listeners further suggests variation in the representation of these dialect-specific variants. Local, native variants are the most robustly encoded, as a result of frequent exposure (Clopper, 2014). Non-local, socially enregistered variants are robustly encoded as a result of social salience due to stereotyping (see also Sumner and Kataoka, 2013), although their encoding may be weaker overall than the encoding of local variants. Non-local, non-enregistered variants are weakly encoded as a result of limited exposure and/or limited social salience. These non-local, non-enregistered variants are predicted to include both the variants associated with a non-enregistered variety, such as the Northern /ɛ, æ, ɑ/ examined in this study, as well as non-stereotyped variants associated with enregistered varieties, such as Southern /u/-fronting. This interpretation therefore allows for differential encoding of dialect-specific variants within and across dialects as a function of their social salience due to stereotyping and enregisterment.

Listeners from different regions are therefore expected to exhibit different cross-dialect lexical processing patterns. For example, Southern listeners are expected to show near-ceiling performance for their native Southern dialect (Labov and Ash, 1997; Mason, 1946), as well as for the relatively standard Midland dialect (Clopper and Bradlow, 2008). However, like the Midland listeners, they are expected to perform more poorly on the non-local, non-enregistered Northern dialect. By contrast, Northern listeners are expected to perform like the Midland listeners on the non-local, enregistered Southern dialect and the familiar, standard Midland dialect. Given that the Northern dialect is non-enregistered even for Northern listeners (Clopper and Bradlow, 2008; Niedzielski, 1999), however, they may exhibit intermediate performance for the local, Northern dialect rather than the expected near-ceiling performance for familiar, native varieties (see also Clopper, 2014).

Performance on other non-local dialects is predicted to pattern along the dimensions of exposure and social enregisterment, as observed in the current study. Thus, American English listeners are predicted to exhibit better performance on socially
stereotyped varieties relative to varieties that are not socially marked, as in experiment 2. In addition, given Sumner and Kataoka’s (2013) findings that prestigious non-local varieties, such as standard Southern British English, are processed differently by General American listeners than non-prestigious, non-local varieties, such as New York City English, the prestige associated with the socially enregistered varieties may also impact speeded lexical processing. Thus, whereas Southern American English enjoys some covert prestige among non-Southerners as sounding pleasant (Niedzielski and Preston, 2000), which may facilitate processing relative to the non-enregistered Northern dialect, a variety like New York City English, which is generally not perceived as prestigious by non-locals, may exhibit intermediate performance between a more prestigious variety and a non-enregistered variety.

The comparisons in the current study further suggest that familiarity can result from many different kinds of exposure, including native experience with the local variety and experience with enregistered varieties through learned social stereotypes (see also Sumner et al., 2014). This second aspect of familiarity is especially important from a theoretical perspective because, unlike familiarity, it does not derive directly from experience. That is, whereas familiarity due to native exposure is directly modeled in an exemplar framework as a large number of relevant previously encountered exemplars, socially marked variants may not require many specific exemplars to be robustly represented. Further, although recent research has shown that children acquire social stereotypes about regional dialects of their native language during elementary school (Kinzler and DeJesus, 2013), how those stereotypes are acquired and the extent to which they are learned based on exposure to talkers from different regions is unknown. The results of experiment 2 suggest that these social stereotypes impact not only our attitudes about talkers from different regions, but also cross-dialect lexical processing. Thus, this work points to a critical need to understand how social information about linguistic categories is acquired and represented so that differences in cross-dialect processing as a function of social enregisterment can be explained.

4.3. Conclusions
The results of these experiments provide further evidence for a model of speech processing in which linguistic and social information is inextricably linked. In particular, interference effects on lexical processing due to normalization for acoustic variability between talkers are compounded by normalization for phonologically ambiguous variants across dialects, whereas within-dialect talker variability does not differ across varieties. The cross-dialect interference effects can be observed for both local/non-local and non-local/non-local dialect pairs, although the effects themselves differ as a function of dialect familiarity. In particular, the local dialect is robust to interference effects on accuracy, whereas non-local dialects are not.

The current results also suggest that social enregisterment of regional dialects, or their salient variants, plays a critical role in lexical processing. Lexical classification is faster and more accurate for local variants than for non-local variants, and faster for non-local, stereotyped variants than for non-local, non-stereotyped variants. These patterns are observable even when dialect classification accuracy based on these variants is not significantly different from chance. These findings suggest differential encoding of non-local variants, such that stereotyped variants are more robustly represented than non-stereotyped variants, even in the absence of explicit dialect indexing of these variants.
Together, the results of this study have implications for our understanding of the nature and structure of the cognitive representations of socioindexical variation and point to the need for further research on hierarchical structure, gradience, and social stereotypes in indexical representations and their interface with the lexicon.

Acknowledgements

I would like to thank John Pate, Erin Loxley, Divya Maran, and Dosia Kumbe for research assistance and Bridget Smith, Rory Turnbull, and Kodi Weatherholtz for helpful discussion of this work. Portions of this research were conducted at the Language Sciences Research Lab at the Center of Science and Industry in Columbus, Ohio. Dosia Kumbe was supported by NSF REU site grant SBE-1451039.

References


