Applied Psycholinguistics **36** (2015), 729–746
doi:10.1017/S0142716413000489

# The effects of dialect variation on speech intelligibility in a multitalker background

EWA JACEWICZ and ROBERT ALLEN FOX
*Ohio State University*

Received: April 18, 2012 Accepted for publication: August 30, 2013

ADDRESS FOR CORRESPONDENCE
Ewa Jacewicz, Department of Speech and Hearing Science, Ohio State University, 1070 Carmack Road, Columbus, OH 43210. E-mail: jacewicz.1@osu.edu

ABSTRACT
Speech intelligibility in a multitalker background can be affected by the language of both the talker and the interfering speech. This study investigated whether this interaction is modulated by dialect variations of the same language. American English listeners were presented with target sentences in either their own General American English (GAE) or a different accent (Southern American English [SAE]) masked by either GAE or SAE two-talker babble at three sound to noise ratios (SNRs): +3, 0, and –3 dB. All speech materials were produced by male talkers. Across all conditions, SAE target was more intelligible than GAE. Intelligibility of either target decreased as the level of the interfering babble noise increased. Target accent interacted with masking accent: at +3 dB SNR, GAE (and not SAE) was the more effective masker. The target-masker interaction was different as listening conditions deteriorated: at 0 and –3 dB SNR, masking accent did not affect GAE target, but when the target was SAE, the SAE masker (and not GAE) was more effective. Thus, at increased noise levels, listeners benefited from the mismatch between the target and masking accents only when the target was in a nonnative accent. These results demonstrate that dialect variation can influence listeners' performance in a multitalker environment. The apparent asymmetry in intelligibility of accents may be in part related to dialect-specific prosodic and phonetic features.

Listening to speech in noisy environments, such as in the presence of multiple talkers, is challenging as these conditions limit its intelligibility. The growing literature on the subject repeatedly points to two factors which contribute most to the perceptual degradation of the target speech as determined in laboratory testing: (a) the increased sound level of the multitalker babble relative to the target and (b) the increased number of interfering talkers in the babble (e.g., Bronkhorst, 2000; Bronkhorst & Plomp, 1992; Brungart, Chang, Simpson, & Wang, 2009; Brungart, Simpson, Ericson, & Scott, 2001; Cooke, Garcia Lecumberri, & Barker, 2008; Miller, 1947; Simpson & Cooke, 2005; Van Engen & Bradlow, 2007). Performance also decreases when the target and interfering talkers share similar vocal characteristics (Brungart, 2001; Brungart et al., 2001). In general, multitalker environments produce two types of masking effects: energetic and informational. Energetic masking is assumed to occur at the periphery of the auditory system where the target and interfering voices "compete," that is, overlap in time and frequency. Informational masking is thought to be more central, when the segregation of the target and the interfering voices becomes difficult due to stimulus uncertainty although both types of signals are audible.

The intelligibility of the target in the presence of the masking speech is also affected by the linguistic system of each. Recent work has shown that the language of the masker can differentially affect the intelligibility of the target. In particular, a mismatch between the languages of the target and the masker produces intelligibility benefit compared to when both the target and the masker share the same language (Freyman, Balakrishnan, & Helfer, 2001; Garcia Lecumberri & Cooke, 2006; Rhebergen, Versfeld, & Dreschler, 2005; Van Engen & Bradlow, 2007). Exploring these effects, a number of studies have compared a range of listener–target–masker combinations that also included foreign or "accented" language either in the target sentence (Gordon-Salant, Yeni-Komshian, & Fitzgibbons, 2010; Pinet, Iverson, & Huckvale, 2011; Rogers, Dalby, & Nishi, 2004; Van Engen, 2010) or in the masking babble (Calandruccio, Dhar, & Bradlow, 2010; Freyman et al., 2001; Van Engen, 2010) and with native or nonnative listeners (e.g., Cooke et al., 2008; Garcia Lecumberri & Cooke, 2006; Mayo, Florentine, & Buus, 1997; Pinet et al., 2011). In general, native listeners performed better compared to nonnative listeners in a variety of listener–target–masker combinations, and the intelligibility of accented speech decreased in the presence of the masker (see Garcia Lecumberri, Cooke, & Cutler, 2010, for a comprehensive review). These findings prompted the question that motivated the present study: Can dialect variations of the same language differentially affect intelligibility of the target speech in the presence of the multitalker masker? The goal is to establish whether varying dialect in the target speech and in the masker produces effects comparable with those reported for native and nonnative (or accented) languages. In this pursuit, the study aims to determine whether speech recognition is also affected by fine-grained acoustic phonetic details that are present in dialect-specific pronunciation patterns in different regional varieties of the same language.

The existing experimental evidence shows that listeners are sensitive to pronunciation differences across dialect regions in the United States (Clopper & Pisoni, 2004, 2007) and that familiarity with a dialect improves its processing and comprehensibility. That is, comprehension difficulties and processing costs decrease as listener's familiarity with features of that dialect increases (Adank, Evans, Stuart-Smith, & Scott, 2009; Floccia, Goslin, Girard, & Konopczynski, 2006; Scott & Cutler, 1984; Sumner & Samuel, 2009). Two recent studies are of direct relevance to our present investigation because both examined intelligibility of regional dialects in masking noise, thus increasing cognitive effort of listeners due to difficult listening conditions, which tend to reduce redundant speech cues that are otherwise abundant in stimuli presented in quiet (e.g., Meador, Flege, & Mackay, 2000; Pisoni, Nusbaum, & Greene, 1985; van Wijngaarden, Steeneken, & Houtgast, 2002). In the first study, Clopper and Bradlow (2008) presented sentence stimuli produced by talkers representing four broad American English dialect regions (Mid-Atlantic, North, South, and General American) to three groups of listeners who lived, prior to age 18, in either Northern or General American English (GAE) dialect regions or were mobile, living in more than one region. In Experiment 1, the stimuli were presented in speech-shaped noise at –2 and –6 dB sound to noise ratios (SNRs) and in Experiment 2 at –2 and +2 dB. The results showed that, in the most difficult listening condition (–6 dB), the intelligibility scores were significantly different among all four dialects. As the overall performance improved in the easiest +2 dB condition, some of the significant differences disappeared, showing that only Mid-Atlantic talkers were less intelligible than the talkers from the other three dialects. In both experiments, however, the highest intelligibility scores were obtained from productions by the GAE talkers, indicating that GAE was the most intelligible dialect for all listeners in all testing conditions. Overall, this study shows that dialect differences among talkers have a differential effect on intelligibility of degraded speech and tend to be attenuated in more favorable listening conditions that provide listeners with greater redundancy of cues.

In the second study, Adank et al. (2009) examined processing cost (measured in terms of response times) associated with comprehension of sentences produced by Southern Standard British English talkers and Scottish English (Glasgow) talkers by native speakers of these two regional varieties. The Standard English listeners were unfamiliar with the Scottish variety, whereas the Scottish listeners were equally familiar with both Standard and Scottish English. The stimuli were presented in quiet and in speech-shaped noise at three SNRs: +3, 0, and –3 dB. The results of Experiment 1 showed that the Standard English listeners were significantly slower when responding to sentences in an unfamiliar dialect (Scottish) at moderate SNR levels (+3 and 0 dB), but the difference between the two accents became much smaller at –3 dB, that is, as the listening conditions deteriorated and the response times were the slowest. Thus, Standard English listeners showed less efficient speech processing for an unfamiliar regional accent in moderate listening conditions, most likely due to their lack of experience with the Scottish English. The Scottish English listeners were equally fast when responding to both varieties, apparently due to their familiarity with both accents. It needs to be underscored that the difference between the familiar and unfamiliar accents was not significant in quiet, suggesting that both groups of listeners benefited from the redundancy of the speech signal in favorable listening conditions.

In the present study, we investigate whether listeners' familiarity or unfamiliarity with the dialect of American English (General or Southern) affects intelligibility scores when the target speech is presented in two-talker babble. While varying the dialect in the masking babble, conditions were created for additional interactions involving the dialect of the masker and dialect of the target sentence. Based on the results of the two studies reviewed above, we can tentatively predict that listeners familiar with GAE will perform better when target sentences are produced in that variety and will perform worse when responding to productions in a local Southern American English (SAE) dialect, which is relatively unfamiliar to them. This scenario may not hold true when the familiarity with the dialect in the target interacts with familiarity with the dialect in the masking babble.

To our knowledge, no study has as yet explored the effects of dialect of the babble on intelligibility of the target. Therefore, to spell out predictions about possible outcomes of the interaction between the two, we turned to available literature that examined such effects for native and nonnative language configurations. Van Engen and Bradlow (2007) found that for American English listeners responding to an American English target, English two-talker babble was more detrimental than Mandarin two-talker babble, but this effect disappeared in six-talker babble. Calandruccio et al. (2010) provided American English listeners responding to American English targets with five distinct two-talker babble maskers, which used American English, three levels of Mandarin-accented English (mildly, moderately, and heavily accented) and Mandarin. There was a significant difference in performance among all five maskers in a more difficult listening condition (–5 dB). Listener performance was poorest in English babble, followed by that in Mandarin babble. However, their performance was better in the presence of the three Mandarin-accented English maskers. Altogether, American English listeners performed best in the presence of nonnative, heavily Mandarin-accented English masker and worst in the presence of their native American English babble.

In another study, Van Engen (2010) used native American English listeners and native Mandarin listeners who spoke English as a second language (L2). The listeners were presented with American English target sentences masked by a two-talker babble in either English or Mandarin. In this design, L2 listeners responded in the presence of either a native language (Mandarin) or L2 (English) masker and native English listeners responded in the presence of their native language (English) and foreign language (Mandarin) masker. The results showed that English listeners performed poorly in the presence of their native language (English) masker but were significantly better in Mandarin babble, thus replicating an earlier finding by Van Engen and Bradlow (2007). However, performance of L2 listeners was also worse in the presence of the English masker (a masker in their L2) and better in the Mandarin masker, which was in their native language. Thus, the English masker was more detrimental for both groups of listeners, regardless of their native language background (while processing English sentences).

In the present study, we varied two regional dialects, GAE and SAE, in target sentences and in the masker (two-talker babble) and presented the speech material to GAE listeners at three SNRs (+3, 0, and –3 dB). Taking into consideration the main findings of the research reviewed above, we predict the following pattern of responses. First, we expect that intelligibility of the target will decrease as the SNR decreases. Second, listeners are expected to perform better when responding to GAE target (Adank et al., 2009; Clopper & Bradlow, 2008) masked by SAE babble rather than by GAE babble (based on Van Engen, 2010). This prediction, however, rests on the premise that the native-dialect status and target–masker similarity effects correspond to the effects reported for the native and L2 languages (Garcia Lecumberri & Cooke, 2006). Finally, the target–masker dialect interactions are predicted to have a greater effect on listener performance at the more difficult SNR (–3 dB) than in moderate listening conditions (+3 and 0 dB).

## METHODS

### *Listeners*

Forty-one listeners initially participated in the study. Two participants were excluded from data analysis because, due to a design error, their responses were obtained using a different version of the MATLAB program that controlled the experiment. The remaining 39 listeners (30 females, 9 males) were included in all analyses in the present study. The participants were born and raised in Central Ohio (Columbus area) and ranged in age between 19 and 25 years ($M = 21.15$, $SD = 1.7$). They were monolingual students enrolled at Ohio State University who spoke the local regional variety as verified by the research staff in an informal interview. This variety of English, known as the Midland dialect (Labov, Ash, & Boberg, 2006), is considered a part of the GAE classification by Clopper, Levi, and Pisoni (2006). In the present study, we adopt the term GAE for this dialect variety spoken in Central Ohio. All participants had normal hearing, as determined by a pure-tone screening, and were paid for their efforts. The recruitment of participants and study protocol was approved by the Ohio State University Institutional Review Board.

### *Stimuli*

*Target sentences.* Target sentences were taken from the Revised Bamford–Kowal–Bench Standard Sentence Tests (Bench, Kowal, & Bamford, 1979). Six sentence lists were randomly selected for use in the experiment (Lists 3, 6, 7, 9, 13, and 18). In addition, list 4 was recorded in order to create a practice run. Each list contained 16 short, meaningful sentences (e.g., *The book tells a story*) for a total of 96 used in the study. There were three or four keywords per sentence (a total of 50 per list) for a total of 300 keywords analyzed in the study. A complete set of stimuli used in the study is included in Appendix A. The sentences were produced by eight middle-age male talkers, four from Columbus, Ohio (who spoke the Central Ohio variety of GAE), and four from Sylva, North Carolina (who spoke the local Western North Carolina variety of SAE). This variant of SAE is typical of the dialect region identified as Inland South in Labov et al. (2006). All talkers were long-time residents of their respective regions, were born and raised in those areas, spoke the respective regional variety as verified by the research staff, had a comparable speech tempo, and did not demonstrate any marked peculiarities in their voice characteristics. The final set of sentences used in the experiment consisted of 12 sentences for each speaker ($12 \times 8 = 96$). The selected sentences were spoken fluently, that is, without hesitations or pauses, with a similar falling intonation, at comparable rates ($M = 3.97$ syllables/s for the GAE talkers and $M = 4.02$ syllables/s for the SAE talkers) and with comparable fundamental frequency (f0; $M = 132.73$ Hz for the GAE talkers and $M = 137.14$ Hz for the SAE talkers) and f0 range (118–148 Hz). The recordings were completed under laboratory conditions at Ohio State University and at Western Carolina University.

*Multitalker masking babble.* Previous recordings of spontaneous conversations from a large corpus of cross-dialectal productions (Jacewicz, Fox, & Wei, 2010) were utilized in babble creation. The talkers selected for the babble included four men aged 50–60 years (two from Central Ohio and two from Western North Carolina). They spoke the same dialect varieties (GAE and SAE) as the talkers who produced the target sentences. All talkers (i.e., those used in the babble and in the target) were matched for speech tempo and f0. Due to the complexity of the study, it was decided not to include talker gender as an additional variable, and all talkers in the present study were male.

Short individual phrases of the recorded spontaneous conversations were extracted from each talker's discourse. Each phrase contained approximately 5–10 syllables (e.g., *quite a few more restaurants; not as well educated and*) and was produced fluently without any pauses. The number of selected phrases differed among talkers. One long file was then created for each talker in which his phrases were reordered using a custom MATLAB program to make his speech semantically anomalous. While combining the reordered individual phrases into the long file, the effort was put to eliminate syntactic and semantic effects so that one could not follow any "story" when extracting individual words. There were no silent intervals between the phrases and the phrases were equated for root mean square amplitude. Two different talker files from each dialect were chosen for creation of the two-talker masker. Under the control of a MATLAB program, the two files were mixed into a single two-talker babble file. Each two-talker babble file was then equated for root mean square amplitude and divided into 120 sections of equal duration, 4-s long each. In this way, 120 samples were obtained for each masker dialect (GAE and SAE) for a total of 240.

Next, each of the 96 target sentences, 48 for each dialect, was mixed with a randomly chosen babble sample so that half of the GAE or SAE sentences were mixed with 24 GAE and 24 SAE babble samples. This was done using a different custom MATLAB program that controlled the experiment. On each trial, the target sentence was positioned within the central 75% portion of the 4-s babble sample. The actual position of the sentence was randomly varied within this central location so that listeners could not predict exactly when the sentence was to begin. The onset and offset of the mixed token was ramped from/to zero using a Hanning window over the first and last 5 ms, respectively. Prior to mixing the target sentence and the babble sample for each trial, the level of the babble was adjusted relative to the fixed level of the target sentence to create three SNRs: +3 dB (the level of the babble was 3 dB less than the level of the target), 0 dB (both levels were equal), and –3 dB (the level of the babble was 3 dB more than the level of the target). Following Adank et al. (2009), these three SNRs represented moderate (+ 3 and 0 dB) and difficult (–3 dB) listening conditions.

### *Procedure*

Each listener was tested individually at Ohio State University, seated in a sound-attenuating booth and facing a computer monitor. Sound was delivered diotically over Sennheizer HD 600 headphones at a comfortable listening level. The experiment was administered under the control of a custom MATLAB program. Each participant listened to all 96 sentences in one session, which took approximately 45 min. The order of experimental blocks was the same for each listener, proceeding from the easiest (+3 dB) to the most difficult (–3 dB) SNR. At each SNR level, there were 16 sentences produced by GAE talkers (4×4) and 16 by SAE talkers (4 × 4). Half of the sentences for each dialect were masked by either GAE or SAE babble (8 × 2) so that each listener responded to four sentences from each talker, two of which were masked by GAE babble and two by SAE babble. All 32 sentences within each SNR block were presented in a random order. To reiterate, the presentation order of target sentence dialect and dialect in the masking babble was randomized for each SNR level. The listeners were told that they would listen to sentences produced by several different talkers and that listening conditions would deteriorate during the experiment so that the target sentences would not always be easy to detect. None of the listeners was told about dialect variation in the stimulus speech. For familiarization purposes, eight practice trials were first presented at +3 dB SNR (one sentence for each speaker) that were different from those used in the experiment. After the practice run, each listener responded to three experimental blocks of 32 sentences each. The experiment was self-paced. Upon hearing each stimulus, listeners typed what they heard as the target sentence into a box that appeared on the computer screen. No repetitions were allowed. Listeners could take breaks between the testing blocks upon request.

### *Data analysis*

The digitally recorded responses were scored on the basis of keywords by three different experimenters, each of whom scored a subset of sentences. A reliability check was then performed on all responses by each of the three scorers, who exchanged their scored spreadsheets. Scoring discrepancies that occurred because of error or oversight (in 8 sentences out of a total of 3,744) were resolved among all three experimenters following discussions and individual reevaluation of responses in question. The affected responses were then rescored prior to data analysis. The scoring system was adapted from Van Engen and Bradlow (2007) so that words with added or deleted morphemes were counted as incorrect and those containing spelling errors were counted as correct. Raw scores for each participant were first converted to percent correct and then to rationalized arcsine units (RAUs; Studebaker, 1985) to ensure valid assessment of differences across the entire range of the scale (from –14 to 114) after normalizing for ceiling and floor effects.

Repeated-measures analysis of variance (ANOVA) was used to analyze the arcsine transformed performance scores; SNR level, target dialect, and masker dialect were the within-subject factors. We also report a measure of the effect size, partial eta squared ($\eta^2$), in addition to the significance values for all significant main effects and interactions. Partial $\eta^2$ should be considered a measure of the proportion of variance explained by a dependent variable when controlling for other factors. Pairwise post hoc *t* tests (with Bonferroni adjustment) were used to explore the nature of significant main effects and interactions.

## RESULTS

Listeners' responses to all target and masker combinations are displayed in Figure 1. As expected, higher SNRs yielded higher intelligibility scores and the

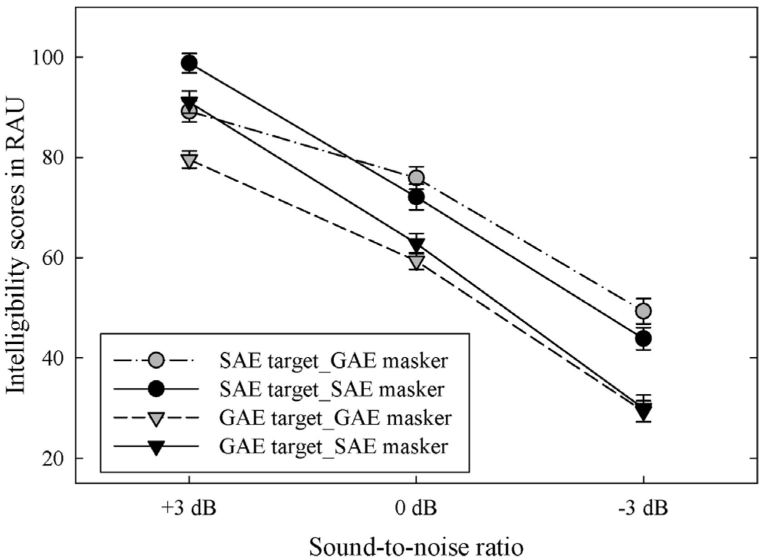


Figure 1. Mean (*SE*) arcsine transformed intelligibility scores for General American English (GAE) and Southern American English (SAE) targets as a function of dialect in the masker (GAE and SAE) and sound to noise ratio. RAU, rationalized arcsine unit.

performance worsened with each lower SNR ($M$ = 89.7, 67.5 and 38.1 RAU at +3, 0, and –3 dB, respectively). The main effect of SNR level was significant, $F$ (2, 76) = 629.4, $p < .001$, $\eta^2 = 0.943$. Post hoc comparisons indicated that all conditions were significantly different from one another ($p < .001$). The main effect of target dialect was significant, $F$ (1, 38) = 90.8, $p < .001$, $\eta^2 = 0.705$, indicating higher intelligibility for SAE talkers ($M$ = 71.5 RAU) than GAE talkers ($M$ = 58.7 RAU). Also significant was the main effect of masker dialect, $F$ (1, 38) = 7.58, $p = .009$, $\eta^2 = 0.166$, with lower responses in GAE babble ($M$ = 63.8 RAU) compared to SAE babble ($M$ = 66.4 RAU).

There was a significant SNR × Target Dialect interaction, $F$ (2, 76) = 4.13, $p = .020$, $\eta^2 = 0.098$. As illustrated in Figure 2, the differences between the SAE and GAE talkers increased as listening conditions deteriorated: they were smallest at +3 dB (8.7 RAU) and greatest at –3 dB (16.9 RAU). Post hoc analyses revealed that the differences between the SAE and GAE talkers were significant at each SNR ($p < .001$), which is attributable to the significant main effect of talker dialect. However, pairwise comparisons of GAE and SAE differences between SNR levels indicated that the size of the differences was significantly larger only between +3 and –3 dB SNR ($p = .005$). The locus of the interaction was that the SAE advantage was greatest at –3 dB compared to +3 dB SNR, whereas the size of the SAE and GAE talker differences was not significant for comparisons between 0 versus +3 dB and for 0 versus –3 dB SNR.

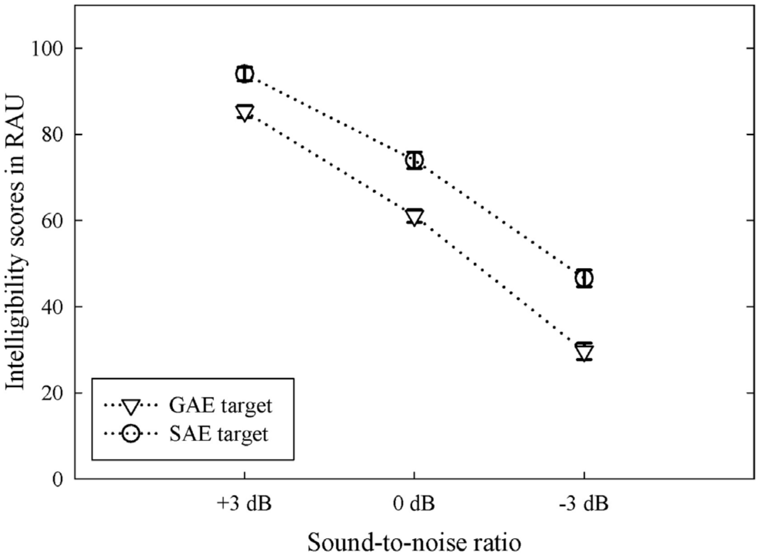


Figure 2. Mean (*SE*) arcsine transformed intelligibility scores for target dialect (General American English [GAE] and Southern American English [SAE]) at each sound to noise ratio. RAU, rationalized arcsine unit.

The second significant interaction was between SNR and masker dialect, $F$ (2, 76) = 11.52, $p < .001$, $\eta^2 = 0.233$, which is illustrated in Figure 3. Post hoc comparisons indicated that the responses in SAE babble were significantly higher than in GAE babble only at +3 dB SNR. There was no significant effect of the dialect in the babble in the more challenging listening conditions at 0 and –3 dB SNR.

The interaction between target dialect and masker dialect was also significant, $F$ (1, 38) = 5.78, $p = .021$, $\eta^2 = 0.132$. Post hoc analyses of intelligibility differences between GAE and SAE talkers in either GAE or SAE babble indicated significantly greater differences for GAE babble compared to SAE babble ($p$ = .022). Intelligibility of GAE talkers decreased in GAE babble compared to SAE babble, whereas intelligibility of SAE talkers was comparatively high and was unaffected by the dialect in the babble. The three-way interaction between target dialect, masker dialect, and SNR was not significant.

As reported above, the significant target dialect by masker dialect interaction indicated that dialect in the masker had no effect on intelligibility of the SAE target, while the SAE was a less effective masker than the GAE for the GAE target. However, an inspection of Figure 1 suggests a differential performance of both maskers as a function of SNR. At +3 dB SNR, the GAE was a more effective masker for either target than was the SAE, whereas at the two lower SNRs, the GAE seemed to be a less effective masker for the SAE target. To better understand these results, a separate ANOVA was conducted in which the +3 dB SNR was excluded from the SNR factor and only 0 and –3 dB SNR levels were included in

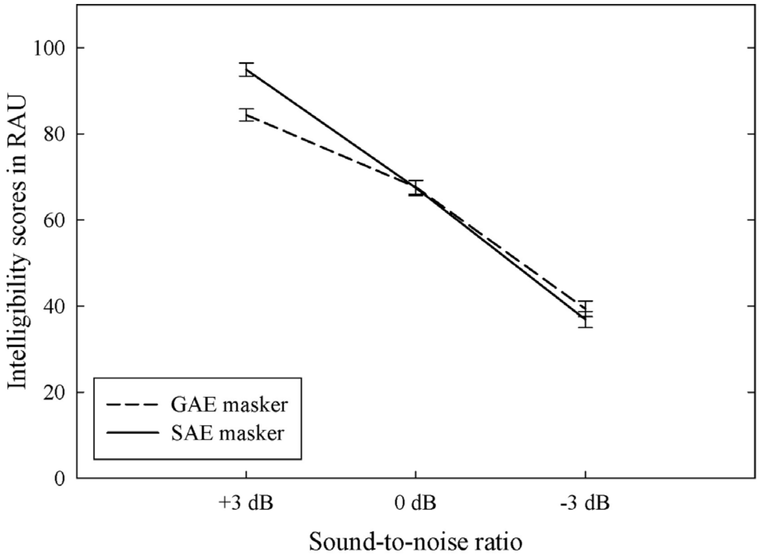


Figure 3. Mean (*SE*) arcsine transformed intelligibility scores for masker dialect (General American English [GAE] and Southern American English [SAE]) at each sound to noise ratio. RAU, rationalized arcsine unit.

the analysis. As might be expected, the main effect of SNR was significant, $F$ (1, 38) = 349.0, $p < .001$, $\eta^2 = 0.902$, as well as the main effect of talker dialect, $F$ (1, 38) = 88.4, $p < .001$, $\eta^2 = 0.699$, which again indicated higher intelligibility for SAE talkers than for GAE talkers. The effect of masker dialect was not significant. There was a significant interaction between target dialect and masker dialect, $F$ (1, 38) = 5.70, $p = .022$, $\eta^2 = 0.699$. No interaction with SNR was significant.

The significant target dialect by masker dialect interaction in this ANOVA was followed by Bonferroni-adjusted post hoc tests, which revealed that the difference between the GAE and SAE maskers was significant only for the SAE target ($p = .026$). It was not significant for the GAE target. This interaction indicates that at more difficult SNRs, the GAE was a less effective masker for the SAE target, whereas dialect in the masker did not differentially affect listeners' performance for the GAE target. In other words, at moderate to difficult SNRs, listeners benefited from hearing their own dialect in the babble when responding to sentences in a different dialect, but a mismatched target-masker did not produce better performance than did the matched target-masker for sentences in listeners' own dialect.

In summary, the results showed that SAE talkers were more intelligible than GAE talkers, even more so when listening conditions deteriorated. Dialect in the masker interacted with the target dialect, but this interaction was SNR-dependent. GAE babble was a more effective masker than SAE babble at the easiest listening condition at +3 dB SNR. However, at more challenging conditions at 0 dB SNR and –3 dB SNR, the GAE was a less effective masker than the SAE for the SAE target, whereas dialect in the babble had no significant effect on intelligibility of the GEA target.

## DISCUSSION

As an elaboration of the effects of dialect variation on speech intelligibility, the present study tested target-masker combinations for two regional varieties representing GAE and SAE. Our current interest was in the intelligibility of the target in the presence of an interfering speech that either shared or did not share the same dialect features. Several findings emerged.

The expected effects of SNR found here are in agreement with a number of previous studies showing that perceptibility of the target speech decreases as the level of the masking noise increases (e.g., Bronkhorst, 2000; Bronkhorst & Plomp, 1992; Brungart et al., 2001; Rhebergen & Versfeld, 2005; Simpson & Cooke, 2005; Van Engen & Bradlow, 2007). The widely observed effects resulting from manipulations of SNR levels are also applicable to listening environments that involve dialect variation of the same language.

Based on results of studies that used native and nonnative (or foreign-accented) speech combinations (e.g., Calandruccio et al., 2010; Garcia Lecumberri & Cooke, 2006; Van Engen, 2010; Van Engen & Bradlow, 2007), we expected that GAE listeners would perform worse when responding to the GAE target masked by GAE babble rather than by SAE babble. This prediction was confirmed only at the easiest SNR level (+3 dB). However, even in this condition, GAE babble also had a detrimental effect on intelligibility of the SAE target. This indicates that, at more favorable listening conditions, there is more interference when the babble is in the listener's own dialect (GAE). Intelligibility further declines when the target speech also shares the same acoustic characteristics. We can interpret these results as showing that listeners' familiarity with the features of their own dialect has a more detrimental effect on listening in a multitalker environment than when the target and the masker do not share the same dialect.

The present results also differ for the masking effects in more challenging listening conditions. We predicted that the target-masker dialect would have a greater effect on listener performance at a more difficult SNR (–3 dB) than in moderate listening conditions (+3 and 0 dB). Instead, dialect in the babble had no significant effect on intelligibility of the target at either 0 or –3 dB SNR when the target was in listeners' own dialect. However, dialect in the babble did significantly affect intelligibility of the target when the target was in a different dialect. Listeners benefited from hearing their own dialect in the babble when processing target sentences in a different dialect. These results imply that, as conditions deteriorate, listeners devote more cognitive resources to processing the target speech and pay less attention to the dialect in the babble when hearing their own dialect in the target sentences. However, the mismatch between the target and the masker can produce an intelligibility benefit at moderate to difficult conditions for combinations with listeners' own dialect in the masker. These effects are not in the direction as previously found in the literature. Rather, accent interactions in the target *and* in the masker imply a more complex relationship affecting the processing of phonetic variation in a native language in noisy multitalker environments.

To some extent, differences in design and experimental conditions need to be taken into consideration while comparing the results across the literature. In the present design, each listener was randomly presented with the dialect in the target and in the masker from trial to trial. The listeners in Clopper and Bradlow (2008) responded to dialect and speaker variation in the target masked by noise, and thus no interactions between dialects in the target and in the masker were tested in that study. Adank et al. (2009) used a similar design for testing dialect effects and Rogers et al. (2004) for testing intelligibility of Mandarin-accented English. Therefore, the finding that in the most difficult listening condition (–6 dB) the intelligibility scores are significantly affected by speaker dialect (Clopper & Bradlow, 2008) may not be replicated if multitalker babble, rather than noise, is used as a masker. This possibility needs to be verified in future research. We also need to bear in mind that cross-language patterns may not be entirely pertinent to cross-dialect patterns (e.g., Cutler, Smits, & Cooper, 2005; Tuinman, Mitterer, & Cutler, 2011) and may not provide an exclusive basis for predictions about the outcome of dialect interactions in speech-on-speech masking. A direct comparison of cross-language and cross-dialect effects also warrants future exploration.

A robust finding of the present study was that intelligibility was significantly higher when the target speech was not in the listeners' own dialect. This effect persisted at all SNR levels, and the advantage of the different dialect increased as listening conditions deteriorated. It is most puzzling, however, that it was the SAE and not the GAE that listeners found easier to understand in a noisy environment. This result is clearly different from Clopper and Bradlow (2008), who found that GAE was the most intelligible variety for listeners coming from four different dialect backgrounds. The differences in design could have contributed to this discrepancy, as their stimuli were degraded by speech-shaped noise and ours by multitalker babble. In addition, their GAE talkers were selected from broader dialect regions encompassing several states as were the listeners, who were grouped on the basis of their prior residential history in several dialect regions. Our talkers came from two relatively homogenous speech communities and represented a particular dialect variant spoken in each narrowly defined region. Therefore, although the Midland (Central Ohio) dialect variety in our study meets the definition of GAE, it may not match the exact phonetic characteristics as in Clopper and Bradlow (2008). By the same token, our variety of SAE may not correspond to their SAE dialect, which included talkers and listeners from a much broader geographic region in the United States.

The question arises as to which acoustic characteristics of SAE could have contributed to the higher intelligibility of the SAE target. Although we can only speculate at present, our current understanding leads us to conjecture that, to some extent, listeners could have utilized dialect-specific prosodic and temporal cues. During an informal exchange after the experimental sessions between several participants and the study personnel, some of them remarked that "these Southerners were much easier to understand." Because the research staff was specifically instructed not to share with any of the participants any information related to the nature of the voices used in the experiment, it became apparent that at least some of them detected and recognized the Southern accent.

Recently, on the basis of Midland and Southern dialects using a different corpus, Clopper and Smiljanic (2011) reported significant dialect effects on prosodic patterns, including pause distribution and the f0 patterns associated with prominent syllables and phrase boundaries. While examination of prosodic features of the two dialects used in our study is currently underway, preliminary results show that there are significant dialectal differences in the use of the f0 change in stressed syllables. In particular, SAE variants exhibit a comparatively greater pitch movement than GAE variants (Hart, 2013; Fox, Jacewicz, & Hart, 2013). Looking at time-normalized changes in f0 from vowel onset to vowel offset, it was found that f0 contours in the SAE dialect were significantly different from those in the GAE dialect. For example, SAE vowels had an earlier rise to f0 maximum, a greater f0 decrement from maximum f0 frequency to final f0 frequency, and a lower f0 offset. These salient acoustic features produced more "exaggerated" or "melodic" pitch contours than those found in GAE vowels.

Notably, the greater pitch movement interacts with a greater dynamic formant movement in SAE vowels compared with GAE vowels. That is, the monophthongal SAE vowels are more "diphthongized" and their distinct qualities have been termed Southern breaking or Southern drawl (Sledd, 1966). An in-depth analysis of formant dynamics in SAE and GAE vowels in a large corpus can be found in Fox and Jacewicz (2009) and Jacewicz, Fox, and Salmons (2011). Vowels produced by SAE talkers also have comparatively longer durations (Fox & Jacewicz, 2009; Jacewicz, Fox, & Salmons, 2007). Although the talkers selected for this study were comparable in terms of their overall speech tempo, the dialect-related differences in vowel durations may have contributed to the perceived dialectal differences. Considering all the prosodic, spectral, and temporal patterns, we have a basis to hypothesize that dynamic cues may be a robust feature of this variety of the Southern speech and may provide listeners with a greater amount of redundancy in the signal compared with the GAE. The exact contributions of these dynamic cues to the perceptibility of the target speech remains to be explored in future research.

There is also another possibility awaiting experimental exploration: the spectral properties of SAE and GAE speech (whether in the target or in the babble) could have influenced intelligibility of the target and masking effectiveness. For example, Callandrucio et al. (2010) showed that spectral differences between masker signals for a speech-in-speech recognition task using two-talker maskers contributed to differences in performance. In their study, when the auditory task was not too difficult, maskers with greater spectral energy in selected frequency regions resulted in greater masking. It cannot be ruled out that long-term spectrum could have been a factor contributing to the differences between SAE and GAE target and masker performance in the present study. However, this possibility can be best explored in light of responses from both GAE *and* SAE listeners so that factors related to experience and familiarity with each dialect have been sufficiently controlled for.

In conclusion, the present study demonstrates that dialect variation does influence listeners' performance in a multitalker environment. Under current experimental conditions, dialect in the masker interacted with the target dialect, but this interaction was very SNR-dependent. The regional variety of SAE used here showed clearly an intelligibility benefit for listeners unfamiliar with this dialect compared to their own variety of GAE. This asymmetry in intelligibility resembles the asymmetry between two related languages, such as Danish and Swedish, reported by Gooskens, van Heuven, van Bezooijen, and Pacilly (2010), showing that Danish is less intelligible as a language than Swedish. While the relationship between the orthography and pronunciation is the suggested source of the asymmetry in these two languages, we see the likely causes of the asymmetry in our study in the dialect-specific prosodic and phonetic features. It needs to be determined which dialect-specific features contribute most to greater intelligibility of individual dialects and how they interact in the speech-in-speech perception.

## APPENDIX A

The 96 target sentences used in the experiment at each SNR level were selected from the Revised Bamford–Kowal–Bench Standard Sentence Tests (Bench, Kowal, & Bamford, 1979), Lists 3, 6, 7, 9, 13, and 18. The keywords are in boldface.

### *Set 1 (+3 db SNR)*

1. The **old man** is **worried**.
2. **Rain falls** from the **clouds**.
3. The **cook** is **making** a **cake**.
4. **They** are **kneeling down**.
5. The **jelly jar** was **full**.
6. **She drinks** from her **cup**.
7. The **strawberry jam** was **sweet**.
8. The **lady washed** the **shirt**.
9. **She** had her **spending money**.
10. A **boy ran down** the **path**.
11. **He needed** his **vacation**.
12. The **dog** is **eating** some **meat**.
13. **He listened to** his **father**.
14. The **farmer keeps** a **bull**.
15. The **apple pie** was **good**.
16. The **boy** has **black hair**.
17. The **bus left early**.
18. The **teapot** is **very hot**.
19. The **woman cleaned** her **house**.
20. The **new road** is on the **map**.
21. The **house** had a **nice garden**.
22. The **young people** are **dancing**.
23. **They** are **crossing** the **street**.
24. The **boy** is **running away**.
25. **They** had **two empty bottles**.
26. The **snow is** on the **roof**.
27. A **girl kicked** the **table**.

28. The **mud stuck** on his **shoe**.
29. The **child grabbed** the **toy**.
30. The **broom stood** in the **corner**.
31. A **boy broke** the **fence**.
32. **They ate** the **lemon pie**.

*Set 2 (0 db SNR)*

1. **They** are **playing** in the **park**.
2. The **bakery is open**.
3. The **fruit came** in a **box**.
4. Some **animals sleep** on **straw**.
5. **They painted** the **wall**.
6. **Potatoes grow** in the **ground**.
7. The **truck climbed** the **hill**.
8. **He found** his **brother**.
9. The **husband brought** some **flowers**.
10. **They** are **climbing** the **tree**.
11. **She talked** to her **doll**.
12. The **towel dropped** on the **floor**.
13. The **lady wore** a **coat**.
14. **Father forgot** the **bread**.
15. The **sun melted** the **snow**.
16. **They laughed** at his **story**.
17. The **lady packed** her **bag**.
18. The **children** are **walking home**.
19. **He** is **holding** his **nose**.
20. The **cup** is **hanging** on a **hook**.
21. **They say** some **silly things**.
22. The **mother heard** the **baby**.
23. A **sharp knife** is **dangerous**.
24. **They like orange marmalade**.
25. The **room** is **getting cold**.
26. The **children dropped** the **bag**.
27. **He** is **skating with** his **friend**.
28. The **children** are **all eating**.
29. The **paint dripped** on the **ground**.
30. **They waited** for **one hour**.
31. The **yellow pears** tasted **good**.
32. The **train** is **moving fast**.

*Set 3 (–3 db SNR)*

1. **Father looked** at the **book**.
2. A **letter fell** on the **floor**.
3. The **ball** is **bouncing** very **high**.
4. The **machine** was **very noisy**.
5. The **five men** are **working**.
6. A **man told** the **police.**
7. **She stood** near her **window**.
8. The **dog came back**.

9. The **police helped** the **driver**.
10. The **fruit is** on the **ground**.
11. The **lady went to** the **store**.
12. **She argued** with her **sister**.
13. The **young boy left home**.
14. The **book tells** a **story**.
15. The **dinner plate** is **hot**.
16. **Mother got** a **saucepan**.
17. **Milk comes** in a **carton**.
18. The **smart girls** are **reading**.
19. The **orange** was **very sweet**.
20. **He** is **bringing** his **raincoat**.
21. The **dog played** with a **stick**.
22. The **plant** is **hanging above** the **door**.
23. The **baby broke** his **cup**.
24. The **big dog** was **dangerous**.
25. **They washed** in **cold water**.
26. **Mother shut** the **window**.
27. **He** is **cleaning** his **car**.
28. The **wife helped** her **husband**.
29. The **girl** has a **picture book**.
30. **Mother stirred** her **tea**.
31. The **father** is **coming home**.
32. The **shoes** were **very dirty.**

## REFERENCES

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

Bench, J., Kowal, A., & Bamford, J. (1979). The BKB (Bamford-Kowal-Bench) sentence lists for partially-hearing children. *British Journal of Audiology, 13*, 108–112.

Bronkhorst, A. W. (2000). The cocktail party phenomenon: A review of research on speech intelligibility in multiple-talker conditions. *Acustica united with Acta Acustica, 86*, 117–128.

Bronkhorst, A. W., & Plomp, R. (1992). Effect of multiple speechlike maskers on binaural speech recognition in normal and impaired hearing. *Journal of the Acoustical Society of America, 92*, 3132–3138.

Brungart, D. S. (2001). Informational and energetic masking effects in the perception of two simultaneous talkers. *Journal of the Acoustical Society of America, 109*, 1101–1109.

Brungart, D. S., Chang, P. S., Simpson, B. D., & Wang, D. L. (2009). Multitalker speech perception with ideal time-frequency segregation: Effects of voice characteristics and number of talkers. *Journal of the Acoustical Society of America, 125*, 4006–4022.

Brungart, D. S., Simpson, B. D., Ericson, M. A., & Scott, K. R. (2001). Informational and energetic masking effects in the perception of multiple simultaneous talkers. *Journal of the Acoustical Society of America, 110*, 2527–2538.

Calandruccio, L., Dhar, S., & Bradlow, A. R. (2010). Speech-on-speech masking with variable access to the linguistic content of the masker speech. *Journal of the Acoustical Society of America, 128*, 860–869.

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

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

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

Clopper, C. G., & Smiljanic, R. B. (2011). Effects of gender and regional dialect on prosodic patterns in American English. *Journal of Phonetics, 39*, 237–245.

Clopper, C. G., Levi, S., & Pisoni, D. (2006). Perceptual similarity of regional dialects of American English. *Journal of the Acoustical Society of America, 119*, 566–574.

Cooke, M., Garcia Lecumberri, M. L., & Barker, J. (2008). The foreign language cocktail party problem: Energetic and informational masking effects in non-native speech perception. *Journal of the Acoustical Society of America, 123*, 414–427.

Cutler, A., Smits, R., & Cooper, N. (2005). Vowel perception: Effects of non-native language vs. non-native dialect. *Speech Communication, 47*, 32–42.

Floccia, C., Goslin, J., Girard, F., & Konopczynski, G. (2006). Does a regional accent perturb speech processing? *Journal of Experimental Psychology: Human Performance and Perception, 32*, 1276–1293.

Fox, R. A., & Jacewicz, E. (2009). Cross-dialectal variation in formant dynamics of American English vowels. *Journal of the Acoustical Society of America, 125*, 2603–2618.

Fox, R. A., Jacewicz, E., & Hart, J. (2013). *Pitch pattern variations in three regional varieties of American English.* Paper presented at the 14th Annual Conference of the International Speech Communication Association (Interspeech 2013), Lyon, France.

Freyman, R. L., Balakrishnan, U., & Helfer, K. S. (2001). Spatial release from informational masking in speech recognition. *Journal of the Acoustical Society of America, 109*, 2112–2122.

Garcia Lecumberri, M. L., & Cooke, M. (2006). Effect of masker type on native and non-native consonant perception in noise. *Journal of the Acoustical Society of America, 119*, 2445–2454.

Garcia Lecumberri, M. L., Cooke, M., and Cutler, A. (2010). Non-native speech perception in adverse conditions: A review. *Speech Communication, 52*, 864–886.

Gooskens, C., van Heuven, V. J., van Bezooijen, R., & Pacilly, J. J. A. (2010). Is spoken Danish less intelligible than Swedish? *Speech Communication, 52*, 1022–1037.

Gordon-Salant, S., Yeni-Komshian, G. H., & Fitzgibbons, P. J. (2010). Recognition of accented English in quiet and noise by younger and older listeners. *Journal of the Acoustical Society of America, 128*, 3152–3160.

Hart, J. (2013). Prosodic differences among dialects of American English. (Senior honors thesis, Ohio State University, 2013).

Jacewicz, E., Fox, R. A., & Salmons, J. (2007). Vowel duration in three American English dialects. *American Speech, 82*, 367–385.

Jacewicz, E., Fox, R. A., & Salmons, J. (2011). Cross-generational vowel change in American English. *Language Variation and Change, 23*, 45–86.

Jacewicz, E., Fox, R. A., & Wei, L. (2010). Between-speaker and within-speaker variation in speech tempo of American English. *Journal of the Acoustical Society of America, 128*, 839–850.

Labov, W., Ash, S., & Boberg, C. (2006). The Midland. In *Atlas of North American English* (pp. 263–278). Berlin: Mouton de Gruyter.

Mayo, L. H., Florentine, M., & Buus, S. (1997). Age of second-language acquisition and perception of speech in noise. *Journal of Speech Language and Hearing Research, 40*, 686–693.

Meador, D., Flege, J. E., & Mackay, I. R. A. (2000). Factors affecting the recognition of words in a second language. *Bilingualism: Language and Cognition, 3*, 55–67.

Miller, G. (1947). Sensitivity to changes in the intensity of white Gaussian noise and its relation to masking and loudness. *Journal of the Acoustical Society of America, 191*, 609–619.

Pinet, M., Iverson, P., & Huckvale, M. (2011). Second-language experience and speech-in-noise recognition: Effects of talker-listener accent similarity. *Journal of the Acoustical Society of America, 130*, 1656–1662.

Pisoni, D., Nusbaum, H., & Greene, B. (1985). Perception of synthetic speech generated by rule. *Proceedings of the IEEE, 73*, 1665–1676.

Rhebergen, K. S., & Versfeld, N. J. (2005). A Speech Intelligibility Index-based approach to predict the speech reception threshold for sentences in fluctuating noise for normal-hearing listeners. *Journal of the Acoustical Society of America, 114*, 2181–2192.

Rhebergen, K. S., Versfeld, N. J., & Dreschler, W. A. (2005). Release from informational masking by time reversal of native and non-native interfering speech. *Journal of the Acoustical Society of America, 118*, 1274–1277.

Rogers, C. L., Dalby, J., & Nishi, K. (2004). Effects of noise and proficiency on intelligibility of Chinese-accented English. *Language and Speech, 47*, 139–154.

Scott, D. R., & Cutler, A. (1984). Segmental phonology and the perception of syntactic structure. *Journal of Verbal Learning and Verbal Behavior, 23*, 450–466.

Simpson, S. A., & Cooke, M. (2005). Consonant identification in N-talker babble is a nonmonotonic function of N. *Journal of the Acoustical Society of America, 118*, 2775–2778.

Sledd, J. (1966). Breaking, umlaut, and the southern drawl. *Language, 42*, 18–41.

Studebaker, G. (1985). A "rationalized" arcsine transform. *Journal of Speech, Language, and Hearing Research, 28*, 455–462.

Sumner, M., & Samuel, A. G. (2009). The effect of experience on the perception and representation of dialect variants. *Journal of Memory and Language, 60*, 487–501.

Tuinman, A., Mitterer, H., & Cutler, A. (2011). Perception of intrusive /r/ in English by native, cross-language and cross-dialect listeners. *Journal of the Acoustical Society of America, 130*, 1643–1652.

Van Engen, K. J. (2010). Similarity and familiarity: Second language sentence recognition in first- and second-language multi-talker babble. *Speech Communication, 52*, 943–953.

Van Engen, K. J., & Bradlow, A. R. (2007). Sentence recognition in native- and foreign-language multi-talker background noise. *Journal of the Acoustical Society of America, 121*, 519–526.

Van Wijngaarden, S. J., Steeneken, H. J. M., & Houtgast, T. (2002). Quantifying the intelligibility of speech in noise for non-native listeners. *Journal of the Acoustical Society of America, 111*, 1906–1916.