Free-classification of American dialects in three conditions: natural, monotonized, and low-pass filtered speech

Undergraduate Research Thesis

Presented in partial fulfillment of the requirements for graduation

"with research distinction" in Linguistics in the undergraduate colleges of

The Ohio State University

By Erin Walpole

The Ohio State University

May 2017

Project Advisor:

Professor Cynthia Clopper, Department of Linguistics

Abstract

The dialects of American English have distinct features: these features include vowel shifts – the Northern Cities Chain Shift and the Southern Chain Shift (Labov, Ash, & Boberg 2006; Clopper, Pisoni, & deJong 2005) − and prosodic variation, including intonation and rhythm (Clopper & Smiljanic 2011, 2015). In the current study, I ran three conditions to test which prosodic cues listeners were using when classifying talkers by regional dialect. American English has six distinct dialects: Northern, Southern, Midland, Mid-Atlantic, Western, and New-England (Labov, Ash, & Boberg 2006). Participants listened to 60 talkers, 10 from each of the six regional American English dialects, and were asked to sort the talkers into groups by dialect using free-classification. All of the talkers read the same sentence, which was manipulated in two of the three conditions. The first condition left the talkers' voices natural and un-manipulated. The second condition monotonized all of the talkers' voices. The third condition ran all of the talkers' voices through a low-pass filter, which removed everything above 400 Hz. Results indicated that all participants, regardless of condition, made about 9 groups of talkers on average. Results also revealed effects of condition and talker dialect on accuracy. For the condition accuracy, the monotonized condition had the most accurate groupings, while the low-pass filtered condition had the least accurate groupings. For the talker dialect accuracy, the Western dialect had the most accurate groupings while the Southern dialect had the least accurate groupings. Multidimensional scaling (MDS) plots visualized the groupings made for each condition. In both the natural and monotonized condition, participants were using dialect and gender to sort talkers. In the low-pass filtered condition, participants were using gender and not dialect to sort talkers, and the MDS plot looked different from the other MDS plots indicating that intonation alone was not effective for dialect classification.

Introduction

Previous Work

When an utterance is received by a listener, the listener takes in information about the speaker. They are able to identify where the speaker is from, their age, gender, ethnicity, and background (Abercrombie 1967). The speaker's dialect is what delivers most of this information. Previous studies have looked at how listeners use the information that they hear from talkers to classify them by their dialect.

Van Bezooijen and Gooskens (1999) found that prosodic features only played a minor role in dialect identification. They conducted three experiments: one for British English and two for Dutch. In the Dutch experiments, they found that listeners were more insecure about classifying dialects of Dutch with fewer prosodic cues. In the English experiment, speech utterances with noticeable or marked prosodic differences were included along with utterances without marked differences. The prosodic special condition, which was the condition with marked prosodic differences, was added because for that condition, the utterances highlighted specific prosodic features for the different dialects of British English. The results revealed that the absence of intonation cues, such as pitch, negatively affected classification accuracy. However, removal of verbal cues in their experiment also resulted in significantly lower means for accuracy. They noted that when the verbal information was removed, the identification suffered. The identification suffered because, they hypothesized, verbal information contains many different cues, which are needed for identification.

Van Bezooijen and Gooskens (1999) also found that more linguistic information did not lead to more accurate identification. On the contrary, for some dialects of Dutch, listeners were able to make better distinctions with only pronunciation information, than with all of the

available linguistic information. They did find that, for British English, the participants were still able to classify dialects significantly above chance on the basis of prosodic cues alone. However, the accuracy that the participants had for the integral version (all cues present), and the verbal version (monotonized but had all of the verbal information) was better in terms of identification and classification than the prosodic versions. That result indicates that listeners could identify and classify talkers using prosodic cues, but not as well as they could when they had all of the cues, and all of the verbal information.

Clopper and Pisoni (2004b) conducted a study, which looked at how listeners could categorize talkers by regional dialect of American English based on vowel variation. They had two experiments, which used the six regional dialects of American English. They also used forced-choice classification. In experiment one, they conducted an acoustic analysis, which revealed the phonetic features in the stimulus materials that distinguish the dialects of American English. In experiment two, they found that listeners were able to categorize the talkers into three broad groups (New England, South, and North/West). They also found that the labels that they had provided for the forced-choice classification task might have led to response biases on the listeners' part − biases that would not have been present in a free-classification task. The biases, they believed, might suggest that dialect or accent awareness may also play a role in dialect classification. They showed that listeners could complete the task, but they only studied the effects of vowel variation on categorization, not prosody.

Clopper and Pisoni (2007) used free-classification as a means of sorting talkers of American English. They also had two experiments, both of which involved free-classification of American English talkers and used six regional dialects of American English. The two experiments differed in gender − experiment one was only male talkers, while experiment two

included male and female talkers. They found that listeners were able to make distinctions based on the vowel differences for each of the American English dialects. For all of their experiments, gender emerged as a dimension. They noted that that result was interesting because the participants were asked to ignore the talkers' gender, and yet it still managed to arise. Clopper and Pisoni (2007) noted that result occurred because the listeners seemed to be sensitive to the interaction between gender and regional dialect in speech perception. They also noted that gender, being a more salient talker feature, simply could not be ignored. The current study investigated free-classification of American English with a focus on the role of prosodic cues in classification.

Previous work on prosodic variation, such as Munro et al.'s (2010) study, suggest that speaking rate differences can be important for differentiating dialects, and native-ness versus nonnative-ness. In their study, they found that listeners could still differentiate between native and nonnative speakers, even when the speaking rate differences were removed. Bent et al. (2016) argued that that differentiation could be evidence that other prosodic features, such as pitch, rhythm, and voice quality, could all play a role. Thus, a range of prosodic features may contribute to dialect classification.

Akbik, Atagi, and Bent (2013) found that the Southern dialect, having the most distinct perceptual features, was the most frequently identified American dialect. There is a myth that Southern talkers speak at a slower rate than other dialects of American English. Clopper and Smiljanic (2011) commented on that myth and they noted that Southern male speakers did not have slower speech, but rather they had more pauses than speakers from other dialects. That result revealed that the regional dialects of American English, specifically the Southern dialect in that case, vary in speech rate and pausing.

Sundara and Vicenik (2013) used free-classification to allow listeners to sort different talkers into groups based on their differing dialects and language. Their study used talkers from Australian English, American English, and German. They found that the listeners used prosodic cues, such as intonation and rhythm, of each dialect/language to make their classifications. They also found that free-classification allowed for more freedom in labeling, while forced-choice labels were less likely to represent the perceptions of non-linguists.

Bent, Atagi, Akbik, and Bonifield (2016) had listeners complete two tasks, including a free-classification task, where they asked the listeners to group talkers based on the talkers' perceived region of origin, and the ladder task, where they had listeners group talkers based on the talkers' perceived distance from standard American English. Their study included six American English dialects, six international native dialects, and twelve international non-native dialects. They found that native American English listeners were able to make distinctions between the native and non-native accents. They also found that the listeners were able to use various cues to sort the talkers into groups in the free-classification task, including their distance from the local standard, acoustic-phonetic characteristics, and speaking rate. Their ladder task also revealed that the participants could identify the six regional dialects of American English as closer to standard American English, but they did find the Southern dialect was perceived as the most distant from standard American English. In the ladder task, they found that the faster speaking talkers were perceived as more native than the slower speaking talkers and that that characteristic was used to differentiate between native and non-native speakers. The present study investigated similar concepts, but only for American English.

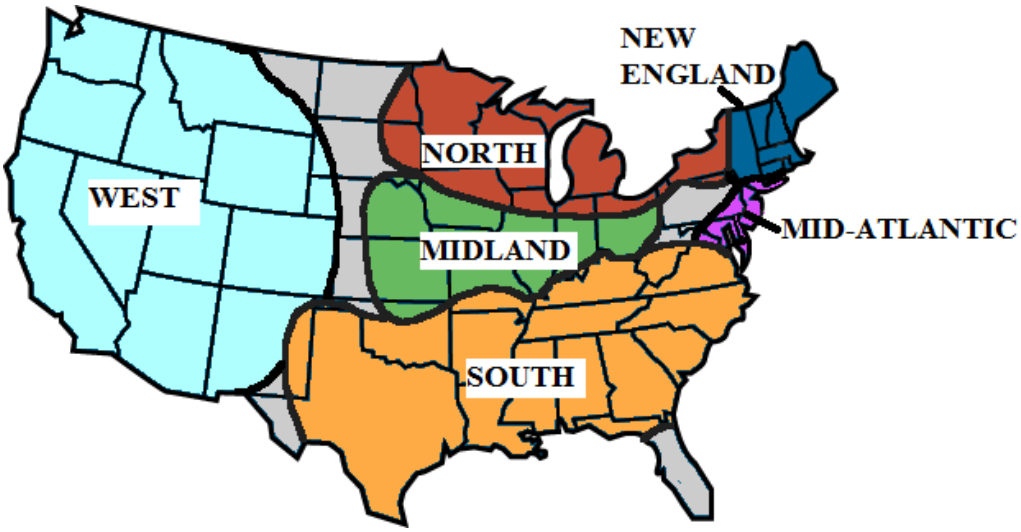

Figure 1. Regional Dialects of American English (based on Labov et al.'s 2006 Atlas of North American English).

In this study, the regional dialects of American English were analyzed (Northern, Southern, Mid-Atlantic, West, New-England, and Midland). The map in Figure 1 shows the dialect regions in the United States. Each of the regional dialects of American English have distinct vowel system differences. The Northern Cities Chain Shift, as shown in Figure 2, affects the Northern dialect. Vowels of the Northern dialect are raised and fronted, such as /ae/, and lowered, such as /ɑ/. The vowel /ɔ/ is lowered, and /ɛ/ and /ʌ/ are both backed. In some instances /ɪ/ was reported to have been backed similar to /ɛ/ (Labov et al. 2006; Labov 1998).

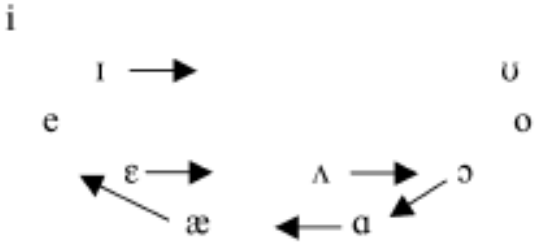

Figure 2. The Northern Cities Vowel Shift (Labov 1998).

The Southern Chain Shift, shown in Figure 3, affects the Southern dialect of American English. The vowels /u/ and /o/ are fronted in this shift. The vowels /ɪ/ and /ɛ/ are raised and the vowels /i/ and /e/ are lowered. The Southern dialect also has shown monphthongization of the diphthongs /ɑi/ and /oi/ (Thomas 2001).

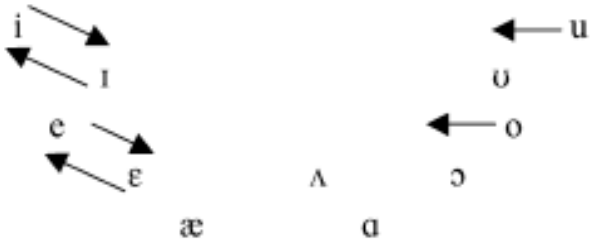

Figure 3. The Southern Chain Shift (Labov 1998).

The New England dialect has diphthongs /ɑi/ and /ɑw/ which are raised in the Eastern New England dialect (Thomas 2001) and the vowels /ae/ and /ɑ/ are fronted in the Western New England dialect (Boberg 2001; Thomas 2001). The Western New England dialect also has evidence of /ɛ/ backing (Boberg 2001; Thomas 2001).

The Western dialect is characterized by the fronting of the vowel /u/ and the low-back merger of /ɑ/ and /ɔ/ as seen in the “third dialect” (Labov et al. 2006; Thomas 2001). In the “third dialect” merger, words such as “caught” and “cot” become homophones (Labov 1998). The Mid-Atlantic dialect raises the vowel /ɔ/. It also raises /ae/ in some words, and not others due to the history of the Mid-Atlantic dialect’s contrast between long and short /ae/ (Labov 1994; Thomas 2001). The Midland dialect is considered to be the least marked and only has the low-back merger of /ɑ/ and /ɔ/, as seen in the “third dialect” (Labov 1994; Thomas 2001).

Research Question

For this study, I investigated the extent to which listeners can use prosodic cues to sort talkers by regional dialect of American English, and which cues listeners take into account when they sort talkers into groups based on their dialect. I also wanted to assess how similar the groupings were for each of the conditions (using multidimensional scaling), and how accurate the groupings were by dialect and by condition.

There is evidence that American listeners can group and distinguish dialects/languages using just prosodic cues (Sundara & Vicenik 2013). That distinction is possible because American English dialects show differences in prosody (Clopper & Smiljanic 2015). In the present study, talkers' voices were put through a low-pass filter and monotonized. Low-pass filtering deletes everything above 400 Hz, which made the speech unintelligible but left pitch information, specifically intonation, intact. Monotonizing takes the mean pitch of each speaker and fixes the sentence at that pitch, which removed the intonation that is normally present in speech (Van Bezooijen & Gooskens 1999). I was interested in seeing which of the three conditions (natural, monotonized, and low-pass filtered) had the most accurate groupings. I expected that the natural condition would have the most accurate groupings because all of the cues (vocalic and prosodic) were present. I also expected that the modified talkers' voices would be more difficult to classify, although I expected that accurate groupings would still be possible with the low-pass filtered speech and the monotonized speech, as was seen in Van Bezooijen and Gooskens' (1999) experiments with Dutch and British listeners. However, if the monotonized condition was grouped most accurately or similarly to the natural condition, it would indicate that intonation is, in fact, not an important prosodic cue and it may even be disruptive. If the low-pass filtered condition was grouped the most accurately, then intonation (a prosodic cue) was

used successfully to identify and sort talkers by dialect, suggesting that it is an important cue for classification of American English dialects.

Methods

Talkers

Our study used 60 different talkers, all from the Nationwide Speech Project Corpus (Clopper & Pisoni 2006). The talkers were all between the ages of 20-29 years old when they were recorded, and white. They were from six different dialect regions: New England, Mid-Atlantic, North, Midland, South, and West. Five male talkers and five female talkers were from each of the six regional dialects of American English. The six regions were chosen based on Labov et al.'s (2006) dialect classifications, as shown in Figure 1. Each of the talkers lived in their dialect region exclusively until the age of 18 and both parents were also from the same dialect region. The talkers were primarily undergraduate students from Indiana University, though five had completed their bachelor's degrees, and one had completed a master's degree.

Participants

The participants for this study were LOC (Linguistics Outside the Classroom) students from The Ohio State University. The mean age of the participants for all of the conditions was 20 years with 18 years as the youngest age and 50 years as the oldest. Each condition of the study had at least twenty usable participants. The natural condition included twenty participants, the monotonized condition included twenty-one participants, and the low-pass filtered condition included twenty-three participants. The participants needed to have no speech/hearing deficits, and needed to be native speakers of American English, because this study is about American

English dialects. Previous residential history for each participant was collected. The age and gender of each participant were also collected. For this thesis, I did not take participants' information into account, but future analyses can be conducted to analyze those demographic factors. In total, data from one hundred and four participants were collected and sixty-four participants met the requirements for this study. Among the excluded participants, twenty-eight reported that they were not native speakers of American English, three reported that they had speech therapy, five adjusted the volume for the low-pass filtered experiment, two were given the wrong instructions due to experimenter error, one reported that they had hearing loss, and one reported that they had a speech impairment.

Stimulus Materials

The experiment had three conditions. Each of the three conditions used the same 60 talkers from the NSP Corpus. The talkers were recorded reading the beginning of *Goldilocks and the Three Bears*, which included the sentence, "they lived in a cottage deep in the woods". This sentence was extracted from each talker's recording.

In the first condition, the talkers' voices were completely unaltered. The natural condition served as the baseline group. In the second condition, the talkers' voices were monotonized. The mean f0 of each speaker was taken and the sentence was fixed at that f0. When the speech was monotonized, intonation, an aspect of prosody, was removed (Van Bezooijen & Gooskens 1999). In the third condition, the low-pass filtered condition, the talkers' voices were put through a low-pass filter. The talkers' voices were filtered at 400 Hz. Anything that was above the 400 Hz filter was deleted. This filter made the speech unintelligible, but kept the prosody, including intonation and rhythm, of the speech intact (Van Bezooijen & Gooskens 1999).

Digital movies were created using iMovie to make the auditory recordings into visual representations. The present study had the visual track of each movie created with a single frame digital image, as was seen in Clopper's (2008) study. The initials of the original talkers' names were used for the movies' single frame digital image to give each file a trackable unique quality. However, the initials were in no way helpful to the participants in the experiment in terms of identification. The audio recordings were imported for the audio track and the digital image was imported for the video track (Clopper 2008). The digital images' duration needed to match the duration of the audio track in order to present the same image throughout the duration of the audio track. The movies were then exported as .avi files individually to be played back in PowerPoint on Microsoft Windows.

Procedure

The stimulus materials were presented to the participants using a single PowerPoint slide. All 60 of the talkers' movies were arranged in columns on the left-hand side of the slide, while a 22x22 grid was on the right-hand side of the slide. Figure 4 shows a blank free-classification display that was used for all three conditions. Figure 5 shows a participant's completed free-classification display for the monotonized condition showing 9 talker groups.

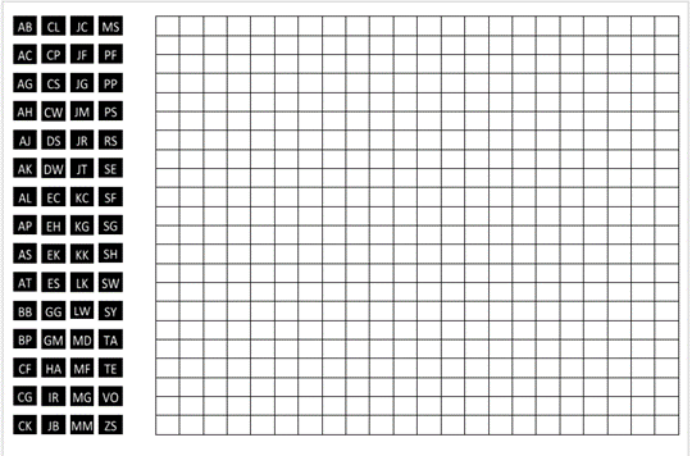

Figure 4. Free-classification display used for all three conditions.

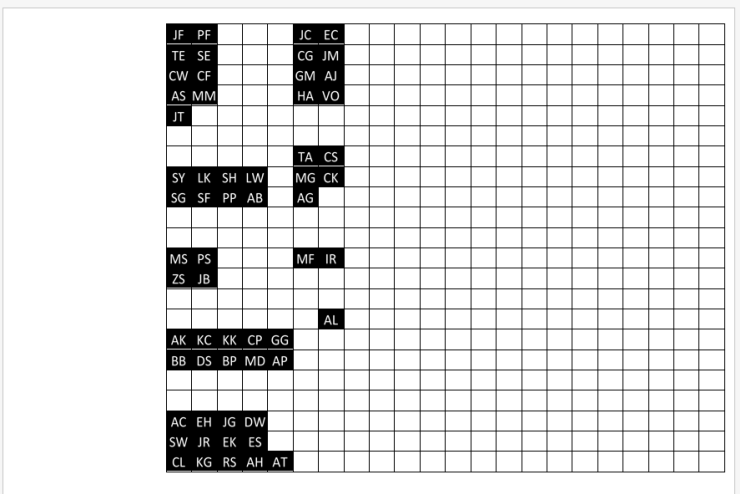

Figure 5. Example of a participant's completed free-classification display.

The movies matched the grid cells perfectly in size so that they could be neatly fit into place. To play the movies, participants needed to double-click each individual movie box. To move the movies from the columns to the grid, participants needed to click the movie once and the movies could then be dragged to the grid in any place that the participant desired. That process allowed participants to manipulate the location of the movies in the grid so that they would have each movie aligned properly. To be aligned properly, the cells needed to fit perfectly into the grid.

Participants were told that each square represented a different talkers' voice. Instructions on how to play and move each movie were given. The participants were asked to listen to each talker and sort all of the talkers into groups. The participants were told that they could have as many groups as they wanted, with as many talkers in those groups as they wanted, but with one stipulation: if the talkers sounded like they came from the same place, they needed to be sorted into the same group. These instructions were given to all of the participants regardless of their condition. For the monotonized and low-pass filtered conditions, the participants were warned that the talkers' voices might sound unusual, but they would still need to sort the talkers into groups based on regional background.

Analysis

A custom macro analyzed all of the usable files for all of the conditions and outputted three files for each condition – groupings, similarity, and debugging. The text file for the groupings kept track of how many groups each participant made, and which talkers they had put into those groups. An ANOVA was calculated to compare the average numbers of groups that were made by all participants for all conditions. The dependent variable was the number of groups made for each participant, while the independent variable was the condition for each participant.

The similarity output included a 60 by 60 talker matrix showing the similarity for all of the talkers who were grouped together for each participant. If the talkers were put into the same group together, then they were given a 1. If the talkers were not put into the same group together, then they were given a 0. The groupings were considered accurate if talkers from the same dialect were in the same group. The "hits", or the number of talkers grouped together from the

same dialect, were calculated out of all the possible same-dialect pairings. The “misses”, or the number of talkers that were grouped together from different dialects, were calculated out of all of the possible incorrect dialect pairings. To find the overall proportion correct, the misses of each participant were subtracted from the hits of each participant. This calculation was done for each of the talker dialects for each participant, and for all of the dialect groupings for each participant. The calculation is similar to the V-measure calculations by Rosenberg and Hirschburg (2007).

The proportions correct were then analyzed using an ANOVA to compare the mean proportion that the subjects had gotten correct for each talker dialect, as well as for each condition. The dialects of the talkers and condition were tested to see if they had a significant effect on the accuracy of the groupings that participants had made. A repeated measures ANOVA was calculated for the accuracy data. The repeated measures ANOVA allowed for each participant to contribute more than one value to the dependent variable. The proportion that each participant got correct served as the dependent variable, while the condition and talker dialect served as the independent variables. Condition was a between-subjects variable and talker dialect was a within-subject repeated measures variable in the analysis.

An MDS (multidimensional scaling) display was created for each condition, which showed, visually, the perceptual similarity of the dataset (Clopper 2008). The MDS analysis provided the best fit for the dataset and the number of dimensions that the participants used when they completed the task. In regards to MDS analyses, the visualization of similarity groupings for each of the conditions was important for showing the groupings that were made by listeners. The similarity matrixes were summed for each condition and an MDS analysis was run to determine the stress value for models with different numbers of dimensions for each condition. The stress values were used to determine how many dimensions were needed to maximize the

interpretability of the MDS plots. A stress plot was created for each condition to identify where the “elbow” was for the dataset. The “elbow” reveals the number of dimensions that maximize the fit of the data while minimizing the number of model parameters (Cox & Cox 2001). With more dimensions, the badness of fit (or stress) of the dataset is reduced and the elbow indicates where model improvement is offset by an increase in parameters (Clopper 2008). Four dimensions were considered and their stress values were compared for each condition. The current dataset only required two dimensions to be modeled because that is where the stress plots indicated that the elbow was for each condition.

Results

Groupings

Figure 6 shows the number of groups that were made for each of the three conditions. In the natural condition, participants made an average of nine groups with a standard deviation of 5. In the monotonized condition, participants made an average of eight groups with a standard deviation of 4. In the low-pass filtered condition participants made an average of nine groups with a standard deviation of 5. The ANOVA on groupings did not reveal a significant effect of condition ([$F(2,61)=.77$, n.s.]). This result means that the groupings made for each condition were not significantly different from one another. That is, the participants in all of the conditions made about the same number of groups.

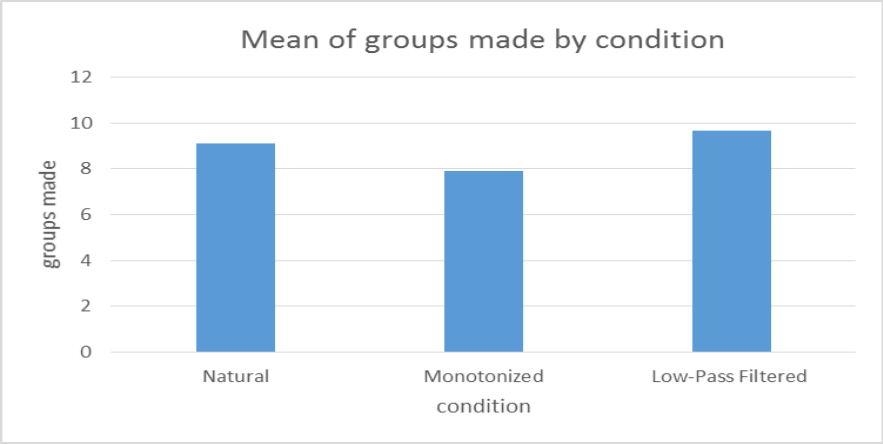

Figure 6. Mean number of groups made for the three conditions (natural, monotonized, low-pass filtered).

Accuracy

Figure 7 shows the proportion correct by condition. The monotonized condition had the most accurate groupings (with a mean of 0.09 and standard deviation of 0.24), while the low-pass filtered condition had the least accurate groupings (with a mean of -0.01 and standard deviation of 0.07). The natural condition had intermediate accuracy (with a mean of 0.04 and standard deviation of 0.14). The repeated measures ANOVA on accuracy revealed a significant main effect of condition ([$F(2,61)=3.30$, $p=.044$]). This result indicates that the condition had a significant impact on the accuracy of groupings that were made, including more accurate performance by participants in the monotonized condition and less accurate performance by participants in the low-pass filtered condition.

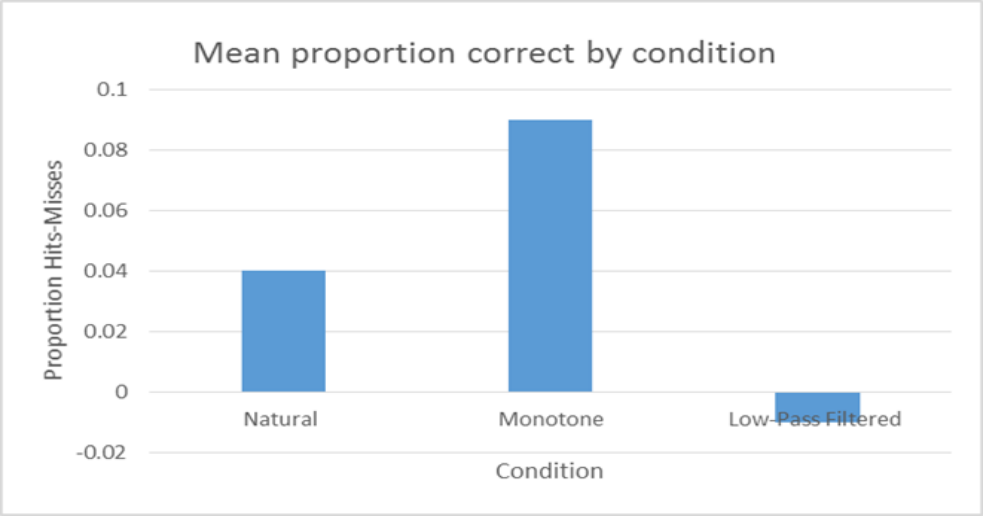

Figure 7. Mean proportion correct by condition.

Figure 8 shows the mean proportions that participants got correct by dialect. The Western dialect was grouped the most accurately with a mean proportion of 0.07 (SD=0.21). The North and Midland dialects were grouped similarly with the same mean proportion of 0.05 (SD=0.16 for North, SD=0.19 for Midland). The Mid-Atlantic dialect had a mean proportion of 0.03 (SD=0.14). The New England dialect had a mean proportion of 0.02 (SD=0.12). The Southern dialect's mean proportion was 0 (SD=0.16), indicating that the difference between the proportion of hits and the proportion of misses for the Southern dialect was the same. The repeated measures ANOVA also found that there was a main effect of talker dialect on accuracy ([F(5,305)=3.07, p=.010]). This result indicates that talker dialect had an impact on the accuracy of the groupings made based on talker dialect as shown in Figure 8. The interaction between condition and dialect was not significant ([F(10,305)=1.238, n.s.)]. This result indicates that while dialect and condition did independently have significant effects, the effect of dialect was not affected by condition or vice versa.

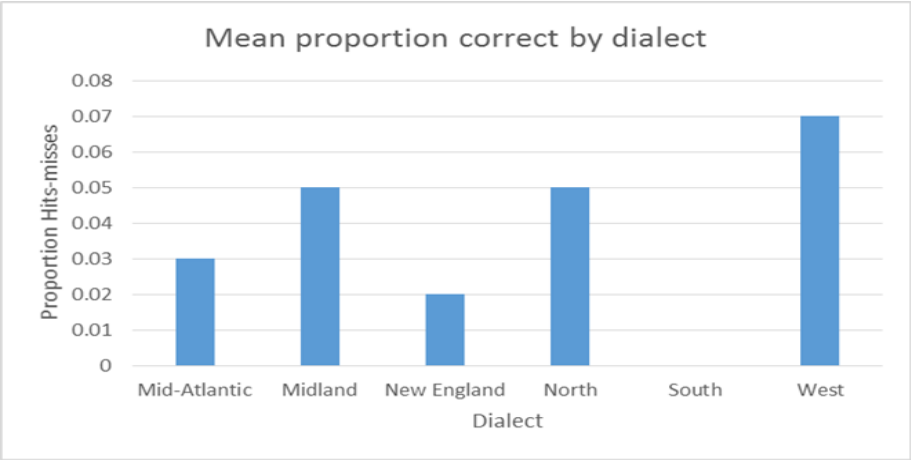

Figure 8. Mean proportion correct by dialect

Talker Similarity

Multidimensional scaling plots were created for the talker similarities within each of the conditions (natural, monotonized, and low-pass filtered) as shown in Figures 9, 10, and 11. The symbols represent the talkers from each of the dialects (North, South, Midland, New England, Mid-Atlantic, or West), as well as each of the talkers' genders (male or female). Each of the dialects was assigned a color to help differentiate it from the other dialects. The colors for each dialect are the same colors from Figure 1's dialect map. The gender of each of the talkers is shown by their shape. If the talker's shape is a triangle, then the talker was male. If the talker's shape is a circle, then the talker was female. Each condition's multidimensional scaling plot has two dimensions, which was determined by a stress plot.

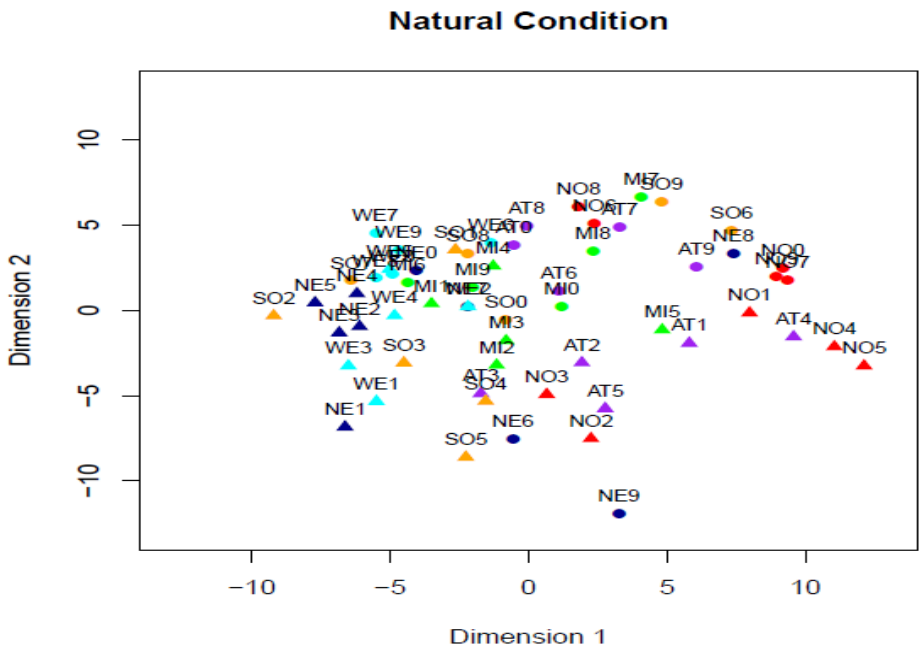

Figure 9. MDS plot of the natural condition.

Figure 9 shows the multidimensional scaling plot for the natural condition. Dimension one for the natural condition was the talker dialects. In red, the Northern dialect (NO) can be seen in a cluster above 10 on dimension one. The Mid-Atlantic dialect (AT) can be seen in purple in the middle of the plot. The Midland dialect (MI) can be seen in green also in the middle of the chart. The Southern dialect (SO) can be seen in orange to the left of the Midland dialect, and at approximately -2.5 on dimension one. The Western dialect (WE) can be seen in the light blue at about -5 on dimension one. Last, the New England dialect (NE) can be seen in the dark blue at about -7 on dimension one. Dimension two was gender. The majority of the male talkers (triangles) were grouped below 0 on dimension two, while the female talkers (circles) were grouped above 0 on dimension two.

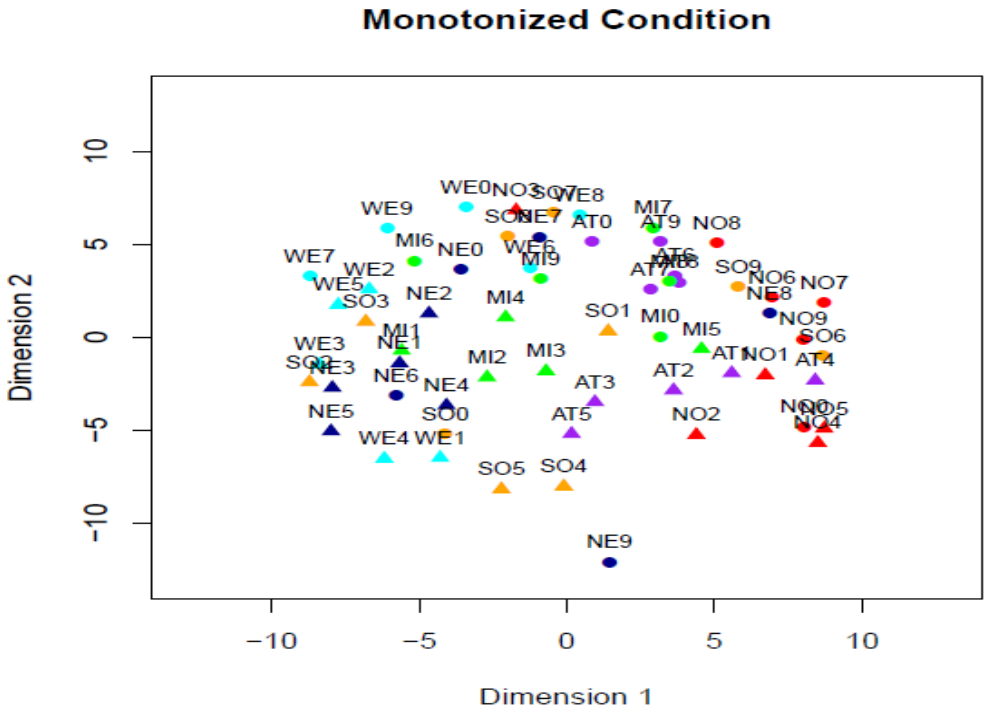

Figure 10. MDS plot for the monotonized condition.

Figure 10 shows the multidimensional scaling plot for the monotonized condition. Dimensions one and two for the MDS for the monotonized condition were the same as in the natural condition. The dialect groupings were similar, though slightly different, to the natural condition. In both the natural condition's MDS plot and the monotonized condition's MDS plot, the Northern talkers were above 10 on dimension one and between the -5 and 5 on dimension two. The other dialects were, for the most part, grouped in similar positions for both the natural and monotonized condition's MDS plots. The Western dialect seemed to have been grouped in a slightly tighter cluster in the natural condition's MDS plot. The Western talkers spread as far as -10 on dimension one for the monotonized condition, but only spread as far as about -8 on dimension one for the natural condition. The genders were also grouped similarly. The females were predominantly on the upper half of the plot (for dimension two) and the males were

predominately on the lower half of the plot. The similar groupings indicate that the prosodic cue of intonation is not an important cue for dialect classification. Listeners were still able, despite the monotonization, to group talkers by dialect and gender, which had emerged as the second dimension of similarity. This plot suggests that, while intonation was not used to group talkers by dialect, another cue, such as vowel information, could have helped listeners group the talkers. Clopper and Pisoni (2007) had similar findings, when they found that listeners were specifically using vowels to group talkers. Clopper and Pisoni (2007) isolated the vowels of the talkers by running acoustic analyses to pick out distinct features for each of the dialects. The chosen sentences needed to contain dialect-specific vowel shifts. Distinct features, such as the Northern dialect's shifted /ae/ needed to be present for the Northern talkers, while the monophthongization of the /ɑi/ diphthong needed to be present in the speech of the Southern talkers.

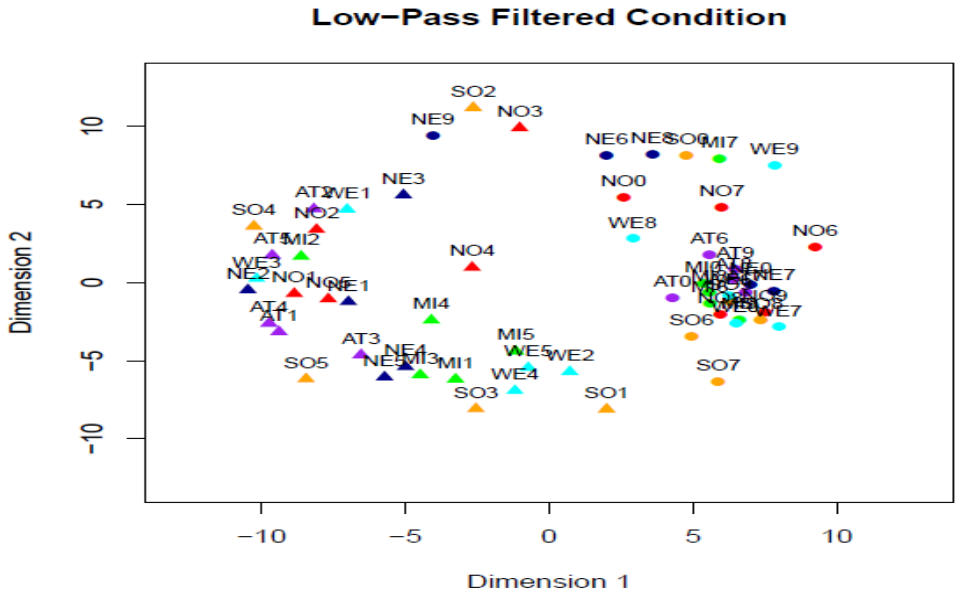

Figure 11. MDS plot for the low-pass filtered condition.

Figure 11 shows the multidimensional scaling plot for the low-pass filtered condition. The low-pass filtered MDS had different dimensions than the MDS plots for the natural condition and the monotonized condition. Dimension one was gender, rather than dialect. The female talkers were grouped predominantly on the right hand side of the plot, while the males were grouped predominantly on the left hand side of the plot. Dimension two, however, was not dialect because the dialects were not grouped together. For instance, the Western dialect can be seen in various sections on the plot. They are seen at the bottom of the plot at around -5 on dimension two, and also around 0, and 5. All of the dialects follow this pattern of not being grouped together. Lack of a dimension corresponding to dialect indicates that prosodic cues are not sufficient for dialect classification in American English. Unlike the natural condition and the monotonized condition, in the low-pass filtered condition, the participants did not have access to segmental information. The plot looked different than the natural and monotonized condition plots and the dialects were not effectively grouped together.

Discussion

Summary of the Findings

Similar to Clopper and Pisoni's (2007) study, the participants showed that they were able to make fine-grained distinctions when they were creating regional dialect groups. Free-classification was used to allow the participants to make as many distinctions as they liked. Despite the freedom that free-classification allowed, across all three conditions participants made about the same number of groups. The average number of groups made for the natural condition and the low-pass filtered condition was nine groups. The average number of groups made for the monotonized condition was about eight groups. An ANOVA found that the groupings for the

three conditions were not significantly different. Similarly, in Clopper and Pisoni's (2007) free-classification study, participants made about ten groups for experiment one and about eight groups for experiment two.

The multidimensional scaling plots for each of the conditions showed the dimensions that listeners were using to sort the talkers. The analysis of each condition revealed two dimensions of talker similarity. When the higher frequency information was removed, by deleting everything above 400 Hz, participants were unable to sort the talkers by dialect, but they were able to use gender to sort the talkers. In the low-pass filtered condition, the dialects were not clustered together, which gave the indication that dialects were not being classified when only prosodic cues were available to them. Further, because of the similarity between the natural condition and the monotonized condition, intonation, which had been removed, was shown to not be a necessary prosodic cue for the sorting task because the dialects could still be classified and sorted in the monotonized condition. This finding suggests that segmental cues were being used to classify talkers by dialect, as was seen in Clopper and Pisoni's (2007) experiment. Similarly, in Van Bezooijen and Gooskens' (1999) study, they found, for both the Dutch and British experiments, the verbal features of speech seemed to have more origin-identifying cues than prosodic features had. While the dialect groupings were not accurate, it would be interesting to try to determine which cues the listeners were using to group the talkers.

In the natural condition and the monotonized condition, the participants were able to sort talkers by their gender, but also were able to classify the talkers by dialect. Gender probably emerged as one of the dimensions because it simply cannot be ignored. In Clopper and Pisoni's (2007) study, it was shown that gender emerged despite the fact that listeners were told to ignore gender when they were making their groups. Labov (1990) said that women may produce more of some

phonological variants than men, or (due to stigma) might produce fewer of particular variants than men, but overall women tend to lead phonological change. Clopper and Pisoni (2007) noted, too, that the emergence of gender might also account for listeners' sensitivity to the interaction between gender and dialect in production. There were significant effects of condition and talker dialect on accuracy. Using the proportion correct for each participant, I found that the monotonized condition was grouped the most accurately. Talkers from the same dialect were grouped with other talkers from the same dialect the most in the monotonized condition. The natural condition's groups were less accurate than the monotonized condition, and the low-pass filtered condition was the least accurate of all of the conditions. This result was surprising because I had expected that the natural condition, which had all of the vowel information and prosodic cues, would have the most accurate groupings. I hypothesize that this result happened because the monotonization isolated the vowel cues (by removing the intonational cues) and made it easier for participants to make distinctions by dialect. This finding could be backed by Clopper and Pisoni (2007), who found that vowel cues were used to effectively make distinctions between dialects of American English.

For the talker dialect effect on accuracy, I also used the "hits-misses" method to find the proportion that each participant had gotten correct for each dialect of American English. The Western dialect was the most accurately grouped, while the Southern dialect was the least accurately grouped. This result was also surprising because the Northern and Southern dialects both have qualities which distinguish them from the other dialects − including their vowel shifts and prosodic variation (Labov 1998; Labov et al. 2006). I had expected that the Southern and Northern groups would have the most accurate groupings.

Clopper, Pisoni, and deJong (2005) found that the New England males and females, Midland females, and the Western males and females in the NSP corpus share the low back merger (/ɑ/ and /ɔ/) in words such as “frogs/logs” and “hod”. They found that the Western dialect had /u/ fronting as well. The fronting of /u/ was also seen in the Midland and Mid-Atlantic dialects. The shared features that the Western dialect has with the other dialects might account for why the Western dialect talkers were classified better than the other six dialects. The Western dialect, from the results of Clopper, Pisoni, and deJong’s (2005) study, seems to have the most distinctive features, including features from the Northern and Southern vowel shifts, and the “third dialect” merger. However, if the listeners were relying on segmental information to classify the Western dialect, it would not explain why the Southern dialect, having had its own distinct chain shift, would have been grouped less accurately. In addition, the Midland and Northern dialects had less accurate groupings than the Western dialect as well. The Midland dialect, which shares the low back merger and /u/ fronting with the Western dialect, should have been grouped just as accurately, if not more accurately (given the area where the study took place).

Clopper, Pisoni, and deJong (2005) also found that the NSP corpus talkers from the Midland dialect shared features with the Southern dialect’s vowel shift. Both the Midland and Southern dialect showed /u/ fronting. Some Midland talkers also shared features of /ae/ fronting, as was seen in the Northern dialect. The shared features with the Northern and Southern dialects might have made the Midland dialect more difficult to classify for participants. In turn, the shared features might have made the “distinct” features of the Northern and Southern dialects less distinct to participants, which also might have been why they had less accurate groupings for those dialects.

Clopper and Smiljanic (2011) compared the prosodic features of the Midland and Southern dialects. They found that the Southern talkers' pauses, which were longer and more frequent than the pauses in Midland American English, might have contributed to the stereotype of slowness of speech. Clopper and Smiljanic (2015) found that Southern talkers also had longer vowels on average, in comparison to the Northern dialect, which might have also contributed to the speech differences that listeners had perceived. In addition, Clopper and Smiljanic (2011) noted that the Midland dialect and Southern dialect had some similar pitch accents, including the three-way contrast between H*, L+H* and L*+H. They found that H- phrase accents were preferred by Southern female talkers more so than Midland talkers, which shows a difference in prosody between the dialects. They found that female talkers for all dialects had a preference for the H* pitch accent in the *Goldilocks* passage. Analysis could be done to see if pitch accent differences were relevant in this study as well by comparing an analysis of the talkers' pitch accent patterns to the perceptual results.

Clopper and Smiljanic (2015) noted that the Southern dialect and Midland dialect did not strongly differ in prosody or segmental properties, such as the fronting of back vowels. Their results indicated that the Midland dialect, in a sense, acts as an intermediate dialect (between the Northern dialect and the Southern dialect). It is interesting that the Midland dialect, which is so prosodically and segmentally similar to the Southern dialect would have better performance in the current study than the Southern dialect. In Figure 8, which showed the average proportions correct for each dialect, it can be seen that the Midland dialect had the same proportion correct (on average overall) as the Northern dialect (with a mean of 0.05). The shared features that the Midland dialect has with the Western dialect could be the culprit for its more accurate groupings. The Western dialect, which had the most accurate groupings, also shares features with the

Midland dialect including the /ɑ/~/ɔ/ merger. Further analysis can be done to analyze what specific features the talkers had, and a comparison can be done to see why some dialects had more accurate groupings than other dialects.

Mobility of the participants could have also played a role in the classification process. More mobile participants would have been more exposed to other dialects and could, therefore, potentially classify various dialects more accurately. Clopper and Pisoni (2006) conducted a study which analyzed listeners' groupings that they had made using forced-choice classification. They found that less mobile participants were unable to make distinctions between dialects that more mobile participants could make. In particular, they saw that non-mobile Northern listeners perceived the Midland and Northern dialect as more similar to one another than other listeners. Mobility of the participants in this study might have played an integral role in their ability to make accurate classifications. Further analysis is needed to interpret if that could have occurred in this study. In addition, Labov and Ash (1998) found that listeners from Birmingham could better identify the Southern dialect shifted vowels than listeners from Chicago or Philadelphia. In general, the listeners from Birmingham would have had more experience with the Southern dialect and would, therefore, be able to better identify it. Further analysis is needed to look at where the listeners are from. Then, the dialect accuracy could be analyzed for each participant and the dialect which each participant is most familiar with could also be taken into account.

Implications

This research helped pave the way to better understanding the dialects of American English − particularly which prosodic cues listeners use when distinguishing and classifying people by dialect. Understanding such information is important for understanding speech

processing and dialect perception in general. Understanding perception is particularly important for the perception of dialects for computer systems. Perception could also account for why some dialects are preferred over others. Comprehension of dialects could also be better understood, such as why some dialects are comprehended better than others. More research can be done in the future to look at other dialects of American English, such as AAVE (African American Vernacular English), and what vowel and prosodic differences they have with the other American English dialects. More work could also be done to look at the individual identities of the participants from this experiment to see if their background (residential history, and mobility) had an effect on their perception and/or their classifications of the American English dialects that they heard. The consonant sounds could be removed to allow the focus to be on the vowels. The different dialect speaking rates could also be made the same to look at listeners' ability to sort dialects when that prosodic feature is removed.

References

Abercrombie, D. (1967). Elements of general phonetics. *Edinburgh University Press*, Edinburgh.

Akbik, A., Atagi, E., & Bent, T.(2013).Categorization of regional, international, and nonnative accents. *Journal of the Acoustical Society of America*, 133.5, 3566.

Bent, Tessa, et al. (2016). Classification of regional dialects, international dialects, and nonnative accents. *Journal of Phonetics*, 58, 104-117.

Bezooijen, R. Van, and C. Gooskens. (1999). Identification of Language Varieties: The Contribution of Different Linguistic Levels. *Journal of Language and Social Psychology*, 18.1,31-48.

Boberg, C. (2001). The phonological status of Western New England. *American Speech*, 76, 3–29.

Cox, Trevor F., and Michael A. A. Cox. (2001). Multidimensional Scaling. *Boca Raton: Chapman & Hall/CRC*. Print.

Clopper, Cynthia G. (2008). Auditory Free Classification: Methods and Analysis. *Behavior Research Methods*, 40.2, 575-81.

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

Clopper, Cynthia G., and David Pisoni B. "Free Classification of Regional Dialects of American English." *Journal of Phonetics*, 35.3 (2007): 421-38.

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

Clopper, C. G., Pisoni, D. B., & deJong, K. (2005). Acoustic characteristics of the vowel systems of six regional varieties of American English. *Journal of the Acoustical Society of America*, 118, 1661-1676.

Clopper, Cynthia G., and Rajka Smiljanic. (2011). Effects of Gender and Regional Dialect on Prosodic Patterns in American English. *Journal of Phonetics*, 39.2, 237-45.

Clopper, Cynthia G., and Rajka Smiljanic. (2015). Regional Variation in Temporal Organization in American English. *Journal of Phonetics*, 49, 1-15.

Labov, W. (1990). The intersection of sex and social class in the course of linguistic change. *Language Variation and Change*, 2, 205-254.

Labov, W. (1994). Principles of linguistic change: Social factors. Malden, MA: *Blackwell*.

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

Labov, W., Ash, S., & Boberg, C. (2006). Atlas of North American English. *Mouton de Gruyter*, Berlin.

Munro, M.J., Derwing, T .M.,& Burgess, C.S. (2010). Detection of nonnative speaker status from content-masked speech. *Speech Communication*, 52, 626–637.

Rosenberg, A., Hirschburg, J. (2007). V-measure: A conditional entrophy-based external cluster evaluation measure. *Columbia University*, New York City, New York.

Thomas, E.R. (2001). An Acoustic Analysis of Vowel Variation in New World English. *Duke University Press*, Durham, NC.

Vicenik, Chad, and Megha Sundara. (2013). The Role of Intonation in Language and Dialect Discrimination by Adults. *Journal of Phonetics*, 41.5, 297-306.