Regional accents affect speech intelligibility in a multitalker environment

Ewa Jacewicz, Robert Allen Fox

Department of Speech and Hearing Science, The Ohio State University, Columbus, OH, USA.

jacewicz.1@osu.edu, fox.2@osu.edu

Abstract

The current understanding of listener sensitivity to regional accents comes from examination of speech processing in quiet and in noise. This study had two aims: 1) to examine the intelligibility of regional accents in a multitalker environment, and 2) to explore a methodological question of whether systematic regional features can be detected in the productions of only one representative talker or whether several talkers are necessary to provide the suitable sample. Two American English dialects, General American English and Southern American English, were systematically varied both in the target speech and in the masking babble at three sound-to-noise ratios. The results showed that regional accents did influence listeners' performance in a multitalker environment. Intelligibility was hampered when the target and the masker shared common dialect features or when listeners' heard their own dialect in the masking babble. Southern American was a more intelligible variety than General American, which can be attributable to a set of specific acoustic phonetic features. The study found that systematic regional features can be reliably detected in the production of only one representative talker.

Index Terms: regional accent, speech intelligibility, multitalker masker, sentence processing, American English.

1. Introduction

Listening to speech in noisy environments—including background voices—is challenging as these conditions limit intelligibility. Two factors contribute most to the perceptual degradation of the target speech masked by the interfering voices: 1) the increased sound level of the voices relative to the target, and 2) the increased number of interfering talkers [1]. Typically, laboratory testing uses multitalker babble as an imitation of a noisy background in order to assess the effectiveness of a specific masker. The stimuli are often chosen so that the target and the masker would share similar or different voice characteristics [2], the same or a different language [3] or the same or different accent of the same language [4]. For example, exploring intelligibility of foreign-accented speech, an "accented" language was used either in the target sentence [5, 6, 7] or in the masking babble [8] in listening tests with both native and non-native listeners [9]. Native listeners were shown to outperform non-native listeners in a variety of combinations of target and different masking conditions. Also, the intelligibility of accented speech decreased more in the presence of the masker compared to non-accented speech.

In another line of research, the effects of accent in the target sentences were examined in the context of regional dialect of the same language. For example, significant differences were found in processing cost associated with the comprehension of Southern British English and Scottish English talkers by the corresponding listener groups [10]. When the speech was masked by speech-shaped noise, the Southern British English listeners were slower in their responses to Scottish sentences whereas the Scottish listeners were equally fast when responding to either English variety. This discrepancy was explained on the basis of familiarity and experience with each regional accent: The Scottish listeners were familiar and experienced with both English varieties whereas the Southern British English listeners were unfamiliar and inexperienced with Scottish English. Another study demonstrated sensitivity to regional dialect variation in American English [11]. It was shown that when listening conditions deteriorated, such as in the presence of increased noise level in the masker, listeners responded differentially to sentences produced in four regional varieties (Mid-Atlantic, North, South and General American). However, as the intelligibility increased in more favorable listening conditions, the significant differences among dialects in target sentences mostly ceased, indicating that listeners benefitted from greater redundancy of cues in speech. Across testing conditions, the intelligibility was highest for the General American English as it was the variety all listeners were experienced with.

The current study further examined the intelligibility of regional accents while varying the regional dialects both in the target speech and in the masker. The interaction between the dialect in the target and in the masker has not yet been widely explored and our knowledge about listener sensitivity to regional accents comes from processing of speech in quiet and in noise. Our primary aim was to determine the extent to which regional accents affect speech intelligibility in a multitalker environment. As a second aim, the study addressed a methodological question which often arises in studying the perceptual effects of dialect variation in the stimulus speech. Namely, since regional accents are thought to supply coherent deviations in phonetic information [12], using several talkers in the stimulus material is preferred in order to ensure that listeners attend to the systematic dialect features and not to individual talker characteristics. The use of one representative talker only is thought to carry an inherent confound since the listeners have no opportunities to disentangle dialect-specific features from talker idiosyncrasies. While comparing the effectiveness of the two approaches, the present study seeks to determine if dialect features are salient enough to cause similar intelligibility disruptions regardless of the number of talkers used to elicit the perceptual response.

2. Methods

Two regional dialects of American English were selected for the current investigation: A Midland variety spoken in central Ohio which is referred to in the literature as General American English (GAE) [11] and Southern American English (SAE) spoken in the Appalachian region in western North Carolina.

2.1. Previous work

In a previous experiment [13], GAE listeners responded to sentences produced by multiple talkers (4 GAE and 4 SAE) masked by a 2-talker babble (either GAE or SAE) and presented at three sound-to-noise ratios (SNR): +3 dB, 0 dB

and -3 dB. In the easiest condition (+3 dB SNR), listeners' performance to the GAE target was lower when masked by GAE babble than by SAE babble. GAE babble had also a detrimental effect on responses to the SAE target, showing that intelligibility disruption was greater when the babble was in the listeners' own dialect. However, at more challenging conditions at 0 dB SNR and -3 dB SNR, the responses were more variable. Dialect in the babble had no significant effect on intelligibility of the GEA target, whereas intelligibility of the SAE target was higher when masked by GAE babble rather than SAE babble. This suggests that, when conditions deteriorated, listeners were less differentially affected by the dialect in the babble when responding to sentences in their own dialect whereas target-masker dialect mismatch increased intelligibility of sentences produced in a different dialect. .

In terms of intelligibility of the target speech, there was a greater benefit when the target sentences were produced by the SAE talkers (and not by the GAE talkers). The advantage of the SAE dialect increased at more difficult SNRs, indicating that GAE listeners found some acoustic features of SAE easier to process despite their relative inexperience with SAE compared to their own GAE variety.

2.2. Experiment 1: 2-talker masker

2.2.1. Stimulus materials

96 target sentences were selected from the Revised Bamford-Kowal-Bench Standard Sentence Tests [14]. The sentences were read by two carefully selected middle-age male talkers, representing GAE and SAE. The GAE talker was from Columbus, Ohio and the SAE talker was from Sylva, North Carolina. These talkers were different from the talkers used in the previous work (described in Section 2.1) but both were born and raised in the same areas and spoke the same local GAE and SAE varieties. The target sentences were mixed with a 2-talker babble, which was created from recordings of spontaneous conversations taken from a large corpus. The talkers selected for the babble included 4 men (2 for GAE and 2 for SAE) who were of comparable origin, age, speech tempo and fundamental frequency as the talkers who read the target sentences. This 2-talker babble was the same as used in the previous work. For presentation in the experiment, the level of the babble was adjusted relative to the fixed level of the target sentence to create three SNRs: +5 dB (the level of the babble was 5 dB less than the level of the target), 0 dB (both levels were equal) and -5 dB (the level of the babble was 5 dB more than the level of the target). These levels were selected to represent easy, moderate and difficult listening conditions, respectively [3]. The +5 dB SNR (and not +3 dB) was used to examine potential interaction between the dialect in the target and in the masker when listening conditions were most favorable and the -5 dB SNR (instead of -3 dB) was used to test this interaction under adverse listening conditions.

2.2.2. Listeners and procedure

21 monolingual GAE listeners participated in one session. They were students enrolled at The Ohio State University, matched the GAE talkers for dialect and ranged in age from 20 to 28 years. 13 listeners responded to 96 target sentences read by the GAE talker and 8 listeners to the same 96 sentences read by the SAE talker. Half of the sentences were masked by GAE babble and half by SAE babble. The order of experimental blocks was the same for each listener: +5 dB SNR (babble: SAE, GAE), 0 dB SNR (babble: SAE, GAE), and -5 dB SNR (babble: SAE, GAE). Upon hearing each stimulus, listeners typed what they heard as the target sentence into a textbox that appeared on the computer screen. No repetitions were allowed. The experiment was self-paced. Listeners were tested in a sound-attenuating booth. Sound was delivered diotically over Sennheiser HD 600 headphones at a comfortable listening level. The digitally recorded responses were scored by two experimenters on the basis of keywords using the scoring system as in [3].

2.3. Results

Raw scores for each participant were first converted to percent correct and then to rationalized arcsine units, or RAUs [15] to ensure valid assessment of differences across the entire range of the scale (from -17 to 117) after normalizing for ceiling and floor effects. Repeated-measures ANOVA was used to analyze the arcsine transformed scores. SNR level and masker dialect were the within-subject factors; target sentence dialect was the between-subject factor. Pairwise post hoc *t*-tests (using Bonferroni adjustment) were used to explore the nature of significant main effects and interactions.

The main effect of SNR was significant [$F(2,86)=874,3$ $p<.001$] and post hoc tests showed that performance at all three levels differed significantly from one another. The significant main effect of masker dialect [$F(1,43)=43.2$, $p<.001$] indicated higher intelligibility in SAE babble than in GAE babble. The significant main effect of target dialect [$F(1,43)=16.9$, $p<.001$] indicated higher intelligibility of the SAE target than the GAE target.

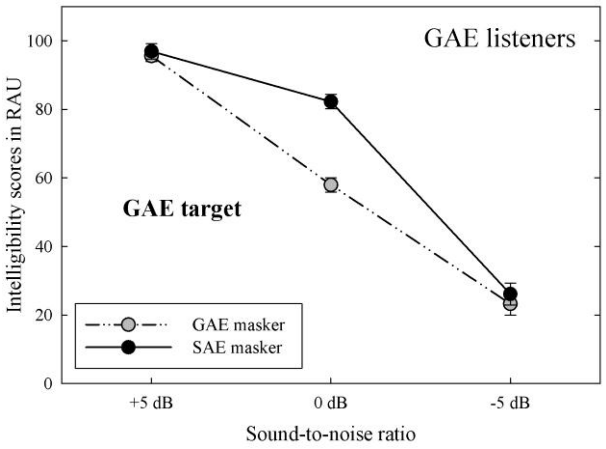

Figure 1: *Means (s.e.) for GAE target in 2-talker babble.*

In response to the GAE target, listeners scored near the ceiling at the easiest +5 dB SNR and the performance worsened with each lower SNR. As Figure 1 shows, the effects of dialect in the masker were evident only at 0 dB SNR. In this moderate listening condition, GAE babble produced more masking for the GAE target than SAE babble. However, dialect in the babble did not significantly affect listeners' performance when the conditions deteriorated (-5 dB SNR) or were very easy (+5 dB SNR).

Figure 2 shows listeners' responses to the SAE target sentences. Unlike for the GAE target, the masker dialect *did* affect intelligibility at all SNRs. Although dialect in the babble did not produce the dramatic differences at 0 dB SNR found in responses to the GAE target, the SAE dialect

advantage was evident both in more accurate responses to SAE target sentences and in less detrimental effects of SAE babble across all SNRs, including the most difficult -5 dB SNR. Clearly, the SAE talker was more intelligible than the GAE talker and listeners' own dialect in the babble was a more effective masker. These differences in listeners' responses produced a significant interaction between SNR, masker dialect and target dialect [$F(2,86)=14.2$, $p<.001$].

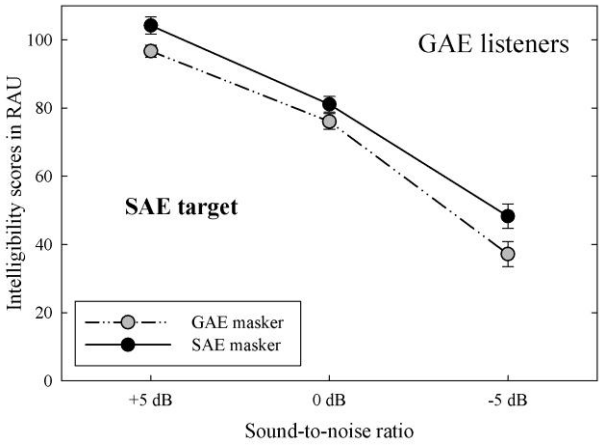

Figure 2: *Means (s.e.) for SAE target in 2-talker babble.*

2.4. Experiment 2: 6-talker masker

Experiment 2 tested if the dialect effects found in Experiment 1 persisted when more talkers were added to the babble. As repeatedly shown in past research, the greater number of interfering voices causes perceptual degradation of the target speech. We therefore expected lower intelligibility of either the GAE or SAE target at all SNRs but our predictions regarding the effects of dialect in a 6-talker masking babble were more circumspect.

2.4.1. Stimulus materials, listeners and procedure

The same 96 target sentences read by the same two talkers were used in Experiment 2. The only difference in the stimulus materials was in the number of talkers in the babble. The talkers selected for the new babble included 12 men (6 for GAE and 6 for SAE). The selection criteria were the same as in Experiment 1 and the talkers for the babble were selected from the same corpus. 18 listeners participated. All listeners were new but were recruited from the same subject pool as in Experiment 1. Each listener was tested in one session. 10 listeners responded to 96 sentences read by the GAE talker and 8 listeners to the same 96 sentences read by the SAE talker. Half of the sentences were masked by GAE babble and half by SAE babble. The presentation of the stimuli and the experimental procedures were exactly as in Experiment 1.

2.5. Results

As in Experiment 1, repeated-measures ANOVA with the within-subject factors SNR level and masker dialect and the between-subject factor target sentence dialect was used to analyze the arcsine transformed scores for the 6-talker babble. All main effects were significant. The main effect of SNR was significant [$F(2,80)=2182.2$, $p<.001$]. Performance at all three levels differed significantly from one another and intelligibility decreased as listening conditions deteriorated. A significant main effect of masker dialect [$F(1,40)=24.3$, $p<.001$] showed again a significantly higher intelligibility in SAE babble than in GAE babble. Target dialect was also significant [$F(1,40)=14.9$, $p<.001$], indicating that intelligibility was higher for the SAE target than for the GAE target. Of our particular interest is a significant interaction between SNR, masker and target dialect [$F(2,80)=8.6$, $p<.001$], which manifested itself similarly as in Experiment 1. As Figure 3 shows, for the GAE target, dialect in the babble affected listeners' performance only at 0 dB SNR. Again, GAE babble was a more effective masker than SAE babble. The pattern of responses for the 6-talker babble (Figure 3) was very similar to that for the 2-talker babble (Figure 1) although intelligibility was generally lower, particularly at -5 dB SNR.

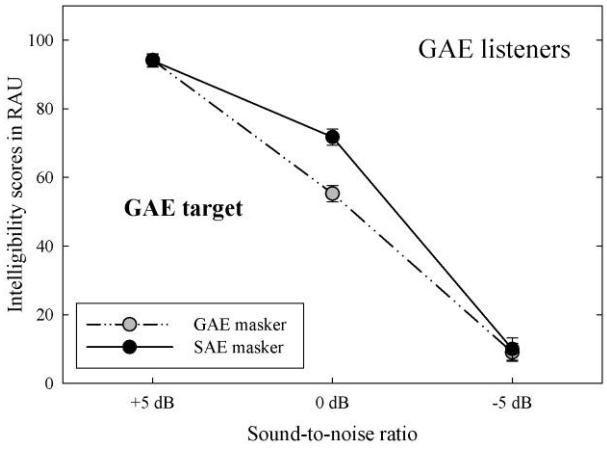

Figure 3: *Means (s.e.) for GAE target in 6-talker babble*.

Response pattern for the SAE target shown in Figure 4 was also consistent with the pattern for the 2-talker babble (Figure 2), showing again that intelligibility was higher in SAE babble at all SNRs.

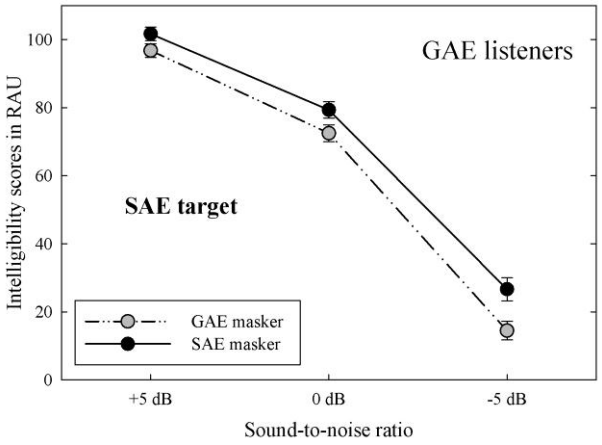

Figure 4: *Means (s.e.) for SAE target in 6-talker babble.*

There were two notable disparities in the response patterns to the 2-talker and 6-talker masker. The first was at 0 dB SNR where the 6-talker masker reduced the difference between the effects of dialects in the babble compared to the 2-talker masker. The second was in the most difficult condition at -5 dB SNR where intelligibility dropped sharply in the 6-talker babble, suggesting that adding more interfering voices to the

babble produced a greater amount of energetic masking which had a greater detrimental effect on intelligibility.

3. Discussion

This paper had two aims. First, it sought to determine the extent to which interacting acoustic cues in regional accents affect speech intelligibility. Second, it explored a methodological question of whether systematic regional features can be detected in the productions of only one representative talker or whether several talkers are necessary to provide the suitable sample. In other words, the study examined whether dialect features are perceptually salient to the extent that they can provide primary indexical information regardless of other peculiarities of individual talkers' voices.

In the previous work (Section 2.1), the dialect effects were most evident at +3 dB SNR. The masker in the listeners' own dialect (GAE) had a more detrimental effect on both the GAE and the SAE target sentences than a masker in a different dialect (SAE). Another important finding was that SAE (and not GAE) was the more intelligible variety. This result was rather surprising since GAE listeners were relatively inexperienced with this regional variety although they might have been familiar with general features of the broadly distributed SAE. The American South is currently the most populous region in the U.S. approaching 115 million (37% of all U.S. residents) living in 16 states. The two current experiments explored these dialect effects in a different design. A consistent result was that SAE (and not GAE) was the more intelligible dialect. Also, in general, listeners scored higher in the presence of the SAE masker, indicating that hearing their own regional accent in the GAE babble (whether 2-talker or 6-talker) had a more detrimental effect on intelligibility of the target (whether GAE or SAE).

The differences between the two studies were found in performance at specific SNRs, which can be spelled out as follows. In the current study, at +5 dB SNR, the responses to target sentences in the listeners' own dialect (GAE) were at the ceiling and listeners' responses were unaffected by maskers in either dialect. However, when the target sentences were in a different dialect (SAE), listeners were more affected by the masker in their own dialect. As the SNR decreased to +3 dB SNR in the previous study, the GAE in the babble had a detrimental effect regardless of the dialect of the target. The differences between the two studies were further evident at the common 0 dB SNR. The detrimental effect of the GAE in the babble persisted when listeners responded to one talker in the current Experiments 1 and 2. However, when responding to multiple talkers in the previous study, the differential effects of dialect in the babble ceased, most likely because the task was more cognitively demanding.

Once the intelligibility of the target further dropped at -3 dB SNR in the previous study, listeners continued to be unaffected by the masker when listening to the GAE target. Finally, at -5 dB SNR in the current tasks, intelligibility of the GAE target was low and was not influenced by the dialect in the babble. This was not the case for the SAE target, however, which was comparatively more intelligible. Furthermore, listeners were still sensitive to the dialect differences in the babble and, even at this condition, performed more poorly when hearing their own dialect in the masker.

Inevitably, these results raise the question of which acoustic features contributed to the greater perceptual salience of SAE compared to GAE. This particular Southern dialect region is classified as Inland South and is considered to represent the most advanced features of Southern English spoken in the U. S. [16]. This variety, spoken in several adjacent counties in western North Carolina, is known as Southern Appalachian English. The current speech samples were collected in Jackson County and this speech community remained mostly homogenous until relatively recently. The largest group of settlers, in terms of the ethnic origin, was the Scotch-Irish from Northern Ireland. The Scotch-Irish contingent was important in shaping the development of Southern Appalachian culture, including language. It is the residual phonetic features of the Scotch-Irish of the settlers that contributed to the distinctiveness of the SAE spoken in Inland South. This variant has a slow speech rate, prolonged vowels and abundance of diphthongal changes in vowels known as "Southern Drawl" [17, 18, 19]. There are also notable prosodic differences [20]. Although the selected talkers for the current study were matched for speech rate with the GAE talkers, it is highly plausible that longer vowels, enhanced spectral dynamics and different pitch movement supplied salient cues related to the melodic component of language. These prominent features could captivate listeners' attention and facilitate processing in a multitalker background.

The methodological question addressed in this paper pertained to whether systematic dialect features can be reliably detected in the production of one representative talker of that dialect as opposed to multiple talkers. The results provide evidence that one representative talker is sufficient: the Southern AE dialect was more intelligible than the GEA regardless of the number of talkers used to produce the sentence material. However, interaction between the dialects in the target and in the masker may be affected by the number of talkers producing the target, most likely due to the differences in processing demands as a function of the number of talkers (and thus the number of different voices) in the target speech.

In part, the current results are in agreement with studies which used native and non-native (or foreign-accented) speech combinations, showing that English listeners performed poorly in the presence of their native language masker but were significantly better when the babble was in a different language [3, 8]. The current study shows that similar effects apply also to the dialect of the same language. That is, listeners perform most poorly when both target and the interfering babble are in their own dialect, most likely due to their familiarity with the dialect features which they can detect both in the target and in the masker.

4. Conclusions

Regional accent does influence listeners' performance in a multitalker environment. Intelligibility can be hampered when the target and the masker share common dialect features or when listeners' hear their own dialect in the masking babble. Under difficult listening conditions, some regional accents can be more intelligible than others, which can be attributable to a set of specific acoustic phonetic features. Systematic regional features can be reliably detected in the production of only one representative talker. However, differences in design and experimental conditions may yield some variations and these factors need to be taken into consideration while comparing the results across the literature.

5. References

[1] Bronkhorst, A. W., "The cocktail party phenomenon: A review of research on speech intelligibility in multiple-talker conditions", Acustica united with Acta Acustica, 86: 117–128, 2009.

[2] Brungart, D. S., Simpson, B. D., Ericson, M. A., and Scott, K. R., "Informational and energetic masking effects in the perception of multiple simultaneous talkers", Journal of the Acoustical Society of America, 110: 2527–2538, 2001.

[3] Van Engen, K. J., and Bradlow, A. R., "Sentence recognition in native- and foreign-language multi-talker background noise", Journal of the Acoustical Society of America, 121: 519–526, 2007.

[4] Calandruccio, L., Dhar, S., and Bradlow, A. R., "Speech-on-speech masking with variable access to the linguistic content of the masker speech", Journal of the Acoustical Society of America, 128: 860-869, 2010.

[5] Gordon-Salant, S., Yeni-Komshian, G. H., and Fitzgibbons, P. J., "Recognition of accented English in quiet and noise by younger and older listeners", Journal of the Acoustical Society of America, 128: 3152-3160, 2010.

[6] Rogers, C. L., Dalby, J., and Nishi, K., "Effects of noise and proficiency on intelligibility of Chinese-accented English", Language and Speech, 47: 139-154, 2004.

[7] Pinet, M., Iverson, P., and Huckvale, M., "Second-language experience and speech-in-noise recognition: Effects of talker-listener accent similarity", Journal of the Acoustical Society of America, 130: 1653-1662, 2011.

[8] Van Engen, K. J., and Bradlow, A. R., "Sentence recognition in native- and foreign-language multi-talker background noise", Journal of the Acoustical Society of America, 121: 519–526, 2010.

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

[10] Adank, P., Evans, B. G., Stuart-Smith, J., and Scott, S. K., "Comprehension of familiar and unfamiliar native accents under adverse listening conditions", Journal of Experimental Psychology: Human Performance and Perception, 35: 520-529, 2009.

[11] Clopper, C. G., and Bradlow, A. R., "Perception of dialect variation in noise: Intelligibility and classification", Language and Speech, 51: 175-198, 2008.

[12] Wells, J. C., "Accents of English, volume 1: An introduction", Cambridge University Press, 1982.

[13] Jacewicz, E., and Fox, R. A., "The effects of dialect variation on speech intelligibility in a multitalker background", Applied Psycholinguistics, forthcoming.

[14] Bench, J., Kowal, A., and Bamford, J., "The BKB (Bamford-Kowal-Bench) sentence lists for partially-hearing children. British Journal of Audiology", 13: 108-112, 1979.

[15] Studebaker, G., "A 'rationalized' arcsine transform. Journal of Speech, Language, and Hearing Research", 28: 455–462.

[16] Labov, W., Ash, S., and Boberg, C., "Atlas of North American English: Phonetics, phonology and sound change", Mouton de Gruyter, 2006.

[17] Jacewicz, E., Fox, R. A., and Wei, L., "Between-speaker and within-speaker variation in speech tempo of American English", Journal of the Acoustical Society of America, 128: 839-850, 2010.

[18] Fox, R., A., and Jacewicz, E. "Cross-dialectal variation in formant dynamics of American English vowels", Journal of the Acoustical Society of America, 126: 2603-2618, 2009.

[19] Jacewicz, E., Fox, R. A., and Salmons, J., "Cross-generational vowel change in American English", Language Variation and Change, 23:45-86, 2011.

[20] Fox, R. A., Jacewicz, E., and Hart, J., "Pitch pattern variations in three regional varieties of American English", Interspeech, Lyon, France, 2013.