Published in final edited form as:
Lang Speech. 2008 ; 51(Pt 3): 175–198. doi:10.1177/0023830908098539.

Perception of Dialect Variation in Noise: Intelligibility and Classification

Cynthia G. Clopper and **Ann R. Bradlow**
Northwestern University

Abstract

Listeners can explicitly categorize unfamiliar talkers by regional dialect with above-chance performance under ideal listening conditions. However, the extent to which this important source of variation affects speech processing is largely unknown. In a series of four experiments, we examined the effects of dialect variation on speech intelligibility in noise and the effects of noise on perceptual dialect classification. Results revealed that, on the one hand, dialect-specific differences in speech intelligibility were more pronounced at harder signal-to-noise ratios, but were attenuated under more favorable listening conditions. Listener dialect did not interact with talker dialect; for all listeners, at a range of noise levels, the General American talkers were the most intelligible and the Mid-Atlantic talkers were the least intelligible. Dialect classification performance, on the other hand, was poor even with only moderate amounts of noise. These findings suggest that at moderate noise levels, listeners are able to adapt to dialect variation in the acoustic signal such that some cross-dialect intelligibility differences are neutralized, despite relatively poor explicit dialect classification performance. However, at more difficult noise levels, participants cannot effectively adapt to dialect variation in the acoustic signal and cross-dialect differences in intelligibility emerge for all listeners, regardless of their dialect.

Keywords

dialect variation; speech intelligibility; speech perception in noise

1. Introduction

Dialect variation is one of the many important sources of talker-specific variation in the speech signal (Abercrombie, 1967). Recent research in sociolinguistics and speech science has shown that listeners can judge the regional or ethnic background of unfamiliar talkers based on short speech samples with above-chance accuracy, suggesting that naïve listeners have cognitive categories for dialect variation (Clopper, Conrey, & Pisoni, 2005; Clopper & Pisoni, 2006a; Purnell, Idsardi, & Baugh, 1999; Van Bezooijen & Gooskens, 1999; Williams, Garrett, & Coupland, 1999). However, performance in these tasks is typically relatively poor in comparison to other types of speech categorization tasks; while gender or phoneme classification accuracy can exceed 90 percent under some degraded listening conditions (Lass, Hughes, Bowyer, Waters, & Bourne, 1976; Miller & Nicely, 1955), dialect classification accuracy ranges from 25% to 50% under ideal listening conditions. This disparity between

Address for correspondence. Cynthia G. Clopper, Department of Linguistics, Ohio State University, Oxley Hall 222, 1712 Neil Avenue, Columbus, OH 43210, USA; < clopper.1@osu.edu>.

dialect classification and gender or phoneme classification performance may reflect somewhat poorly specified cognitive representations for social variation.

In one study, Clopper and Pisoni (2007) conducted an auditory free classification experiment in which they asked listeners in the United States to classify talkers by regional dialect based on short, sentence-length utterances. In a free classification task, participants are asked to classify a set of stimulus items in the absence of experimenter-defined category labels. Both the number of categories and the dimensions along which the stimulus materials should be classified may be specified by the experimenter or left unspecified (Imai & Garner, 1965). Clopper and Pisoni (2007) asked their participants to group the talkers by regional dialect of American English, and the listeners were permitted to make as many groups of talkers with as many talkers in each group as they wished. The talkers came from six dialect regions of the United States: New England, Mid-Atlantic, North, Midland, South, and West. The listeners exhibited relatively fine-grained perception of regional variation and produced an average of 8–10 groups of talkers. However, an analysis of the participants' classifications revealed that the listeners produced approximately two-thirds as many errors (19%) as they did correct talker pairings (28%), suggesting that even under ideal listening conditions, the classification task was quite difficult.

A multidimensional scaling analysis of the classification data revealed three primary dimensions of perceptual talker similarity: geography (Northern vs. Southern dialects), linguistic markedness (many vs. few characteristic regional dialect properties), and gender. The first two dimensions create a perceptual similarity space with four quadrants that distinguish the marked Northern dialects (Mid-Atlantic and North), the marked Southern dialect (South), and the relatively unmarked Northern dialects (New England, Midland, and West). These three relatively unmarked dialects share many phonological properties and collectively make up the General American dialect (see Clopper, Levi, & Pisoni, 2006). The unmarked Southern quadrant in the similarity space is empty because Southern varieties of American English tend to be phonologically marked. The results of the scaling analysis suggest that the perceptual dialect similarity spaces of listeners without explicit linguistic training reflect the linguistic properties of regional variation. Thus, despite the overall difficulty of the task, the participants' classification judgments were highly interpretable with respect to phonological dialect variation.

Very little research has examined the extent to which listeners can normalize for or adapt to dialect variation in speech processing, and the effects of dialect variation on speech processing are only poorly understood. It is well documented that both linguistic and indexical sources of variation significantly affect speech intelligibility. For example, meaningful sentences are more intelligible than semantically anomalous sentences in quiet and in noise (Miller & Isard, 1963), target words in highly predictable sentence contexts are more intelligible than target words in semantically unpredictable contexts in noise (Lieberman, 1963), and clearly produced speech is more intelligible than plain speech in quiet (Picheny, Durlach, & Braida, 1985) and in noise (Payton, Uchanski, & Braida, 1994). In addition, familiar talkers are more intelligible in noise than unfamiliar talkers (Nygaard, Sommers, & Pisoni, 1994), and familiar foreign accents are more intelligible than unfamiliar accents (Bradlow & Bent, 2003; Clarke, 2003). For native listeners, native speech is more intelligible than foreign-accented speech, but non-native listeners can perform as well or better on foreign-accented speech than native speech when they share a native language with the non-native talker (e.g., Bent & Bradlow, 2003; Stibbard & Lee, 2006).

With respect to dialect variation, Mason (1946) reported that talkers and listeners who shared a dialect were more mutually intelligible in noise than talkers and listeners from different dialect regions. In addition, talkers from the Northern Midwest (Illinois, Wisconsin, and Michigan) were somewhat more intelligible overall than talkers from the South. Labov and Ash (1997) also found evidence of a local dialect preference in a transcription task with phrase-length utterances in good listening conditions. Listeners from Birmingham, Alabama, were more accurate overall than listeners from Philadelphia or Chicago when the speech samples were produced by a talker from Birmingham. These results suggest that when a listener encounters a familiar talker, accent, or dialect, the appropriate mapping between acoustic-phonetic and lexical information is already in place and speech intelligibility is improved relative to unfamiliar talkers, accents, or dialects.

Research on speech intelligibility has also shown that the level of degradation can interact with linguistic and indexical factors to produce differential effects of noise under various talker, listener, and stimulus conditions. For example, children with learning disabilities show a greater decline in performance as listening conditions become more difficult than children without learning disabilities (Bradlow, Kraus, & Hayes, 2003). Similarly, speech intelligibility performance by normal-hearing native adult listeners is less affected by increases in noise or reverberation levels than performance by either hearing-impaired (Plomp, 1994) or non-native (Nábelek & Donahue, 1984; Takata & Nábelek, 1990) listeners. Of particular interest for the present study is the finding that disproportionate effects of noise for highly proficient non-native listeners relative to native listeners emerge only under difficult listening conditions (Mayo, Florentine, & Buus, 1997; Meador, Flege, & Mackay, 2000; Rogers, Lister, Febo, Besing, & Abrams, 2006). With respect to dialect variation, these results suggest that differences in intelligibility between familiar and unfamiliar dialects might only become evident under relatively unfavorable listening conditions.

The goal of the current study was to further explore the effects of noise on the perception of dialect variation. In a series of four experiments, we examined the intelligibility and perceptual classification of four regional varieties of American English by listeners with three distinct residential histories. The talkers came from the Mid-Atlantic, Northern, Southern, and General American regions of the United States. The listeners came from the Northern and General American regions. Listeners who had lived in multiple different dialect regions (Mobile listeners) were also included. In Experiments 1 and 2, we used a sentence transcription task to examine the effects of talker dialect and talker-listener dialect mismatch on speech intelligibility at three signal-to-noise ratios. We predicted that performance would be best when the talker and the listener shared a dialect, and that the Mobile listeners would be more accurate overall, given their increased exposure to dialect variation. We also predicted that differences in intelligibility between the dialects would increase under more difficult listening conditions. In Experiments 3 and 4, we used an auditory free classification task to explore the effects of noise on the perception of variation. We expected that performance would be relatively poor, given that the task is quite difficult even under ideal listening conditions. However, we wanted to be able to compare the effects of noise on the intelligibility and classification of regional dialects of American English and we expected that we would observe a parallel decline in classification and intelligibility performance at increasingly difficult signal-to-noise ratios. A parallel decline across the two tasks would suggest that the effects of degradation on both classification and intelligibility could be accounted for by a single model of perceptual adaptation to noise.

2 Experiment 1

2.1 Listeners

Seventy-six listeners were recruited from the Northwestern University community for participation in Experiment 1. The listeners received partial course credit in an introductory Linguistics course for participating. Data were excluded from 34 listeners prior to analysis: 15 were bilingual, two reported a history of a hearing or speech disorder at the time of testing, two were non-native speakers of English, and 15 did not meet the residency requirements described below. The remaining 42 listeners were all monolingual native speakers of American English, with no history of hearing or speech disorders. They ranged in age from 17–30 years old.

The listeners were divided into three groups based on their residential history. The North group included the 14 listeners who had lived exclusively in the Northern dialect region prior to the age of 18. The General American group included the 14 listeners who had lived exclusively in the General American dialect regions (including the New England, Midland, and Western regions) prior to the age of 18. The Mobile group included the 14 listeners who had lived in more than one dialect region prior to the age of 18. Twelve of the Mobile listeners had lived in two different dialect regions and two had lived in three different regions. In addition, three of the Mobile listeners had lived abroad. No other dialect regions were represented by an adequate number of listeners for inclusion in this experiment.

2.2 Talkers

Twenty-four talkers were selected from the Nationwide Speech Project corpus (Clopper & Pisoni, 2006b) for use in Experiment 1. Three male and three female talkers from each of four dialect regions in the United States (Mid-Atlantic, North, South, and General American) were selected. The talkers were 18–25 years old at the time of recording and had lived exclusively in their respective dialect region before the age of 18. Both parents of each talker were also from the same dialect region. Just as residential history was used to classify the listeners by dialect, residential history was also the only factor used to classify the talkers by dialect.

However, previous acoustic analyses of isolated words produced by this set of talkers confirmed that they reliably exhibited dialect-specific vowel variants in their speech (Clopper, Pisoni, & de Jong, 2005). In particular, the Mid-Atlantic talkers produced /ɑ/, //, and /ɔ/ with overlapping second formant values, distinguishing them only in terms of vowel height. The Northern talkers produced vowels consistent with the Northern Cities Chain Shift, an ongoing vowel shift in the Northern dialect region (Labov, 1998), including the raising and fronting of /æ/, the fronting of /ɑ/, the backing and lowering of /ɛ/, and the backing of //. The Southern talkers produced vowels consistent with the Southern Vowel Shift, an ongoing vowel shift in the Southern dialect region (Labov, 1998), including the centralization of /e/, the peripheralization of /ɛ/, and the fronting of /u/ and /o/. The General American talkers produced a merger of /ɑ/ and /ɔ/ and variable fronting of /u/ and /o/.

We conducted an additional acoustic analysis of the speech of this set of talkers to confirm that no common acoustic correlates of speech intelligibility were significantly different across the four dialects. Vowel space area and vowel space dispersion were not significantly affected by talker dialect, although the female talkers exhibited significantly greater vowel space area and vowel space dispersion than the males. Speaking rate was also not significantly affected by dialect or gender. Given that males and females were distributed evenly across the dialects, and that the gender × dialect interactions were not significant for any of the acoustic measures, we expected that talker gender would not significantly affect our results. Nevertheless, we included talker gender as a within-subject factor in our statistical analysis of the data.

2.3 Stimulus materials

Four different meaningful English sentences were selected for each talker, for a total of 96 sentences. The sentences were taken from the high predictability section of the Speech Perception in Noise (SPIN) test (Kalikow, Stevens, & Elliott, 1977). Each sentence was five to eight words in length, including two to five content words, and the final word in each sentence was highly predictable from the preceding semantic context. The sentences were mixed with speech-shaped noise at two signal-to-noise ratios (SNRs): −2 dB and −6 dB. Each stimulus item was preceded by a 900 ms leader and followed by a 500 ms tail. The leader was composed of 400 ms of silence followed by 500 ms of noise. The tail was composed of 500 ms of noise. The addition of the noise leader and noise tail resulted in an impressionistically less abrupt onset and offset to the stimulus materials and prevented the initial and final segments of the stimulus utterances from being further masked by the onset and offset of the noise, respectively.

2.4 Procedure

The listeners were seated in individual sound-attenuated booths in front of personal computers equipped with a keyboard and headphones (Sony HD580). The sentences were presented one at a time over the headphones and the participants were asked to write down as much of the sentence as they could understand on a prepared answer sheet. The experiment was self-paced and the listeners pressed the spacebar on the keyboard to proceed to the next trial. They could take as much time as they needed to respond, but they were only permitted to listen to each sentence one time.

Twelve practice trials produced by a different set of General American talkers were presented first at the easier (−2 dB) SNR. The listeners then heard half of the sentences (two sentences per talker) in a fixed random order at −2 dB SNR, followed by the other half of the sentences in a fixed random order at −6 dB SNR. The easier SNR was presented first to all participants, so that any effects of practice on transcription performance would reduce the difference between the two noise levels and result in a more conservative evaluation of the effects of noise on intelligibility. The sentence lists were counterbalanced such that each sentence was presented to half of the listeners at each SNR. No sentence was repeated over the course of the experiment.

The participants' responses were scored by content words correct, including all nouns, main verbs, adjectives, and adverbs. Incorrect morphological endings resulted in an incorrect score. However, homophones (such as piece for peace) were scored as correct.

2.5 Results

A summary of the results of Experiment 1 is shown in Table 1. Percent correct scores were converted to rationalized arcsine units (rau) prior to analysis because performance in the −6 dB condition was close to the floor (Studebaker, 1985). A repeated measures ANOVA on accuracy with talker dialect, talker gender, and SNR as within-subject factors and listener dialect as a between-subject factor revealed significant main effects of talker dialect, $F(3, 117) = 156.0$, $p < .001$, and SNR, $F(1, 117) = 582.7$, $p < .001$, a significant talker dialect × SNR interaction, $F(3, 117) = 3.8$, $p = .01$, a significant talker dialect × talker gender interaction, $F(3, 117) = 114.1$, $p < .001$, and a significant talker dialect × talker gender × SNR interaction, $F(6, 117) = 2.8$, $p = .04$. The main effect of listener dialect and all other interactions were not significant.

As expected, performance was better at the easier SNR (−2 dB) than the harder SNR (−6 dB). Paired comparisons across the four talker dialects at each SNR revealed the source of the talker dialect × SNR interaction. At −2 dB, the General American talkers were more intelligible than any of the other three dialects and the Mid-Atlantic talkers were less intelligible than each of the other three dialects, all $p < .02$. Intelligibility did not differ between the Northern and Southern talkers. At −6 dB, all pairwise comparisons were significant, $p < .02$. The General American talkers were the most intelligible, followed by the Southern, Northern, and Mid-Atlantic talkers. A summary of these pairwise comparisons is shown for each SNR in Figure 1.

An inspection of the talker dialect by talker gender interaction revealed that the female talkers were more intelligible than the male talkers for the Mid-Atlantic and General American dialects, but that the male talkers were more intelligible than the female talkers for the Northern and Southern dialects. In addition, pairwise comparisons across talker gender for each talker dialect at each SNR revealed the locus of the three-way interaction: pairwise gender differences were significant for all talker dialect × SNR pairs, $p < .001$, except the Northern talkers at the harder (−6 dB) SNR.

2.6 Discussion

We observed a main effect of talker dialect on intelligibility that was mediated by SNR. At the more difficult SNR (−6 dB), we observed significant differences between all four dialects with respect to intelligibility. However, at the more moderate SNR (−2 dB), intelligibility did not differ between the Northern and Southern dialects, suggesting an attenuation of the talker dialect effect in the more favorable listening condition. Contrary to our predictions, the effect of listener dialect was not significant. For this set of talkers and listeners, the General American dialect was the most intelligible and the Mid-Atlantic dialect was the least intelligible, regardless of the listener's residential history. Finally, despite the overall effects of talker gender on vowel space area and vowel space dispersion, the effect of talker gender on intelligibility was mediated by dialect. The Mid-Atlantic and General American female talkers were more intelligible than the Mid-Atlantic and General American male talkers, respectively, but the Northern and Southern male talkers were more intelligible than the Northern and Southern female talkers, respectively. Experiment 2 was designed to explore the effects of talker dialect and talker gender on intelligibility at a higher SNR (+2 dB) to determine whether the attenuation of dialect differences from −6 dB SNR to −2 dB SNR would continue under more favorable listening conditions.

3 Experiment 2

3.1 Listeners

Fifty-six listeners were recruited from the Northwestern University community for participation in Experiment 2. The listeners received partial course credit in an introductory linguistics course for participating. Data from 32 listeners were excluded prior to analysis: 13 were bilingual, three reported a history of a hearing or speech disorder at the time of testing, one was a non-native speaker of English, 14 did not meet our residency requirements, and one set of data was lost due to experimenter error. The remaining 24 listeners were all monolingual native speakers of American English with no history of hearing or speech disorders. They ranged in age from 18–22 years old. As in Experiment 1, the listeners were divided into three groups based on their residential history: North ($N = 4$), General American ($N = 12$), and Mobile ($N = 8$). Five of the Mobile listeners had lived in two different dialect regions and three had lived in three regions. In addition, three of the Mobile listeners had lived abroad.

3.2 Talkers

The same set of 24 talkers was used in Experiment 2 as in Experiment 1.

3.3 Stimulus materials

The stimulus materials in Experiment 2 were the same as in Experiment 1, except that the sentences were mixed with speech-shaped noise at −2 dB and +2 dB SNRs.

3.4 Procedure

The same procedure was used in Experiment 2 as in Experiment 1, except that after 12 practice trials at +2 dB SNR, the listeners heard half of the sentences (two sentences per talker) in a fixed random order at +2 dB SNR, followed by the other half of the sentences in a fixed random order at −2 dB SNR.

3.5 Results

A summary of the results of Experiment 2 is shown in Table 2. Percent correct scores were converted to rationalized arcsine units (rau) prior to analysis because performance in the +2 dB condition was close to ceiling (Studebaker, 1985). A repeated measures ANOVA on accuracy with talker dialect, talker gender, and SNR as within-subject factors and listener dialect as a between-subject factor revealed significant main effects of talker dialect, $F(3, 63) = 70.7$, $p < .001$, talker gender, $F(1, 63) = 7.8$, $p = .01$, SNR, $F(1, 63) = 133.9$, $p < .001$, and listener dialect, $F(2, 63) = 3.5$, $p = .048$, a significant talker dialect × SNR interaction, $F(3, 63) = 5.7$, $p = .002$, and a significant talker dialect × talker gender interaction, $F(3, 63) = 71.9$, $p < .001$. All other interactions were not significant.

As expected, performance was better at the easier SNR (+2 dB) than the harder SNR (−2 dB). Post-hoc paired comparisons across the four talker dialects at each SNR revealed that at +2 dB, the General American talkers were significantly more intelligible than the Northern and Mid-Atlantic talkers, and the Mid-Atlantic talkers were less intelligible than the other three dialects, all $p < .02$. Intelligibility did not differ between the Northern and Southern talkers or between the General American and Southern talkers. At −2 dB, the General American and Southern talkers were more intelligible than the Northern and Mid-Atlantic talkers and the Northern talkers were significantly more intelligible than the Mid-Atlantic talkers, all $p < .03$. Intelligibility did not differ between the General American and Southern talkers. A summary of these pairwise comparisons is shown for each SNR in Figure 2.

In addition, the female talkers were more intelligible than the male talkers, but as in Experiment 1, this effect was mediated by talker dialect. The female talkers were more intelligible than the male talkers for the Mid-Atlantic and General American dialects, whereas the male talkers were more intelligible than the female talkers for the Northern and Southern dialects. Finally, while the main effect of listener dialect was significant, none of the post-hoc paired comparisons reached significance, suggesting that any effects of listener dialect on intelligibility performance were quite small.

In Experiment 1, the Northern and Southern talkers were equally intelligible at the −2 dB SNR and all other pairwise comparisons were significant. At the same SNR in Experiment 2, the Southern and General American talkers were equally intelligible and all other pairwise comparisons were significant. To determine if the experimental condition was a significant factor in intelligibility, we conducted a repeated measures ANOVA on accuracy at the −2 dB SNR with talker dialect as a within-subject factor and experiment (Experiment 1 or 2) as a between-subject factor. The main effect of talker dialect was significant, as expected based on the previous analyses. However, the main effect of experiment and the experiment × talker dialect interaction were not significant. Thus, in general, the same results were obtained at the moderate (−2 dB) SNR, regardless of whether it was the “easy” or “hard” SNR in the experiment.

3.6 Discussion

Once again, we observed a main effect of talker dialect that was mediated by SNR. While we observed some attenuation of the dialect effect on intelligibility at the moderate (−2 dB) SNR in Experiment 1, we observed even more attenuation at the easier (+2 dB) SNR in Experiment 2. In particular, the number of significant pairwise dialect comparisons decreased from five at the moderate (−2 dB) SNR to four at the easier (+2 dB) SNR. These results suggest that as listening conditions became more favorable, participants were more able to adapt to dialect variation in the transcription task. Even at the most favorable SNR that we tested, however, the Mid-Atlantic talkers were still less intelligible than the talkers from the other dialects, and the Northern talkers were less intelligible than the General American talkers. Additional research is needed to determine the SNR at which all effects of dialect variation on intelligibility are neutralized.

Unlike in Experiment 1, talker gender and listener dialect also emerged as significant factors in intelligibility performance. As in Experiment 1, the effect of talker gender was mediated by talker dialect. The Mid-Atlantic and General American female talkers were more intelligible than the Mid-Atlantic and General American male talkers, respectively, but the Northern and Southern female talkers were less intelligible than the Northern and Southern male talkers, respectively.

Post-hoc analysis of the main effect of listener dialect revealed that none of the pairwise comparisons were significant. However, we did not fully sample the listener dialect space due to the limits of the Northwestern University student population, and it is possible that listeners from the Mid-Atlantic or Southern regions might have shown a preference for their own dialect, relative to listeners from other regions. While the Northern talkers reliably produced the Northern Cities Chain Shift in their speech, Northern listeners do not explicitly perceive this shift (Niedzielski, 1999). This mismatch between perception and production may explain why the Northern listeners failed to perform more accurately on the Northern talkers than the General American or Mobile listeners.

We obtained some preliminary data from eight Mid-Atlantic listeners in Experiment 2. As shown in Table 3, the performance of the Mid-Atlantic listeners is consistent with the results obtained from the Northern, General American, and Mobile listeners. This finding is in direct contradiction to previous findings on the effects of a shared dialect or native language on speech intelligibility performance (e.g., Bradlow & Bent, 2003;Labov & Ash, 1997;Mason, 1946). In interpreting our results, however, it is important to bear in mind that both the talkers and the listeners were relatively mobile, educated young adults from predominantly middle-class backgrounds. We expect that the preference for local dialects would emerge if the talkers and the listeners came from more entrenched local communities with more extreme regional variation and/or less exposure to regional variation in the United States.

In both Experiments 1 and 2, more pairwise differences in intelligibility were observed between the four dialects in the harder SNR condition than the easier SNR condition. These findings suggest that noise differentially affects listeners' ability to adapt to dialect variation in the acoustic signal in a speech intelligibility task. In particular, at the more difficult SNRs, more differences between the dialects emerged because the listeners were less able to rapidly normalize for dialect differences than when the signal was less degraded at the easier SNRs. The finding that increased noise impaired listeners' ability to rapidly adapt to dialect variation also suggests that the listeners still had access to dialect-specific information in the acoustic signal at the harder SNRs. However, the extent to which listeners can use this source of variation in an explicit dialect classification task in noise has not been examined. With respect to gender, the data from Experiments 1 and 2 suggest that this effect may be more robust to the effects of noise than dialect variability. That is, under both relatively easy and hard SNRs, gender-based intelligibility differences emerged, suggesting that listener adaptation to idiolectal differences may be less subject to interference from background noise than dialect differences.

Experiments 3 and 4 were designed to explore how noise would affect explicit dialect classification performance. Previous research has shown that listeners perform forced-choice dialect classification tasks with relatively poor performance under ideal listening conditions (Clopper et al., 2005; Clopper & Pisoni, 2006a), but that highly interpretable perceptual dialect similarity spaces can be obtained using free dialect classification tasks, despite relatively low accuracy (Clopper & Pisoni, 2007). We used a free classification task in Experiments 3 and 4 to obtain both a measure of classification accuracy in noise and perceptual dialect similarity spaces for comparison to the previous results without noise. In Experiment 3, the stimulus materials were presented at the moderate (−2 dB) SNR.

4 Experiment 3

4.1 Listeners

Thirty-five listeners were recruited from the Northwestern University community for participation in Experiment 3. The listeners received partial course credit in an introductory Linguistics course for participating. Data from 15 listeners were excluded prior to analysis: seven were bilingual, five reported a history of a hearing or speech disorder at the time of testing, and three did not meet our residency requirements. The remaining 20 listeners were all monolingual native speakers of American English with no history of a hearing or speech disorder. They ranged in age from 18–22 years old. As in Experiments 1 and 2, the residential history of the listeners was restricted to residents of the North ($N = 4$) and General American ($N = 7$) dialects, and Mobile listeners ($N = 9$). Eight of the Mobile listeners had lived in two dialect regions and one had lived in three different dialect regions. In addition, one of the Mobile listeners had lived abroad.

4.2 Talkers

The same set of 24 talkers was used in Experiment 3 as in the previous experiments.

4.3 Stimulus materials

One unique sentence of moderate intelligibility was selected for each talker for use in Experiment 3, for a total of 24 different sentences. The sentences were mixed with speech-shaped noise at −2 dB SNR. Each stimulus item was preceded by a 900 ms leader and followed by a 500 ms tail. The leader was composed of 400 ms of silence followed by 500 ms of noise. The tail was composed of 500 ms of noise. The addition of the noise leader and noise tail resulted in an impressionistically less abrupt onset and offset to the stimulus materials and prevented the initial and final segments of the stimulus utterances from being further masked by the onset and offset of the noise, respectively.

4.4 Procedure

The listeners were seated in individual sound-attenuated booths in front of personal computers equipped with a mouse and headphones (Sony HD580). Each talker was represented on the computer screen by a small black square with a unique alphabetic identifier (the talker's initials). Each square was linked to the sound file containing the sentence produced by that talker. The squares were arranged in two columns on the left-hand side of the screen. On the right-hand side of the screen was a 16 × 16 cell grid. The listeners could listen to each sentence by double-clicking on the black square with the mouse. They could move the squares around the screen by dragging them with the mouse.

The listeners were instructed to listen to each of the talkers and then move them onto the grid and put them into groups based on regional dialect. They were permitted to make as many groups as they wanted with as many talkers in each group as they wished. They were not required to put the same number of talkers in each group. The participants were encouraged to listen to and rearrange the talkers as many times as they wanted until they were satisfied with their classification (see Clopper & Pisoni, 2007).

All of the participants completed the free classification task after completing an unrelated transcription experiment.

4.5 Results

A summary of the listeners' classification behavior is shown in Table 4. On average, the listeners made approximately five to six groups of talkers with four to five talkers per group. Accuracy was assessed by calculating the percentage of correct talker pairings (hits) produced by each listener out of all possible correct pairings ($N = 60$) and the percentage of incorrect talker pairings (false alarms) produced by each listener out of all possible incorrect pairings ($N = 216$). The accuracy measures were not normalized for total number of talker groups produced, which means that participants who made few groups with many talkers per group typically had a higher percentage of both hits and false alarms than participants who made many groups with few talkers per group. Overall accuracy is therefore captured by the difference between hits and false alarms. In Experiment 3, the listeners produced an average of 22% of the total number of possible correct pairs of talkers and an average of 22% of the total number of possible incorrect pairs of talkers. A *t*-test confirmed that the difference between the percentage of correct and incorrect pairings was not significant, suggesting relatively poor performance overall.

A multidimensional scaling (MDS) analysis was conducted to examine the participants' classification behavior in more detail. A 24 × 24 matrix was computed by summing the total number of times each pair of talkers was grouped together across all 20 listeners. The resulting talker similarity matrix reflected the gradient similarity of the 24 talkers, such that pairs of talkers who were consistently placed in different groups received a score of 0 and pairs of talkers who were consistently grouped together received a score of 20. The talker similarity matrix was submitted to a MDS analysis in one and two dimensions. The two-dimensional solution was selected for interpretation and discussion because stress was greatly reduced from the one- to two-dimensional solution (one-dimension stress = .39, two-dimension stress = .26) and the two dimensions were interpretable.

The two-dimensional MDS solution is shown in Figure 3. The male talkers are represented by filled symbols and the female talkers are represented by open symbols. Clockwise rotation of the space approximately 60 degrees reveals two orthogonal dimensions of perceptual similarity: gender and geography. Most of the male talkers are on the right-hand side of the figure, and most of the female talkers are on the left-hand side of the figure. While the participants were instructed to classify the talkers by regional dialect, they were apparently unable to ignore talker gender in making their classifications. The second dimension, geography, roughly separates the Northern dialects (including the Mid-Atlantic and Northern talkers) on the bottom right from the non-Northern dialects (including the Southern and General American talkers) on the upper left, suggesting that the listeners had some success classifying the talkers by regional dialect. However, a closer inspection of Figure 3 reveals that many of the talkers appear on the "wrong" side of the geography dimension, confirming the conclusion from the summary data that performance was relatively poor overall.

4.6 Discussion

In the free classification task with moderate noise (−2 dB SNR), the listeners made on average five to six groups of talkers (range: 3–10) with four to five talkers per group (range: 1–19). They also made as many errors as they did correct pairings. In similar experiments with larger groups of talkers and without noise, Clopper and Pisoni (2007) found that participants made eight to ten groups of talkers (range: 3–30) with seven to nine talkers per group (range: 1–38). Clopper and Pisoni's (2007) listeners also made more correct pairings (28%) than incorrect pairings (19%) in the no-noise condition. Thus, while overall classification strategy was comparable between the earlier experiments without noise and the current experiment, performance was much less accurate with moderate (−2 dB SNR) noise than in the no-noise condition.

Despite the participants' low overall accuracy in the current experiment, however, the two-dimensional MDS solution was interpretable, with gender and geography emerging as the two primary dimensions of perceptual similarity. As in the earlier studies of perceptual dialect similarity (Clopper et al., 2006; Clopper & Pisoni, 2007), participants were unable to ignore talker gender when making regional dialect classification judgments and gender emerged as the first perceptual dimension of similarity. Geography emerged as the second, much weaker, dimension of perceptual similarity. In the earlier free classification studies without noise, the geography dimension separated the Southern talkers from all of the other dialects, whereas in the current experiment, the General American talkers were also perceived as Southern. General American and Southern dialects share back vowel fronting, suggesting that the listeners may have been relying on different acoustic-phonetic cues to dialect region when noise was present than when it was absent. The results of this experiment should be interpreted with caution, however, given that the overall perceptual similarity space obtained in the MDS analysis in the current experiment was relatively noisy and the second dimension (geography) did not clearly differentiate all of the Northern talkers from all of the non-Northern talkers. Given the listeners' difficulty with this task at the moderate (−2 dB) SNR, in Experiment 4 we examined free classification performance at the easier (+2 dB) SNR.

5 Experiment 4

5.1 Listeners

Twenty-three listeners were recruited from the Northwestern University community for participation in Experiment 4. The listeners received partial course credit in an introductory Linguistics course for participating. Data from six participants were excluded prior to analysis: four were bilingual, one reported a history of a hearing or speech disorder at the time of testing, and one did not meet our residency requirements. The remaining 17 listeners were all monolingual native speakers of American English with no history of hearing or speech disorders. They ranged in age from 18–23 years old. As in the previous experiments, the residential history of the listeners was restricted to Northern ($N = 3$) and General American ($N = 10$) dialects, and Mobile ($N = 4$) listeners. All four of the Mobile listeners had lived in two different dialect regions and none had lived abroad.

5.2 Talkers

The same set of 24 talkers were used in this experiment as in the previous experiments.

5.3 Stimulus materials

The same set of stimulus materials were used in Experiment 4 as in Experiment 3, except that the sentences were mixed with speech-shaped noise at a SNR of +2 dB.

5.4 Procedure

The same procedure was used in Experiment 4 as in Experiment 3.

5.5 Results

A summary of the listeners' classification behavior is shown in Table 5. On average, the listeners made approximately seven groups of talkers with three to four talkers per group. In addition, they produced an average of 17% of the total number of possible correct pairs of talkers and an average of 15% of the total number of possible incorrect pairs of talkers. A *t*-

test confirmed that the difference between the percentage of correct and incorrect pairings was not significant, suggesting relatively poor performance overall.

As in Experiment 3, a 24 × 24 talker similarity matrix was constructed from the listeners' responses and was submitted to one- and two-dimensional MDS analyses. The two-dimensional solution was selected for interpretation and discussion because stress was greatly reduced from the one- to two-dimensional solution (one-dimension stress = .48, two-dimension stress = .29) and the two dimensions were interpretable. The two-dimensional MDS solution is shown in Figure 4. The male talkers are represented by filled symbols and the female talkers are represented by open symbols. Clockwise rotation of the space approximately 60 degrees reveals two interpretable dimensions of perceptual similarity: gender and linguistic markedness. Most of the male talkers are at the bottom of the space and most of the female talkers are at the top of the space. Once again, the listeners could not ignore this important dimension of talker similarity in the regional dialect free classification task. The second dimension, linguistic markedness, distinguished those dialects with many characteristic properties (Mid-Atlantic and South) on the right from those dialects with few characteristic perceptual properties (North and General American) on the left. As in Experiment 3, the dimension related to regional dialect is less effective at discriminating talkers than the gender dimension.

5.6 Discussion

As in Experiment 3, the listeners made approximately the same number of correct talker pairings as incorrect pairings. Both the correct and incorrect scores were lower at the easier (+2 dB) SNR in the current experiment than at the moderate (−2 dB) SNR in Experiment 3, as a result of the difference in the number of groups produced by the listeners in the two experiments. The listeners in Experiment 3 made approximately six groups of talkers, whereas the listeners in Experiment 4 made seven groups of talkers. When more groups of talkers are produced, fewer talker pairs are produced overall, which lowers both the percentage of correct pairings and the percentage of incorrect pairings. A *t*-test comparing the difference scores (hits–false alarms) across the two experiments was not significant, confirming that accuracy did not significantly improve at the more favorable (+2 dB) SNR.

The MDS solution revealed two dimensions of perceptual similarity: gender and linguistic markedness. These dimensions are consistent with previous research on the perceptual free classification of regional dialects of American English (Clopper & Pisoni, 2007) and suggest that listeners were able to extract some dialect-specific information from the acoustic signal at +2 dB SNR. Previous research on the perception of the Northern dialect of American English suggests that listeners from the Midwest have difficulty distinguishing the Northern dialect from General American, despite the many acoustic properties that are unique to the Northern region (Clopper et al., 2005; Niedzielski, 1999). The Northern and General American talkers were perceptually similar in the current experiment along the linguistic markedness dimension, suggesting that these listeners also failed to distinguish the Northern dialect from General American. However, in Experiment 3, at the harder (−2 dB) SNR, the Northern and General American dialects were differentiated along the geography dimension, suggesting that noise may highlight some phonetic differences between the dialects, while attenuating others, resulting in different patterns of perceptual similarity across the different conditions. The markedness dimension that emerged in Experiment 4 was only somewhat more successful at distinguishing the dialects than the geography dimension that emerged in Experiment 3, however, so these results should also be interpreted with some caution. Classification in noise is clearly a very difficult task, even at the most favorable (+2 dB) SNR that we tested. Additional research is needed to determine the noise threshold for more accurate dialect classification and

NIH-PA Author Manuscript

NIH-PA Author Manuscript

to explore how noise masking affects the perception of the various acoustic-phonetic cues to dialect region.

6 General discussion

The results of this set of experiments make four contributions to our understanding of the perception and representation of talker dialect variation. First, the finding that the General American dialect was the most intelligible variety for all listeners at all three SNRs suggests a new criterion for determining the standard variety of a given language. Sociolinguists generally balk at the idea of labeling any variety as the standard, given the normative values that such a label typically entails. However, Labov's (1998) division of American English into Northern, Southern, and "The Third Dialect" seems unsatisfactory and some method for applying the "General American" label is necessary. General American English could be defined historically as the variety that has undergone relatively little phonological change over the last 50 or 100 years. Alternatively, it could reflect the most common variants, or least common denominator, of all regional, social, and ethnic varieties. However, it may also be useful to consider perception and intelligibility in determining the standard variety. We have shown here and in previous research (Clopper & Pisoni, 2007) that linguistic markedness is an important perceptual dimension of dialect similarity for naïve listeners. In addition, the intelligibility data in Experiments 1 and 2 of the current study showed that the least marked variety was also the most intelligible. Thus, General American English may be most usefully defined based on a convergence of historical, production, and perceptual evidence.

The intelligibility benefit for General American English has several possible sources. First, the social dominance or prestige associated with General American English may make it a desirable variety for all listeners to understand (Casad, 1974). That is, listeners may be (consciously or unconsciously) socially motivated to understand General American English, even if they are a native speaker of a different dialect. Second, the widespread use of General American English in the media or its geographic distribution over much of the United States may make it one of the most frequently encountered varieties of American English for all listeners and, therefore, one of the most intelligible. Finally, it may be the case that standard varieties in general evolve to contain acoustic-phonetic or structural properties that make them inherently more intelligible. Additional research is needed to determine if other standard varieties of English (such as Received Pronunciation in the British Isles) or standard varieties of other languages are also relatively more intelligible than non-standard regional or social varieties and the extent to which an intelligibility benefit for the standard variety is due to social, geographic, or linguistic factors.

Second, we observed an effect of talker gender on intelligibility that was mediated by talker dialect. The Mid-Atlantic and General American women were more intelligible than the men, whereas the Northern and Southern women were less intelligible than the men. These results are surprising given the acoustic analysis of vowel space expansion and dispersion, which revealed significant main effects of talker gender (the women had more expanded and disperse vowel spaces overall), but no talker dialect × talker gender interactions. Thus, the findings from the acoustic analysis would lead to the prediction that all of the women should be more intelligible than the men, regardless of dialect (Bradlow, Torretta, & Pisoni, 1996). The results of the current study suggest that these effects might be dialect specific, however, and further research is needed to explore the interaction between talker gender and talker dialect in speech intelligibility.

Third, performance on the free classification task in noise was relatively poor, while intelligibility at the same noise levels was well above the floor. At the moderate (−2 dB) SNR, participants in the free classification task made as many incorrect pairings as they did correct pairings and the perceptual similarity structure of their classifications only weakly reflected the North versus South division of American English regional dialects. In addition, General American was classified as non-Northern, contrary to previous results in which similar dimensions were obtained (Clopper & Pisoni, 2007). At the more favorable (+2 dB) SNR, participants again made as many incorrect pairings as correct pairings, but the perceptual similarity structure of their classifications roughly reflected a second important dimension of dialect variation in the United States: linguistic markedness. In both free classification experiments, listeners were unable to ignore gender-specific variation and talker gender emerged as a dimension of perceptual talker similarity. In the intelligibility experiments, we observed attenuation of dialect-specific effects on intelligibility as the SNR was made more favorable. At the hardest (−6 dB) SNR, significant pairwise differences in intelligibility emerged across all four dialects, whereas at the easiest (+2 dB) SNR, the Mid-Atlantic dialect was consistently less intelligible than the other dialects, and General American was more intelligible than the Northern dialect.

Taken together, these results initially appear to present a paradox: at the more difficult SNRs, we observed greater sensitivity to dialect variation in the intelligibility task in the form of more significant pairwise differences in transcription accuracy across the different dialects. In the free classification task, however, performance was poor even at the moderate SNRs, suggesting that the participants were unable to use whatever dialect-specific information was available in the noise-masked signal for explicit classification. Thus, the results of the intelligibility task suggest that listeners are sensitive to dialect variation in noise, whereas the free classification data suggest that they are not. Although dialect classification is quite difficult under all listening conditions, comparisons between the results of the classification experiments in noise in the current study and previous free classification tasks using similar stimulus materials revealed that overall accuracy and interpretability of the perceptual similarity spaces were better in the more favorable listening conditions than the degraded conditions.

The apparent paradox in our results can be resolved by considering the specific demands of the two tasks. In particular, the transcription task examined the indirect influence of dialect variation on speech intelligibility, whereas the classification task examined explicit perceptual dialect classification. Under conditions in which overt classification was at best partly successful (+2 dB SNR), listeners were able to normalize for or rapidly adapt to some of the dialect-specific information in the transcription task in which dialect variation was not overtly relevant. However, under more difficult listening conditions (−6 dB SNR), this normalization failed and more dialect-specific differences in intelligibility emerged. That is, even when listeners could not accurately classify talkers by dialect (+2 dB SNR), they could still adjust their perception to take into account some dialect-specific information to accurately transcribe the sentences, particularly for the Southern and Northern talkers. However, under more difficult conditions (−6 dB SNR), the listeners could not effectively adapt to dialect-specific phonetic realization patterns, and intelligibility differences between the General American, Southern, and Northern dialects emerged. The transcription tasks in the current study were specifically designed to reduce the ability of the listeners to rapidly adapt online to dialect variation. In particular, both talker and dialect were randomized across trials, the sentences were never repeated over the course of the experiment, and the participants were not given any feedback about their accuracy. However, given the growing literature on perceptual learning of indexical properties such as dialect and foreign accent (Clarke & Garrett, 2004; Evans & Iverson, 2004) and the role of noise in perceptual learning (Davis, Johnsrude, Hervais-Adelman, Taylor, & McGettigan, 2005), dialect adaptation and perceptual learning in noise is an important direction for future research.

Fourth, the interaction between SNR and talker dialect in intelligibility revealed that when more information was available in the signal, listeners were better able to normalize for dialect variation. In particular, at the moderate SNRs, the listeners were able to activate appropriate dialect-specific mappings between acoustic-phonetic and lexical information to perform the transcription task. However, when less information was available in the signal, listeners relied on "standard" representations for the acoustic-phonetic mapping and produced more transcription errors, particularly for the more marked dialects (South, North, and Mid-Atlantic). Thus, the mechanism for adaptation or normalization lies in the selection and activation of an appropriate acoustic-phonetic mapping and the differential effect of noise on intelligibility for the four dialects is a result of the relative success of this process.

These differences in perceptual adaptation across noise conditions may reflect properties of the signal, the mental representations of dialect variation, or the perceptual processes required to extract linguistic information from a degraded signal. With respect to the properties of the signal, it may be that dialect-specific acoustic information is disproportionately masked at the higher noise levels. On the one hand, given that variation in American English is described almost exclusively with respect to vowels (e.g., Labov, Ash, & Boberg, 2005) and that vowel information is less masked than consonant information by noise, the properties of the acoustic signals themselves would be expected to contribute little to the observed attenuation effects. On the other hand, the different patterns of perceptual similarity observed across different levels of degradation in the free classification experiments suggest that dialect-specific information may be differentially masked by noise. A possible signal-related source for these differential, dialect-specific effects of noise masking is at the level of prosodic structure. For example, recent work has shown that regional dialects of American English exhibit differences in the acoustic realization of prosodic structure, such as pitch accent alignment and speaking rate (Arvaniti & Garding, 2007; Clopper & Smiljanic, 2006). Further research is needed to explore the effects of noise on the perception of dialect-specific phonetic information.

A second possibility is that as noise levels are increased, participants are less able to successfully match the incoming signal to stored representations of dialect variation because those mental representations are already poorly specified. Clopper and Pisoni (2006a) argued that the poor performance observed in perceptual dialect categorization and classification tasks reflects the nature of the listeners' perceptual dialect categories. In particular, dialect categories, unlike talker, gender, or linguistic categories, are relatively weak, which makes explicit dialect perception tasks difficult. The results of Experiments 3 and 4 in the present study, however, suggest that perceptual dialect categories are somewhat robust to noise, such that listeners can use their representations of dialect variation to complete the free classification task with weakly interpretable results with moderate amounts of noise. Poorly specified representations may contribute to the interaction between noise and dialect intelligibility, but they are probably not the only source of this interaction.

Finally, the greater processing and memory load constraints imposed on the listener in the less favorable listening conditions may force them to adopt a different processing strategy for integrating information from multiple levels of representation (Mattys, White, & Melhorn, 2005). The task of extracting linguistic information from the highly degraded signals may cause the participants to "skip" dialect-level representations altogether in an effort to perceive, encode, and transcribe as much of the utterance as possible before the auditory trace disappears. As a result, appropriate acoustic-phonetic mappings are not always used and the participants produce more errors, particularly for the more phonologically marked dialects. Analogous findings consistent with this interpretation have been reported in the productions of bi-dialectal talkers (Howell, Barry, & Vinson, 2006). The talkers were able to maintain a second dialect under normal auditory feedback conditions, but reverted to their native dialect in both frequency-shifted and delayed auditory feedback conditions. Taken together, these results suggest that for both perception and production, participants increasingly rely on a single dialect target when the testing conditions create a heavy processing load. Processing costs associated with perception in noisy conditions may therefore be a primary contributing factor to the interaction we observed between talker dialect and SNR in the current set of experiments. Additional research is needed to tease apart the sources of this interaction between signal degradation and talker dialect in speech intelligibility to better understand the relationships between speech processing and cognitive representations of linguistic and sociolinguistic categories in memory.

Acknowledgments

This work was supported by NIH NRSA Postdoctoral Fellowship F32 DC007237 and NIH Research Grant R01 DC005794 to Northwestern University. The authors would like to thank Jennifer Alexander, Arim Choi, and Judy Song for their assistance with data collection.

References

ABERCROMBIE, D. Elements of general phonetics. Edinburgh: Edinburgh University Press; 1967.

ARVANITI, A.; GARDING, G. Dialectal variation in the rising accents of American English. In: Cole, J.; Hualde, J., editors. Laboratory Phonology. Vol. 9. Berlin: Mouton de Gruyter; 2007. p. 547-576.

BENT T, BRADLOW AR. The interlanguage speech intelligibility benefit. Journal of the Acoustical Society of America 2003;114:1600–1610. [PubMed: 14514213]

BRADLOW AR, BENT T. Listener adaptation to foreign-accented English. Proceedings of the International Congress of Phonetic Sciences 2003;15:2881–2884.

BRADLOW AR, KRAUS N, HAYES E. Speaking clearly for children with learning disabilities: Sentence perception in noise. Journal of Speech, Language, and Hearing Research 2003;46:80–97.

BRADLOW AR, TORRETTA GM, PISONI DB. Intelligibility of normal speech I: Global and fine-grained acoustic-phonetic talker characteristics. Speech Communication 1996;20:255–272.

CASAD, EH. Dialect intelligibility testing. Norman, OK: Summer Institute of Linguistics, University of Oklahoma; 1974.

CLARKE, CM. Processing time effects of short-term exposure to foreign-accented English. University of Arizona; Tucson: 2003. Doctoral dissertation

CLARKE CM, GARRETT MF. Rapid adaptation to foreign-accented English. Journal of the Acoustical Society of America 2004;116:3647–3658. [PubMed: 15658715]

CLOPPER CG, CONREY BL, PISONI DB. Effects of talker gender on dialect categorization. Journal of Language and Social Psychology 2005;24:182–206.

CLOPPER CG, LEVI SV, PISONI DB. Perceptual similarity of regional dialects of American English. Journal of the Acoustical Society of America 2006;119:566–574. [PubMed: 16454310]

CLOPPER CG, PISONI DB. Effects of region of origin and geographic mobility on perceptual dialect categorization. Language Variation and Change 2006a;18:193–221.

CLOPPER CG, PISONI DB. The Nationwide Speech Project: A new corpus of American English dialects. Speech Communication 2006b;48:633–644.

CLOPPER CG, PISONI DB. Free classification of regional dialects of American English. Journal of Phonetics 2007;35:421–438.

CLOPPER CG, PISONI DB, de JONG K. Acoustic characteristics of the vowel systems of six regional varieties of American English. Journal of the Acoustical Society of America 2005;118:1661–1676. [PubMed: 16240825]

CLOPPER, CG.; SMILJANIC, R. Regional prosodic variation in American English: Midland vs. South. Poster presented at New Ways of Analyzing Variation 35; Columbus, OH. 2006.

DAVIS MH, JOHNSRUDE IS, HERVAIS-ADELMAN A, TAYLOR K, McGETTIGAN C. Lexical information drives perceptual learning of distorted speech: Evidence from the comprehension of noise-vocoded sentences. Journal of Experimental Psychology: General 2005;134:222–241. [PubMed: 15869347]

EVANS BG IV, ERSON P. Vowel normalization for accent: An investigation of best exemplar locations in northern and southern British English sentences. Journal of the Acoustical Society of America 2004;115:352–361. [PubMed: 14759027]

HOWELL P, BARRY W, VINSON D. Strength of British English accents in altered listening conditions. Perception and Psychophysics 2006;68:139–153. [PubMed: 16617838]

IMAI S, GARNER WR. Discriminability and preference for attributes in free and constrained classification. Journal of Experimental Psychology 1965;69:596–608. [PubMed: 14304312]

KALIKOW DN, STEVENS KN, ELLIOTT LL. Development of a test of speech intelligibility in noise using sentence materials with controlled word predictability. Journal of the Acoustical Society of America 1977;61:1337–1351. [PubMed: 881487]

LABOV, W. The three dialects of English. In: Linn, MD., editor. Handbook of Dialects and Language Variation. San Diego: Academic Press; 1998. p. 39-81.

LABOV, W.; ASH, S. Understanding Birmingham. In: Bernstein, C.; Nunnally, T.; Sabino, R., editors. Language Variety in the South Revisited. Tuscaloosa, AL: Alabama University Press; 1997. p. 508-573.

LABOV, W.; ASH, S.; BOBERG, C. Atlas of North American English. New York: Mouton de Gruyter; 2005.

LASS NJ, HUGHES KR, BOWYER MD, WATERS LT, BOURNE V. Speaker sex identification from voiced, whispered, and filtered isolated vowels. Journal of the Acoustical Society of America 1976;59:675–678. [PubMed: 1254794]

LIEBERMAN P. Some effects of semantic and grammatical context on the production and perception of speech. Language and Speech 1963;6:172–187.

MASON HM. Understandability of speech in noise as affected by region of origin of speaker and listener. Speech Monographs 1946;13(2):54–68.

MATTYS SL, WHITE L, MELHORN JF. Integration of multiple speech segmentation cues: A hierarchical framework. Journal of Experimental Psychology: General 2005;134:477–500. [PubMed: 16316287]

MAYO LH, FLORENTINE M, BUUS S. Age of second-language acquisition and perception of speech in noise. Journal of Speech Language and Hearing Research 1997;40:686–693.

MEADOR D, FLEGE JE, MacKAY IR. Factors affecting the recognition of words in a second language. Bilingualism 2000;3:55–67.

MILLER GA, ISARD S. Some perceptual consequences of linguistic rules. Journal of Verbal Learning and Verbal Behavior 1963;2:217–228.

MILLER GA, NICELY PE. An analysis of perceptual confusions among some English consonants. Journal of the Acoustical Society of America 1955;27:338–352.

NÁBELEK AK, DONAHUE AM. Perception of consonants in reverberation by native and non-native listeners. Journal of the Acoustical Society of America 1984;75:632–634. [PubMed: 6699302]

NIEDZIELSKI N. The effect of social information on the perception of sociolinguistic variables. Journal of Language and Social Psychology 1999;18:62–85.

NYGAARD LC, SOMMERS MS, PISONI DB. Speech perception as a talker-contingent process. Psychological Science 1994;5:42–46.

PAYTON KL, UCHANSKI RM, BRAIDA LD. Intelligibility of conversational and clear speech in noise and reverberation for listeners with normal and impaired hearing. Journal of the Acoustical Society of America 1994;95:1581–1592. [PubMed: 8176061]

PICHENY MA, DURLACH NI, BRAIDA LD. Speaking clearly for the hard of hearing I: Intelligibility differences between clear and conversational speech. Journal of Speech and Hearing Research 1985;28:96–103. [PubMed: 3982003]

PLOMP R. Noise, amplification, and compression: Considerations of three main issues in hearing aid design. Ear and Hearing 1994;15:2–12. [PubMed: 8194676]

PURNELL T, IDSARDI W, BAUGH J. Perceptual and phonetic experiments on American English dialect identification. Journal of Language and Social Psychology 1999;18:10–30.

ROGERS CL, LISTER JJ, FEBO DM, BESING JM, ABRAMS HB. Effects of bilingualism, noise, and reverberation on speech perception by listeners with normal hearing. Applied Psycholinguistics 2006;27:465–485.

STIBBARD RM, LEE JI. Evidence against the mismatched interlanguage intelligibility benefit hypothesis. Journal of the Acoustical Society of America 2006;120:433–442. [PubMed: 16875239]

STUDEBAKER GA. A “rationalized” arcsine transform. Journal of Speech and Hearing Research 1985;28:455–462. [PubMed: 4046587]

TAKATA Y, NÁBELEK AK. English consonant recognition in noise and in reverberation by Japanese and American listeners. Journal of the Acoustical Society of America 1990;88:663–666. [PubMed: 2212289]

Van BEZOOIJEN R, GOOSKENS C. Identification of language varieties: The contribution of different linguistic levels. Journal of Language and Social Psychology 1999;18:31–48.

WILLIAMS, A.; GARRETT, P.; COUPLAND, N. Dialect recognition. In: Preston, DR., editor. Handbook of Perceptual Dialectology. Philadelphia: John Benjamins; 1999. p. 345-358.

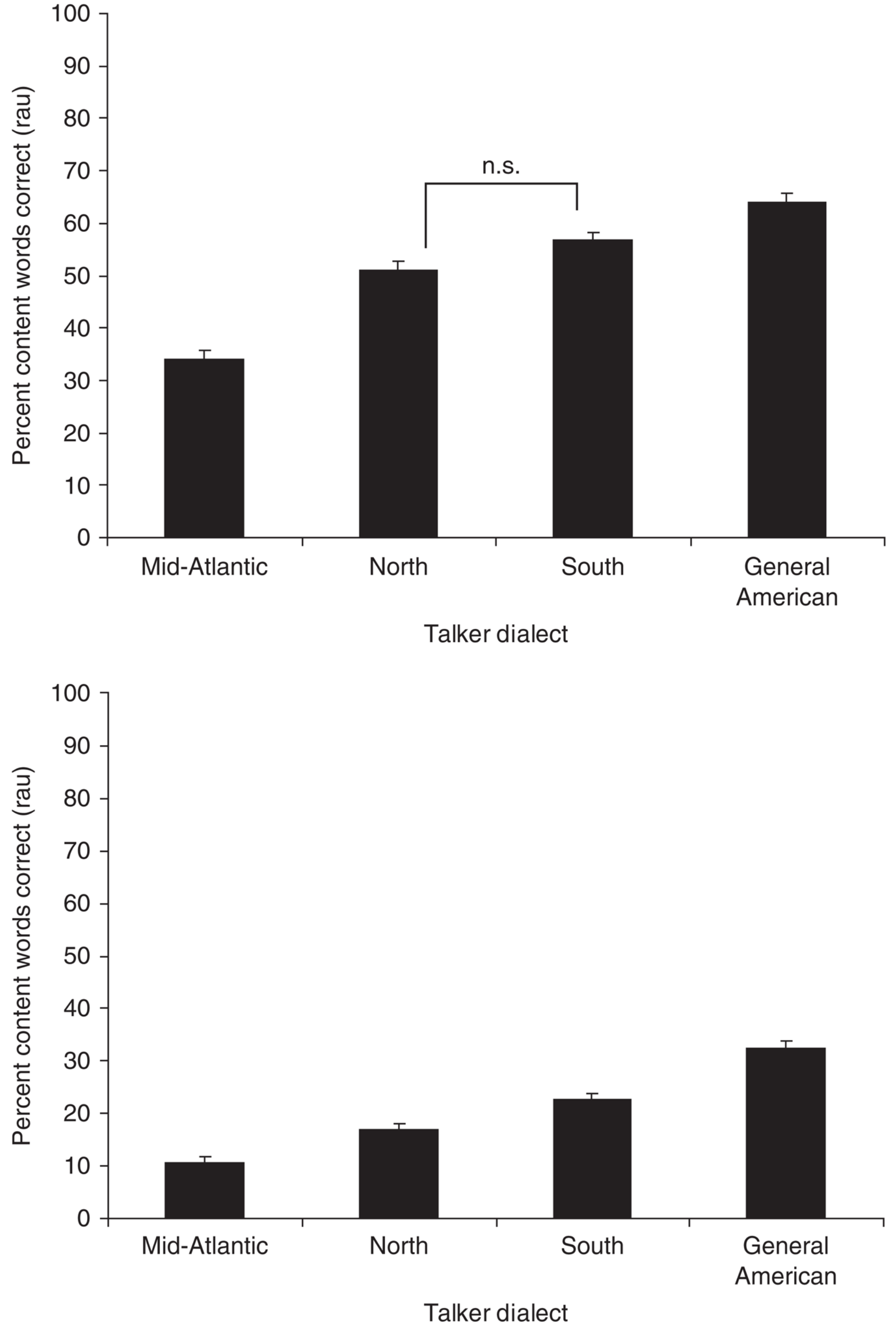

Figure 1.

Mean percent correct target words in rau for each talker dialect at −2 dB SNR (top) and −6 dB SNR (bottom) in Experiment 1. Error bars indicate one standard error. Non-significant pairwise comparisons have been indicated by brackets. All other pairwise comparisons are significant (all p <.02)

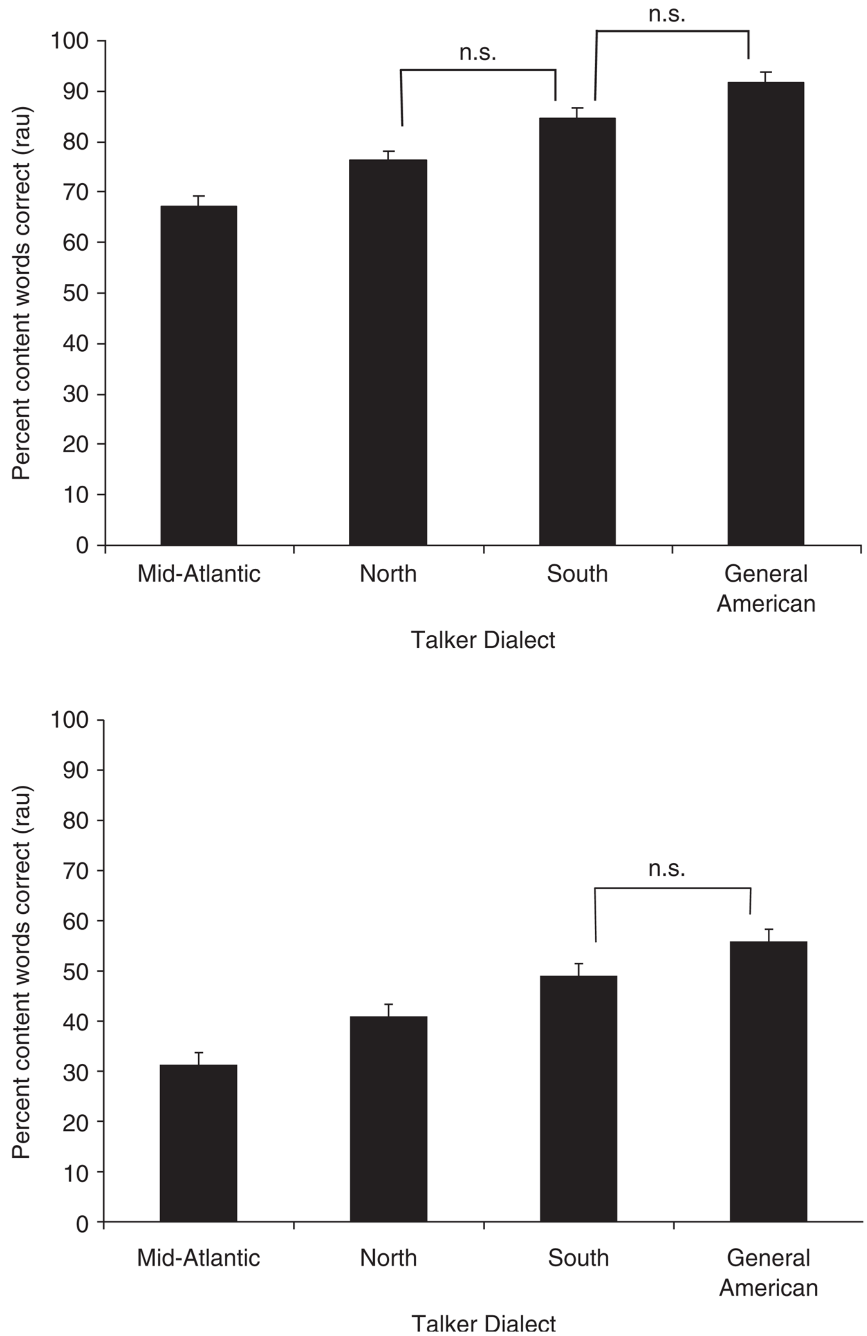

Figure 2.
Mean percent correct target words in rau for each talker dialect at +2 dB SNR (top) and −2 dB SNR (bottom) in Experiment 2. Error bars indicate one standard error. Non-significant pairwise comparisons have been indicated by brackets. All other pairwise comparisons are significant (all p <.03)

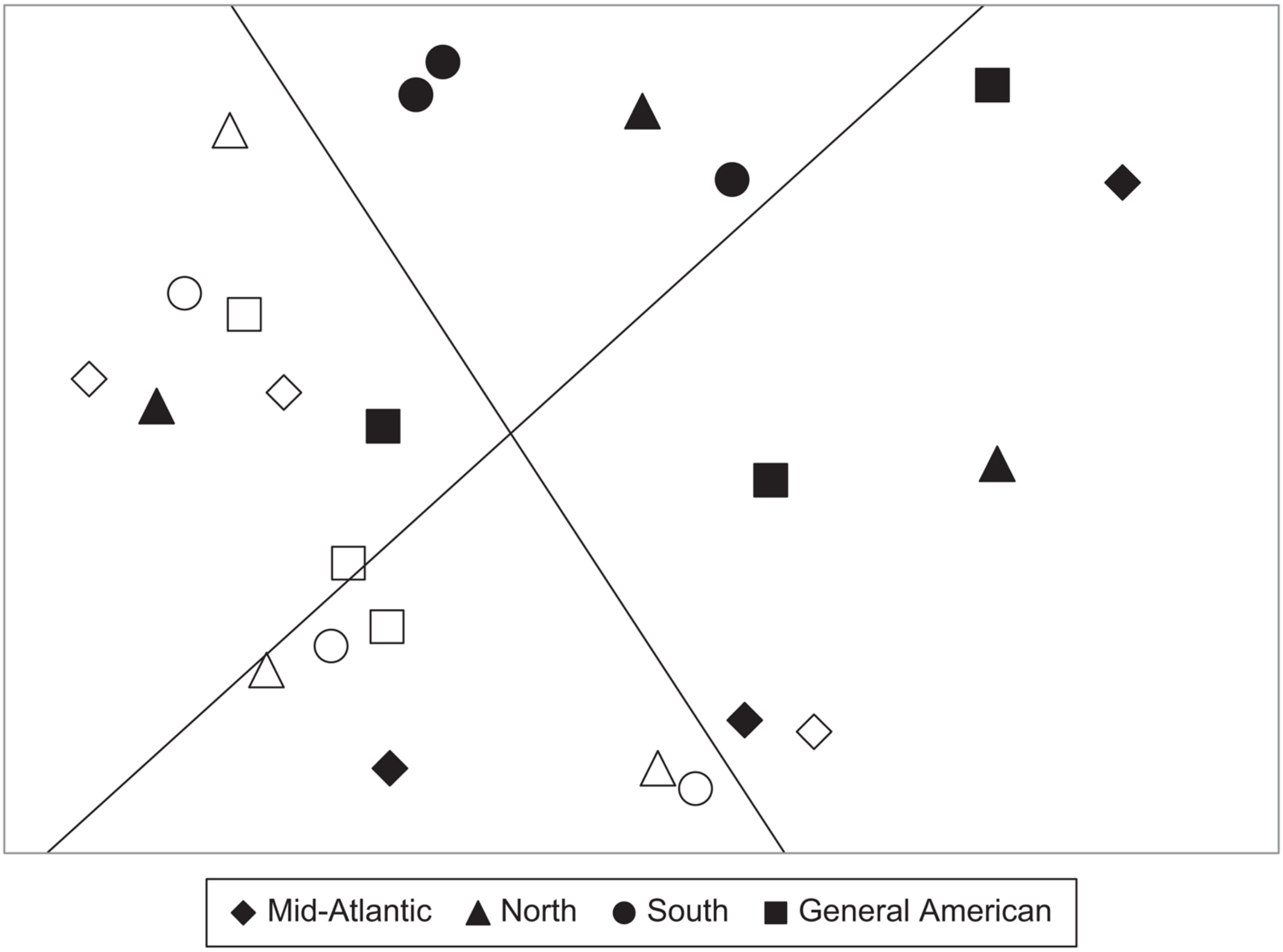

Figure 3.
Two-dimensional MDS solution for the free classification data obtained in Experiment 3. Filled symbols represent male talkers and open symbols represent female talkers

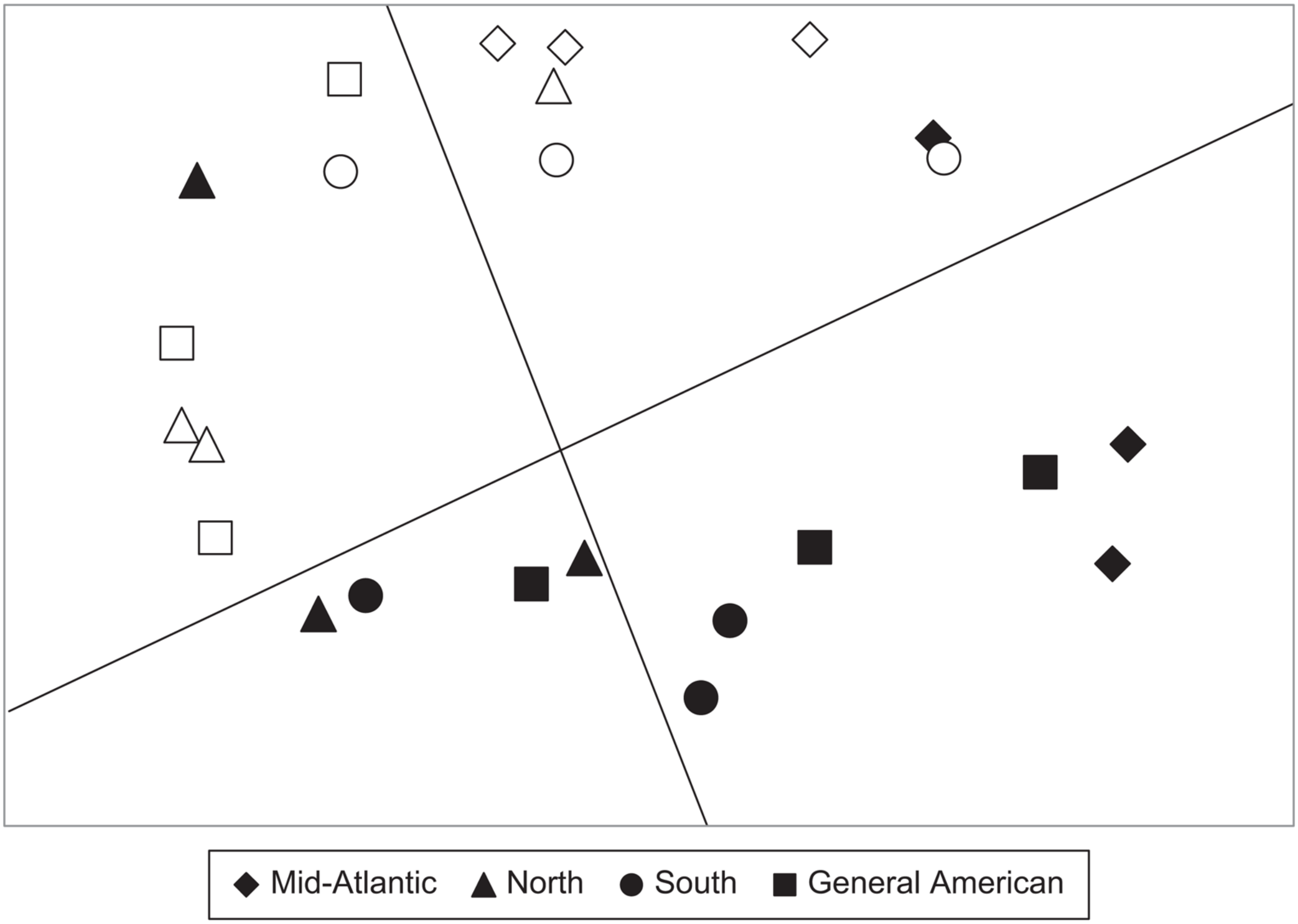

Figure 4.
Two-dimensional MDS solution for the free classification data obtained in Experiment 4. Filled symbols represent male talkers and open symbols represent female talkers

Table 1

Mean percent correct target words in rau for the Mid-Atlantic (AT), North (NO), South (SO), and General American (GA) talker dialects at each SNR for each listener group in Experiment 1. Standard deviations are shown in parentheses

	−2 dB SNR				−6 dB SNR			
	AT	**NO**	**SO**	**GA**	**AT**	**NO**	**SO**	**GA**
North	37 (9)	53 (15)	57 (11)	65 (10)	10 (10)	21 (10)	21 (6)	34 (12)
General American	32 (11)	48 (15)	56 (12)	62 (9)	12 (8)	15 (9)	22 (6)	33 (9)
Mobile	33 (8)	51 (7)	56 (7)	65 (12)	10 (8)	15 (10)	24 (7)	31 (7)

Table 2

Mean percent correct target words in rau for the Mid-Atlantic (AT), North (NO), South (SO), and General American (GA) talker dialects at each SNR for each listener group in Experiment 2. Standard deviations are shown in parentheses

	+2 dB SNR				−2 dB SNR			
	AT	**NO**	**SO**	**GA**	**AT**	**NO**	**SO**	**GA**
North	67 (9)	76 (10)	85 (8)	92 (3)	31 (11)	41 (9)	49 (4)	56 (8)
General American	64 (14)	81 (7)	85 (9)	91 (7)	29 (11)	41 (12)	54 (13)	66 (11)
Mobile	65 (13)	84 (10)	85 (12)	88 (10)	37 (13)	59 (12)	68 (12)	69 (11)

NIH-PA Author Manuscript

NIH-PA Author Manuscript

Table 3

Mean percent correct target words in rau for the Mid-Atlantic (AT), North (NO), South (SO), and General American (GA) talker dialects at each SNR for the Mid-Atlantic (N = 8) and the General American (N = 12) listeners in Experiment 2. Standard deviations are shown in parentheses

	+2 dB SNR				−2 dB SNR			
	AT	**NO**	**SO**	**GA**	**AT**	**NO**	**SO**	**GA**
Mid-Atlantic	57 (19)	72 (10)	84 (8)	84 (14)	30 (7)	50 (10)	51 (10)	57 (11)
General American	64 (14)	81 (7)	85 (9)	91 (7)	29 (11)	41 (12)	54 (13)	66 (11)

Table 4

Summary of the listeners' classification behavior in Experiment 3

	Mean	Median	Minimum	Maximum
Groups	5.75	5.5	3	10
Talkers per group	4.47	4	1	19
% correct pairings	22	20	8	62
% errors	22	18	13	65

Table 5

Summary of the listeners' classification behavior in Experiment 4

	Mean	Median	Minimum	Maximum
Groups	7.00	7	4	10
Talkers per group	3.71	3	1	10
% correct pairings	17	17	7	30
% errors	15	14	7	25