

**The effects of indexical and phonetic variation on vowel perception in
typically developing 9- to 12-year-old children**

Ewa Jacewicz, Robert Allen Fox

Department of Speech and Hearing Science, The Ohio State University

Contact author

Ewa Jacewicz
The Ohio State University
Department of Speech and Hearing Science
1070 Carmack Road
Columbus, OH, 43210-1002.
Email: jacewicz.1@osu.edu
Phone: 614-292-4874
Fax: 614-292-7504

This is an author-produced manuscript that has been peer reviewed and accepted for publication in the *Journal of Speech, Language, and Hearing Research (JSLHR)*. As the "Just Accepted" version of the manuscript, it has not yet undergone copyediting, proofreading, or other quality controls associated with final published articles. As the publisher and copyright holder, the American Speech-Language-Hearing Association (ASHA) disclaims any liability resulting from use of inaccurate or misleading data or information contained herein. Further, the authors have disclosed that permission has been obtained for use of any copyrighted material and that, if applicable, conflicts of interest have been noted in the manuscript.

ABSTRACT

Purpose: To investigate how linguistic knowledge interacts with indexical knowledge in older children's perception under demanding listening conditions created by extensive talker variability.

Method: Twenty five 9- to 12-year-old children, 12 from North Carolina (NC) and 13 from Wisconsin (WI), identified 12 vowels in isolated *hVd*-words produced by 120 talkers representing the two dialects (NC and WI), both genders and three age groups (generations) of residents from the same geographic locations as the listeners.

Results: Identification rates were higher for responses to talkers from the same dialect as the listeners and for female speech. Listeners were sensitive to systematic positional variations in vowels and their dynamic structure (formant movement) associated with generational differences in vowel pronunciation resulting from sound change in a speech community. Overall identification rate was 71.7%, which is 8.5% lower than for the adults responding to the same stimuli in Jacewicz and Fox (2012).

Conclusions: Typically developing older children are successful in dealing with both phonetic and indexical variation related to talker dialect, gender and generation. They are less consistent than the adults most likely due to their less efficient encoding of acoustic-phonetic information in the speech of multiple talkers and relative inexperience with indexical variation.

KEY WORDS: Speech perception, Children, Dialect

Introduction

Research on children's development of speech perception has naturally focused on the early formative years, from infancy through middle childhood. Admittedly, our knowledge of perceptual development in infants and young children has grown over the past few decades (e.g., Gerken & Aslin, 2005; Jusczyk, 1997; Kuhl, Stevens, Hayashi, Deguchi, Kiritani, & Iverson 2006; Nittrouer, 1996; Sundara, Polka, & Genesee, 2008; Walley, 2005). However, relatively little is known about the perceptual abilities of older children, those above 7 years of age, who appear to perform almost as well as adults but are still lacking adult-like consistency. Addressing this gap, the current study makes inquiries into perception in typically developing older children and, assuming their advanced speech processing abilities, examines their vowel recognition performance in the face of extensive talker variability in the speech stimuli.

Speech perception involves extracting two types of information from the acoustic signal: linguistic and indexical. The term *indexical* refers to those aspects of the speech signal that provide information about the talkers themselves, including their identity, social status or health condition (Abercrombie, 1967). Voice quality of the talkers conveys biological information related to their gender, age, size and medical state, may index their psychological characteristics such as personality and may also supply cues to any social characteristics such as regional origin, social values, attitudes and education. In this study, we are interested how linguistic knowledge interacts with indexical knowledge in older children and how the relationship between the two types of knowledge is manifested in their vowel recognition performance. In particular, we examine older children's sensitivity to fine phonetic detail in vowel quality (the linguistic knowledge) entailed in generational changes in vowel pronunciation as a result of sound change in a speech community (Labov, 1994). In this study, these generational changes correspond to

talkers' age in that older adults produce somewhat different vowel variants than do young adults and children, respectively. These generational variations are tested in the social context of two regional dialects so that children are exposed to indexical features of their own dialect and of a different dialect. Biological information in talker gender is also of interest to the study, which uses both male and female talkers. Altogether, talker's regional dialect, age group (generation) and gender are the three types of indexical information investigated.

The participants of our study are typically developing children 9 to 12 years of age. Presumably, children of this age are generally successful in interpreting linguistic information in static and dynamic cues in vowels, a skill they developed earlier in life (Nittrouer, 2007; Ohde & German, 2011; Ohde & Haley, 1997; Sussman, 2001). While speech pattern recognition is shown to be still maturing in children 5 to 7 years of age (e.g., Walley, 1993), we expect that 9-12-year old children will be able to use the spectral and temporal cues in categorizing vowels in adult-like manner. Recurrent results of studies reporting lack of significant difference between older children and adults suggest that these children have developed mature perceptual representations for both vowels and consonants (Parnell & Amerman, 1978; Walley & Flege, 1999) although they may still be less consistent than adults in categorizing phonemic contrasts (Flege & Efting, 1986; Hazan & Barrett, 2000), especially under adverse listening conditions (Eisenberg, Shannon, Schaefer Martinez, Wygonsky & Boothroyd, 2000; Johnson, 2000).

The extent to which older children can successfully cope with indexical information when listening to voices of multiple talkers remains largely unknown. Indexical learning appears to begin very early, even *in utero*, as neonates and then young infants begin to recognize the maternal voice and its unique pitch patterns (Bergeson & Trehub, 2007; Mehler, Bertoncini, Barriere, & Jassik-Gerschenfeld, 1978; Moon, Lagercrantz, & Kuhl, 2013). However, little

research has explored how knowledge of indexical variation develops in acquisition. From a handful of related studies on accent acquisition, we learn that children 4 to 5 years old are aware of accent variation among talkers and have developed social preferences for native-accented speakers of their native language (American English) to foreign-accented speakers (Kinzler, Corriveau, & Harris, 2011; Kinzler, Shutts, DeJesus, & Spelke, 2009). The earliest traces of these preferences were already found in infants, who looked longer at native-accented speakers and preferred to reach for objects and foods that were offered by native speakers (Kinzler, Dupoux, & Spelke, 2007; Shutts, Kinzler, McKee, & Spelke, 2009). However, while children 5-6 years old were found to successfully categorize accents into native (French) versus foreign (British English-accented French) in sentences spoken by two talkers of each accent, they were less sensitive to the dialect variation in their native French (Girard, Floccia, & Goslin, 2008). In particular, they were unable to categorize regional accents into Southern or Northern French on the basis of sentences produced with these regional accents. In another study, similar sentential material was presented to British 5- to 7-year-old children for categorization into regional (British English of the Plymouth area versus Irish) and foreign (French-accented British English) accents (Floccia, Butler, Girard, & Goslin, 2009). While the 5-year-olds were unable to categorize the regional and foreign accents, the 7-year-olds could reliably distinguish between the two regional dialects but were significantly better at detecting the difference between the foreign-accented English and their native Plymouth variety. The asymmetry in attending to regional dialect versus foreign accent cues by 5- and 7-year-olds was interpreted as a developmental effect.

The perception of indexical information by children above 7 years of age was investigated by Hazan and Markham (2004). In that study, word intelligibility was measured

within the same regional accent group for 45 talkers differing in age (adults and children) and gender in a low-level background noise. The younger child listeners (7-8 years old) made more errors than the older children (11-12 years old) and adults. However, a consistent result was that women tended to be more intelligible than men and that child talkers were no less intelligible than adult talkers. The relative intelligibility of individual talkers was also consistent across all listener groups, suggesting that the acoustic-phonetic characteristics of individual utterances (and not the age of either listeners or talkers) were the primary determinants of higher word intelligibility.

Our current investigation focuses on older children's ability to process linguistic and indexical information in the face of extensive talker variability in stimulus speech. To our knowledge, no study has as yet examined older children's perceptual performance in a vowel recognition task under such challenging conditions. The current experiment will thus inform us about their abilities to respond to variations in voice qualities of 120 adult and child male and female talkers, which convey fine acoustic-phonetic details pertaining to three levels of accent strength in two regional dialects. While we expect some confusions because sensitivity to indexical variation may still be maturing in children of this age and their abilities to cope with stimulus uncertainty and task demands could still differ from adults, we expect their overall performance to approximate that of the adults.

Based on the reviewed literature, we expect that the older children in the current study will have already developed awareness for regional variation. Presumably, the 9- to 12-year olds will be able to differentiate between two dialects of their native language when attending to longer passages of speech although they may experience greater difficulty when listening to minimally contrastive individual words. In an earlier work, we found that 8-12 years old children

acquired regional vowel characteristics typical of their native dialect (Jacewicz, Fox, & Salmons, 2011a). For three dialects of American English, the differences were found in general vowel dispersion pattern and in formant dynamics, indicating children's command of dialect-specific features. We reasoned that, if children of that age produce regional accents, they are likely to perceive the dialectal differences, particularly since they can learn a new dialect from available input upon arrival in another speech community (e.g., Chambers, 1992; 2003; Kerswill, 1996; Payne, 1980).

There are two widely known studies which reported adults' perceptual accuracy for vowels in individual words in the *hVd*-context, the original Peterson and Barney (1952) study and its replication by Hillenbrand, Getty, Clark, & Wheeler (1995). Listeners in these studies responded to stimulus materials produced by a large number of talkers including men, women and children, 76 talkers in the first study and 139 in the second. In spite of such considerable talker variability, both reported high identification rates, 94.4% and 95.4%, respectively. However, these high rates may have resulted, in part, from the fact that half of the listeners also served as talkers in the Peterson and Barney (1952) study and because listeners in Hillenbrand et al. (1995) were phonetically trained. While both studies attempted to match talkers and listeners for dialect, Hillenbrand et al. (1995) used more strict criteria and included dialect screening in selecting the listeners. However, much lower identification rates were obtained in a subsequent replication of this experiment reported in Labov (2010), in which dialect was the control variable. In particular, when stimuli produced by talkers from three different dialects were randomly presented to adult listeners from the same dialects, the overall accuracy reached only 77%. The lower rates resulted from dialect misalignment between talker and listener, suggesting that dialect variation can be a strong source of confusion and mislabeling of vowels.

Recently, Jacewicz and Fox (2012) further showed that adult listeners were sensitive to both dialect variation and detailed acoustic cues in vowels reflecting sound change, that is, systematic generational change in pronunciation patterns in a speech community (Jacewicz, Fox, & Salmons, 2011b; 2011c; Labov, 1994). Sound change pertains to both positional relations among vowels in the acoustic space and their dynamic structure and adult listeners were responsive to these generational variations. The overall accuracy for the adults in Jacewicz and Fox (2012) was 80.2%, which is comparable to Labov (2010) and considerably lower than both Peterson and Barney (1952) and Hillenbrand et al. (1995). These results establish an important reference for the current study, in which the same material as in Jacewicz and Fox (2012) is presented to children.

On the basis of that study, we expect that the older children in this study will manifest a native dialect advantage identifying vowels of their own dialect with higher accuracy than the vowels from another dialect. Higher identification rates are also expected for vowels produced by female talkers than by male. In terms of generational differences, we expect more variability in children's patterns which may also differ from those in the adults. In particular, if sensitivity to sound change in vowel production is dependent upon linguistic and social interactions accumulated through the lifetime, children may not be able to associate the acoustic cues with indexical information to the same degree as the adults because of their relative inexperience with these variations. On the other hand, given their developed perceptual abilities to utilize and interpret linguistic information in static and dynamic cues in vowels, we expect them to manifest their knowledge of the relationship between linguistic and indexical variation at least for some vowel categories in the stimulus set.

It needs to be emphasized that perceptually salient acoustic cues which carry indexical information—such as the amount of diphthongization as a function of a specific talker generation—are not distributed evenly across all vowels. Some vowel categories are more affected by cross-generational sound change than others and some are not at all affected. Also, some vowels are more prone to confusions than others. It is of interest how older children interpret the static and dynamic cues in individual vowels in light of their accumulated linguistic and indexical knowledge and how their responses may differ from those of the adults in Jacewicz and Fox (2012).

Method

Participants

Twenty eight children were originally recruited. Three participants were excluded from the final data analyses because their response patterns indicated that they were unable to do the task. Data from the remaining 25 child listeners were included in all analyses. Out of the 25, 12 children (5 M, 7 F) were born and raised in or near Sylva, NC and spoke a Southern variant of American English typical of Western North Carolina. Thirteen children (7 M, 6 F) were born and raised in or near Madison, WI and spoke a Midwestern variety of American English typical of Southeastern Wisconsin. Their age ranged from 9 to 12 years in both NC ($M = 11.1$, $SD = 0.8$) and WI ($M = 10.5$, $SD = 1.3$). The participants were recruited using flyers, bulletin board postings, email and word of mouth. Each child spoke the local dialect as verified by the research team on the basis of short conversations which elicited several markers of the dialect. Based on the background information provided by the parent at the outset of testing, no child had hearing problems or a history of language and reading disorder. Recruitment criteria and testing

procedures were in accord with approved IRB protocols for treatment of study participants at Western Carolina University and University of Wisconsin-Madison.

Stimuli

The stimuli used in the perception experiment were the naturally produced utterances *heed, hid, heyd, head, had, heard, hod, hawed, hoed, who'd, hood, hide* containing 12 vowels: /i, ɪ, e, ε, æ, ɜ, ɑ, ɔ, o, u, ʊ, aɪ/. They were produced as isolated tokens by a total of 120 talkers representing the two dialects (NC and WI), both genders and three age groups (or generations) of residents from the same geographic locations as the listeners. To emphasize the cross-generational aspect of differences in vowel production as a function of talker age, we adopt here the naming convention for the three talker age groups as in Jacewicz and Fox (2012): children (C) for talkers aged 8-12 years, parents (P) for young adults aged 35-50 years and grandparents (GP) for older adults aged 66-91 years. These talkers were selected from a large corpus of recordings completed for a production study of regional and cross-generational variation in American English (see Jacewicz et al., 2011b for details about the talkers and the recording procedure). In order to increase variability in voice quality of the talkers used in this study and, at the same time, restrict the number of experimental sessions, only 6 out of the 12 stimulus tokens were selected from each individual talker. The selected words were pseudorandomized across the talkers in each of the two separate dialect groups to ensure that all 12 different vowel categories appeared an equal number of times in each dialect, gender and age group. Altogether, 60 NC talkers produced 360 NC tokens (6 vowel categories x 10 talkers x 2 genders x 3 ages) and 60 WI talkers produced the corresponding 360 WI tokens for the total of 720. This stimulus material was the same as that used with adult listeners in Jacewicz and Fox (2012). The selection of stimuli was completed by two experienced phoneticians based on acoustic measurements and

common auditory criteria: falling pitch, no obvious idiosyncrasies in talker voice characteristics, and dialect-specific pronunciation pattern as appropriate for each talker's generation. Prior to presentation, all tokens were analyzed acoustically and equalized for mean intensity.

Summary of the acoustic characteristics of stimulus vowels. The acoustic analysis included vowel duration and frequencies of the first two formants measured at 5 equidistant time points (20-35-50-65-80%) in the vowel to estimate the dynamic changes in the formant pattern. While more details of this type of analysis can be found elsewhere (notably Fox & Jacewicz, 2009; Jacewicz et al., 2011b), we briefly summarize here the major characteristics of the current stimulus set. There were notable differences in vowel duration. NC vowels were on average longer than WI vowels ($M = 302$ ms and $M = 254$ ms, respectively) and female vowels were slightly longer than male ($M = 286$ ms versus $M = 270$ ms). The /ɔ/ was the longest vowel in the set ($M = 318$ ms) and the /ɪ/ was the shortest ($M = 226$ ms).

The NC vowels represent a set of features typical of Southern English and some of these features are related to vowel rotations called the Southern Shift (Labov, Ash, & Boberg, 2006). The Southern Shift is manifested primarily in front vowels /i, ɪ, e, ε, æ/ which are heavily diphthongized and tend to overlap in the acoustic space. Furthermore, the /æ/ is raised approximating the position of /ε/, the diphthong /aɪ/ is produced as a monophthong and /u, ʊ/ are fronted. These features were present in the productions of the GP talkers and, to a great extent, in P talkers although their vowel system showed signs of restructuring. The Southern Shift features were clearly receding in children who showed less of the acoustic overlap and lesser diphthongization of the front vowels, lowering of /æ/ to the position of the monophthongal /aɪ/ in GP-generation and greater formant movement in /aɪ/ which reflected production more typical of

a diphthong. On average, children's monophthongs had a reduced formant movement (i.e., were less "diphthongized") compared to those in GP and P generations. Detailed acoustic displays of the vowels, tables reporting duration and formant measurements along with further discussions can be found in Jacewicz and Fox (2012).

The WI vowels had a different dispersion in the acoustic space and a different pattern of formant dynamics. In general, WI vowels were more "monophthongal" than were NC vowels except for the full diphthong /aɪ/. Notable differences included: 1) a lack of the acoustic overlap of the front vowel group, 2) a raised variant of /æ/ whose formant movement called Northern breaking (Labov et al., 2006) had a different direction compared to NC /æ/, 3) a far back position of /u/, and 4) monophthongal versions of /e, o/ which are typical of English spoken in Wisconsin and Minnesota. Overall, WI vowels did not change their positions and the amount of spectral dynamics to the extent as NC vowels. Rather, the cross-generational changes affected individual vowels such as the /ɛ/ in P generation which lost its formant movement and /æ, ɑ/ in children which changed their positions in the vowel space. In particular, the /æ/ lowered its position relative to the GP generation to sound more like an [æɑ] and not [ɛæ] and the /ɑ/ was in a closer proximity to /ɔ/, suggesting that these two back vowels begin to merge in this dialect as they have merged across broad regions of the United States. This relatively recent sound change known as the Low Back Vowel Merger is currently spreading into areas where one might anticipate resistance to merger (Dinkin, 2011; Irons, 2007).

Procedure

Listeners were tested in one session lasting approximately one hour. The 720 stimuli were presented in three blocks of 240 tokens each. In each block, the tokens were randomized

and included instances of all 12 vowels, both genders, both dialects and three generations. Each listener was tested individually in a quiet room and the experiment was administered by two different female research assistants, one at Western Carolina University and one at University of Wisconsin-Madison. Signals were delivered binaurally over Sennheiser HD600 headphones at 70 dB SPL. Prior to the actual experiment, a 20-item practice was given to each listener for familiarization with the task. The tokens in the practice were different from those used in the actual testing but all 12 vowels were included in the practice and the tokens were produced by talkers from both dialects, both genders and three generations. The responses from the practice were not included in the overall total of responses from each listener in the experiment. On the basis of the practice, the experimenter determined whether the child was able to meet the technical demands of the experiment and to match the orthographic form with the sound. For example, an obvious mismatch such as choosing “heed” upon hearing *who’d* was taken as evidence that the child was unable to do the task. In general, children were comfortable with the task except for two from NC and one from WI, whose responses indicated that they were merely guessing. Data from these participants were not analyzed in the present study. No response feedback was provided during any part of the experimental procedure, including the practice.

The experiment was under computer control. Listeners responded by clicking with the mouse on one of 12 boxes on the screen, which displayed a given word such as “head” or “had” (there was one box for each token type). The listeners were told that they would hear one word at a time. Immediately upon hearing the word they were to click on one box on the screen to indicate which word was presented. They were also instructed that they could listen to the same signal one additional time if they missed the word but then they had to guess if they were still uncertain which response to choose. Listeners were not informed about the dialect, age and

gender of the talkers. They were only told they would hear many different voices. The experiment was self-timed. The next token could only be played after listener had responded to the token. Breaks between experimental blocks were allowed and encouraged. The order of presentation of the blocks was counterbalanced across listeners.

Results

Average identification rates (IDRs) for NC and WI listeners are summarized in Table 1. Complete confusion matrices can be found in Appendix 1 for NC listeners and in Appendix 2 for WI listeners. Shown in Table 1 are responses to vowels of listeners' own (native) dialect and of the other (non-native) dialect. For both groups, the overall IDRs were higher for the native dialect although several individual non-native vowels, two for NC listeners and four for WI listeners, were identified with comparatively higher accuracy. The overall IDR for the children was 71.7%. This is 8.5% lower than the overall IDR for the adult listeners in Jacewicz and Fox (2012), which is included in Table 1 for comparison. The lower rates were consistent for the majority of the vowels, indicating somewhat greater uncertainty in children's responses compared to adults. Particularly striking were the IDRs for /u/ and /a/ which, respectively, were 23.5% and 16.9% percentage points lower than adults. These results will be discussed in greater detail below.

[TABLE 1 about here]

Overall IDRs were assessed by a repeated-measures analysis of variance after transformation of the proportional data to RAU (Studebaker, 1985). The within-subjects factors were talker dialect, gender and age group and listener dialect was included as the between-subjects factor. The degrees of freedom for the *F*-tests were Greenhouse-Geisser adjusted to address significant violations of sphericity. We also report a measure of the effect size – partial

eta squared (η_p^2). The relevant post hoc analyses were completed using Bonferroni-adjusted *t*-tests for all pairwise comparisons.

Of particular interest in this ANOVA was that the main effect of listener dialect was not significant, $F(1,23) = 0.095$, $p = .760$, $\eta_p^2 = .004$, indicating no overall differences in children's performance as a function of their dialect. However, a significant listener dialect by talker dialect interaction showed native dialect advantage in that IDRs were higher for NC children responding to NC talkers and for WI children responding to WI talkers, $F(1,23) = 53.4$, $p < .001$, $\eta_p^2 = .699$. Similar results were obtained for adult listeners in Jacewicz and Fox (2012), indicating that knowledge of the native dialect features facilitates vowel perception by both adults and children. This interaction further showed that the difference between the IDRs for the native and non-native dialect was greater for NC listeners than for WI listeners. This suggests that NC children's experience with the Southern dialect may have contributed to their notably higher IDRs in response to NC talkers ($M = 76.7$ RAU) than to WI talkers ($M = 64.3$ RAU). WI children's IDRs were also higher in response to their native WI vowels ($M = 73.4$ RAU) than to non-native NC vowels ($M = 69.8$ RAU) but the difference between the two was smaller. These results are somewhat surprising because one might expect the opposite: NC children should be more familiar with the Midwestern vowel features which are less distant from a general American variety which they hear daily in the media, than WI children with Southern features, which are more distinct from "standard" varieties. More focused experiments are needed to shed more light on this discrepancy.

All three other main effects were significant. The significant effect of talker dialect revealed that NC vowels yielded on average higher IDRs than WI vowels, $F(1,23) = 15.92$, $p = .001$, $\eta_p^2 = .409$. We will return to this result in General Discussion. The significant effect of

talker gender showed that IDRs for females were significantly higher than for males, $F(1,23) = 23.08, p < .001, \eta_p^2 = .501$. This finding reflects the general tendency for female speech to be more intelligible than male speech which has also been reported elsewhere (Bradlow, Toretta, & Pisoni, 1996; Ferguson, 2004). Similarly, the main effect of gender was significant for adult listeners in Jacewicz and Fox (2012) and overall IDRs were also significantly higher for female talkers, which is consistent with the present results for children.

There was a significant effect of age group, $F(1.8,40.5) = 4.15, p = .027, \eta_p^2 = .153$. Post hoc testing using a Bonferroni correction for multiple tests showed lower IDRs for GP versus P talkers ($p = .002$) and lower IDRs for GP versus C talkers ($p = .022$). The pairwise comparisons between P and C talkers were not significant. However, a significant talker dialect by age group interaction, $F(2,46) = 17.56, p < .001, \eta_p^2 = .433$, revealed that this pattern was only true for NC dialect but not for WI dialect, in which there were no significant differences between age groups. This implies that the dialect features produced by the old NC talkers were salient enough to cause a greater number of confusions compared to younger NC talkers whereas accuracy for WI dialect was unaffected by talker generation. However, these results must be interpreted with caution because of a relatively small sample of the current study and the fact that the stimuli were randomized across dialects. While the acoustic analysis indicated that NC vowels of GP talkers were indeed highly affected by the Southern Shift whose features were receding in P and C talkers, the generational changes in WI were also evident but comparatively less distinct. It could be the case that generational differences would become more evident if more speakers and listeners were added to the study or if the presentation of speech material were blocked by talker dialect. These possibilities need to be addressed in future work.

One more significant interaction arose between talker dialect, age, and gender, $F(2,46) = 9.72, p < .001, \eta_p^2 = .297$. The locus of the interaction was in higher IDRs for female talkers than for male in P and C groups in NC and in P and GP groups in WI. There is no obvious reason for why the anticipated female talker advantage was not manifested in either NC GP or WI C groups. It is possible that the results would differ if more speakers and listeners were included in the study. No other interactions were significant.

Identification of Individual Vowel Categories

Thus far, we have examined general trends in vowel identification and our analyses focused on overall patterns for all vowel categories combined. However, as shown in the confusion matrices in Appendices 1 and 2, the effects of talker dialect and age group were highly variable across individual vowels and, in many cases, there were large gender-related differences. Given the extensive variability in the present stimulus set, is it difficult to determine which acoustic cues were used by the present listeners in recognition of individual tokens. However, we can explain results for selected vowels by linking the perceptual response to the acoustic characteristics that seem to best account for the obtained patterns. Acoustic analysis of the current stimuli pointed to variation in the dynamic vowel structure as one important source of cues for the listeners. Formant dynamics conveyed not only linguistic information about vowel quality but also provided cues to indexical variation related to dialect and talker generation. As shown and argued elsewhere (Jacewicz et al. 2011b), vowel inherent dynamic structure does not remain intact in the process of cross-generational sound change but the primary indicator of sound change –the positional change in the acoustic space– also involves corresponding changes in formant dynamics. To illustrate the apparent sensitivity of child listeners to the dynamic

structure in addition to the positional vowel change, we will now consider perceptual responses to the acoustic variation in NC variants of /aɪ/.

[FIGURE 1 about here]

Shown in the left panel of Fig. 1 are mean IDRs for NC /aɪ/. For NC listeners, the IDRs were low and the lowest rate was for GP talkers who produced the vowel as a monophthong, which was fronted relative to the variants of the younger generations (see the right panel). Consequently, the confusions were mostly with /æ/ - a front and relatively monophthongal vowel. IDRs increased in P talkers with the increase in the amount of formant movement in their productions and the confusions were also with back vowels /ɑ/ and /ɔ/, most likely in response to the backing of the vowel. However, when formant movement increased in C females becoming more typical of a diphthong, the IDRs were comparatively higher and reached 80%. The same general pattern was maintained for WI listeners although their IDRs were lower (except for C females), presumably due to their relative inexperience with the NC dialect. Overall, these results show that both NC and WI listeners were sensitive to formant dynamics. The increased formant movement disambiguated the vowel and reduced its confusions with the monophthongs. Notably, the pattern of confusions was similar for both listener groups and was also reflecting their sensitivity to the positional vowel changes in the acoustic space irrespective of their amount of experience with the NC dialect.

[FIGURE 2 about here]

Figure 2 provides another example of listeners' sensitivity to both formant dynamics and positional vowel changes. In the Midwestern WI variety, the variants of /ɛ/ and /æ/ are in close proximity in the acoustic vowel space (see the right panels). The vowels undergo cross-

generational changes both in their positions and dynamic characteristics. NC listeners (upper left) identified /ɛ/ with highest accuracy in response to GP talkers who produced it with a considerable formant movement. Apparently, the combination of cues in the dynamic change along with the direction of formant movement and the position of the vowel created the clearest example of /ɛ/. However, the IDRs dropped and the confusions with /æ/ increased for P talkers most likely in response to a drastic reduction of formant movement and the lowering of the vowel in the acoustic space to a position close to /æ/. There was a further drop in IDRs for C talkers whose /ɛ/ was even closer to /æ/. Clearly, NC listeners were sensitive to this cross-generational sound change in a non-native dialect. The comparatively higher IDRs for WI listeners in response to P and C talkers indicated their greater familiarity with these variants of /ɛ/ but, consistent with NC listeners, their accuracy was also higher for GP talkers.

Responses to WI /æ/ are shown in the lower left panel. The greatest difference between the NC and WI listeners was in their respective responses to GP talkers. NC listeners confused /æ/ comparatively more often with /ɛ/ when it was produced as a raised variant in the acoustic space (i.e., closer to /ɛ/) in GP talkers. WI listeners also confused this raised variant with /ɛ/ but mostly when it was produced by males (compare the acoustic displays in the right panels). As the vowel “descended” in the vowel space with each younger generation, it was confused with /ɛ/ less often by both listener groups, which indicates its greater perceptual separation from /ɛ/.

The above examples illustrate that the children were sensitive to both positional variations in the vowel space and to variations in the dynamic vowel structure in the productions of multiple talkers. However, it needs to be emphasized that the IDRs varied greatly as a function

of vowel category and some vowels were more prone to confusions than others. For example, inspecting the confusion matrices, we find that NC and WI /ɑ/ and /ɔ/ were highly confused and their IDRs were lower compared with adults in Jacewicz and Fox (2012). This suggests that the proximity of the vowels in the back corner of the vowel space and a mixture of spectrally overlapping exemplars from two dialects including the monophthongal NC /aɪ/ elevated stimulus uncertainty for the children. Presumably, their comparatively lesser experience with the subtle distinction among these vowels –which were also problematic for the adults– contributed to a greater amount of mutual confusions. It may also be the case that separability of the vowels in this region of the vowel space is severely reduced, as suggested elsewhere (Neel, 2008).

The IDRs for another vowel, the WI /e/, were also low and both WI and NC children confused the vowel with the front monophthongs /i, ɪ, ε/. This general pattern is consistent with that of the adults in Jacewicz and Fox (2012). The greater number of confusions with the monophthongs can be explained on the basis of listeners' sensitivity to the reduced formant movement in WI variant of /e/, which was more typical of a monophthong. Accuracy in identifying /u/ was surprisingly low, on average 23.5% lower than for the adult listeners. However, there was a common pattern of confusions in that both adults and children confused NC /u/ with /ʊ/ and /o/ and WI /u/ with /ʊ/ only. These two sets of results suggest that the greater formant movement in NC /u/ most likely contributed to mislabeling of the vowel as /o/ and a more “monophthongal” version of WI /u/ was mislabeled more often as /ʊ/. Again, children's inexperience with the extensive variation in /u/ in terms of positional differences (fronted NC variants versus far back WI variants) and dialect-specific dynamic structure (greater and reduced formant movement in NC and WI, respectively) is a plausible explanation of their comparatively

lower IDRs. On the other hand, when acoustic cues were perceptually unambiguous and seemingly unaffected by talker variability such as in the word *heard*, children's accuracy was almost at the ceiling (96%) suggesting their attentiveness to the stimulus material.

Comparing the Performance of Children and Adults

We expected that the overall performance of the 9- to 12-year-old children would approximate that of the adult listeners. We found that children's IDRs were on average 8.5% lower than those for adults in Jacewicz and Fox (2012). To assess if this difference was significant, a separate ANOVA was used which included identification responses for adults from that study. In this ANOVA, talker dialect, age and gender were the within-subject factors and listener dialect and age (children and adults) were the between-subject factors. Most importantly, the main effect of listener age was significant, $F(1,51) = 22.13, p < .001, \eta_p^2 = .303$, showing significantly poorer performance by children compared with adults. This finding demonstrates that when faced with extensive talker variability, older children still do not perform as well as adults. Their abilities to cope with indexical information are still maturing and their comparatively lesser experience with dialectal and generational variation adversely affects their identification decisions.

General Discussion

This study investigated vowel recognition performance of typically developing 9- to 12-year-old children under demanding listening conditions created by extensive talker variability in the stimulus speech. The manipulated sources of indexical variation included dialect (native and non-native), talker gender and dialect-specific generational differences in vowel production resulting from sound change in a speech community. Based on the reports in the literature, we expected that children of this age will have developed adult-like abilities to categorize vowels on

the basis of their spectral and temporal cues. They will have also developed awareness for regional variation in speech and will display perceptual preferences for the native-dialect and for female speech. However, we could not predict their ability to deal with extensive talker variability in identifying the vowels, a skill that characterizes speech processing by adults. To determine this, we deliberately introduced variability by including 120 talkers ranging in age from children to old adults and presented this dialect-controlled material in a fashion of the classic Peterson and Barney (1952) vowel identification experiment. In our previous work, we found that under such difficult listening conditions, the accuracy of adult listeners is reduced (Jacewicz & Fox, 2012). This is not surprising since a number of earlier related studies have demonstrated that listeners identify isolated words more accurately when they are produced by a single talker than by multiple talkers (e.g., Mullennix, Pisoni & Martin, 1989) and that dialect variation also reduces the accuracy. The results of the current experiment allowed us to compare whether and how children's performance differs from that of adults for the effects of talker variability.

We found the performance of these 9- to 12-year-olds to be relatively good but significantly worse than that of adults. Their overall identification rate of 71.7% was lower compared to the rates for adults, which were 80.2% in Jacewicz and Fox (2012) and 77% in a similar experiment reported in Labov (2010). However, the pattern of confusions in children's data was similar to that of adults in Jacewicz and Fox (2012), which indicates that children were able to attend to the acoustic cues in vowels and indexical variations in the talkers' voices and interpret them in ways similar to adult listeners. It is important to emphasize that the children's identification accuracy differed greatly as a function of vowel category and some vowels were more easily confused for other vowels than others. Several vowels, /i, e, a, ɔ, ai, u/, were

especially prone to confusions and their overall identification as intended by the talkers was at or below 70%. Except for /u/, these vowels were also those most often confused by adults in Jacewicz and Fox (2012).

Although generally children were consistent with adults in making identification choices, we can indirectly infer from the confusion matrices that some of their responses may have been due to orthographic interference or inability to assign an orthographic form to match a particular pronunciation of the word they heard. This could be particularly true for some low frequency words labeled *hawed*, *hoed*, *heyd* which, on average, were misidentified more often by children than by adults. However, similar discrepancies were found also for some of the high frequency words *hid* and *head* which should not have been as problematic as the low frequency ones. Still some of the responses may have simply been due to “guessing” or lack of attention. At this point, we need to clarify that these children were fluent readers and participated in another experiment which required fluent reading (without hesitations) an extended set of sentences. Thus, there is no immediately obvious reason why orthographic interference should have large effects on reading isolated items displayed in the present perception task. Rather, the most compelling account for the differences between the performance of children and adults lies in the less efficient encoding of acoustic-phonetic information in the speech of multiple talkers and children’s relative inexperience with dialect variation compared with adults.

In our view, generational and dialect-related variations in vowel production contributed most to the greater ambiguity of acoustic information in the speech signal. While we do not rule out that some of the confusions were related to other factors such as children’s shorter attention span or boredom, the contribution of these factors to the overall identification pattern does not appear substantial. To inquire into this possibility, we examined the identification patterns of the

often confused vowel /ɔ/ whose accuracy was particularly low in both children (41.4%) and adults in Jacewicz & Fox (2012) (54.3%). We expected the vowel to be predominantly confused with /ɑ/ and, to a much lesser extent, with any back vowel and perhaps with the front /æ/ due to its spectral proximity in the low region of the vowel space. However, confusions with any of the front vowels /i, ɪ, e, ε/ and /ɜ/ cannot be easily explained on the basis of acoustic similarity and should be thus attributed to other factors affecting identification decisions, including attention, guessing, effects of stimulus order, fatigue, boredom or apathy. For adults, confusions of /ɔ/ with the above vowels totaled 1.4% for NC listeners (a total of 13 responses out of 900) and 1.3% for WI listeners (12 out of 900) whereas for children, the percentages were only slightly higher: 3.5% for NC listeners (25 out of 720) and 2.4% for WI listeners (19 out of 780). The higher number of unexplained confusions in children relative to adults suggests that some of their identification choices were indeed unrelated to the stimulus characteristics. However, confusions of this type were infrequent compared with confusions which can be attributed to acoustic properties of individual vowel categories.

We wish to emphasize that the current design did not allow for a direct assessment of the developmental aspect of the basic perceptual abilities of older children compared with adults. Administering the current task, our interest was in older children's sensitivity to fine phonetic details and their abilities to associate these details with talker characteristics. The results demonstrate that older children can successfully use their linguistic and indexical knowledge in dealing with extensive talker variability when listening to isolated words, but their overall performance is significantly worse compared with adults.

Two additional findings from the current study deserve consideration. First, the significant main effect of talker dialect revealed that NC vowels were identified with higher accuracy than WI vowels. This would suggest that the intelligibility of Southern English is greater than that of Midwestern English for both groups of listeners. Although the main effect of talker dialect was not as strong as the interaction between the listener dialect and talker dialect which indicated the native dialect advantage, we cannot ignore the fact that, on average, the Southern vowels produced higher identification rates than the Midwestern vowels. This finding is rather surprising because Southern English tends to be perceptually more distinct from the general American variety than Midwestern English (Clopper & Bradlow, 2008). One account of this effect is that the current NC vowels were on average longer than WI vowels and the slower articulation rate enabled additional processing time. However, this reasoning may not hold true because, by the same token, the vowel /ɔ/ had the longest duration in the set but produced the highest number of confusions. We have no explanation of this curious result at this time, suggesting only that both groups of children found the combination of longer vowel durations and some yet undetermined spectral characteristics of NC vowels as providing clearer category exemplars. At the same time, we point out that the main effect of talker dialect was not significant for adults in Jacewicz and Fox (2012).

Second, the significance of the main effect of talker gender invites a question of which characteristics of female voices contributed most to the greater intelligibility of vowels in their productions. We can again invoke vowel duration as one possible source of greater intelligibility because their vowels were longer than male, which would imply greater clarity of their productions as suggested elsewhere (Henton, 1995; Whiteside, 1996). However, this explanation is not consistent with the finding that the vowels of older talkers were also longer than those by

young adults and children and yet their productions were identified with a significantly lower accuracy compared to the two other groups. Further research is warranted to explore these gender-related effects in vowel recognition and speech perception in general. At the very least, the current results provide evidence that older children are sensitive to gender-related differences in vowel production as their accuracy rates are higher in identifying vowels produced by females.

In summary, the results of the current study demonstrate that older children are generally successful in dealing with extensive talker variability in a traditional vowel recognition experiment. Our previous work showed that acoustic phonetic and indexical variation related to talker dialect, gender and generation produces reliable differences in the perception of adult listeners. The present results provide further evidence that sensitivity to indexical talker characteristics is shaped in childhood. The fact that older children are less consistent than the adults in making their perceptual decisions most likely reflects their less efficient encoding of phonetic characteristics produced by multiple talkers, which creates greater ambiguity of acoustic information. The reduced encoding efficiency may be in part related to children's lesser experience with regional variation in speech and dialect-specific internal vowel structure, including systematic variation in formant dynamics. The contribution of other factors such as children's shorter attention span, while acknowledged, seems to be of comparatively lesser importance. Further research is needed to determine whether there are still some developmental aspects of the basic perceptual abilities in older children that also affect their vowel recognition performance in addition to the indexical sources found in this study.

Acknowledgments

This publication was made possible by Grant Number R01 DC006871 from the National Institute of Deafness and Other Communication Disorders, National Institutes of Health. Its contents are solely the responsibility of the authors and do not necessarily represent the official views of NIH. We thank Joseph Salmons for his contributions to this research. Many thanks go to Janaye Houghton and Dilara Tepeli for help with data collection in North Carolina and Wisconsin, respectively, and to Leigh Smitley for help with data management and processing.

References

- Abercrombie, D. (1967). *Elements of general phonetics*. Edinburgh University Press.
- Bergeson, T. R. & Trehub, S. E. (2007). Signature tunes in mothers' speech to infants. *Infant Behavior & Development, 30*, 648-654.
- Bradlow, A. R., Toretta, G. M., & Pisoni, D. B. (1996). Intelligibility of normal speech I: Global and fine-grained acoustic-phonetic talker characteristics. *Speech Communication, 20*, 255-272.
- Chambers, J. K. (1992). Dialect acquisition. *Language, 68*, 673-705.
- Chambers, J. K. (2003). *Sociolinguistic theory: Linguistic variation and its social significance* (2nd ed.). Oxford, United Kingdom: Blackwell.
- Clopper, C. G., & Bradlow, A. (2008). Perception of dialect variation in noise: Intelligibility and classification. *Language and Speech, 51*, 175-198.
- Dinkin, A. J. (2011). Weakening resistance: Progress toward the low back merger in New York State. *Language Variation and Change, 23*, 315-345.
- Eisenberg, L. S., Shannon, R. V., Schaefer Martinez, A., Wygonsky, J. & Boothroyd, A. (2000). Speech recognition with reduced spectral cues as a function of age. *The Journal of the Acoustical Society of America, 107*, 2704-2710.
- Ferguson, S. H. (2004). Talker differences in clear and conversational speech: Vowel intelligibility for normal-hearing listeners. *The Journal of the Acoustical Society of America, 116*, 2365-2373.
- Flege, J. E. & Eefting, W. (1986) Linguistic and developmental effects on the production and perception of stop consonants. *Phonetica, 43*, 155-171.
- Floccia, C., Butler, J., Girard, F. & Goslin, J. (2009). Categorization of regional and foreign accent in 5- to 7-year-old British children. *International Journal of Behavioral Development, 33*, 366-375.
- Fox, R. A., & Jacewicz, E. (2009). Cross-dialectal variation in formant dynamics of American English vowels. *The Journal of the Acoustical Society of America, 126*, 2603-2618.
- Gerken, L. A., & Aslin, R. N. (2005). Thirty years of research on infant speech perception: The legacy of Peter W. Jusczyk. *Language Learning and Development, 1*, 5-21.
- Girard, F., Floccia, C., & Goslin, J. (2008). Perception and awareness of accents in young children. *British Journal of Developmental Psychology, 26*, 409-433.

- Hazan, V. & Barrett, S. (2000). The development of phonemic categorization in children aged 6-12. *Journal of Phonetics*, 28, 377-396.
- Hazan, V., & Markham, D. (2004). Acoustic-phonetic correlates of talker intelligibility for adults and children. *Journal of the Acoustical Society of America*, 116, 3108-3118.
- Henton, C. (1995). Cross-language variation in the vowels of female and male speakers. *Proceedings of the 13th Congress of Phonetic Sciences* (pp. 420-423). Stockholm, Sweden.
- Hillenbrand, J. M., Getty, L. A., Clark, M. J., & Wheeler, K. (1995). Acoustic characteristics of American English vowels. *The Journal of the Acoustical Society of America*, 97, 3099–3111.
- Irons, T. L. (2007). On the status of low back vowels in Kentucky English: More evidence of merger. *Language Variation and Change*, 19, 137-180.
- Jacewicz, E. & Fox, R. A. (2012). The effects of cross-generational and cross-dialectal variation on vowel identification and classification. *The Journal of the Acoustical Society of America*, 131, 1413–1433.
- Jacewicz, E., Fox, R. A., & Salmons, J. (2011a). Regional dialect variation in the vowel systems of typically developing children. *Journal of Speech, Language, and Hearing Research*, 54, 448-470.
- Jacewicz, E., Fox, R. A., & Salmons, J. (2011b). Cross-generational vowel change in American English. *Language Variation and Change*, 23, 45-86.
- Jacewicz, E., Fox, R. A., & Salmons, J. (2011c). Vowel change across three age groups of speakers in three regional varieties of American English. *Journal of Phonetics*, 39, 683–693.
- Johnson, C. E. (2000). Children's phoneme identification in reverberation and noise. *Journal of Speech, Language and Hearing Research*, 43, 144-157.
- Jusczyk, P. W. (1997). *The discovery of spoken language*. Cambridge, MA: MIT Press.
- Kerswill, P. (1996). Children, adolescents, and language change. *Language Variation and Change*, 8, 177–202.
- Kinzler, K.D., Corriveau, K.H., & Harris, P.L (2011). Children's selective trust in native-accented speakers. *Developmental Science*, 14, 106-111.

- Kinzler, K.D., Dupoux, E., & Spelke, E.S. (2007). The native language of social cognition. *Proceedings of the National Academy of Sciences of the United States of America*, 104, 12577–12580.
- Kinzler, K.D., Shutts, K., DeJesus, J., & Spelke, E.S. (2009). Accent trumps race in children's social preferences. *Social Cognition*, 27, 623–634.
- Kuhl, P. K., Stevens, E., Hayashi, A., Deguchi, T., Kiritani, S. & Iverson, P. (2006). Infants show a facilitation effect for native language phonetic perception between 6 and 12 months. *Developmental Science*, 9, F13-F21.
- Labov, W. (1994). *Principles of linguistic change. Volume 1: Internal factors*. Oxford, United Kingdom: Blackwell.
- Labov, W. (2010). *Principles of linguistic change. Volume 3: Cognitive and cultural factors*. Oxford, United Kingdom: Blackwell.
- Labov, W., Ash, S., & Boberg, C. (2006). *Atlas of North American English: Phonetics, phonology, and sound change*. Berlin, Germany: Mouton de Gruyter.
- Lobanov, B. (1971). Classification of Russian vowels spoken by different speakers. *The Journal of the Acoustical Society of America*, 49, 606–608.
- Mehler, J., Bertoncini, J., Barriere, M., & Jassik-Gerschenfeld, D. (1978). Infant recognition of mother's voice. *Perception*, 7, 491–497.
- Moon, C., Lagercrantz, H., & Kuhl, P.K. (2013). Language experienced in utero affects vowel perception after birth: A two-country study. *Acta Paediatrica*, 102, 156-160.
- Mullennix, J. W., Pisoni, D. B., & Martin, C. S. (1989). Some effects of talker variability on spoken word recognition. *The Journal of the Acoustical Society of America*, 85, 365–378.
- Neel, A. (2008). Vowel space characteristics and vowel identification accuracy. *Journal of Speech, Language, and Hearing Research*, 51, 574–585.
- Nittrouer, S. (1996). Discriminability and perceptual weighting of some acoustic cues to speech perception by 3-year-olds. *Journal of Speech and Hearing Research*, 39, 278–297.
- Nittrouer, S. (2007). Dynamic spectral structure specifies vowels for children and adults. *The Journal of the Acoustical Society of America*, 122, 2328–2339
- Ohde, R. N., & German, S. R. (2011). Formant onsets and formant transitions as developmental cues to vowel perception. *The Journal of the Acoustical Society of America*, 130, 1628–1642

- Ohde, R. N., & Haley, K. L. (1997). Stop consonant and vowel perception in 3- and 4-year-old children. *The Journal of the Acoustical Society of America*, *102*, 3711–3722.
- Parnell, M. M., & Amerman, J. (1978). Maturation influences on perception of coarticulatory effects. *Journal of Speech and Hearing Research*, *21*, 682–701.
- Payne, A. (1980). Factors controlling the acquisition of the Philadelphia dialect by out of-state children. In W. Labov (Ed.), *Locating language in time and space* (pp. 143–178). New York, NY: Academic Press.
- Peterson, G. E., & Barney, H. L. (1952). Control methods used in a study of the vowels. *The Journal of the Acoustical Society of America*, *24*, 175–184.
- Shutts, K., Kinzler, K.D., McKee, C.B., & Spelke, E.S. (2009). Social information guides infants' selection of foods. *Journal of Cognition and Development*, *10*, 1–17.
- Studebaker, G. A. (1985). A “rationalized” arcsine transform. *Journal of Speech and Hearing Research*, *28*, 455-462.
- Sundara, M., Polka, L. & Genesee, F. (2006). Language experience facilitates discrimination of /d-ð/ in monolingual and bilingual acquisition of English. *Cognition*, *100*, 369-388.
- Sussman, J. E. (2001). Vowel perception by adults and children with normal language and specific language impairment: Based on steady states or transitions. *The Journal of the Acoustical Society of America*, *109*, 1173–1180.
- Walley, A. C. (1993). The role of vocabulary growth in children's spoken word recognition and segmentation ability. *Developmental Review*, *13*, 286-350.
- Walley, A. C. (2005). Speech perception in childhood. In D. B. Pisoni, & R. E. Remez, (Eds.), *The handbook of speech perception* (pp. 449-468). Malden, MA: Blackwell Publishing.
- Walley, A. C. & Flege, J. E. (1999). Effects of lexical status on the perception of native and nonnative vowels: A developmental study. *Journal of Phonetics*, *27*, 307-332.
- Whiteside, S.P. (1996). Temporal-based acoustic-phonetic patterns in read speech: Some evidence for speaker sex differences. *Journal of the International Phonetic Association*, *26*, 23-40.

Appendix captions

Appendix 1

Identification rates and confusion matrix for North Carolina listeners.

Appendix 2

Identification rates and confusion matrix for Wisconsin listeners.

Appendix 1: Identification rates and confusion matrix for North Carolina listeners.

Speaker gender	Vowel intended by speaker	Vowel identified by listeners												
		/i/	/ɪ/	/e/	/ɛ/	/æ/	/ɑ/	/ɔ/	/ɜː/	/aɪ/	/o/	/ʊ/	/u/	
(a) North Carolina listeners: Responses to North Carolina GP speakers														
Male	/i/	56.7	6.7	30.0	6.7	
	/ɪ/	...	56.7	...	35.0	3.3	3.3	1.7	
	/e/	3.3	...	85.0	3.3	1.7	1.7	1.7	1.7	1.7	
	/ɛ/	...	8.3	...	83.3	5.0	1.7	1.7	
	/æ/	...	1.7	1.7	8.3	86.7	1.7	
	/ɑ/	...	1.7	1.7	...	5.0	58.3	26.7	...	5.0	1.7	
	/ɔ/	...	3.3	5.0	36.7	43.3	...	5.0	3.3	1.7	1.7	
	/ɜː/	98.3	1.7
	/aɪ/	1.7	5.0	35.0	58.3
	/o/	3.3	3.3	86.7	...	6.7	
	/ʊ/	1.7	...	1.7	91.7	5.0	
	/u/	3.3	5.0	6.7	10.0	75.0	
Female	/i/	73.3	8.3	11.7	3.3	3.3	
	/ɪ/	1.7	51.7	...	40.0	1.7	3.3	...	1.7	...	
	/e/	8.3	1.7	80.0	3.3	3.3	...	1.7	1.7	
	/ɛ/	1.7	5.0	...	86.7	3.3	1.7	1.7	
	/æ/	3.3	93.3	1.7	1.7	...	
	/ɑ/	1.7	...	13.3	50.0	18.3	...	16.7	
	/ɔ/	16.7	70.0	3.3	1.7	5.0	...	3.3	
	/ɜː/	1.7	96.7	1.7	
	/aɪ/	...	1.7	3.3	...	35.0	3.3	3.3	...	53.3	
	/o/	1.7	1.7	93.3	1.7	1.7	
	/ʊ/	3.3	1.7	91.7	3.3	
	/u/	1.7	...	1.7	...	3.3	10.0	16.7	66.7	

(b) North Carolina listeners: Responses to North Carolina P speakers

Male	/i/	71.7	8.3	8.3	8.3	3.3
	/ɪ/												

	/e/	6.7	...	80.0	3.3	1.7	1.7	5.0	...	1.7	...
	/ɛ/	1.7	10.0	...	83.3	3.3	1.7	...
	/æ/	1.7	91.7	1.7	3.3	...	1.7	...
	/ɑ/	...	1.7	6.7	51.7	30.0	...	1.7	3.3	...	5.0
	/ɔ/	...	1.7	38.3	46.7	...	1.7	8.3	...	3.3
	/ɜː/	98.3	1.7
	/aɪ/	...	3.3	15.0	15.0	3.3	...	61.7	1.7
	/o/	1.7	3.3	...	1.7	...	85.0	1.7	6.7
	/u/	1.7	1.7	95.0	1.7
Female	/u/	1.7	1.7	6.7	5.0	16.7	68.3
	/i/	78.3	1.7	5.0	8.3	1.7	...	1.7	...	1.7	1.7
	/ɪ/	1.7	68.3	...	25.0	1.7	1.7	1.7
	/e/	1.7	3.3	88.3	...	1.7	...	1.7	...	1.7	1.7
	/ɛ/	93.3	6.7
	/æ/	...	1.7	98.3
	/ɑ/	...	1.7	8.3	50.0	21.7	...	15.0	1.7	...	1.7
	/ɔ/	1.7	20.0	68.3	...	1.7	5.0	...	3.3
	/ɜː/	100.0
	/aɪ/	...	5.0	...	1.7	16.7	3.3	3.3	...	70.0
	/o/	6.7	88.3	1.7	3.3
	/u/	3.3	...	3.3	3.3	1.7	85.0	3.3
	/u/	3.3	6.7	1.7	5.0	83.3

(d) North Carolina listeners: Responses to North Carolina C speakers

Male	/i/	81.7	5.0	1.7	10.0	1.7
	/ɪ/	1.7	70.0	...	15.0	8.3	...	5.0	...
	/e/	3.3	1.7	76.7	3.3	5.0	...	1.7	3.3	3.3	1.7
	/ɛ/	...	15.0	...	76.7	...	1.7	1.7	5.0	...
	/æ/	1.7	1.7	93.3	3.3
	/ɑ/	1.7	...	3.3	63.3	25.0	...	5.0	1.7
	/ɔ/	3.3	...	5.0	...	6.7	35.0	41.7	3.3	...	5.0
	/ɜː/	1.7	98.3
	/aɪ/	1.7	3.3	1.7	...	13.3	13.3	13.3	...	50.0	3.3
	/o/	1.7	5.0	1.7	83.3	1.7	6.7
	/u/

Female	/u/	...	1.7	3.3	5.0	1.7	1.7	86.7
	/i/	81.7	8.3	1.7	6.7	1.7
	/ɪ/	1.7	65.0	...	30.0	3.3
	/e/	3.3	...	86.7	1.7	...	1.7	...	1.7	5.0
	/ɛ/	...	3.3	1.7	86.7	6.7	1.7
	/æ/	1.7	1.7	3.3	...	90.0	3.3
	/ɑ/	...	1.7	1.7	1.7	5.0	36.7	25.0	...	26.7	1.7
	/ɔ/	1.7	...	8.3	25.0	55.0	...	6.7	3.3
	/ɜː/	100.0
	/aɪ/	...	5.0	3.3	...	8.3	3.3	80.0
	/o/	3.3	88.3	1.7	6.7
	/ʊ/	3.3	93.3	3.3
/u/	1.7	3.3	3.3	6.7	85.0	

(e) North Carolina listeners: Responses to Wisconsin GP speakers

Male	/i/	81.7	8.3	1.7	6.7	1.7
	/ɪ/	...	58.3	...	33.3	3.3	1.7	3.3	...
	/e/	13.3	8.3	43.3	25.0	6.7	1.7	1.7
	/ɛ/	...	8.3	...	86.7	1.7	1.7	1.7	...
	/æ/	1.7	1.7	...	25.0	65.0	...	1.7	1.7	1.7	1.7
	/ɑ/	...	1.7	3.3	...	11.7	38.3	13.3	...	31.7
	/ɔ/	3.3	...	10.0	46.7	30.0	...	6.7	1.7	1.7	...
	/ɜː/	1.7	96.7	1.7	...
	/aɪ/	1.7	1.7	13.3	1.7	1.7	...	78.3	1.7
	/o/	5.0	1.7	1.7	5.0	3.3	3.3	70.0	3.3	6.7
Female	/u/	1.7	10.0	1.7	8.3	28.3	50.0
	/i/	76.7	11.7	...	8.3	1.7	1.7	...
	/ɪ/	...	90.0	...	6.7	3.3
	/e/	43.3	6.7	23.3	21.7	1.7	1.7	...	1.7
	/ɛ/	1.7	3.3	...	93.3	1.7
	/æ/	...	1.7	...	38.3	58.3	1.7
	/ɑ/	...	6.7	10.0	41.7	11.7	...	25.0	1.7	...	3.3
	/ɔ/	...	1.7	1.7	1.7	3.3	48.3	41.7	...	1.7
	/ɜː/

/aɪ/	1.7	3.3	13.3	1.7	3.3	76.7
/o/	1.7	...	1.7	5.0	5.0	1.7	...	76.7	1.7	6.7
/ʊ/	1.7	...	1.7	1.7	95.0	...
/u/	3.3	3.3	6.7	21.7	65.0

(f) North Carolina listeners: Responses to Wisconsin P speakers

Male	/i/	76.7	11.7	1.7	3.3	...	3.3	1.7	...	1.7
	/ɪ/	...	55.0	...	45.0
	/e/	16.7	3.3	30.0	48.3	1.7
	/ɛ/	...	1.7	...	51.7	43.3	...	1.7	...	1.7
	/æ/	8.3	90.0	...	1.7
	/ɑ/	...	6.7	16.7	10.0	1.7	...	65.0
	/ɔ/	...	1.7	5.0	55.0	33.3	...	3.3	...	1.7	...
	/ɜ:/	1.7	96.7	1.7	...
	/aɪ/	...	13.3	10.0	...	1.7	...	1.7	...	73.3
	/o/	10.0	1.7	1.7	...	78.3	3.3	5.0
Female	/ʊ/	96.7	3.3	3.3
	/u/	5.0	1.7	3.3	25.0	65.0	
	/i/	71.7	20.0	...	3.3	1.7	1.7	1.7	...
	/ɪ/	...	78.3	...	15.0	1.7	5.0	
	/e/	35.0	10.0	35.0	11.7	...	3.3	3.3	1.7	...	
	/ɛ/	73.3	23.3	1.7	...	1.7	
	/æ/	18.3	80.0	1.7	
	/ɑ/	...	3.3	1.7	...	13.3	5.0	8.3	...	68.3	
	/ɔ/	...	1.7	1.7	...	1.7	61.7	26.7	...	3.3	3.3	...	
	/ɜ:/	1.7	...	1.7	1.7	95.0	
/aɪ/	1.7	3.3	10.0	1.7	3.3	...	3.3	...	76.7		
/o/	3.3	5.0	5.0	1.7	...	81.7	...	3.3	
/ʊ/	...	1.7	3.3	93.3	1.7	
/u/	5.0	5.0	5.0	23.3	61.7	

(g) North Carolina listeners: Responses to Wisconsin C speakers

Male	/i/	83.3	8.3	1.7	5.0	1.7
	/ɪ/	5.0	48.3	...	28.2	1.7	...	6.7	...

	/e/	21.7	3.3	61.7	8.3	1.7	1.7	1.7
	/ɛ/	1.7	48.3	48.3	1.7
	/æ/	5.0	91.7	1.7	1.7	...
	/ɑ/	1.7	5.0	11.7	38.3	8.3	...	33.3	...	1.7	...
	/ɔ/	...	1.7	6.7	50.0	21.7	1.7	16.7	1.7
	/ɜː/	1.7	96.7	1.7
	/aɪ/	...	8.3	16.7	1.7	1.7	...	1.7	1.7	68.3
	/o/	6.7	1.7	90.0	...	1.7
	/u/	...	1.7	3.3	...	3.3	91.7	...
	/ʊ/	...	1.7	1.7	6.7	1.7	5.0	6.7	76.7
Female	/i/	86.7	5.0	...	6.7	1.7
	/ɪ/	1.7	45.0	...	48.3	1.7	1.7	...	1.7	...
	/e/	38.3	3.3	41.7	11.7	1.7	1.7	1.7
	/ɛ/	1.7	1.7	1.7	33.3	60.0	1.7
	/æ/	8.3	90.0	1.7
	/ɑ/	1.7	3.3	1.7	...	23.3	10.0	1.7	...	58.3
	/ɔ/	3.3	...	5.0	58.3	21.7	3.3	6.7	1.7
	/ɜː/	1.7	...	96.7	1.7	...
	/aɪ/	...	3.3	8.3	3.3	1.7	1.7	81.7
	/o/	1.7	10.0	11.7	73.3	...	3.3
	/u/	3.3	6.7	...	3.3	...	1.7	83.3	1.7
	/ʊ/	1.7	...	1.7	1.7	8.3	6.7	11.7	68.3

Appendix 2: Identification rates and confusion matrix for Wisconsin listeners.

Speaker gender	Vowel intended by speaker	Vowel identified by listeners											
		/ɪ/	/ɪ/	/ɪ/	/ɪ/	/ɪ/	/ɪ/	/ɪ/	/ɪ/	/ɪ/	/ɪ/	/ɪ/	/ɪ/
(a) Wisconsin listeners: Responses to Wisconsin GP speakers													
Male	ɪ	89.2	4.6	...	1.5	4.6	...
	ɪ	1.5	64.6	...	27.7	1.5	1.5	3.1
	ɪ	10.8	10.8	44.6	29.2	4.6
	ɪ	1.5	96.9	1.5	...
	ɪ	26.2	72.3	1.5	...
	/ɪ/	1.5	3.1	13.8	50.8	12.3	1.5	16.9
	/ɪ/	3.1	10.8	55.4	27.7	...	3.1
	/ɪ/	1.5	1.5	...	1.5	93.8	...	1.5
	ɪ	...	4.6	12.3	1.5	80.0	1.5
	ɪ	1.5	...	4.6	1.5	89.2	3.1	...
Female	ɪ	1.5	1.5	1.5	90.8	4.6
	/ɪ/	1.5	4.6	4.6	32.3	56.9
	ɪ	89.2	1.5	1.5	3.1	4.6	...
	ɪ	...	86.2	...	4.6	...	1.5	4.6	...	3.1	...
	ɪ	27.7	3.1	46.2	15.4	7.7
	ɪ	...	1.5	...	93.8	1.5	1.5	...	1.5
	ɪ	3.1	6.2	90.8
	/ɪ/	1.5	...	1.5	1.5	18.5	40.0	12.3	...	24.6
	/ɪ/	1.5	7.7	41.5	46.2	1.5	1.5	...
	/ɪ/	1.5	98.5
ɪ	1.5	...	4.6	3.1	1.5	4.6	84.6	
ɪ	3.1	...	1.5	7.7	3.1	80.0	1.5	3.1	
ɪ	...	1.5	1.5	96.9	...	
/ɪ/	...	1.5	1.5	...	4.6	3.1	1.5	23.1	64.6	

(b) Wisconsin listeners: Responses to Wisconsin P speakers

Male	ɪ	93.8	1.5	1.5	1.5
	ɪ	...	66.2	1.5	21.5	1.5	7.7	...	1.5	...

		10.8	7.7	41.5	38.5	1.5
		...	4.6	...	73.8	20.0	1.5
		16.9	80.0	1.5	1.5
	/ /	1.5	26.2	36.9	4.6	...	30.8
	/ /	9.2	53.8	33.8	1.5	1.5	...
	/ /	1.5	3.1	93.8	1.5
		...	6.2	1.5	1.5	...	1.5	89.2
		1.5	1.5	...	3.1	84.6	6.2	3.1
		3.1	95.4	1.5
	/ /	1.5	7.7	26.2	64.6
Female		93.8	1.5	...	3.1	1.5
		1.5	72.3	...	16.9	3.1	1.5	...	3.1	1.5
		21.5	6.2	52.3	12.3	3.1	1.5	1.5	1.5
		1.5	84.6	13.8
		1.5	93.8	4.6
	/ /	...	3.1	1.5	...	18.5	30.8	4.6	...	41.5
	/ /	1.5	...	10.8	55.4	30.8	1.5
	/ /	1.5	1.5	1.5	93.8	...	1.5
		...	3.1	3.1	...	3.1	90.8
		3.1	1.5	1.5	4.6	81.5	...	7.7
		...	1.5	...	1.5	1.5	...	1.5	92.3	1.5
	/ /	6.2	20.0	73.8

(c) Wisconsin listeners: Responses to Wisconsin C speakers

Male		95.4	1.5	1.5	1.5
		1.5	66.2	...	24.6	1.5	3.1	...	3.1	...
		13.8	4.6	64.6	12.3	1.5	3.1	...
		6.2	66.2	23.1	3.1	...	1.5
		1.5	...	90.8	1.5	...	1.5	4.6	...
	/ /	1.5	1.5	12.3	47.7	21.5	...	15.4
	/ /	1.5	7.7	46.2	33.8	1.5	9.2
	/ /	...	1.5	92.3	1.5	4.6
		...	1.5	6.2	1.5	1.5	87.7	1.5
		...	1.5	3.1	90.8	3.1	1.5
		1.5	3.1	...	4.6	...	3.1	86.2	1.5

Female	/ɪ/	1.5	1.5	1.5	1.5	23.1	70.8
	■■■	92.3	1.5	...	3.1	1.5	1.5
	■■■	...	70.8	...	23.1	1.5	1.5	1.5	1.5
	■■■	12.3	1.5	63.1	16.9	1.5	1.5	1.5	1.5
	■■■	...	1.5	...	92.3	6.2
	■■■	...	1.5	1.5	1.5	92.3	1.5	1.5
	/ɪ/	1.5	1.5	23.1	29.2	1.5	...	43.1
	/ɪ/	...	1.5	10.8	52.3	29.2	3.1	3.1	...
	/ɪɪ/	1.5	98.5
	■■■■	3.1	...	3.1	93.8
	■■■	1.5	...	1.5	1.5	1.5	1.5	...	87.7	...	4.6
	■■■	1.5	...	3.1	...	1.5	...	3.1	89.2	1.5
	/ɪ/	6.2	24.6	69.2

(d) Wisconsin listeners: Responses to North Carolina GP speakers

Male	■■■	55.4	1.5	26.2	10.8	3.1	1.5	1.5
	■■■	...	60.0	...	33.8	1.5	3.1	1.5
	■■■	1.5	...	76.9	10.8	1.5	...	3.1	...	3.1	...	1.5	1.5
	■■■	...	6.2	6.2	72.3	12.3	1.5	1.5	...
	■■■	3.1	20.0	73.8	1.5	...	1.5	...
	/ɪ/	12.3	58.5	29.2
	/ɪ/	4.6	7.7	38.5	41.5	...	3.1	1.5	1.5	1.5
	/ɪɪ/	1.5	...	1.5	95.4	1.5	...
	■■■■	...	3.1	4.6	6.2	35.4	3.1	1.5	...	46.2
	■■■	...	1.5	3.1	1.5	89.2	3.1	1.5
Female	■■■	1.5	...	3.1	1.5	1.5	92.3	...
	/ɪ/	1.5	3.1	1.5	4.6	20.0	24.6	44.6
	■■■	75.4	3.1	12.3	6.2	1.5	1.5
	■■■	4.6	36.9	3.1	50.8	1.5	3.1
	■■■	4.6	...	58.5	12.3	7.7	1.5	3.1	1.5	6.2	1.5	1.5	1.5
	■■■	3.1	78.5	15.4	1.5	...	1.5
	■■■	...	1.5	...	3.1	92.3	3.1	...
	/ɪ/	1.5	...	15.4	50.8	24.6	...	6.2	...	1.5	...
	/ɪ/	1.5	3.1	16.9	56.9	1.5	1.5	9.2	1.5	7.7
	/ɪɪ/	4.6	1.5	90.8	...	3.1

	...	1.5	...	1.5	40.0	9.2	1.5	...	46.2
	3.1	...	7.7	78.5	6.2	4.6
	4.6	...	1.5	92.3	1.5
/ /	1.5	1.5	1.5	35.4	20.0	40.0

(e) Wisconsin listeners: Responses to North Carolina P speakers

Male		73.8	3.1	15.4	4.6	1.5	1.5
		...	58.5	1.5	29.2	4.6	3.1	...	3.1	...
		4.6	...	61.5	18.5	3.1	3.1	9.2
		1.5	12.3	1.5	72.3	6.2	...	1.5	...	1.5	...	1.5	1.5
		1.5	4.6	93.8
	/ /	10.8	52.3	33.8	1.5	1.5	...
	/ /	...	1.5	4.6	32.3	44.6	1.5	...	9.2	...	6.2
	/ /	3.1	1.5	93.8	1.5
		...	3.1	13.8	21.5	3.1	...	56.9	1.5
		...	1.5	...	1.5	1.5	1.5	1.5	87.7	3.1	1.5
Female		1.5	...	1.5	1.5	90.8	4.6	
	/ /	1.5	1.5	4.6	1.5	1.5	9.2	63.1	
		90.8	...	3.1	1.5	1.5	1.5	1.5	
		...	44.6	1.5	40	1.5	4.6	3.1	...	4.6	
		78.5	10.8	6.2	3.1	...	1.5	
		3.1	76.9	13.8	3.1	...	1.5	...	1.5	...	
		1.5	98.5	
	/ /	1.5	15.4	52.3	26.2	1.5	3.1	
	/ /	1.5	...	4.6	16.9	61.5	1.5	...	6.2	...	7.7
	/ /	...	1.5	1.5	...	96.9	
	...	3.1	1.5	...	24.6	13.8	4.6	...	50.8	1.5	...		
	...	3.1	1.5	1.5	...	1.5	87.7	3.1	1.5	
	3.1	...	6.2	...	3.1	84.6	3.1	
/ /	1.5	...	1.5	1.5	1.5	12.3	16.9	64.6	

(f) Wisconsin listeners: Responses to North Carolina C speakers

Male		84.6	4.6	6.2	...	1.5	1.5	1.5
		1.5	67.7	...	23.1	1.5	1.5	...	4.6	...

		66.2	10.8	6.2	1.5	1.5	9.2	3.1	1.5
		...	13.8	1.5	76.9	3.1	1.5	1.5	1.5	...
		3.1	86.2	...	1.5	...	4.6	...	4.6	...
	/ /	1.5	1.5	4.6	40.0	46.2	1.5	3.1	...	1.5	...
	/ /	3.1	35.4	43.1	12.3	1.5	4.6
	/ /	1.5	...	96.9	...	1.5
		1.5	15.4	29.2	9.2	1.5	41.5	1.5
		1.5	...	1.5	1.5	83.1	7.7	4.6
		...	1.5	1.5	...	6.2	3.1	1.5	1.5	81.5	3.1
	/ /	...	1.5	1.5	3.1	7.7	27.7	58.5
Female		86.2	...	4.6	4.6	1.5	1.5	...	1.5
		...	66.2	...	24.6	4.6	3.1	...	1.5	...
		...	1.5	75.4	9.2	3.1	1.5	3.1	...	4.6	1.5
		...	7.7	...	86.2	4.6	1.5
		81.5	16.9	...	1.5	...
	/ /	1.5	1.5	...	3.1	6.2	41.5	27.7	...	18.5
	/ /	1.5	7.7	38.5	43.1	...	1.5	6.2	...	1.5
	/ /	1.5	4.6	1.5	92.3
		...	1.5	1.5	1.5	7.7	1.5	83.1	...	3.1	...
		1.5	1.5	3.1	89.2	4.6	...
		3.1	...	12.3	...	1.5	83.1	...
	/ /	9.2	21.5	69.2

Table I. Overall identification rates (in % correct) by vowel category for each NC and WI child listener group responding to both native and non-native dialect vowels. Results for adult listeners reported in Jacewicz & Fox (2012) are included for comparison.

Vowel intended by speaker	NC listeners		WI listeners		Total	Jacewicz & Fox (2012)
	Native dialect	Non-native dialect	Native dialect	Non-native dialect		
■■■	73.9	79.5	92.3	77.7	80.9	86.9
■■■	67.2	62.5	71.1	55.7	64.1	73.8
■■■	83.3	39.2	52.1	69.5	61.0	69.8
■■■	85.0	64.4	84.6	77.2	77.8	84.8
■■■	92.2	79.2	86.7	87.7	86.5	84.1
/ɪ/	51.7	23.9	39.2	49.2	41.0	57.9
/ɪ/	54.2	29.2	33.6	48.5	41.4	54.3
/iɪ/	98.6	96.1	95.1	94.4	96.1	98.7
■■■■	62.2	75.8	87.7	54.1	70.0	69.2
■■■	87.5	78.3	85.6	85.9	84.3	96.5
■■■	91.4	90.3	91.8	87.4	90.2	96.9
/ɪ/	77.5	64.5	66.7	56.7	66.4	89.9
Total	77.1	65.3	73.9	70.3	71.7	80.2

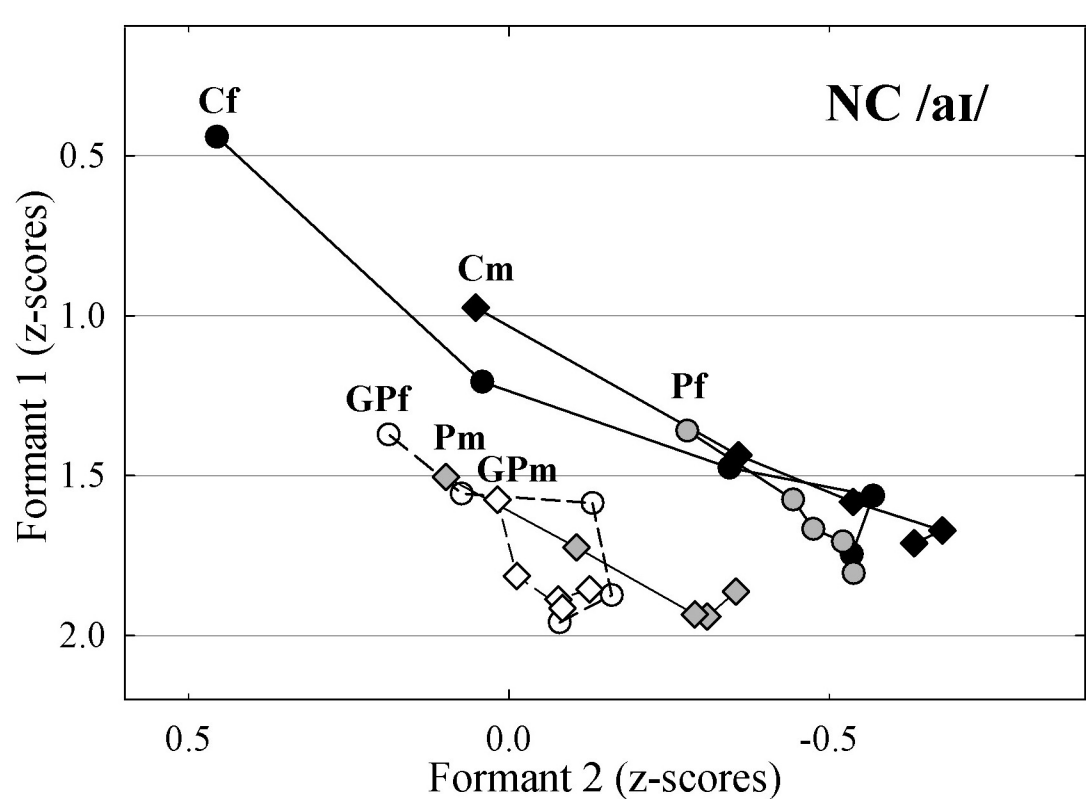
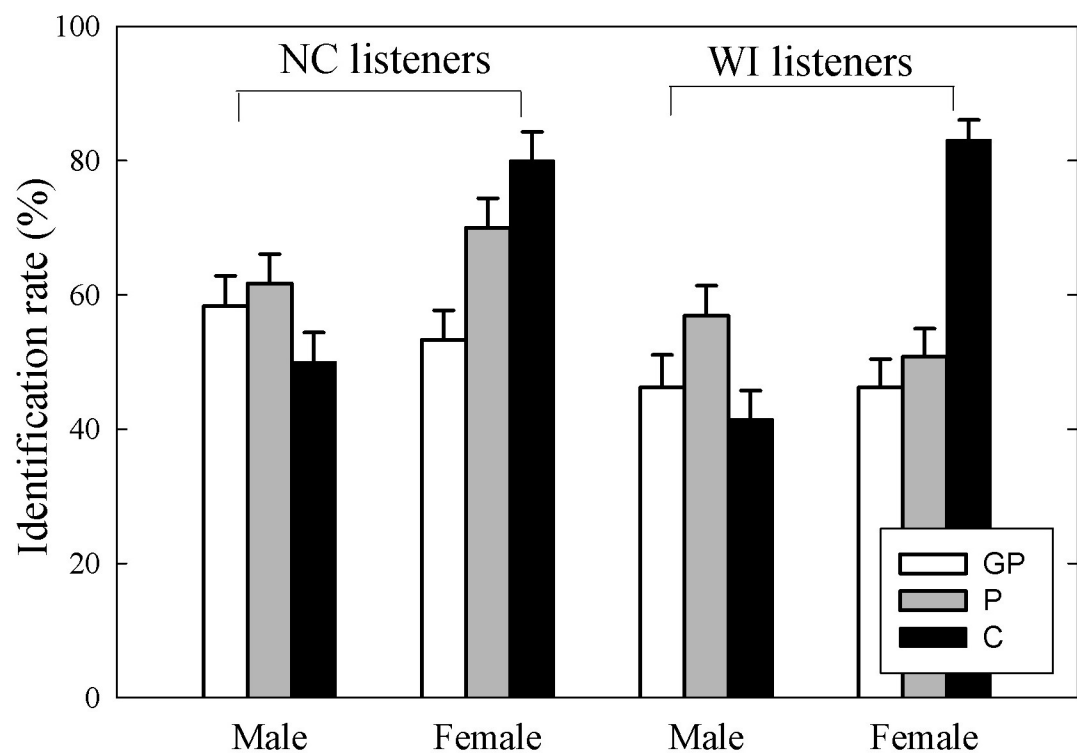
Figure captions

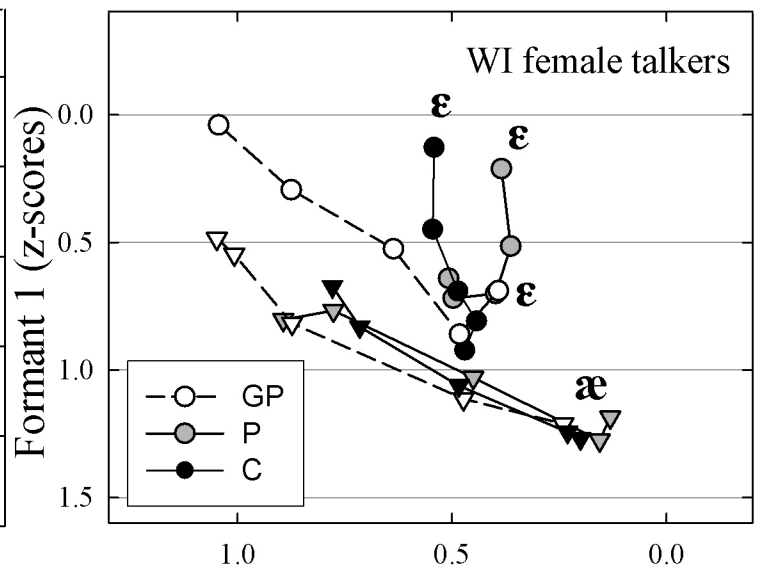
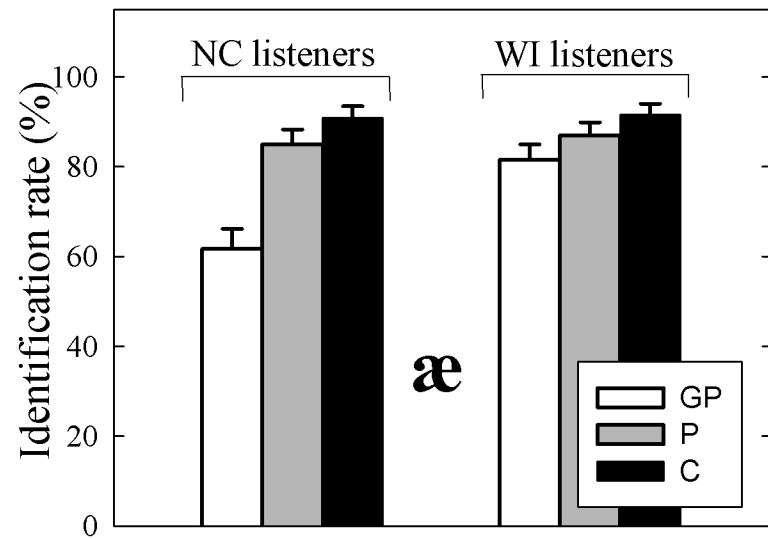
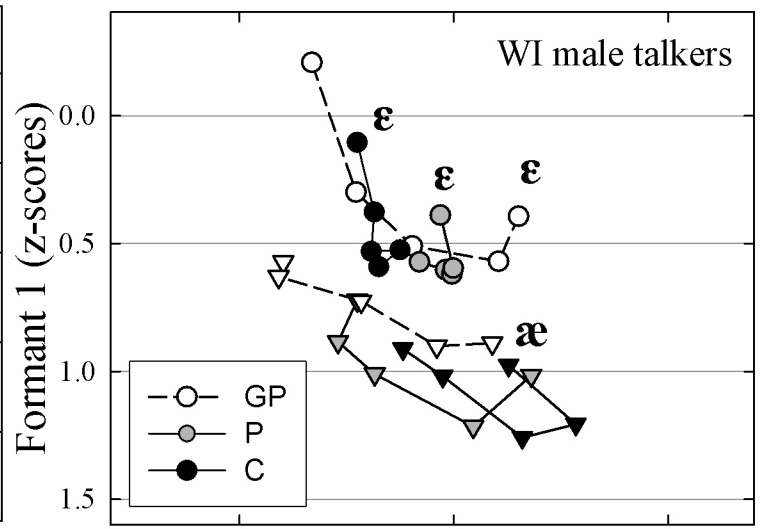
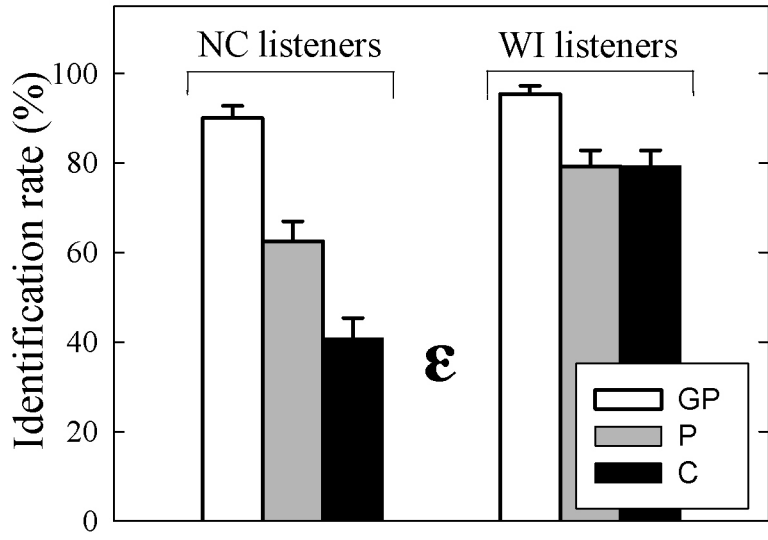
Figure 1

Left panel: Mean identification rates (with standard errors) for NC and WI listeners responding to exemplars of NC /h/ in *hide* split by talker gender. Right panel: Mean relative positions of NC /h/ in the acoustic vowel space produced by the talkers and their dynamic formant pattern (F1 and F2) measured at five equidistant time points (20-35-50-65-80%) over the course of the vowel's duration. The z-scores values reflect formant frequencies normalized for anatomical differences in the length of the vocal tract between males and females using the approach recommended by Lobanov (1971). The symbol for each talker group is placed close to the 80%-point in a vowel, indicating the direction of formant movement. GP = Grandparents, P = Parents, C = Children, m = male, f = female.

Figure 2

Left panels: Mean identification rates (with standard errors) for NC and WI listeners responding to exemplars of WI /h/ in *head* (top) and /h/ in *had* (bottom) collapsed across talker gender. Right panels: Mean relative positions of WI /h/, /h/ in the acoustic vowel space produced by the male and female talkers and their dynamic formant pattern sampled at five time points in the vowel.





Talkers

Formant 2 (z-scores)